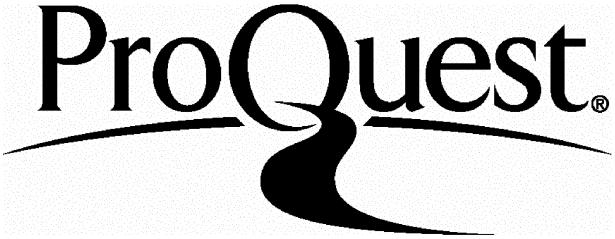
Structural Studies of factor B of the Alternative Pathway of the Complement System

Thesis Presented for the Degree of
Doctor of Philosophy
By

Justin Hinshelwood

Department of Biochemistry and Molecular Biology
Royal Free Campus
Royal Free and University College Medical School
University College London

May 2000



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

ABSTRACT

Factor B of the complement system is a five-domain, 90 kD serine protease proenzyme which is activated by the protease factor D in the context of the Mg^{2+}-dependent complex C3bB during complement activation. The cleavage of factor B into the Ba and Bb fragments during activation facilitates the assembly of the alternative pathway C3 convertase, C3bBb. The Ba fragment contains three short consensus repeat (SCR) domains, while the Bb fragment contains a von Willebrand factor type A (vWF-A) domain and a serine protease (SP) domain. All the domain types of factor B have been implicated in interactions with activated C3 and the assembly of the alternative pathway C3 convertase.

This thesis describes a number of structural and functional studies using six different preparations of factor B and its fragments and domains. Surface-enhanced laser desorption-ionization affinity mass spectrometry (SELDIAMS) was used to investigate the reaction of factor B with immobilised activated C3(NH_3) in the presence of Mg^{2+}. From this it was determined that (i) a major function of the vWF-A domain is to bind to activated C3 during the formation of the C3 convertase, which it does at its active site cleft, and that (ii) SELDIAMS provides an efficient means of identifying residues involved in protein-protein interactions. Next, ^{1}H NMR, FTIR and circular dichroism spectroscopy studies were performed to provide unequivocal evidence to confirm that the factor B vWF-A domain has allosteric properties and the conformation is dependent upon the presence of bound metal. This implies that the vWF-A interaction with C3b may alter its Mg^{2+}-bound coordination. Further ^{1}H NMR spectroscopic evidence is presented to confirm that the solution structure of factor B was not in an extended arrangement, but its five domains were involved in close interactions. It is proposed that the conformation of the Ba fragment is affected by its incorporation into factor B. The Ba fragment is involved in decoupling an interaction between the vWF-A and SP domains and the proximity of the vWF-A and SP domains within the Bb fragment leads to a conformational change in which conserved charged residues may be important. Allosteric structural rearrangements in the SP domain as the result of its interactions with the vWF-A domain or the Ba fragment provide an explanation of the regulation of the catalytic activity of factor B.

ACKNOWLEDGEMENTS

I wish to acknowledge the following people for their assistance and support during the course of my studies at the Royal Free and University College Medical School (RFUCMS). I would like to thank my supervisor Professor S. J. Perkins for providing me with the opportunity to study for this thesis and to the members of the protein structure group, past and present, for their help in all aspects of my work. In particular I would like to thank Dr C. Ullman and Dr M. K. Boehm for their careful guidance and patience and Mr J. Gor for his continual assistance. I would like to thank Dr K. Srai and Dr B. Ramesh for allowing me access to the Wolfson laboratory and the department of Oncology for accommodating us within their work space. Many thanks are due to Dr R. B. Sim (MRC Immunochemistry Unit, Oxford) for his collaboration and many useful discussions; Mr A. C. Willis for the protein sequence analysis (MRC Immunochemistry Unit, Oxford); Professor J. E. Volanakis (Dept. of Medicine, University of Alabama at Birmingham, U. S. A.) for the provision of factor D; Professor R. H. J. Begent and Dr D. I. R. Spencer (Dept. of Oncology, Royal Free Hospital) for access to the PBS-1 mass analyser; Dr J. Feeney, Dr T. Frenkiel, Dr M. Gradwell and the Medical Research Council for the provision of ^{1}H NMR facilities at the National Institute of Medical Research, Mill Hill, London; Dr A. J. Beavil (Kings College, London) for access to the CD spectrometer; Dr K. J. Welham for access to MALDI and electrospray equipment (School of Pharmacy, London); Dr. P. I. Haris (De Montfort University, Leicester) for useful discussions and Dr S. V. L. Narayana for the factor B SP coordinates (PDB code 1dle). Many thanks are also due to Dr. G. Brooker (Wadham College, Oxford) for his assistance with this document. I offer my apologies for the many people whom I have inevitably forgotten to mention, but may I offer solace by confirming that you are all wonderful people and that your influences were undoubtedly crucial for helping me to get this far. I would like to thank the Wellcome Trust for project grant support.

Finally I would like to thank my family and friends for all their support during these studies. I would especially like to thank Caroline for her remarkable strength and encouragement; during this time she has given birth to Lois and Oona and has in the most part been responsible for their care.

Contents **Page**

List of Abbreviations

2D-NMR	two dimensional nuclear magnetic resonance
α2m	α2macroglobulin
APS	ammonium persulphate
AMC	7-amino-4-isocoumarin
B_o	applied magnetic field
Boc	t-butyloxycarbonyl
BSA	bovine serum albumin
C8bp	C8 binding protein (synonym of homologous restriction factor)
cDNA	complementary DNA
C1/2/3/4/5/6/7/8/9	complement components 1/2/3/4/5/6/7/8/9
C1-inh	complement component 1 inhibitor
C1q	component of complement component 1
C1r	protease component of complement component 1
C1s	protease component of complement component 1
C3a/4a/5a	anaphylatoxins
C3b	proteolytically activated form of C3
C3i	hydrolytically activated form of C3 (synonym of C3(H_2O))
C3(H_2O)	hydrolytically activated form of C3 (synonym of C3i)
C3(NH_3)	amidated C3 (activated form of C3)
C4bp	C4 binding protein
C8bp	C8 binding protein (synonym of HRF)
CCP	complement control protein (synonym of SCR)
CD	circular dichroism
CDxx	cluster of differentiation or determinant (xx is a number)
CD11a/b/c/d	α subunit of the β2 integrins
CD18	β subunit of the β2 integrins
CD21	complement receptor 2
CD55	complement receptor 1
CD46	membrane cofactor protein

CD59	protectin
CI	chemical ionization
CR1/2/3/4/5	complement receptor 1/2/3/4/5
CR3	complement receptor 3 (composed of CD11b and CD18)
CRP	C reactive protein
DAF	decay accelerating factor
DEAE	diethylaminoethyl
DNA	deoxyribonucleic acid
DSS	2,2-dimethyl-2-silapentane-5-sulphonate
DSSP	dictionary of secondary structure of proteins
DTT	dithiothreitol
EBV	Epstein Barr virus
EDTA	ethylenediaminetetraacetic acid
EGF	epidermal growth factor
EI	electron ionization
ELISA	enzyme linked immunosorbent assay
ESI	electrospray ionization
Fab	antigen binding region of an immunoglobulin molecule
FB	factor B
Fc	C-terminal halves of two heavy chains of an Ig molecule
FI	factor I
FID	free induction decay
FIM	factor I module
FIMAC	factor I module/membrane attack complex
FPLC	fast performance liquid chromatography
FT-IR	Fourier transform infrared
GAL	galactose
GST	glutathione S-transferase
HIV	human immunodeficiency virus
HPLC	high performance liquid chromatography

HTH	helix-turn-helix
HRF	homologous restriction factor (synonym of C8bp)
HRF20	protectin (synonym of CD59, P-18 and MIRL)
IAM	iodoacetamide
Ig	immunoglobulin
IPTG	isopropyl-β-D-galactoside
LB	Luria Bertani
LDLr	low density lipoprotein receptor
LIMS	laser ionization mass spectrometry
LFA-1	cartilage matrix protein (composed of CD11a and CD18)
MAC	membrane attack complex
MALDI	matrix-assisted laser desorption ionisation
MASP-1/-2	mannose binding protein associated serine protease-1/-2
MBL	mannose binding lectin
MBP	mannose binding protein
MCP	membrane cofactor protein
MIDAS	metal ion dependent adhesion site
MHC	major histocompatibility complex
MIRL	membrane inhibitor of reactive lysis (synonym of protectin, CD59 and HRF20)
MPD	2-methyl-2,4-pentanediol
mRNA	messenger RNA
MTPBS	hypertonic phosphate buffered saline
NMR	nuclear magnetic resonance
nOE	nuclear Overhauser effect
OD	optical density
ORF	open reading frame
P	properdin
P-18	protectin (synonym of CD59, HRF20 and MIRL)
PBS	phosphate buffered saline

PC	phosphocholine
PCR	polymerase chain reaction
PDB	protein databank
PEG	polyethylene glycol
PHD	profile network system from Heidelberg
PLR	perforin-like region
PMSF	phenylmethylsulphonylfluoride
PNH	paroxysmal nocturnal haemoglobinuria
PVDF	polyvinyl difluoride membrane
RMS	root mean squared
rpm	revolutions per minute
PPM	parts per million
RaRF	Ra reactive factor (synonym of MBL and MBP)
RCA	regulation of complement activation
rFI	recombinant factor I
RMS	root mean-squared
RNA	ribonucleic acid
SCR	short consensus (complement) repeat (synonym of CCP) or the structurally conserved regions (relating to homology modelling)
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SDW	sterile deionized and distilled water
SELDI	surface enhanced laser desorption ionisation
SERPIN	serine protease inhibitor protein
SLE	systemic lupus erythematosus
SP40,40	clusterin
SPA	surfactant protein A
S-protein	vitronectin
sFI	serum factor I
SP	serine protease
TEMED	NNN'N'-Tetramethylenediamine

T_m	melting temperature
TOF-MS	time-of-flight mass spectrometry
TSR	thrombospondin repeat
UV	ultraviolet
vWF-A	von Willebrand factor type-A domain

AMINO ACID ABBREVIATIONS

Amino acid	3 letter format	1 letter format
alanine	Ala	A
arginine	Arg	R
aspartic acid	Asp	D
asparagine	Asn	N
cysteine	Cys	C
glutamic acid	Glu	E
glutamine	Gln	Q
glycine	Gly	G
histidine	His	H
isoleucine	Ile	I
leucine	Leu	L
lysine	Lys	K
methionine	Met	M
phenylalanine	Phe	F
proline	Pro	P
serine	Ser	S
threonine	Thr	T
tryptophan	Trp	W
tyrosine	Tyr	Y
valine	Val	V

NUCLEOTIDE ABBREVIATIONS

adenine	A
guanine	G
thymidine	T
cytosine	C

List of Tables **Description** **Page**

Chapter 1

The Complement System: The Major Effector of the Humoral Immune System is The First Line of Defence of the Immune System

1.1 The biological immune system

Immunity describes a complex process within the body, rather than a specific group of organs. Certain organs and tissue types do have a predominantly immunological function, including reticulo-endothelial tissue, bone marrow, lymph glands, the thymus, spleen and Peyer's patches (Kuby, 1997). The classic biological definition of immunity includes all of the physiological mechanisms that give an organism the ability to recognize foreign substances and neutralize or degrade them, with or without injury to the organism's own tissue. The roles of the humoral and cell-mediated responses of the immune system are interrelated and so provide the activities necessary for the immune response. It is important to see immunity in its broad biological context and the different aspects of immunity have been grouped as described below.

Innate immunity is immunity present from before birth and consists of many non-specific factors, together with blood-based immunity inherited from the mother, which operate against almost any substance that threatens the body. Four types of defensive barriers are involved: (a) Anatomical barriers which include the skin and mucous membranes. (b) Physiological barriers involve body temperature which inhibits the growth of some pathogens, acidic pH of the stomach and chemical mediators such as complement, interferon and lysozyme. Other barriers are provided by (c) endocytosis and phagocytosis by specialised cells and (d) the inflammatory response where vascular fluid and phagocytic cells leak into tissues.

Acquired immunity is a more specialised form of immunity found only in the vertebrates. It is a result of an encounter with a new substance, which triggers events that induce an immune response specifically against that particular substance. This involves B lymphocytes, T lymphocytes and macrophages and thus highlights the importance of lymphatic tissue, the site of lymphocyte maturation and differentiation. The primary functions of acquired immunity can be summarised in three categories: (a) the production of antibodies; (b) the stimulation of specialised cells, which destroy invading cells or organisms and neutralize their toxic products, and (c) the removal of damaged or dying cells by leukocytes. Acquired

immune responses are adaptive and display four characteristic attributes: antigenic specificity; diversity; immunologic memory and self/nonself recognition.

1.2 The complement system

The complement system is an important component of immune defence against infection. It is a multi-component protein system found in plasma, that is responsible for defence and clearance in the blood stream, and it was named as such because it helps the antibody response. It is part of the innate (non-adaptive) immune system and can respond to challenges by micro-organisms before an adaptive response has developed. Four major functions of complement are: (a) the recognition of target surfaces; (b) the opsonization and removal of complement-coated antigens via complement receptors on phagocytic cells; (c) the recruitment of phagocytic cells into the site of complement activation; (d) the destruction of membrane integrity of target organisms (reviewed in Law and Reid, 1995). Complement comprises over 30 proteins which function as enzymes, binding proteins, regulators or membrane bound receptors. The complement cascade proteins are a major constituent in plasma with a total concentration of around 4 g/l. C3 and C4 are the most abundant complement proteins at 1.3 and 0.6 g/l, respectively, which is consistent with their central role in the cascade, whereas factor D is the least abundant (1 μg/l). The properties of the major soluble proteins are summarised in Table 1.1. These proteins acquire the ability to interact with one another, with antibody and with cell membranes after activation of the system. These interactions directly generate the various biological interactions that are mediated by complement and which range from lysis of different kinds of cells, bacteria and viruses to direct involvement in inflammatory processes. In addition complement is able to enlist the participation of other humoral and cellular effector systems and to induce histamine release from mast cells, to direct migration of leukocytes and phagocytes, and to release lysosomal constituents from phagocytes. Complement is involved in the regulation of B lymphocyte activity, clearance of host cell breakdown products and the presentation of some antigens to antigen presenting cells.

Table 1.1 Properties of the water soluble complement proteins

Protein	Relative molecular mass	Serum concentration μg/ml	Number of polypeptide chains	Homologous proteins
C1q	465	80-100	18	MBP, SPA
C1r	85	35-50	1, cleaved to 2 on activation	trypsin like serine protease
C1s	85	35-50	1, cleaved to 2 on activation	trypsin like serine protease
MBP	540	0.1-5	18	C1q, SPA
MASP-1	85	2-15	1	trypsin like serine protease
MASP-2	85	uncertain	1	trypsin like serine protease
C4	195	300-450	3	C3, C5, α2m
C2	110	15-25	1, cleaved to 2 on activation	serine protease: factor B
C3	185	1000-1350	2	C4, C5, α2m
C5	185	60-90	2	C3, C4, α2m
C6	120	60-90	1	C7, C8, C9 α and β chains
C7	115	50-80	1	C6, C8, C9 α and β chains
C8	160	180-250	3	C6, C7, C9 α and β chains
C9	75	50-80	1	C6, C7, C8 α and β chains
factor B	90	180-250	1, cleaved to 2 on activation	serine protease: C2
factor D	25	2	1	trypsin like serine protease
properdin	220	20-30	oligomeric, usually tetramer of 56 kDa subunit	thrombospondin
factor H	155	100-150	1	RCA family: CR1, CR2, MCP, DAF
factor I	88	30-40	2	none
C4bp	540	200-400	7 × 70 kDa plus 1 × 50 kDa	RCA family: CR1, CR2, MCP, DAF
C1-Inh	110	150-300	1	serpin

Abbreviations used: α2m, α2macroglobulin; CR1/2, complement receptor 1/2; DAF, decay accelerating factor; MBP, mannose binding protein; MASP, MBP associated serine protease, MCP, membrane cofactor protein; RCA, regulation of complement activation; SPA, surfactant protein A.

Various proteins of the complement system detect "targets" and bind to them, usually by mechanisms that involve the recognition of charge distribution patterns or of carbohydrate on the surface of the target. Recognition and activation occur by three major routes (Figure 1.1): (i) the classical pathway, primarily by immunoglobulin and immune complexes (IgG and IgM); (ii) a branch of the classical pathway known as the lectin pathway, activated by complex carbohydrates; and (iii) the alternative pathway, by the molecular structures on the target cell surface that disrupt the delicate balance of proteins involved. The binding of complement proteins to the target results in activation of the complement system and formation, on the target, of unstable complex proteases, the C3 convertases. The C3 convertases (designated C4b2a and C3bBb in Figure 1.1) are each made up of two protein components. One component (C3b or C4b) is covalently bound to the surface of the complement activator and the other (C2a or Bb) is a serine protease that is able to cleave and activate C3. The component C3 is a major plasma glycoprotein and plays a central role in the system, being common to all three pathways. The major fragment of activated C3, C3b, binds covalently to complement activating surfaces such as cells, viruses and immune complexes. When large amounts of C3b have been deposited, phagocytosis of the substance is greatly enhanced. This occurs partly through the interaction of surface-bound C3 fragments with C3 receptors located on phagocytic cells. If the complement activator has a lipid bilayer, lysis can occur through interaction with the membrane of components C5, C6, C7, C8 and C9. The formation of a C5 convertase enzyme (C4b2a3b classical/lectin and C3bBb3b alternative) can initiate the terminal or lytic pathway. The end result of the cascade is the formation of the membrane attack complex (MAC) with an M_r of the order of 1,000,000. The MAC forms transmembrane channels which displace lipid molecules and other constituents, thus disrupting the phospholipid bilayer of target cells and leading to osmotic cell lysis.

Other biologically important functions mediated by the complement system involve the low molecular mass anaphylatoxins C3a, C4a and C5a. While these have effects on a wide range of cells their major effect is to increase vascular permeability. They can also promote smooth muscle contraction, and C5a is involved in neutrophil chemotaxis.

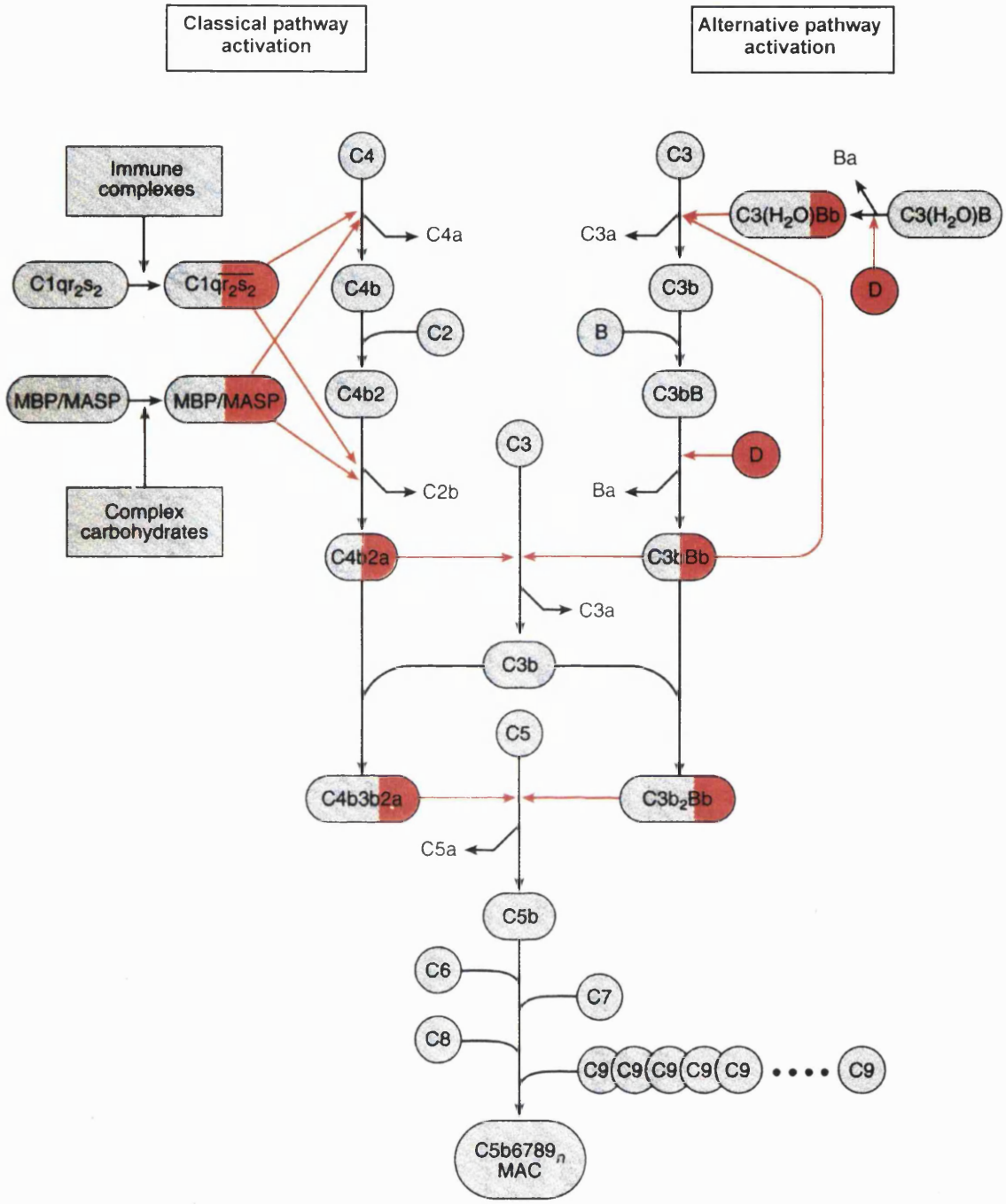


Figure 1.1 The activation steps of the complement system. The classical pathway (left) is triggered by immune complexes (via C1) or by complex carbohydrates in the lectin pathway (via MBP/MASP), while the alternative pathway (right) is triggered by a wide variety of compounds and cell surfaces (via factor B). The number of C9 molecules (*n*) within the $C5b6789_n$ complex can vary between 1 and 18. Enzymatic cleavage is indicated as solid red lines and the enzymatically active components are shaded red. Adapted from Law and Reid, 1995.

The large C3b and C4b fragments, and degradation fragments of C3 bound to the complement activator, are important for clearing immune aggregates and so have an influence on the regulation of immune complex size.

The complement system in humans is well characterised at the biochemical level. The primary structures of the proteins in the system are mostly known, and knowledge of the three-dimensional structures is progressing. The activities of the proteins *in vitro* are generally known in considerable detail, but many uncertainties remain about their activities *in vivo*. Similar complement systems occur in all mammals and variants of the system have been found in all other vertebrate classes. A few proteins have been described in invertebrates that are similar to complement proteins, and these represent stages in the development of the activities of the mammalian complement system (Dodds and Day, 1993).

Since the system is involved in the removal of materials from the circulation and tissues, it has the potential to opsonise or lyse host cells. In addition to the beneficial effects of complement, undesirable complement-mediated tissue damage may occur in many situations. These situations include mechanical injury, viral infection, tissue damage initiated by autoantibodies in chronic inflammatory conditions and rheumatoid arthritis. Complement is responsible for the acute rejection of xenotransplants. Diminished complement activity that arises through consumption of complement or from genetic deficiencies is associated with susceptibility to infection and inadequate removal of immune complexes from the circulation, which may lead to damage of the small blood vessels, particularly of the skin and kidneys.

1.3 Classical pathway

The classical pathway of complement consists of the glycoproteins C1q, C1r and C1s and C2-C9 (Figure 1.1). One molecule of C1q and two of C1s and C1r associate in the presence of Ca^{2+} ions to form a large protein complex known as C1. C1q is the molecule that interacts with most potential targets. Classical pathway activation has been mostly studied with immune complexes that contain IgG or IgM antibodies as the activator.

1.3.1 The structure of C1q, C1r and C1s

The complement classical pathway C1 enzyme complex is composed of the subunits C1q, C1r and C1s. C1q is made from 18 polypeptide chains of three types, 6A, 6B and 6C arranged in a complex structure often described as a "bunch of tulips". Three of these chains (1A, 1B and 1C) form a collagen-like triple helix at the N-terminal end of the molecule (Reid and Porter, 1976). The six triple helices are aligned in parallel for half the collagen-like length and then diverge to terminate at the C-terminus in six globular heads (Brodsky-Doyle *et al.*, 1976). The subcomponents C1r and C1s have homologous amino acid sequences and similar tertiary structures (Kishumoto *et al.*, 1989). C1r and C1s interact to form a tetrameric chain, C1s-C1r-C1r-C1s in which the two catalytic domains of C1r are orientated in a head to tail manner (Arlaud *et al.*, 1986). C1s interacts with C1r via the C1r/C1s specific domain. The activated $C\overline{1}r_2C\overline{1}s_2$ complex has been found to have an asymmetric X structure (Weiss *et al.*, 1986; Perkins and Nealis, 1989). A variety of models have been proposed for the interaction of the $C\overline{1}r_2C\overline{1}s_2$ tetramer with C1q to form the C1 complex (Perkins, 1989). These models can be divided into two groups: those which propose that the $C\overline{1}r_2C\overline{1}s_2$ tetramer is positioned on the outside of C1q, summarized in the "W-model" (Perkins and Nealis, 1989) and the "O_2-model" (Cooper, 1985); and those proposing that the $C\overline{1}r_2C\overline{1}s_2$ subunits are interwoven between the arms of C1q, summarized in the "S-model" (Schumaker *et al.*, 1986) and the "8-model" (Colomb *et al.*, 1989).

1.3.2 Activation

Activation of the classical pathway is initiated by the binding of C1q to a variety of substances (Sim and Reid, 1991). The most familiar is the formation of immune complexes of either IgG or IgM with antigens. Many other substances can activate the classical pathway, without a requirement for antibody (Sim and Malhorta, 1994). These include:

1. nucleic acid and chromatin
2. cytoplasmic intermediate filaments
3. mitochondrial membranes possibly via cardiolipin or mitochondrial proteins
4. some viruses, e.g. murine leukaemia virus (MuLV)
5. gram positive bacteria, e.g. some pneumococci, streptococci, via capsular

polysaccharide

6. gram negative bacteria via the lipid A component of the lipopolysaccharide of the cell wall.

The initial step of activating the classical pathway is the binding of C1q via the globular heads to the activator. The interaction between C1q and the Fc region of immunoglobulin (IgG or IgM) is facilitated by the close proximity of many IgG molecules (Borsos, 1989) or a conformational change in the Fab arms of IgM (Perkins *et al.*, 1991) to expose the C1q binding sites. It is thought that the domain types of the globular heads are likely to have differing specificities for charge groupings on the surface of the activator. Activation of complement requires multiple interactions between a single molecule of C1q and the activator, and therefore the activator is usually of high molecular mass and has a repetitive structure such as exposed lipid A on bacterial surfaces or multiple antibody molecules bound to a particulate antigen.

When two or more of the globular heads bind the activator, a conformational change in the collagenous stalks of C1q is induced (Heinz, 1989) which increases the affinity of C1q for $C\bar{1}r_2C\bar{1}s_2$. $C\bar{1}r_2C\bar{1}s_2$ will interact with C1q only in the presence of Ca^{2+} ions, once C1q has bound to an efficient activator. The catalytic domain of one C1r autoactivates and cleaves the corresponding domain in the neighbouring C1r molecule (Dodds *et al.*, 1978). This in turn activates C1s by a proteolytic cleavage (Ziccardi, 1976). The activated $C\bar{1}s$ is then able to cleave the classical pathway component C4 into C4a and C4b (Thielens *et al.*, 1984). C4 has sequence homology to the complement components C3 and C5 and the non-complement proteins α_2-macroglobulin and pregnancy zone protein. The proenzyme form of C4 is composed of three chains: α, β and γ. $C\bar{1}s$ cleaves a single peptide bond in the α-chain of C4 to produce a 9 kDa fragment C4a and a metastable fragment C4b. The C-terminal portion of the α-chain (α'-chain) of C4b remains disulphide linked to the β-chain. The larger fragment, C4b, has an exposed thiol ester in the α' chain that is able to react non-specifically with any available nucleophiles (Harrison *et al.*, 1981). These nucleophiles may, for example, be the hydroxyl groups on sugars that form an ester bond to the carbonyl of the thiol ester or

amino groups located on a variety of surfaces, which form an amide bond. A proportion, usually less than 10%, of the C4b ends up covalently bound to the complement activator. The remainder reacts with water and diffuses away from the site of complement activation. Proenzyme C2 binds to the surface-bound C4b and, if it is appropriately positioned close to activated C1s, it is cleaved to form a Mg^{2+} dependent complex, C4b2a (Horiuchi *et al.*, 1991). The C4b2a complex is the classical pathway C3 convertase enzyme which is able to cleave and activate C3, a homologue of C4.

1.3.3 Mannose binding protein (MBP)

Molecules other than C1q can also participate in the activation of the classical pathway of complement (Sim and Malhotra, 1994; Holmskov *et al.*, 1994; Malhotra *et al.*, 1994, 1995). The activation of the classical pathway via complex carbohydrates is often referred to as the lectin pathway. Mannose binding protein (MBP), also known as Ra reactive factor (RaRF) and mannose- or mannan binding lectin (MBL), is able to substitute for C1q after interaction with mannose-rich structures on yeasts, bacteria and viruses (Figure 1.1). It does not bind to normal IgG, but can activate complement on interaction with the carbohydrate groups of a glycosylation variant of IgG, which is present at elevated levels in rheumatoid arthritis (Malhotra *et al.*, 1995). The structure of MBP resembles that of C1q in that it has collagenous segments and 6 globular heads. In MBP, each globular head is made up of three identical C-type lectin domains which interact with carbohydrate in a Ca^{2+} dependent manner. MBP can activate C1r and C1s and molecules that are structurally similar to C1r and C1s that are termed MBP associated proteases (MASPs) (Malhotra *et al.*, 1994). This antibody-independent pathway may be important for immunodeficient individuals and the very young who have not yet developed a mature immune system.

1.3.4 C-reactive protein (CRP)

C-reactive protein (CRP) is a protein which, in mammals, is expressed during the acute phase response to tissue injury or inflammation. CRP displays several functions associated with host defense: it promotes agglutination, bacterial capsular swelling, phagocytosis and complement fixation. CRPs have also been sequenced in an invertebrate, the Atlantic horseshoe crab, where they are a normal constituent of the hemolymph. CRP can form complexes with charged groups, including the phosphate groups in choline phosphate of pneumococcal C-type polysaccharide and microbial polysaccharides (both those containing phosphocholine (PC) and not containing PC) and lipids, polyanion/polycation complexes and chromatin. Bound CRP can interact with C1q to activate the classical pathway.

The classical pathway is therefore activated by a wide range of stimuli. C1q and CRP are principally involved in recognising charge clusters including carbohydrates and lipids. In contrast MBP recognises neutral sugars.

1.3.5 Control proteins

Regulation of the classical pathway is controlled at two stages: formation of the $C\bar{1}$ enzyme complex; and control of the C3 convertase, C4b2a. Many serine proteases in the blood have relatively specific natural inhibitors that belong to the "serpin" family. Of the complement proteases, only C1r, C1s and possibly both MASPs are controlled by a serpin, named $C\bar{1}$-inhibitor ($C\bar{1}$-inh). $C\bar{1}$ inhibitor controls $C\bar{1}r_2$ activation through non-covalent interactions (Arlaud *et al.*, 1989; Ziccardi and Cooper, 1979). $C\bar{1}$ inhibitor can also covalently interact with the four serine protease active sites of the $C\bar{1}$ complex, resulting in the dissociation of the complex to form two molecules of $C\bar{1}$ inhibitor, $C\bar{1}r$-$C\bar{1}s$-$C\bar{1}$ inhibitor (Ziccardi and Cooper, 1979). The free collagen-like stalks of C1q are then free to bind to cell-surface C1q receptors which may then initiate phagocytosis (Malhotra *et al.*, 1990; Malhotra and Sim, 1989; Ghebrehiwet, 1989). Hereditary or acquired lack of $C\bar{1}$-inhibitor causes angiooedema. Factor J also controls the formation of the $C\bar{1}$ complex by inhibiting the association of $C\bar{1}r_2C\bar{1}s_2$ with C1q (Lopez-Trascasa *et al.*, 1989).

Control of the C3 convertase is governed by its inherently unstable nature and cleavage by factor I. C2a spontaneously dissociates from the complex. This dissociation is enhanced by C4 binding protein (C4bp) (Gigli *et al.*, 1979) and decay accelerating factor (DAF) (Nicholson-Weller *et al.*, 1982). Factor I cleaves C4b in two positions to give the fragments C4c and C4d. C4bp (Fujita *et al.*, 1978; Gigli *et al.*, 1979), membrane cofactor protein (MCP) (Seya *et al.*, 1986) (Table 1.2) or complement receptor 1 (CR1) (Table 1.3) act as cofactors for this factor-I mediated cleavage. In most physiological circumstances, the concentration of C4bp exceeds that of C4b generated by activation, and therefore further activation of the complement components is prevented.

1.4 Alternative pathway

The proteins of the alternative pathway of complement are factor D, factor B, C3 and C5 (Figure 1.1). Components C5-C9 are common to both pathways and will be discussed in the terminal components section. C3 has a central role in both pathways. The regulatory proteins factor H, factor I and properdin have major roles in controlling activation of the alternative pathway. Activation of the alternative pathway can be both antibody dependent and antibody independent. Antibody dependent activation can occur via IgG, IgA and (rarely) IgE immune complexes. Antibody independent activation can be effected by a whole spectrum of substances located on the surfaces of bacteria, fungi, viruses, multicellular parasites and tumour cells.

1.4.1 Activation

The mechanism by which the targets are recognised by the alternative pathway is less well understood than for the classical pathway. In the alternative pathway, several proteins are involved simultaneously in recognition. Perturbations in the interaction between C3b deposited on the activating surface and regulatory molecules is important for determining whether complement is activated or not.

Table 1.2 Membrane bound regulators of complement.

Regulator	Action	Characteristics
DAF	Dissociates C3 convertases.	70 kDa glycoprotein present on the membranes of peripheral blood cells, vascular endothelial cells, placenta and epithelial cells. Extracellular portion contains 4 SCR domains. Soluble forms occur.
MCP (CD 46)	Binds C3b and acts as a cofactor to factor I mediating C3 cleavage. Cellular receptor for measles virus.	50-60 kDa integral membrane glycoprotein present on all circulating cells except erythrocytes and most other cell types. Contains 4 SCR domains.
Protectin (CD 59) (p18) (MACIF) (HRF20) (MIRL)	Binds C5d-8 complex and prevents formation of the polymeric C9 complex.	20 kDa glycoprotein present on the membranes of all circulating blood cells, endothelial cells, epithelial cells and spermatozoa.
Homologous restriction factor (HRF) (C8 binding protein) (C8bp)	HRF/C8bp prevents the binding of C9 to C8αγ and its insertion into membranes.	Poorly characterised to date.

Abbreviations used: DAF, decay accelerating factor; MCP, membrane cofactor protein; MIRL, membrane inhibitor of reactive lysis; HRF, homologous restriction factor; C8bp, C8 binding protein; SCR, short consensus repeat

Native C3, like C4, contains an internal thiolester in the α-chain which can be hydrolysed to form C3i (also known as C3(H_2O)) and is capable of attaching to membranes (Pangburn and Müller-Eberhard, 1980; Pangburn *et al.*, 1981). Although C3i is structurally different from C3b in that it has an intact α-chain, both C3i and C3b are functionally similar. C3 is continually activated at a slow rate in the fluid phase by three mechanisms: (i) C3 may be cleaved to C3b by serum proteases; (ii) small nucleophiles, or water, may gain access to C3 and react with the thiolester; (iii) C3 may be subject to non-specific perturbation leading to exposure and hydrolysis of the thiolester. Thus the alternative pathway is in a permanently activated state. If activated C3 is deposited on an activating surface of the alternative pathway, it can serve as a seed for the positive amplification loop which operates explosively. Activated C3 is able to form a C3 convertase with factor B in the presence of factor D. Removal of C3a by proteolytic cleavage by the C3 convertase induces a conformational change in C3b which leads to the exposure of the internal thiolester, which is normally buried in native C3. The exposed thiolester is extremely reactive with nucleophiles, including water and molecules bearing hydroxyl or amino groups. C3b, activated at the surface of a foreign cell, is largely restricted to binding to the surface of the same cell or to being inactivated by water. This puts a limit on the spatial range of the activated C3b. The deposition of C3b is minimised on host cells because of their inability to activate the host's own complement pathways. C3 activation is kept to a low level by the control proteins, factor H and factor I in the blood. The C3 and C5 convertases decay quite rapidly by the dissociation of the enzymatic components in the absence of control proteins. The survival of the first C3b molecule deposited on the surface of a substance determines whether it will be an activator or non-activator of the alternative pathway

Activation of the alternative pathway does not depend upon antibodies recognising specific molecules on the target cell surface: rather it relies on molecular structures on the target cell to upset the delicate balance of proteins involved to focus their activation and deposition on its surface. Activators include polysaccharides, fungi, bacteria, viruses, parasites and certain mammalian cells. It is not known what structure all the activators have in common but it is possible that C3b deposited on activators is protected from proteolytic degradation,

enabling the formation of the C3bBb complex (Fearon, 1978; Pangburn *et al.*, 1980; Fearon and Austen, 1975). C4b bound to a surface can also activate the alternative pathway by binding to C3bBbP, which may be important in cases of C2 deficiency (Farries *et al.*, 1990).

The most obvious similarity between the classical and alternative pathway is the C3 convertase enzyme. The C3 convertase of the classical pathway is C4b2a which is very similar to the alternative pathway C3 convertase, C3bBb. C4b is a homologue of C3b and C2 a homologue of factor B. Factor D is a serine protease that has a role similar to the classical pathway C1s. The alternative pathway does not have proteins similar to C1q or C1r. A paradoxical situation arises in which C3b is required in order to assemble the enzymes responsible for C3 cleavage. *In vivo* the classical pathway is also present and the amplification of the C3b deposition via the alternative pathway occurs if the classical pathway C3 convertase (C4b2a) is the source of the initial C3b molecule. Multiple C3b molecules are deposited in clusters on the complement activator. This cluster formation is important in mediating multiple interactions with C3 receptors on phagocytic cells.

1.4.2 Control proteins

Control of the alternative pathway is comparable to that of the classical pathway in that homologous proteins fulfill similar roles. Both positive and negative control mechanisms exist. The C3bBb convertase is stabilized by the binding of the control protein properdin (P) to form a C3bBbP complex (Fearon and Austen, 1975; Medicus *et al.*, 1976; Farries *et al.*, 1987), which facilitates the binding of factor B and prevents cleavage of the convertase by factor I (Farries *et al.*, 1988a). The versatility of complement to recognise all types of foreign material is desirable, but its binding to host cells must be minimised. Negative control of activation is governed by factor H which acts in a similar way to C4bp in the classical pathway by acting as a cofactor for factor I-mediated cleavage of C3b and C3i to form the fragments iC3b + C3f and iC3i + C3f respectively (Pangburn and Müller-Eberhard, 1983). The factor H molecule can interact with C3b via three binding sites and so prevent the formation of the C3bBb convertase and it can dissociate the complex (decay accelerating activity). Factor H is also able to discriminate between activator and non-activator bound C3 molecules. Non-

activators, which include host cells, carry polyanionic substances like sialic acid or glycosaminoglycans on the cell surface. The high affinity of factor H for non-activator bound C3b seems to be a joint recognition of both C3b and surface structures. Various microbes have capsules rich in sialic acid (group B streptococcus, *E.coli* capsule type K1 and group B meningococcus) which allows them to escape alternative pathway activation, opsonisation and phagocytosis.

1.4.3 Protection mechanisms

The majority of C3b and C4b molecules (70 to 90%) formed as a result of complement activation do not become covalently linked to surfaces but remain in the fluid phase (Hourcade *et al.*, 1989). Of those that do bind, some will do so to host cells. These are protected against complement-mediated cytolysis by the membrane proteins DAF and MCP (Atkinson and Farries, 1987). DAF prevents assembly of C3 convertase on membranes and will dissociate enzymes already bound. MCP has cofactor activity for factor I mediated cleavage (Liszewski *et al.*, 1991; Fearon, 1979).

1.5 Terminal pathway

1.5.1 C5 convertase

The binding of one or more C3b molecules to either the classical or the alternative C3 convertase (C4b2a or C3bBb) changes the specificity of the complex to a C5 convertase, C4bC2aC3b or C3bBbC3b respectively (Medicus *et al.*, 1976; Daha *et al.*, 1976). The alternative pathway C5 convertase is regulated by MCP and factor I by the conversion of one C3b dimer to C3bi (Seya *et al.*, 1991), whereas the classical C5 convertase is controlled by factor H (Ito and Tamura, 1983). C5 binds to the C3b fragment in the C5 convertase complex. This convertase cleaves a single peptide bond in the α-chain of C5 to yield the fragments C5a and metastable C5b (Nilsson *et al.*, 1975). Although C5 is very similar in structure to C3, it does not have an internal thiolester bond and therefore does not bind covalently to surfaces after activation (Wetsel *et al.*, 1988; Law *et al.*, 1980).

1.5.2 Biological activity

The C5b fragment contains a labile binding site with specificity for component C6. The complex C5bC6 remains loosely bound to C3b on the target cell surface until interaction with C7. The C5b67 complex undergoes a conformational change in which it dissociates from C3b, and a metastable membrane binding site is formed on the C7 subunit (Preissner *et al.*, 1985). C5b67 binds to a membrane to become an integral membrane protein. C8 is composed of three chains (α, β and γ) and binds to C5b67 via the β-chain (Monahan and Sodetz, 1981; Stewart *et al.*, 1987). C8 is then able to undergo a conformational change which allows penetration into the membrane of the disulphide linked α- and γ-chains. C9 binds to the C5b678 complex via the α-subunit on C8 (Stewart *et al.*, 1987). This catalyses the polymerisation of up to 18 molecules of C9. C5b678 forms small functional channels of approximately 3 nm in diameter which can enlarge to 10 nm upon incorporation of C9 molecules (Ramm *et al.*, 1985). The composition of the membrane attack complex (MAC) depends on the availability of monomeric C9. If sufficient C9 is present, binding will continue until it forms a typical cylindrical membrane lesion as seen in electron micrographs.

The assembly of a functional MAC on a cell membrane leads to cytolysis. Two major hypotheses exist to explain cell death, the "doughnut" hypothesis where the MAC components form a protein-lined channel or the "leaky patch" hypothesis where there is a destabilisation of the lipid bilayer due to the insertion of the MAC (reviewed in Bhakdi and Tranum-Jensen, 1991 and in Esser, 1991). The MAC can destroy the membrane of an enveloped virus and can kill many types of bacteria. The formation of an MAC on the surface of a nucleated cell will stimulate a rapid increase in the concentration of intracellular Ca^{2+} which in turn stimulates the recovery processes of the cell. This allows a cell to survive a mild complement attack and also initiates an inflammatory response. If a much stronger attack occurs, the recovery processes are overwhelmed and the cell dies (Morgan *et al.*, 1986).

1.5.3 Control proteins

Control of the terminal components occurs at two points. The complement associated protein SP40,40 (clusterin) binds to C5b6 to prevent the formation of the complete MAC (Choi *et al.*, 1989). The S-protein (vitronectin) and serum low density lipoprotein (LDL) will compete with membrane lipid for the binding site on C5b-7 and hence can prevent attachment to the surface of bystander cells (Podack *et al.*, 1978). Attack on host cells by the MAC is inhibited mainly by the cell surface regulatory protein CD59, which binds to the MAC and alters its mode of interaction with the membrane.

1.5.4 Protection mechanisms

Complement-mediated cell lysis is dependent on homologous restriction where complement from one species is inefficient at lysing cells of the same species. The interaction of specific membrane proteins with the terminal complement components ensures that the self-inflicted complement damage is kept to a minimum (Lachmann, 1990). Table 1.2 summarises the actions and characteristics of the membrane-bound regulators. Homologous restriction factor (HRF) or C8 binding protein (C8bp) (Zalman *et al.*, 1989) prevents the binding of C9 to C8$\alpha\gamma$ and its insertion into membranes (Schönermark *et al.*, 1988). Another of these protective proteins is protectin which is known by many other names; P-18 (Sugita *et al.*, 1988), HRF20 (Okada *et al.*, 1989), membrane inhibitor of reactive lysis (MIRL) (Holguin *et al.*, 1989) or CD59 (Davies *et al.*, 1989) (Table 1.2). CD59 bound to the C5b-8 complex is thought to prevent the binding of the first molecule of C9 which is required to initiate polymerization and MAC formation (Lachmann, 1991; Rollins and Sims, 1990).

1.5.5 Complement receptors

In addition to the membrane-bound regulatory proteins of the complement system, there are several receptors for the complement proteins or their activation fragments. These receptors are involved in a wide range of biological activities. There are one or more distinct complement receptors on the surface of most types of cells and on many tissue cells. Interaction of the complement ligands with these specific cellular receptors triggers various responses including cellular activation, secretion of mediators, ingestion, directed migration,

or other activities depending on the receptor engaged or cell type. A number of these are well characterised at the biochemical level, but others are identified only by their binding function. The receptors for the C3 fragments CR1, CR2, CR3 and CR4 have a number of regulatory and receptor functions (Table 1.3) (Sim *et al.*, 1987; Ross, 1989). CR1 and CR2 are anchored to the membrane via hydrophobic transmembrane segments (Klickstein *et al.*,1988). CR1 will inhibit complement activation either by accelerating the decay or dissociation of the C3 convertases or by acting as a cofactor for factor I-mediated cleavage of C3b. It is involved in processing of immune complexes and in the mediation of binding and phagocytosis of C3b-coated particles by phagocytic cells. CR2 bound to C3b-bearing particles enhances B-cell proliferation in the presence of T-cell derived factors (Lambris, 1988). CR3 adheres to and promotes phagocytic uptake of iC3b-coated particles into circulating phagocytic cells. CR4 mediates binding of iC3b and C3dg (Gaither *et al.*, 1987).

A C1q receptor (also called the collectin receptor) may be principally involved in phagocytosis of C1q-bearing particles. It also serves as a receptor for MBP and for a small family of proteins, the collectins, which are related in structure to C1q and MBP (Malhorta *et al.*, 1990, 1992). Other receptors of interest in the complement system are the receptors through which the small anaphylotoxins C4a, C3a and C5a exert their effects (see next section).

Table 1.3 Receptors for complement proteins and fragments.

Receptor	Ligand	Characteristics
CR1 (CD55)	C3b C4b	Polymorphic in size. The main variant contains 30 SCR domains in an elongated string-like structure. Found on erythrocytes and most leukocytes, tissue macrophage and kidney. Major role in transport and phagocytosis of C3b-bearing immune complexes.
CR2 (CD21)	iC3b, C3d, EBV	Elongated structure like CR1 which contains 15 or 16 SCR domains. Distribution is limited to B and T lymphocytes and follicular dendritic cells. Major role is regulation of B cell activities. Receptor for Epstein Barr Virus (EBV).
CR3 (CD11b, CD18)	iC3b	Member of the integrin family. Found on phagocytic cells. Major role in phagocytosis. Has binding sites for non-complement ligands.
CR4 (CD11c, 18, p150, 95)	iC3b	Homologous to CR3 with similar distribution and possibly a similar role. Has binding sites for non-complement ligands.
CR5	C3d/C3dg	Uncharacterised at the molecular level. Binding activity for soluble C3d on polymorphs.
C3a receptor	C3a	Histamine release from mast cells.
C5a receptor	C5a	Mediates inflammatory response. Found on neutrophils, monocytes, leukocytes, macrophages, mast cells and smooth muscle
C1q receptor	C1q	Mediates cellular cytotoxicity, opsonization and immune complex localization. Found on B lymphocytes, neutrophils, monocytes, and endothelial cells.

Abbreviations used: CR1-5, complement receptor 1-5; EBV, Epstein Barr Virus; SCR, short consensus repeats.

1.6 The physiological effects of complement

1.6.1 The inflammatory response

The cleaved peptides, C3a, C4a and C5a (anaphylatoxins) are functionally and structurally homologous (Greer, 1986). These can bind to specific receptors, C3a and C5a receptors (Lambris, 1988), to elicit a variety of cellular responses (Damerau, 1987; Rollins and Springer, 1985; Huey and Hugli, 1985; Johnson and Chenoweth, 1985). Signal transduction of the C5a receptor is mediated by a GTP-binding protein (G protein) (Siciliano *et al.*, 1990). The sequence of the C5a receptor (Boulay *et al.*, 1991; Gerard and Gerard, 1991) has significant homology to other G protein coupled receptors (Boulay *et al.*, 1991). C5a is the most potent mediator of the inflammation response. This leads to chemotactic migration of neutrophilic granulocytes and monocytes, cell adhesion, enzyme release and the formation of arachidonic metabolites and active oxygen species. C3a and C5a can induce histamine release from mast cells and prostaglandins from macrophages, which have vasodilator activity and increase the efflux of macromolecules from the plasma. The C-terminal arginine of the anaphylatoxins has been found to be essential for activity. Control of the inflammatory response is mediated by the anaphylatoxin inactivator (carboxypeptidase N) which selectively removes this arginine (Bokisch and Müller-Eberhard, 1970; Plummer and Hurwitz, 1978).

1.6.2 Defence against infection

Complement can defend against infection, either in combination with antibodies, or independently of antibodies. C3b produced after alternative pathway activation binds to the surfaces of foreign cells. Since many host phagocytic cells have receptors for C3b (Fearon and Wong, 1983), these are brought into close contact with the foreign organism. This opsonizes the bacterium and encourages endocytosis by polymorphonuclear leukocytes and monocytes. The Fc, CR1 and CR3 receptors on the phagocytes recognize the Fc portion on the antibody, C3b and iC3b respectively (Micklem and Sim, 1984; Gresham and Volanakis, 1986). If the amount of antibody attached via the Fc receptor is high enough, phagocytosis of the immune complex will commence. If there is insufficient antibody present, activation of the complement cascade may occur. This will increase efficiency of the binding of phagocytes

due to the deposition of C3b on foreign cell surfaces. It is generally considered that opsonization is more important than lysis in defence against infections.

1.6.3 Involvement in other systems

The complement cascade does not function in isolation, but also affects both the kallikrein/kinin and the coagulation/fibrinolysis systems (Kalter *et al.*, 1985; Ziccardi, 1983). Activated factor XII (Hageman factor) splits prekallikrein to kallikrein and plasminogen to plasmin, and all these enzymes are able to cleave C1̄ and initiate the complement cascade. Kallikrein and plasmin both cleave C3, and plasmin can cleave factor B. C1q can bind to fibrinogen and fibrin to localize C1q at the site of a wound to provide efficient opsonization (Entwhistle and Furcht, 1988). Control over the three systems is exerted by C1̄ inhibitor which is active against C1̄r and C1̄s, factor XII, plasmin and kinin-generating kallikrein, and also by carboxypeptidase N which inactivates C3a, C5a and kinin by splitting off the terminal arginine. Although protein S is a vitamin K dependent cofactor in the coagulation system (Comp *et al.*, 1984), it also can form a complex with the complement protein C4bp (Hessing, 1991), although it does not affect the function of C4bp as a regulator of the C3 convertase (Dahlbäck *et al.*, 1983).

1.6.4 The role of complement in disease

Although some deficiencies of complement components have no obvious deleterious effects, many produce considerable morbidity and mortality. The analysis of disorders involving deficiencies in the complement cascade can highlight the functions of the components concerned. Most commonly, deficiencies in the components of the classical pathway proteins are identified in patients with symptoms resembling systemic lupus erythematosus (SLE) (Morgan and Walport, 1991), leading to the conclusion that perhaps a major role of complement *in vivo* is in the prevention of immune-complex disease (Lachmann, 1990). In addition to this, manifestations of renal disease, arthritis, Reynaud's phenomenon and mucosal lesions are observed in individuals with deficiencies of the classical pathway components.

C1̄ inhibitor deficiency leads to the disease hereditary angioedema (HANE) which is

characterised by swelling of the mucous membranes. This is due to the increased activity of C4, C2 and kallikrein to produce the anaphylatoxin C4a, C2-kinin fragment and bradykinin. These fragments increase vascular permeability and hence oedema (Mollnes and Lachmann, 1988).

Factor D deficiencies are associated with recurrent bacterial infections (Kluin-Nelemans *et al.*, 1984). Properdin deficiencies are inherited as an X-linked trait and lead to *Neisseria* infections in affected males (Nusinow *et al.*, 1985). C3 deficiencies are characterized by recurrent pyrogenic infections, thus confirming the essential role that C3 plays in opsonizing bacteria for removal by phagocytic cells (Alper *et al.*, 1972; Berger *et al.*, 1983). Deficiency in factor I causes an unrestricted amplification of the alternative pathway which depletes factor B and C3 (Nusinow *et al.*, 1985). Factor H deficiencies have the same characteristics as those for factor I (Thompson and Winterborn, 1981) due to the fact that factor H is an obligate cofactor for factor I.

Deficiencies in the terminal pathway are linked with a susceptibility to recurrent infections by the bacterium *Neisseria*. Such infections are rare in Europe and the U.S.A. but relatively common in developing countries (Orren *et al.*, 1987; Schlesinger *et al.*, 1990). Deficiencies of the complement receptors CR1 (Walport and Lachmann, 1988), CR2, CR3 and CR4 (Kishimoto *et al.*, 1989) lead to a number of disease states, including immune-complex diseases, leukocyte adhesion and deficiency syndrome, hypertrophied gums and a tendency to have indolent staphylococcal skin infections. The lack of an inflammatory response is because the emigration of polymorphs and monocytes into the inflammatory region does not occur (Lachmann, 1990).

Patients with paroxysmal nocturnal haemoglobinuria (PNH) lack the ability to synthesize glycosylphosphatidylinositol (GPI) anchors which link a variety of host protection proteins to the cell membrane. These include DAF, HRF and CD59 (Table 1.2). This makes the cells highly susceptible to complement-mediated haemolysis (Lublin and Atkinson, 1989).

1.7 Domain structures and protein families in complement

1.7.1 The definition of a domain

A domain is defined as a part of a protein that can fold independently of its neighbouring sequences and functions as an individual entity within the whole protein (Wetlaufer, 1981; Doolittle, 1995). The underlying basis of domain organisation is the protein folding that stabilises an unique structure. The minimum size of a domain has been estimated at around 40 residues which reflects the smallest stable structure whereas the largest domain is around 400 residues and represents the largest domain that can fold up efficiently as a single unit (Wetlaufer, 1981).

DNA replication is given to a variety of iterative errors and, through homologous recombination, there is a tendency for duplication to beget more duplication (Doolittle, 1995). It seems that, during evolution, regions of similar sequences have arisen in different proteins (Doolittle, 1995), and it is assumed that all these will have homologous tertiary structures, although the exact function of the domain may differ. Domain shuffling has occurred throughout the course of protein evolution. It has been proposed that the movement of these domains from the original gene was due to exon insertion and duplication (Doolittle, 1995). There is a widespread, but mistaken, belief that each exon encode a single domain. The shuffling of domains of recently-evolved proteins has been greatly promoted by introns, but this does not imply that domainal rearrangements involve introns. Many evolutionarily-mobile domains, while genetically flanked by introns, also contain internal introns that do not participate in the rearrangement.

Identification of domains from sequence analysis has become a productive research area as the number of protein sequences in the databases far exceeds the number of experimentally determined protein structures, by nearly two orders of magnitude. The problem is complicated because amino acid sequences are eroded during evolution at a much faster rate than three-dimensional structures (Doolittle, 1995). Often domains are identified from the arrangement of cysteine residues in the sequence, but as more sophisticated analysis is developed even domains without any disulphide bridges can be identified (Edwards and

Perkins, 1996). Many domains have been named on the basis of the first protein in which they were found, arbitrarily by function, by acronyms, or they may be known by several names. Since domains usually have a stable structure when isolated from the multi-domain protein, the three-dimensional structure can be determined by multidimensional NMR or X-ray crystallography, and the structure of homologous domains in other proteins can be predicted (Šali *et al.*, 1990).

The proteins which constitute the complement cascade are mostly multi-domain proteins composed of a number of well-defined autonomously folding domains (Figure 1.2). The proteins may consist entirely of the same type of domain (factor H), or have a mosaic structure, where different domain types are present. Although some domains in extracelluar proteins may have a universal function, it is generally thought that domains acquire different functions depending on their structural setting (Bork *et al.*, 1996).

The solution structures of the SCR, EGF and EGF-like domains, LDLr, C3a and C5a, and the extracellular domain of CD59 have been determined by multi-dimensional NMR spectroscopy. The crystal structures of the EGF, LDLr, SP, SERPIN, pentraxin, C3a and C3d structures have been solved by X-ray diffraction studies. The structures of some other complement protein domains have been analysed from low resolution structural studies such as Fourier transform infra-red spectroscopy (FT-IR), electron microscopy and solution scattering and also sequence analysis and molecular graphics studies. These include: the thrombospondin repeat (TSR); perforin-like region (PLR); factor I/ membrane attack complex (FIMAC) domain; CUB domain; clusterin domain; C1q domain; and the C-type lectin domain. Some of the domain types associated with the complement proteins are discussed below and a summary is presented in Table 1.4.

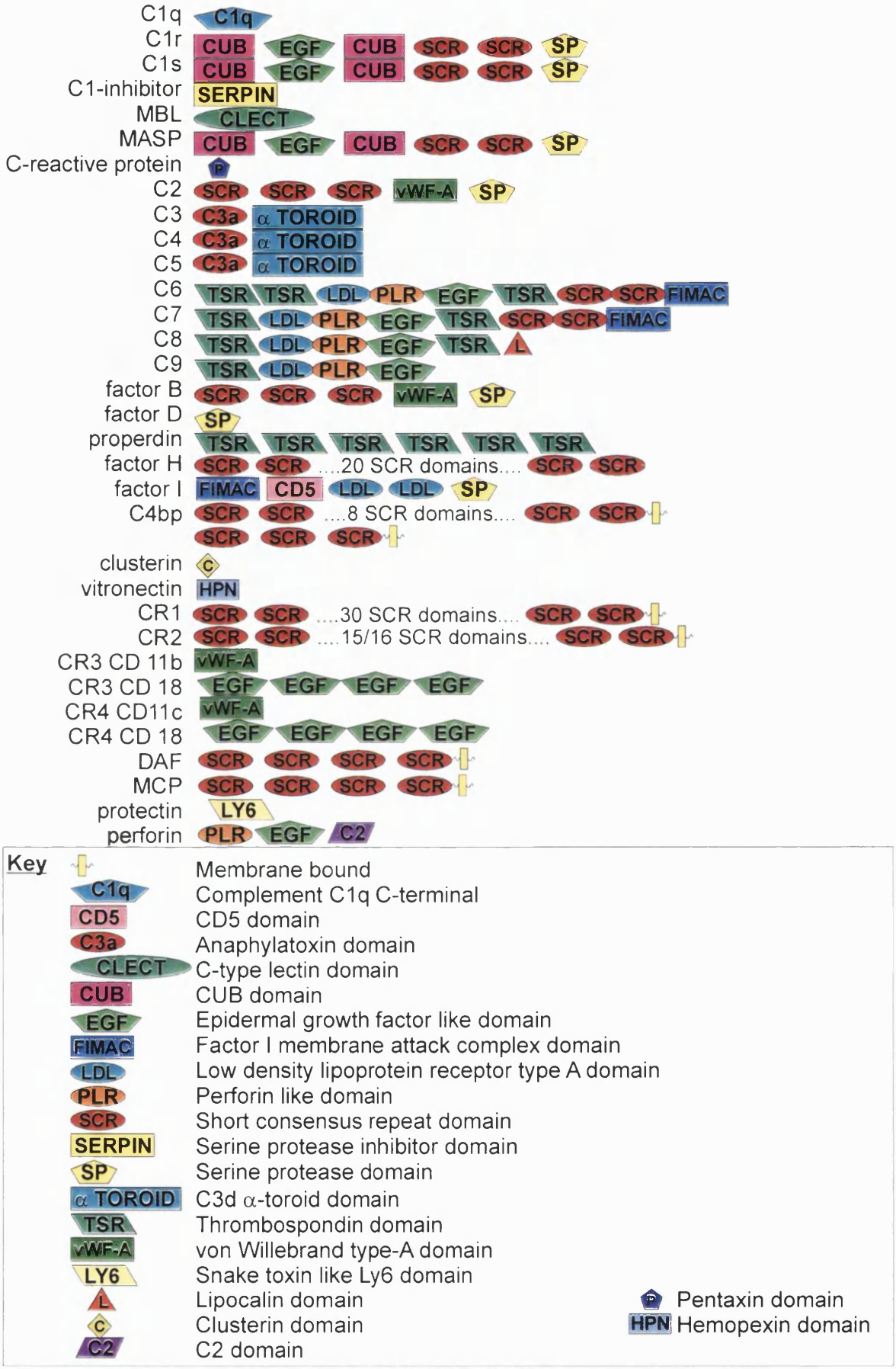


Figure 1.2 Summary of the domain types that occur in complement protein structures. The domain representations are not to scale and do not relate to actual domain sizes.

Table 1.4 Domain types found in complement proteins

Domain type	Typical function(s)	Occurrence in complement proteins	Size*
Short consensus repeat (SCR)	mediate protein interactions	C1r, C1s, MASPs, C2, C6, C7, CR1, CR2, factor H, C4bp, DAF, MCP, factor B	60
von Willebrand factor type A (vWF-A)	adhesive function	C2, factor B, CR3 (Corbi *et al.*, 1988) and CR4 (Corbi *et al.*, 1987)	200
Serine protease (SP)	proteolytic	C1r,C1s, MASP, C2, factor B, factor D, and control factor I (Perkins and Smith, 1993)	250
Anaphylatoxin	mediators of the local inflammatory process (Morgan *et al.*, 1985)	C3a, C4a and C5a (Huber *et al.*, 1980)	75
Snake toxin or Ly6	inhibitor of complement-mediated lysis	CD59 (Kieffer *et al.*, 1994)	100
Epidermal growth factor (EGF)	Ca^{2+} binding	C1r, C1s, C6, C7, C8α, C8β, C9 and MBP	40
factor I/membrane attack complex (FIMAC)	unknown	factor I, C6 and C7	74
C3d	contains reactive thiolester	C3, C4, C5	300
LDL receptor (LDLr)	binding site (Scott, 1989; Tschopp & Masson, 1987).	factor I, C6, C7, C8a, C8b and C9 (Goldberger *et al.*, 1987; Catterall *et al.*, 1987)	40
CD5	antigen binding (Van de Velde *et al.*, 1991)	factor I	100
Thrombospondin (TSR)	protein interactions (Holt *et al.*, 1990; Prater *et al.*, 1991).	properdin, C6, C7, C8α (Rao *et al.*, 1987) C8β (Haefliger *et al.*, 1987), C9 (DiScipio *et al.*, 1984)	60
Perforin (PLR)	formation of pore-like lesions	C6, C7, C8α, C8β and C9 (Lichtenheld & Podack, 1989)	360
Serine protease inhibitor (SERPIN)	extracellular irreversible serine protease inhibitors (Carrell & Travis, 1985)	C1-inhibitor	400 to 500
CUB	functionally diverse, mostly developmentally regulated proteins	C1s/C1r and MASPs (Bork, 1991; Bork and Beckmann, 1993)	110
Clusterin	bind to cells, membranes, and hydrophobic proteins and associated with programmed cell death	complement-associated protein (SP-40,40)	2 × 200
Pentraxins	calcium binding, complement fixation and acute phase response	C-reactive protein	
C1q	activates the serum complement system	C1q	136
C-type lectin	calcium-dependent carbohydrate recognition (Drickamer, 1993)	Mannan (mannose)-binding proteins (MBP)	110 to 130
Lipocalin	transport small hydrophobic molecules (Igaraishi *et al.*, 1992; Godovac-Zimmermann 1988;)	C8 gamma chain (Haefliger, 1991)	
Hemopexin	binds heme, cell adhesion	Vitronectin (Stanley, 1986)	200
C2	calcium-dependent phospholipid binding (Davletov et al., 1993)	perforin (Brose et al.,1995)	116

Footnote: * The size relates to the typical number of residues within the domain.

1.7.2 Short consensus repeat domain

Many complement regulatory proteins consist almost entirely of multiple, tandemly arranged, stretches of residues that share a consensus sequence known as a short consensus repeat (SCR) (Reid and Day, 1989). This consensus sequence has been identified in many other complement proteins and in some non-complement proteins (Figure 1.2).The SCR structure occurs over 140 times in more than 20 extracellular proteins including 12 complement proteins (Perkins *et al.*, 1988; Barlow *et al.*, 1991) and is thought to mediate protein-protein interactions. There are four invariant cysteines, several glycines and prolines and a single tryptophan along with other highly conserved residues. The solution structures of the 5th, the 16th and the 15th and 16th pair of SCR repeats of factor H have been studied by NMR (Norman *et al.*, 1991; Barlow *et al.*, 1991, 1992). These three individual domains have very similar backbone structures indicating that the consensus sequence can dictate a consensus structure. The SCR structure is an independently folded domain with four β-strands on one face and two β-strands on the other, and the region between the six β-strands is composed of turns and loops (Norman *et al.*, 1991; Barlow *et al.*, 1991,1992). This structure comprises a β-sandwich arrangement whereby hydrophobic side chains contribute to a hydrophobic core, and two intradomain disulphide bonds at each end of the elongated domain structure (between the first and the third cysteine and between the second and fourth cysteine) stabilise the structure surrounding this hydrophobic core.

1.7.3 von Willebrand factor type A (vWF-A) domains

The von Willebrand factor type A (vWF-A) domain is widespread and is the prototype for a superfamily of at least 75 related sequences, many of which have been implicated in binding to collagen (Colombatti and Bonaldo, 1991; Perkins *et al.*, 1994). It was termed the vWF-A domain since it corresponded to one of the four different domain types (A to D) in von Willebrand Factor (Sadler *et al.*, 1985; Jenkins *et al.*, 1998). Other groups have termed this domain as an A-domain to correspond to its "adhesive" properties or as an "inserted" or I-domain as it occurs in some integrins. It occurs in the complement proteins C2, factor B, CR3 (Corbi *et al.*, 1988) and CR4 (Corbi *et al.*, 1987), LFA-1, cartilage matrix protein (Kiss *et al.*, 1989) and the collagen types VI (Chu *et al.*, 1990), VII,

XII and XIV. It also occurs in many cell surface receptor proteins such as the integrins (Takada and Hemler, 1989). It is approximately 200 residues long (25 kDa) and there are no conserved cysteine residues. The vWF-A protein fold was first identified by structure prediction analyses (Edwards and Perkins, 1995; 1996). Crystal structures of the vWF-A domain of complement receptor type 3 (CR3; CD11b), leukocyte integrin lymphocyte function associated antigen-1 (LFA-1; CD11a), very late activation protein receptor-2 (VLA-2; CD49b) which is the collagen binding site in the integrin α2 β1 (GPIa/IIa), the von Willebrand factor A1 domain and the A3 domain have since been determined to confirm this prediction (Lee *et al.*, 1995a, 1995b; Baldwin *et al.*, 1998; Qu & Leahy, 1995, 1996; Emsley *et al.*, 1997, 1998; Celikel *et al.*, 1998; Huizinga *et al.*, 1997; Bienkowska *et al.*, 1997). The salient features of this domain are that it resembles a dinucleotide-binding structure, or Rossmann fold, with seven amphiphatic α-helices surrounding a central hydrophobic, mostly parallel β-sheet. The integrin vWF-A domains contain an unusual Mg^{2+} coordination site described as the metal ion dependent adhesion site (MIDAS) which is critical for the adhesive function. Binding of the divalent cation Mg^{2+} appears to induce conformational changes within the integrin vWF-A domain structure (Lee *et al.*, 1996; Qu and Leahy, 1995), although this has recently been disputed (Baldwin *et al.*, 1998). Sequence alignments suggest that the vWF-A1 and A3 domains contain a MIDAS site, but crystal structures show that the motif is modified and no metal ion is bound (Bienkowska *et al.*, 1997; Emsley *et al.*, 1998).

1.7.4 Serine protease domain

The serine protease domain is widespread and its archetype is represented by the digestive enzymes β-trypsin and α-chymotrypsin. The three-dimensional structure is well characterised to consist of two subdomains, each constructed from a six-stranded antiparallel β-sheet barrel structure (Chothia and Janin, 1982). The sub-domains are linked by a polypeptide hinge and by one disulphide bridge (chymotrypsinogen A) or two disulphide bridges (trypsinogen). The catalytic site is located between the two sub-domains. It is the functional domain involved in enzymatic cleavage of peptide bonds. The catalytic triad (histidine, aspartic acid, serine) involved in the active site is conserved amongst all serine

proteases, but disulphide bridge connectivity is not fully conserved. The nucleophilicity of the serine residue is increased by an electronic rearrangement of the catalytic residues within the structure. All the enzymes participating in the major steps concerned with complement activation or control (C1r,C1s, MASPs, C2, factor B, factor D, and factor I) belong to the family of mammalian serine proteases (Johnson *et al.*, 1984; Perkins and Smith, 1993). On the basis of their primary structure, complement SP domains belong to the chymotrypsin family, and on the basis of their specificity for Arg residues, to the trypsin superfamily. The SP domains of the complement system, as with those of the blood clotting system, often differ from the broadly specific digestive serine proteases by having a large N-terminal polypeptide chain (as distinct from a small activation peptide) added to the domain.

1.7.5 LDL receptor (LDLr)

The LDLr repeat type A was first identified in the LDL receptor protein. The LDLr domain is found as two contiguous repeats in the factor I sequence and one repeat in terminal components C6, C7, C8α, C8β and C9 (Goldberger *et al.*, 1987; Catterall *et al.*, 1987). It also occurs in LDLr as 7 repeats, in LDLr-related protein as 31 repeats and in GP330 as 15 repeats. The LDLr is a cysteine-rich domain (Yamamoto *et al.*, 1984), containing 6 conserved cysteines , with the exception of the first LDLr domain of factor I which contains only four cysteines (Catterall *et al.*, 1987). It is approximately 40 amino acids in length and has substantially more residue conservation than the other complement domains. The homologous domain in the LDLr acts as a binding site and it has been proposed that it has the same function in the complement proteins (Scott, 1989; Tschopp and Masson, 1987). The repeat has been shown (Daly *et al.*, 1995) to consist of a β-hairpin structure followed by a series of β- turns. The binding of calcium seems to induce no significant conformational change.

1.7.6 Factor I/membrane attack complex (FIMAC) domain

The factor I/membrane attack complex (FIMAC) domain is a cysteine-rich repeat of consensus length 74 amino acids (Ullman and Perkins, 1997). It was first identified by sequence comparisons of the N-terminal domain in complement factor I with the two C-

terminal domains in C6 and C7 of the complement membrane attack complex (Catterall *et al.*, 1987; Goldberger *et al.*, 1987; Haefliger *et al.*,1989; DiScipio and Hugli, 1989). This domain has unknown function in the factor I-mediated cleavage of C3b and C4b or the initial formation of the membrane attack complex by C5b, C6, C7, C8 and C9. Both the disulphide arrangement and the structure of the FIMAC domain are unknown. Database searches have revealed a distant sequence similarity with the follistatin superfamily of connective tissues and the endocrine system. This similarity was supported by consensus secondary structure predictions for both superfamilies (Ullman and Perkins, 1997). Concurrently, the crystal structure of the follistatin domain in SPARC protein (Secreted Protein Acidic Rich in Cysteines) (also known as BM-40 or osteonectin) of the extracellular matrix was determined (Hohenester et al., 1997). This turned out to be a hybrid of two other domains, the epidermal growth factor and the ovomucoid protease inhibitor. Several extracellular matrix proteins contain FS domains whose structures have been predicted by homology to the BM40 FS domain (Bork *et al.*, 1996). Follistatin contains three tandem repeats and agrin contains nine repeats of this domain. The follistatin domain is elongated and consists of an N-terminal β-hairpin and a small core of α/β structure stabilised by a small hydrophobic core and by a total of five disulphide bonds (Hohenester *et al.*, 1996). The domain is assembled from two rather weakly interacting substructures, a highly twisted N-terminal β-hairpin and a pair of antiparallel α-helices connected to a small three-stranded β-antiparallel sheet. This subdivision is represented in a disulphide linking pattern of two non-overlapping sets (1-3, 2-4 and 5-9, 6-8, 7-10).

1.7.7 The CD5 domain

Amongst the complement proteins the CD5 domain is found solely in the sequence of factor I. Elsewhere, this domain is found in membrane proteins including the macrophage scavenger receptor (Freeman *et al.*, 1990), in differentiation antigens CD5 and Ly-1 located on T-lymphocytes of human origin and specialized B-lymphocytes of mouse origin (de Bruijn and Fey, 1985; Huang *et al.*, 1987), and in the speract receptor of sea urchin sperm (Dangott *et al.*, 1989). The extracellular part of the CD5 molecule is composed almost entirely of CD5 modules and may be responsible for binding to CD72, the human B-cell differentiation antigen

(Van de Velde *et al.*, 1991). In factor I the domain is approximately 100 residues long and contains 6 conserved cysteines (Kunnath-Muglia *et al.*, 1993).

1.7.8 Serine protease inhibitor (SERPIN)

C1 inhibitor is a member of the SERine Protease INhibitor (SERPIN) superfamily (Carrell and Travis, 1985). SERPINs are high molecular mass (400 to 500 amino acids), extracellular, irreversible serine protease inhibitors with a well defined structural-functional characteristic: a reactive region that acts as a "bait" for an appropriate serine protease. Structural comparisons with the crystal structure of the homologous protein α_1-antitrypsin show conservation of those features important in inhibitory function and in maintaining tertiary structure (Bork *et al.*, 1986).

1.7.9 C3d domain

C3d is a 35 kDa fragment of C3 which has sequence homology to regions of C4, C5 and α-2 macroglobulin. It has an α–α barrel structure which also occurs in glucoamylase, endoglycosylase and the β subunit of farnesyltransferase (Naga *et al.*, 1998). Consecutive helices alternate from outside to inside resulting in a core of six parallel helices surrounded by a second set of six parallel helices.

1.7.10 Anaphylatoxin domain

Anaphylatoxins (Morgan *et al.*, 1985) are mediators of the local inflammatory process that act by inducing smooth muscle contraction. There are three different anaphylatoxins: C3a, C4a and C5a. They are peptides of about 75 amino-acid residues that are derived from the proteolytic degradation of complement C3, C4 and C5 and which contain six disulphide-bonded cysteines (Huber *et al.*, 1980). This cysteine-rich region shares similarity with a three-times repeated domain found in the mammalian extracellular matrix proteins fibulins 1 and 2 (Argraves *et al.*, 1990; Pan *et al.*, 1993). The three disulphide bonds are conserved in the first and last repeats, but the first disulphide bond is missing in the second repeat.

Chapter 2

Methods for Protein Expression, Purification, Characterisation and Crystallisation

2.1 Expression methods

Proteins have many crucial biological functions which include enzyme catalysis, transport and storage, movement, mechanical support, immune protection, signal transduction and growth and differentiation (Stryer, 1995). The study of proteins is one of the most fundamental exercises in biochemistry, whether it be to discover the effect a protein has on a whole cell or its structure at atomic resolution. The problem faced by biochemists over the years has been the isolation of sufficient amounts of the protein of interest for functional and structural studies.

Proteins may exist in minute quantities within cells and organisms and it is then not always practical to purify these proteins from the natural host in quantities sufficient to determine their structure and function. With the advent of *in vitro* manipulation of deoxyribonucleic acid (DNA), the ability to transfer genetic information that encodes for proteins between organisms has become relatively easy via a variety of methods. The adaptation of these techniques, and a more thorough understanding of the molecular biology of the cell of both prokaryotic and eukaryotic organisms, has led to the advent of foreign genes, or parts thereof, being expressed in heterologous systems and the resulting products characterised (Alberts *et al.*, 1989). However, these systems are extremely protein specific and the nature of the protein and the use of the final product must be taken into consideration. The available systems make use of both eukaryotic cells (including vertebrate cells, insect cells, plant cells and fungi) and prokaryotic cells. In each case, the level of transcription of the recombinant gene and ultimately the level of expression of the recombinant protein, is dependent upon conserved sequences of DNA known as promoters, located on the 5' side (upstream) of the cloned sequence (although factors such as the stability of the mRNA and the protein within the cell are also important). RNA polymerase and associated transcription factors bind to the promoter to initiate synthesis of RNA which is subsequently translated by ribosomes in the cytoplasm. The rate of initiation is proportional to the strength of the binding of the RNA polymerase. Promoter sequences that mediate strong binding of RNA polymerase, and therefore have a higher rate of RNA synthesis, are known as strong promoters. RNA synthesis that is initiated in the presence of an external stimulus is said to be

controlled by an inducible promoter, whereas constant synthesis is due to constitutive initiation. Comparison of *E. coli* promoter regions has shown that two areas within this region are highly conserved. Both are hexamers with the first region being 35 base pairs (bp) upstream and the second 10 bp upstream (the Pribnow box) from the start of transcription. The spacing of these two regions is important in the regulation of transcription levels as well as the regulation of transcription initiation. The resulting transcribed RNA is translated into an amino acid chain by ribosomal assemblies (Alberts *et al.*, 1989). There are significant differences in RNA processing between prokaryotes and eukaryotes which will not be discussed here.

2.1.1 *E. coli* as a host

Bacterial expression systems are advantageous for the production of large quantities of heterologous proteins from a small quantity of inexpensive media. Both Gram-positive and Gram-negative bacteria have been engineered. The Gram-positive bacteria, such as *Bacillus brevis*, secrete proteins into the culture medium. Expression-secretion vectors that use the signal peptide sequences of the cell wall proteins (middle wall protein and outer wall proteins) have been developed and used successfully (Udaka and Yamagata, 1993). Although product solubility and purification may cause problems, the Gram-negative bacterium *E. coli* has been an invaluable organism for the over-expression of recombinant proteins. A multitude of expression systems exists. These systems can express recombinant proteins as either discrete domains or in fusion with a highly expressed protein. The latter enables higher levels of solubility to be reached in some cases, and can ease purification and detection of the expressed product. Careful selection of the *E. coli* host strain and expression system can give protein levels of up to 25% of the total cell protein (Marston, 1986).

The three main requirements for expression of foreign genes in *E. coli* are: (i) the gene must not be interrupted by intervening sequences; (ii) the gene must be placed under the control of an *E. coli* promoter that is efficiently recognised by *E. coli* RNA polymerase; (iii) the transcribed mRNA must be relatively stable and efficiently translated.

2.1.2 *E. coli* plasmids and promoters

Transfer of the recombinant gene into *E. coli* is usually via a plasmid which is an independent genetic element capable of replicating outside the host genome. As many bacterial promoter regions are relatively weak, plasmids have been engineered that possess stronger promoters, and thereby greater controls have been introduced in the initiation of transcription. The plasmids used for expression of recombinant proteins in *E. coli* are often derived from the general cloning vector pBR322 (Bolivar *et al.*, 1977), which contains an origin of replication as well as an antibiotic resistance gene. This resistance enables the selection of host cells carrying the plasmid when they are grown on a medium containing the respective antibiotic. Promoters such as the *lac*UV5 (*lac*), *trp*, *tac*/*trc* (*trp-lac* hybrid promoters), the λ phage P_L promoter and the T7 phage promoter are now used to control and initiate recombinant gene expression in *E. coli*. These promoters are essential as eukaryotic promoters have very poor function in *E. coli*. Many of these promoters are chosen because they can be regulated. If the recombinant gene is toxic to the host upon expression, coupling to a strong unregulated promoter is not recommended. High levels of constitutive transcription have been shown to interfere with plasmid DNA replication and lead to plasmid instability (Remaut *et al.*, 1987).

2.1.3 *trp*/*lac* hybrid promoters

The *trp* and *lac* promoters have been used to direct expression of foreign gene products within *E. coli* (Danley *et al.*, 1989). By themselves these promoters are relatively weak, but by fusing the -35 region of the *trp* promoter and the -10 region of the *lac*UV5 promoter along with the *lac* operator, the *trp-lac* hybrids (known as the *tac* and *trc* promoters) can be made relatively strong. Regulation can then be controlled by the addition of isopropyl-β-D-thiogalactoside (IPTG) (Amann *et al.*, 1985; de Boer *et al.*, 1983). What distinguishes the two types of hybrid promoter from each other is that the nucleotide spacing in between the -35 and -10 regions is 17 for the *trc* promoter and 16 for the *tac* promoter (Amann, 1988; Amann, 1985; DeBoer, 1983). Both promoters have the same relative strength, being up to eleven times more efficient than *lac*UV5 and three times more efficient than *trp* promoters alone (DeBoer, 1983). Many expression vectors are based on these

promoter hybrids including the pGEX expression systems.

2.1.4 Expression systems for fusion proteins

Direct expression of foreign polypeptides within *E. coli* can be limited as the protein may be degraded by the host cell. This difficulty is most apparent with small peptides (Marston, 1986). Other problems associated with direct expression are poor solubility and the difficulty of subsequent purification of the protein product. In an attempt to circumvent these problems, fusion of the sequence with a highly expressed non-toxic protein has been used. A number of fusion systems exist for *E. coli* which use such proteins (e.g. β-galactosidase, chloramphenicol acetyltransferase, glutathione transferase, maltose binding protein, phosphate binding protein, protein A, protein G, streptavidin and thioredoxin) or synthetic peptides (e.g. poly-Arg, -Glu or -His residues). These fusion partners may be placed either N-terminal or C-terminal to the recombinant gene sequence (Uhlén and Moks, 1990).

2.1.5 The pGEX expression system

The *E. coli* pGEX bacterial expression system is versatile for high yields of recombinant protein. Preparation of the gene constructs and generation and growth of recombinants are less labour intensive than other expression systems so this system is ideal for studies of protein domains. The pGEX expression system has been widely used and citations in the literature describe the versatility of the system (Table 2.1). The pGEX expression system is based upon the inducible expression of the 26 kDa *Schistosoma japonicum* glutathione *S*-transferase (GST) protein (Sj26) within bacterial cells (GST; E.C. 2.5.1.18). The gene encoding the protein is carried on a plasmid vector with expression of GST under the control of the *tac* promoter to provide chemically inducible high-level expression. Unique restriction endonuclease sites exist at the 3' end of the GST gene to allow the insertion of heterologous gene sequences.

2.1.6 pGEX expression plasmids

The plasmid pGEX-1 contains the *tac* promoter followed by the complete coding sequence of Sj26 in which the normal termination sequence is replaced by a

Table 2.1 Examples of recombinant proteins expressed in various *E.coli* expression systems

system	recombinant protein	references
λP_L promoter	SV40 small antigen	Derom *et al.*, 1982
	human interferon-γ	Remault *et al.*, 1987
	human tumour necrosis factor	Remault *et al.*, 1987
	human interleukin 2	Remault *et al.*, 1987
T7 promoter	HIV-1 protease	Cheng *et al.*, 1990
	HSV-1 ribonucleotide reductase small unit (R2)	Yang *et al.*, 1991
	human retinoic acid receptor	Lankinen *et al.*, 1991
	papain	Taylor *et al.*, 1992
tac promoter (pKK vector)	hepatitis B pre-S antigen	Amann and Brosius, 1985
	chicken triose phosphate isomerase	Straus and Gilbert, 1985
	protein-tyrosine phosphatase	Hashimoto *et al.*, 1992
	high molecular mass kininogen	Kunapuli *et al.*, 1992
lipoprotein secretion system	human growth hormone	Duffaud *et al.*, 1987
	human galactosyltransferase	Aoki *et al.*, 1990
phosphatase secretion system	anti-digoxin antibody variable region fragment	Anthony *et al.*, 1992
	human epidermal growth factor	Oka *et al.*, 1985
Staph. protein A secretion system	insulin-like growth factors I and II	Löwenadler *et al.*, 1987 Uhlén and Abrahamsen, 1989
β-galactosidase fusion	influenza virus haemagglutinin	Heiland and Gething, 1981
	human interleukin-6	Snouwaert *et al.*, 1991

polylinker containing unique sites for the restriction enzymes *Bam* HI, *Sma* I and *Eco* RI. This is followed by a TGA translation termination codon in all three reading frames. pGEX-1 also contains the gene encoding β-lactamase that enables ampicillin resistance, along with the pB322 ori (origin of DNA replication), and the fragment of the *lac* operon which contains the over-expressed $lacI^q$ allele of the *lac* repressor and part of *lacZ*. The existence of the *lac* operon fragments allows the tight control of transcription in host *E. coli* bacteria that do not encode this type of repressor in their genome (Smith and Johnson, 1988). Based on pGEX-1, two other plasmids were assembled; these vectors, pGEX-2T and pGEX-3X shift the reading frame of the multiple cloning site (MCS) by one nucleotide (Smith and Johnson, 1988). The pGEX-2T and pGEX-3X vectors contain a thrombin and Factor Xa cleavage site, respectively, at the C-terminus of the GST protein (Figure 2.1).

2.1.7 Purification and isolation of protein products

By exploiting the ability of GST to bind to glutathione, a simple purification procedure for this expression system has been created. Affinity chromatography using immobilised glutathione enables GST fusion proteins to be purified under non-denaturing conditions (Smith and Johnson, 1988). The GST fusion protein is absorbed onto the resin of immobilised glutathione and most of the non-specific bacterial proteins remain in solution. By removing the mobile phase and washing the beads with fresh buffer, a high proportion of the product recovered will be the GST fusion protein. The purified fusion protein can be eluted from the affinity resin by competition with free reduced glutathione. As glutathione has a low relative molecular mass, it is easily dialysed away, and a relatively pure protein preparation can be achieved with little effort (Smith and Johnson, 1988). Alternatively, the fusion protein can be cleaved on the immobilised glutathione matrix (Abath and Simpson, 1991; Olsen and Mohapatra, 1992). Cleavage of the fusion protein by the appropriate protease separates GST from the recombinant domain (Smith and Johnson, 1988). pGEX-2T and pGEX-3X contain within the polylinker sequence specific cleavage sites for thrombin and factor X_a respectively. Inclusion of the protease cleavage sites enables the removal of the GST portion from the fusion protein after isolation by affinity chromatography.

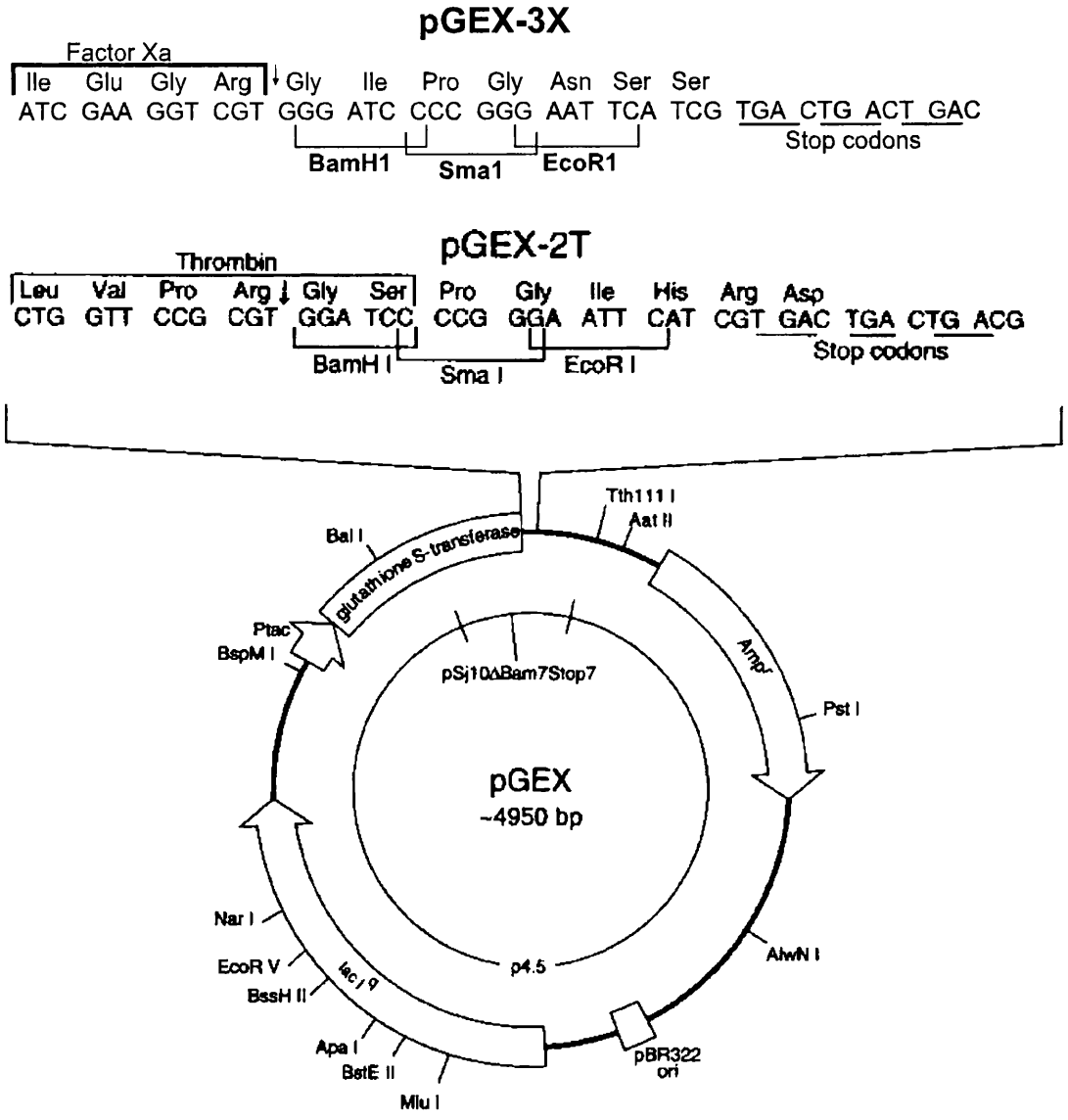


Figure 2.1 Map of the glutathione-S-transferase fusion vectors pGEX-2T and pGEX-3X showing the reading frames and main features.

Table 2.2 Examples of the uses of GST Fusion Proteins

Function	Description	Reference
immunological studies	development of an ELISA sandwich technique using immobilized glutathione	Rabin *et al.*, 1992
	epitope mapping of HPV-33 minor capsid protein epitopes	Volpers *et al.*, 1993
	protection of sheep against larval tapeworms	Johnson *et al.*, 1989
	as antigens to raise antibodies against proteins to determine orientations and locations of proteins	Baines *et al.*, 1993 Tomlinson *et al.*, 1993 Vanscheeuwijck *et al.*, 1993
studies of protein-nucleic acid interaction	recombinant GST:GCN4 (yeast DNA-binding protein) used in detection of HIV in blood samples	Kemp *et al.*, 1989
	GST:HIV p15 protein (containing zinc binding domains) stimulated dimerization of viral DNA	Weiss *et al.*, 1992
	identification of the specific DNA consensus sequence bound by GST:N-Myc protein	Alex *et al.*, 1992
studies of protein-protein interactions	screening of cell lysates for proteins that interact with GST:pRb	Kaelin *et al.*, 1991
	interactions of SH2 domains with tyrosine-phosphorylated peptides of PDGF β-receptor	Panayotou *et al.*, 1993
structural studies	crystals of *E. coli* F_1 portion of ATPase	Codd *et al.*, 1992
	NMR structure of the T lymphocyte CD2 antigen	Driscoll *et al.*, 1991
	X-ray crystal structure of the integrin CR3 type -I domain	Lee *et al.*, 1995a 1995b

2.1.8 Suitability of pGEX for protein expression

Due to the specificity of the affinity interaction involved within this system, single step purifications can achieve more than 90% purity (Smith and Corcoran, 1995). Normally, the products of the pGEX expression system are soluble under non-denaturing conditions and the affinity of GST for immobilised glutathione is not reduced when the fusion protein is solubilised using detergents (Smith and Johnson, 1988; Frankel *et al.*, 1991). Because of this, the foreign polypeptide may retain native structure as well as functional activity. If stronger conditions are needed for solubilisation of the fusion protein, then the affinity is affected and poor or no binding to the resin is observed (Frangioni and Neel, 1993). Insolubility of fusion proteins has been associated with the presence of either strongly hydrophobic regions, or a high proportion of charged residues or with proteins of sizes greater than 100 kDa (Smith and Johnson, 1988).

2.2 Purification of protein preparations

For structural studies of proteins the preparations have to be of high purity, generally more than 95%. Crude purifications are possible by fractionation with one or more precipitants (either by addition of salts, organic solvents, organic polymers), or by physical treatments such as changing pH or temperature in order to decrease the solubility or denature unwanted macromolecules. It is also possible to sediment macromolecules by ultracentrifugation and isolate the protein of interest at a specific position in the supernatant (size exclusion). High purity generally involves more resolute separation techniques. Often the first stage in the purification scheme may involve an affinity absorbent whereby a specific ligand is immobilised on a matrix. There are many choices for the further purification of a protein sample which depend upon the chemical and physical characteristics of the proteins concerned. These methods rely upon the charge of the molecule (ion-exchange), the isoelectric point (chromatofocusing), the hydrophobicity of the molecular surface (hydrophobic interaction and reverse phase). Electrophoretic methods are also employed for protein purification in preparative SDS-PAGE and isoelectric focussing. Size exclusion chromatography or gel filtration is often used as the final stage of purification.

2.3 Characterisation of the preparation

Gel electrophoresis is the most convenient method to visualise quickly the macromolecular content of a sample. Proteins are separated by virtue of their size using polyacrylamide gel electrophoresis (PAGE) with sodium dodecyl sulphate (SDS). The proteins migrate at a rate inversley proportional to their molecular masses under the influence of an electric current and are visualised by a number of methods. Isoelectric focussing is a technique whereby the isoelectric point of the proteins in a sample may be visualised. In Western blotting the proteins from the gel can be electrophoretically transferred to a membrane and probed with specific antibodies for identification. Visualisation is achieved by the use of a second antibody enzyme conjugate that catalyses a colour change in a substrate. Blotting can be employed to identify specific properties of the proteins such as glycosylation. Further characterisation may require amino acid sequence determination and disulphide mapping. An important aspect of the characterisation is the functionality of the protein preparation. This would ideally be demonstrated *in vitro* and *in vivo*. Spectrophotometric analyses of the protein sample may be useful to determine the purity and concentration.

2.3.1 Mass spectrometry

Mass spectrometers use the difference in mass-to-charge ratio (*m/z*) of ionized atoms or molecules to separate them from each other. Mass spectrometry is therefore useful for identification of atoms or molecules and also for determining chemical and structural information about molecules. Molecules have distinctive fragmentation patterns that provide structural information which can be used to identify structural components. The general operation of a mass spectrometer is to create gas-phase ions, to separate the ions in space or time based on their mass-to-charge ratio, and to measure the quantity of ions of each mass-to-charge ratio. The ion separation power of a mass spectrometer is described by the resolution, which is defined as:

$$R = \frac{m}{\delta m} \qquad \text{Equation 2.1}$$

where m is the ion mass and δm is the difference in mass between two resolvable peaks in

a mass spectrum (e.g. a mass spectrometer with a resolution of 1000 can just resolve an ion with a *m/z* of 100.0 from an ion with a *m/z* of 100.1).

2.3.2 Instrumentation

In general a mass spectrometer consists of an ion source, a mass-selective analyser, and an ion detector. Since mass spectrometers create and manipulate gas-phase ions, they operate in a high-vacuum system. The magnetic-sector, quadrupole, and time-of-flight designs also require extraction and acceleration ion optics to transfer ions from the source region into the mass analyser. The details of mass analyser designs and basic descriptions of sample introduction and ionization and ion detection are discussed below.

2.3.3 Electron ionization (EI) and chemical ionization (CI)

Electron ionization (EI) is widely used in mass spectrometry for relatively volatile samples that are insensitive to heat and have relatively low molecular mass. An EI source uses an electron beam, usually generated from a tungsten filament, to ionize gas-phase atoms or molecules. The spectra, usually containing many fragment-ion peaks, are useful for structural characterisation and identification, and small impurities in the sample are easy to detect. Chemical ionization (CI) may be used to enhance the abundance of a chosen molecular ion and can be applied to similar samples as EI. For both ionization methods, the molecular mass range is 50 to 800 Da. In rare cases it is possible to analyse samples of higher molecular mass. Chemical ionization (CI) uses a reagent ion to react with the analyte molecules to form ions by either a proton or hydride transfer. The reagent ions are produced by introducing a large excess of methane (relative to the analyte) into an electron impact (EI) ion source.

2.3.4 Electrospray ionization (ESI)

Electrospray ionization (ESI) allows production of molecular ions directly from samples in solution. The ESI source consists of a very fine needle and a series of skimmers. A sample solution is sprayed into the source chamber to form droplets. The droplets carry charge when they exit the capillary. As the solvent evaporates, the droplets disappear leaving highly charged analyte molecules. ESI is particularly useful for large biological molecules that

are difficult to vaporize or ionize. It can be used for small and large molecular-mass biopolymers (peptides, proteins, carbohydrates, and DNA fragments), and lipids. The sample must be soluble, stable in solution, polar, and relatively clean (free from nonvolatile buffers, detergents, salts). Unlike MALDI, which is pulsed, it is a continuous ionization method that is suitable for using as an interface with HPLC or capillary electrophoresis where multiply charged ions are usually produced. ESI should be considered a complement to MALDI.

2.3.5 Laser ionization (LIMS)

A laser pulse can ablate material from a surface and create a microplasma that ionises some of the sample constituents. The laser pulse accomplishes both vaporization and ionization of the sample.

2.3.6 Matrix-assisted laser desorption ionization (MALDI)

Matrix-assisted laser desorption (MALDI) is a LIMS method of vaporizing and ionising large biological molecules such as peptides, proteins, oligonucleotides, and other compounds of biological origin as well as of small synthetic polymers. The macromolecule is typically dispersed in a crystal of a matrix material (Table 2.3). MALDI ion sources all consist of a sample stage (or probe) that is used to carry the analyte into the vacuum system of the spectrometer. A UV laser pulse ablates the matrix which carries some of the large molecules into the gas phase in an ionized form so that they can be extracted into the mass spectrometer. Commercial instruments use laser light in pulses of less than 10 ns from either a nitrogen laser (337 nm) or a *Q*-switched neodymium: yttrium aluminium garnet (Nd: YAG) laser (226 or 354 nm). Empirical evidence has suggested that the mass analyser and detectors work best when the laser energy is near the threshold level required to produce desorption of the macromolecule (Beavis and Chait, 1996). The matrix material is selected for the type of macromolecule under investigation. Most matrix compounds produce satellite signals called adduct peaks at slightly higher mass than the analyte molecule peaks. These result from the photochemical breakdown of the matrix into a more reactive species which can add to the analyte. The best matrices have low intensity photochemical adduct peaks.

MALDI has enabled the analysis of large biomolecules that were previously only amenable to study using conventional techniques (SDS-PAGE, etc.). The amount of sample needed is very low (ρmoles or less). The analysis can be performed in the linear mode (high mass, low resolution) up to a molecular mass of 300,000 (in rare cases) or reflectron mode (lower mass, higher resolution) up to a molecular mass of 10,000 and the analysis is relatively insensitive to contaminants. The mass accuracy (0.1 to 0.01%) is not as high as for other mass spectrometry methods.

2.3.7 Time-of-flight mass spectrometry (TOF-MS)

A time-of-flight mass spectrometer (TOF-MS) uses the differences in transit time through a drift region to separate ions of different masses (Figure 2.2). It operates in a pulsed mode so ions must be produced or extracted in pulses. An electric field accelerates all ions into a field-free drift region with a kinetic energy of qV, where q is the ion charge and V is the applied voltage. Since the ion kinetic energy KE is $0.5mv^2$ (mass m and velocity v), lighter ions have a higher velocity than heavier ions and reach the detector at the end of the drift region sooner:

$$KE = qV \qquad \text{Equation 2.2}$$

$$\frac{1}{2}mv^2 = qV \qquad \text{Equation 2.3}$$

$$v = \sqrt{\frac{2qV}{m}} \quad . \qquad \text{Equation 2.4}$$

The transit time (t) through the drift tube is L/V where L is the length of the drift tube.

$$t = L\left(\sqrt{\frac{m}{2qV}}\right) \qquad \text{Equation 2.5}$$

$$m = 2qV\left(\frac{t}{L}\right)^2 \quad . \qquad \text{Equation 2.6}$$

Table 2.3 Properties of MALDI matrices (from Beavis and Chait, 1996).

Matrix	**Analytes**[a]		**Suggested Solvent (water: organic)**	**Ionization** [b]	**Adduct** [c]
	Peptides	**Proteins**			
gentisic acid, 2,5-dihydroxybenzoic acid	+	+/-	9:1	+	M + 136
sinapic acid, sinapinic acid, *trans*-3,5-dimethoxy-4-hydroxycinnamic acid	+/-	0	2:1	+	M + 206
3-indoleacrylic acid	+	+	2:1	++	M + 185
4-HCCA, a-cyano-4-hydroxycinnamic acid	+	+	2:1	+++	-

Footnotes:

[a] +, Matrix may be used for most peptides and proteins; +/-, matrix may (or may not) work.

[b] The more + signs, the more intense the signal and the higher charge state of the most intense peak.

[c] Expected mass of the most intense satellite peak.

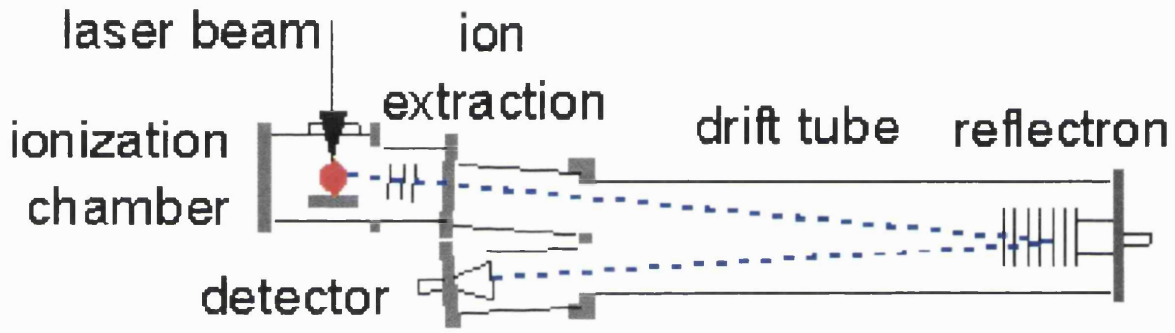


Figure 2.2 This schematic shows ablation of ions from a solid sample with a pulsed laser in a reflectron time-of-flight mass spectrometer (TOF-MS). The reflectron is a series of rings or grids that act as an ion mirror. This mirror compensates for the spread in kinetic energies of the ions as they enter the drift region and improves the resolution of the instrument. The output of an ion detector is displayed on an oscilloscope as a function of time to produce the mass spectrum.

2.3.8 Surface enhanced laser desorption ionization spectrometry (SELDI)

The advent of MALDI-TOF mass spectrometry has allowed the development of new bioanalytical methods which offer accurate characterisation of proteins and peptides. Molecules have distinctive fragmentation patterns that provide structural information. The technique of biomolecular interaction analysis has been interfaced with MALDI-TOF mass spectrometry to provide a powerful method for the investigation of protein interactions (Krone *et al.*, 1997). This has been described as surface enhanced laser desorption ionization (SELDI) (Hutchens and Yip, 1993; Leung *et al.*, 1998).

2.4 Crystallisation

The crystallisation of biological macromolecules is a multi-parametric process involving the three steps of nucleation, growth and cessation of growth. The main difference from small molecule crystal growth arises from the conformational flexibility and chemical versatility of macromolecules and their consequent greater sensitivity to external conditions. For a rational design of growth conditions, physical and biological parameters need to be controlled.

2.4.1 Purity

Purity is not an absolute requirement for crystallisation as crystals can be obtained from mixtures. Poor purity is however unlikely to result in high quality monocrystals of the size required for the purpose of X-ray crystallography. Typically crystals should be of the order of 0.2 mm in size, but the use of synchrotron radiation sources and cryogenics means that crystals of much smaller dimensions may yield data suitable for structural analysis. For structural analysis, the macromolecular sample has to be of high purity, not only in terms of lack of contaminants but also conformationally pure. Denatured macromolecules or macromolecules with structural microheterogeneities adversely affect crystal growth more than unrelated molecules do.

2.4.2 Solubilities and supersaturation

To grow crystals, a solution containing the molecules of interest has to be brought to

a supersaturated, thermodynamically unstable state, which may develop into a crystalline or amorphous phase when it returns to equilibrium. Supersaturation can be achieved by slow evaporation of the solvent or by varying specific parameters. Intrinsic physical and chemical parameters that affect the protein solubility are: the concentration of protein and precipitants; temperature; pH; rate of crystal growth; ionic strength and purity of chemicals; volume and geometry of samples; wall and interface effects; density and viscosity; pressure; electric and magnetic fields; vibrations and sound; and the sequence of events that arrive at a specific set of parameters. Biochemical parameters may include the sensitivity of the protein conformations to physical parameters, the binding of ligands, specific additives such as reducing agents, polyanionic detergents and polyamines. These are also related to the properties of the protein. The source and age of the protein preparation are also parameters that may affect solubility. The solubility of a macromolecule can be represented in a phase diagram as a function of one parameter, all other parameters being constant (Figure 2.3).

The solubility curve divides the under-saturated and supersaturated zones. Under the solubility curve the protein will never crystallise. Above the solubility curve the concentration of the protein is higher than the concentration at equilibrium for the given parameter. This corresponds to the supersaturation zone. The level of saturation is defined as the ratio of protein concentration over the solubility value. Depending on the kinetics to reach equilibrium and the level of supersaturation, this region may be subdivided into three zones. The precipitation zone is where excess protein immediately separates from the solution in an amorphous state. The nucleation zone is where excess protein separates into a crystalline form. Near the precipitation zone, crystallisation may occur as a shower of tiny microcrystals which can be confounded with amorphous precipitate. In a metastable zone a supersaturated solution may not nucleate for a long period of time, unless the solution is mechanically shocked or a seed crystal is introduced. This zone corresponds to the growth of crystals without the nucleation of new crystals.

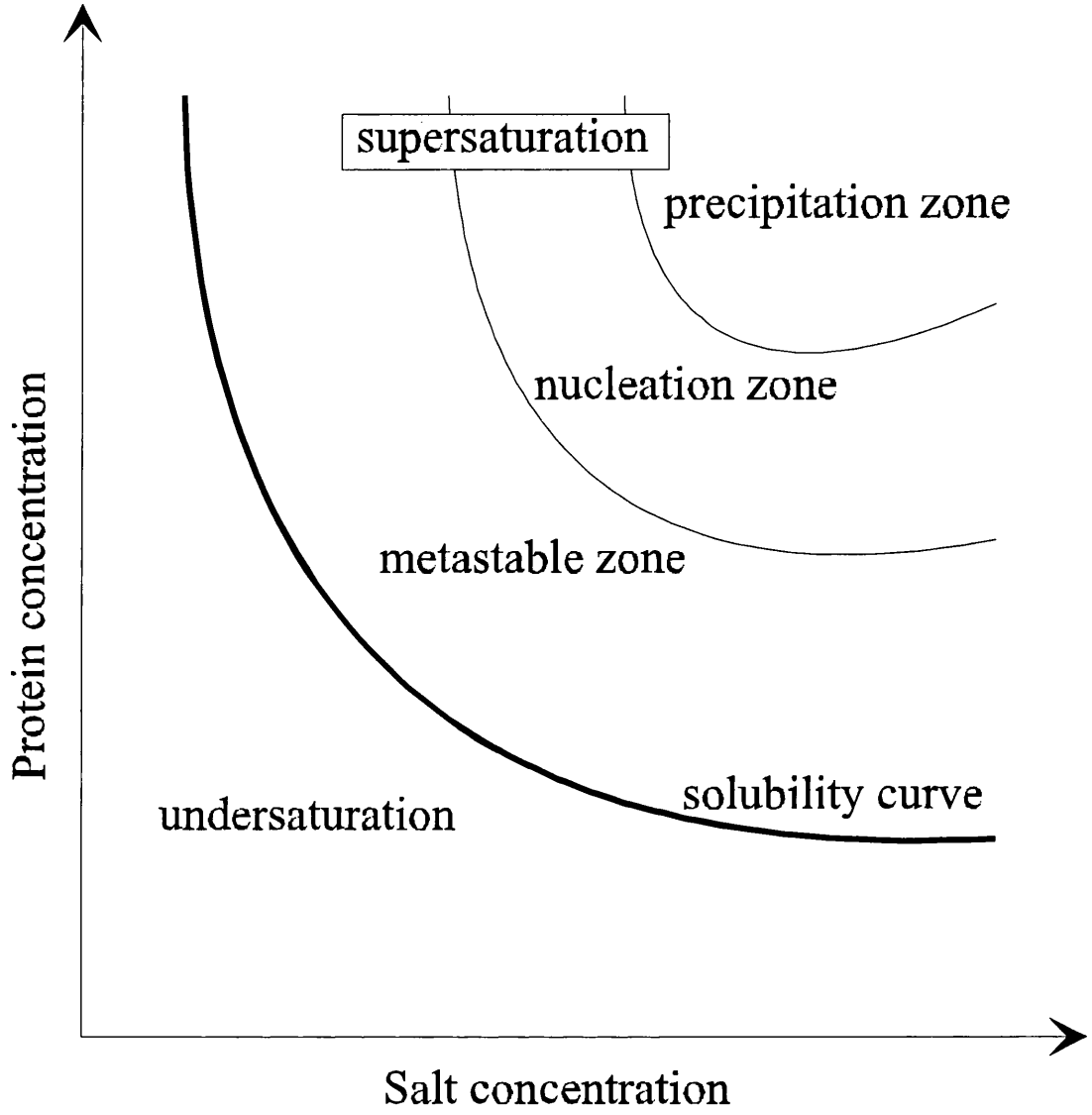


Figure 2.3 Phase diagram representing the solubility of a protein as a function of a single parameter, in this case the salt concentration.

2.4.3 Nucleation and cessation of growth

Proteins require defined pH and ionic strength for stability and function and so crystallisation often requires complex aqueous solutions. Crystallisation starts with a nucleation phase followed by a growth phase and nucleation requires a greater supersaturation than the growth phase. Cessation of growth occurs when the protein has been depleted from the medium, or because of growth defects, poisoning of the faces or ageing of the protein. The crystal quality may be correlated with the packing of the molecules into the crystal lattice and the external crystal morphology. The forces that are involved in the packing of macromolecules are weak compared to those maintaining the cohesion of small-molecule crystals. These forces involve salt bridges, hydrogen bonds, van der Waals, dipole-dipole and stacking interactions. The weak cohesion of macromolecule crystals results from the fact that only small parts of the macromolecular surfaces participate in intermolecular contacts, the remainder being in contact with the solvent. This explains the commonly observed polymorphism of biological macromolecular crystals.

2.4.4 Typical trial arrays

Screening the physical and chemical parameters required to achieve the specific conditions in which a particular macromolecule will crystallise can be a daunting task. Figure 2.4 summarises the systematic approach that was employed in the projects described in this thesis to grow protein crystals.

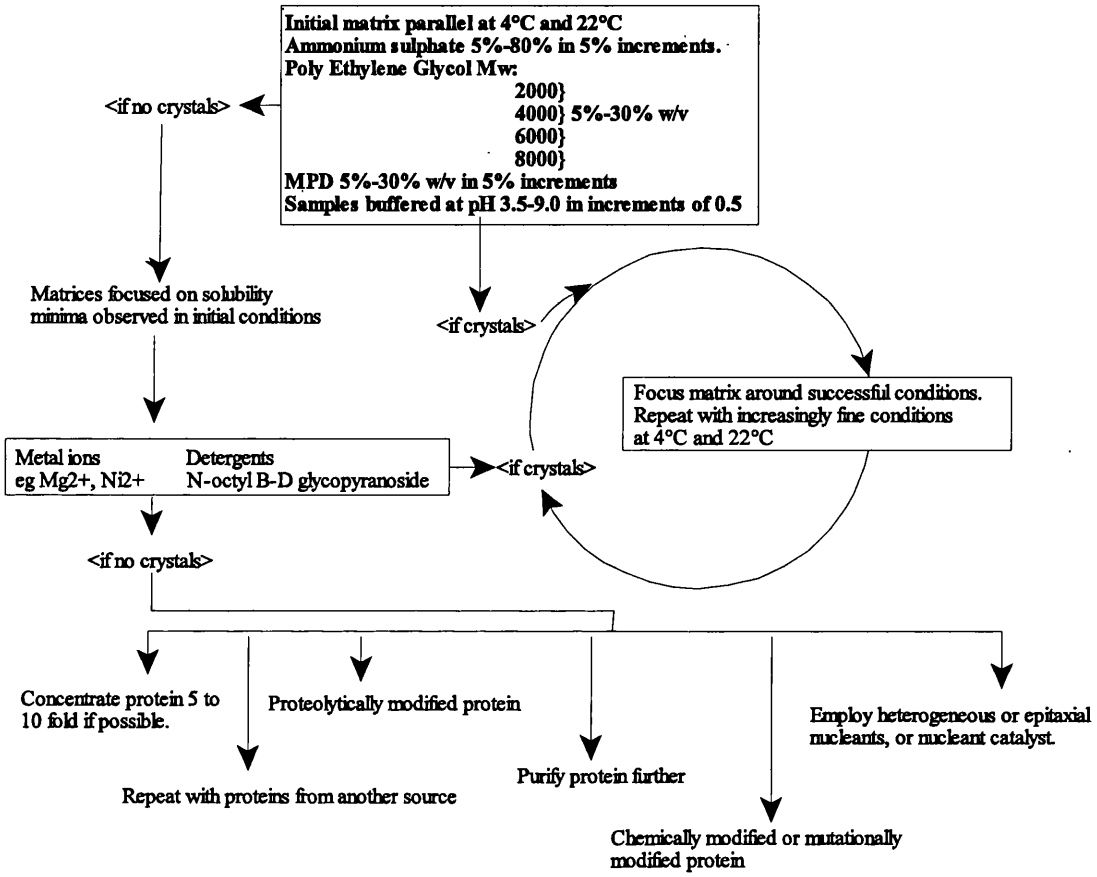


Figure 2.4 Flow diagram showing the succession of variables and procedure in the investigation of crystallisation conditions.

Chapter 3

Methods for Protein Structure Determination

3.1 Importance of protein structure information

The three-dimensional structure of a protein has important implications for how it functions and interacts with its surrounding environment. Knowledge of a protein structure is therefore crucial for us to understand fully its biological functions. Protein structural information studies thus provide us with techniques with which we can develop our understanding of biology from the atomic level to that of the whole organism.

3.2 Overview of protein structure

3.2.1 Primary structure

Protein structure can be divided into various levels of organisation. The primary structure, or amino acid sequence, relates to the covalent structure of the protein. This amino acid sequence determines the resulting protein structure. All the 20 amino acids have in common a central carbon atom (C_α) to which are attached a hydrogen atom, an amino group (NH_2) and a carboxyl group (COOH). What distinguishes one amino acid from another is the side chain attached to the C_α at the fourth valency. The addition of any carbohydrate chains, metal and other prosthetic groups completes the definition of primary structure.

3.2.2 Secondary and tertiary structure

The peptide bond is a rigid planar unit, but there is a large degree of rotational freedom about the two bonds on either side, and the different amino acids convey different properties to the polypeptide chain. The main chain of the polypeptide is polar and therefore hydrophilic and the polar groups are neutralised by hydrogen bond formation. This results in the formation of regular secondary structures such as 3_{10} and α-helices, β-strands, sheets and turns.

Tertiary structure describes how the secondary structure elements are organised into domains. Protein structure formation depends very heavily on non-covalent forces. These include repulsion between non bonded electron shells, coupling between oscillating dipoles giving rise to attraction, electrostatic attraction and repulsion (of full and partial charges) and hydrogen bonds. The interaction of the polypeptide chain with the solvent system has a large

influence on the structure, and the strongest contribution to protein stability comes from hydrophobic forces. The main driving force for folding water-soluble globular proteins is the energy gain in packing the hydrophobic side chains into the interior of the molecule.

3.2.3 Post translational modification

Proteins often undergo covalent modifications after the peptide bonds have been formed by the ribosomal synthetic machinery. Specific proteolysis is a common and convenient tool for protein activation and the construction of protein complexes. The physiological advantages of proteolytic cleavage are well demonstrated in the complement system. The susceptibility to inactivation and degradation is also necessary for the routine turnover of proteins. The destination of a polypeptide may be encoded in its sequence. A signal peptide is an N-terminal extension containing hydrophobic residues which attaches to the micrososomal membrane during secretion before its removal by membrane bound proteases.

The side chains of many amino acids may be covalently modified by the addition of various chemical groups. These modifications may extend the chemical capabilities of the 20 amino acid residues or modify the biological activity of the protein. Modification of the N-terminal group is frequently observed by acetylation and formylation. The N-terminus may also be masked by conversion of an N-terminal glutamate to a pyrrolidone carbonyl group. The biological significance of masking the N-terminal group is not clear but it may protect the protein from attack by aminopeptidases, or it may be important for anchoring the N-terminus in an apolar environment, or it may have physiological significance.

Attachment of carbohydrates is one of the most prevalent posttranslational modifications of eukaryotic proteins, yet the process has no defined universal purpose. Those functions that have been determined seem to be specific to each protein. These may relate to the biological activity, but are more often important for the physical properties of the protein. Glycosyl groups attached to proteins are variable in structure, hydrophilic in nature and bulky. There are two types of glycosylation, called N-type and O-type depending on the atom of the

protein to which the carbohydrate is attached. N-type glycosylation occurs exclusively on the nitrogen atom of Asn side chains in the endoplasmic reticulum. The Asn residue always occurs in a characteristic sequence: Asn-Xaa-Ser or Asn-Xaa-Thr where Xaa can be any residue except Pro. The glycan chain is extensively modified during passage of the protein from the endoplasmic reticulum to the Golgi apparatus. O-type glycosylation involves attachment of carbohydrate to the side chain oxygen of Ser and Thr residues in the Golgi apparatus.

Disulphide bond formation between Cys residues is common in extracellular proteins. Two cysteine residues in different parts of the polypeptide chain, but adjacent in the folded structure of a protein, can be oxidised to form a disulphide bridge. Disulphide bridges stabilise the three dimensional structure of a protein and may hold together different polypeptide chains. The mechanism of disulphide bond formation *in vivo* is uncertain, but probably involves thiol disulphide exchange between the protein and small molecules in solution. The predominant thiol compound in most cells is glutathione which exists in both the reduced (GSH) and oxidised forms (GSSG). Formation of one disulphide bond in a protein requires two sequential disulphide exchanges. An enzyme, protein disulphide isomerase, is present in the endoplasmic reticulum to catalyse disulphide rearrangements in proteins.

3.2.4 Motifs and domain structures

Protein structures can be characterised by defining their secondary structural elements in the interior of the molecule and the way that they are connected by loop regions at the surface. Certain combinations of secondary structure elements occur frequently in protein structures and are called motifs (Branden and Tooze, 1999). In large proteins, polypeptide chains are folded into independent discrete structural units known as domains, the cores of which are built up from combinations of motifs of secondary structure.

3.3 Protein structure determination

Methods for protein structural studies either characterise the general properties of the protein (circular dichroism (CD) and Fourier transform infrared (FT-IR) spectroscopy), examine their atomic structures (crystallography and nuclear magnetic resonance (NMR)

spectroscopy), or predict their atomic structures. Even though the use of CD and FT-IR gives only low resolution information about protein structure at the secondary structure level, these techniques are very useful to test if a structure prediction correlates with the native structure and to determine which of the five structural classes the protein belongs to (see section 3.7).

3.3.1 Circular dichroism

Circular dichroism (CD) is observed when optically active matter absorbs left and right circularly polarised light differently. Linearly polarized light can be viewed as a superposition of opposite circularly polarized light of equal amplitude. A projection of the combined amplitudes perpendicular to the propagation direction thus yields a line (Figure 3.1a). When this light passes through an optically active sample with a different absorbance for the two components, the amplitude of the more strongly absorbed component will be smaller than that of the less absorbed component. The consequence is that a projection of the resulting amplitude now yields an ellipse instead of the original line (Figure 3.1b). The generating of ellipticity is called Circular Dichroism.

Amino acids are chiral molecules which occur in proteins in the L-form. The application of CD is based upon the optical activity of the individual amino acids within the peptide backbone of the macromolecule. In practice, two types of CD spectra can be distinguished. There is CD that is associated with the peptide backbone (derived from amide-amide interactions). Superposed onto this is the optical activity of the chromophores (such as the phenol of tyrosine, the phenyl of phenylalanine, the indole of tryptophan, and disulphide bonds). The CD will be associated with the normal absorption spectrum of the chromophore under study and will be characteristic of its position and orientation with respect to the molecule.

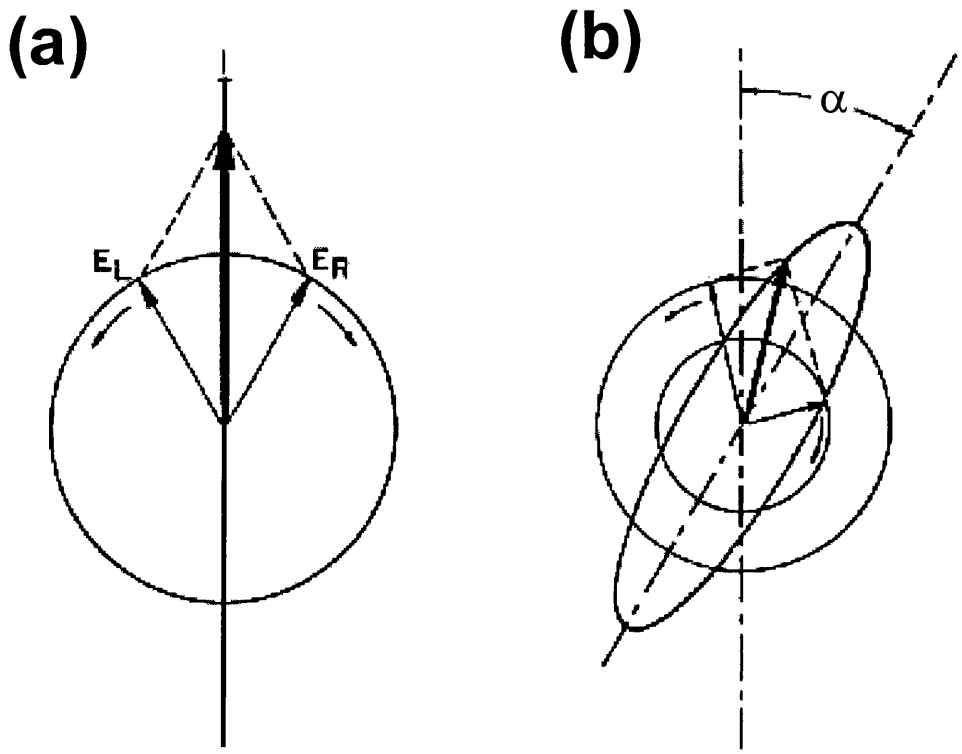


Figure 3.1 (a) Linearly polarized light can be viewed as a superposition of opposite circular polarized light of equal amplitude and phase. E_L and E_R are the electric vectors of the transmitted L and R waves. (b) If the left- and right-hand polarized components are differently absorbed, the transmitted light is elliptically polarised. The effect is minute and using actual numbers the ellipse would still resemble a line. If the L component is retarded with respect to R, the resultant is plane polarised, like the incident light, but the plane is rotated by α.

The amide interactions that give rise to CD are due to two types of electronic absorptions, electric dipole allowed (π-π^*) and magnetic dipole allowed (n-π^*) (Figure 3.2). Electric dipole allowed (magnetic dipole forbidden) π-π^* transitions give a strong absorption around 190 nm, while magnetic dipole allowed (electric dipole forbidden) n-π^* transitions give a weak absorption around 210 nm. Because of the weaker absorption of the n-π^* transition, masking by the π-π^* transition often occurs. These two transitions are observed over a spectral range of 250-170 nm, and it is this range that is used for the analysis of secondary structure. As the optical activity due to the coupling chromophores is related to the relative orientations of the transition moments, conformation (and hence secondary structure) can be inferred.

The CD spectrum is considered as the measured absorption spectrum for right circularly polarised light minus the absorption spectrum for left circularly light. Thus:

$$\Delta\varepsilon = \varepsilon_L - \varepsilon_R = \frac{A_L - A_R}{cl} = \frac{\Delta A}{cl} \qquad \text{Equation 3.1}$$

as derived from Beer's law where: $\Delta\varepsilon$ is the differential molar extinction coefficient; ΔA is the differential absorbance between left circularly polarised light (A_L) and right circularly polarised light (A_R); c is the concentration in molarity; and l is the path length. When applied to the CD spectra of a protein the ΔA^{λ}_{obs} observed at every wavelength becomes:

$$\Delta A^{\lambda}_{obs} = \Delta A^{\lambda}_{\alpha\text{-}helix} + \Delta A^{\lambda}_{\beta\text{-}sheet} + \Delta A^{\lambda}_{random} \qquad \text{Equation 3.2}$$

where $\Delta A^{\lambda}_{element}$ is the CD contribution of the indicated structural component. This is based on the assumption that the measured spectrum can be reduced to a linear combination of fundamental spectra. Equation 3.2 in practice is a general oversimplification. When both parallel and anti-parallel β-sheets are taken into account as well as the various β-turns that have different CD spectra, along with a reassessment of the random contribution, equation 3.2 can be restated:

$$\begin{aligned}\Delta A^{\lambda}_{obs} = \Delta A^{\lambda}_{\alpha\text{-helix}} &+ [\Delta A^{\lambda}_{\text{parallel }\beta\text{-sheet}} + \Delta A^{\lambda}_{\text{antiparallel }\beta\text{-sheet}}] \\ &+[\Delta A^{\lambda}_{\beta\text{-turn(III)}} + \Delta A^{\lambda}_{\beta\text{-turnII}} + \Delta A^{\lambda}_{\gamma\text{-turn}} + \ldots] \\ &+[\Delta A^{\lambda}_{LH} + \text{Dalirregular}]\end{aligned} \qquad \text{Equation 3.3}$$

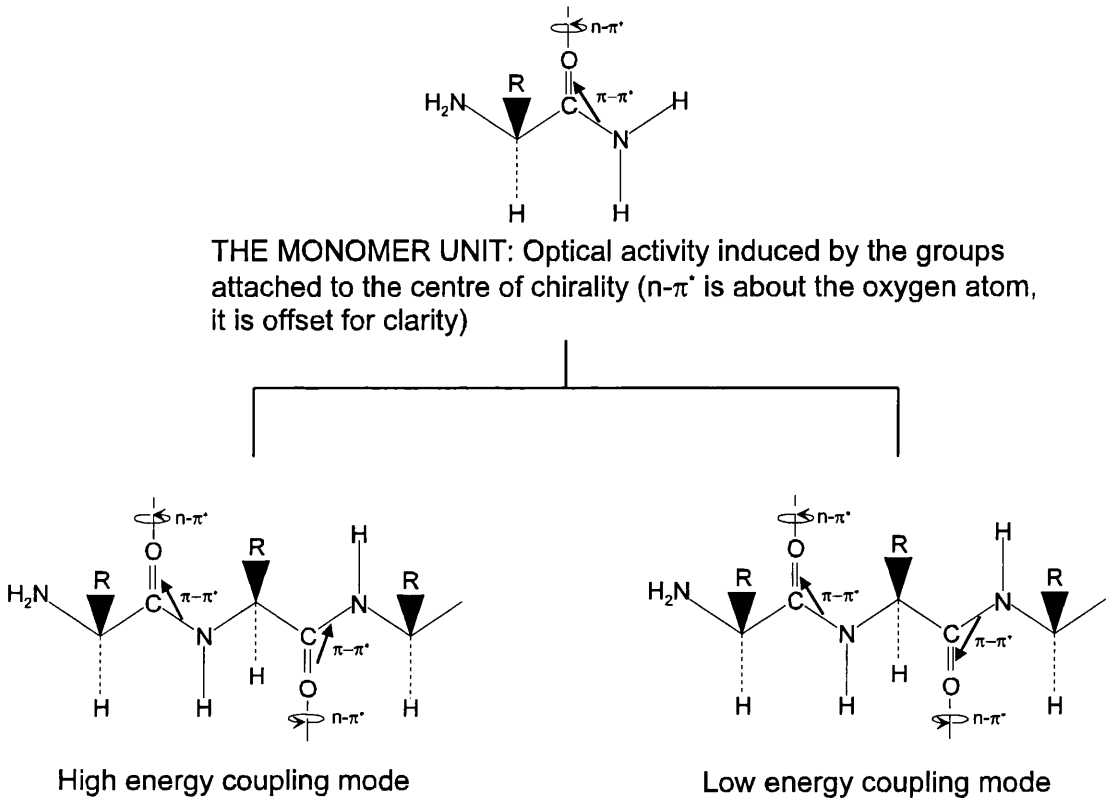


Figure 3.2 A representation of the optical activity induced by the chiral centres that occur in amino acids and peptides. The figure illustrates the effect of the polypeptide on the CD contributions induced by the centres of chirality. The induced optical activity derived from coupling chromophores is relative to the orientations of the transition moments. This means that there is a direct relationship between conformation (secondary structure) of the polypeptide and optical activity (adapted from Drake, 1994).

The reassessed random component now includes a contribution from the polyproline II type left-handed helix conformation (ΔA^{λ}_{LH}) with the remainder assigned as irregular conformation (Yang *et al.*,1986; Toumadje *et al.*, 1992).

Although ΔA is measured, the parameter $[\theta]_{\lambda}$, the molar ellipticity is often used. These parameters are related by the expression:

$$[\theta]_{\lambda} = 3300\Delta\varepsilon \qquad \text{Equation 3.4}$$

The molar ellipticity $[\theta]_{\lambda}$ is defined as:

$$[\theta]_{\lambda} = \frac{\left(\theta_{obs} M_{w}\right)}{100lc} \qquad \text{Equation 3.5}$$

where θ_{obs} is the obseved ellipticity in degrees and M_w is the molecular mass. For the mean residue ellipticity, M_w is replaced by the mean residue mass.

To obtain a CD spectrum, a beam of monochromatic and linearly polarised light passes through a polarisation modulator onto a photomultiplier detector. The light beam now possesses periodic polarisation (at any fixed wavelength), with the intensity of the beam not varying. When an optically active sample (that absorbs at this wavelength) is introduced into the beam, preferential absorption will be seen with the periodicity of the modulation. The intensity of the transmitted light now varies during the cycle of periodic polarisation and this variation is directly related to the circular dichroism of the sample at the specified wavelength. By scanning through the wavelengths, a full CD spectrum can be generated (Drake, 1994).

Experimental conditions are limited by the measurement method, as the more light that falls on the photomultiplier, the lower the noise. In contrast, increasing the amount of sample (increasing concentration or path length) gives a greater CD signal but also less light falls onto the photomultiplier as a result of sample absorption. For optimum sensitivity relatively low concentrations of proteins in solution are needed for this technique. As a result of this, the protein has to be pure, and the solvent buffer has to be transmissive across the spectral range 250-170 nm.

Reliable calculations of the CD of a specific protein from first principles remain difficult and reference spectra of proteins with known X-ray structures, giving a consensus set of spectra, are used to infer the appropriate secondary structure elements (Johnson, 1990). Figure 3.3 illustrates the CD spectra associated with protein secondary structures. CD analysis can yield results with a precision of 5% for α-helix proteins but this relies on the occurrence of one dominant conformational feature. With this in mind it has been suggested that CD spectra should be used to assign the fold class to the particular protein and to monitor processes such as protein unfolding or interactions (Drake, 1994; Campbell and Dwek, 1984). Computer programs such as CONTIN (Provencher, 1982) can identify the secondary structure content of a sample from its CD spectrum.

3.3.2 Fourier transform infrared spectroscopy

Fourier transform infrared (FT-IR) spectroscopy is a technique that allows the elucidation of secondary structure elements within a peptide or protein by study of the amide I band. Before the introduction of the Fourier transform technique, analysis of protein spectra was hampered by the strong absorption of H_2O at the position of the amide I band within the infrared region. This difficulty has been circumvented in the past by the use of ^2H_2O (D_2O) solvents, but this can introduce artefacts of interpretation due to solvent effects. The power of FT-IR spectroscopy is that it allows the background absorption of H_2O (in the range 1700-1500 cm^{-1}) to be subtracted from the spectrum under study to yield the infrared protein spectrum in H_2O.

With the advent of FT-IR the analysis of secondary structure components within infrared spectra of proteins has become commonplace. The amide bond in the peptide backbone gives the main infrared active vibrational modes (Table 3.1). The two most important absorption bands for secondary structure analysis are the amide I and amide II bands (Haris and Chapman, 1994). The amide I band arises due to the C=O bond stretch which is weakly coupled with the in-plane N-H bending and C-N stretching modes. The frequency of this band depends on the hydrogen bonding scheme of the C=O and N-H groups, and will depend on the secondary structure. This has been shown to be the case with

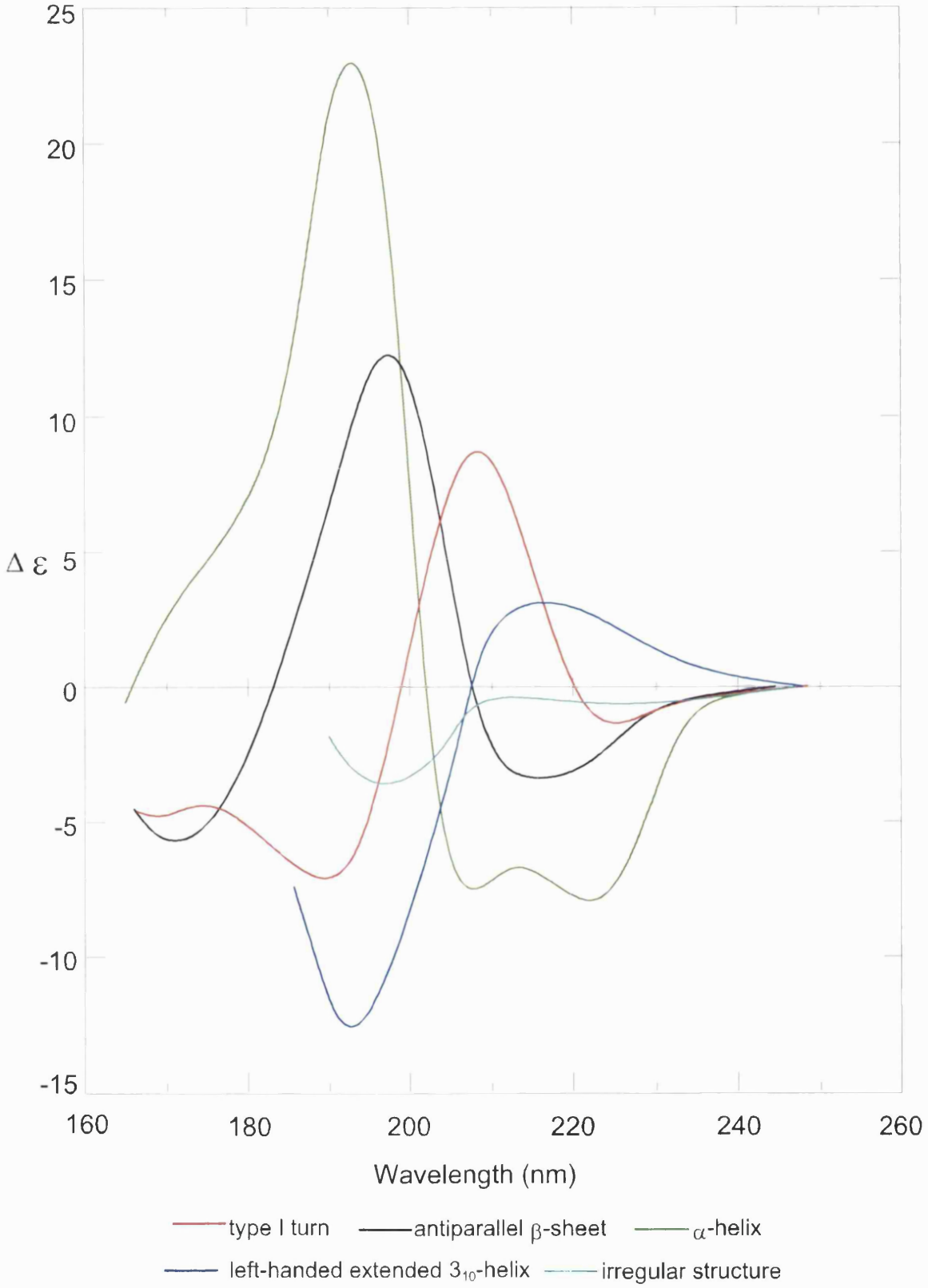


Figure 3.3 The CD spectra associated with various protein secondary structures (Drake 1994).

Table 3.1 Characteristic infrared bands of the peptide linkage

Designation	Approximate frequency, cm^{-1}	Origin
A	3300	N-H(s)
B	3100	N-H(s)
I	1690-1600	C=O(s) 80%, N-H(b) 10%, C-N(s) 10%
II	1575-1480	N-H(b) 60%, C-N(s) 40%
III	1301-1229	C-N(s) 30%, N-H(b) 30%, C=O(s) 10%, O=C-N(b) 10%
IV	767-625	O=C-N(b) 40, other 60%
V	800-640	N-H(b)
VI	606-537	C=O(b)
VII	200	C-N(t)

Adapted from Haris and Chapman 1994. Amides IV to VII are out-of-plane modes; the others are in-plane modes: (s)-stretching, (b)-bending, (t)-torsion.

experiments carried out on proteins of known three-dimensional structure, which showed good correlations between amide I band frequencies and secondary structure elements (Haris and Chapman, 1994).

As opposed to traditional dispersive infrared spectrometers that irradiate the sample while scanning through the spectral range, FT-IR spectrometers irradiate the sample with all wavelengths simultaneously. This simultaneous irradiation comes about when a polychromatic beam encounters a beam splitter which redirects the beam onto two mirrors at right angles to each other (Michelson interferometer). One of these mirrors is translated in a direction perpendicular to its axis. This has the effect of introducing a path length difference between the two beams. The beams subsequently recombine at the beam splitter. If the two path lengths are the same, then there is no path length difference and the beams constructively combine for all frequencies present in the original beam. For different path lengths, the amplitude of the recombined signals will depend on the frequency and distance the mirror moves. The recombined beam is directed through the sample and onto the detector. The detector output is in the form of an interferogram, i.e., the sum of the sine waves of all frequencies present. Using an unattenuated polychromatic beam allows for a high signal to noise ratio in the measured spectrum. FT-IR has several advantages over dispersive spectrometers, the most important of which is the efficient and rapid collection of data. It is also less susceptible to stray radiation and it is easy to perform many scans to improve the signal to noise ratio.

When obtaining the spectrum of a protein or polypeptide, a reference spectrum of the solvent/buffer (under exactly the same conditions) is needed so as to subtract the background absorption due to the solvent/buffer. The routine for subtraction of H_2O based solvents is via an interactive difference method. Here, the H_2O combination band near 2150 cm^{-1} (where there is no protein absorbance) is cancelled to achieve a flat baseline between 2000-1710 cm^{-1}. Spectra using ^2H_2O buffers require the purging of atmospheric H_2O vapour using dried air or N_2 gas.

To get around the problem of a broad amide I band in protein structure with unresolved secondary structure elements, mathematical techniques are applied to the spectral data such as derivative and deconvolution procedures in order to resolve the overlapping components. Various studies of these amide I components (Table 3.1) for model peptides and proteins of known secondary structure has led to the following assignments; 1620-1640 cm^{-1} are β-sheet; 1670-1695 cm^{-1} is associated with antiparallel β-sheet; 1648-1657 cm^{-1} is normally α-helical. In H_2O there can be an overlap of absorptions due to α-helical and unordered structure. This can be remedied by taking the equivalent protein spectrum in D_2O where the unordered structure will absorb at 1644 cm^{-1}. The N-H bending vibration that contributes the amide II band is most useful for studying the hydrogen deuterium exchange that occurs at the amide group.

The use of FT-IR spectroscopy for quantitative analysis of secondary structure has been commonly carried out using the method of identifying the amide I component bands from the deconvoluted spectra, then applying curve fitting to the band contours. This method suffers owing to the assumption that all the components have the same shape. Another problem is that the process of deconvolution may bring about the appearance of side lobes as well as a disproportionate enhancement of random noise. An alternative method utilises a calibration set of IR spectra of 18 soluble proteins of known crystal structures (Lee *et al.*, 1990). The errors of analysis using this method are 3.9% for α-helix, 8.3% for β-sheet and 6.6% for turns when these analyses are compared with X-ray diffraction values. Secondary structures can be determined using this method without the need for deconvolution and with little pretreatment of data (Haris and Chapman, 1994).

3.4 X-ray crystallography

Unlike CD and FT-IR, X-ray crystallography provides information about molecular structures at atomic resolution levels and is the most popular method of determining the three-dimensional structure of proteins. X-ray crystallography relies on the basic principle of diffraction. The molecular structure that is determined is an approximation of all the molecules that occur within the crystal (Creighton, 1993). Diffraction patterns can be detected when a

wave is scattered by a periodic structure such as a crystal or fibre. A single crystal is mounted and illuminated with an intense, well collimated source of X-rays of a single wavelength. The X-rays are generated either by electron bombardment of a copper target giving a wavelength of 0.1542 nm, or from synchrotron radiation sources. X-rays are scattered by electrons; thus the scattering intensity is proportional to the atomic number of each atom. The resultant diffraction pattern is a function of the electron density of the crystal. This pattern is captured on photographic film or on an area detector and from this the dimensions of the unit cell and the symmetry of the molecule can be determined. Before molecular information can be determined, the phase problem which arises from loss of phase information during detection needs to be solved. Solutions that are commonly used are multiple isomorphous replacement, anomalous scattering, and molecular replacement. From any of these solutions, an electron density map can be constructed, from which a contour diagram is drawn and the structure finally refined (Campbell and Dwek, 1984; Atkins, 1987; Branden and Tooze, 1999; Creighton, 1993).

Even though this technique is well established, there are a number of drawbacks. Crystal production can be difficult, time consuming and sometimes impossible. The crystallisation procedure may render the protein in an unnatural, nonphysiological environment. Other problems are that the end model represents a time-averaged structure, meaning that flexible regions of proteins are unresolved through disorder (MacArthur *et al.*, 1994). Despite these drawbacks, X-ray crystallography has provided many protein structure determinations.

3.5 Nuclear magnetic resonance spectroscopy

Nuclear magnetic resonance (NMR) allows structural determination of molecules in solution at atomic resolution. The magnetic resonance phenomenon occurs as a result of the quantum mechanical property of "spin" (angular momentum) intrinsic to a number of different nuclei. Although the nucleus can be described as spinning about an axis, spin is a unfortunate choice of words since the source of the angular momentum is purely a quantum mechanical phenomenom and could equally well be described as "bitterness" or "sweetness" (Evans, 1995). NMR spectroscopy can be carried out only on nuclei that possess a magnetic

moment. Common biological nuclei such as ^{12}C or ^{16}O do not possess a magnetic moment and so do not give NMR spectra. The nuclear magnetic moment (μ) is given by:

$$\mu = \gamma I \hbar \qquad \text{Equation 3.6}$$

where γ is the gyromagnetic ratio which determines the resonant frequency of the nucleus and $\hbar$ is Planck's constant. The nuclear spin quantum number I can have the values $I = 0, {}^1/_2, 1, 1{}^1/_2$, etc (Table 3.2). A nucleus with spin I may take $(2I+1)$ different orientations in a magnetic field, which are distinguished by the magnetic quantum number m_I. For a proton two energy states, α (lower) and β (upper), correspond to the spin orientations (Figure 3.4). The ratio of the populations of the two energy states in equilibrium is given by:

$$\frac{N_\beta}{N_\alpha} = e^{-\frac{\Delta E}{k_B T}} \qquad \text{Equation 3.7}$$

(the Boltzmann equation) where N_α is population in the lower energy level (α state) and N_β is the population in the upper energy level (β state). At thermal equilibrium there will be a surplus of nuclei in the lower energy state (Boltzmann surplus). This gives rise to a net magnetisation parallel to the applied field, along the z-axis (Figure 3.5).

If a proton is placed in an external magnetic field B_0, its magnetic moment has a component either parallel or anti-parallel to the direction of the applied field, corresponding to the lower and upper energy levels (Figure 3.6). During transition between the α and β states the proton will be in a mixed quantum state where there will be a sideways component of the magnetic moment which precesses around the B_0 axis. The angular frequency of precession is known as the Larmor frequency, which is its NMR absorption frequency ν:

$$2\pi\nu = -\gamma B_o \qquad \text{Equation 3.8}$$

where γ is the gyromagnetic ratio and B_0 is the applied magnetic field.

In the NMR experiment there are many spins precessing about B_0 and at equilibrium these spins are randomly distributed so that the net sideways component of the magnetic moment (M_{xy}) is zero. In order to observe NMR signals, the system must be perturbed by

Table 3.2 Properties of some nuclei

Nucleus	Spin (in units of ħ)	Spin orientations m_I, (2I +1) values	Gyromagnetic ratio γ (10^7 rad $s^{-1}T^{-1}$)	Relative sensitivity *1	Frequency at 1T (MHz) *2	Natural abundance (%)
^1H	$^1/_2$	$-^1/_2$, $^1/_2$	26.75	1.00	42.58	99.98
^2H	1	-1,0,1	4.11	9.65×10^{-3}	6.54*3	1.5×10^{-2}
^{13}C	$^1/_2$	$-^1/_2$, $^1/_2$	6.73	1.59×10^{-2}	10.71	1.1
^{14}N	1	-1,0,1	1.93	1.01×10^{-3}	3.08	99.6
^{15}N	$^1/_2$	$-^1/_2$, $^1/_2$	-2.71	1.04×10^{-3}	4.32	0.4
^{17}O	$^5/_2$	$-^5/_2$, $-^3/_2$, $-^1/_2$, $^1/_2$, $^3/_2$, $^5/_2$	-3.63	2.91×10^{-2}	5.76	3.7×10^{-2}
^{19}F	$^1/_2$	$-^1/_2$, $^1/_2$	25.18	0.83	40.01	100
^{23}Na	$^3/_2$	$-^3/_2$, $-^1/_2$, $^1/_2$, $^3/_2$	7.07	9.25×10^{-3}	11.26	100
^{31}P	$^1/_2$	$-^1/_2$, $^1/_2$	10.84	6.63×10^{-2}	17.24	100
^{35}Cl	$^3/_2$	$-^3/_2$, $-^1/_2$, $^1/_2$, $^3/_2$	2.62	4.7×10^{-2}	4.17	75.5

Footnotes:

*1 At constant field.

*2 The value of γ can be obtained from $\omega = -\gamma B_o/2\pi$ Equation 3.8

*3 The ^2H frequency is the method of choice for locking modern NMR spectrometers, and for this purpose NMR samples require at least 5% ^2H_2O. Often the amide protons of proteins are exchanged with deuterium to simplify the proton spectra.

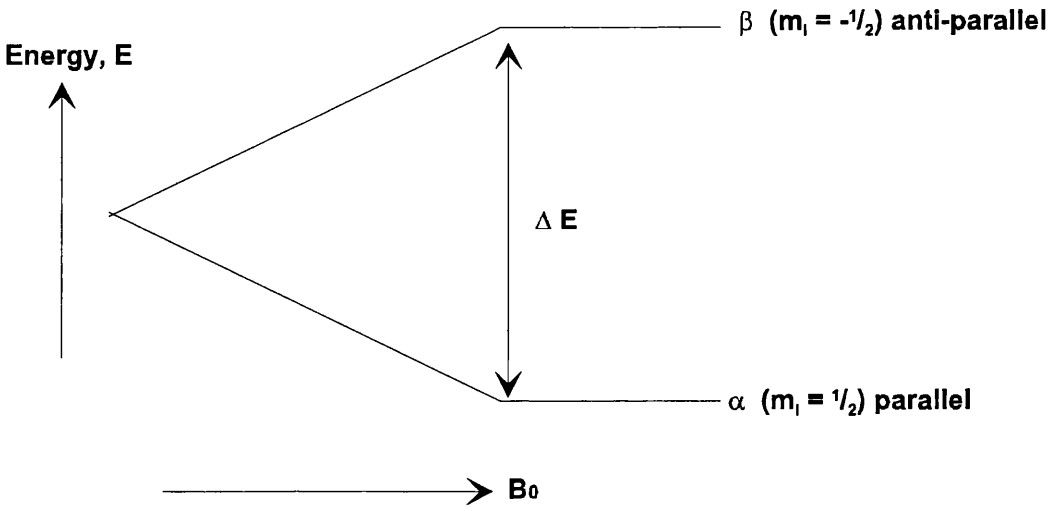


Figure 3.4 Nuclear spin energy levels of a spin 1/2 nucleus with a positive value of γ (eg ^{1}H or ^{13}C) placed in an increasing external magnetic field B_0. The angular momentum of the nucleus will align along an axis either parallel or anti-parallel to the direction of the applied field, and this corresponds to the lower (α) and upper (β) energy levels respectively.

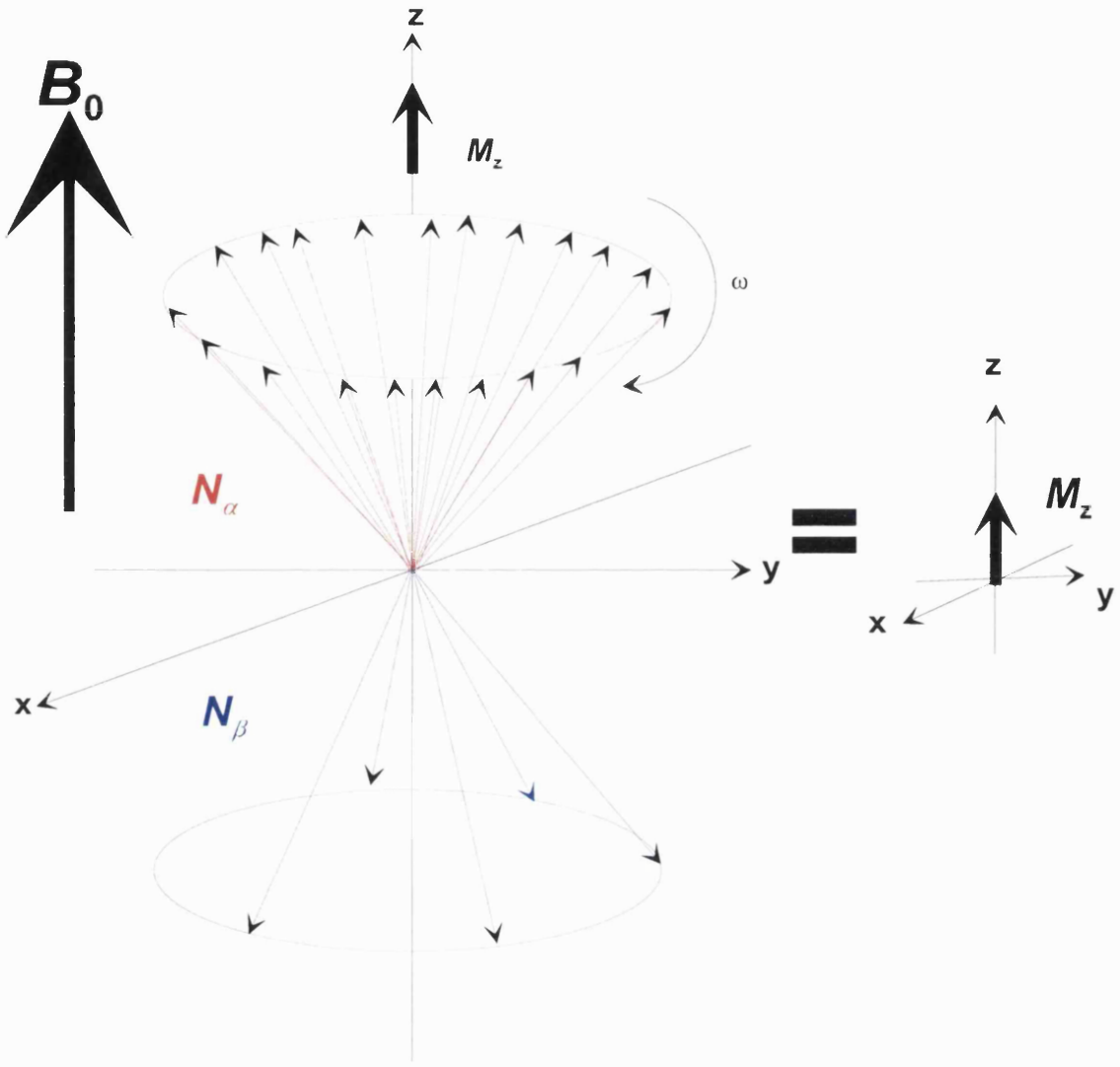


Figure 3.5 At thermal equilibrium there will be a surplus of nuclei in the lower energy state (Boltzmann surplus). The ratio of the populations of the two energy states will be given by the Boltzmann equation (Equation 3.6). N_α describes the population in the lower energy level (α state) and N_β the population in the upper energy level (β state). This gives rise to a net magnetisation (M_z) parallel to the applied field, along the z-axis. The many spins in a sample placed in an external magnetic field B_0 may conveniently be represented by bringing them to a common origin. The diagram is drawn with the nuclear magnetic moments lying in the surface of one of two cones whose axis are aligned either parallel or anti-parallel to the direction of the applied field corresponding to the lower and upper energy levels respectively (for positive γ). There is no measurable sideways component of the angular momentum or magnetic moment unless the nuclei are forced to undergo transitions between the states. During an NMR transition the magnetic moment precesses at an angular frequency ω, with the sense shown (for positive γ).

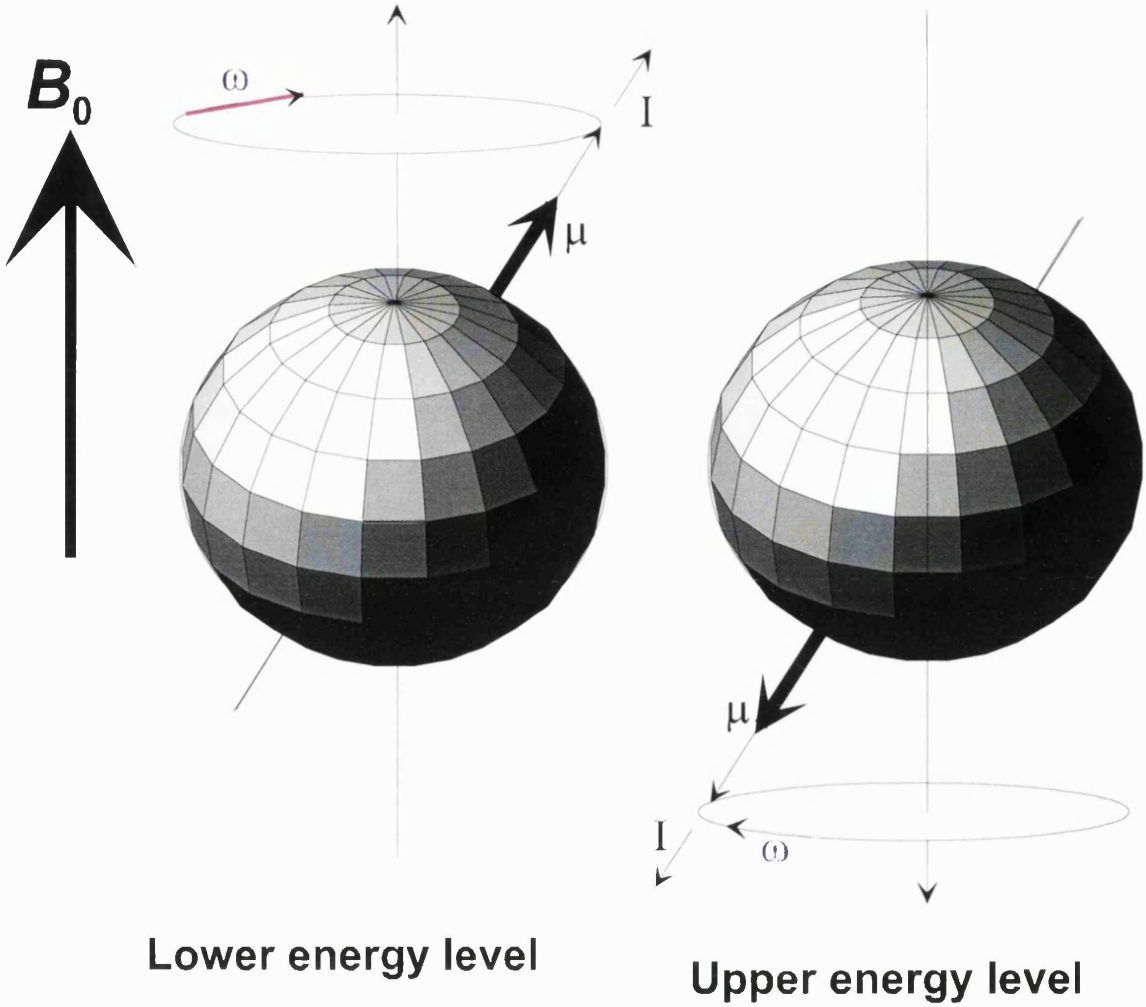


Figure 3.6 Schematic representation of a proton placed in an external magnetic field B_0. The proton has its spin angular momentum quantized along an axis, either parallel or anti-parallel to the direction of the applied field, corresponding to the lower (α state) and upper energy levels (β state) respectively (for positive γ). When the nucleus is either the α or β state, the angular momentum I and the magnetic moment μ are randomly distributed on the surface of a cone and there is no organised sideways component. During a NMR transition the nuclear magnetic moment acquires a sideways component which precesses in the sense indicated above. The rate of precession of the nucleus around the B_0 axis is known as the Larmor frequency, which is its NMR absorption frequency $\omega/2\pi$.

radiofrequency electromagnetic radiation near the Larmor frequency. The displacement of magnetisation from equilibrium is produced by a second magnetic field B_1, induced from a current in a coil so as to introduce field B_1 perpendicular to B_o. The B_1 field is usually applied as a pulse of duration t_p. The angle θ through which the magnetisation is tipped from the z-axis is calculated thus:

$$\theta = \gamma B_1 t_p \qquad \text{Equation 3.9}$$

Resonance occurs when nuclear-spin reorientation occurs between the spin states. The most common method of measurement is by application of a pulse of radiofrequency electromagnetic radiation to the magnetically aligned sample and detection of a transient signal from the nuclear spins (Figure 3.7a). In this one-pulse experiment the bulk magnetisation is typically moved through a tip angle t_p of 90°, or π/2 away from the z-axis leading to a transverse magnetisation. The pulse is then referred to as a 90° (π/2) pulse. This results in a maximal signal as the NMR probe detects transverse magnetisation only (Figure 3.7b). The signal is called a free induction decay (FID): free of any further influence of the radio frequency field; induced in the probe coil and decaying as the spins return to equilibrium. This decay process is exponential and is known as relaxation. Using Fourier transform analyses, multiple transients can be used and the signals processed to give a spectrum of intensity against frequency as opposed to the transient being intensity against time (Campbell and Dwek, 1984; Evans, 1995). The value of NMR results in the fact that the magnetic field experienced by a nucleus is not simply the applied field, since it is modified by small induced fields of neighbouring atoms. The five basic parameters of NMR are the signal intensity, chemical shift δ, spin-spin coupling J, and the two relaxation times T_1 and T_2.

3.5.1 Signal intensity

If a system of nuclear spins is initially in equilibrium, then the amplitude of the transient signal after a pulse is proportional to the number (concentration) of nuclei in the sample. If the spins do not start out in equilibrium, because a previous pulse has equalised the populations (saturation), then the amplitude of the signal will be reduced.

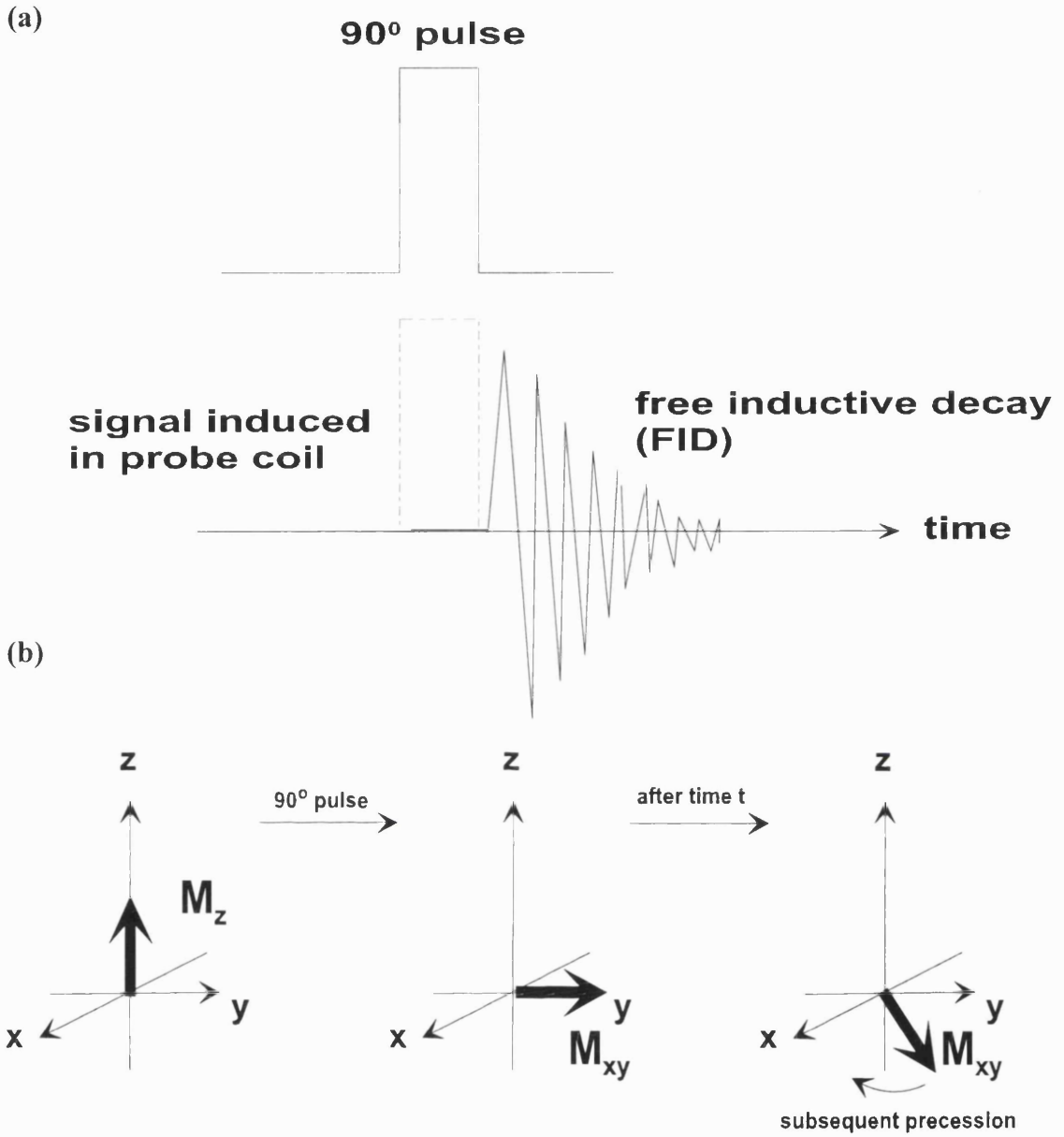


Figure 3.7 (a) The application of a radio frequency electromagnetic pulse to an ensemble of spins creates a superposition of the different spin states. The free inductive decay (transient) is a superposition of the signals from the nuclei and can be resolved into a frequency spectrum by Fourier transformation. (b) The 90°, or $\pi/2$, radio frequency electromagnetic pulse equalises the populations in the upper and lower energy states and induces coherence between the spins. The net magnetisation parallel to the applied field (M_z) disappears and the bulk magnetisation moves by 90° into the *xy* plane where it can be detected by the NMR spectrometer probe. The probe is so constructed that it can detect transverse magnetisation only.

3.5.2 Chemical shift

Nuclei are surrounded by electrons, which partially shield them from the applied field B_0. B_0 induces currents in the electron clouds that reduce the effective field experienced by the nucleus. These induced currents are directly proportional to B_0. The scale used in NMR is called the chemical shift δ scale and is given by:

$$\delta = 10^6\left(\frac{\delta_{ref} - \delta_{obs}}{\delta_{ref}}\right) \qquad \text{Equation 3.10}$$

where δ_{ref} is the position, in Hertz, observed for a reference compound and δ_{obs} is the position of the signal of interest. The scale, which is measured in units of parts per million (ppm), is independent of B_0. Intrinsic shifts are characteristic of a particular chemical group, and induced shifts arise from the influence through space of neighbouring magnetic centres.

3.5.3 Spin-spin coupling and multiplet structure

NMR spectra exhibit multiplet structures that arise from weak interactions between magnetic nuclei, communicated by the electrons in a chemical bond. The size of the interaction is defined as the spin-spin coupling constant J and is usually expressed in Hertz.

3.5.4 Spin-lattice or longitudinal relaxation time T_1

Nuclei in the upper energy state can lose energy to their surroundings thereby increasing the thermal motions of the molecules in solution or the crystal lattice. This relaxation process is characterised by the T_1 relaxation time. The T_1 relaxation time is described as the time constant for the recovery of M_z after a perturbation. The component M_z can be perturbed in several ways and T_1 depends on the molecular environment. Relaxation times are short in solids and relatively long in liquids.

3.5.5 Transverse or spin-spin relaxation time T_2 and linewidth

One of the dynamic effects of the dipolar interaction between nuclei is to cause mutual spin flips which limit the lifetime of a nucleus in the upper state. The interaction between nuclei and paramagnetic ions is an efficient way of relaxing nuclei. The spin-spin relaxation time T_2

is defined as the time constant for the randomisation of the M_{xy} components. T_1 processes will cause the M_{xy} components to disappear, but in general T_2 is less than T_1 due to dipole-dipole interaction between different environments. It is related to the linewidth by the following relationship:

$$\Delta \nu_{\frac{1}{2}} = \frac{1}{\pi T_2} \qquad \text{Equation 3.11}$$

where $\Delta\nu_{1/2}$ is the half-height linewidth.

3.5.6 Protein NMR

NMR is a relatively insensitive technique and requires pure protein at millimolar concentrations. The most common constituent of proteins that generate an NMR signal is the ^{1}H nucleus (proton), with a nuclear spin of ½ (Table 3.2). Labile protons (eg NH, NH_2, OH, and SH) can be replaced with deuterium to simplify the NMR spectrum and so D_2O is a common solvent in NMR. The low natural abundance of ^{13}C (nuclear spin ½) means that it can also be used to some extent but recombinant protein expression systems are required to exploit active nuclei such as ^{13}C and ^{15}N.

3.5.7 Structural information

Structural information contained within a NMR spectrum can take a number of forms. The most regularly used effects are the nuclear Overhauser effect (nOE) and coupling constants *J*. nOEs arise from through-space proton-proton interactions (for protons <5Å apart) that are proprtional to r^{-6}, where *r* is the distance between the two protons, and they give information about distance separations. The *J* coupling constants arise through bond spin-spin interactions (for protons within three chemical bonds of each other) and give dihedral angle data. Study of amide proton exchange rates also gives information about regular secondary structures (Evans, 1995). By repeating a sequence of pulses, data can be processed to obtain a spectrum as a function of two frequencies to give a two-dimensional (2-D) spectrum. Heteronuclear experiments provide further dimensions. Molecular structures can be determined to the atomic level by the assignment of the NMR signals to individual

atoms and the combination of the distance and angular NMR constraints using molecular modelling software.

The general advantages of NMR over X-ray crystallography are that the technique can be performed in conditions closer to the biological environment without the artefacts that result from the crystallisation process. NMR also provides information on dynamics and can identify individual sidechain motions (MacArthur *et al.*, 1994). Disadvantages of NMR are that the sample for analysis has to be in solution at high concentration (around 1 mM concentration) to give good signal-to-noise ratios. The determination of protein structure with reasonable ease has been limited to small molecules. Larger proteins have also been determined but only after extensive work and isotopic labelling using ^{13}C and ^{15}N.

3.6 Structure prediction

Homology modelling offers a means of obtaining atomic coordinate models in the absence of X-ray crystallographic or NMR data (Lattman, 1995) provided that an homologous structure can be identified. This technique is based on the assumption that domains which appear to share the same ancestor gene will adopt a common fold structure; a sequence similarity of more than 30 % is usually required. An homology model will not be as reliable as an experimentally determined structure and in particular surface loop structures can be difficult to predict. In its favour, homology modelling is able to provide reasonable representations of the core structures of proteins where the general localisation of residues may be good enough to enable interpretations and predictions of functionality. The approach that is used to obtain an atomic coordinate model from a sequence of interest is outlined in Figure 3.8. Obviously, the crucial stage in this approach is the identification of a structure that has the same fold as the target sequence. In the following description of homology modelling, the fold prediction methods that are used for analogy modelling will also be discussed because they can be beneficial for supporting predictions from sequence similarities alone, especially when sequence identity is in the ambiguous 20 to 30% range.

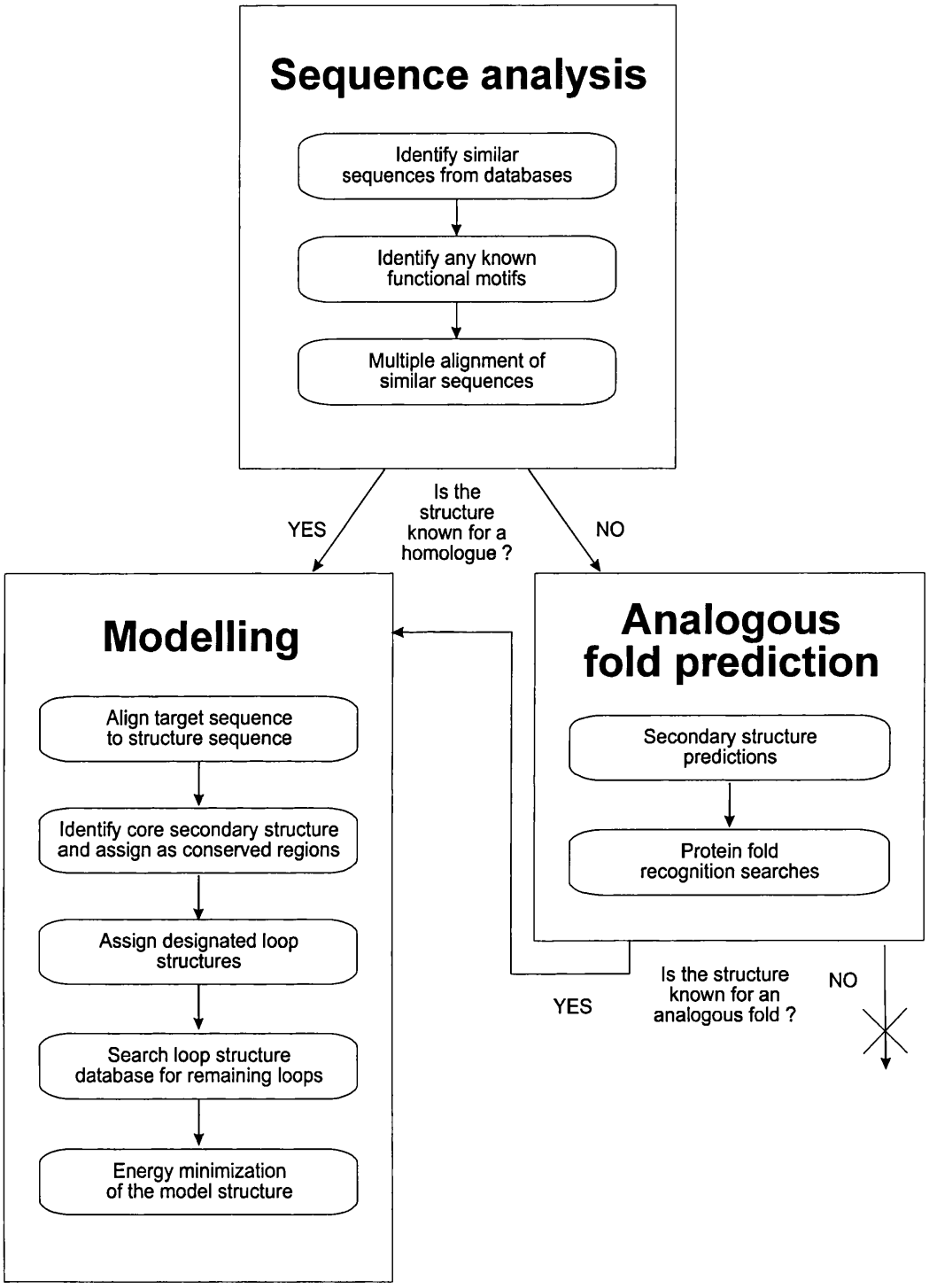


Figure 3.8 Flow chart of the procedures used to generate an atomic coordinate model for a target sequence based on a known structure that is assumed to have the same fold. There are two alternative routes for identifying a structure with the same fold: either a fold structure is shown to have significant sequence similarity (homology) to the target sequence, or secondary and tertiary structure predictions for the target sequence are used to identify an analogous fold. Once the fold structure has been identified, an atomic coordinate model can be built for the target sequence by means of a rigid fragment assembly method.

3.6.1 Sequence analysis

A protein sequence of unknown structure can be analysed in many ways in the process of predicting its structure. Initial analysis is usually carried out at the primary structure level. This involves the search for similar sequences and their subsequent alignment with the "probe sequence". The identification of homologues can be performed using algorithms such as FASTA (Pearson and Lipman, 1988) and BLAST (Altschul *et al.*, 1990), which scan sequence databases for matches to a target sequence. Alternatively, databases such as Entrez store existing sequences with links to their close homologues. However, it may not be possible to find homologous sequences because the residues that are essential for the domain fold represent only a small fraction of the sequence. In these circumstances, short sequence motifs that are characteristic of a protein family or superfamily can be invaluable for predicting homologues. Known sequence motifs are compiled in the Prosite database (Bairoch, 1991). Comparisons are made using sequence alignment algorithms that score residue differences between equivalent positions in two sequences. The physio-chemical properties of the 20 amino acids are summarised in Figure 3.9 and this identifies similarity relationships. The most commonly used alignment scoring schemes have been derived by examining the substitution frequencies observed in sequence alignments, e.g. the Dayhoff matrices (Dayhoff *et al.*, 1978) and BLOSUM mutation matrices (Henikoff and Henikoff, 1992). The next step involves the prediction of secondary structure elements, namely α-helices, β-strands and loops (sections 3.6.2 to 3.6.5). The reliability of the technique is improved by applying this step to the whole sequence alignment, i.e. calculation of an average. Also associated with this step is the prediction of solvent accessibility of residues within the sequence (section 3.6.6). Knowledge of the arrangement and the proportion of secondary structure elements and their solvent exposure assists the assignment of the "probe sequence" to a fold class based on a known atomic structure which may help in the overall definition of its tertiary structure.

Secondary structure arrangements (topology) define the folding pattern of a globular protein domain. If the fold structure has not been determined for the superfamily or if the sequences are very divergent, prediction of secondary structure from the amino acid sequence presents a major means for identifying the correct fold.

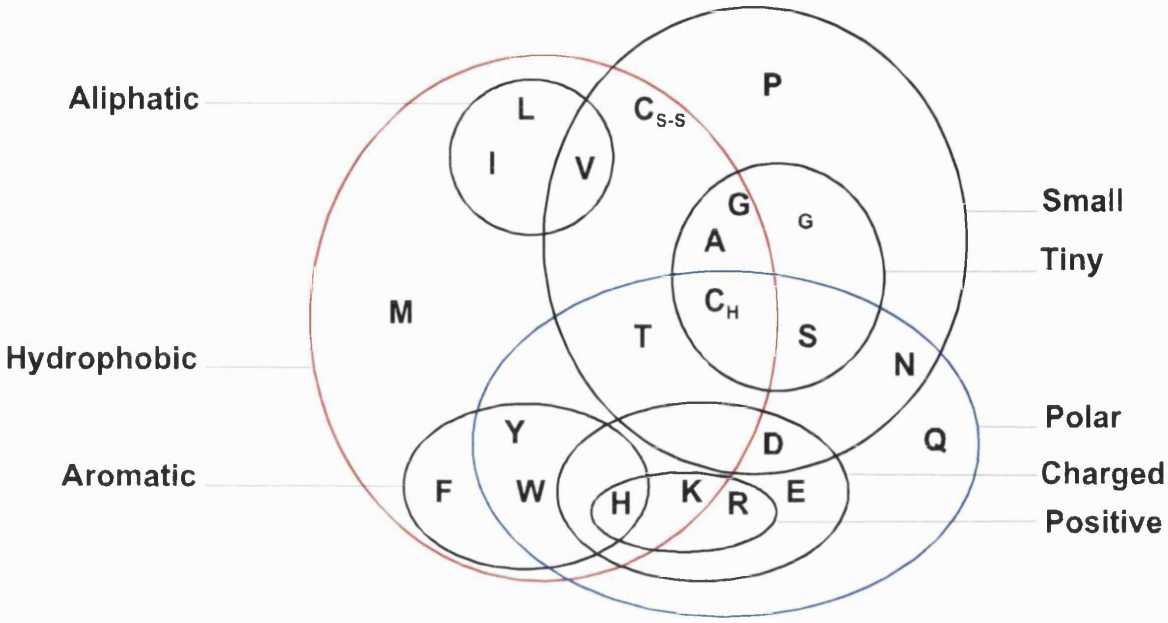


Figure 3.9 Venn diagram shows the relationship of the 20 naturally occurring amino acids to a selection of physio-chemical properties which are important in the determination of protein structure. The diagram is dominated by properties relating to size and hydrophobicity. The amino acids are divided into two major sets, one containing all amino acids which contain a polar group (blue) and a set which exhibit a hydrophobic effect (red). A third major set (small) is defined by size and contains the nine smallest amino acids. Within this is an inner set of smaller residues (tiny) which have at most two side-chain atoms. The localisation of Cys is ambiguous as the reduced form (C_H) has properties similar to serine, while the oxidised form (C_{S-S}) may be more equivalent to Val. Other sets include full charge (charged) which contains the subset positive (negative is defined by implication) and aromatic and aliphatic. The latter set is not as general as the name implies and contains only residues containing a branched aliphatic side chain. Because of its unique backbone properties, proline was excluded from the main body of the diagram. An equivalent position is suggested for Gly by a small G.

An alternative way of finding the fold class is to put the "probe sequence" into a fold recognition process (sections 3.6.7). This is used in cases where the sequence similarity with a known structure is very low (less than 15 %). Here, assignment of fold class or even a direct structural homologue can be found.

The final step in structure prediction is the model building process by homology. This is possible only if any of the previous steps identifies an amino acid sequence that has a known three dimensional structure. Homology modelling can be used to test the reliability of the secondary structure assignments, solvent accessibility predictions and the output from the fold recognition process.

3.6.2 The Chou-Fasman method

Numerous methods have been published for predicting secondary structure and this section and the following outline some of those that are commonly used. The Chou-Fasman method is a statistical prediction method that was derived from the propensities for each amino acid to occur either in an α-helix or in a β-sheet in 15 protein structures (Chou and Fasman, 1978). According to these propensities, each amino acid is allocated to one of six classes depending on its likelihood of forming an α-helix, which vary from strong α-helix former to strong α-helix breaker, and to one of six classes depending on its likelihood of forming a β-sheet, which range from strong β-sheet former to strong β-sheet breaker. A series of rules is used to assign secondary structure elements to clusters of probable α-helix and β-sheet residues in an amino acid sequence. Although this method is fairly uncomplicated in its concept, it has been criticised because of its simple statistical approach, its arbitrary prediction rules, and because it does not consider the chemical and physical properties of the amino acids (King *et al.*, 1996). Its accuracy is improved if it is used with a multiple sequence alignment.

3.6.3 The GOR method

The GOR method is another statistical method (GOR-I; Garnier *et al.*, 1978). The method was developed using a database of 26 protein structures, later updated to a database containing 75 structures (GOR-III; Gibrat *et al.*, 1987). Each residue is unambiguously assigned to one of 4 possible conformations, α-helix, β-sheet, β-turn (2-residue turn) or random coil. The basis of this method is that the amino acid sequence and the secondary structure are two distinct messages, related by a translation process that can be examined using information theory. Although in theory the conformation of any particular residue is dependent on every other amino acid in the protein, the most significant influence on the conformation of a residue is exerted by the eight residues either side of it (Robson and Pain, 1971; Robson and Suzuki, 1976). Structure prediction uses the information a residue carries about its own secondary structure, the information a residue carries on the secondary structure of a second residue within eight residues along the sequence that is independent of the second residue's type, and the information a residue carries about the secondary structure of a second residue that depends on the second residue's type. This method is theoretically elegant and it allows the separation of the different types of information involved in the folding of a protein. However, it also neglects the physical and chemical properties of the amino acids and it deliberately neglects protein folding (King *et al.*, 1996). Its accuracy is again improved if it is used with a multiple sequence alignment.

3.6.4 The PHD method

The **P**rofile network system from **H**ei**D**elberg (PHD) is a secondary structure prediction algorithm based on the output from neural networks. PHD uses the profiles from multiple sequence alignments as an input into a three-layered network. The first layer is a neural network that has been trained to classify residues according to three states of secondary structure, α-helix, β-strand and loop. In the second layer, stretches of predicted residues are analysed and contiguous regions of residues that are predicted to have the same structure are assigned as secondary structure elements, and unlikely stretches of secondary structure elements are discarded. At this stage, the agreement of predicted segment lengths with those observed in protein structures is noticeably improved. The third layer averages the

predictions from 12 networks that have been trained on different datasets and so acts to reduce the "noise" associated with the predictions. PHD had an accuracy greater than 70% when cross-validated on more than 100 unique structures (Rost and Sander, 1993). Based on the statement that homologous proteins have the same three-dimensional fold and approximately equivalent secondary structure profiles at around a similarity level of 25-30% identical residues, a multiple sequence alignment of the protein family can contain more structural information than a single sequence (Rost and Sander, 1993*a*).

3.6.5 The SAPIENS method

SAPIENS (**S**econdary structure and **A**ccessibility class **P**rediction **I**ncluding **EN**vironment-dependent **S**ubstitution tables) is a prediction method based on the evaluation of mean propensities and environment-dependent substitution tables for amino acid residues in aligned sequences. The program uses these data in a three-step method with a multiple sequence alignment as its primary input. Unlike PHD, SAPIENS evaluates four conformational states (α-helix [H], β-strand [E], buried coil [i], and exposed coil [o]). These assignments are modified for neighbouring residue cooperativity and according to the positions of residues that are typically found at the N- and C-terminal caps of secondary structure elements. Next, the secondary structure assignments are altered using predicted solvent accessibility patterns, which are compared to those observed for secondary structure elements in known protein structures. Finally, the conformational state at each residue position is averaged across the multiple sequence alignment and the most dominant state is used.

3.6.6 Solvent accessibility

The exterior portion of a protein that is in contact with solvent molecules is referred to as the accessible surface area. This surface is generally described as that which is in van der Waals contact with a probe as it moves across the surface of the molecule. The probe is a sphere of radius 1.4 Å which approximately represents a water molecule. The accessibility of atoms or groups of atoms within a protein crystal structure to solvent molecules has been assessed using an algorithm developed by Lee and Richards (1971). This showed that about 40-50% of the surface area of a typical globular protein is occupied by nonpolar atoms.

3.6.7 Tertiary structure predictions

Fold recognition by threading or inverse folding methods attempt to match the one-dimensional information contained in protein sequences directly to three-dimensional fold structures (Bowie and Eisenberg, 1993). The rationale behind these methods is that by aligning a residue from the sequence of interest with a target residue in a known fold structure, the tertiary location of the target residue can specify not only amino acid type but also the secondary structure, the exposure and the spatial interaction with other residues. These techniques enable the detection of more distant sequence-to-structure relationships than sequence comparisons alone, which can fail to recognise highly similar protein structures that have a sequence similarity between 20% and 30%. In general these methods use two principal measures of residue-to-environment compatibility, both of which are based on the analysis of known structures. The first principle is referred to as the solvation preference and it characterises a residue according to its degree of exposure to solvent. Typically, polar residues are expected to occur at exposed sites and apolar residues are expected to occur at buried sites. The second principle or the neighbour preference measure characterises the residue by what type of residues are found nearby. Thus, if a residue is surrounded by apolar residues then that residue is expected itself to be apolar.

THREADER is a fold recognition program that uses a dynamic programming algorithm based on a combination of neighbour and solvation preferences (Jones *et al.*, 1992). A library of 102 unique protein fold structures at 2.8 Å resolution or better was constructed. Each fold is considered only as a chain tracing through space. The test sequence is optimally fitted ("threaded") to the backbone coordinates of each fold and the pseudo-energy of each fitting is evaluated by the summing of pairwise interaction potentials between amino acids. This pseudo-energy evaluation is performed using a set of knowledge-based potentials that are derived from statistical analysis of known protein structures. The interaction energies in THREADER are expressed as Z-scores:

$$\text{Z-score} = (\text{Energy} - \text{mean}) / \text{standard deviation} \qquad \text{Equation 3.12}$$

for the pairwise or solvation energies. The pairwise energy Z-score is used for ranking the library folds and the lowest pairwise interaction energy identifies the most probable match. A Z-score less than -3.5 is regarded as very significant and is probably a correct prediction, but any such prediction needs to be substantiated by other methods, i.e secondary structure prediction and solvent accessibilties.

3.7 Analysis of known crystal structures

From the analysis of existing crystal structures, protein folds have been classified into five classes according to their secondary structure (Richardson, 1981; Kabsch and Sander 1983*a*). These classes are: (i) all α, with all α helices and no β-strands; (ii) all β, containing only β-strands and no α-helices; (iii) α/β, in which the polypeptide chain alternates between α-helices and β-strands; (iv) α+β, where α-helices and β-strands occur in separate parts of the structure; and (v) coil, in which there is little or no regular secondary structure. Procedures exist for the analysis of secondary structures and solvent accessibilities.

3.7.1 DSSP (Dictionary of Secondary Structure of Proteins)

DSSP is a program that gives the secondary structure elements of a known crystal structure. The output from the program allows the comparison of secondary structure elements between crystal structures and predicted elements. DSSP works by analysing the pattern of hydrogen bonds, inter-C^{α} distances, and inter-C^{α} torsion angles plus solvent exposure of each residue within a file of protein atomic coordinates. A dictionary of eight conformational states was defined by analysing a set of atomic coordinates for 62 different globular proteins using this program (Kabsch and Sander, 1983). The output from DSSP for a known structure can be compared to the predicted secondary structure elements and gives a measure of the reliability of the secondary structure prediction.

3.7.2 COMPARER

PSA, which is part of the of the COMPARER suite of programs, can be used to assess the solvent contact areas of atomic coordinate protein models. PSA uses an algorithm (Richmond and Richards, 1978) with side-chain accessibility calculations and normalisation

carried out as described by Hubbard and Blundell (1987).

3.8 Model building

Once an homologous fold adopted by a target sequence has been identified, the three-dimensional structure can be modelled. INSIGHT II (release 97.0; Biosym/MSI, San Diego, USA) is a suite of molecular graphics and computational chemistry software that is used for building and refining domain models. The homologous (or analogous) fold structure is used as a template, from which an atomic coordinate model is built by a rigid fragment assembly method that is implemented using the HOMOLOGY program. HOMOLOGY imports the template structure and facilitates the alignment of its sequence to that of the protein structure to be modelled.

The structurally conserved regions (SCRs) in the core of the fold must be defined. The atomic coordinates of the template SCRs are then copied to the model, except when differences occur between the sidechain atoms of the two proteins, in which case these are modelled from a library of amino acid structures.

The loops which join the SCRs are modelled next. The conformations of SCRs are restricted by hydrogen bonding constraints, but such conformational restriction is not usually observed for loops and they therefore exhibit a higher degree of sequence and structural divergence between homologous proteins. Consequently loops are more difficult to model. Where possible, loops from the template structure are used. This type of loop is termed a designated loop in HOMOLOGY and the template coordinates for designated loops are copied to the target model.

Other loops are termed searched loops, and these occur when the loop length is different between the template and target proteins. Searched loops are modelled using a database of compatible loop structures (Hobohm and Sander, 1994). HOMOLOGY provides an algorithm that enables the database to be searched for loops that have the desired length and which best satisfy the geometric constraints required to join two SCRs. This search

algorithm calculates an α-carbon distance matrix for the model residues on either side of a loop and this is compared to α-carbon distance matrices from structures in the database. The database loop structures that give the best fit to a model loop are defined as those with the lowest root mean-squared (RMS) distance values:

$$\text{RMS distance} = \left(\sum_{i=1}^{N} \frac{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2}{N} \right)^{\frac{1}{2}} \qquad \text{Equation 3.8}$$

where the summation is performed for 1 to N pairs of α-carbon atoms between the two structures; and x, y, z and x_0, y_0, z_0 are the atomic coordinates of each corresponding α-carbon atom between the two structures. The selection of a loop structure is typically governed by conditions such as: suitable mainchain torsion angles and distances between α-carbon atoms at the points where loops are joined to SCRs; the loop does not contain secondary structure; there is no gross steric overlap of the loop with other regions of the model. When a searched loop is selected, its coordinates are copied to the model.

After atomic coordinates have been assigned to every model residue, manual rotamer searches can be used to relieve steric overlap. HOMOLOGY contains a library of commonly occurring sidechain rotamers and these are tested individually in an attempt to reduce the steric overlap of poorly modelled sidechains.

3.8.1 Model refinement

A model that has been built using a rigid fragment assembly method may contain structural artefacts. These include the substitution of large sidechains for small ones, strained peptide bonds between segments taken from different reference proteins, and non-optimum conformations for the loops. In order to overcome these artefacts, models can be refined using energy minimization.

DISCOVER is a molecular simulation program within INSIGHT II that can perform many routines, including energy minimisation, template forcing, torsion forcing, and dynamic

trajectories as well as the calculation of interaction energies, derivatives, mean square displacements, and vibrational frequencies (DISCOVER 2.9.7/95.0/3.0.0 user guide). The consistent-valence forcefield was used to represent the potential energy of the molecular system. This forcefield contains terms which describe the energies necessary to stretch a bond, distort the angle between three atoms, rotate atoms about their bond axis, and move an atom out of the plane defined by the three atoms to which it is bonded, as well as extra terms that describe the energies representing the coupling effects of one of the above energies with another (cross terms), and associated with the attractive, repulsive and electrostatic forces between atoms that are not bonded to one another (charges). Energy minimization is used to produce a model that is chemically and conformationally reasonable.

The first stage of protein minimization would commonly be *splice repair*. This minimizes the energy of the peptide bonds at the junctions between regions from different reference structures. Bond lengths and mainchain omega torsion angles are displayed by the program so that the progression of the minimization can be monitored. The *relax* option enables different minimization strategies to be set up. The model is divided into individual fragments from different reference structures, SCRs and loops, mainchains and sidechains, and mutated sidechains and non-mutated sidechains, and from these different regions can be selected for minimization.

3.8.2 Structure validation

After building an atomic coordinate model, whether by means of X-ray crystallography, NMR or homology modelling techniques, it is important to assess its quality. PROCHECK (Laskowski *et al.*, 1993) is a suite of programs that assesses a PDB-format atomic coordinate file using stereochemical parameters derived from high-resolution protein crystal structures (Morris *et al.*, 1992), and bond lengths and bond angles derived from a comprehensive analysis of small-molecule structures (Engh and Huber, 1991). The stereochemical quality of the model is output as a residue-by-residue listing that enables the clear identification of regions that are in error. A useful feature of PROCHECK is that it produces a Ramachandran plot of the ϕ (phi) and ψ (psi) mainchain torsion angles

(Ramachandran and Sassiekharan, 1968).

3.8.3 Surface electrostatic potentials

Electrostatic interactions play a central role in a variety of biological processes (Honig and Nicholls, 1995). DELPHI is part of the INSIGHT II suite of programs and is used to calculate the electrostatic properties of charged molecules. The output from DELPHI can be mapped onto the molecular surface of the protein via its interface with INSIGHT II. The electrostatic potential in and around macromolecules can be calculated using a finite difference solution to the Poisson-Boltzmann equation:

$$\nabla\cdot[\varepsilon(\mathbf{r})\nabla\cdot\varphi(\mathbf{r})] - \varepsilon(\mathbf{r})\kappa(\mathbf{r})^2 \sinh[\varphi(\mathbf{r})] + 4\pi\rho^f(\mathbf{r})/k_BT = 0 \qquad \text{Equation 3.13}$$

where $\varphi(\mathbf{r})$ is the electrostatic potential at any point in space, and has units of k_BT/e (k_B is the Boltzmann constant; T is the absolute temperature and e is the elementary charge); ε is the relative permittivity; and ρ^f is fixed charge density (in proton charge units). The term $\kappa^2 = 1/\lambda^2 = 8\pi q^2 I/ekT$, where λ is the Debye length and I is the ionic strength of the bulk solution. The variables φ, ε, κ and ρ are all functions of the vector $\mathbf{r}$.

The molecular surface is defined as the contact surface formed between the van der Waals envelope of the molecule and a probe molecule of 1.4 Å radius. All internal regions are assigned a low value of ε (around 2-4), whereas exterior regions are assigned the standard relative permittivity of water (ε of around 80). Using iterative procedures for the solution of equation 3.13, $\varphi(\mathbf{r})$ can be calculated for a molecule in solution of arbitrary ionic strength. In the context of proteins, unique patterns of $\varphi(\mathbf{r})$ are seen in many cases to have an important functional role (Honig and Nicholls, 1995).

Chapter 4

Experimental:
Preparation of factor B derived proteins from blood and by recombinant expression.

4.1 Introduction

Factor B is a key component of the alternative pathway of complement. It is structurally and functionally homologous to the classical pathway protein C2 (Mole *et al.*, 1984). Factor B comprises the serine protease proenzyme which is the zymogen of the protease component of the alternative pathway C3 and C5 convertases. When the complement system is activated, the major opsonin of complement C3 is cleaved into the fragments C3a and C3b. Removal of C3a by proteolytic cleavage by the C3 convertase induces a conformational change in C3b which leads to the exposure of the internal thiolester, which is normally buried in native C3. C3b, or C3(H_2O), forms a Mg^{2+} ion dependent complex with factor B, whereupon factor D, another serine protease, cleaves factor B in the the C3bB complex, to form two fragments, Ba and Bb. Factor D cleaves factor B by hydrolysis of the Arg233-Lys234 peptide bond to release the Ba (~30 kDa) and Bb (~61 kDa) fragments. Fragment Bb remains, transiently, bound to C3b to form the C3 convertase complex, C3bBb, which activates more C3 (Law and Reid, 1995). The C3bBb complex can then bind to a second C3b molecule, to form the C5 convertase, C3bBb3b. C5 is a homologue of C3, but without a reactive thiolester. The cleavage of C5 links the alternative pathway to the formation of the membrane attack complex. If activated C3 is deposited on an activating surface of the alternative pathway, it can serve as a seed for the positive amplification loop which operates explosively. The amplification of C3 activation which occurs by this mechanism is essential for opsonisation of complement-activating materials. The activities of factor B are thus essential for activation of the alternative pathway.

Factor B is a single chain glycoprotein (~90 kDa) composed of five domains (Mole *et al.*, 1984) (Figure 4.1) When examined by electron microscopy, factor B appears to have a three-lobed globular structure (Smith and Müller-Eberhard,1984). The amino-terminal lobe, the Ba fragment (~30 kDa) consists of three independently-folded short consensus repeat domains (also known as complement control protein domains) (Morley and Campbell,1984), each approximately 60 amino acids in length. Ba has been shown to be involved in the initial interaction of factor B and C3 during the formation of the C3 convertase (Hourcade, 1995; Pryzdial and Isenman, 1987; Ueda *et al.*, 1987). The Ba fragment has

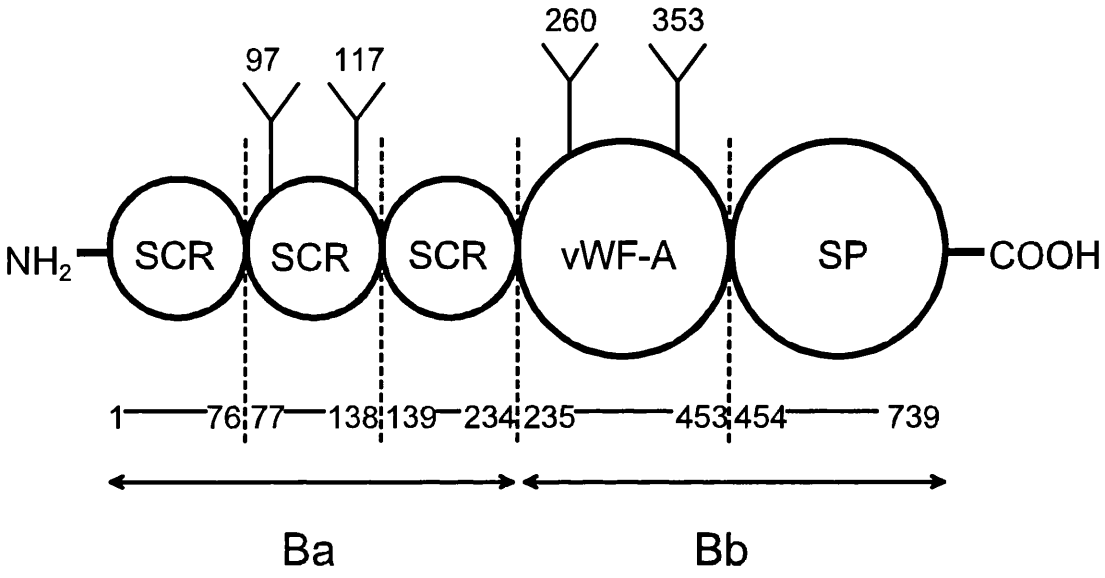


Figure 4.1 Schematic outline of the domain structure of Factor B; SCR (short consensus/ complement repeat) domains, vWF-A (type A von Willebrand Factor) domain and SP (serine protease) domain. Potential N-linked oligosaccharide sites are shown on the second SCR and the vWF-A domains. The residue numbering corresponds to the primary sequence reported by Mole *et al.* (1984).

been reported to suppress B lymphocyte functions in vitro (Ambrus *et al.*, 1990). Elevated levels of Ba in the plasma of patients with chronic renal failure may contribute to the defective immune response in these patients (Opperman *et al.*, 1988) The Bb fragment (~60 kDa) corresponds to two lobes of the factor B structure. The carboxyl-terminal lobe of Bb is a serine protease (SP) domain (Christie and Gagnon, 1983; Campbell *et al.*, 1984). The SP domain is the functional domain involved in the enzymatic cleavage of peptide bonds in complement by factor B, factor D, factor I, C1r,C1s and C2 (Reid *et al.*, 1986). The SP (~35 kDa) domain of factor B is the site of catalytic action of the C3 convertase of the alternative pathway. The α-chain of C3 is cleaved between Arg77 and Ser78, producing two biologically active fragments: the activation peptide C3a (~8,900) and the activated protein C3b (~171 kDa) (Hugli, 1975; Tack and Prahl, 1976). This domain belongs to the same superfamily as β-trypsin and α-chymotrypsin with the double sub-domain β-sheet motif structure very similar to that of other serine proteases, but contains large surface polypeptide insertions which are atypical of this overall protein fold (Perkins and Smith, 1993). The SP domains of the complement system, as with those of the blood clotting system, often differ from the broadly specific digestive serine proteases by having a large N-terminal polypeptide chain (as distinct from a small activation peptide) added to the domain. The SP domain of factor B, as with that of C2, is unusual in that the cleavage leading to the activation of the proenzyme occurs at the amino-terminal region of the vWF-A domain. Ordinarily such a cleavage site is at the immediate N-terminus of the SP domain. All three domain types have been implicated in the interactions of factor B with C3 (Sim and Perkins, 1990).

The middle lobe of factor B (amino-terminal domain of Bb) has sequence homology to about 210 amino acids present in a wide variety of other proteins (reviewed in Colombatti and Bonaldo, 1991; Perkins *et al.*, 1994). It was termed the vWF-A domain since it corresponded to one of the four different domain types (A to D) in von Willebrand Factor (Sadler *et al.*, 1985; Jenkins *et al.*, 1998). This domain type also occurs in proteins of the extracellular matrix (cartilage matrix protein, and collagen types VI, VII and XIV) and other proteins of the immune system (the leucocyte integrins CR3, CR4, LFA-1; the VLA-1, VLA-2 integrins; complement C2). It seems likely that the vWF-A domain in many of these

proteins is implicated in ligand binding. A total of 75 sequences in 25 proteins have been analysed to show a relatively low degree of sequence conservation and no conserved disulphide bridges (Perkins *et al.*,1994).

In vitro site directed mutagenesis is a valuable tool for the study of protein structure-function relationships. Methods for performing this technique generally require single stranded DNA (ssDNA) as the template and are labour intensive or technically difficult. The "QuickChange" method allows site specific mutation in virtually any double-stranded plasmid and requires no specialised vectors. The rapid four step procedure (Figure 4.2) is simple to perform and uses plasmid DNA purified by miniprep (see Methods 4.2 below). The basic procedure utilises a supercoiled, double stranded DNA (dsDNA), two oligonucleotide primers containing the desired mutation and a temperature cycler. The oligonucleotide primers, each complementary to opposite strands of the vector, extend during temperature cycling by means of *Pfu* DNA polymerase. *Pfu* DNA polymerase will replicate strands of supercoiled double stranded DNA (dsDNA) without displacing the mutant oligonucleotides. On incorporation of the oligonucleotide primers, a mutant plasmid containing staggered nicks is generated. The reaction mixture is treated with *Dpn* I endonuclease which is specific for methylated and hemimethylated DNA (target sequence 5'-G^{m6}ATC-3'). This will digest the parental DNA template and so will select for the mutated synthesised DNA. This works because DNA isolated from almost all *E.coli* strains is *dam* methylated and therefore susceptible to cleavage by *Dpn* I. The nicked vector DNA can then be transformed into *E.coli* competent cells.

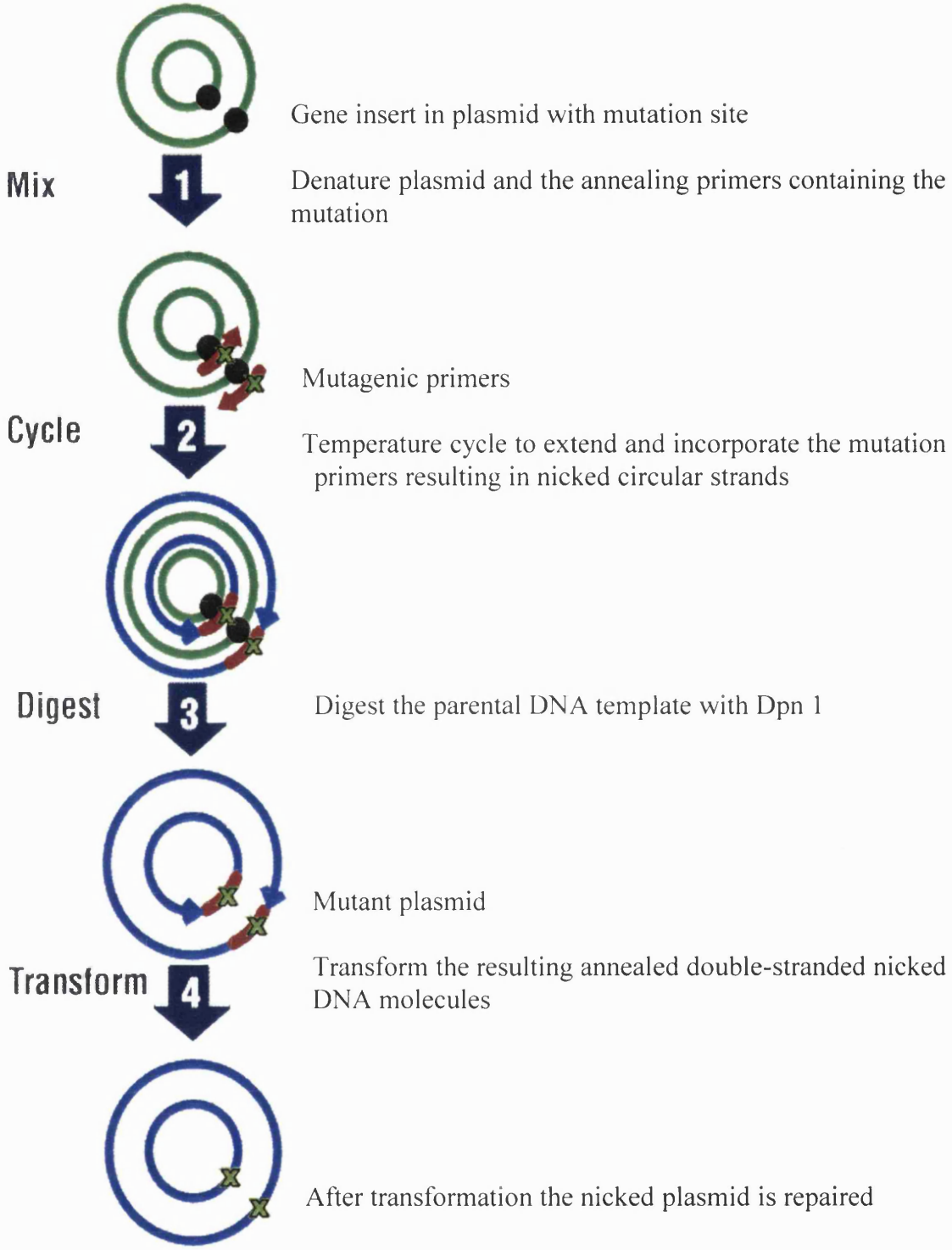


Figure 4.2 Schematic representation of the "QuickChange" mutagenesis method (Stratagene, La Jolla, USA). A four-step process of mixing, cycling, digestion and transformation is described. The original and mutant plasmids are represented by green and blue circles respectively. The mutation target site is represented by a black dot and the mutagenic primers by brown arrows. A green X represents the desired mutation. Adapted from the Stratagene web resource (http://www.stratagene.com).

4.2 Materials

Glutathione-agarose beads (G-4510) and human thrombin (T-3010) were obtained from Sigma Chemical Co., Poole, Dorset, U. K., as were the constituents of the buffer used to lyse *E. coli* cells [DNAse (D-7291), lysozyme (L-6876), Triton X100 (T-6878) and benzamidine (B-6506)]. Chromatography columns were from Pharmacia, Milton Keynes, Bucks., U.K. Fresh-frozen plasma for the purification of complement components came from the Oxford Regional Blood Transfusion Service, John Radcliffe Hospital, Oxford, U.K. The serine protease inhibitor Pefabloc-SC (4-(2-aminoethyl)-benzene-sulphonyl-fluoride) was obtained from Pentapharm AG, Basel, Switzerland.

4.3 Methods

4.3.1 GST-vWF-A-218 expression system

A system for the expression of a recombinant vWF-A domain of factor B in *E. coli* MC1061 cells was kindly provided by Dr. R.B. Sim, (Williams *et al.*, 1994). This system expressed the protein as a fusion product with glutathione-S-transferase (GST). A summary of this work done by Williams *et al.* (1994) follows. The region encoding the vWF-A domain was amplified from a full length factor B cDNA clone named pFB-2 (Steinkasserer *et al.*,1992) by polymerase chain reaction (PCR). The factor B cDNA clone was generously provided by Dr W. Schwaeble, Dept of Microbiology and Immunology, University of Leicester. The middle five exons of the factor B gene encode the vWF-A domain of factor B (Steinkasserer *et al.*, 1992). The first exon begins with Gly229 and the last ends with Ile444. Oligonucleotides were designed to flank the region of interest with the introduction of a *Bam* HI restriction site at the 5′ end and an *Eco* RI restriction site at the 3′ end to enable sub-cloning into the pGEX-2T vector in the correct orientation (5′ oligonucleotide CG GGA TCC GGG GAA CAA CAG AAG CGG AAG; 3′ oligonucleotide GGA ATT CTA GAT CAT TTG GTA GAA AAC ATC TTC). The amplified cDNA insert (663 bp) produced was digested with *Eco* RI and *Bam* HI and ligated into the pGEX-2T vector pre-cut with the same enzymes. The ligation mixture was used to transform MC1061 competent cells. Individual bacterial clones containing cloned cDNA were identified by the production of correct-sized PCR products using the oligonucleotides described above from DNA mini-preps (Birnboim

et al., 1979). Sub-cloned DNA from a large-scale plasmid purification (Sambrook *et al.*, 1989) was sequenced using a 'T^7 Sequencing Kit' following the manufacturer's protocol. The sequence was found to be identical to that found in the pFB-2 clone from which the fragment was amplified. The modifications to the ends of the amplified product were found to be as intended. As this recombinant product contained 218 residues, it was denoted vWF-A-218. The *E. coli* MC1061 cells containing the vWF-A insert were stored in LB broth containing 15% glycerol at -70°C.

4.3.2 Preparation of the vWF-A-218 domain

vWF-A-218 was expressed and purified from *E. coli* (provided by Dr. R.B. Sim) by a combination of the methods in Steinkasserer *et al.* (1992) and Smith and Johnson (1988). A starter culture was grown up at 37°C overnight in Lennox broth with 100 μg/ml ampicillin. This was diluted 1:10 with fresh Lennox broth/ampicillin and incubated for 90 min. The cells were then induced with 1 mM isopropyl-β-D-galactoside and grown for a further four to five h. The cells were pelleted by centrifugation using a 850 S rotor in an IEC-PR7000 centrifuge at 6,000 r.p.m. (3000 *g*) for five min and then resuspended in lysis buffer on ice for one h. The lysis buffer was a hypertonic phosphate buffered saline (MTPBS) (16 mM NaH_2PO_4, 54 mM Na_2HPO_4, 150 mM NaCl, pH 7.2) containing lysozyme (0.2 mg/ml), DNase (1 μg/ml), Triton X-100 (0.1% v/v), Pefabloc (1 mM), dithiothreitol (2 mM), benzamidine (50 mM) and EDTA (ethylenediaminetetraacetic acid)(5 mM). This was followed by sonication on ice for 3 × 20 sec using a MSE sonicator. The cell lysate was spun down using a Sorvall SS-34 rotor for 25 min at 17,000 rpm (22,500 *g*). The fusion protein was then bound to glutathione-agarose (10 ml per litre of culture) by incubating it with the supernatant overnight at 4°C. This was washed on a column with degassed hypertonic phosphate buffered saline and then subjected to cleavage by thrombin (250 ml per five litres of culture) at 37°C for four h to release the recombinant vWF-A-218 domain. The cleaved vWF-A-218 was washed from the beads using degassed hypertonic phosphate buffered saline containing 2 mM dithiothreitol. The thrombin was removed by passing the glutathione-agarose wash over an arginine agarose column (Affinity Chromatography, 0030-100). The material that passed through the column was concentrated under nitrogen using an Amicon

ultra filtration stirred cell with a YM-10 membrane. To prevent dimerisation occurring via the free thiol group, the protein was either kept in reducing conditions in 10 mM DTT (vWF-A-218_{DTT}) or the Cys was oxidised with iodoacetamide (vWF-A-218_{IAM}). In the latter case iodoacetamide was added to a final concentration of 4 mM and the sample was left at room temperature for two min prior to gel filtration. In either case the sample was loaded onto a HiLoad 16/60 Superdex 75 gel filtration column (Pharmacia).

4.3.3 Mutagenesis of vWF-A-218 with the "QuickChange" method

For this thesis project a second recombinant vWF-A domain was prepared by site directed mutagenesis of the pGEX-2T plasmid containing the vWF-A-218 clone using the "QuickChange" site-directed mutagenesis kit (QuickChange 200518, Stratagene, Cambridge, UK). Two cycles of mutagenesis were employed as summarized in Table 4.1. In the first, Cys267 was replaced by Ser and a *Spe* I restriction site (A/CTAGT) was introduced into the mut1 sequence. In the second mutation, the C-terminus was extended by adding the next four residues of the factor B sequence (DESQ), and one *Bfa* I restriction site (C/TAG) was lost in the mut2 sequence (Figure 4.3). Oligonucleotide primers were synthesised at Perkin-Elmer Applied Biosystems (Warrington, Cheshire, U.K.).

4.3.4 Mutagenic primer design

Both the mutagenic primers had to contain the desired mutation and anneal to the same sequence on opposite strands of the plasmid. The primers should be between 25 and 45 bases in length and the melting temperature (T_m) should be approximately 10°C above the extension temperature of 68°C. The following formula was used for estimating the T_m of the primers:

$$T_m = 81.5 + 0.41(\%GC) - \frac{675}{N} - \%mismatch \qquad \text{Equation 4.1}$$

where N was the primer length (number of bases). The desired mutation should be centrally located within the primer with approximately 15 bases of original sequence either side. The

Table 4.1 Primers used for the mutation of the vWF-A-218 cDNA to vWF-A-222 by the "Quick Change" method

Identification*	**Amino acid and DNA sequences****	Direction	**Length (N)**	GC	**T_m (°C)*****
mut1	T G A K K **S** L V N C ACA GGA GCC AAA AAG T**CA** CTA GTC AAC	forward	28	13	74
antimut1	GTT GAC TAG **TG**A CTT TTT GGC TCC TGT G	reverse	28	13	74
mut2	V F Y Q M I **D E S Q** Stop GTT TTC TAC CAA ATG ATC **GAC GAG AGC CAG** TAG AAT TCA TCG TGA CTG	forward	48	21	60
antimut2	CAG TCA CGA TGA ATT CTA **CTG GCT CTC GTC** GAT CAT TTG GTA GAA AAC	reverse	48	21	60

* The mut1/antimut1 and mut2/antimut2 sequences correspond to the Cys267Ser mutation and the four-residue C-terminal extension respectively.

** Characters in bold represent alterations from the original vWF-A cDNA sequence. The underlined TAG motif represents the stop codon required for termination of translation.

*** The melting temperature T_m was calculated according to the Stratagene QuickChange manual:

$$T_m = 81.5 + 0.41(\%GC) - 675/N - \%\ \text{mismatch}$$

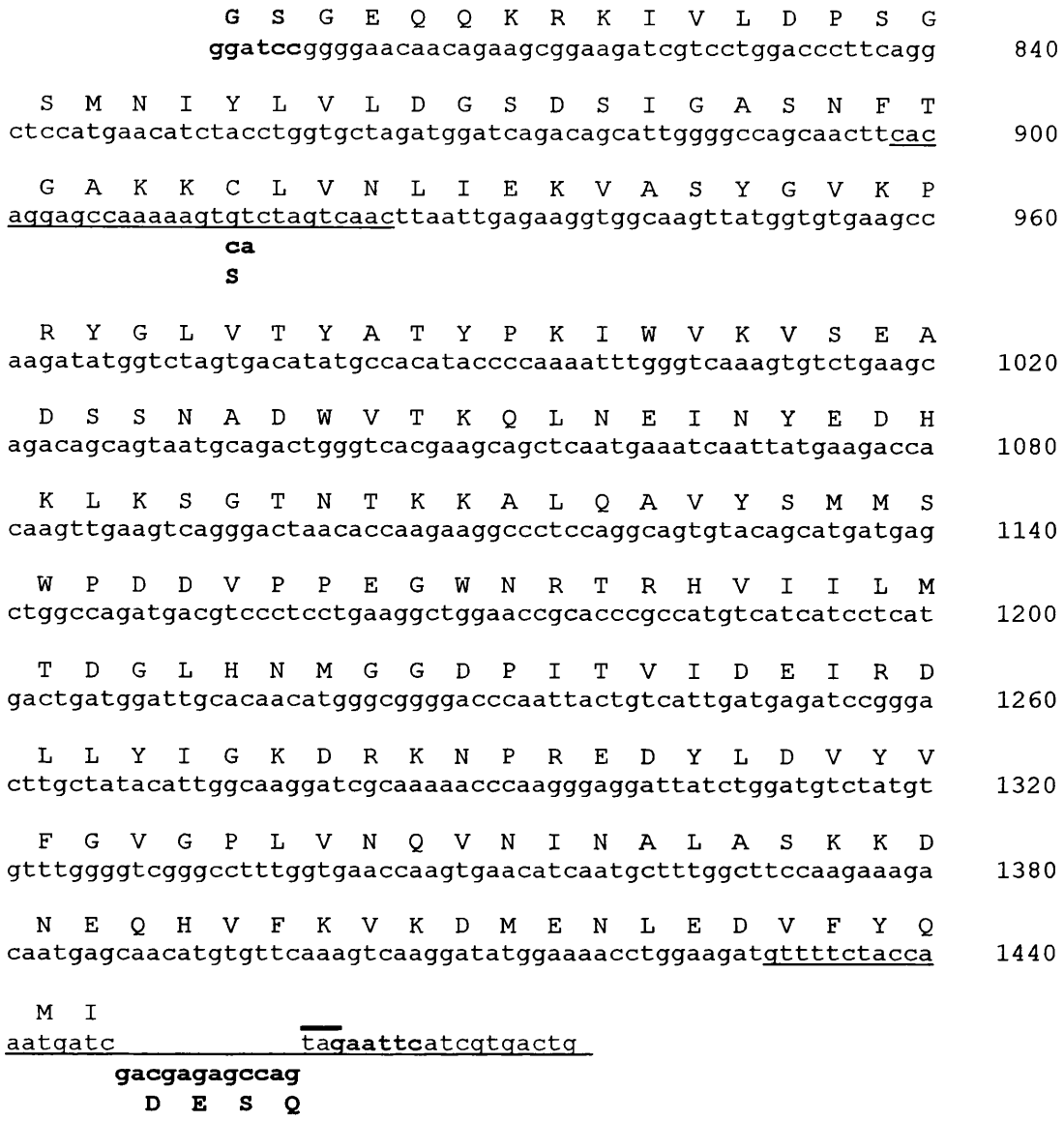


Figure 4.3 The sequence of the cDNA insert (lowercase) of the vWF-A-218 domain in the pGEX-2T expression vector and the translated protein sequence (uppercase). The numbering corresponds to the factor B cDNA sequence numbering. Characters in red and blue represent the *BamH*1 and the *EcoR*1 restriction enzyme sites respectively and characters in bold directly below the vWF-A cDNA sequence represent the desired mutations in vWF-A-222. The residues in green (GS) are derived from the thrombin cleavable linker and not from the factor B sequence. The underlined characters describe the position that the mutant primers will anneal to. The above-lined motif represents the stop codon.

primers should have a minimum GC content of 40% and should terminate in one or more G or C bases.

4.3.5 Preparation of the vWF-A-218 bacterial culture

30 ml of sterile LB broth (containing 100 mg/ml of ampicillin (Sigma)) were inoculated with100 µl of the vWF-A-218 glycerol stock and incubated overnight on an orbital shaker at 37°C at 180 rpm. The bacteria were pelleted by centrifugation on a bench top centrifuge at 3000 rpm. The supernatant was discarded and the tubes were inverted to dry the pellets.

4.3.6 Isolation and purification of vWF-A-218 plasmid DNA

The plasmid was purified using a "Promega Wizard plus minipreps DNA purification system (A7130304)" following the manufacturer's protocol. The plasmid DNA pellets were resuspended in 50 µl of sterile TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0). The optical density of the DNA was measured at both 260 nm and 280 nm in a spectrophotometer to assess concentration and purity of the DNA (a solution containing 50 µg/ml of double stranded DNA has an absorbance of 1.0 at 260 nm; Sambrook *et al.*, 1989). The vWF-A-218 cDNA plasmid concentration was calculated as 2 ng/µl and had a 260/280 ratio of 1.7 (pure DNA has a ratio of 1.8).

4.3.7 Mutagenesis reactions

The vWF-A-218 plasmid was used as the template for the first reaction and vWF-Amut1 plasmid (see later) was selected as the template for the second mutation. The mutagenesis reactions were prepared in sterile snap-fit tubes as follows:

2, 5, 10 or 20 µl of cDNA template

2 µl of mut1 or mut 2 primer (approximately100 µg/ml)

2 µl of antimut 1or antimut 2 primer (approximately100 µg/ml)

5 µl of 10 × buffer

1 µl of dNTP mix primer

made up the volume to 50 µl with dH_2O

1 µl of *Pfu* polymerase (2.5 U/µl)

The reaction mistures were overlaid with mineral oil and subjected to the following temperature cycle:

(1a) 95°C for 30 s

Then 18 cycles of (2a) 95°C for 30 s

(2b) 55°C for 1 min

(2c) 68°C for 11 min

The reaction mixtures were then placed on ice for several minutes and 1 µl of *Dpn* I was added below the mineral oil and mixed by pipetting up and down. This mixture was spun down and incubated at 37°C for two h to digest the non-mutated DNA. The 10 × reaction buffer was composed of the following:

100 mM KCl

100 mM $(NH_4)_2SO_4$

200 mM Tris-HCl (pH 8.8)

20 mM $MgSO_4$

1% Triton X-100

1 mg/ml nuclease free bovine serum albumin (BSA)

4.3.8 Transformation of the mutant plasmid into competent cells

200 µl of Epicurian Coli XL1-Blue supercompetent cells were thawed on ice and 50 µl of these were aliquoted into four precooled 1.5 ml snap-fit tubes. 1 µl of the *Dpn* I treated DNA from each sample reaction was added to each of the separate aliquots of cells. The transformation reactions were gently mixed and incubated on ice for 30 min. The mixtures were then heat pulsed at 42°C for 45 min and replaced on ice for two min. The reaction mixture was added to 0.5 ml of NZY broth preheated to 42°C and incubated at 37°C with shaking (220 rpm) for one h. The entire volume of each transformation was plated out on a separate agar plate containing 100 mg/ml ampicillin and these were incubated at 37°C overnight.

The recipe for NZY broth (1 litre) filter sterilised before use follows:

10 g of NZ amine (casein hydrolysate)

5 g of yeast extract

5 g of NaCl

Autoclave

The following supplement was added prior to use:

12.5 ml of 1 M $MgSO_4$ and 12.5 ml of 1 M $MgCl_2$

10 ml of a 2 M filter sterilised glucose solution

4.3.9 Mini prep of mutant plasmids

Eight colonies were selected from the agar plates of the mutant transformations and the plasmid DNA was purified in a miniprep using the Promega kit. The DNA was resuspended in 40 µl of Tris-EDTA (TE) buffer (10 mM Tris, 1 mM EDTA, pH 7.5) and labelled as vWFCysmut 1-8 for the first mutation and vWFCysext 1-8 for the second.

4.3.10 Restriction digests of the mutant plasmids

10 µl of each sample of the first mutant plasmid DNA (vWFCysmut 1-8) were incubated with 2 µl of 10 × multicore buffer (Promega R999B), 0.5 µl *Spe* 1, with an activity of 18 U/µl (Promega R659A) and 8 µl of dH_2O at 37°C for one h. The digests were analysed by agarose gel electrophoresis on a 1% agarose gel to verify the presence of the mutation. This restriction digest was repeated with 0.5 µl *Spe* I and 0.5 µl *Eco* RI together.

5 µl of each sample of the second mutant plasmid DNA (vWFCysext 1-8) was incubated with 3 µl of 10 x buffer, 0.5 µl *Bfa* I and 25 µl of dH_2O at 37°C for two h. The vWFCysmut7 plasmid, which was known to have contained the first mutation, was also subjected to this reaction. The digests were analysed by agarose gel electrophoresis on a 1% agarose gel to verify the presence of the mutation.

4.3.11 Transformation of the double mutant plasmid into competent cells

MC1061/P3 competent cells (Invitrogen) were thawed on ice and 100 µl of these were aliquoted into a precooled 1.5 ml snap fit tube. 5 µl of 0.5 M β-mercaptoethanol were added and the tube was incubated on ice for 10 min, swirling the tube every two min. 5 µl of

vWFCysext8 was added to the cells, swirled to mix, and incubated on ice for 30 min. The reactions were then heat pulsed at 42°C for 45 min and replaced on ice for two min. The reaction mixture was added to 0.9 ml of SOC medium preheated to 42°C and incubated at 37°C with shaking (220 rpm) for one h. 100 µl of the transformation reaction was plated out on an agar plate containing 100 mg/ml ampicillin and this was incubated at 37°C overnight. Eight colonies were selected from the plate and the plasmid DNA was purified in a miniprep. The plasmid DNA was resuspended in separate aliquots of 40 µl of TE buffer and labelled as mutant 1-8. Glycerol stocks were prepared from each of these colonies and labelled similarly to the plasmid DNA.

4.3.12 Screening of the MC1061 mutants for expression

A culture was grown from each of the glycerol stocks in 15 ml of sterile LB broth containing 100 mg/ml ampicillin until it reached an OD_{600} (1 cm) of 0.6. IPTG was added to a concentration of 1 mM and the cultures were grown for a further three h. A culture was also prepared from the untransformed MC1061 cells. Samples from these cultures were screened for expression by Western blot analysis from a 10% SDS PAGE gel. A polyclonal antiserum to factor B was used to identify the GST-vWF-A-222 fusion protein.

4.3.13 Western blot analysis of the bacterial cell lysates

Two 10% SDS-PAGE gels were prepared (Table 4.2). After electrophoresis one of the gels was stained with Coomassie blue. The other was blotted onto PVDF (BioRad) membrane using a semi-dry blotter (Pharmacia) in 39 mM glycine, 48 mM Tris, 0.0375% (w/v) SDS at a current of 2.5 mA/cm^2 for one h. 5% Marvel in phosphate buffered saline (PBS; 0.5 mM $MgCl_2$, 2.5 mM KCl, 1.5 mM KH_2PO_4, 140 mM NaCl, 8.1 mM Na_2HPO_4) was used to block unbound sites on the PVDF membrane. An anti-factor B polyclonal rabbit serum (1 in 1000 dilution) was incubated with the membrane in PBS with 5% Marvel on a rotary mixer for one hour at 20°C. The antibody solution was then removed and the membrane was rinsed four times. An anti-rabbit IgG alkaline phosphatase conjugate (Sigma, A-3812) was diluted to 1/6500 and incubated with the membrane for one hour at 20°C (second antibody step). The membrane was rinsed four times with PBS with 5% Marvel and

Table 4.2 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) system for the resolution of proteins by their molecular mass

Component	**Stacking gel 15 mls**					**Resolving gel 5 mls**
	0.06	0.08	0.1	0.12	0.15	0.05
30% acrylamide	3	4	5	6	7.5	0.83
1.5 M Tris-HCl (pH 8.8)	3.8	3.8	3.8	3.8	3.8	-
1.0 M Tris-HCl (pH 6.8)	-	-	-	-	-	0.63
distilled water	7.9	6.9	5.9	4.9	3.4	3.4
10% SDS	0.15	0.15	0.15	0.15	0.15	0.05
10% APS	0.15	0.15	0.15	0.15	0.15	0.05
TEMED	0.01	0	0	0	0	0.005

Abbreviations: APS, ammonium persulphate (Sigma); SDS, sodium dodecyl sulphate (BDH); TEMED, NNN'N'-Tetramethylenediamine (BDH))

Acrylamide solution

29% acrylamide (BDH)

1% NN'-Methylenebisacrylamide (BDH)

5X Tris-glycine electrophoresis buffer (25 mM Tris, 191.5 mM glycine, 0.1% SDS):

45 g Tris base

216 g glycine (Sigma)

15 g SDS (BDH)

volume made up to three litres with distilled water

(Adapted from Sambrook *et al.*, 1989).

then rinsed with 0.15 M Tris-acetate, pH 9.6. The substrate solution for the alkaline phosphatase was prepared as follows:

50 ml 0.15 M Tris-acetate pH 9.6 (Sigma)

50 ml H_2O

5 ml 0.1% (w/v) 4-nitro-blue-tetrazolium chloride (Sigma) in 0.15 M Tris-acetate pH 9.6

0.5 ml 0.5% (w/v) 5-bromo-4-chloro-3-indolyl-phosphate p-toluidine (Sigma)

0.2 ml 2 M $MgCl_2$ (BDH)

The substrate was added to the membrane and incubated at 20°C for approximately 20 minutes until a purple (positive) colour appeared.

4.3.14 Preparation of the vWF-A-222 domain

A large scale culture was prepared in a similar manner to that of the original vWF-A-218 domain (see section 4.3.2) using the mutant 1 (the double mutant transformation in the MC1061 cells). The method was identical with the exception that neither DTT nor IAM was used in the preparation and it was not necessary to maintain a reducing environment with degassed buffers.

4.3.15 Preparation of factor B, Bb, Ba, the SP domain and C3

Factor B was purified using dye-ligand affinity chromatography of human plasma following the procedure of Williams and Sim (1993), followed by ion exchange chromatography on a mono Q anion exchange column (dimensions 5 × 50 mm). C3 was purified from human plasma following the procedure of Dodds (1993). Bb, Ba and the SP domain were prepared by proteolytic cleavage of factor B following the procedure of Lambris and Müller-Eberhard (1984). The optimum conditions for the production of the SP domain were determined by a time-course digestion using a range of concentrations of elastase between 0 and 8% w/w. The domain was then produced by an overnight digestion using a 0.5% w/w ratio of elastase to factor B at 37°C in 25 mM Tris, 100 mM NaCl, pH 7.4. The digest was chilled on ice, diluted 10 times with 25 mM Tris, pH 7.5 and loaded onto a SP-

Sepharose fast flow ion-exchange column (dimensions 26 × 130 mm) which had been pre-equilibrated with 25 mM Tris, pH 7.5. The column was developed with a linear gradient of 0 to 500 mM NaCl in 400 ml and the SP domain eluted as a single peak at a concentration of about 300 mM NaCl. All the protein preparations were concentrated under nitrogen using an Amicon ultra filtration stirred cell with a YM-10 membrane before being loaded onto a HiLoad 16/60 Superdex 75 gel filtration column (Pharmacia).

4.3.16 Characterisation of protein preparations

The protein preparations were analysed by SDS-PAGE (8% gel for factor B and 10% gels for the other preparations) using an ATTO mini gel apparatus (Genetic Research Instrumentation ltd., Essex, UK) to determine the approximate molecular masses, the sample purity and an approximation of the concentrations. Samples were denatured in 8 M urea in the presence of either 20 mM dithiothreitol (reducing) or 40 mM iodoacetamide (non-reducing) and loaded onto the gel. Rainbow protein molecular weight markers were used as a reference (Amersham). Isoelectric focusing-PAGE was also used to analyse the protein preparations using PhastGel IEF 3-9 gels on a Phast system (Pharmacia) and with Ampholine PAG plate 3.5-9.5 (Pharmacia) on a multiphor system. Isoelectric focusing markers (Sigma, I-3018) were used as a reference. The protein bands were stained using Coomassie blue. Protein concentrations were determined spectrometrically using a Beckman DU-70 spectrophotometer by their absorbance at 280 nm. The extinction coefficients, A_{280} (1cm, 1%), were calculated from the total of Trp, Tyr and Cys residues in the sequence (Perkins 1986) to be: Ba, 16.7; vWF-A-218, 15.6;-222, 15.2; SP, 11.1; Bb, 12.5; factor B, 13.8.

The samples of the vWF-A-218 and the SP domain destined for N-terminal sequence analysis were run on SDS-PAGE under reducing conditions and then electroblotted onto a Problott membrane (Applied Biosystems) in a Bio-Rad mini Trans-Blot electrophoretic transfer cell. The blots were stained with Coomassie Brilliant Blue and the bands corresponding to the protein of interest were excised. The N-terminal amino-acid sequences were obtained using an Applied Biosystems 470A protein sequencer and Applied Biosystems 120A analyser. To determine the N-terminal sequence of the SP domain, 15 cycles of Edman

degradation were performed using an Applied Biosystems (Perkin-Elmer) 494A Procise protein sequencer, courtesy of Dr A. Willis, MRC Immunochemistry Unit, Department of Biochemistry, University of Oxford.

To verify the absence of free sulphydryl groups, the Ellman assay for thiols was performed on the SP domain (Lundblad, 1995). The change in absorbance of 0.15 mM 5,5'-dithiobis-2-nitrobenzoic acid on the addition of glutathione in concentrations between 0 to 2 mM was measured at 412 nm in a 1 cm path length on a Beckman DU-70 spectrophotometer. The resulting calibration curve was used to determine the number of thiol groups in the SP domain (Figure 4.4). 1 ml of the SP domain at 1.2 mg/ml (0.03 mM) in PBS was added to 1 ml of the 5,5'-dithiobis-2-nitrobenzoic acid stock, and the absorbance at 412 nm was recorded after it had reached a maximum. The assay was repeated with the SP domain denatured using 6 M guanidine hydrochloride. All the buffers were pre-filtered with a 0.2 μm filter and then deoxygenated by bubbling N_2 gas through them.

4.3.17 Mass spectroscopy data collection

Matrix-assisted laser desorption ionization (MALDI) mass spectrometry (Beavis and Chait, 1996) was performed with vWF-A-218 domain preparation on a VG-TOF spectrometer at the University of London facility at the School of Pharmacy. Samples were in 20 mM acetic acid and were loaded in a 1: 1 ratio with the matrix sinapinic acid (*trans*-3,5-dimethoxy-4-hydroxycinnamic acid) to aid in desorption. The instrument was calibrated externally using cytochrome c and insulin. To measure the mass of vWF-A-222, electrospray ionization mass spectrometry was performed on a Finnigan Navigator spectrometer with a Waters Alliance liquid chromatography system for sample delivery at the School of Pharmacy. The sample was in deionized H_2O and was loaded in a 1:1 ratio with 50% acetonitrile. The instrument was calibrated using H_2O clusters.

The cleavage of C3 by the SP domain was assayed using SDS-PAGE to quantify the production of the C3a fragment (M_r 8,900). 5 μl of the purified SP domain (2.5 mg/ml in PBS) was mixed with 5 μl of purified C3 (3 mg/ml in PBS) and 5 μl of the appropriate

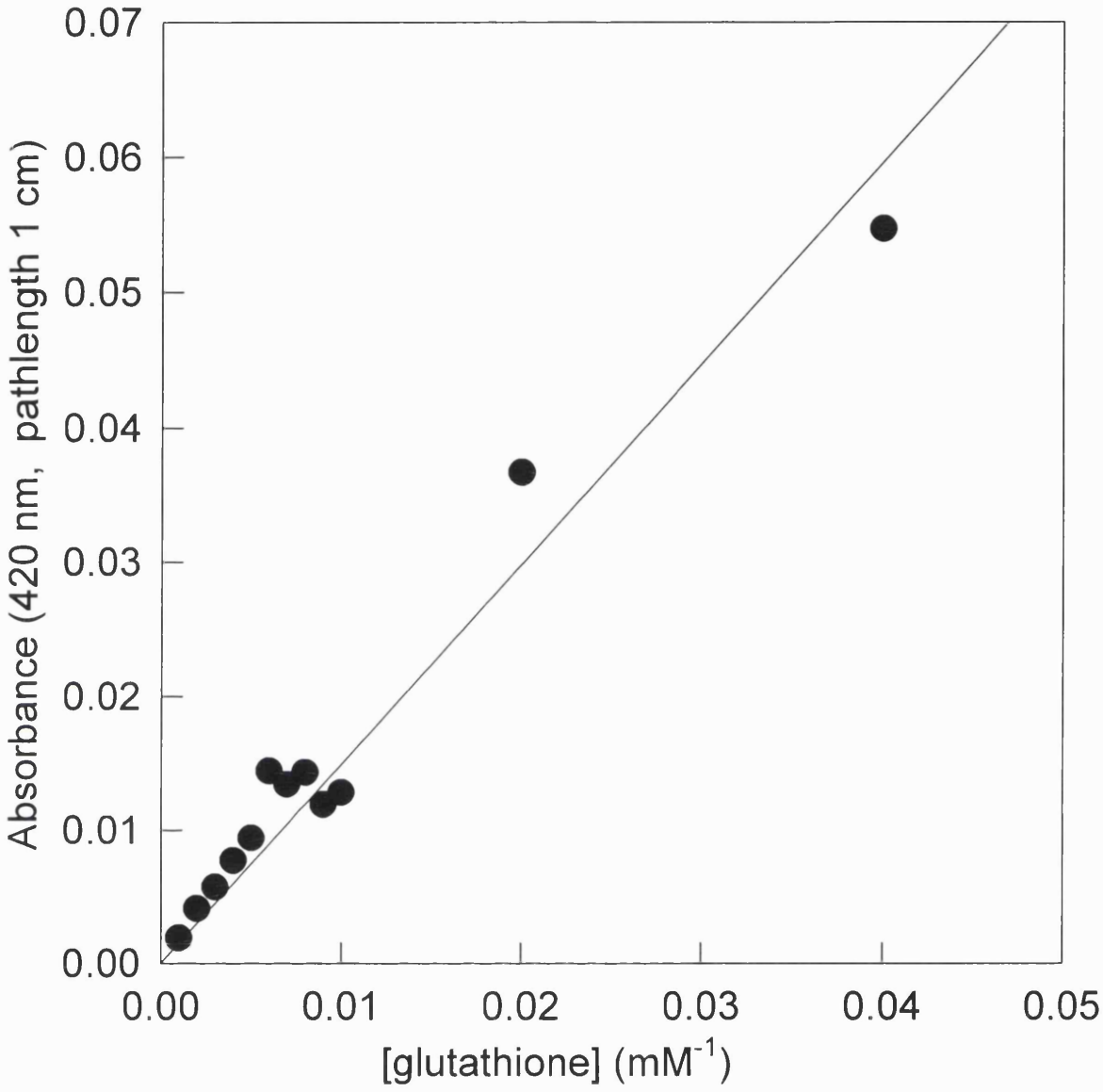


Figure 4.4 Standard curve for the change of absorbance of DTNB at 412 nm against free thiol (glutathione) concentration.

100 mM buffer solution. The samples were incubated at 37°C for two h and overnight. The samples were reduced with dithiothreitol, denatured with urea and loaded onto a 16% tricine gel. The protein bands were visualised by Coomassie blue staining using low-range protein molecular markers (Promega, V5231) as a reference. The C3, and SP preparations incubated in the absence of C3, were used as controls. Band densities were determined by scanning the gel with a HP ScanJet 4C scanner and analysing the image using SigmaGel software (Jandel).

A fluorescent assay for the cleavage of C3 using the synthetic C3 analogue, t-butyloxycarbonyl (Boc)-Leu-Gly-Arg-AMC (Bachem) (Caporale *et al.*, 1981; Gutierrez *et al.*, 1983) was used to assay the activity of the SP domain over the pH range 5-10 in 0.5 unit steps. The C3 cleavage releases a fluorescent coumarin derivative,7-amino-4-isocoumarin (AMC) and so the reaction can be monitored by fluorescence spectroscopy. Stock buffer solutions were prepared similar to those above. 100 µl of the 20 mM solution of the analogue in 0.9 M dimethyl sulphoxide were mixed with 3 ml of the 100 mM buffer stocks and 10 µl of the purified SP domain (2.5 mg/ml in PBS). The reaction mixture was incubated at 37°C in a Perkin-Elmer 50B spectrofluorimeter. The sample was excited with a wavelength of 380 nm and the intensity of the emission at 460 nm was measured for 20 min.

4.3.18 Crystallisation of the factor B proteins

Preparations of the vWF-A-218 domain, the vWF-A-222 domain, the SP domain and whole factor B were subject to crystallisation trials using the hanging drop vapour diffusion technique (Ducruix and Giege, 1992). The vWF-A-218 domain was concentrated to 2 mg/ml only and whole factor B to 3 mg/ml only because the proteins were not soluble above these concentrations. Preparations of both the domain prepared with IAM (vWF-A-218_{IAM}) and the unreduced domain (vWF-A-218_{DTT}) were used. In the latter case the protein had to be maintained in a non-reducing environment with either 10 mM DTT or 10 mM β-2-mercaptoethanol. The vWF-A-222 domain and the SP domain were prepared at two concentrations 3 mg/ml and 7 mg/ml. Trials in the presence of 1 mM $MgCl_2$ or 2 mM EDTA were conducted in parallel as were trials at room temperature, 18° and 4°C. Both a systematic approach, screening with the precipitants ammonium sulphate, polyethylene glycol

(PEG) and 2-methyl-2,4-pentanediol (MPD) and a sparse matrix approach using the Hampton Research crystal screening kits 1 and 2 (Figures 4.5 and 4.6) were employed to screen a wide range of precipitants and additives. All the buffers were filtered prior to use using 2 μm filters. Stock buffer solutions were made using 100 mM Tris, 0.1% NaN_3 at intervals of 0.5 pH unit from pH 3.5 to 9.0. Ammonium sulphate was dissolved in 100 mM Tris, 0.1% NaN_3, to 100% saturation. 80% saturated solutions were then prepared at intervals of 0.5 pH unit from pH 3.5 to 9.0. PEG 3000, 4000, 6000 and 8000 were dissolved in the stock buffer solutions to achieve 30% w/v solutions. Dilutions of PEG 1500, 2000 and MPD were made as required.

The proteins were dialysed into 20 mM Tris, 0.1% NaN_3, pH 7.2 buffer and the solution was centrifuged immediately before use in a microfuge to remove aggregates. Hanging drops were set up on siliconised cover slips by pipetting an equal amount (1-5μl) of the protein solution with the precipitant and inverting the cover slip over the precipitant reservoir. The precipitant reservoir was prepared in the wells of linbro boxes and the cover slips were sealed with grease. The cover slips were examined under crossed polarisers using a Wild stereo microscope after one week, and then on a monthly basis. Any precipitate that formed within the drop was attributed a Q score according to its appearance from the following scheme: cloudy precipitates, 1.0; gelatinous or particulate precipitates, 2.0; spherulites, 3.0; needles, 4.0; plates, 5.0 and prisms, 6.0.

A micro seeding technique was employed whereby a bristle was brushed through the granular precipitate from the drops of a particular screen and then stroked through a freshly prepared drop with similar conditions. This was repeated to screen systematically a whole set of parameters. Macroseeding techniques were also employed whereby whole crystals were transferred to freshly prepared drops.

If precipitant was not observed in the drop, the protein or precipitant concentrations were increased and the screens were repeated. If crystals were not observed in the screen, further systematic screens were conducted covering other parameters, such as precipitant

1. 30% MPD, 0.1 M sodium acetate pH 4.6, 0.02 M calcium chloride
2. 0.4 M sodium tartrate
3. 0.4 M ammonium phosphate
4. 2.0 M ammonium sulphate, 0.1 M Tris HCl pH 8.5
5. 30% MPD, 0.1 M sodium Hepes pH 7.5, 0.2 M sodium citrate
6. 30% PEG 4000, 0.1 M Tris HCl pH 8.5, 0.2 M magnesium chloride
7. 1.4 M sodium acetate, 0.1 M sodium cacodylate pH 6.5
8. 30% 2-Propanol, 0.1 M sodium cacodylate pH 6.5, 0.2 M sodium citrate
9. 30% PEG 4000, 0.1 M sodium citrate pH 5.6, 0.2 M ammonium acetate
10. 30% PEG 4000, 0.1 M sodium acetate pH 4.6, 0.2 M ammonium acetate
11. 1.0 M ammonium phosphate, 0.1 M sodium citrate pH 5.6
12. 30% 2-Propanol, 0.1 M sodium Hepes pH 7.5, 0.2 M magnesium chloride
13. 30% PEG 400, 0.1 M Tris HCl pH 8.5, 0.2 M sodium citrate
14. 28% PEG 400, 0.1 M sodium Hepes pH 7.5, 0.2 M calcium chloride
15. 30% PEG 8000, 0.1 M sodium cacodylate pH 6.5, 0.2M ammonium sulphate
16. 1.5 M Li sulphate, 0.1 M sodium Hepes pH 7.5
17. 30% PEG 4000, 0.1 M Tris HCl pH 8.5, 0.2 M Li sulphate
18. 20% PEG 8000, 0.1 M sodium cacodylate pH 6.5, 0.2 M magnesium acetate
19. 30% 2-Propanol, 0.1 M Tris HCl pH 8.5, 0.2 M ammonium acetate
20. 25% PEG 4000, 0.1 M sodium acetate pH 4.6, 0.2 M ammonium sulphate
21. 30% MPD, 0.1 M sodium cacodylate pH 6.5, 0.2 M magnesium acetate
22. 30% PEG 4000, 0.1 M Tris HCl pH 8.5, 0.2 M sodium acetate
23. 30% PEG 400, 0.1 M sodium Hepes pH 7.5, 0.2 M magnesium chloride
24. 20% 2-Propanol, 0.1 M sodium acetate pH 4.6, 0.2 M Ca chloride
25. 1.0 M sodium acetate, 0.1 M Imidazole pH 6.5
26. 30% MPD, 0.1 M sodium citrate pH 5.6, 0.2 M ammonium acetate
27. 20% 2-Propanol, 0.1 M sodium Hepes pH 7.5, 0.2 M sodium citrate
28. 30% PEG 8000, 0.1 M sodium cacodylate pH 6.5, 0.2 M sodium acetate
29. 0.8 M K, sodium tartrate, 0.1 M sodium Hepes pH 7.5
30. 30% PEG 8000, 0.2 M ammonium sulphate
31. 30% PEG 4000, 0.2 M ammonium sulphate
32. 2.0 M ammonium sulphate
33. 4.0 M sodium formate
34. 2.0 M sodium formate, 0.1 M sodium acetate pH 4.6
35. 1.6 M sodium phosphate, 0.1 M Hepes pH 7.5
36. 8% PEG 8000, 0.1 M Tris HCl pH 8.5
37. 8% PEG 4000, 0.1 M sodium acetate pH 4.6
38. 1.4 M sodium citrate, 0.1 M sodium Hepes pH 7.5
39. 2% PEG 400, 2.0 M ammonium sulphate, 0.1 M sodium Hepes pH 7.5
40. 20% 2-Propanol, 20% PEG 4000, 0.1 M sodium citrate pH 5.6
41. 10% 2-Propanol, 20% PEG 4000, 0.1 M sodium Hepes pH 7.5
42. 20% PEG 8000, 0.05 M sodium phosphate
43. 30% PEG 1500
44. 0.2 M magnesium formate
45. 18% PEG 8000, 0.1 M sodium cacodylate pH 6.5, 0.2 M zinc acetate
46. 18% PEG 8000, 0.1 M sodium cacodylate pH 6.5, 0.2 M calcium acetate
47. 2.0 M ammonium sulphate, 0.1 M sodium acetate pH 4.6
48. 2.0 M ammonium phosphate, 0.1 M Tris HCl pH 8.5
49. 2% PEG 8000, 1.0 M lithium sulphate
50. 15% PEG 8000, 0.5 M lithium sulphate

Figure 4.5 A list of precipitants in Hampton Research crystal screen 1

1. 10% PEG 6000, 2.0 M sodium chloride
2. 0.5 M NaCl, 0.01 M CTAB, 0.01 M magnesium chloride
3. 25% Ethylene glycol
4. 35% Dioxane
5. 5% Isopropanol, 2.0 M ammonium sulphate
6. 1.0 M Imidazole pH 7.0
7. 10% PEG 1000, 10% PEG 8000
8. 10% Ethanol, 1.5 M sodium chloride
9. 2.0 M sodium chloride, 0.1 M sodium acetate pH 4.6
10. 30% MPD, 0.1 M sodium acetate pH 4.6, 0.2 M NaCl
11. 1.0 M 1,6 Hexanediol, 0.1 M sodium acetate pH 4.6, 0.01 M cobalt chloride
12. 30% PEG 400, 0.1 M sodium acetate pH 4.6, 0.1 M cadmium chloride
13. 30% PEG MME 2000, 0.1 M sodium acetate pH 4.6, 0.2 M ammonium sulphate
14. 2.0 M ammonium sulphate, 0.1M sodium citrate pH 5.6 0.2 M K/Na tartrate
15. 1.0 M lithium sulphate, 0.1M sodium citrate pH 5.6, 0.5 M ammonium sulphate
16. 2% Polyethyleneimine, 0.1 M sodium citrate pH 5.6, 0.5 M sodium chloride
17. 35% tert-butanol, 0.1 M sodium citrate pH 5.6
18. 10% Jeffamine M-600, 0.1 M sodium citrate pH 5.6, 0.01M ferric chloride
19. 2.5 M 1,6 Hexanediol, 0.1 M sodium citrate pH 5.6
20. 1.6 M magnesium sulphate, 0.1 M MES pH 6.5
21. 2.0 M sodium chloride, 0.1 M MES pH 6.5, 0.2 M Na/K phosphate
22. 12% PEG 20,000, 0.1 M MES pH 6.5
23. 10% Dioxane, 0.1 M MES pH 6.5, 1.6 M ammonium sulphate
24. 30% Jeffamine M-600, 0.1 M MES pH 6.5, 0.05 M caesium chloride
25. 1.8 M ammonium sulphate, 0.1 M MES pH 6.5, 0.01 M cobalt chloride
26. 30% PEG MME 5000, 0.1 M MES pH 6.5, 0.2 M ammonium sulphate
27. 25% PEG MME 550, 0.1 M MES pH 6.5, 0.01 M zinc sulphate
28. 1.6 M sodium citrate pH 6.5
29. 30% MPD, 0.1 M Hepes pH 7.5, 0.5 M ammonium sulphate
30. 10% PEG 6000, 0.1 M Hepes pH 7.5, 5% MPD
31. 20% Jeffamine M-600, 0.1 M Hepes pH 7.5
32. 1.6 M ammonium sulphate, 0.1 M Hepes pH 7.5, 0.1 M sodium chloride
33. 2.0 M ammonium formate, 0.1 M Hepes pH 7.5
34. 1.0 M sodium acetate, 0.1 M Hepes pH 7.5, 0.05 M cadmium sulphate
35. 70% MPD, 0.1 M Hepes pH 7.5
36. 4.3 M sodium chloride, 0.1 M Hepes pH 7.5
37. 10% PEG 8000, 0.1 M Hepes pH 7.5, 8% Ethylene glycol
38. 20% PEG 10,000, 0.1 M Hepes pH 7.5
39. 3.4 M 1,6 Hexanediol, 0.1 M Tris pH 8.5, 0.2 M magnesium chloride
40. 25% tert-butanol, 0.1 M Tris pH 8.5, 0.1 M calcium chloride
41. 1.0 M lithium sulphate, 0.1 M Tris pH 8.5, 0.01 M nickel chloride
42. 12% Glycerol, 0.1 M Tris pH 8.5, 1.5 M ammonium sulphate
43. 50% MPD, 0.1 M Tris pH 8.5, 0.2 M ammonium phosphate
44. 20% Ethanol, 0.1 M Tris pH 8.5
45. 20% PEG MME 2000, 0.1 M Tris pH 8.5, 0.01 M nickel chloride
46. 30% PEG MME 550, 0.1 M Bicine pH 9.0, 0.1 M sodium chloride
47. 2.0 M magnesium chloride, 0.1 M Bicine pH 9.0
48. 10% PEG 20,000, 0.1 M Bicine pH 9.0, 2% Dioxane

Figure 4.6 A list of precipitants in Hampton Research crystal screen 2

type, temperature and additives to the precipitant. If crystals were observed in the drop, it was important to verify whether they were formed from protein. The most obvious test was to focus the microscope down into the precipitant reservoir below the drop to search for similar crystals. When they were present in the reservoir the crystals must have been derived from the precipitant mixture and not protein. When crystals were present only in the drop and not in the reservoir the only truly satisfactory method of determining their constitution was by X-ray diffraction, but a number of methods were employed to filter out small molecule crystals. The crush test was employed whereby a small probe was used to crush the suspect crystal. A click or a solid crunch would indicate a salt or low molecular mass crystal while a shattering of the crystal would indicate a protein crystal. Another method was to employ the Hampton Research "izit" dye (cat no. HR4-710) which was supposed to penetrate the solvent channels within the protein crystal and stain it blue, thereby identifying it as protein, but not enter into salt crystals.

4.4 Results

4.4.1 Purification and cleavage of the GST-vWF-A-218 protein

The expression of the 218-residue vWF-A domain (denoted as vWF-A-218) as a fusion protein with glutathione-S-transferase, was induced by the addition of isopropyl-β-D-thiogalactoside to cultures of the *E. coli* transformant during the log phase of growth. The expected molecular mass for the fusion protein was 50.6 kDa; 26 kDa for the fusion partner glutathione-S-transferase and 24.6 kDa for the recombinant vWF-A-218 domain (calculated from the expected amino acid sequence). A major protein product was present at the molecular mass expected for the glutathione-S-transferase-vWF-A-218 fusion protein in cell lysates from induced cells. Cleavage of the fusion protein bound to glutathione-agarose beads with thrombin resulted in the release of the recombinant domain into the supernatant (~25 kDa). The glutathione-agarose beads were suspended in two volumes of MTPBS prior to cleavage with thrombin, as smaller volumes resulted in the presence of protein precipitate in the supernatant. The yield (7.3-10.4 mgs per litre of culture) was estimated from the absorbance at 280 nm using an absorption coefficient of 15.6 (Materials and Methods). Minor contaminating *E. coli* proteins were removed by gel filtration on an FPLC HiLoad

16/60 Superdex 75 column (Pharmacia). On storage of the purified domain, a protein band was occasionally apparent at approximately twice the molecular mass of the recombinant domain when assessed under non-reducing conditions by SDS-PAGE on a 10% (w/v) polyacrylamide gel. This band was not seen under reducing conditions, and was therefore probably a disulphide-linked dimer of the recombinant domain, as factor B has a single free cysteine at position 267 (Parkes *et al.*, 1983). Treatment of the purified vWF-A-218 domain with 2 mM iodoacetamide or 5 mM DTT prior to storage prevented formation of the larger protein. N-terminal sequence analysis of the recombinant product gave a homogeneous sequence NH_2-GS^{229}GEQQKRKIV..... The residues GS preceding the vWF-A domain sequence were derived from the thrombin recognition site LVPRGS in the vector sequence. This sequence also contained the site at which factor D cleaves intact factor B between Arg234 and Lys235 (Lesavre *et al.*, 1979).

4.4.2 Development of a vWF-A-222 expression system

For the large-scale production of the vWF-A-218 fusion protein and purified cleaved vWF-A-218 for structural studies, the glutathione-agarose purification was efficient, and over 90% purity was obtained by thrombin cleavage. The purification protocol was adapted for large-scale preparations. It was noticed that the solubility of the fusion protein was impaired if the *E. coli* cells were stored at -20 °C between harvest and lysis. Accordingly, the maximisation of the yield involved the growth of cells, their lysis and the binding to glutathione-agarose within one day. The use of 5 mM dithiothreitol as a reducing agent to replace iodoacetamide during the lysis stage also inhibited dimer formation. Dithiothreitol was washed out, using degassed buffer after binding of protein to the glutathione-agarose, and before cleavage by thrombin. Freeze-drying was found to be a promoter of aggregation, and was replaced by Amicon stirred cells under N_2 pressure for concentration purposes. As shown by SDS-PAGE, dimers were formed in oxidising conditions, and minor aggregation and precipitation were observed during storage at 4°C. Further purification, to better than 95% by size exclusion chromatography on a HiLoad 16/60 Superdex 75 column successfully separated the monomer and dimer. Immediately prior to this, the vWF-A-218 sample was exposed to 10 mM iodoacetamide for one min at 37°C to modify the thiol at Cys267. Since

the aggregates could not be resolubilised, the net yield of purified vWF-A-218 was reduced to 3 mg/ litre of culture during storage, and its solubility did not exceed 3 mg/ml.

Two modifications were made to the vWF-A-218 expression system in order to improve yields and solubility:

(a) Dimerisation was attributable to the single Cys267 residue within vWF-A-218. Cys267 was not found elsewhere in the vWF-A superfamily, although another one occurs elsewhere, namely in the vWF-A domain of complement C2 (Perkins *et al.*, 1994). While this thiol group in C2 was essential for the normal assembly and decay of the classical pathway C3 convertase, the function of Cys267 in native factor B was uncertain (Parkes *et al.*, 1983). Thus a Cys267Ser mutation should eliminate dimerisation with only a minor perturbation of the vWF-A structure, as both residues are polar and small.

(b) The two C-terminal residues of the vWF-A-218 domain were Met and Ile, both of which are hydrophobic and solvent-exposed. In addition, comparison of the vWF-A-218 sequence with vWF-A sequences in known crystal structures of this superfamily showed that the C-terminal α-helix A7 had been truncated (Figure 6.5 of chapter 6), possibly reducing its structural stability. The addition of the next four residues in factor B (DESQ) to the vWF-A-218, all of which are hydrophilic, would be expected to assist α-helix formation and the stability and solubility of this domain.

Primers were designed to adapt the vWF-A-218 domain to incorporate the Cys267Ser mutation and increase its length by four residues (denoted vWF-A-222) in two stages (Table 4.1). The first and second mutations were verified by experimental observation of the expected changes in the restriction digestion pattern of the plasmid by *Eco* RI and *Spe* I (first mutation) and by *Bfa* I (second mutation).

4.4.3 Characterisation of the first mutant (vWFCysmut) cDNA plasmids

The mutation of Cys267 to a Ser267 also incorporated a *Spe* I restriction site into the plasmid, whereas the original vWF-A-218 plasmid contained no *Spe* I restriction sites. This

restriction site was used as a marker to screen the eight mutant plasmids (vwfCysmut 1 to 8). The *Spe* I restriction digests of plasmids vwfCysmut 1, 3 and 7 contained linearised DNA of 5.5 kbp (Figure 4.7a), whereas the other plasmids were not cut. This suggested that plasmids 1, 3 and 7 contained the desired mutation. These results were confirmed by the double digest using *Spe* I and *Eco* RI restriction enzymes. The *Eco* RI site was present in both the mutated sequence and original vWF-A-218 plasmid. The restriction digests of plasmids 1, 3 and 7 produced a 532 bp fragment confirming the presence of the mutation (Figure 4.7b). The vwfCysmut 7 plasmid preparation produced two bands in the first restriction digest (Figure 4.7a) suggesting both the vWF-222 and original vWF-A-218 clones were present in the bacterial cell culture.

4.4.4 Characterisation of the second mutant (vWFCysext) cDNA plasmids

The restriction site of *Bfa* I was chosen to mark the second mutation. There were multiple occurrences of this site in both the vWF-A-222 and vWF-A-218 DNA sequences. The digest with *Bfa* I of the vWF-A-218 plasmid (or that of the first mutant) should have produced fragments of 54, 253, 335, 476, 729, 1030 and 3012 bp in length. The vWF-A-222 plasmid (with both mutations) should have produced fragments of 54, 253, 335, 729, 1506, 3012 bp in length. The digests of plasmids vWFCysext 2, 4, 6 and 8 contained the 1506 bp fragment (Figure 4.8) and lacked the 476 bp and 1030 bp fragments and so contained the second mutation.

The efficiency of the mutagenesis method was estimated by the number of mutants that occurred in the eight plasmid purifications (minipreps) after the transformations. The presence of any undigested starting material would have limited the efficiency of this method as it might have been successfully transformed and would have survived on the antibiotic medium. The *Dpn* I endonuclease was specific for methylated and hemimethylated DNA. DNA isolated from almost all *E.coli* strains is *dam* methylated and therefore susceptible to *Dpn* I digestion. The first mutation was only 38% (3/8) efficient which showed that there was a significant amount of the original plasmid present. After the second mutation the DNA was incubated with the *Dpn* I for two h instead of one h and this resulted in 50% (4/8) efficiency.

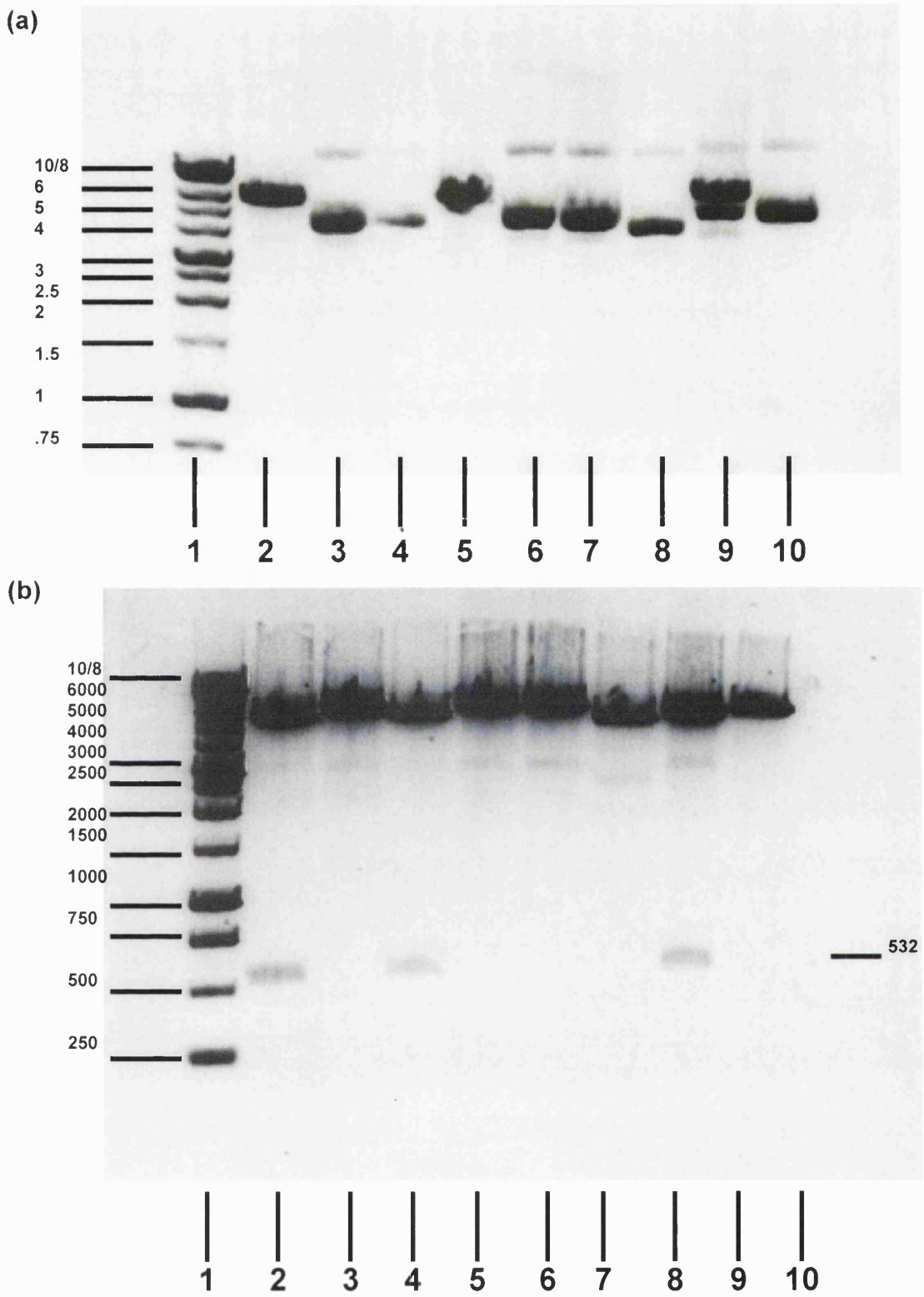


Figure 4.7 1% agarose gels of the restriction digests of the first mutation of the vWF-A plasmid. DNA molecular weight markers are from Promega (G5711). Bands are visualised with ethidium bromide under ultra violet light. (a) Lane 1 is the DNA marker, lane2 to lane 10 are the *Spe*I restriction digests of the plasmids vWFCysmut 1 to 8 respectively with the exception of lane 3 which only contains some of the lane 2 sample. (b) Lane 1 is the DNA marker, lane2 to lane 9 are the double restriction digest of the plasmids vWFCysmut 1 to 8.

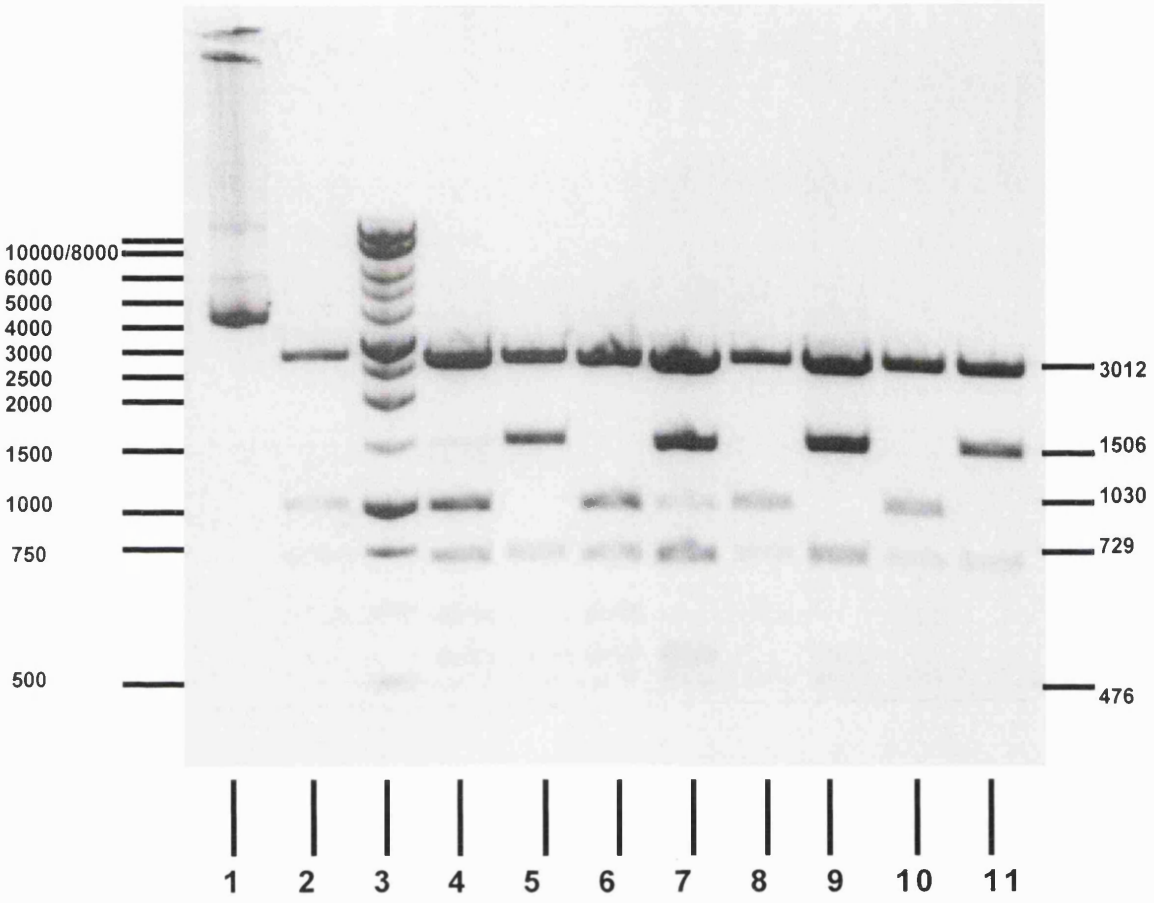


Figure 4.8 1% agarose gel of the *Bfa*I restriction digests of the mutation of the vWFCysmut1 plasmid (vWFCysext). DNA molecular weight markers are from Promega (G5711). The bands are visualised with ethidium bromide under ultra violet light. Lane 1 is the uncut plasmid vWFCysmut7, lane2 is the vWFCysmut7 *Bfa*I restriction digest, lane 3 is the DNA marker, lanes 4 to11 are the *Bfa*I restriction digests of the plasmids vWFCysext 1 to 8 respectively. The bands corresponding to the 3012 bp, 1506 bp, 1030 bp, 729 bp and 476 bp DNA fragments are labelled accordingly.

4.4.5 Western blot analysis of the MC1061 cultures

The bacterial cell lysates of the eight MC1061 colonies (mutants 1-8) contained many proteins, but a major band was observed at around 50 kDa (Figure 4.9a). This band was detected by the factor B polyclonal antiserum (Figure 4.9b) so identifying the GST-vWF-A-222 fusion protein. The lane one proteins, the negative control lane containing the MC1061 cell lysate with no pGEX expression vector, were not identified by the factor B antiserum in the Western blot.

4.4.6 Purification and cleavage of the GST-vWF-A-222 protein

The vWF-A-222 domain preparation was considerably more soluble (~8 mg/ml in PBS) than the vWF-A-218 (~3 mg/ml in PBS) domain which improved the yield to about 10 mg of purified protein per litre of bacterial culture. The preparation was less troublesome than with the original protein because reducing conditions were no longer required and also the solubility was greatly improved. Isolation of this new protein required no further development of the previous purification procedure.

4.4.7 Characterisation of the protein preparations

In SDS-PAGE analysis of the MC1061 vWF-A-222 culture, the bacterial cell lysate contained a major band at around 50 kDa. A Western blot showed that this major band was detected by the factor B polyclonal antiserum. After purification, using a similar approach to that of vWF-A-218 (Materials and Methods), the cleaved vWF-A-218 and vWF-A-222 domains were recognised by Western blot analyses using a polyclonal antiserum to factor B. The vWF-A-218 preparation showed a prominent peak in the MALDI spectrum with a mass of 24,548 (Figure 4.10). The peak at half this value corresponded to a doubly charged species. The vWF-A-222 preparation showed a prominent peak in the electrospray spectrum with a mass of 25,002. These masses were in good agreement with the predicted masses of the vWF-A-218 and vWF-A-222 domains of 24,562 and 25,005 calculated from their amino

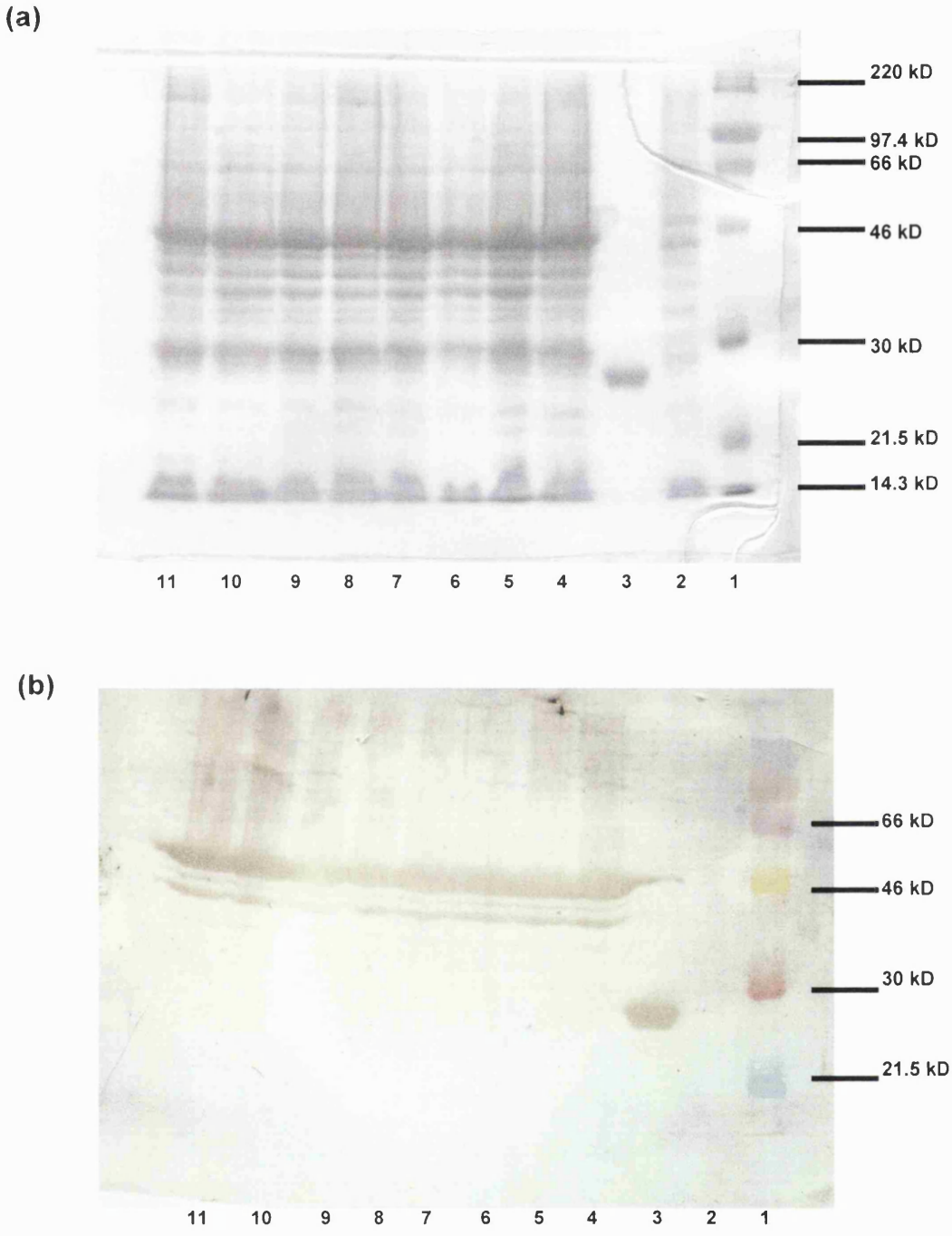


Figure 4.9 (a) 10 % SDS PAGE gel of the bacterial cell lysates from MC1061 cultures of mutants 1 to 8. Samples were boiled in 8 M urea containing 40 mM DTT prior to loading on the gel. Lane 1 contains rainbow protein molecular weight markers (Amersham RPN756), lane 2 contains MC1061 cells with no plasmid, lane 3 contains previously purified vWF-A-218 protein, lane 4 to lane 11 contain the bacterial cell lysate of mutants 1 to 8. (b) Western blot of the above SDS-PAGE gel onto PVDF membrane. The membrane was probed first with a rabbit polyclonal antiserum raised against whole factor B and then with an anti-rabbit IgG alkaline phosphatase conjugate (Sigma, A-3812). A purple colour indicates a positive reaction.

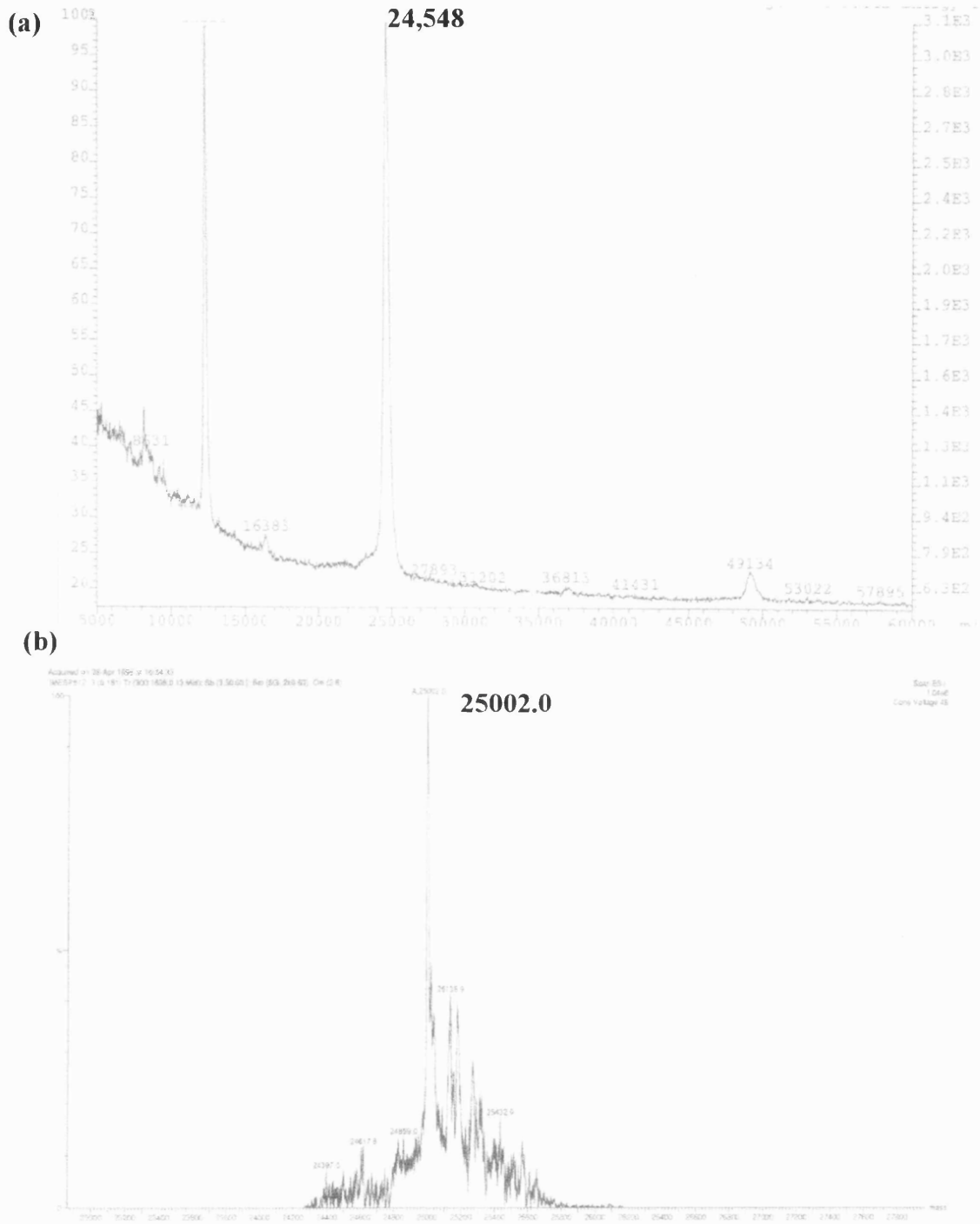


Figure 4.10 Mass spectra of the vWF-A domain preparations. The mass value for each domain is labelled to the right of the signal. (a) Matrix-assisted laser desorption ionisation (MALDI) mass spectrum of the vWF-A-218 sample using a VG-TOF spectrometer. The sample was in 20 mM acetic acid and loaded in a 1: 1 ratio with the matrix sinapinic acid (*trans*-3,5-dimethoxy-4-hydroxycinnamic acid) to aid in desorption. The instrument was calibrated externally using cytochrome c and insulin. (b) Electrospray ionisation mass spectrum of the vWF-A-222 domain using a Finnigan Navigator spectrometer with a Waters Alliance liquid chromatography system for sample delivery. The sample was in deionized H_2O and was loaded in a 1:1 ratio with 50 % acetonitrile. The instrument was calibrated using H_2O clusters. Mass spectrometry was conducted at the University of London facility at the School of Pharmacy.

acid sequences using standard atomic masses with the PAWS program (http://www.proteometrics.com). In large-scale preparations of vWF-A-222, yields were 10 mg/litre of culture, no dimerisation or precipitation was observed during storage at 4°C, and the solubility was increased to about 8 mg/ml in PBS. The vWF-A-218 preparation was shown previously to be biologically active by its ability to inhibit the binding of factor B to C3u (Williams *et al.*, 1999). Circular dichroism spectroscopy, Fourier transform infrared spectroscopy and NMR spectroscopy on the vWF-A-218 and vWF-A-222 domains demonstrated nearly identical results (Chapter 6) and suggested that the three dimensional vWF-A structure had not been perturbed in modifying the original expression vector.

The preparations of Ba, vWF-A, SP, Bb and factor B each eluted as a single peak from the HiLoad 16/60 Superdex 75 gel filtration column and each migrated as single bands on SDS-PAGE under both reducing and non-reducing conditions. The approximate relative molecular masses of these protein preparations were estimated from the SDS-PAGE gels as: Ba, 30,000; vWF-A, 25,000; SP, 35,000; Bb, 60,000; factor B, 90,000. The isoelectric analyses were useful for further investigation into the purity of the protein preparations. Factor B was polymorphic due to the occurrence of a number of allotypes and a variable carbohydrate composition (Campbell and Bentley, 1985), and so the factor B, Ba and Bb preparations produced multiple bands on the isoelectric focusing gels. The vWF-A-218_{DTT} (Figure 4.11b), vWF-A-222 (Figure 4.11c) and the SP (Figure 4.12b) domain preparations each migrated as single bands on the isoelectric focusing gels, whereas the vWF-A-218_{IAM} (Figure 4.11b) preparation migrated as multiple bands. The isoelectric points of the domains were predicted using the GCG software at the HGMP computing resource as: vWF-A-218, pH 6.15; vWF-A-222, pH 5.43 and the SP domain, pH 8.48 (Figures 4.13 and 4.14). These were determined experimentally from the IEF gels as approximately: vWF-A-218, pH 6.2; vWF-A-222, pH 5.6 and the SP domain, pH 9.1.

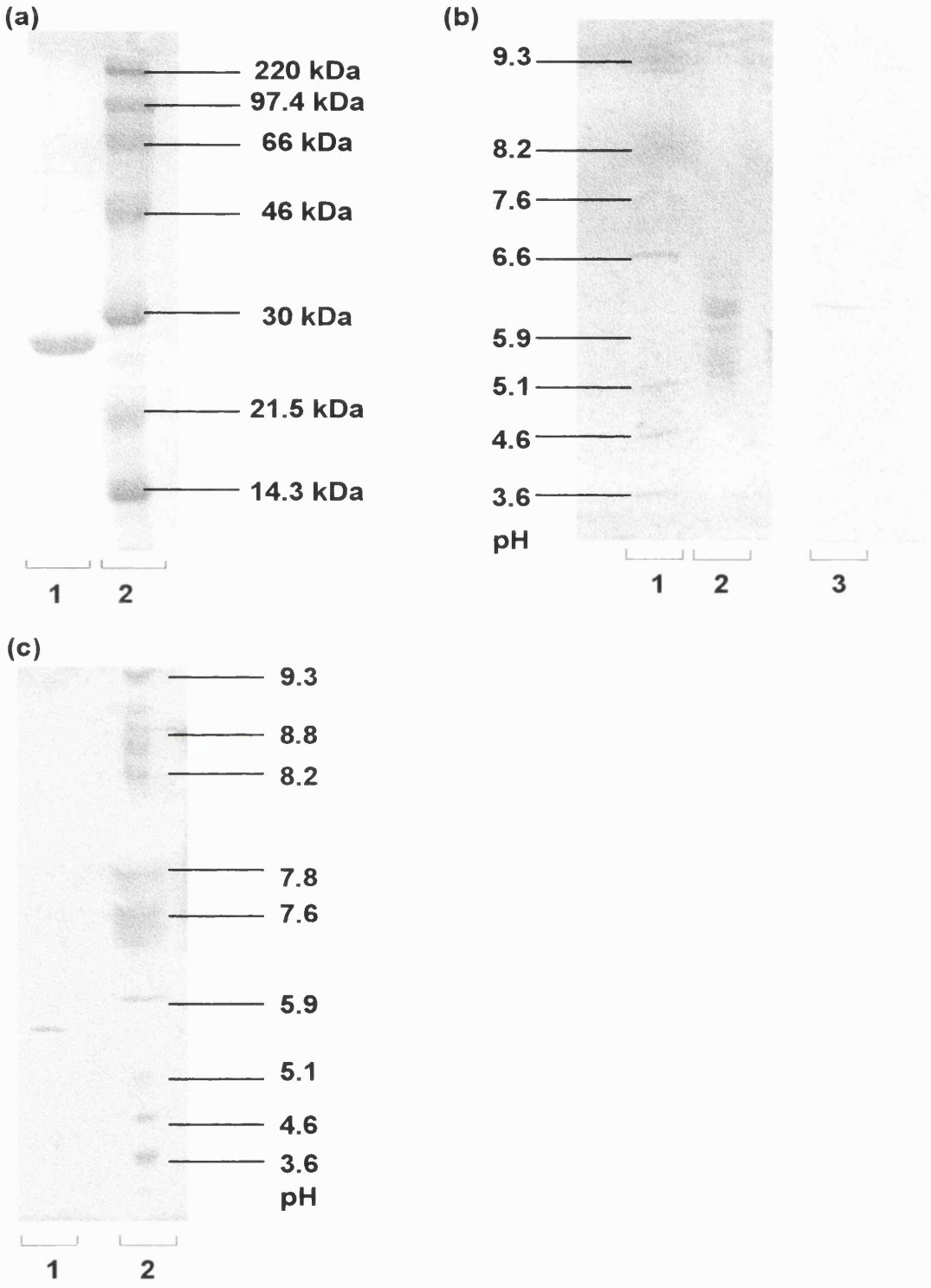


Figure 4.11 (a) 10% SDS-PAGE gel; lane 1 contains purified vWF-A-222 domain and lane 2 contains rainbow protein molecular weight markers (Amersham RPN756). Samples were boiled in 8 M urea containing 40 mM DTT prior to loading on the gel. (b) Isoelectric focussing polyacrylamide gels (PhastGel IEF 3-9, Pharmacia 17-0543-01) of vWF-A-218 domain preparations. Lane 1 contains 5 µl of IEF markers (Sigma IEF-mix, I-3018), lane 2 contains 5 µl of vWF-A-218 prepared with IAM (1 mg/ml) and lane 3 contains 5 µl of vWF-A-218 prepared with DTT (1 mg/ml). (c) Isoelectric focussing polyacrylamide gel (Ampholine PAG plate 3.5-9.5, Pharmacia 18-1124-80); lane 1 contains 20 µl of vWF-A-222 (1 mg/ml), lane 2 contains 10 µl of IEF markers (Sigma IEF-mix, I-3018).

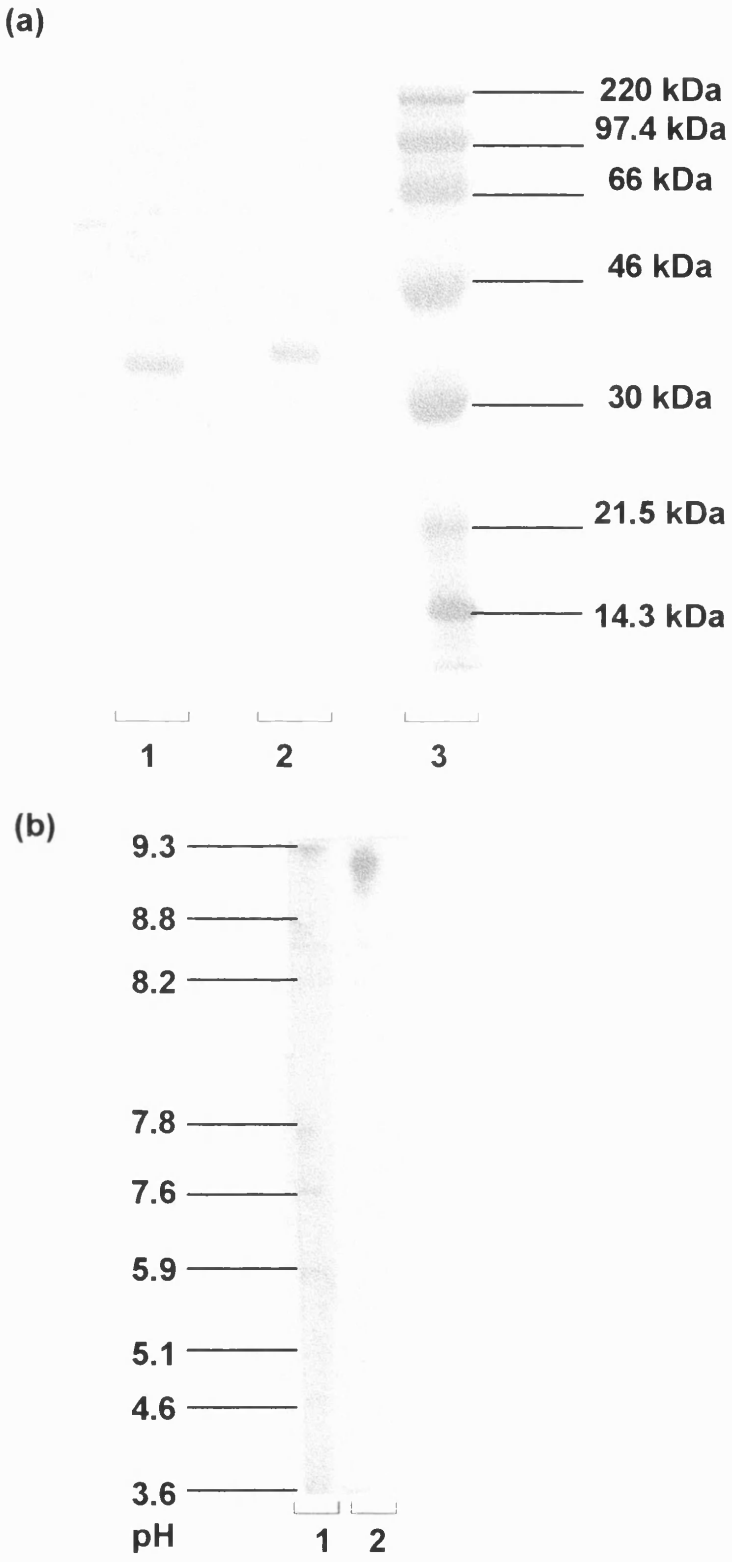


Figure 4.12 (a) 10% SDS-PAGE gel; lane 1 contains purified SP domain (non-reduced), lane 2 contains purified SP domain (reduced) and lane 3 contains rainbow protein molecular weight markers (Amersham RPN756). Samples were boiled in 8 M urea containing 40 mM DTT (reducing) or 80 mM IAM (nonreducing) prior to loading on the gel. (b) Isoelectric focussing polyacyrlamide gel (Ampholine PAGplate 3.5-9.5, Pharmacia 18-1124-80); lane 1 contains 10 µl of IEF markers (Sigma IEF-mix, I-3018) and lane 2 contains 20 µl of the SP domain (1 mg/ml).

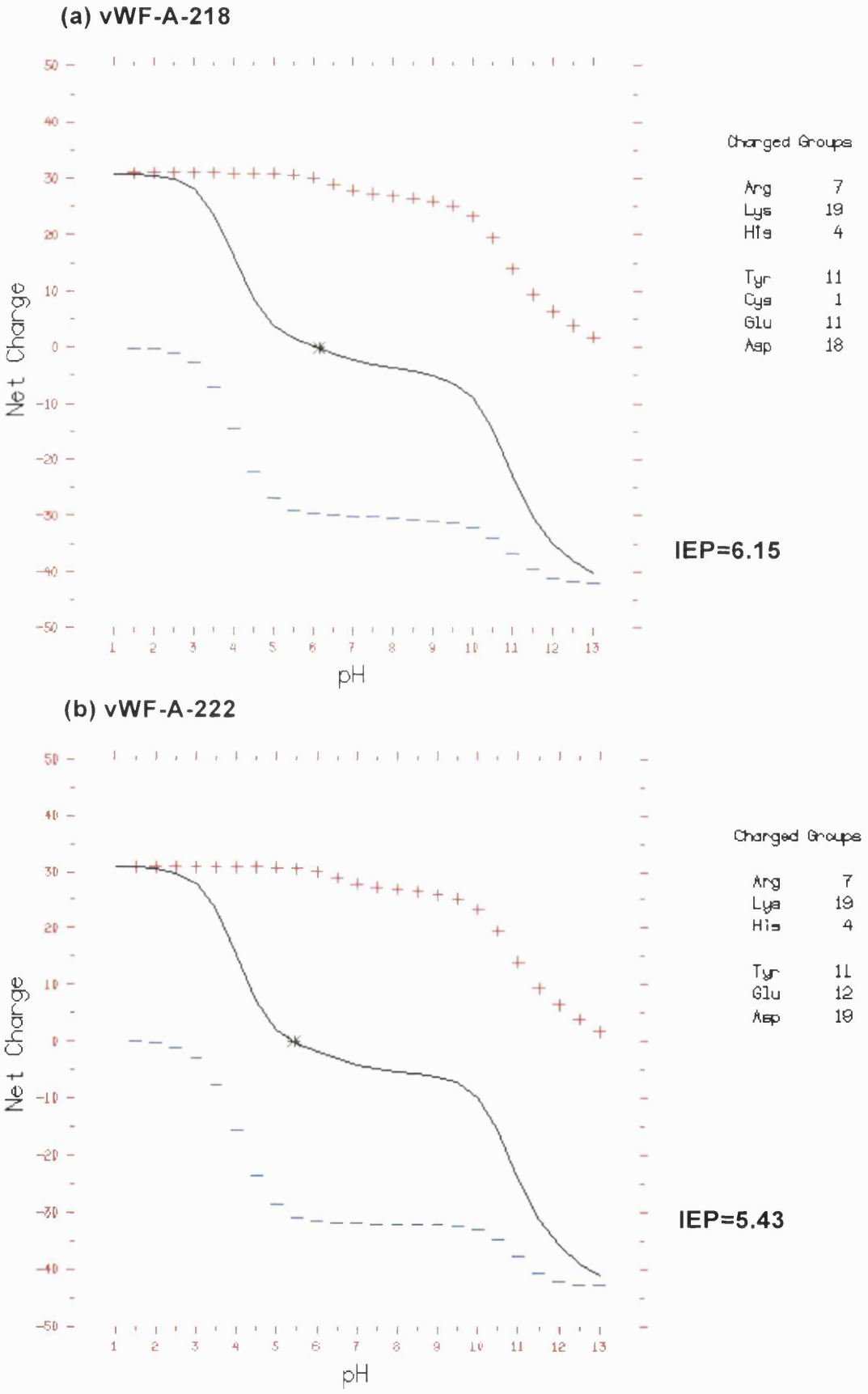


Figure 4.13 Theoretical electrophoretic titration curves as determined by the ISOELECTRIC software on GCG at the HGMP computing resource. (a) The plot for the vWF-218 domain and (b) the plot for the vWF-A-222 domain.

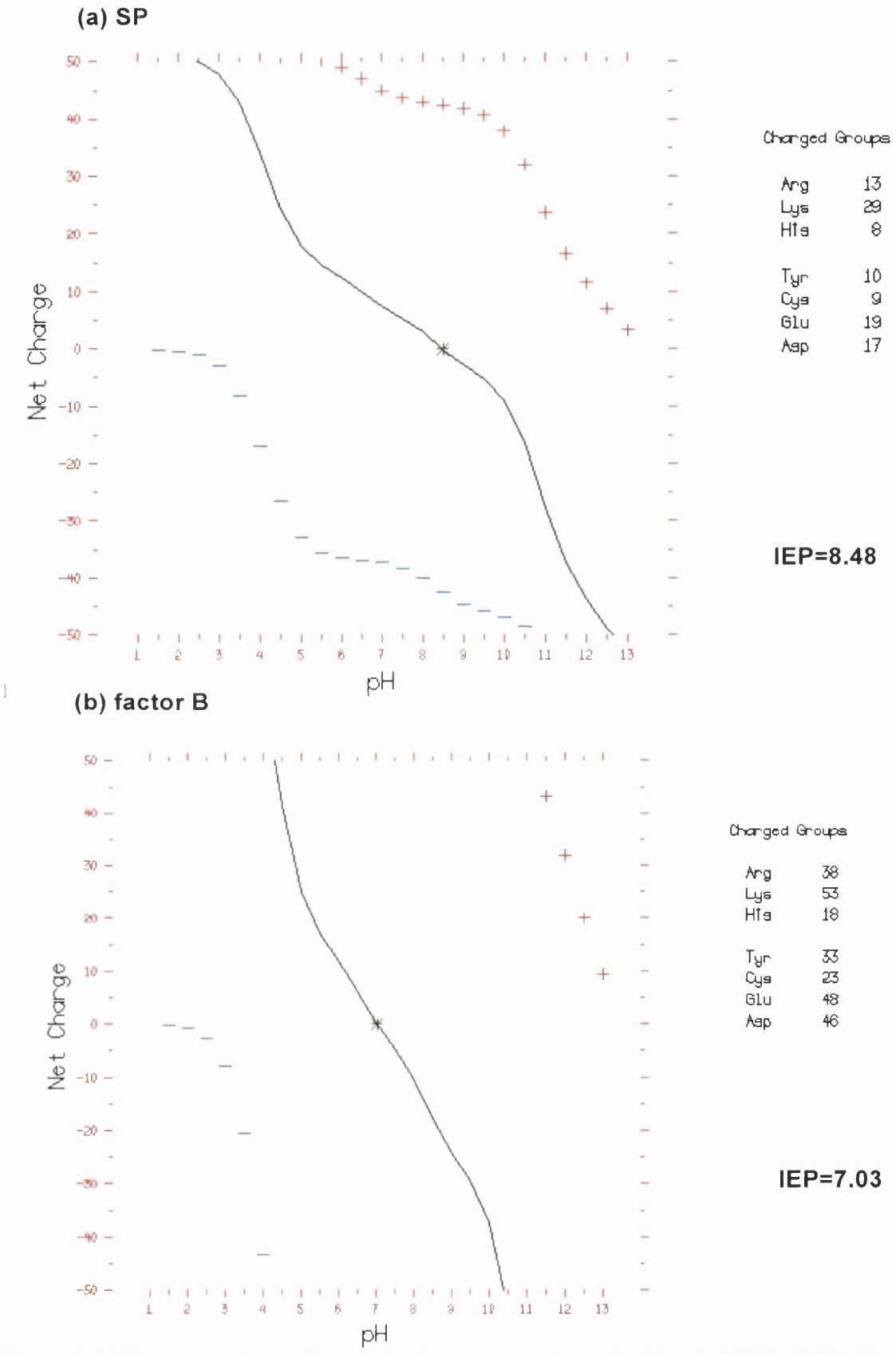


Figure 4.14 Theoretical electrophoretic titration curves as determined by the ISOELECTRIC software on GCG at the HGMP computing resource. (a) The plot for the SP domain and (b) the plot for whole factor B.

The factor B, Ba and Bb fragment preparations were identified by the fact that factor D cleaved the factor B preparation in the presence of C3(H_2O) and Mg^{2+} to produce two smaller fragments of comparable sizes to the Ba (30 kDa) and Bb (60 kDa) fragments (Figure 4.15). The optimum activity of factor D was determined to be at around pH 8.0 and its activity dropped off below pH 6.0 and above pH 11.0. The cleavage of factor B was very inefficient in the negative controls where either factor D or C3 were absent, thus illustrating the tight regulation of the factor D proteolytic activity (Figure 4.15). The SP domain was the most soluble of all the protein preparations and concentrations of 30 mg/ml in PBS were stable. It contained no glycosylation sites and so a homogeneous preparation could be produced from the heterogeneous factor B (Figure 4.12).

The first fifteen residues of the SP domain N-terminal sequence were obtained by automated Edman degradation as DESQSLSL?GMVWEH. The question mark was probably a Cys residue that would have needed to be derivatised in some way in order to have identified it. This sequence was identical to the N-terminal sequence previously determined for the SP domain (Lambris and Müller-Eberhard, 1984). The SP caused no appreciable increase in the absorbance of the 5,5'-dithiobis-2-nitrobenzoic acid stock at 412 nm in either the folded or unfolded states. This suggests that there were no free sulphydryl groups within the domain and that all the Cys residues were involved in disulphide bonds. The SP domain had catalytic activity against C3 as measured by SDS-PAGE (Figure 4.16a). A definite low molecular mass band, of about 8,000, was visible in the lanes that contained the incubated C3 and SP that corresponded to the C3a fragment of C3 (Figure 4.16a). The domain had activity through the pH range 6 to 11, with optimal activity around pH 8 (Figure 4.16b). The SP domain did cleave the synthetic C3 analogue, Boc-Leu-Gly-Arg-AMC, as measured by an increase in the fluorescence of the reaction mixture. It was suspected that the pH may have had some effect on the fluorescence of the AMC, because the starting level and also the rate of increase of fluorescence increased

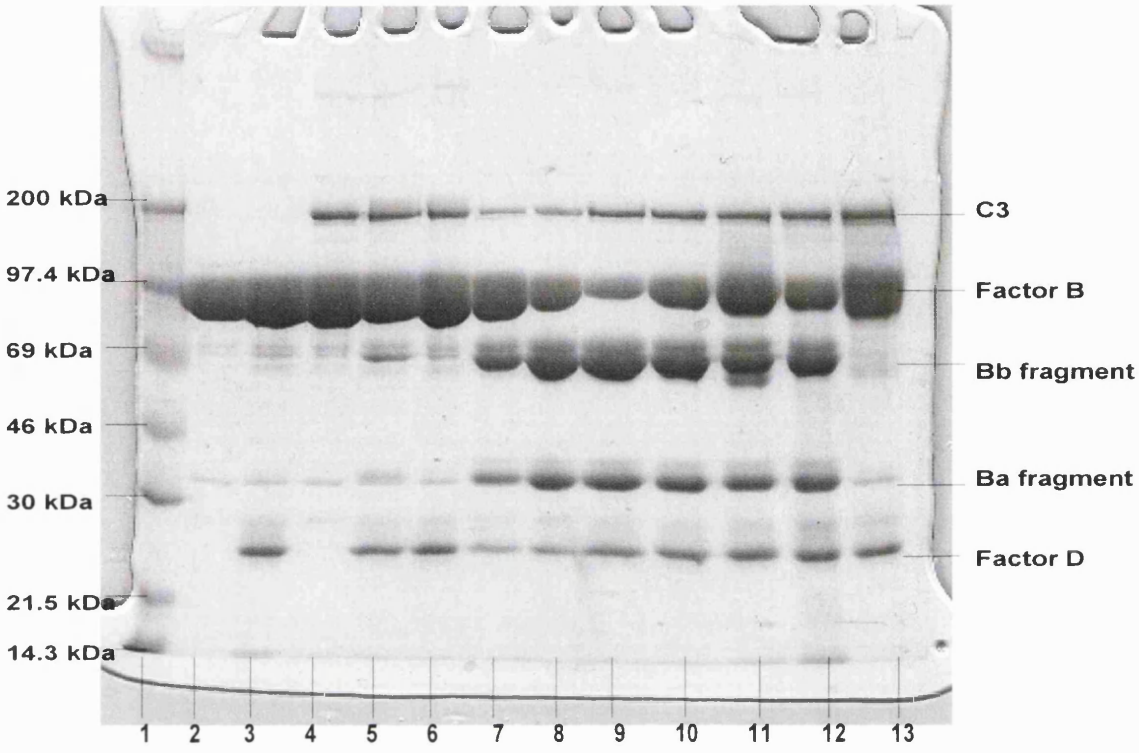


Figure 4.15 8% SDS-PAGE gel of the incubation of factor B, C3 and factor D. Samples were incubated at 37°C for 20 minutes across a pH range from pH 4.0 to 9.0. Lane1, molecular weight markers (Amersham RPN 756); lane 2, factor B (10 µl at 2.5 mg/ml); lane 3, factor B + factor D; lane 4, factor B + C3; lane 5 to lane 13, factor B + factor D factor B + C3 pH 4.0 to pH 12.0 respectively in steps of 1 pH unit.

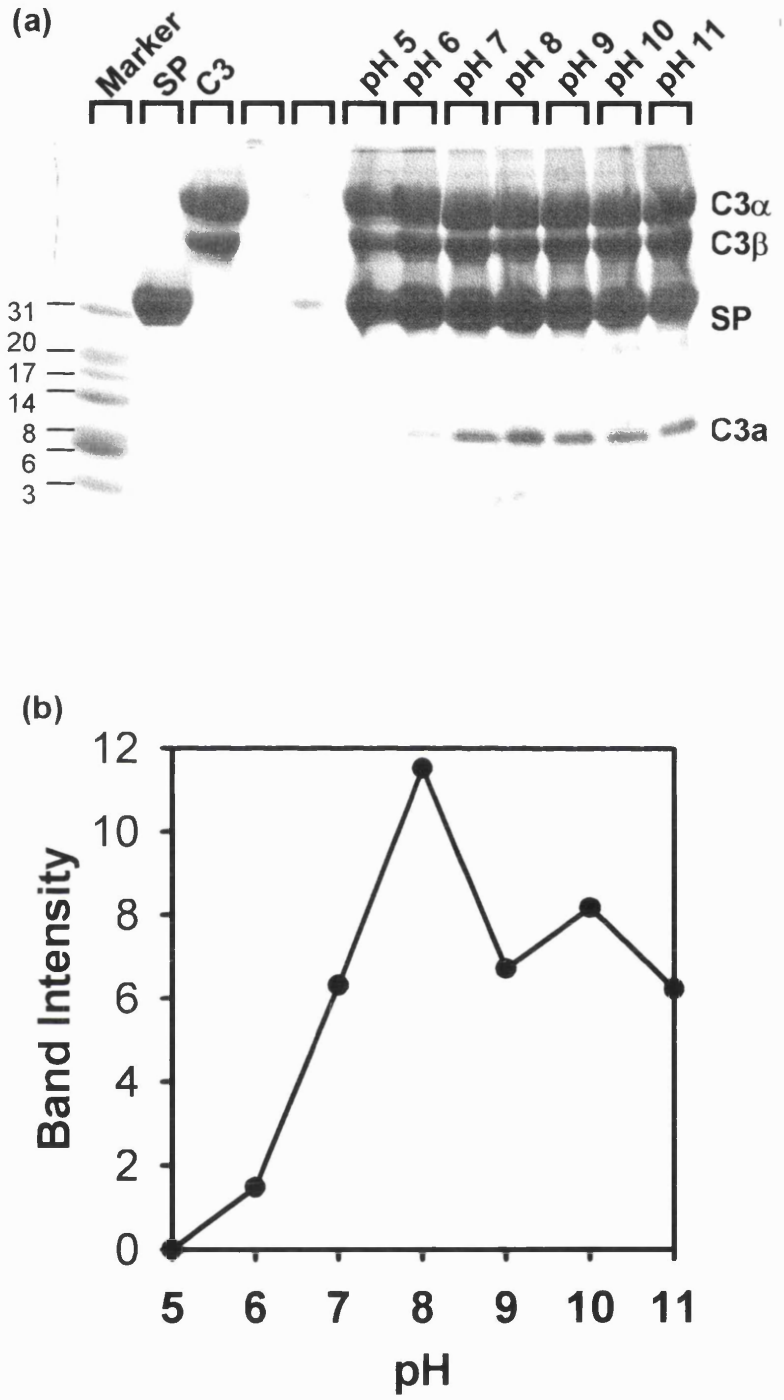


Figure 4.16 Assay of the dependence of the SP domain activity on pH. (a) 16 %Tricine gel run with samples from a 2 hour incubation at 37°C of C3 with the SP domain stained with Coomassie blue. The samples were prepared in reducing conditions where C3 is separated into its α and β chains. Protein molecular weight marker sizes are shown in kD. The SP and C3 lanes are controls. The cleavage product C3a is apparent as a 8,000 M_r band. (b) Intensity of the C3a band as a function of pH where the ordinate represents the relative intensity of the bands.

significantly above pH 7.5. For these reasons the rate of increase of fluorescence did not give an accurate measure of the rate of cleavage of the substrate. The results for the lower pHs were similar to those from the SDS-PAGE assay in that the activity dropped off below pH 6.0.

4.4.8 Crystallisation of the vWF-A-222 domain preparation

Protein crystals of the vWF-A-222 domain were grown in 25% tert-butanol, 0.1 M Tris pH 8.5, 0.1 M $CaCl_2$ (crystal screen II). These crystals were hexagonal plates of 0.1 mm diameter and took three months to grow. They were initially identified by staining blue with the izit dye (Figure 4.17). One crystal was mounted in the X-ray camera, but unfortunately it did not diffract. A crystal was crushed for seeding purposes and it shattered as one would expect a protein crystal would (Ducruix and Giege, 1992). Crystallisation trials on this protein are still in progress.

The presence of a single species, as indicated by gel filtration, SDS-PAGE, mass spectrometry and IEF suggested that the vWF-A-218, vWF-A-222 and the SP domain preparations were structurally homogeneous and so suitable for protein crystallography. A comprehensive approach was taken to try to produce protein crystals of all three preparations. The systematic screening approach did not yield crystals from any of the protein preparations.

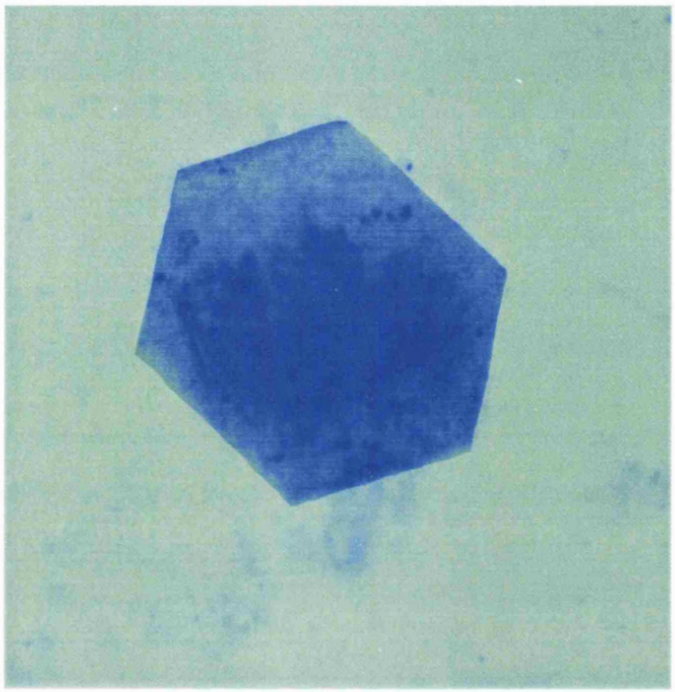


Figure 4.17 Hexagonal crystal of vWF-A-222. The crystal was grown by the method of hanging drop vapour diffusion on a siliconised cover plate. 2 µl of the vWF-A-222 preparation at 5 mg/ml in 25 mM Tris, 0.1 % NaN_3, pH 7.5 was mixed with an equal volume of 25 % tertiary butanol, 0.1 M Tris pH 8.5, 0.1 M $CaCl_2$. The precipitant solution and crystals took three months to grow. The crystal measures approximately 0.1 mm across. The crystal was stained blue using the Hampton Research "izit" crystal dye, which also stained the amorphous precipitate seen in the background.

4.5 Discussion

This chapter describes the development and characterization of two expression systems for isolated vWF-A domains. The previous use of one of these to study the vWF-A secondary structure by Fourier transform infrared spectroscopy (Perkins *et al.*, 1994) showed that the recombinant vWF-A-218 domain was homogeneously folded, and was therefore suitable for further functional and structural studies. A GST-vWF-A-218 fusion protein of the correct molecular mass was expressed (50.3 kDa) and the yield was found to be high (7.3-10.4 mg/litre of culture). This recombinant system corresponded to the exon structure of the gene, and gave monomeric protein in gel filtration, but led to slow aggregation and precipitation during storage. Storage of the vWF-A-218 domain was carried out in the presence of iodoacetamide or DTT as it was found that disulphide linked dimers of the recombinant domain formed. This probably occurred through Cys267 which has been reported to exist as a free cysteine (Parkes *et al.*, 1983). Given the limited solution properties of this first domain, a second one was generated. The inspection of an alignment for 75 vWF-A sequences and analysis of the vWF-A crystal structure (Perkins *et al.*, 1994; Edwards and Perkins 1996; Lee *et al.*, 1995a; Lee *et al.*, 1995b) enabled the rational development of a modified recombinant system, in which a Cys267Ser mutation was performed and the C-terminus of the vWF-A-218 domain was lengthened by four residues to incorporate a C-terminal α-helix. Site directed mutagenesis was efficiently performed and protein characterisation was sufficient to confirm these changes even though the sequence of the plasmid DNA was not directly determined. The C-terminus of the vWF-A-218 domain contained the hydrophobic residues Met443-Ile444. This region of the molecule is likely to be solvent exposed which would put the side chains of these hydrophobic residues into energetically unfavourable positions. The poor solubility of the vWF-A-218 domain was partly attributed to the exposure of these residues. This hypothesis was confirmed by the increased solubility of the vWF-A-222 domain. The sequence and structural data (Perkins *et al.*, 1994; Edwards and Perkins 1996; Lee *et al.*, 1995a; Lee *et al.*, 1995b) would suggest that the mutation incorporating the four residue insertion at the C-terminal of vWF-A-218 would add an extra turn to the C-terminal α-helix thus burying the previously exposed hydrophobic residues. This greatly stabilised the vWF-A-222 domain preparation. After the

Cys267Ser mutation had been performed and the C-terminus of the vWF-A-218 domain was lengthened by four residues, dimerisation did not occur and the protein stability was improved. Mutation of this domain to produce the vWF-A-222 domain proved useful to aid in its crystallisation. It was hoped that the quality of the vWF-A-222 crystals would be improved to allow for an X-ray structure determination.

All the protein preparations were characterised by SDS-PAGE, IEF-PAGE and eluted as monomeric protein in gel filtration. The preparations of native factor B, the Ba and Bb fragments and the SP domain from human plasma were straightforward. The SP domain sequence was verified by N-terminal sequencing analysis. This domain was highly soluble, showed function against C3 and was prepared to high purity. This suggested that it would be a good candidate for crystallisation trials. The parallel approaches of systematic and random screening provided no crystals of this domain, and in order to achieve protein crystals it would be sensible to try protein derived from another source, either animal or recombinant. Indeed preliminary X-ray data from a recombinant factor B SP domain have recently been reported (Narayana *et al.*, 1998).

Study of the function of factor B requires purified protein for characterisation. These protein preparations have been well characterised by the studies described here. These characterisations were necessary before the proteins could be used in the structural and functional studies described in the following chapters of this thesis.

Chapter 5

Experimental:
Hydrodynamic properties and functional activity of the recombinant vWF-A domain preparations.

5.1 Introduction

In this chapter further evidence is presented to show that the recombinant vWF-A domain preparations were folded and monomeric, and were therefore suitable for functional and structural studies. The experimental work carried out for this thesis in this chapter chiefly comprised the analytical ultracentrifuge characterisations using the Beckman Xli instrument and the mass spectrometry analysis of the factor D cleavage of the vWF-A218 and vWF-A-222 domains. The sucrose density centrifugation studies and the functional studies using C3b presented here was work done by Dr S. C. Williams and Dr R. B Sim (Williams *et al.*, 1999). These studies provided essential information about the recombinant vWF-A domain preparations. The functional activity of the two recombinant vWF-A domains from factor B have provided new insights on the role of vWF-A domains in complement activation.

The vWF-A protein fold was first identified by structure prediction analyses (Edwards and Perkins, 1995; 1996). Crystal structures of the I-domain of complement receptor type 3 (CR3) (Lee *et al.*, 1995; Baldwin *et al.*, 1998), leucocyte integrin lymphocyte function associated antigen-1 (LFA-1) (Qu and Leahy, 1995 and 1996), the collagen binding site in integrin α2 β1 (GPIa/IIa) (Emsley *et al.*, 1998), the von Willebrand factor A1 domain (Celikel *et al.*, 1998; Emsley *et al.*, 1998) and the A3 domain (Huizinga *et al.*, 1997; Bienkowska *et al.*, 1998) have since been determined to confirm this prediction. The salient features of this domain are that it adopts a structure related to a flaxodoxin dinucleotide binding domain structure, or Rossmann fold, with seven amphiphatic α-helices surrounding a central hydrophobic, mostly parallel β-sheet. The integrin domains contain an unusual Mg^{2+} coordination site described as the metal ion dependent adhesion site (MIDAS) which is critical for the adhesive function. Binding of the divalent cation Mg^{2+} appears to induce conformational changes within the integrin vWF-A domain (I domain) structure (Lee *et al.*, 1996; Qu and Leahy, 1995). Sequence alignments suggest that the vWF protein A domains contain a MIDAS site, but crystal structures show that the motif is modified and no metal ion is bound (Bienkowska *et al.*, 1997; Emsley *et al.*, 1998).

Electron microscopy studies indicate that only one of the two domains of the Bb

fragment interacts with C3 in the C3bBb structure (Smith *et al.*, 1984). A monoclonal anti-Ba antibody was capable of inhibiting the binding of ^{125}I-labelled factor B to red cell-bound C3b by up to 50% (Ueda *et al.*, 1987). A specific metal-ion independent interaction between Ba and C3b was also identified with the use of a crosslinking agent (Pryzdial and Isenman 1988). Evidence also exists for the existence of a C3b binding site in the SP domain of factor B. After elastase digestion of factor B, direct binding of the isolated ^{125}I-labelled SP domain to C3b bound to zymosan was demonstrated in the presence of Mg^{2+}ions (Lambris and Müller-Eberhard1984). The SP domain was also demonstrated to have Mg^{2+}-ion independent binding affinity for fluid-phase C3b, and a comparison of the Bb fragment and the SP domain provided indirect evidence that the vWF-A domain interacted with C3b (Sanchez-Corral *et al.*, 1990). Mutagenesis of recombinant factor B suggested that surface loops adjacent to the Mg^{2+}-site in the vWF-A domain determined its C3b binding specificity (Tuckwell *et al.*, 1997; Hourcade *et al.*, 1999).

The analytical ultracentrifuge permits observations of the behaviour of macromolecules subjected to a centrifugal field. Such experiments permit determination of solution molecular masses, association constants, and studies of homogeneity, shape, and other molecular parameters. The instrument consists of a centrifuge and a rotor with a windowed sample compartment, so that the distribution of macromolecules along the radial dimension can be determined at any time by means of an appropriate optical system. The analytical ultracentrifuge permits studies of molecules at varying concentrations in many buffers including widely varying salt and pH condition. The conclusions reached are based directly upon first principles, and do not require comparisons to standards, which must be assumed to behave in like manner to the molecules under investigation. Two basic types of experiment can be performed with the analytical ultracentrifuge: (1) a sedimentation velocity experiment and (2) a sedimentation equilibrium experiment. In a sedimentation velocity experiment, the speed with which a molecule moves toward the bottom of the cell is determined. This yields the sedimentation coefficient s which can be related to the molecular mass and to the frictional coefficient or shape of a particle. In a sedimentation equilibrium experiment, a steady state condition is allowed to develop in which a sample's tendency to sediment in the centrifugal

field is counterbalanced by its tendency to diffuse against the concentration gradient so established. This is the preferred method for accurate molecular mass determinations and for studies of homogeneity and molecular associations.

5.2 Methods

5.2.1 Hydrodynamic analyses of vWF-A-218 and vWF-222

In order to estimate the diffusion coefficient of vWF-A-218, its elution relative to those for standards was studied on a Superose 12 gel filtration column. Diffusion coefficients for these standard proteins were plotted against their elution volumes (human ovalbumin, 7.8 $\times 10^{-7}$ cm^2/sec; soya bean trypsin inhibitor, 9.05 $\times 10^{-7}$ cm^2/sec; hen egg lysozyme, 11.2 $\times 10^{-7}$ cm^2/sec (Martin and Ames, 1961)). The diffusion coefficient yields the Stokes' radius a (Ackers, 1964):

$$a = kT / 6\pi \eta D, \qquad \text{Equation 5.1}$$

where k is the Boltzmann constant (1.4×10^{-23} J/°C), T is the temperature (293 °K), η is the viscosity of water at 20 °C (1.002 centipoise), and D is the diffusion coefficient at 20 °C in water.

The sedimentation coefficient of vWF-A-218 was estimated by linear sucrose-density gradients (12 ml) of 5-40% sucrose in 10 mM Tris-HCl, pH 7.4 (Martin and Ames, 1961). vWF-A-218 (50 µl; 300 µg/ml) in PBS was loaded onto the gradient. Centrifugation was carried out in a Beckman SW40 Ti rotor at 37 000 rpm for 12 h at 4 °C. Gradients were fractionated by peristaltic pumping from the base of the gradient, and fractions were analysed by measuring the absorbance at 280 nm and also by SDS-PAGE analysis on 10% (w/v) polyacrylamide gels. The sedimentation coefficient was calculated by comparing the mobility of vWF-A-218 with those of standard proteins run under identical conditions (thyroglobulin, 19.2 S; bovine liver catalase, 11.2 S; bovine serum albumin, 4.22 S; myoglobin, 2.04 S) (Smith, 1970). Knowledge of the Stokes' radius a and the sedimentation coefficient s gives the molecular mass M_r (Siegel and Monty, 1966):

$$M_r = (6\pi\eta N a\, s^o_{20,w})/(1 - \bar{v}\rho), \qquad \text{Equation 5.2}$$

where N is Avogadro's number, $s^o_{20,w}$ is the sedimentation coefficient at 20 °C in water at infinite dilution of the sedimenting particle, $\bar{v}$ is the partial specific volume (calculated to be 0.739 ml/g from Perkins (1986)), and ρ is the density of water at 20 °C. The frictional ratio f/f_0 was calculated from the ratio of Stokes radii a/a_0, where a_0 is the radius of the sphere corresponding in volume to the hydrated volume of vWF-A-218 in which the hydration was 0.3 g H_2O/ g protein.

Analytical ultracentrifugation experiments on vWF-A-218 and vWF-A-222 were performed on a Beckman XLi instrument, in which the sample was monitored using its absorbance at 280 nm, and its refractive index as measured by interferometry. Sedimentation equilibrium data were acquired over 48 h in six-sector cells with column heights of 3 mm at rotor speeds of 6000, 8500, 11000 and 15000 r.p.m. until equilibrium had been reached at each speed. The data were analysed on the basis of a single species within the Beckman software provided as an add-on to Origin Version 4.1 (Microcal Inc.), where the $\bar{v}$ values for vWF-A-218 and vWF-A-222 were calculated to be 0.739 and 0.738 ml/g respectively (Perkins, 1986). Sedimentation velocity data were acquired over 8 h at rotor speeds of 42000 r.p.m. in two-sector cells with column heights of 12 mm, where scans were recorded at 15 min intervals. Boundary sedimentation experiments were conducted where the sample was mixed uniformly throughout the cell at the start of the experiment. During sedimentation, molecules were depleted from the top of the solution column which resulted in the formation of a trailing boundary for the concentration distribution. The sedimentation coefficient was determined by following the rate of motion of the boundary midpoint r_b.

$$\delta \ln(r_b) / \delta t = \omega^2 s \qquad \text{Equation 5.3}$$

where t is the time, ω is the angular velocity of the sample and s is the sedimentation coefficient. Sedimentation coefficients were derived using the transport method, also provided as an add-on within the Origin program.

To predict the sedimentation coefficient of vWF-A from the crystal structure of the homologous vWF-A domain in CR3 (PDB code 1ido (Lee *et al.*, 1995a)), the atomic coordinates of this structure were converted into a sphere model of the same total volume as the vWF-A domain from factor B, using a grid with cubes of side 0.43 nm. If the number of atoms within a cube exceeded a user-defined cut-off, a sphere of the same volume as the cube was placed at the centre of the cube. This cut-off was determined by the requirement that the total volume of spheres was within 1% of the dry volume of 29246.40 nm^3 calculated from the protein sequence (Chothia, 1975; Perkins 1986). A hydration of 0.3 g H_2O per g of protein and an electrostricted volume of 0.0245 nm^3 per bound water molecule (Perkins, 1986) was used to estimate the hydrated volume of the vWF-A domain. The estimated hydration volume was used to determine the parameters for the HYDRO program. The sphere model was hydrated using the HYPRO program to give a total of 535-545 spheres, from which the frictional coefficient was calculated using the GENDIA method (Ashton *et al.*, 1997 and 1998).

5.2.2 Assay of functional (C3b-binding) activity of vWF-A-218

An assay system was developed to study factor B binding to C3b, using C3b immobilised on thiol-Sepharose via the free SH group in C3b (Williams and Sim, 1994). C3 was purified and converted to C3b (Dodds, 1993). Factor B was purified by dye-ligand affinity chromatography from fresh-frozen human plasma (Williams and Sim, 1993). Aliquots of purified factor B (100 μg in 0.2-1 ml of PBS) were radioiodinated, using 0.5 mCi of Na ^{125}I (Amersham International, Bucks, U.K.) by lactoperoxidase-catalysed iodination using the method of Marchalonis (Marchalonis, 1969) to a specific activity of 1×10^6 cpm/μg.

C3b-thiol Sepharose was produced as follows. Briefly, the resin was reduced by adding dithiothreitol (final concentration, 50 mM) to a 1:1 (v/v) slurry (i.e. 1 volume of packed resin: 1 volume of buffer) in PBS. The mixture was then incubated for 30 min at 37 °C. The resin was then washed with at least five volumes of PBS prior to incubation with C3b. Purified C3b in 25 mM Tris-HCl, 140 mM NaCl, pH 8.2 was incubated with pre-swollen thiol-Sepharose (0.5 mg C3b/ml of resin) for 2 h at room temperature using a rotary stirrer. The

percentage of C3b bound was estimated by measuring the OD_{280} of the original solution and three subsequent washes with PBS. Approximately 20-25% of the C3b bound to the resin, in which the final content of C3b was routinely 100-125 μg/ml of resin. Remaining free SH groups on the thiol-Sepharose resin were blocked by incubation with 5 mg of iodoacetamide per ml of resin for 1 h at room temperature. Non-specific binding sites on the C3b-thiol Sepharose were blocked by incubation of the resin with 1 mg/ml bovine serum albumin in PBS for 1 h at room temperature.

The effect of vWF-A-218 on the binding of radioiodinated factor B to C3b-thiol Sepharose was assessed. In each assay, a mixture consisting of 50 μl of C3b-thiol Sepharose suspended in 50 μl of binding buffer [10 mM Pipes (Piperazine-N,N'-bis-2-ethanesulphonic acid) (sodium salt), 30 mM NaCl, 1 mg/ml bovine serum albumin, pH 7.0 containing 0.2 mM $MgCl_2$ was pre-incubated with serial two-fold dilutions of vWF-A-218 (200 μl; maximum concentration 125 μg/ml) for 2 h at 4 °C. ^{125}I-labelled factor B (4×10^5 cpm) was then added and radioactivity bound to the C3b-thiol Sepharose resin was measured after 1 h incubation at room temperature and five washes each with 0.85 ml of binding buffer.

5.3 Results and Discussion

5.3.1 Hydrodynamic properties of the recombinant vWF-A-218 and vWF-A-222 domains

Given the occurrence of aggregation and precipitation during storage, it was necessary to demonstrate for the functional analyses below that the vWF-A domain was monomeric in solution. Based on the elution positions of three standard proteins in gel filtration (Materials and Methods), the diffusion coefficient for vWF-A-218 was found to be between 8.4×10^{-7} and 9.9×10^{-7} cm^2/sec. From this, the Stokes' radius a of vWF-A-218 was found to be 2.4 ± 0.2 nm. Using sucrose density gradient centrifugation (Materials and Methods) (Martin and Ames, 1961), the sedimentation coefficient was found to be 2.4 ± 0.2 S based on comparison with four standard proteins. These data gave a vWF-A-218 molecular mass between 22 and 30 kD, where the range reflects the errors in the determinations of each of a and $s^{o}_{20,w}$. The comparison with the expected molecular mass of 25 kD calculated from the sequence showed that vWF-A-218 was monomeric. The frictional ratio f/f_0 was calculated as 1.1 (Methods); since this was close to 1.0 (the value for a sphere), it was inferred that vWF-A-218 possessed a compact structure.

To confirm quantitatively that the vWF-A-218 and vWF-A-222 domains were monomeric over the concentration range used for functional and structural studies, both were subjected to sedimentation equilibrium analysis (Figure 5.1). Curve fits on the basis of a single monomeric species in a concentration range between 0.3-0.9 mg/ml gave molecular masses of 24000 ± 2000 for vWF-A-218 and 27000 ± 2000 for vWF-A-222, both of which agree with the expected values of 24,500 and 25,000 respectively. Sedimentation velocity experiments gave an $s^{o}_{20,w}$ values of 2.5 ± 0.3 S for the vWF-A-218 and vWF-A-222 domains (Figures 5.2 and 5.3). Calculations of the sedimentation coefficient from the homologous vWF-A crystal structure from CR3 gave values between 2.5-2.6 S. This comparison with the experimental values showed that the recombinant domains possessed monomeric structures that were similar to the CR3 crystal structure.

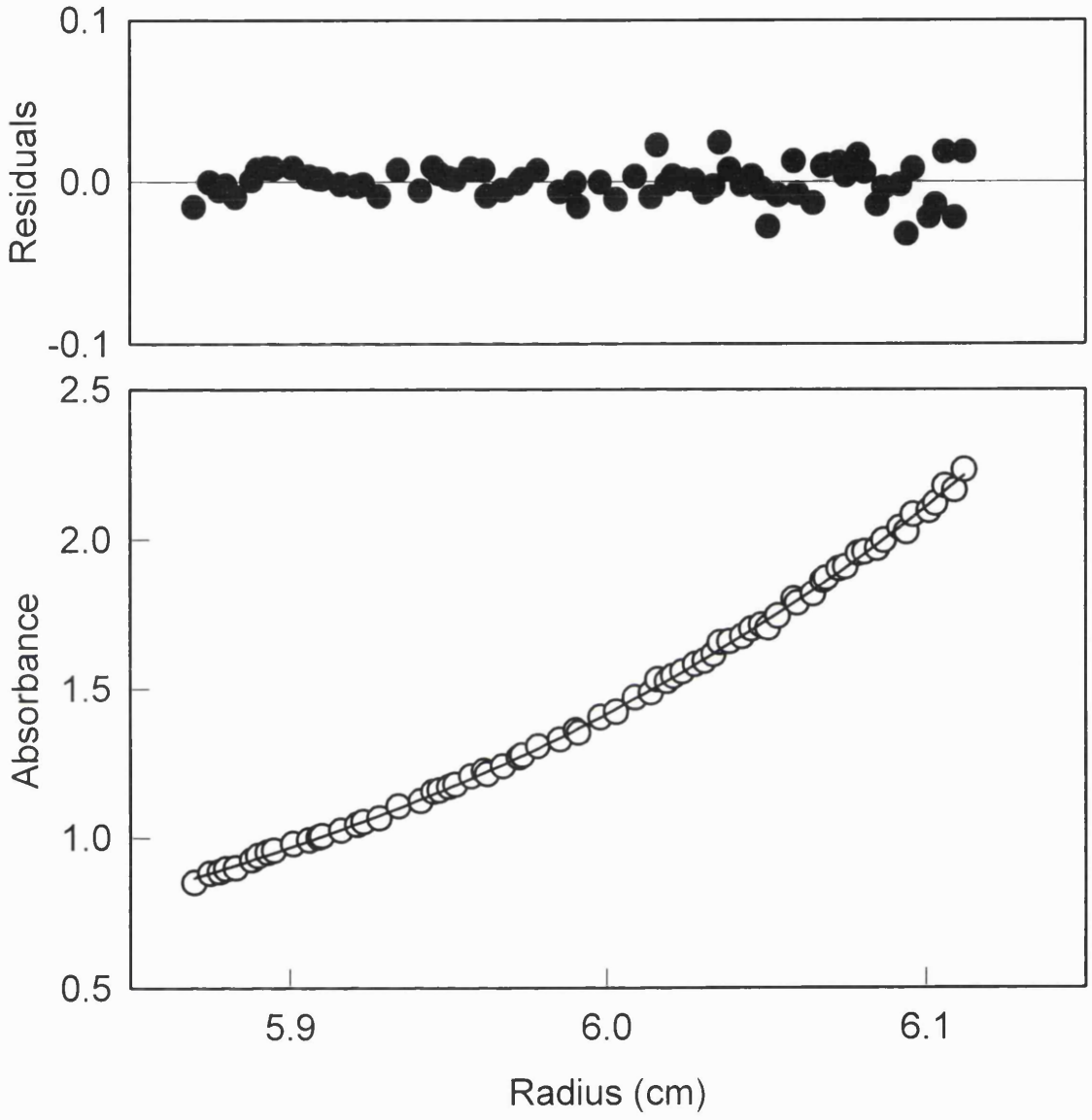


Figure 5.1 Sedimentation equilibrium data for the vWF-A-222 domain. The absorbance values at 280 nm are shown as a function of radial distribution at equilibrium for a loading concentration of 0.8 mg/ml at 20°C and a rotor speed of 15,000 r.p.m. The data are fitted to a monomer model which is shown as a line through the experimental data points (○). The corresponding distribution of the residuals in the upper panel (●) is small and random, indicating that a good fit solution had been obtained.

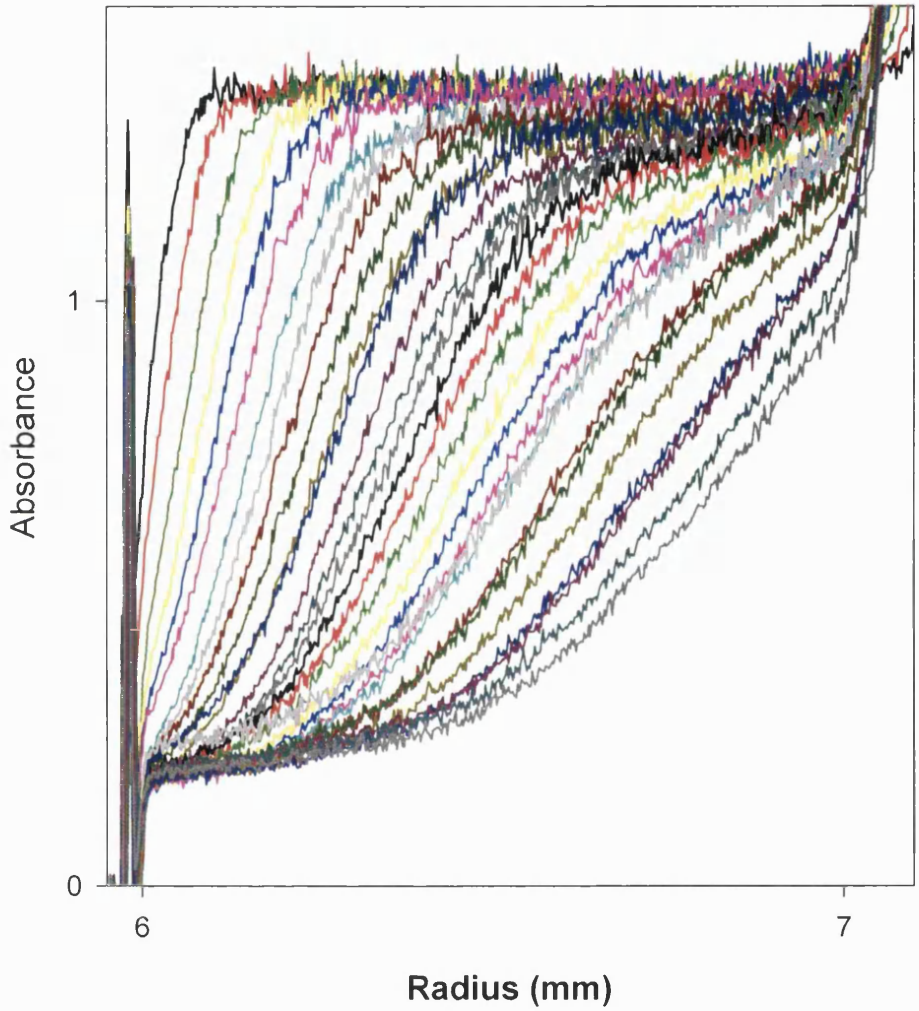


Figure 5.2 Sedimentation velocity data for the vWF-A-218 domain acquired during a boundary sedimentation experiment. The loading concentration of the sample was 0.8 mg/ml at 20°C and the rotor speed was 42,000 r.p.m. Sedimentation velocity data were acquired over 8 h in two-sector cells with column heights of 12 mm, where scans were recorded at 15 min intervals. The scans are represented by the coloured plots starting on the left of the graph. Each successive plot is offset to the right as the sample sediments. The sedimentation coefficient was determined by following the rate of motion of the boundary midpoint.

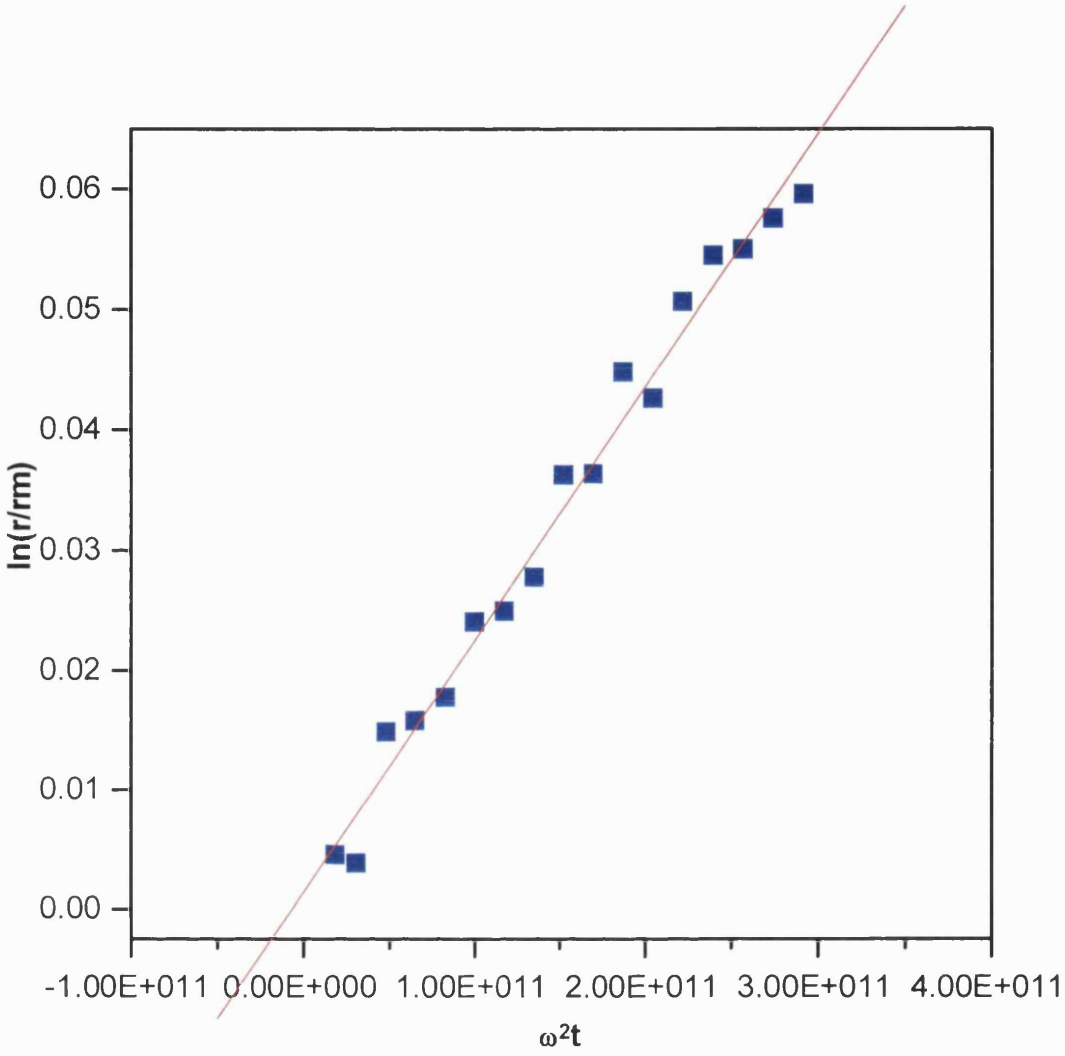


Figure 5.3 Plot of data obtained from a boundary sedimentation experiment for the vWF-A-218 domain. Sedimentation coefficients were derived using the transport method provided as an add-on within the Origin program. The data are fitted to a monomer model which is shown as a line through the experimental data points. The sample was applied with a loading concentration of 0.8 mg/ml at 20°C and a rotor speed of 42,000 r.p.m.

5.3.2 Functional binding of the vWF-A-218 domain to C3b

The availability of a structurally-characterised recombinant vWF-A-218 domain provided a unique opportunity to study its functional activity, as work on the isolated vWF-A domain had not previously been possible. Direct binding of ^{125}I-labelled vWF-A-218 to C3b-thiol Sepharose was first investigated. Approximately 100 μg of purified vWF-A domain was radiolabelled to a specific activity of 1.2×10^6 cpm/μg using the lactoperoxidase method and a test assay was set up to investigate whether direct binding of the vWF-A domain to C3b-thiol Sepharose could be demonstrated. No direct binding of ^{125}I-labelled vWF-A was noted. Binding to C3b-thiol Sepharose was no higher than to a control preparation of thiol-Sepharose which contained no C3b.

The potential interaction of the vWF-A-218 domain with C3b was studied by a more sensitive indirect binding assay based on the assessment of its ability to compete with ^{125}I-labelled factor B for binding to C3b attached to thiol-Sepharose. Using this, direct binding of ^{125}I-labelled factor B to C3b-thiol-Sepharose was observed in the presence of Mg^{2+} ions. Competition binding studies were carried out. The assay was validated using unlabelled factor B as a positive control and an irrelevant protein (ovalbumin) as a negative control. The maximum inhibition obtained with unlabelled factor B was close to 100% as expected. The recombinant vWF-A-218 domain was able to inhibit the binding of ^{125}I-labelled factor B to C3b-thiol-Sepharose to a maximum of 60%. This was achieved with 6.25 μg of vWF-A domain that was approximately an 8-fold molar excess over C3b. The same level of inhibition was achieved with approximately 4 μg of factor B, a 1.7-fold molar excess over C3b.

5.3.3 Effect of factor D on vWF-A-218 and vWF-A-222

To test whether factor D could cleave the vWF-A-218 domain in the presence of C3b in the fluid phase, conditions were used in which factor D cleaves intact factor B. A mixture of vWF-A-218, C3b and factor D (in 10 mM Pipes, 30 mM NaCl, 0.2 mM $MgCl_2$, pH 7.0) was placed in a 1.5 ml tube in a 50:25:1 ratio by mass in a final volume of 200 μl [vWF-A (100 μg):C3b (50 μg): factor D (2 μg)] and incubated for 4 h at 37 °C. Cleavage of vWF-A-218 by factor D would result in the removal of an 8 amino acid peptide

(GSGEQQKR; Figure 5.4) so that the cleaved domain would have the new N-terminus KIVLDPSG. N-terminal sequence analysis (Materials and Methods) was carried out on vWF-A-218 which had been incubated with C3b and factor D. Comparative sequencing was also carried out on vWF-A-218 which had been incubated with C3b only. The results obtained indicated that 10% of the vWF-A domain had been cleaved in the sample incubated with both C3b and factor D, while no cleavage was detected in the control in which Factor D was omitted. These results suggest that factor D is capable of cleaving the vWF-A domain in the presence of C3b.

To test whether the vWF-A domain could be cleaved by factor D in the absence of C3b, vWF-A-218 and vWF-A-222 (50 µl at 1 mg/ml) were incubated with recombinant factor D (1 µl factor D at 1 mg/ml) in PBS at 37°C and pH 7.5 for 24 h. The products were dialysed into water, then mixed with matrix solution, and applied to a mass analyser (Materials and Methods). In both cases, the incubation resulted in the slow production of a cleaved vWF-A domain with a mass reduction of 796 ± 73 (mean ± S. D.; three measurements) (Figure 5.5). This agreed well with the expected mass reduction of 871 based on the removal of the N-terminal 8 residue peptide in the recombinant domains (Figure 5.4). In both cases, peaks corresponding to the uncleaved domain were also visible in Figure 5.5. Even if it was assumed that the efficiency of desorption was equal for cleaved and uncleaved vWF-A-218 and vWF-A-222, this result suggested that cleavage had occurred, and that factor D was active against the free vWF-A domain in the absence of C3b.

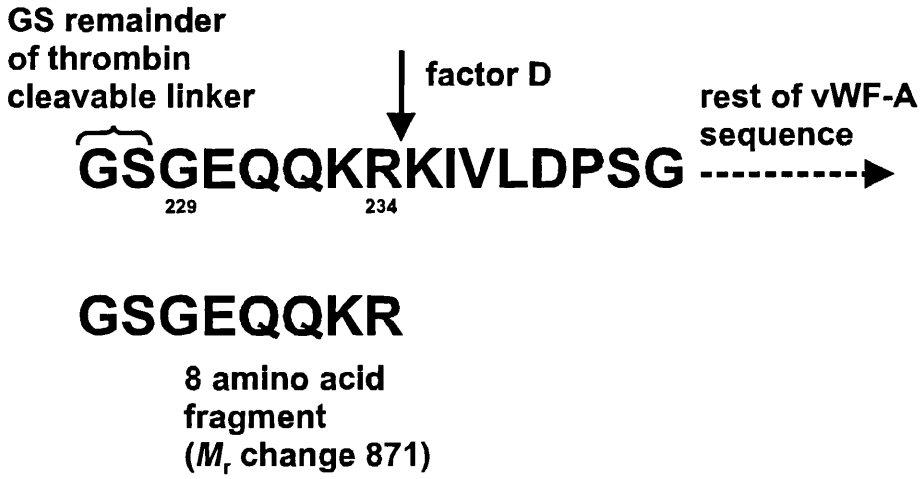


Figure 5.4 Representation of the factor D cleavage site in the recombinant vWF-A domain sequence.

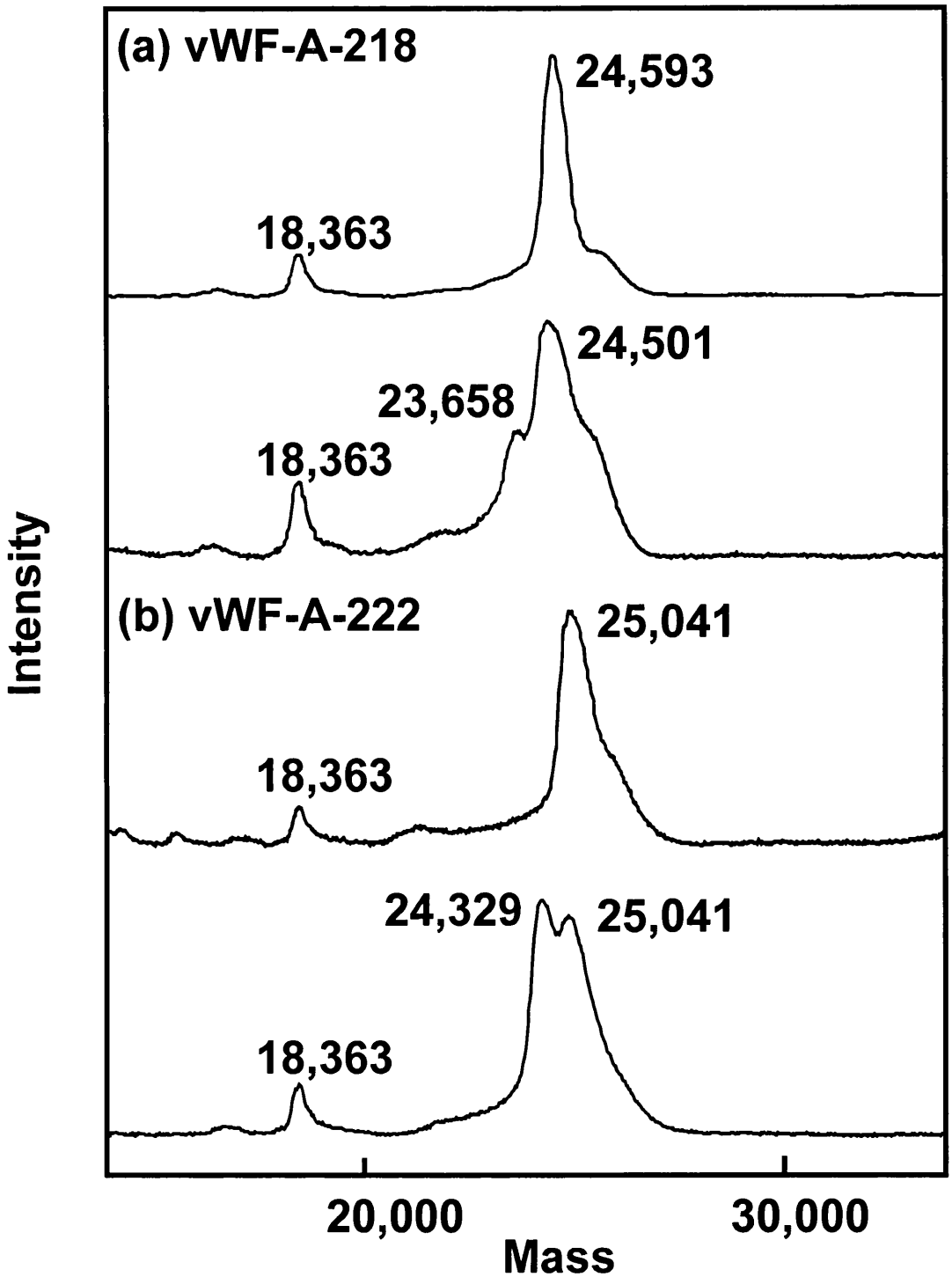


Figure 5.5 The effect of factor D on vWF-A-218 and vWF-A-222 by mass analysis. Mass spectra from (a) the vWF-A-218 and (b) the vWF-A-222 domains are shown. Mass values M_r are reported for each peak, based on internal calibrations against bovine β-lactoglobulin A (M_r 18363). In each of (a) and (b), the upper trace corresponds to the purified domain, and the lower trace to that of the domain after treatment with factor D for 24 h. In both lower traces, a second peak with a reduced molecular mass is visible and corresponds to the cleaved vWF-A domain when each is compared with its corresponding upper trace.

5.4 Conclusions

The vWF-A-218 recombinant system corresponded to the exon structure of the gene and the vWF-A-222 is a four residue C-terminal extension of this with Cys267 replaced with a Ser residue. Both protein preparations were monomeric after gel filtration. The observation of a compact solution structure is in agreement with the crystal structures of homologous vWF-A domains (Lee *et al.*,1995a; Lee *et al.*,1995b). It is also consistent with the observation of three and two-lobed structures for intact factor B and the Bb fragment by electron microscopy, since it can be assumed that the vWF-A domain would form one of these lobes. The production of eight contiguous vWF-A domains from the α-3 chain of collagen VI has been achieved using an eukaryotic expression system (Specks *et al.*,1992). There, the structure of the expressed fragment was visualized to contain eight small globules using electron microscopy, which is consistent with the data on vWF-A domains and factor B.

The production of the vWF-A-218 domain enabled its functional binding to C3b to be studied. The studies carried out here agree with earlier suggestions for the role of the vWF-A domain in factor B (Sanchez-Corral *et al.*,1990). Since the present work shows that the vWF-A domain can specifically inhibit the binding of factor B to C3b, it contains a binding site for C3b. Although the domain was unable to bind to C3b-thiol Sepharose directly, it was able to compete with ^{125}I-labelled factor B for binding, the inhibition being maximally ~60% in the presence of 0.2 mM Mg^{2+}. The negative results from the direct binding studies are best explained by postulating that the binding of the vWF-A domain to C3b is weak and that the complex dissociates during the washing steps. This result is analogous to other work with factor B, in which the Ba fragment was shown to bind to C3b, but the direct binding of Ba to C3b was also not observed (Pryzdial and Isenman 1988). These observations are characteristic of a complex formed by multiple weak binding interactions, not one formed by a single-site high affinity binding interaction.

The availability of the recombinant vWF-A domain also provided insight on the cleavage of factor B by factor D. Factor D is known to circulate in its active form as there is

an apparent absence of a zymogen form in blood. The enzyme is also highly specific, as it cleaves its only known substrate, factor B, only when factor B is bound to C3b or C3(H_2O) (Lesavre *et al.*,1979). Several hexapeptides corresponding to the factor B sequence surrounding the bond that is cleaved by factor D were studied (Lesavre *et al.*, 1982). The peptides were assessed for their ability to inhibit factor D enzymatic activity and for their susceptibility to cleavage by several serine proteases including factor D. The peptides were all able to inhibit factor B cleavage by factor D, but were not substrates for factor D. Active site mapping of factor D with peptide thioesters revealed some interesting features (Volanakis, 1989; Kam *et al.*,1987). Factor D was able to express esterolytic activity against some Arg thioesters, but its catalytic efficiency was found to be three or four orders of magnitude below that of trypsin and C1s. It was suggested on the basis of these results that the active site of factor D, as it exists in serum, has a zymogen-like conformation, which has an obstructed binding site. This view is supported by recent crystal structures for factor D which revealed a self-inhibitory loop in its structure (Jing *et al.*, 1998). It has been proposed that initial recognition of the substrate C3bB may involve the self-inhibitory loop of factor D. The active conformation of the catalytic triad geometry is then induced by its substrate, the C3bB complex, by lowering the self-inhibitory loop thus opening the S1 pocket (Jing *et al.*, 1998). In light of its pivotal role in the alternative pathway of complement activation, much attention is being focused on structure-based drug design of novel small molecule inhibitors to factor D (Cole *et al.*, 1997).

Since the recombinant vWF-A domains contain factor D cleavage sites, the effect of factor D on vWF-A in the presence and absence of C3b was studied by two independent methods. It is significant that factor D was observed to cleave vWF-A slowly on its own in both cases, since this indicates that the free vWF-A domain is an adequate substrate for factor D. In contrast, if intact factor B was incubated with factor D in the presence of C3b, more than 95% of the factor B was efficiently converted to Ba and Bb. One possible explanation for the different extents of cleavage of free vWF-A domain and factor B is that the affinity of vWF-A for C3b is much less than that of intact factor B by reason of the existence of fewer contact points. This would mean that the concentration of the C3bvWF-A complex is much

lower than that of the C3bB complex, so the cleavage of the C3bvWF-A complex would be slower. Another possible explanation is based on the existence of two alternative conformations of the vWF-A domains. General evidence for these in the vWF-A superfamily is found from the distinct locations of mutation sites on the same vWF-A domain that correspond to type 2B and 2M von Willebrand's disease (Jenkins *et al.*, 1998), and from several crystallographic studies of different structures in the presence and absence of metal in the vWF-A active site (Lee *et al.*,1995b; Baldwin *et al.*, 1998). Thus, if the vWF-A domain of factor B has two conformations that depend on whether it is bound to C3b or to the remainder of the factor B structure, and only the C3b-bound vWF-A conformation is efficiently cleaved by factor D, then the less efficient cleavage of the isolated vWF-A domain may correspond to the existence of both conformations of the isolated vWF-A domain when this is free in solution, only one of which is cleavable by factor D. Further kinetic and structural studies are required to resolve these questions.

The production of the isolated vWF-A domain from factor B has provided interesting insights into its previously uncharacterised function. Proteins which are members of the vWF-A domain superfamily have a wide variety of ligand-binding roles (Colombatti and Bonaldo, 1991). The vWF-A domain from factor B can now be included in the list of vWF-A domains which have ligand-binding functions.

Chapter 6

Experimental:
Identification of the C3b binding site in factor B by surface enhanced laser desorption-ionisation affinity mass spectrometry and homology modelling.

6.1 Introduction

A model for the formation of the C3b factor B complex has been proposed which involves interaction via two (or more) low affinity sites on factor B (Pryzdial and Isenman, 1988; Ueda *et al.*, 1987; Volanakis, 1989; Sanchez-Corral *et al.*, 1990; Lambris and Müller-Eberhard,1984; Tuckwell *et al.*, 1997). The complex is depicted as having 2-3 points of contact, one in the Ba fragment and 1-2 in the Bb fragment. Mg^{2+} ions act as an allosteric effector of the Bb binding site (Fishelson *et al.*, 1983). The precise topology of the corresponding sites on C3b has not been determined. C3b (or $C3H_2O$) bound factor B interacts with factor D inducing the active conformation of the enzyme, which in turn catalyses the cleavage of the Arg233-Lys234 bond in factor B. Cleavage of factor B results in higher affinity binding to C3b, sequestration of Mg^{2+} ions, and expression of proteolytic activity for C3 (Fishelson *et al.*, 1983). The transient conformation of factor B can be stabilised by properdin. Decay of the C3bBb C3 convertase enzyme occurs rapidly at 37°C by dissociation of Bb. The loss of 99% of proteolytic activity against C3 of dissociated, as compared to the C3b-bound Bb (Fishelson and Müller-Eberhard, 1984), suggests that the C3bBb complex has influence in the conformation of the Bb fragment.

Among the complement proteins which contain vWF-A domains, factor B, C2, CR3 and CR4 all interact with degradation fragments of C3 (C3b or iC3b), or with those of C4, a homologue of C3. Since the vWF-A domain is common to factor B/C2 and CR3/CR4, it is possible that this domain is involved in ligand binding (Sim and Perkins, 1989). In confirmation of this, the vWF-A domain of factor B binds to C3b (Chapter 5, Section 5.3.2; Williams *et al.*, 1999), and mutagenesis of residues in the vWF-A domain of intact factor B affects its binding to C3b (Tuckwell *et al.*, 1997; Hourcade *et al.*, 1999). Competition binding studies have demonstrated that both Ba and Bb fragments can separately inhibit the binding of factor B to immobilised C3u (Williams & Sim, 1994). Since C2 and C4 are close homologues of factor B and C3 respectively, the interactions in the two systems are likely to be very similar. Mutant recombinant C2 molecules indicated that residues 240-244 of human C2, which lie within the vWF-A domain, are important for the binding of C2a to C4b (Horiuchi *et al.*,1991).

The leukocyte integrin complement receptor type 3 (CR3) is the major integrin of phagocytic cells. It is a member of the β2 integrin family, which consists of four glycoprotein heterodimers, each with a distinct α subunit (CD11a, CD11b, CD11c, and CD11d) non-covalently associated with a common β subunit (CD18). The functions of CR3 are mediated by its binding to several physiological ligands including iC3b (Taniguchi-Sidle and Isenman, 1992). CR3 functions as an opsonic receptor, promoting the divalent cation-dependent binding and uptake of iC3b-derivatized particles by myelomonocytic cells. The CR3 vWF-A domain is an independent structural and functional unit having many of the binding functions of the intact receptor (Lee *et al.*, 1995). Mutation of the divalent metal ion-binding sites in the vWF-A domain of CR3 abolished binding to iC3b (Michishita *et al.*,1993). From site-directed mutagenesis of regions near the amino-terminus of the α'-chain of C3b, there is evidence for a similar binding site within the C3 molecule for CR3 and factor B (Taniguchi-Sidle and Isenman, 1994). The mutant C3b complexes were less able to support cleavage of factor B by factor D than normal C3b complexes, implying a reduced C3b-factor B interaction. The mutant forms of iC3b were found to have reduced binding to CR3.

Ba has been shown to be involved in the initial interaction of factor B and C3b, or $C3H_2O$, during the formation of the C3 convertase (Hourcade *et al.*, 1995; Pryzdial and Isenman, 1988; Ueda *et al.*, 1987). The binding site on Ba cannot fully account for the initial interaction of factor B with C3b or $C3(H_2O)$, because Mg^{2+} ions have been shown to be needed for binding of factor B to C3b and the Mg^{2+} binding site is localised within Bb (Fishelson *et al.*, 1983). Neither the isolated Ba nor Bb fragments have substantial affinity to activated C3 and low intrinsic association constants have been proposed for both interactions (Pryzdial and Isenman, 1988; Volanakis, 1990) and Bb remains bound to C3b after cleavage of factor B by factor D. Competition binding studies have demonstrated that both the Ba and Bb fragments can separately inhibit the binding of factor B to immobilised $C3H_2O$ (Williams and Sim,1993). All three domains have been implicated in the interactions of factor B with C3 (Sim and Perkins, 1990). Current opinion is that the interaction of C3b depends upon two low affinity binding sites within factor B, one on Ba and the other on Bb, and bound Mg^{2+} ions apparently act as an allosteric effector of the latter site (Volanakis, 1990). Thus the avidity

effect provides the overall binding force between factor B and C3. Hydrophobic interactions are thought to be dominant in C3bB complex formation (Pryzdial and Isenman, 1988) although ionic forces are likely to contribute since low ionic strength enhances complex formation (DiScipio 1981). The negatively charged residues Glu736 and Glu737 of C3 appear to be required for the interaction of C3b with factor B although it is difficult to determine whether they serve as actual contact residues, maintain a necessary local conformation, or contribute to a metal binding site.

The rates of formation of the alternative pathway C3 convertase depend on the nature of its environment and are subject to up and down regulation by several control proteins (Pangburn and Müller-Eberhard, 1984). The serum protein factor H and the membrane associated proteins, decay accelerating factor (DAF), complement receptor 1 (CR1) and membrane cofactor protein (MCP) all limit the interaction of factor B with C3b and accelerate the decay of the C3bBb complex (Taniguchi-Sidle and Isenman, 1994). Properdin is the final protein participating in the assembly of the alternative pathway C3 convertase. It is necessary for the formation of a stable complex, resulting in a decreased rate of dissociation of Bb, and therefore for efficient activation of the pathway (Volanakis, 1989). A binding site for properdin has been mapped within a 34-residue segment of the α-chain of C3 (Daoudaki *et al.*, 1988) and it seems likely that the Bb fragment has binding sites for properdin (Volanakis, 1990).

Matrix-assisted laser desorption ionisation (MALDI) mass spectrometry has enabled proteolytic fragmentation patterns to be identified in protein-protein complexes (Beavis & Chait, 1996), and this provides structural information on the residues that participate at the interface between the two proteins in the complex. The technique of biomolecular interaction analysis has been interfaced with MALDI mass spectrometry to provide a powerful joint method for the investigation of protein interactions (Krone *et al.*, 1997). The use of an activated chip surface to which one of the two proteins can be bound covalently simplifies the preparation and identification of these proteolytic fragments (Spencer *et al.*, 1999). This method is known as surface enhanced laser desorption ionisation affinity mass spectrometry

(SELDIAMS) (Hutchens & Yip, 1993; Kuwata *et al.*, 1998; Leung *et al.*, 1998). It was employed to study the interactions between C3 and factor B using a recombinant vWF-A domain, the factor B fragments, native Ba and Bb and native factor B. The interpretation of the peptide data using an homology model for the factor B vWF-A domain, based on the knowledge of 14 homologous crystal structures for five different vWF-A domains, enabled the demonstration that the active site crevice at the carboxyl edge of the β-sheet structure of the vWF-A domain binds directly to C3b, and it was inferred that the amino edge of this β-sheet provides the connections with the SCR and SP domains of factor B. The implications of these observations are discussed below.

6.2 Materials

Fresh frozen plasma was obtained from the Royal Free Hospital blood bank, London, UK and stored at -20°C before use. Chromatography columns and equipment were from Pharmacia, Milton Keynes, Bucks, UK. Cibrachrom Blue dye affinity resin (C-1535), phosphate buffered saline (PBS) (D-5773), thrombin (T-3010), lysozyme (L-6876), glutathione-agarose (G-4510), Lennox broth (L-3022), dithiothreitol (D-9779), Triton (X-100), ampicillin (A-9518), bovine serum albumin (BSA) and benzamidine (B-6506) were purchased from Sigma Chemical Co., Poole, Dorset, UK. Pefabloc (1827) was from Pentapharm AG, Basel, Switzerland. Tris (10315X) and EDTA, ethylenediaminetetraacetic acid (100935V) were from BDH, Poole, Dorset, UK. Isopropyl-β-D-galactoside was from Calbiochem, Nottingham, UK. Trypsin (1 418 025) was from Boehringer Manheim, East Sussex, UK. The hydrophobic marker was from Dako (S2002), High Wycombe, Bucks, UK. Ethanolamine (141-43-5) was from Aldrich, Dorset, UK. Arginine agarose was from (0330-0050) Affinity Chromataography ltd, Freeport, Ballasalla, Isle of Man, UK. Protein concentrators and membranes were from Amicon, Gloucestershire, UK. The SELDI pre-activated surfaces (C700-0039) were from Ciphergen, Camberly, Surrey, UK as were the matrix materials sinapic acid (sinapinic acid, *trans*-3,5-dimethoxy-4-hydroxycinnamic acid), 4-HCCA, (α-cyano-4-hydroxycinnamic acid) and the calibration standards for the spectrometer.

6.3 Methods

6.3.1 SELDIAMS analyses

Native factor B, the Ba and Bb fragments, C3 and the recombinant vWF-A-222 were prepared as described previously (Chapter 4: sections 4.14 and 4.15). C3(NH_3) (an activated form of C3) was prepared by adding ammonium bicarbonate to a final concentration of 100 mM at pH 8.0 to purified C3, then incubating this at 37°C for 1 h, then dialysing overnight into phosphate buffered saline (PBS) (137 mM NaCl, 2.7 mM KCl, 8.1 mM Na_2HPO_4, 1.5 mM KH_2PO_4, 0.5 mM $MgCl_2$, pH 7.3) (von Zabern *et al.*, 1981; Ullman *et al.*, 1998).

For the SELDIAMS experiments, the covalent binding of C3(NH_3) to preactivated chips (Ciphergen Biosystems Ltd., Camberley, Surrey, UK) was achieved via a primary amine based coupling to the N-terminal amino acid and lysine residues. To prevent cross-contamination over the SELDI ProteinChip™ surface, a circle was drawn around each spot on the chip using a hydrophobic marker (Dako, Bucks, UK). The chip surface was prepared by adding 1 µl acetonitrile and allowing it to dry. 2 µl C3(NH_3) at 0.1 mg/ml in PBS were then added to the spots on the chip and kept in a humid environment overnight to allow covalent attachment of the protein. Unbound C3(NH_3) was rinsed off the activated chip surface with 4 µl PBS, and any remaining sites were blocked using 4 µl 1 M ethanolamine, pH 7.5 for 30 min. After the spots had been rinsed three times with 4 µl PBS, they were then probed with 2 µl of protein ligand (either the vWF-A domain, the Ba or Bb fragment, or factor B) at 0.1 mg/ml in PBS. This was left at room temperature in a humid environment for 1 h, after which the spots were rinsed with PBS. The protease incubations were conducted using 2 µl of sequencing-grade trypsin (Boehringer Mannheim, East Sussex, UK) at 5 µg/ml in PBS at room temperature for 1 h, after which the spots were rinsed three times with PBS, then three times with water. The method is summarised in Figure 6.1.

Mass analysis was performed with a PBS-1 mass analyser (Ciphergen Biosystems Ltd.). 1 µl of matrix solution (5 mg/ml) was added to the prepared chip, then this was allowed

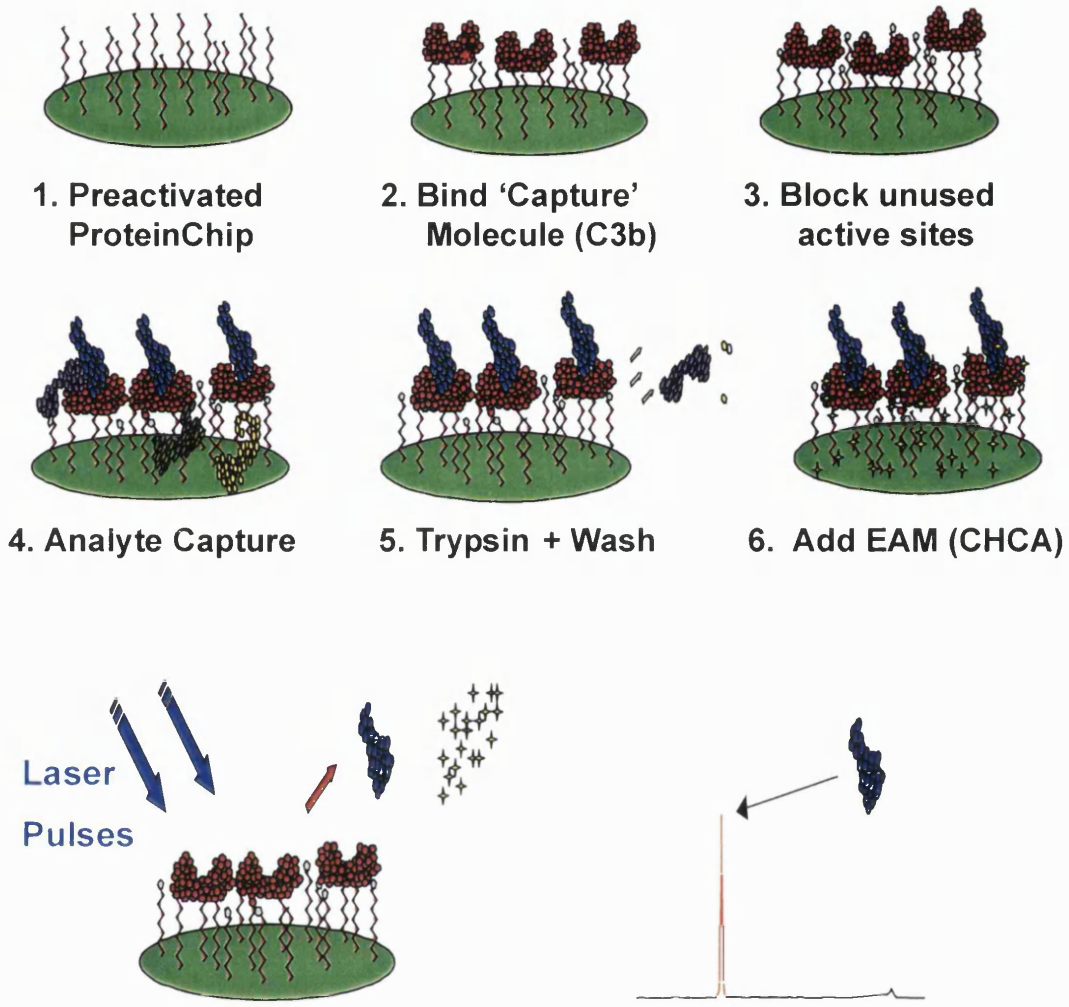


Figure 6.1 Summary of the surface enhanced laser desorption ionisation affinity mass spectrometry technique (SELDIAMS). Step 1 shows the pre-activated surface. Step 2 shows the covalent binding of a protein to the surface. This was typically achieved via a primary amine based coupling to the N-terminal amino acid or lysine residues of the protein. In step 3 any unused sites are blocked. Step 4 shows the capture of the analyte by the surface bound molecule. In step 5 any species that do not associate with the surface bound molecule are washed away. Trypsin is then used to digest the analyte and again the surface is washed to remove any digested analyte that does not remain associated with the surface bound molecule. In step 6 the surface is prepared with the matrix in preparation for laser desorption ionisation where any peptide remnants of the tryptic digest of the analyte molecule will give a signal in the spectrometer.

to dry. The spots were analysed by matrix assisted laser desorption ionisation time of flight mass spectrometry. The protein mass analyses were performed using sinapinic acid (*trans*-3,5-dimethoxy-4-hydroxycinnamic acid) as the matrix in 50% acetonitrile and 0.05% trifluoroacetic acid in distilled water. The peptide mass analyses employed a matrix solution consisting of α-cyano-4-hydroxycinnamic acid in 50% acetonitrile, 50% deionized water and 0.01% trifluoroacetic acid. To determine masses, the instrument was calibrated externally against bovine IgG (mass 147.3 kDa) and internally against bovine ubiquitin (mass 8546.4 Da). The masses from amino acid sequences were calculated using averaged isotopic atomic masses with the Protein Analysis Worksheet (PAWS) program (http://www.proteometrics.com).

6.3.2 Sequence alignment of vWF-A domains

In order to model the vWF-A domain structure of factor B (residues S243 to M443), a total of 23 sets of coordinates from the Protein Data Bank for 14 vWF-A crystal structures (Table 6.1) for CR3, LFA-1, VLA-2 and the A1 and A3 domains of von Willebrand factor were analysed in order to assess the highest similarity with the vWF-A sequence of factor B. The identity of each alignment of two sequences was defined as (the number of identical residues) × 100 / (the total number of topologically-equivalent positions in the alignment, excluding gaps in either sequence). Secondary structure elements were identified using DSSP (Kabsch and Sander, 1983). Residue solvent accessibilities were calculated using a probe of 1.4 Å in the COMPARER program (Lee and Richards, 1971; Šali and Blundell, 1990). For alignment purposes, the DSSP and COMPARER analyses were supplemented by visual inspection using INSIGHT II software (Biosym/MSI, San Diego, U.S.A.) on Silicon Graphics INDY Workstations in conjunction with the superfamily alignment of 75 vWF-A sequences in Perkins *et al.* (1994). Coordinates for 8 CR3 structures were taken from the PDB entries 1ido, 1bho, 1bhq, 1jlm and 1idn. Those for 7 LFA-1 structures were taken from the PDB entries 1zoo, 1zop, 1lfa and 1zon. Those for two VLA-2 structures were taken from the PDB entry 1aox. Those for two vWF-A1 structures were taken from the PDB entries 1auq and 1oak. Those for four vWF-A3 structures were taken from the entries 1ao3 and 1atz.

Table 6.1 The PDB codes of the vWF-A or I-domain structures used for the sequence alignment.

PDB code	Protein	Molecules in the unit cell	Expression system	Metal	Reference
pdb1ido.ent	CD11b (CR3) human	1	*E. coli* PGEX-2T	Mg^{2+}	Lee *et al.*,1995
pdb1bho.ent	CD11b (CR3) human	2	*E. coli* PET3C	Mg^{2+}	Baldwin *et al.*, 1998
pdb1bhq.ent	CD11b (CR3) human	2	*E. coli* PET3C	Cd^{2+}	Baldwin *et al.*, 1998
pdb1jlm.ent	CD11b (CR3) human	1	*E. coli* PGEX-2T	Mn^{2+}	Lee *et al.*,1996
pdb1idn.ent	CD11b (CR3) human	2	*E. coli* PET3C	none	Baldwin *et al.*, 1998
pdb1zoo.ent	CD11a (LFA-1) human	2	*E. coli* PET11C	Mg^{2+}	Qu and Leahy, 1996
pdb1zop.ent mistake in pdb header	CD11a (LFA-1) human	2	*E. coli* PET11C	Mn^{2+}	Qu and Leahy, 1996
pdb1lfa.ent	CD11a (LFA-1) human	2	*E. coli* PET11C	Mn^{2+}	Qu and Leahy, 1995
pdb1zon.ent	CD11a (LFA-1) human	1	*E. coli* PET11C	none	Qu and Leahy, 1996
pdb1aox.ent	α2-β1 human (VLA-2)	2	*E. coli* PGEX-KT	none	Emsley *et al.*, 1998
pdb1auq.ent	vWF-A1 human	1	*E. coli* PET15B	none	Emsley *et al.*, 1998
pdb1oak.ent	vWF-A1 human	1	*E. coli* PET8C	none	Celikel *et al.*, 1998
pdb1ao3.ent	vWF-A3 human	2	*E. coli* pQE9	none	Bienkowski *et al.*, 1998
pdb1atz.ent	vWF-A3 human	2	*E. coli*	none	Huizinga *et al.*, 1997

6.3.3 Homology modelling of vWF-A domain

A homology model of the factor B sequence (residues S243 to M443) which contained the vWF-A domain was readily constructed from the CR3 vWF-A domain (PDB code 1ido; Lee *et al.*, 1995). The sequence alignment of these domains was initially extracted from Perkins *et al.*, (1994) and was modified by eye to optimize the equivalences and minimize the number of gaps in the sequence alignment. The percentage identity between the two sequences was 21.4%. Regions without gaps were defined as structurally conserved regions that formed the secondary structure framework. The insertions and deletions were mainly found in loop positions and were defined as structurally variable regions. This conserved framework that was used for the core of the factor B vWF-A model contained nine peptide fragments from the CR3 structure and comprised 133 residues. Eight peptide fragments of the correct length from known protein structures in the Protein Data Bank were selected for modelling the structurally variable regions in the vWF-A model (factor B residues V275-P282, S305-N306, N315-S326, S342-T355, G365-I374, L384-Y397, V405-L408 and K421-E424). A precalculated Cα distance matrix identified the PDB loops that best fitted the corresponding Cα distance matrix calculated from the structurally conserved regions of the vWF-A framework that defined the starts and ends of these searched loops for a specified number of flanking and intervening residues. These remodelled 68 loop residues comprised 34% of the vWF-A model. Typically for homology models, these loop residues will be less well-defined in position than the 133 framework residues. Sidechain atoms were automatically generated for both the structurally conserved regions and structurally variable regions using the template structures and general rules for residue exchanges. The vWF-A homology model was refined using energy minimization, where 300 steps of steepest descent minimisation were performed with HOMOLOGY and DISCOVER software (MSI/Biosym, San Diego, U. S. A.).

After building an atomic coordinate model, whether by means of X-ray crystallography, NMR or homology modelling techniques, it is important to assess its quality. PROCHECK (Laskowski *et al.*, 1993) is a suite of programs that assesses a PDB-format atomic coordinate file using stereochemical parameters derived from high-resolution protein

crystal structures (Morris *et al.*, 1992), and bond lengths and bond angles derived from a comprehensive analysis of small-molecule structures (Engh and Huber, 1991). The stereochemical quality of the model is output as a residue-by-residue listing that enables the clear identification of regions that are in error. A useful feature of PROCHECK is that it produces a Ramachandran plot of the ϕ (phi) and ψ (psi) mainchain torsion angles (Ramachandran and Sassiekharan, 1968).

Solvent accessibilities of the homology model were calculated using a probe of 1.4 Å in the COMPARER program (Lee and Richards, 1971; Šali and Blundell, 1990). The electrostatic surface charge was calculated using INSIGHT II 97.0 and DELPHI software (Biosym/MSI, San Diego, U.S.A.) on INDY Workstations (Silicon Graphics, Reading, U.K.).

6.4 Results and Discussion

6.4.1 Preparation of the vWF-A domain

The preparations of the 222-residue recombinant vWF-A domain, the native Ba and Bb fragments, and native factor B and C3 each eluted as a single clean peak from the HiLoad 16/60 Superdex-75 gel filtration column. Except for C3, each migrated as a single band on SDS-PAGE under both reducing and non-reducing conditions (not shown), and corresponded to the following expected approximate sizes: vWF-A, 25 kDa; Ba, 30 kDa; Bb, 60 kDa; factor B, 90 kDa. The vWF-A domain was shown to be folded by Fourier transform infrared, circular dichroism and NMR spectroscopy (Perkins *et al.*, 1994; Hinshelwood *et al.*, 1996) (Chapter 7). The vWF-A domain was shown to be functionally active through its ability to inhibit the binding of factor B to C3b (Williams *et al.*, 1999). C3 migrated as a single band in SDS-PAGE under non-reducing conditions, and as two bands in reducing conditions, which corresponded to the expected sizes of 180 kDa for C3, and 110 kDa and 70 kDa for its α and β chains.

6.4.2 Mass spectrometry of immobilised C3(NH_3) with factor B and its components

C3(NH_3) was used as an equivalent of C3b, the activated form of C3 in which nucleophilic attack by free ammonia at the thiolester site converted it into the C3b conformation, which binds to factor B (Sim and Sim, 1983; Pryzdial and Isenman, 1988). C3(NH_3) was covalently immobilised to the preactivated chip (Section 6.1). When this immobilised C3(NH_3) was probed with the recombinant carbohydrate-free vWF-A domain, a peak at a mass of 25,021 Da was observed in the resulting mass spectrum (Figure 6.2a). This peak was present even if the activated chip surface was washed with 50 mM urea (data not shown), and its mass value agreed well with a calculated value of 25,005 Da from the PAWS program (Methods). In control experiments, no peak was observed in this region of the spectrum when the chip was probed with the vWF-A domain in the absence of immobilised C3(NH_3) (Figure 6.2b) or when the immobilised C3(NH_3) was run alone (Figure 6.2c). It was concluded that this vWF-A domain exhibited a specific affinity for C3(NH_3), in agreement with the competition experiments of Williams *et al.* (1999).

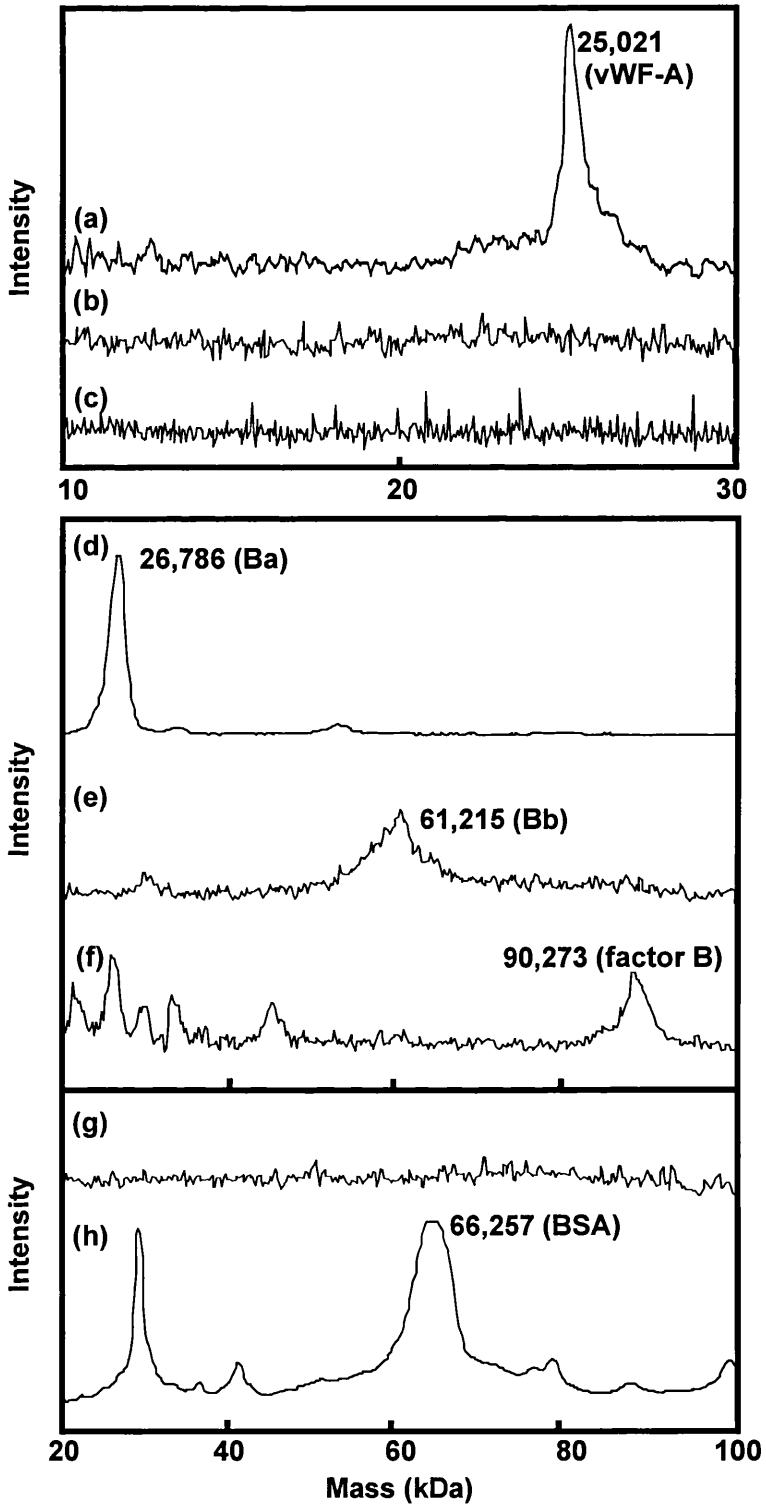


Figure 6.2 Mass analysis of the interaction of the recombinant vWF-A-222 domain, the Ba and Bb fragment, factor B and bovine serum albumin with immobilised C3(NH_3). (a) C3(NH_3) was immobilised on the surface of the activated chip, then probed with vWF-A-222; (b) No C3(NH_3) was present on the surface, but the chip was probed with vWF-A-222; (c) C3(NH_3) was immobilised on the chip, but not probed with vWF-A-222. The same experiment as in (a) was performed for (d) the Ba fragment, (e) the Bb fragment, and (f) native factor B, and also for (g) bovine serum albumin (BSA) as a negative control. The instrument was calibrated externally using bovine IgG, and the protein masses in Da are shown beside each peak. The effect of studying a BSA sample on a normal phase chip surface is shown in (h).

The immobilised C3(NH_3) was likewise probed with the glycosylated native Ba and Bb fragments and factor B. Broad peaks of masses 26,786 Da, 61,215 Da and 90,273 Da respectively with an estimated errors of ±800 Da were observed in the mass spectra (Figures 6.2d, 6.2e and 6.2f respectively). These were in approximate agreement with the expected masses of the three glycoproteins. The calculated masses of the protein component were 25,983 Da, 57,036 Da and 83,001 Da respectively, which would increased to 28-30 kDa, 61 kDa and 89-91 kDa respectively if the masses of one/two, two and three/four biantennary complex-type oligosaccharide chains respectively (each of approximate mass 2,100 Da) were added to these. The appearance of broad peaks was attributed to the presence of two N-linked glycosylation sites in the second SCR domain of the Ba fragment, and two more in the vWF-A domain of the Bb fragment, each of which contain a heterologous population of primarily biantennary complex-type oligosaccharides, together with minor amounts of tri- and tetraantennary types and variable sialic acid contents (Anderson, 1986). When bovine serum albumin was used to probe the immobilised C3(NH_3), no peak was observed at the appropriate mass of about 66 kDa (Figure 6.2g). If bovine serum albumin was noncovalently coated onto a normal phase surface, a peak at a mass of 66,257 Da was observed (Figure 6.2h). It was concluded that the observed peaks in Figures 6.2(d), 6.2(e) and 6.2(f) corresponded to specific interactions of these proteins with activated C3. In particular, this is direct evidence that both Ba and Bb fragments have binding sites for activated C3, in accordance with the multisite model for the assembly of the C3bBb convertase (Lambris and Müller-Eberhard, 1984; Pryzdial and Isenman, 1988; Ueda *et al.*, 1987; Sanchez-Corral *et al.*, 1990).

6.4.3 Mass spectrometry of proteolysed vWF-A domain after interaction with C3(NH_3)

The interaction between the recombinant carbohydrate-free vWF-A-222 domain and immobilised C3(NH_3) was investigated in experiments using proteolysis with trypsin, with and without washes with 50 mM urea (Methods). Trypsin was employed because of the accurate prediction of cleavage sites at Arg or Lys residues, which meant that the proteolytic fragments could be identified by determination of their masses and comparing these with the factor B

sequence. Two sets of mass spectra are shown from these experiments, in which four strong peaks with averaged masses of 3019 Da, 4048 Da, 4174 Da and 6027 Da were observed (Figures 6.3a and 6.3b). There were additional peaks that were observed at mass values above 8,000 Da (data not shown) which were attributed to larger-sized partial digest fragments. It was inferred from these experiments that four peptides from the vWF-A domain had remained associated with the immobilised C3(NH_3) after proteolysis. As controls, a range of different conditions were tested using the permutations summarised to the right in Figure 6.3, including the presence or absence of C3(NH_3), the use of either the vWF-A domain or bovine serum albumin, and treatment of the vWF-A domain or bovine serum albumin with trypsin prior to or after adding the probe to the chip surface. The outcome of 64 runs (including duplicates to confirm reproducibility) showed that the four peaks of Figures 6.3a and 6.3b were absent in the nine experiments outlined in Figures 6.3c to 6.3k. The possibility of nonspecific interactions of the vWF-A peptide fragments with immobilised C3(NH_3) or the chip surface was ruled out with the experiments of Figures 6.3d and 6.3e respectively. The possibility that the tryptic peptides originated from C3(NH_3) or the self degradation of trypsin was eliminated using Figure 6.3g. The controls confirmed that the four peaks did indeed originate from vWF-A peptides with C3b-binding affinity, and indicate that these four peptides remain bound to C3(NH_3) during the post-trypsin wash stage.

In order to assign the four peptide mass peaks to specific residues in the vWF-A domain, the vWF-A sequence was analysed for tryptic fragments corresponding to these masses using the PAWS program (Figure 6.4). A single segment of the vWF-A sequence could be unambiguously identified for each mass peak to within 0.2 % (Table 6.2). This is compatible with the accuracy of matrix-assisted laser desorption ionisation time-of-flight spectrometry (Krone *et al.*, 1997). Thus the 6027 Da, 4174 Da, 4048 Da and 3019 Da mass signals corresponded to the vWF-A residues G229-R283 (+GS), G229-K266 (+GS), G229-K265 (+GS) and T355-R381 respectively in factor B. The N-terminal GS dipeptide results from the remnant of the fusion protein thrombin cleavage site (Smith and Johnson, 1988; Williams *et al.*, 1999). The nearest approach to these mass values by other possible tryptic peptides involved larger mass differences of 150 Da, 22 Da, 20 Da and 24 Da in

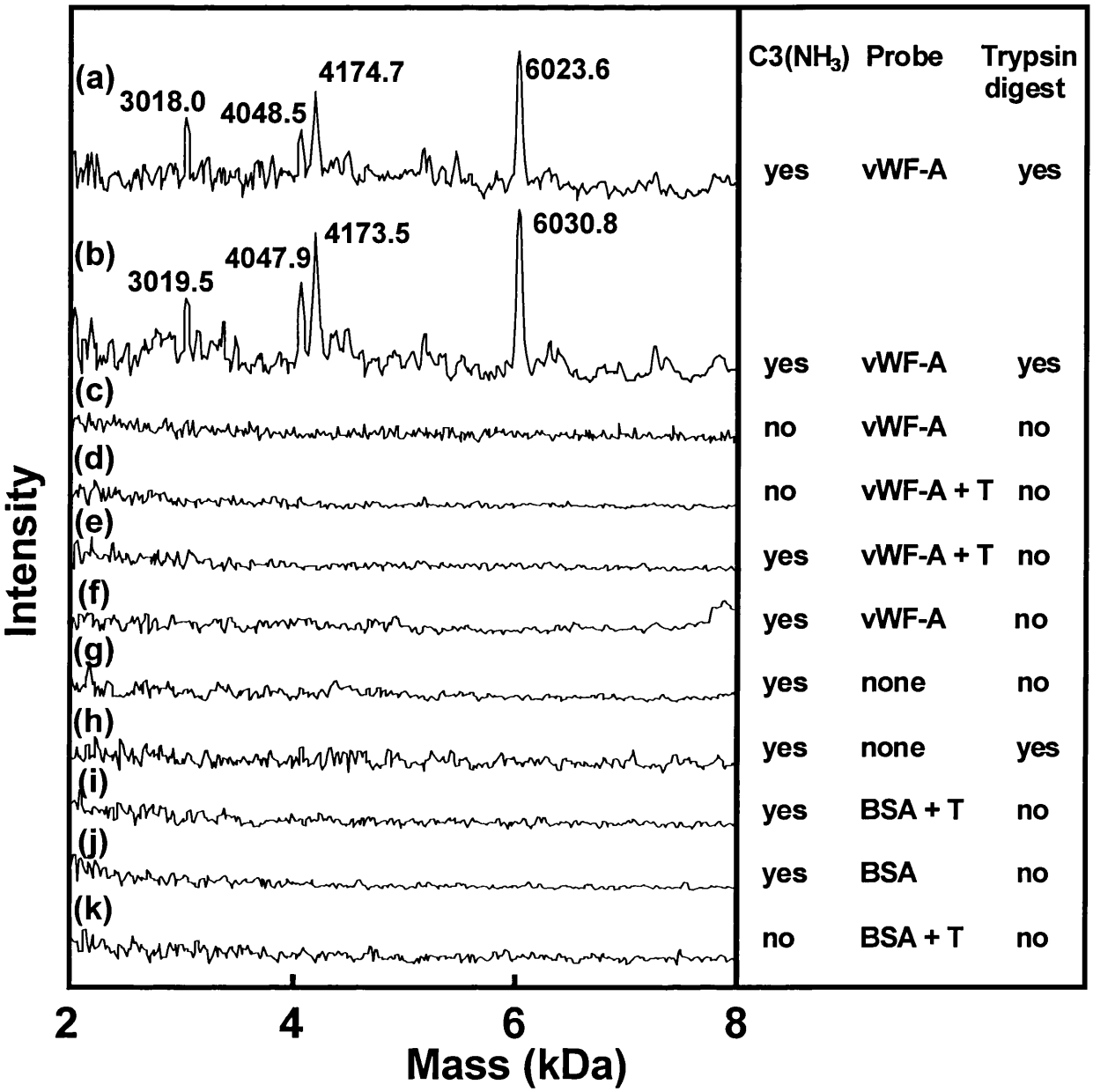


Figure 6.3 Mass analysis of the peptides derived from the trypsin digest of the vWF-A-222 domain bound to immobilised C3(NH_3) on the activated chip. Only the spectrum between 2-8 kDa is shown here, in distinction to Figure 6.2. The four strongest peaks are labelled with their corresponding mass values (Table 6.2). The instrument was calibrated internally against bovine ubiquitin (mass 8564.8 Da).

(a) The complex of vWF-A and C3(NH_3) was digested with trypsin, then washed with PBS, then with water.

(b) The experiment in (a) was repeated after washing with PBS that contained 50 mM urea to act as a more stringent washing agent.

Other experiments are summarised in the right-side panel which defined the components that were present, including the control with bovine serum albumin (BSA) and the pretreatment with trypsin (T).

(a)

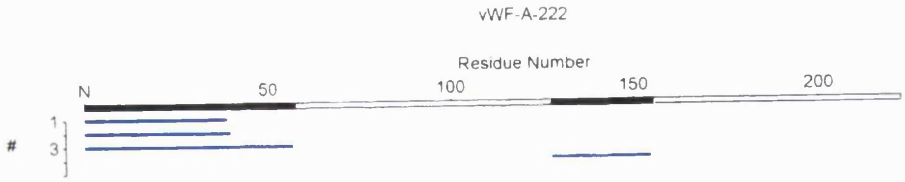
vWF-A-222
Cleavage at KR
[1-222] mass = 25005.3

G[1-39]K = 4041.5 G[1-40]K = 4169.7 G[1-57]R = 6024.9 I[10-39]K = 3042.4

```
  1  G S G E Q Q K r K i v l d p s g s m n i y l v l d g s d s i  30
 31  g a s n f t g a k K s l v n l i e k V A S Y G V K P R y g l  60
 61  v t y a t y p k I W V K v s e a d s s n a d w v t k Q L N E  90
 91  I N Y E D H K l k S G T N T K k A L Q A V Y S M M S W P D D  120
121  V P P E G W N R t r H V I I L M T D G L H N M G G D P I T V  150
151  I D E I R d l l y i g k D R k N P R e d y l d v y v f g v g  180
181  p l v n q v n i n a l a s k K d n e q h v f k V K d m e n l  210
211  e d v f y q m i d e s q                                    222
```

(1)	[1-7] = 732.7	(2)	[8-8] = 174.2	(3)	[9-9] = 146.2
(4)	[10-39] = 3042.4	(5)	[40-40] = 146.2	(6)	[41-48] = 915.1
(7)	[49-57] = 976.1	(8)	[58-68] = 1275.5	(9)	[69-72] = 544.7
(10)	[73-86] = 1508.6	(11)	[87-97] = 1402.5	(12)	[98-99] = 259.3
(13)	[100-105] = 606.6	(14)	[106-106] = 146.2	(15)	[107-128] = 2549.9
(16)	[129-130] = 275.3	(17)	[131-155] = 2760.2	(18)	[156-162] = 821.0
(19)	[163-164] = 289.3	(20)	[165-165] = 146.2	(21)	[166-168] = 385.4
(22)	[169-194] = 2838.2	(23)	[195-195] = 146.2	(24)	[196-203] = 1016.1
(25)	[204-205] = 245.3	(26)	[206-222] = 2106.3		

(b)



#	Mass	Matching sequence (± error)
1	4048.2	G[1-39]K (-5.70)
2	4174.1	G[1-40]K (-3.43)
3	6030.8	G[1-57]R (-4.92)
4	3019.5	T[129-155]R (-0.99)

Figure 6.4 Output from the PAWS program (http://www.proteometrics.com) calculated from the vWF-A-222 amino acid sequence using standard atomic masses. The program can predict all the possible peptides resulting from a tryptic digest (cleavage at K or R) of the protein and then searches for those corresponding to specific masses. The peptide masses that were searched for in this instance were 4048.2, 4174.1, 6030.8 and 3019.5 Da to within 0.2% error. (a) The results can be displayed in terms of the amino acid sequence. The predicted mass of the whole domain is 25005.3 Da. The list of 26 peptides below the sequence represent those produced by complete cleavage by trypsin. (b) Graphical representation of the cleavage data. The peptides of interest were derived from two regions of the vWF-A-222 domain, denoted by the shaded regions of the sequence. The sequence corresponding to the searched mass is described to the right of the mass near the bottom of the page. The discrepancy between the searched mass and the peptide mass is written in parenthesis following the peptide sequence.

Table 6.2 Observed masses of the peptides that remain associated with C3(NH_3) and their assignment to factor B vWF-A peptides

Observed mass (Da)*	**Calculated mass of peptide (Da)**	**Mean difference (Da)**	**Sequence of factor B**
6023.6, 6030.8	6024.9	2.3	G229-R283**
4174.7, 4173.5	4169.7	4.4	G229-K266**
4048.5, 4047.9	4041.5	6.7	G229-K265**
3018.0, 3019.5	3017.5	1.3	T355-R381

* The two mass values correspond to the experiments of Figures 6.3a and 6.3b.

** In the recombinant vWF-A domain, there is an additional N-terminal dipeptide Gly.Ser which is derived from the thrombin cleavage site of the fusion protein (Perkins *et al.*, 1994; Williams *et al.*, 1999).

comparison to the small mass differences of 1 to 7 Da in Table 6.2, and so could be ruled out. It was concluded that the residues G229-K265 (Peptide 1) and T355-R381 (Peptide 2) of factor B were the minimum sizes of peptides that were involved in the interaction of the vWF-A domain with C3(NH_3), although residues in Peptide 1 extending up to R283 may also participate in the interaction with C3(NH_3).

Factor B is polymorphic due to the existence of allotypes and variable carbohydrate compositions (Campbell and Bentley, 1985; Anderson, 1986). No peptide signals were observable in SELDIAMS experiments when the immobilised C3 was probed with the native glycosylated factor B or the Ba or Bb fragments, then subjected to tryptic proteolysis. In native factor B, the two vWF-A peptides identified in the SELDIAMS experiments are associated with the N-linked glycosylation sites at N260 and N353 (Figure 6.5), and the difficulty of identification of any peptide signals from the Bb fragment was partly attributed to the presence of heterologous carbohydrate glycoforms.

6.4.4 Homology model for the vWF-A domain in factor B

The interpretation of the peptide data for the proteolysed vWF-A domain required a structural analysis of this superfamily, as no experimentally-determined structure was available for vWF-A from factor B. A total of 23 sets of coordinates from 14 crystal structures for 5 different homologous vWF-A domains were analysed (Methods), from which a sequence alignment could be derived from the identification of their secondary structures (Figure 6.5). The six hydrophobic β-strands BA, BB, BC, BD, BE and BF were well conserved in position and length to within ± 1 residue in the 23 structures (except for BC in LFA-1 which was shifted by 2 residues). While the amphipathic surface-exposed α-helices A1 and A4 were well conserved to ± 2 or ± 3 residues in length, the other α-helices A2, A3, A5, A6 and A7 showed significant variation in terms of length and conformation between the five structures, especially towards the C-terminus of the protein fold. Thus A2 and A3 were merged into a single α-helix in the VLA-2, vWF-A1 and vWF-A3 structures, A5 was missing in LFA-1 because of a sequence deletion, A6 was sometimes preceded or followed by additional 3_{10}-helices or α-helices, and A7 was variable in its length and position (or was

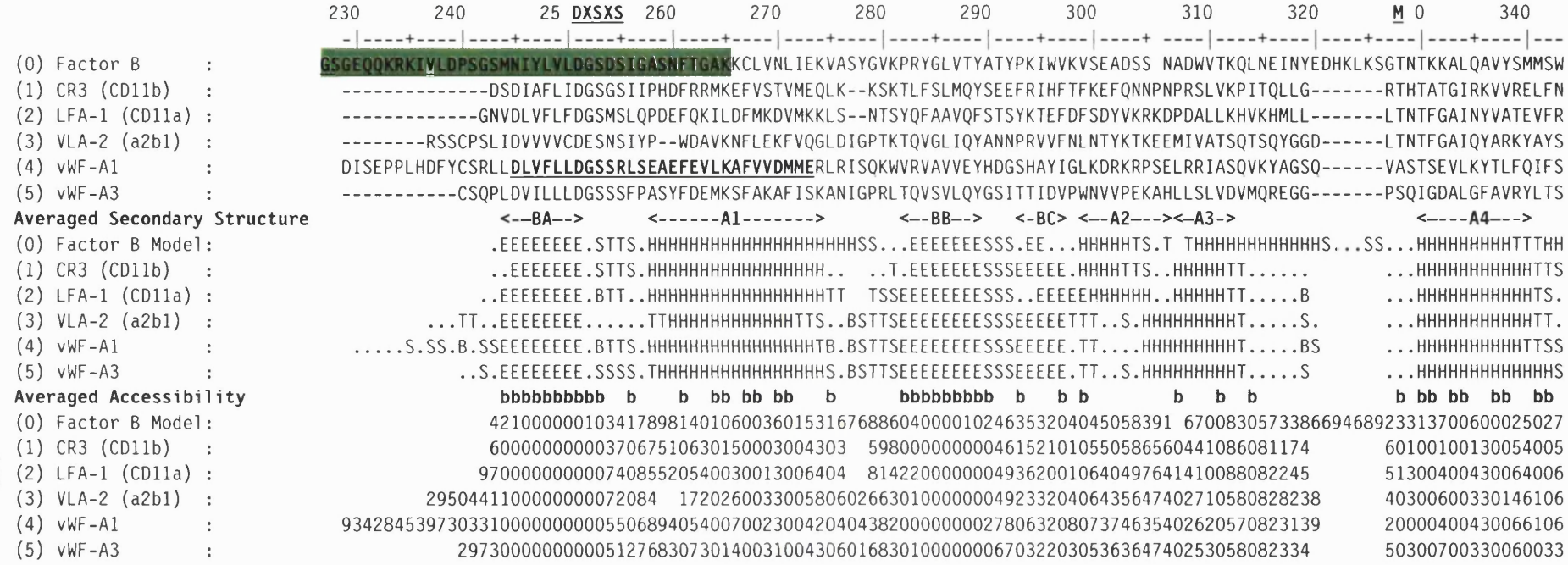


Figure 6.5 Sequence alignments of vWF-A domains. The consensus alignment of known crystal structures used for the homology model of factor B. The SELDI-identified Peptides 1 and 2 in factor B are highlighted in green and yellow respectively. The GpIb receptor site in the von Willebrand factor A1 domain and the iC3b site in CR3 are bolded and underlined. The location of the metal binding residues is denoted by DXSXS and M above the sequences. The secondary structure elements identified by DSSP are denoted as: H, α-helix; G, 3_{10}-helix; E, β-strand; B, single residue β-ladders; T, turns; S, bends, and their positions were labelled A1 to A7 and BA to BF to follow Lee *et al.* (1995a) and Edwards & Perkins (1996). The solvent accessibilities calculated using COMPARER are denoted by values of 0 to 9 to correspond to accessibilities of 0-9%, 10-19%, and so on up to 90-99%, and are denoted by 'b' if the mean crystallographic value is 0 or 1. For each of the five independent vWF-A crystal structures sets (Methods), the secondary structures and accessibilities were averaged from the individual analyses for between 2-8 coordinate sets.

continued overleaf

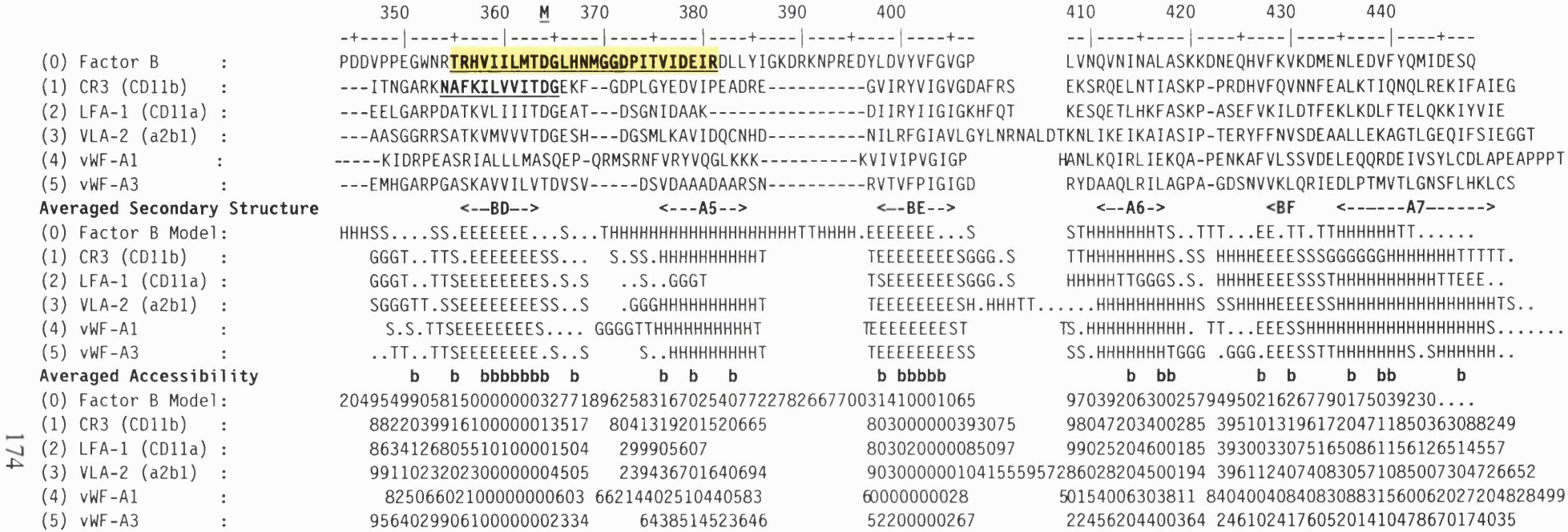


Figure 6.5 Sequence alignments of vWF-A domains

divided into two in the vWF-A3 structure). These large structural variations in α-helices between the five different vWF-A sequences is attributable to their surface location in the vWF-A domain at which packing constraints are absent. This exposure leads, for example, to the noticeable solvent dependence of the α-helix band in Fourier transform infrared spectra of the vWF-A domain (Perkins *et al.*, 1994).

The comparison of the vWF-A domain of factor B with the five different crystal structures in Figure 6.5 was derived from that of Figure 1 in Edwards and Perkins (1996), which was based on comparisons with predicted structures. The sequence identity (defined in Methods) was 23.8% for the comparison of factor B and CR3 using the alignment of Figure 6.5, which was higher than those derived for the other four sequences of Figure 6.5. These were 20.6% for the comparison with LFA-1, 19.8% with VLA-2, 13.7% with vWF-A1 and 15.3% with vWF-A3. The CR3 structure also contained a Mg^{2+} site, unlike the vWF-A1 and vWF-A3 structures, and possessed an iC3b site which may or may not resemble the C3b site of the vWF-A domain of factor B. The CR3 vWF-A crystal structure was thus used as a template for the homology modelling of residues S243-M443 in the vWF-A domain of factor B (SECTION), although the C-terminal α-helix A7 has exhibited variable conformations in the 8 CR3 coordinate sets available to date. Evidence in support of the vWF-A homology model was provided by the conservation of the six β-strands and the α-helices A1 and A4, the location of the conserved metal-binding residues DXSXS, T and D in their expected positions, and by the high solvent accessibilities of N260 (80%) and N353 (50%), on α-helix A1 and between α-helix A4 and β-strand BD respectively, which agrees with their putative glycosylation (Figure 6.5). Likewise the high accessibilities of 60% for the two different glycosylation sites in the homologous C2 vWF-A structure supported this factor B vWF-A model (Figures 6.5 vs. 6.6). The use of PROCHECK to test the model showed that only E423 and Q424 were minor outliers in a Ramachandran plot (Figure 6.7), while residue 232 in CR3 which was a minor outlier in its crystal structure was no longer so in the model. That E423 and Q424 were outliers was attributed to their position next to an insertion that was not present in any of the vWF-A crystal structures (Figure 6.5), which did not facilitate their modelling.

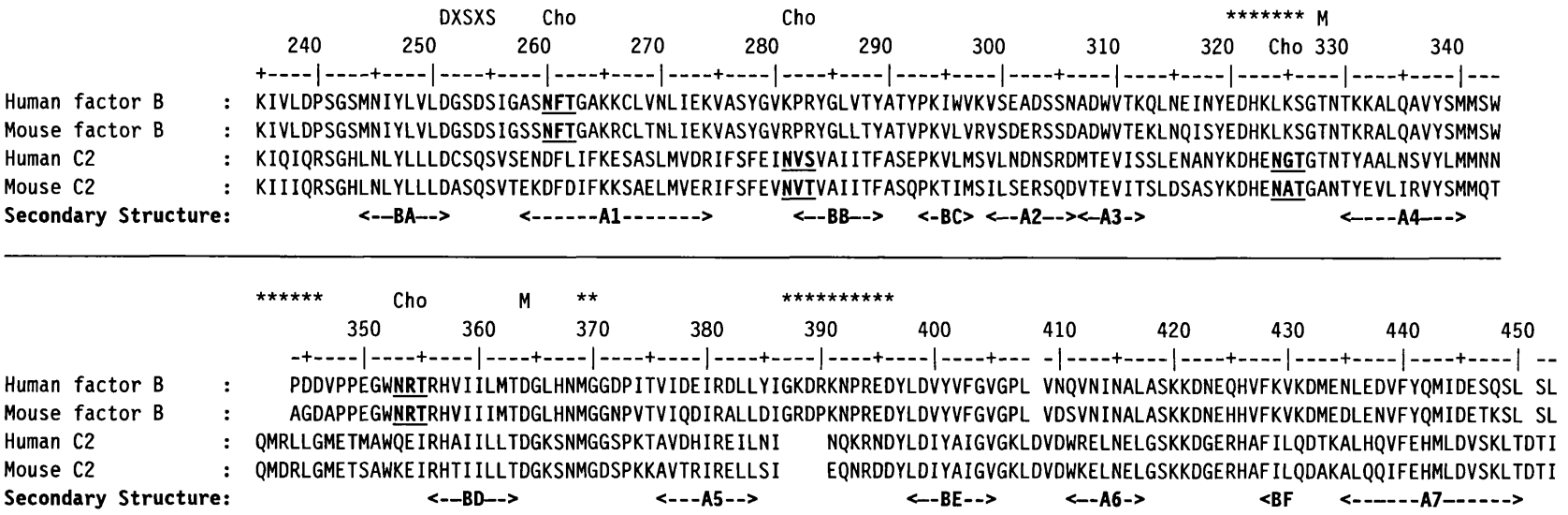
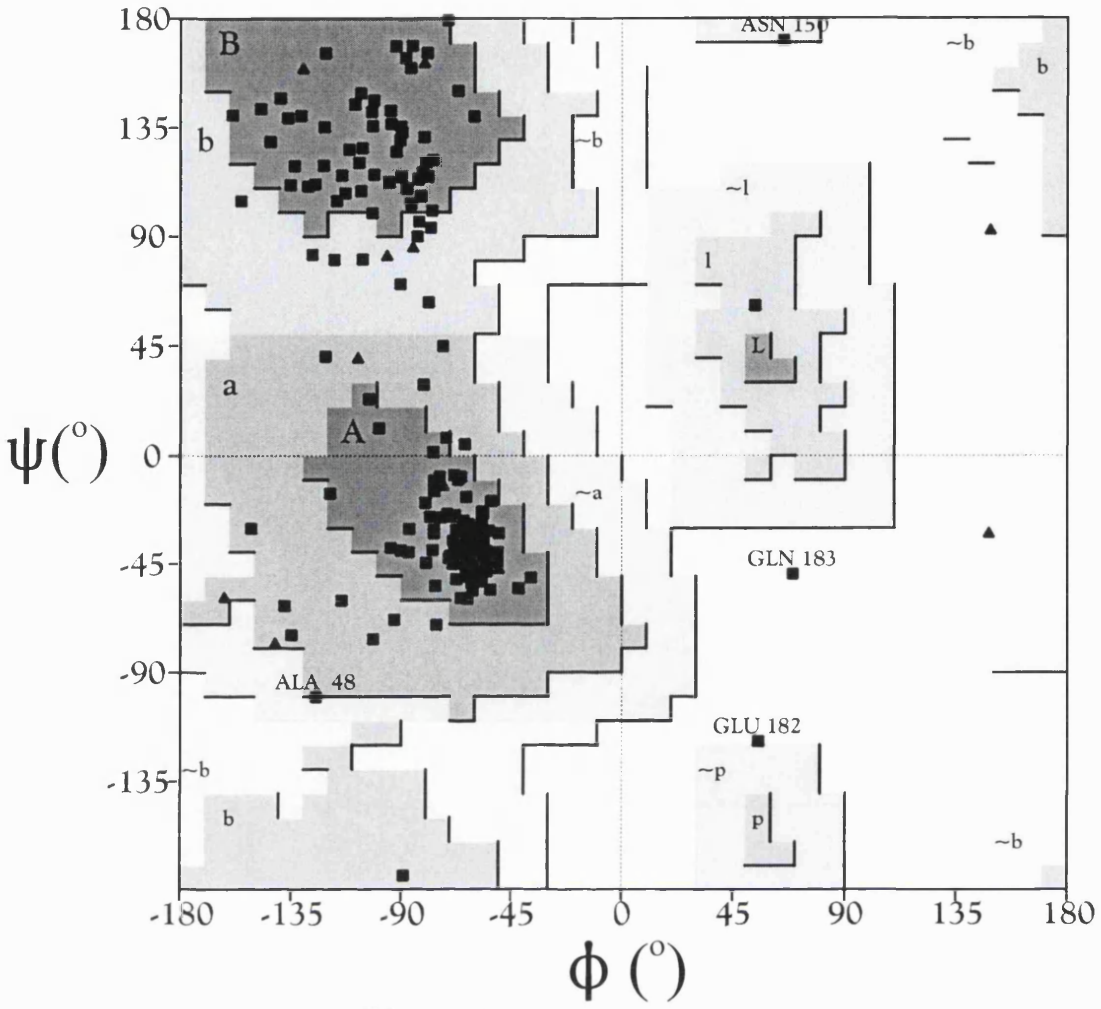


Figure 6.6 Sequence alignments of vWF-A domains from human and mouse factor B and C2 sequences. The four sequence insertions in factor B (see text) are denoted by asterisks. The location of the carbohydrate sites is denoted by Cho above the sequences. The sequences are shown in comparison with the consensus secondary structure from Figure 6.5 for ease of reference.

The most notable result from the vWF-A homology modelling was that four sequence insertions at about E320-S326, P344-D346, M369-G370 and G387-D396 could be identified relative to the positions of conserved sequence/structure features in the five vWF-A crystal structures and were not present in any of these crystal structures (Figure 6.5). Two of these (M369-G370 and G387-D396) occured at the carboxyl-edge of the vWF-A β-sheet near the C3b binding site, while the other two (P344-D346, E320-S326) occurred at the amino-edge of this β-sheet, close to the linkers that join the vWF-A domain to the three SCRs in the Ba fragment and the SP domain in the Bb fragment. The four insertions occur just before and after the α-helices A4 and A5, and have the effect of deepening one side of the active site cleft in the vWF-A domain. It is significant that the four insertions are present also in the homologous vWF-A C2 sequence, although two insertions are different in size (asterisked in Figure 6.6). In factor B, the three larger insertions possess remarkably high proportions of charged residues compared to the rest of the vWF-A domain, although none of them is involved with the C3b-binding site identified by SELDIAMS experiments. In human factor B, the ten-residue insertion G387-D396 possesses three acidic and four basic residues, while the three-residue insertion P344-D346 possesses two acidic residues, and the seven-residue insertion E320-S326 possesses two acidic and three basic residues. Similar totals are present in the insertions of mouse factor B and human and mouse C2 (Figure 6.6). Electrostatic views of the carboxyl-edge face of the model indicates that there is a prominently acidic active site cleft at this side where Mg^{2+} is bound, while the amino-edge face shows no prominent acidic or basic regions (Figure 6.8). There is also a cluster of acidic residues near to α-helix A7 in both factor B and C2. In factor B, the oligosaccharide at N260 is located at the carboxyl-edge face of the β-sheet, while that at N353 is located at its amino-edge (Figure 6.6). In C2, the oligosaccharide at N281 is located at the amino edge face of the β-sheet, while that at N324 is located on the opposite face at the carboxyl edge. Both occur in topologically-similar locations to those observed in factor B (Figure 6.6).

6.4.5 Location of Peptides 1 and 2 in the vWF-A homology model

Both the vWF-A peptides identified from SELDIAMS occurred at structurally well-defined positions in the alignment and in the vWF-A homology model. Peptide 1



Plot statistics

Residues in most favoured regions [A,B,L]	153	86.0%
Residues in additional allowed regions [a,b,l,p]	21	11.8%
Residues in generously allowed regions [~a,~b,~l,~p]	2	1.1%
Residues in disallowed regions	2	1.1%
Number of non-glycine and non-proline residues	178	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	13	
Number of proline residues	8	
Total number of residues	201	

Figure 6.7 A Ramachandran plot of the mainchain torsion angles ϕ and ψ of the vWF-A domain homology model. The plot was generated using PROCHECK. Allowed regions are shown in grey for α-helix (A), β-sheet (B), left-handed helix (L) and poly-proline (P) mainchain conformations, with the most favoured conformations indicated by the darkest regions. Glycine residues are indicated by triangles and all other residues are shown as squares. Glu182 and Gln183 are the only non-glycine residues that occupy disallowed regions of the plot. These correspond to residues Glu423 and Gln424 of the factor B sequence. The plot statistics are presented under the Ramachandran plot.

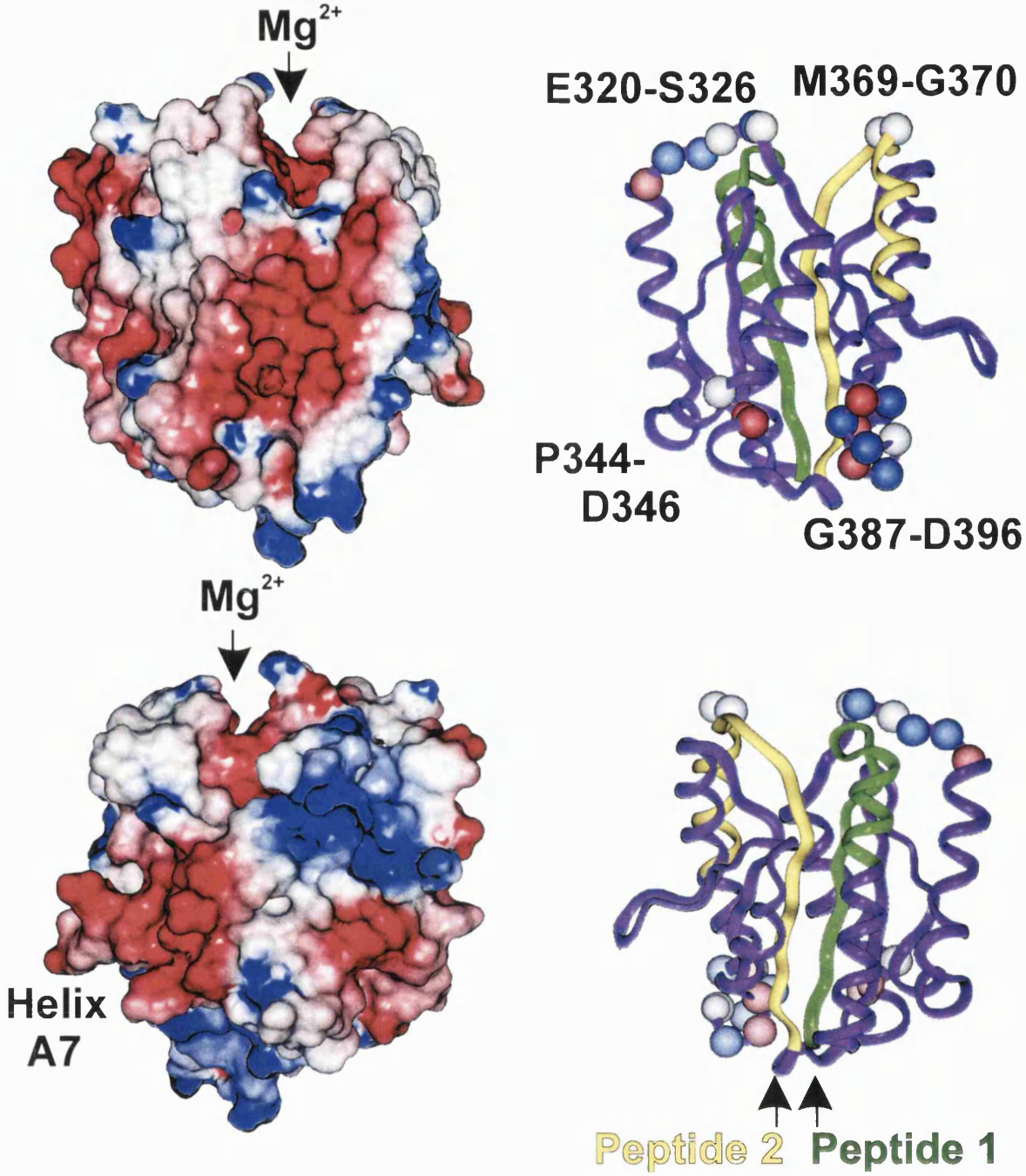


Figure 6.8 Two views rotated by 180° that depict the surface of the factor B vWF-A homology model. Electrostatic and ribbon representations are shown for each view. Peptides 1 and 2 are shown in green (G229-K265) and yellow (T355-R381) respectively. The Mg^{2+} binding site is in a cleft (arrowed) and lined at its base by acidic residues. The N-terminus and C-terminus are at the base of the structure. The positions of four insertions (E320-S326 and M369-G370 next to the Mg^{2+} site; P344-D346 and G387-D396 near the N-terminus and C-terminus) that are not seen in other crystal structures in the vWF-A superfamily are indicated by spheres (blue, basic; red, acidic; white other). The electrostatic views show several other acidic and basic regions, of which a prominent group of acidic residues corresponds to a surface loop (D422 and E424) and the surface of helix A7 (D432, D434, E437, D438, D445 and E446) to the left in the lower view.

corresponded to a region from the N-terminus that included the BA β-strand to half way along the α-helix A1. This region contained the conserved DXSXS motif found in many vWF-A sequences, which corresponds to part of the Mg^{2+} binding site (Perkins *et al.*, 1994). The possible extension of Peptide 1 up to R283 (Table 6.2) includes the remainder of the α-helix A1 up to the start of the β-strand BB. Peptide 2 included the BD β-strand, a two-residue insertion, and the start of the A5 α-helix. It also contained D364 that provided part of the Mg^{2+} binding site (Figure 6.5).

Topologically, Peptides 1 and 2 are seen to be proximate to each other in the factor B vWF-A homology model (yellow and green ribbons respectively, Figure 6.8), and include the crossover point in the parallel β-sheet from where the two central β-strands wind to the opposite ends of the β-sheet. This crossover point forms a cleft at the carboxyl-edge of doubly-wound α/β folds, and contains the Mg^{2+} site in many members of the vWF-A superfamily. Only a limited number of the 39 and 27 residues in Peptides 1 and 2 respectively can be available for protein-protein interactions with C3b. For instance, since the peptide G229-R234 is cleaved from the vWF-A domain on activation by factor D, these residues are ruled out from a C3b binding role. The adjacent residues K235-S243 at the N-terminus of the vWF-A domain are unlikely to bind to C3b, since they will form part of the factor D binding site which has to be accessible to factor D after the vWF-A domain is bound to C3b (Williams *et al.*, 1999). The residues M244-S253 in Peptide 1 and V358-D364 in Peptide 2 correspond to buried sidechains in the β-strands BA and BD respectively, some of which contribute to the Mg^{2+} site, so cannot be available for direct interactions with C3b. From Figure 6.5, other predicted buried residues in the vWF-A model can be discounted also (I256, F261, A264 and K265 in Peptide 1; V376 and E379 in Peptide 2). This leaves 8 and 17 residues in Peptides 1 and 2 respectively that have surface accessible sidechains that can potentially form contacts with C3b (Figure 6.5). Of these 25 residues, 4 are acidic, 3 are basic, 11 are neutral hydrophilic, and 7 are hydrophobic, and the majority of these are located at the carboxyl-edge of the β-sheet in the vWF-A domain. It is thus concluded by elimination that the amino acids that corresponds to the C3b-binding site are located at the active site crossover point of the vWF-A domain.

6.5 Conclusions

The combination of experiments using a new development of SELDI technology with a vWF-A homology model constructed from five different crystal structures has provided novel insight into the functional activity of the vWF-A domain in factor B, i.e. that the function of its active site cleft is to bind to C3b. Mass spectroscopy analysis of tryptic digests of a functionally-active 222-residue recombinant vWF-A domain from factor B (Williams *et al.*, 1999) showed that two peptides G229-K265 and T355-R381 were directly implicated in ligand binding to C3b. This experimental approach is based on a recombinant vWF-A domain that is of high purity, being correctly folded and functionally active. The primary advantage of SELDIAMS is that it corresponds to the biochemical analysis of the whole protein surface. The comparison with the vWF-A homology model showed that both peptides flanked its active site cleft at the carboxyl-edge of the central hydrophobic β-sheet. The vWF-A homology model was used to narrow down the specific residues that participate in C3b-binding. This showed that potential C3b-binding residues in both peptides could be eliminated to leave only about 25 that are primarily located at the exposed surface of this active site cleft. Of interest is that the first of these factor B vWF-A peptides coincides in position with the GpIb platelet receptor binding site at residues 514-542 in the A1 domain of von Willebrand factor (Knott *et al.*, 1992; Sinha *et al.*, 1994; Matsushita and Sadler, 1995), numbered as residues 245-273 in Figure 6.5. The second of these peptides coincides in position with the iC3b active site peptide in the vWF-A domain of CR3 identified by Ueda *et al.* (1994), numbered as residues 354-365 in Figure 6.5.

Additional information on the C3b-binding site of the vWF-A domain was obtained from its homology model. Residues in the metal binding site of the integrin vWF-A domains CR3, LFA-1 and VLA-2 are conserved in factor B. This suggests that the Mg^{2+}-binding site is conserved in the vWF-A domain of factor B. However this observation is complicated by the observation that factor B will bind to C3b in the absence of Mg^{2+} (Pryzdial and Isenman, 1987; Williams and Sim, 1994), the absence of one of these metal-binding residues in mouse C2 (Figure 6.6), and the non-conservation of the metal binding site in all three vWF-A domains of von Willebrand factor (Jenkins *et al.*, 1998). Both hydrophobic and ionic

interactions have been implicated in C3b factor B complex formation, where low ionic strengths favour complex formation (Pryzdial and Isenman, 1988; DiScipio 1981). This is consistent with the 25 potential C3b-contact residues which include both hydrophobic and hydrophilic residues, as well as 7 charged residues. The presence of other charged residues at three of four significant insertions identified in the vWF-A homology model and on α-helix A7 may be relevant also. Of interest is that E736 and E737 in C3 have been implicated in the interaction of C3b with factor B (Taniguchi-Sidle and Isenman, 1994). It can be speculated that these charged groups may complete the Mg^{2+} coordination by analogy with the observed intermolecular lattice contacts in the crystal structure of CR3 where a Glu residue behaves as a ligand mimic (Lee *et al.*, 1995a).

This application of SELDIAMS is complementary to site-specific mutagenesis experiments in the sense that an increase or decrease in activity following a single-residue mutagenetic change does not unambiguously identify the role of that specific residue. It is necessary to show that the protein folding is unaffected by the mutation, so a systematic study would require mutagenesis experiments throughout the protein. Consequently SELDIAMS offers much economy of effort through its ability to analyse the whole protein surface, and is potentially applicable to other protein-protein complexes (Spencer *et al.*, 1999). The results of the present SELDIAMS analysis agree with a mutagenesis study of single-residue changes in factor B (Hourcade *et al.*, 1999) that demonstrated that D254, N260 and D364 were involved for C3b binding (Figure 6.5). The SELDIAMS experiments also agreed with the outcome of experiments with Chimeras 1 and 3 of intact factor B by Tuckwell *et al.* (1997). Chimera 1 corresponded to part of Peptide 1, in which G252-S259 of factor B was mutated to C252-N259 of human C2, and Chimera 3 corresponded to part of Peptide 2, in which L366-D372 of factor B was mutated to K366-S372 of human C2, both of which resulted in loss of C3b binding activity. As Chimera 1 contained a free C252 residue, which could become disulphide-linked with the free C267 residue of factor B, that study was not conclusive in showing that the formation of a potential Cys-Cys link might have caused the lack of interaction between Chimera 1 and C3b.

The role of factor B in binding to C3b and the resulting activation of the alternative pathway of complement activation has been clarified in the present studies. Early crystallographic studies on the vWF-A domain in CR3 suggested that it could exist in two conformations depending on the presence of the bound cation (Lee *et al.*, 1995a, 1995b), although the biological significance of this finding is unclear as the crystal lattice packing contacts may be responsible for some of these observations (Baldwin *et al.*, 1998; Liddington and Bankston, 1998). Nonetheless there is evidence from mutational studies indicating "on" and "off" conformations, as well as the separate locations of the mutation sites associated with the Type 2B and 2M phenotypes in von Willebrand's disease, that the vWF-A1 domain in von Willebrand factor is able to transmit ligand binding signals within its structure (Cooney and Ginsberg, 1996; Jenkins *et al.*, 1998). The vWF-A homology model showed that the factor D cleavage site is 8 residues away from its N-terminus (Figure 6.9). The recombinant vWF-A domain also included an extra 8 residues that are N-terminal of the factor D cleavage site. It was possible for factor D to cleave this recombinant vWF-A domain both in the presence or absence of C3b, although this cleavage is slow in both cases (Williams *et al.*, 1999). In contrast to this, the cleavage of intact factor B by factor D in the presence of C3b occurs efficiently, but does not occur in the absence of C3b. This means that this factor D cleavage site is sterically concealed, and/or it possesses an inactive conformation in intact factor B until it comes into contact with C3b.

If the factor B vWF-A domain is able to transmit conformational signals, one possible explanation of the factor D cleavage properties of intact factor B is that the binding of the active site of vWF-A to C3b induces a conformational change at its opposite face to upregulate factor D cleavage (Figure 6.9). Another possible explanation would be that the independent binding and reorientation of the SCR domains of intact factor B to another region of C3b would cause a conformational change in factor B to make this peptide available for cleavage by factor D (Figure 6.9). Irrespective of which explanation is followed, these experiments show that the role of the vWF-A domain in factor B is to facilitate its cleavage by factor D after binding to C3b, and can mediate the regulation of the activities of factor B. The release of the three SCR domains from factor B or an allosteric structural change in the

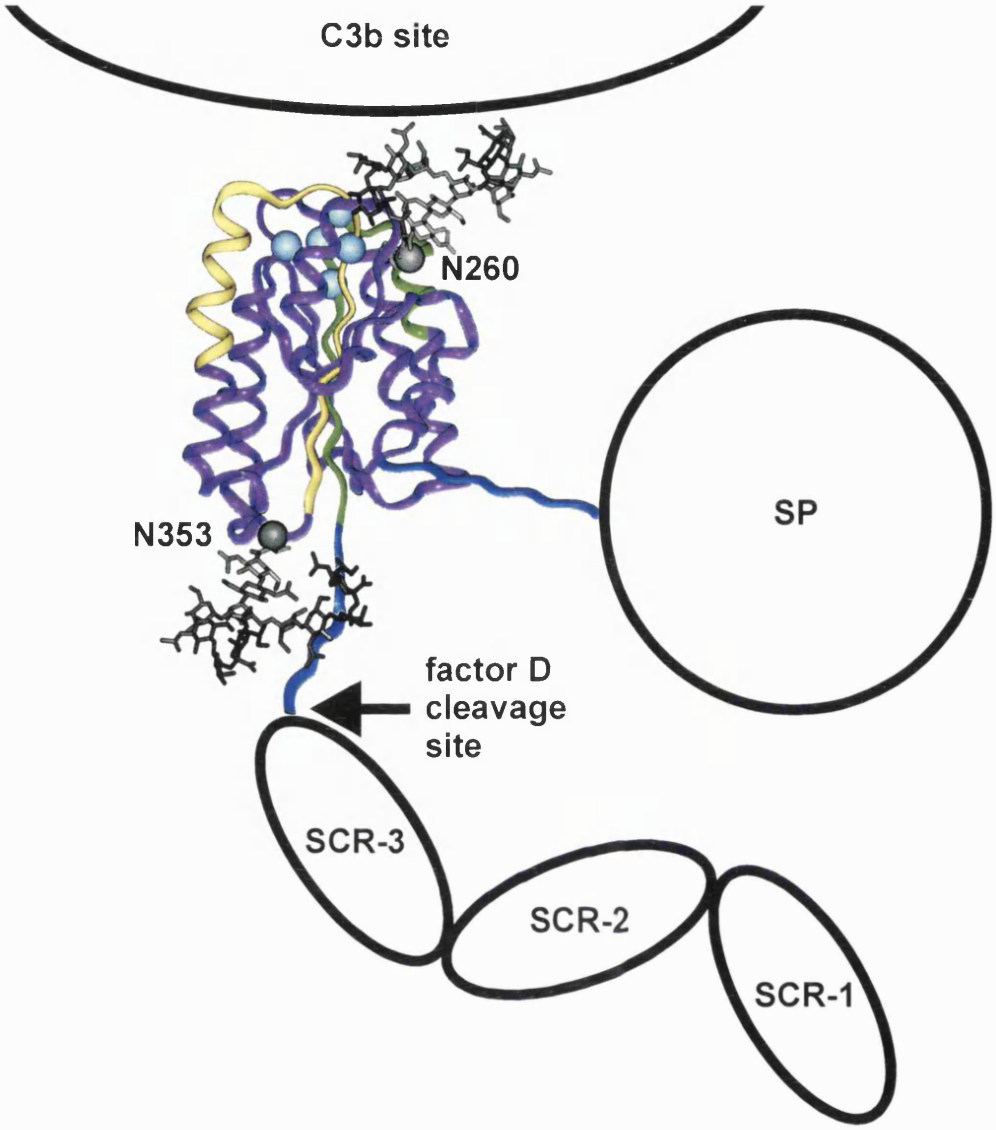


Figure 6.9 Schematic view of the vWF-A structure in relation to C3b and the other domains of factor B. The SCR and SP domains are of approximately correct sizes but their positions are not known. In the vWF-A domain, the Mg^{2+}-cleft is at the top in proximity to the C3b binding site, its N-terminus is connected to the SCR domains, and its C-terminus is connected to the SP domain. This view is rotated by 90° compared to those shown in Figure 4. The five Mg^{2+}-binding residues D251, S253 and S255 in Peptide 1 (green ribbon), T328 (purple ribbon) and D364 in Peptide 2 (yellow ribbon) are shown as blue spheres. The N-terminal vWF-A residues K235-S243 (Figure 6.5) are attached to the homology model in an extended conformation (blue ribbon) up to the factor D cleavage site (arrowed). The C-terminal vWF-A residues I444-L452 up to the first Cys residue of the SP domain are attached as an extended conformation (blue ribbon), to which the SP domain is joined. Biantennary complex-type oligosaccharide chains are shown at N260 (near the Mg^{2+}-site) and N353 (near the N-terminus).

vWF-A domain then brings about the activation of the SP domain. It is interesting that the SP domain of factor B is attached to the α-helix A7 of the vWF-A domain (Figure 6.9), which is in topological proximity at the same α-helix face to the predicted inhibitory binding site at α-helices A2 and A3 of the A1 domain in von Willebrand factor (Perkins *et al.*, 1999). This may suggest a common functional property of both domains.

Chapter 7

Experimental:
Metal dependent conformational changes in the recombinant vWF-A domains.

7.1 Introduction

The active site of open twisted α/β domains is typically formed by a crevice outside the carboxy ends of the β-strands (Figure 7.1). This crevice is formed by two adjacent loop regions that connect the two strands with the α-helices on opposite sides of the β-sheet (Branden and Tooze, 1999). The vWF-A domain structure is an open twisted mostly-parallel β-sheet structure flanked by α-helices above and below the β-sheet, and possesses an active site cleft at the crossover point where the β-sheet winds towards opposite ends of the structure (Edwards and Perkins, 1996). This secondary structure of the vWF-A domain resembles, but is not identical with the dinucleotide fold (Figure 7.2) (Perkins *et al.*, 1999). The metal ion dependent adhesion site (MIDAS) of the integrin vWF-A domains coordinates the divalent cation through two Ser residues within a conserved amino acid sequence motif Asp-X-Ser-X-Ser (DXSXS) located in a crevice at the C terminal of the β–strand (βa), a Thr from the α3-α4 loop located on the other side of the crevice, and an Asp from the βd-α5 loop (Figures 7.2 and 7.3) (Qu and Leahy, 1995). These residues are highly conserved in the vWF-A protein fold family and in factor B sequence it is likely that coordination of the metal ion is similar to that of the integrins.

There are presently 23 sets of vWF-A coordinates for 14 vWF-A crystal structures for five different vWF-A proteins. Metal binding sites were observed in the integrin vWF-A structures of CR3, LFA-1 and VLA-2, but not in the A1 and A3 domains of von Willebrand factor (Lee *et al.*, 1995a, 1995b; Baldwin *et al.*, 1998; Qu and Leahy, 1995, 1996; Emsley *et al.*, 1997, 1998; Celikel *et al.*, 1998; Huizinga *et al.*, 1997; Bienkowska *et al.*, 1997). The sequence alignment of the vWF-A superfamily (Perkins *et al.*, 1994) shows that the metal coordination site corresponds to conserved Asp and Ser residues within a DxSxS motif on one side of the active site cleft, together with conserved Thr and Asp residues on the other side of the cleft.

Crystal structures of two conformations of the CR3 vWF-A domain have been determined. It has been proposed that the vWF-A domain exists in either "on" or "off" states, and that this domain functions to transmit signals of ligand binding through its structure

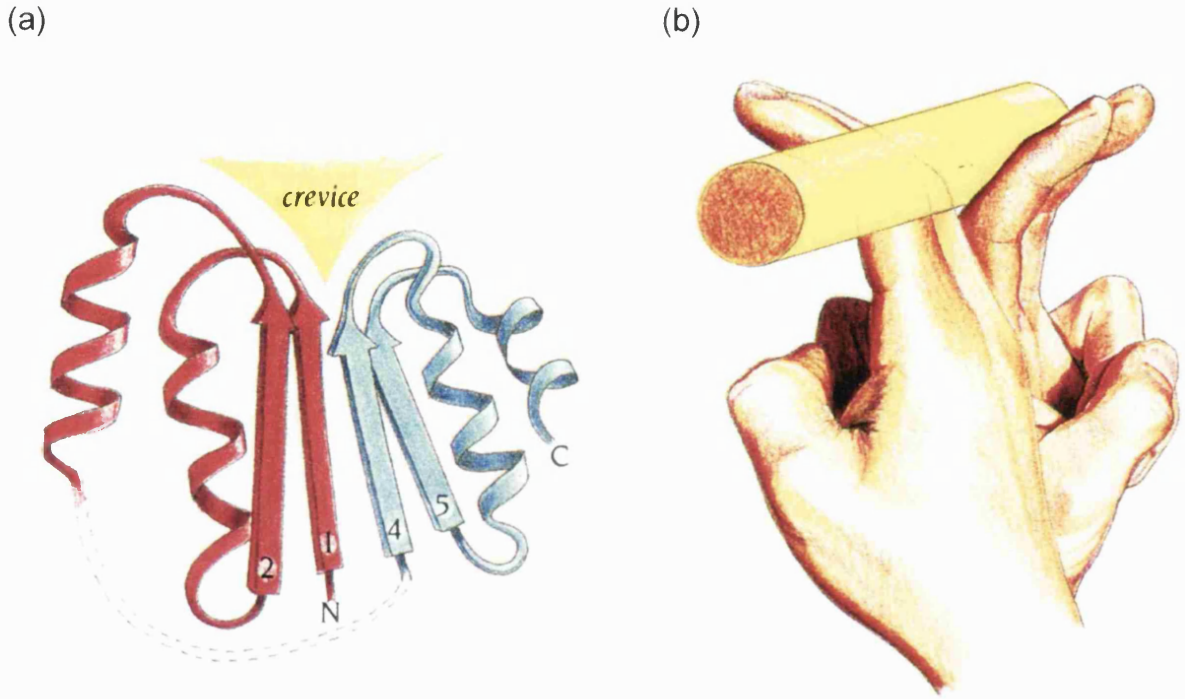


Figure 7.1 (a) The active site in open twisted α/β domains is in a crevice outside the carboxy ends of the β-strands. This crevice is formed by two adjacent loop regions that connect the two strands with the α-helices on opposite sides of the β-sheet. This is illustrated by the curled fingers of the two hands (b) where the top halves of the fingers represent loop regions and the bottom halves represent the β-strands. The rod represents a bound molecule lying in the binding crevice (Adapted from Branden and Tooze, 1999).

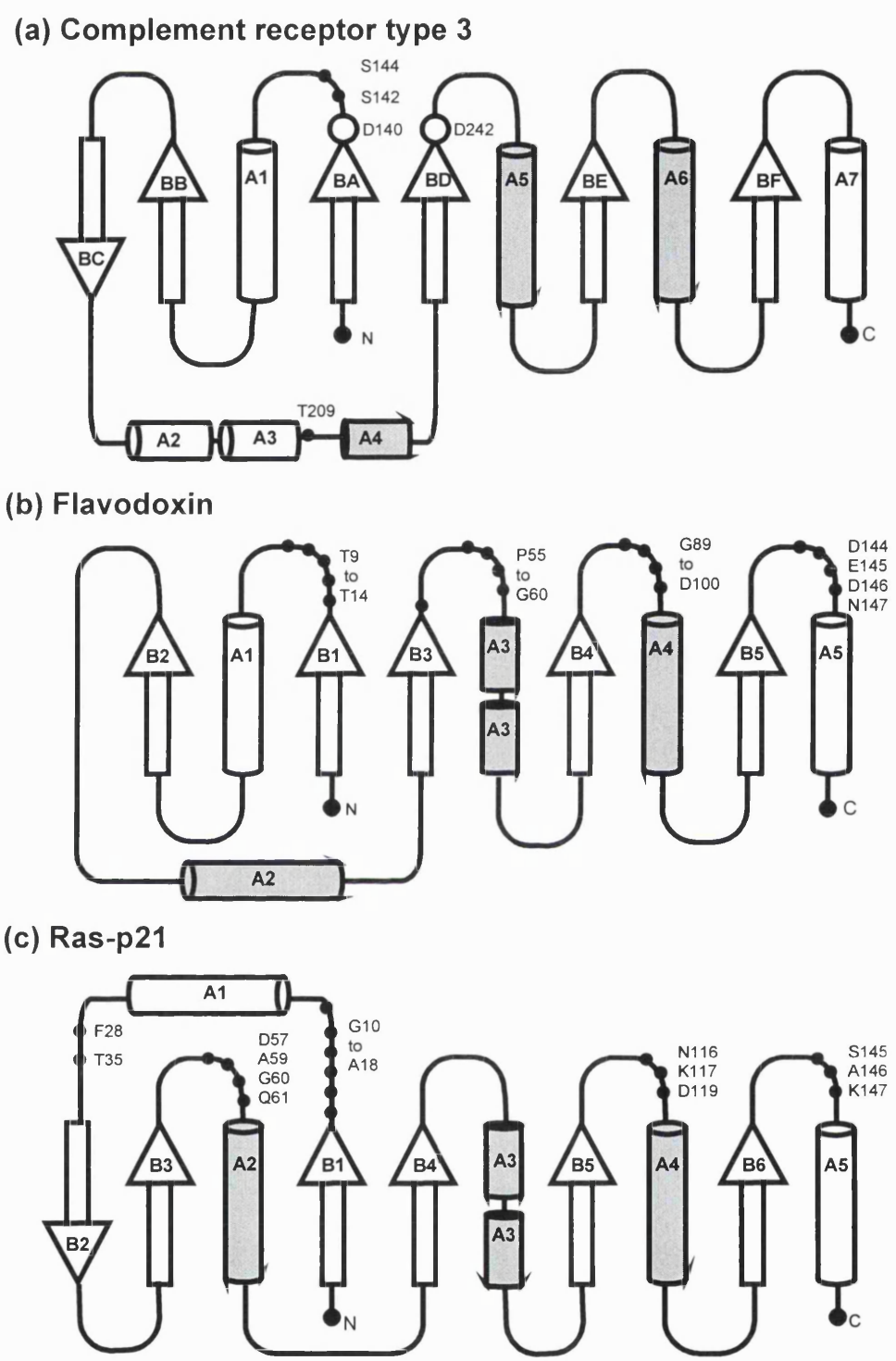


Figure 7.2 Supersecondary structure topologies for doubly-wound open α/β folds. α-helices are represented as cylinders and β-strands are shown as arrows. The labelling follows the scheme in previous publications for each protein. The N- and C-termini in the three structures are all located at topologically equivalent positions. The α-helices above the β-sheet are shaded, while those below the β-sheet are unshaded. (a) Five residues involved in the Mg^{2+} binding site of CR3 are highlighted (D140 and D242 (O); S142, S144 and T209 (●)). (b) Residues on four carboxyl-edge loops that constitute the flavin mononucleotide binding site in flavodoxin are highlighted by (●). (c) Residues on five carboxyl-edge loops that constitute the GTP-binding site in ras-p21 are highlighted by (●).

(Lee *et al.*, 1995a and b; Cooney and Ginsberg, 1996; Jenkins *et al.*, 1998), although this is controversial (Baldwin et al., 1998; Liddington and Bankston, 1998). The effect of metal binding to the vWF-A domain of integrins is presently unclear, since the Mg^{2+} binding site in the vWF-A crystal structure for CR3 differs from those in other integrin vWF-A crystal structures, namely leucocyte-function associated antigen-1 (LFA-1; CD11a), very late activation protein-2 (VLA-2; platelet glycoprotein Ia; CD49a) and other CR3 structures (Figure 7.3). Consequently it is not known whether the integrin vWF-A structure is affected by metal binding, and this is currently an area of discussion (Lee *et al.*, 1995a; Baldwin *et al.*, 1998; Liddington and Bankston, 1998). Binding of the metal ion is considered critical for the adhesive function of the integrins. It may be required to stabilise the active domain conformation or it may participate in interactions with the ligand, possibly through interactions with an acidic residue (Qu and Leahy, 1996). Further uncertainty in the role of the Mg^{2+} in the vWF-A domain fold arises from the absence of a Mg^{2+} site in the vWF-A1 and vWF-A3 domains of von Willebrand factor even though the MIDAS motif in the A3 domain contains only one non-conservative change of Asp to Thr (Celikel *et al.*, 1998; Emsley *et al.*, 1998; Jenkins *et al.*, 1998; Huizinga *et al.*, 1997; Bienkowska *et al.*, 1997). This lack of metal binding is due to the unusual conformation of the Phe residue that follows the DXSXS motif (Bienkowska *et al.*, 1997). There is also uncertainty about the role of the Mg^{2+} ion in the vWF-A domain of factor B. It would appear that the metal ion is not absolutely required for the assembly of a functional C3bBb, C3 convertase, but it does have an enhancing role which appears to be more important in the case of the surface bound enzyme (Pryzdial and Isenman, 1986). The explanation given for this is that C3b in the fluid phase would be able to accommodate the fit of factor B, whereas surface bound C3b would have restricted flexibility making strong Mg^{2+} independent binding no longer possible. Thus the Mg^{2+} ion was necessary for a stable surface bound C3bBb complex.

A clarification of the structural role of bound Mg^{2+} to the vWF-A domain is essential for an understanding of its metal-binding properties. Given the large difference reported between the metal-bound and metal-free forms of vWF-A by crystallography, it is surprising that initial circular dichroism (CD) and Fourier transform infrared (FT-IR)

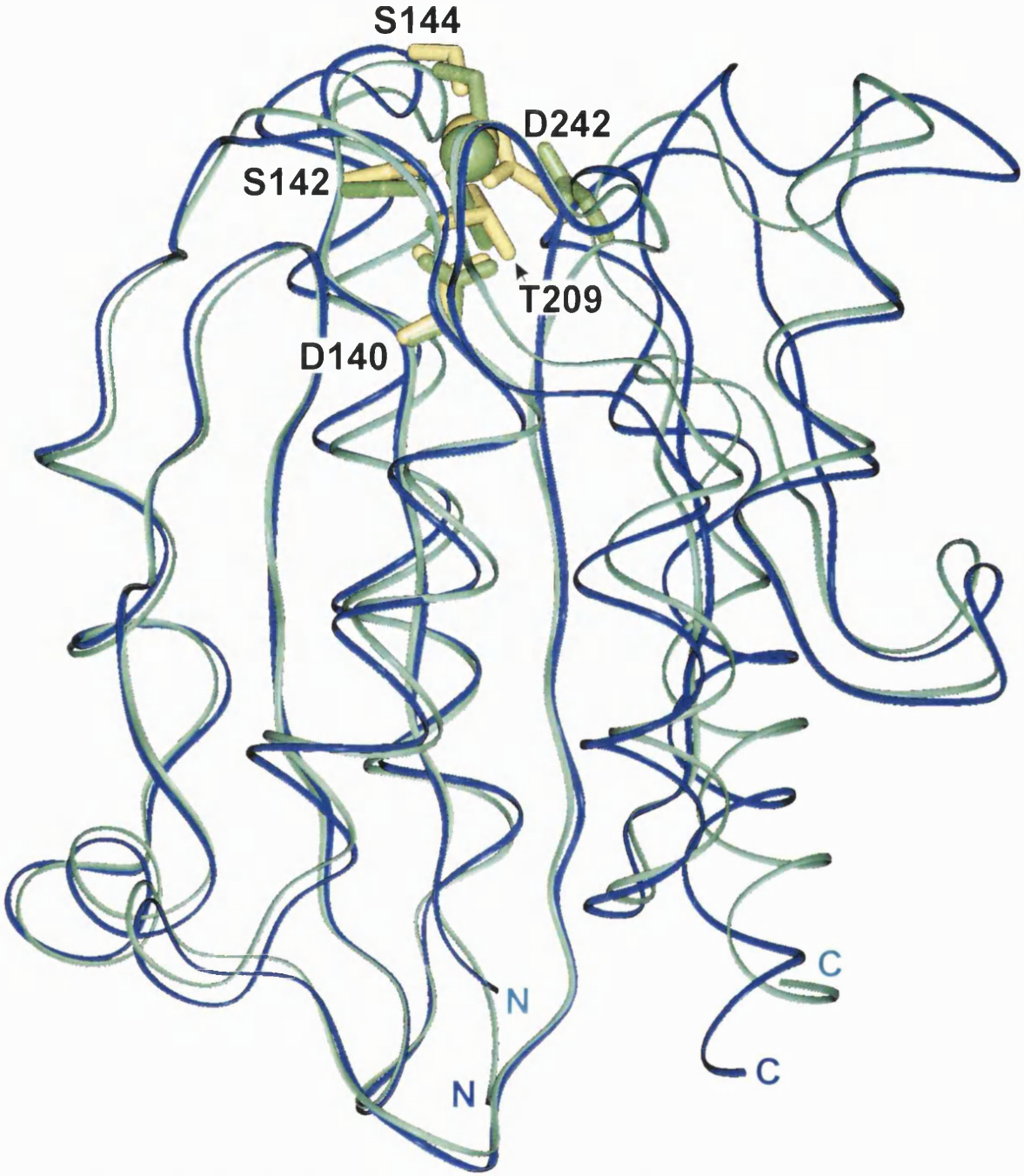


Figure 7.3 Conformational change seen between two superimposed crystal structures for complement receptor type 3 (CR3) in the presence of Mg^{2+}. The central six-stranded β-sheet is viewed face-on. The structure corresponding to the PDB code 1ido is shown as a light blue ribbon with the active site residues depicted using green sidechains and metal atom (Table 7.1). That corresponding to the PDB code 1bho is shown in dark blue with yellow sidechains and metal atom. A large movement in Asp242 and the α-helices relative to an almost invariant β-sheet is seen between the two forms.

spectroscopic studies of vWF-A domains have been unable to detect any significant conformational rearrangement upon metal binding (Perkins *et al.*, 1994; Fairbanks *et al.*, 1995; Nolte *et al.*, 1999). This Chapter describes the application of a multidisciplinary spectroscopic approach to a recombinant vWF-A domain from factor B in the presence of excess metal or EDTA. The two recombinant vWF-A domains described earlier were used, one based on the factor B vWF-A sequence Gly229-Ile444, and a more stable one based on the sequence Gly229-Gln448 (Chapter 4) (Williams *et al.*, 1999). A homology model for the vWF-A domain enabled the data to be interpreted (Chapter 6) (Hinshelwood *et al.*, 1999). The first unequivocal evidence for a significant conformational change between the metal-bound and metal-free forms of a vWF-A domain is presented here, and the functional implications of this for factor B are then discussed.

7.2 Materials and Methods

7.2.1 Molecular graphics analyses of vWF-A domains

The metal binding site in the vWF-A structure was analysed using 23 sets of coordinates from the Protein Data Bank for 14 crystal structures for five vWF-A domains in the following proteins, namely the three integrins CR3 (CD11b), leucocyte function-associated antigen-1 (LFA-1; CD11a), and very late activation protein-2 (VLA-2; platelet glycoprotein Ia; CD49b), as well as the A1 and A3 domains of von Willebrand factor (Lee *et al.*, 1995a, 1995b; Baldwin *et al.*, 1998; Qu and Leahy, 1995, 1996; Emsley *et al.*, 1997, 1998; Celikel *et al.*, 1998; Huizinga *et al.*, 1997; Bienkowska *et al.*, 1997). The Protein Data Bank codes are identified in Table 7.1. The structures were subjected to visual inspection on Silicon Graphics INDY Workstations using INSIGHT II 97.0 software (Biosym/MSI, San Diego, U.S.A.) in conjunction with Crystal Eyes stereo glasses and the superfamily alignment of 75 vWF-A sequences in Perkins *et al.* (1994). The consensus secondary structure was identified using DSSP (Kabsch and Sander, 1983) and taken as the average across the crystal structures (Hinshelwood *et al.*, 1999). The construction of the homology model was based on the CR3 vWF-A domain (Chapter 6) (PDB code 1ido; Lee *et al.*, 1995a; Hinshelwood *et al.*, 1999).

7.2.2 Preparations of factor B, its Bb fragment and the recombinant vWF-A domains

Native factor B, the Ba and Bb fragments, C3 and the recombinant vWF-A-218 and vWF-A-222 domains were prepared as described previously (Chapter 4: sections 4.3.6, 4.4.14 and 4.4.15). The vWF-A-218 contained the factor B sequence G229-I444 and was preceded by the residues GS derived from the thrombin recognition site LVPRGS. The vWF-A-222 sequence contained the factor B sequence G229-Q448, but differed in that a Cys267Ser mutation was made in order to eliminate a free Cys residue, as well as containing the N-terminal residues GS. All the domains were concentrated under pressure on an Amicon YM10 membrane (mass cut-off of 10 kDa) and purified by gel filtration using a Pharmacia FPLC system with a HiLoad 16/60 Superdex 75 column.

Table 7.1 Survey of distances between the metal ion and its coordinating atoms in homologous vWF-A crystal structures

vWF-A[+] (PDB code)	Metal	Coordination residue[++] DxSxS			Thr	Asp
CR3 (CD11b)		Asp140	Ser142	Ser144	Thr209	Asp242
1ido	Mg^{2+}	4.2, 4.2	2.2	2.2	2.1	4.2, 4.2
1jlm	Mn^{2+}	4.0, 4.4	2.3	2.1	4.0	2.3
1bho-1	Mg^{2+}	4.0, 4.1	2.0	2.0	4.2	2.1
1bho-2	Mg^{2+}	3.9, 4.6	2.3	1.7	4.6	2.1
1bhq-1	Cd^{2+}	3.9, 4.4	2.3	2.6	3.8	2.3
1bhq-2	Cd^{2+}	4.1, 4.3	2.2	2.2	3.8	1.9
LFA-1 (CD11a)		Asp137	Ser139	Ser141	Thr206	Asp239
1zoo-1	Mg^{2+}	3.9, 4.0	2.1	2.1	4.2	2.0
1zoo-2	Mg^{2+}	4.0, 4.1	2.1	2.1	4.2	2.1
1zop-1	Mn^{2+}	3.9, 4.0	2.1	2.1	4.2	2.0
1zop-2	Mn^{2+}	4.0, 4.1	2.1	2.1	4.0	2.1
1lfa-1	Mn^{2+}	4.0, 4.0	2.1	2.1	4.0	2.1
1lfa-2	Mn^{2+}	4.0, 4.0	2.1	2.1	4.0	2.1
VLA-2 (CD49b)		Asp151	Ser153	Ser155	Thr221	Asp254
1aox-1	Mg^{2+}	4.1, 4.0	2.2	2.0	4.2	2.1
1aox-2	Mg^{2+}	4.1, 4.0	2.3	2.0	4.2	2.0

[+] Where two molecules appear in the unit cell, they are described by a suffix of -1 or -2.

[++] The crystallographic resolutions range between 1.7 Å to 3.0 Å. Distances are in Å. The ligation site is to either the OD1 and/or OD2 atoms in D residues (via a water bridge in the case of the first D residue and the T residue in 13 of the 14 structures), the OG in S residues, and the OG1 atom in T residues. In the LFA-1 structures, chloride ions and/or water molecules form additional ligands with metal (not shown).

7.2.3 Circular dichroism spectroscopy

Circular dichroism (CD) spectroscopy was performed at 20°C using a Jobin-Yvon CD6 spectropolarimeter with quartz cells of path length 0.2 mm and 0.1 mm. The instrument was calibrated with an aqueous solution of recrystallised D10-camphosulphonic acid ($\theta^{0.1\%,\ 1cm} = 0.308$ at 290 nm). The vWF-A-222 domain, the Bb fragment and native factor B were dialysed into 5 mM Tris-HCl, 0.5 mM $MgCl_2$, pH 7.5, while the vWF-A-218 sample was dialysed into 100 mM $NaClO_4$, 5 mM Tris-HCl, 0.5 mM $MgCl_2$, pH 7.5 because of its lower solubility. Sample concentrations used to derive the differential molar extinction coefficient $\Delta\varepsilon$ were calculated from the absorbance at 280 nm (Drake, 1994) and were in the range of 0.2 to 1.5 mg/ml. The absorption coefficients A_{280} (1cm, 1%) were calculated from the total of Trp, Tyr and Cys residues in the sequence to be 15.6 for vWF-A-218, 15.2 for vWF-A-222, 12.5 for the Bb fragment and 13.8 for factor B (Perkins, 1986). Experiments were performed in the presence of $MgCl_2$ to which EDTA was subsequently added to investigate the effect of Mg^{2+} ions. Alternatively, the CD spectra were recorded in the absence and presence of Mg^{2+} ions without the use of EDTA. Spectral quantification to obtain secondary structure contents was performed using CONTIN (Provencher, 1982). In each case the background spectrum from the sample buffer was subtracted from the protein sample spectrum.

7.2.4 Fourier transform infrared (FT-IR) spectroscopy

Fourier transform infrared (FT-IR) spectroscopy was performed using the vWF-A-222 domain using a 1750 Perkin-Elmer FT-IR spectrometer continuously purged with N_2 gas to reduce water vapour absorption in the spectral region of interest (Haris and Chapman, 1994). The sample was contained between CaF_2 windows fitted with a Teflon spacer to give a path length of 50 μm in a thermostatically-controlled cell housing. vWF-A samples at 5 mg/ml were in PBS buffer in 99.8% ^2H_2O with either 5 mM $MgCl_2$ or 5 mM EDTA. FT-IR spectra were obtained from both the protein sample and the buffer at each temperature. 1000 scans were recorded at 20°C and 100 scans were recorded for each temperature increment thereafter at a resolution of 4 cm^{-1}. The scans were signal averaged and the absorbance spectrum was obtained by digital subtraction of the ^2H_2O buffer spectrum from the sample

spectrum (Haris *et al.*, 1986). Detailed analysis of the spectra was carried out using a second derivative procedure provided by the GRAMS software (Perkin-Elmer) with 13 data point Savitzy-Golay smoothing.

7.2.5 ^{1}H NMR spectroscopy

Each protein preparation was dialysed into PBS in 99.8% ^2H_2O with 1.5 mM $MgCl_2$ over 36 h with 4 buffer changes at 8°C, except for the vWF-A-222 studies for which Mg-free PBS was used. NMR spectra were recorded at 30°C using 500 MHz and 600 MHz Varian spectrometers and analysed using Varian VNMR software at the MRC NMR Centre at the National Institute of Medical Research, Mill Hill. Sample concentrations were in the range 2-10 mg/ml. pH meter readings in ^2H_2O were uncorrected values measured on a Corning 240 pH meter, and additions of stock solutions of NaO^2H or ^2HCl were used to adjust the pH. The H^2HO chemical shift was referenced with respect to the chemical shift of trimethylsilylpropionic acid (1 mg/ml) in the same buffer and temperature range as the protein samples, and the protein NMR spectra were then referenced with respect to the H^2HO signal. A line broadening of 1 Hz was applied before Fourier transformation. The effect of Mg^{2+} was investigated by adding up to 5 mM EDTA to some of the preparations before measuring the spectra. Alternatively spectra were recorded in the absence and presence of Mg^{2+} ions.

7.2.6 Calculation of ring current shifts

In the proton NMR spectra of proteins, resonances upfield of 0.5 ppm correspond to ring current shifts of methyl groups and single protons of amino acid side-chains positioned close to one or several aromatic rings (Perkins, 1982). The ring current shifts are localised to within about 7 Å of the centre of the ring, and affect only a small proportion of proton signals. The program RCCAL (Perkins, 1982) was used on a Silicon Graphics INDY Workstation to calculate the ring current shifts from protein coordinates in Protein Data Bank format. The appearance of the NMR spectra were then predicted by correction of the random coil chemical shifts of the free amino acid residues by these calculated ring current shifts (Bundi and Wüthrich, 1978). The validity of these calculations has been established in calibration studies (Perkins, 1982).

7.3 Results and Discussion

7.3.1 The metal binding site in the vWF-A superfamily

The sequence alignment of the vWF-A superfamily (Perkins *et al.*, 1994) showed that the metal coordination site corresponded to a conserved DxSxS motif on one side of the active site cleft, together with conserved Thr and Asp residues on the other side of the cleft. The metal binding sites in 14 sets of CR3, LFA-1 and VLA-2 coordinates were examined by molecular graphics (Table 7.1). The molecular coordination to Mg^{2+} or to other divalent cations such as Mn^{2+} or Cd^{2+} differed only between the first-determined CR3-Mg^{2+} structure (PDB code 1ido) and the remaining 13 vWF-A-metal structures (Table 7.1). The metal coordination with DxSxS is the same in all 14 structures, and involves a water molecule bridge between the metal and a bidentate linkage with an Asp carboxyl group (distances of 3.9 Å to 4.3 Å), and two direct metal bonds with Ser hydroxyl groups (distances of 1.7 Å to 2.3 Å). In distinction to this, that with the conserved Thr and Asp residues involves a shift of the water molecule bridge from Asp in the first vWF-A structure (Lee *et al.*, 1995a; PDB code 1ido) to Thr in the remaining 13 structures. In the latter, the Thr hydroxyl-metal distance ranges between 3.8 Å to 4.6 Å, and the Asp carboxyl-metal distance ranges between 1.9 Å to 2.3 Å. In the two vWF-A structures that did not bind metal, three (A1 domain) or two (A3 domain) residues in the DxSxS-D-T motif of the active site cleft were not conserved, thus explaining the lack of metal binding (Perkins *et al.*, 1994; Jenkins *et al.*, 1998). These comparisons show that two types of metal-protein structures occur in the vWF-A domain, one found in 13 of the 14 available metal-bound structures, and the other involving a structural change at one side of the active site cleft, which is hereinafter termed the conformationally-flexible side of this cleft. In the light of the recent controversy over CR3 integrin-metal structures (Baldwin *et al.*, 1998; Liddington and Bankston, 1998), in which it was unclear if metal affected the vWF-A structure, a reinvestigation of the binding of Mg^{2+} to the vWF-A domain by independent approaches based on solution structure methods was carried out.

7.3.2 Circular dichroism studies of factor B and its fragments

Circular dichroism (CD) spectroscopy is a monitor of the overall protein secondary structure and is sensitive to conformational changes (Drake, 1994). For CD studies, the

recombinant vWF-A-218 and vWF-A-222 domains, the Bb fragment and factor B were prepared, each of which was eluted as a single peak from the HiLoad 16/60 Superdex-75 gel filtration column. Each migrated as a single band on SDS-PAGE under both reducing and non-reducing conditions. From these, the approximate masses of these proteins were estimated as: vWF-A-218 and vWF-A-222, 25 kDa; Bb fragment, 60 kDa; factor B, 90 kDa, all of which were as expected from their sequences. All the preparations were recognised in Western blots using a polyclonal antiserum to factor B, and were determined to be monomeric by mass spectrometry or analytical ultracentrifugation (Chapter 5) (Williams *et al.*, 1999; Hinshelwood *et al.*, 1999). Both the vWF-A-218 and vWF-A-222 domains were functionally active (Williams and Sim, 1994; Williams *et al.*, 1999). The vWF-A-222 domain was more soluble than the vWF-A-218 domain, and was therefore better suited for CD studies (Chapter 4).

The secondary structures of the above four proteins were examined using their CD spectra. Each CD spectrum showed a large negative ellipticity between 210 nm and 220 nm, with a small trough between these wavelengths, as expected for proteins with α-helix and β-sheet contents (Drake, 1994). The vWF-A-218 and vWF-A-222 spectra in the presence of 0.5 mM $MgCl_2$ were very similar (Figure 7.4), and indicated that these two recombinant vWF-A domains had similar folded structures. They were readily distinguishable from the factor B and Bb spectra by showing a higher negative ellipticity. The latter is consistent with the highest proportion of α-helix in this domain when this is compared to the remaining domains in factor B which possess predominantly β-sheet structures. On the addition of 1 mM EDTA to the vWF-A-218 or vWF-A-222 domains, a significant reduction of 15-24% in the ellipticity at 215 nm from approximately -3.5 M^{-1} cm^{-1} to -2.9 M^{-1} cm^{-1} was observed (Figure 7.5). Similar experiments on the Bb fragment in the presence of Mg^{2+} and EDTA revealed a 4% reduction of the ellipticity when EDTA was added. The direction of this change is the same as that seen for the recombinant vWF-A domains. No change was reproducibly detected within error for factor B under these conditions. These results were obtained in three CD sessions with up to three independent preparations. The same outcome was also obtained in the reversed experiment in which the CD spectrum of the vWF-A-222 domain prepared

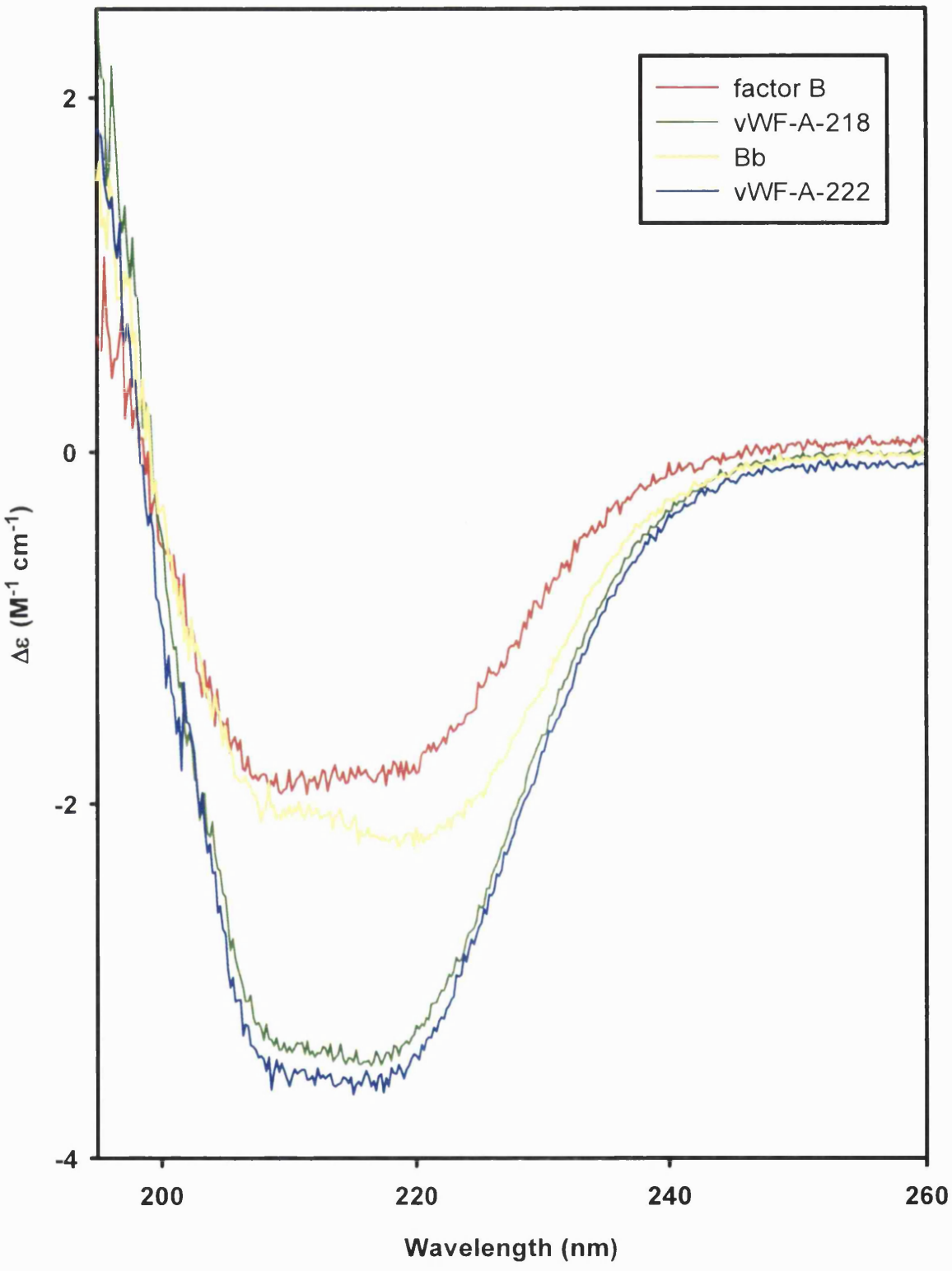


Figure 7.4 CD spectra of the factor B, Bb fragment, vWF-A-218 and vWF-A-222 domain preparations in 5 mM Tris-HCl, 0.5 mM $MgCl_2$, pH 7.5 at 20°C. Sample concentrations were in the range 0.2 to 1.5 mg/ml. Spectral quantification to obtain secondary structures was performed using CONTIN (Provencher, 1983).

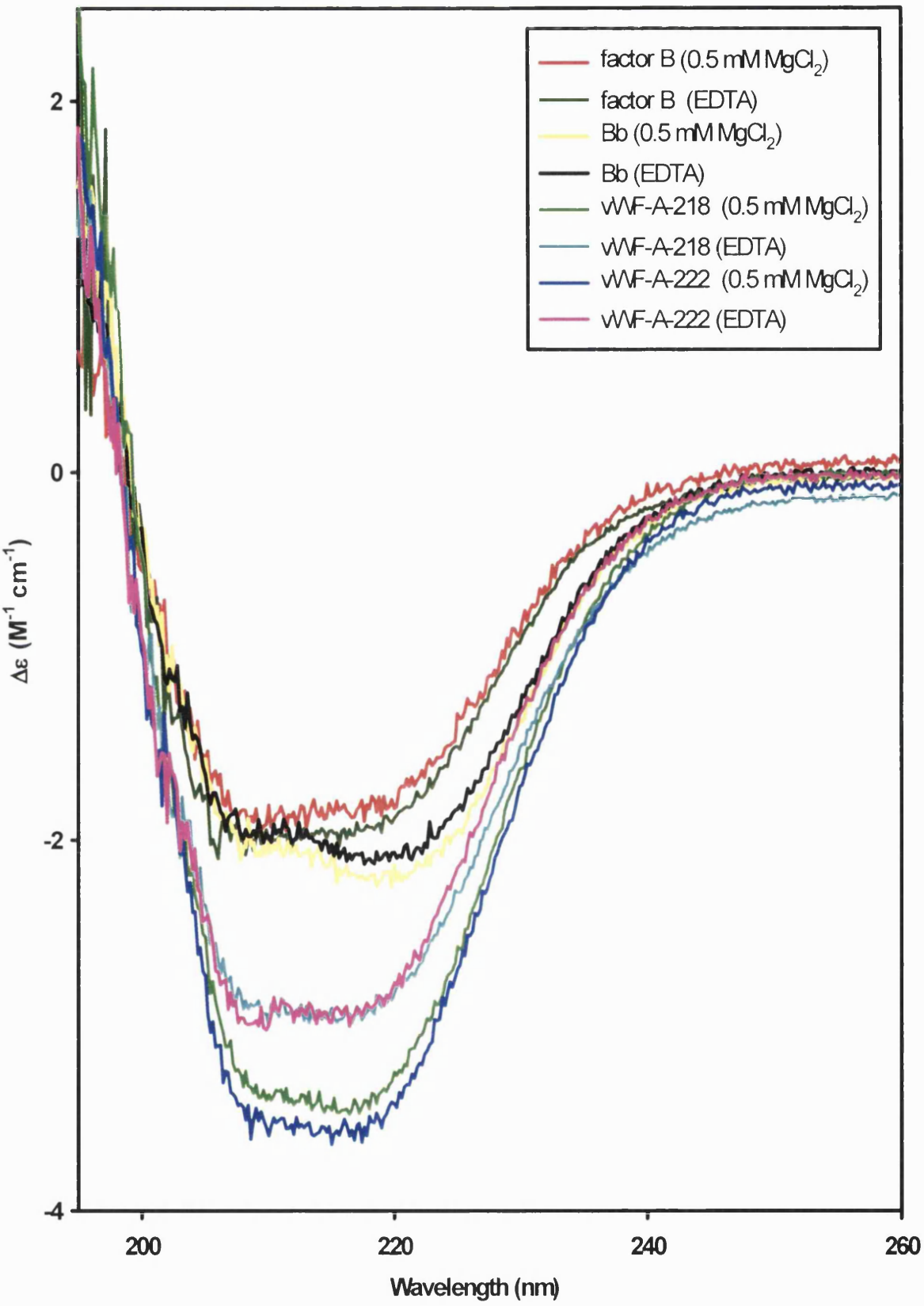


Figure 7.5 CD spectra of the factor B, Bb fragment, vWF-A-218 and vWF-A-222 domain preparations in 5 mM Tris-HCl, 0.5 mM $MgCl_2$, pH 7.5 at 20°C before and after the addition of 1 mM EDTA. Sample concentrations were in the range 0.2 to 1.5 mg/ml. Spectral quantification to obtain secondary structures was performed using CONTIN (Provencher, 1983).

in the absence of Mg^{2+} ions was compared with the CD spectrum of the same sample containing 2 mM $MgCl_2$. In control experiments, no changes occurred in the CD spectrum of the vWF-A-222 sample when this was titrated with $CaCl_2$ instead of $MgCl_2$.

The observed secondary structures for each of the metal-bound and metal-free preparations were quantified from the CD spectra using CONTIN (Provencher, 1982). The quantification was not sufficiently accurate to identify metal-induced changes in secondary structure contents, but should give good estimates of the overall content (Drake, 1994). The CD analysis showed approximately 26% α-helix and 31% β-sheet in the vWF-A domain, in accord with the estimate of approximately 31% α-helix and 36% β-sheet for this domain by Fourier transform infrared spectroscopy (Perkins *et al.*, 1994). In comparison, similar CD analyses gave 45% α-helix and 7% β-sheet for the vWF-A domain of CR3 (Fairbanks *et al.*, 1995), and 38% α-helix and 14% β-sheet for the eight vWF-A domains in the α3 chain of collagen type VI (Specks *et al.*, 1992). The consensus secondary structure content calculated from the mean of the 23 vWF-A crystal coordinates available to date gives 36% α-helix and 19% β-sheet (Hinshelwood *et al.*, 1999). These comparisons show that the CD data are consistent with the presence of both α-helix and β-sheet in vWF-A domains, and that both the vWF-A-218 and vWF-A-222 domains possess folded structures as expected.

The changes in CD ellipticity in Figure 7.5 are best explained by noting that the SP domain is common to both the Bb fragment and factor B, but is absent from the vWF-A samples. The presence of the SP and SCR domains, which contain dominantly β-sheet structures, will reduce the proportion of α-helix by a factor of about 2 for the Bb fragment and about 3 for intact factor B when compared to the vWF-A domain, and will result in smaller Mg^{2+}-induced effects as observed. The reduced effect of Mg^{2+} on the CD spectra of the Bb fragment and factor B may also be affected by contacts between the SP domain and/or the Ba fragment with the vWF-A domain as appropriate. Since α-helices have large effects on CD spectra, the large increase in negative ellipticity for the recombinant vWF-A domain in the presence of Mg^{2+} implies that the bound metal has facilitated the stabilisation of a small amount of ordered α-helix. In this context, it is of interest that the occurrence of α-helices is

variable in crystal structures for the vWF-A superfamily (Figure 6.5 of Chapter 6 and Figure 3 of Hinshelwood *et al.*, 1999), even within different crystal forms of the same molecule (Lee *et al.*, 1995a, 1995b). For example, α-helix A5 becomes a short 3_{10}-helix in the LFA-1 (CD11a) structure, unlike the case in the other vWF-A structures. In the case of factor B, the vWF-A homology model showed that significant insertions were predicted just before and after the α-helices A4 and A5, the effect of which was to deepen the conformationally-flexible side of the metal-binding active site cleft (Figure 7.6). It was inferred that the changes in the vWF-A CD spectra of Figure 7.5 may be best explained by a small increase in the ordered α-helix content in the vicinity of the metal-binding cleft upon the binding of Mg^{2+}, although a contribution from more ordered β-sheet structures cannot be ruled out.

7.3.3 Mg^{2+}-dependence of the vWF-A structure by FTIR spectroscopy

FT-IR spectroscopy monitors the protein secondary structure through the separate observation of signals for α-helix, β-strand and other secondary structures (Haris and Chapman, 1994). In the absorbance spectrum of the vWF-A-222 domain in a ^2H_2O buffer with 5 mM $MgCl_2$, the amide I band appears as a broad signal centered at 1641-1642 cm^{-1} that corresponds principally to mainchain C=O stretching motions (Figure 7.7a). After calculation of the second derivative spectrum of the vWF-A-222 domain, two major bands at 1650 cm^{-1} and 1636 cm^{-1} were well resolved (Figure 7.7b). These two bands were assigned to α-helix and β-sheet structures respectively. There were also small sub-components of the amide I band at 1659 cm^{-1}, 1670 cm^{-1} and 1692 cm^{-1}. The band at 1659 cm^{-1} may be attributed to the presence of either 3_{10}-helix or β-turns. The bands at1670 cm^{-1} and1692 cm^{-1} are likely to correspond to β-turn or anti-parallel β-sheet structures (Haris *et al.*, 1986; Surewicz *et al.*, 1993). These observations agree with the predicted structure of the factor B vWF-A domain and its homology model (Perkins *et al.*, 1994; Edwards and Perkins, 1995, 1996; Hinshelwood *et al.*, 1999). They also agree with the previous FT-IR study of the vWF-A-218 domain in buffers based on both H_2O and ^2H_2O solvents, in which the two major bands were assigned to solvent-exposed α-helices and buried β-strands (Perkins *et al.*,1994).

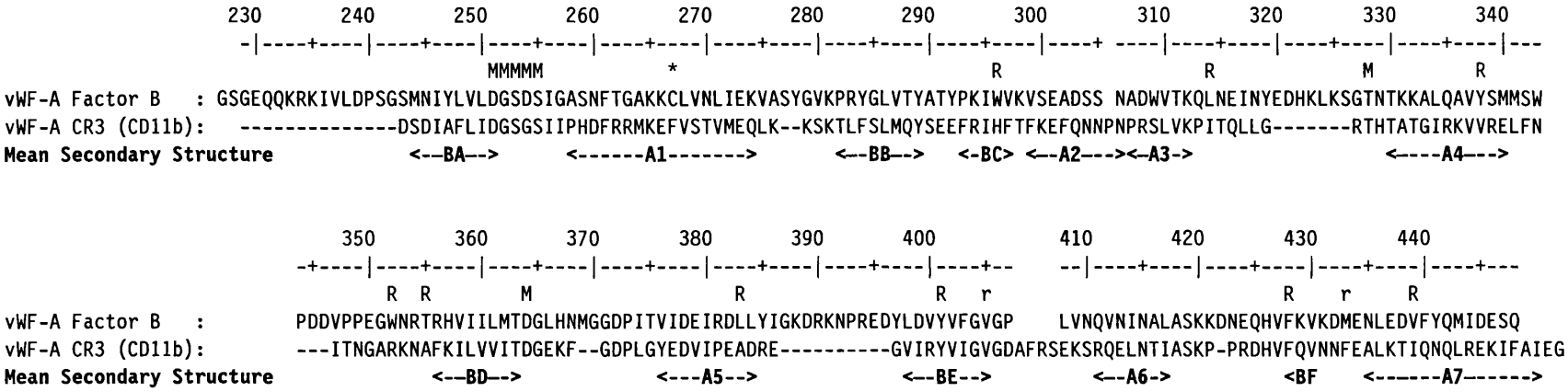


Figure 7.6 Sequence and numbering of the vWF-A domain in factor B. This is aligned with the sequence of CR3 and the mean secondary structure of the vWF-A superfamily (Hinshelwood *et al.*, 1999). The metal binding motif DxSxS-D-T is indicated by M symbols above the sequences. Residues potentially involved in methyl ring current shift interactions are likewise denoted by R for factor B and r for CR3. The position of Cys267 is asterisked.

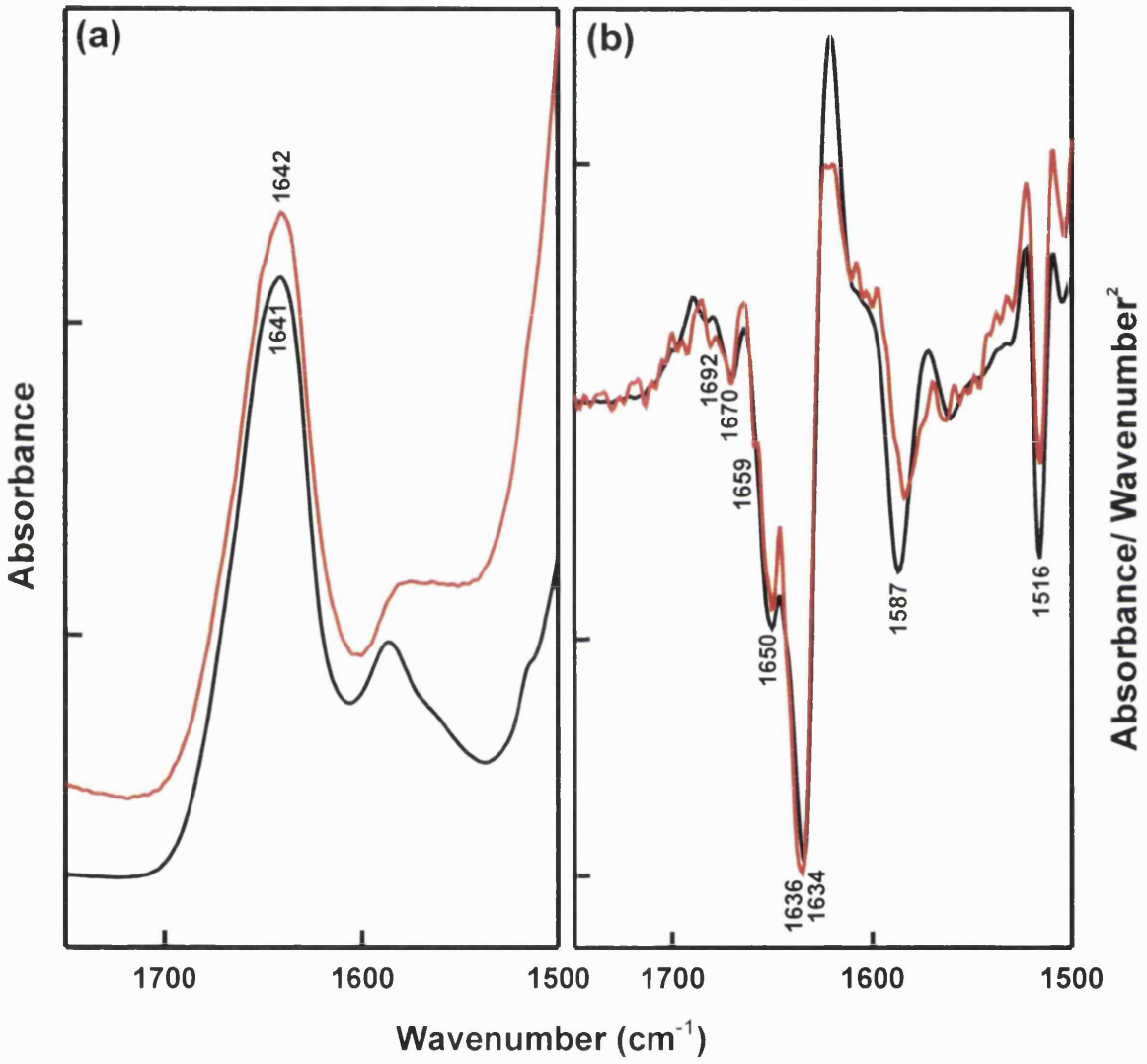


Figure 7.7 FT-IR spectroscopy of the amide I band of vWF-A-222 domain studied in PBS in ^2H_2O with either 5 mM $MgCl_2$ (red) or 5 mM EDTA (black). The concentration of the vWF-A-222 sample was 5 mg/ml for both samples. The absorbance spectra are shown in (a) and the second derivative spectra are shown in (b). Peak positions discussed in the text are identified by wavenumber labels corresponding to their band frequencies in the spectra.

In the present study, the greater stability of the vWF-A-222 domain compared to the vWF-A-218 domain permitted the effect of Mg^{2+} ions on its FT-IR spectrum to be studied in more detail. FT-IR spectra were recorded using ^2H_2O buffers both in the presence and absence of metal. The removal of Mg^{2+} caused the β-sheet band at1636 cm^{-1} to shift reproducibly to a slightly lower frequency of 1634 cm^{-1} with a minor reduction in its intensity (Figure 7.7b). This shift was attributable to a greater ^1H-^2H exchange of the mainchain amide protons of the metal-free vWF-A-222 domain as the result of a more flexible structure (Haris *et al.*, 1986, 1990; Jackson *et al.*, 1991; Perkins *et al.*, 1992). This explanation was based on observations of the extent of ^1H-^2H exchange using the intensity of the amide II band at both 1536 cm^{-1} and 1546 cm^{-1}, where the amide II band arises principally from N-H bending vibrations, which can be used to monitor ^1H-^2H exchange. These intensities were normalised by comparison with that of the well-resolved 1516 cm^{-1} Tyr band. The intensity of the residual amide II bands at both 1536 cm^{-1} and 1546 cm^{-1} were found to be approximately four times higher for the metal-bound vWF-A domain compared to that measured in the presence of EDTA after similar periods of dialysis into ^2H_2O buffers for both samples. In distinction to the changes seen in the CD spectrum, the α-helix band at 1650 cm^{-1} showed less dependence on the presence of Mg^{2+}, where no significant perturbation was observed on chelation of the metal (Figure 7.7b). The possibility that the observed spectral changes resulted from bands from EDTA was eliminated by reason of the use of (sample-buffer) difference spectra throughout and the lack of FT-IR bands from EDTA itself in the spectral range between 1630 cm^{-1} and 1700 cm^{-1}.

The stability of the Mg^{2+}-bound and Mg^{2+}-free forms of the vWF-A-222 domain was investigated using thermal denaturation experiments. The 1634-1636 cm^{-1} band (β-sheet) exhibited a sigmoidal dependence on temperature, where its negative intensity was much diminished above 50°C (Figure 7.8). The melting temperatures of the vWF-A domain were determined from the midpoints of the two graphs to be 46°C for the Mg^{2+}-bound form and 36°C for the Mg^{2+}-free form. Analysis of the negative intensity of the 1650 cm^{-1} band (α-helix) also demonstrated a sigmoidal temperature dependence with a similar denaturation temperature (data not shown). In contrast, a band at 1628 cm^{-1} that corresponded to random

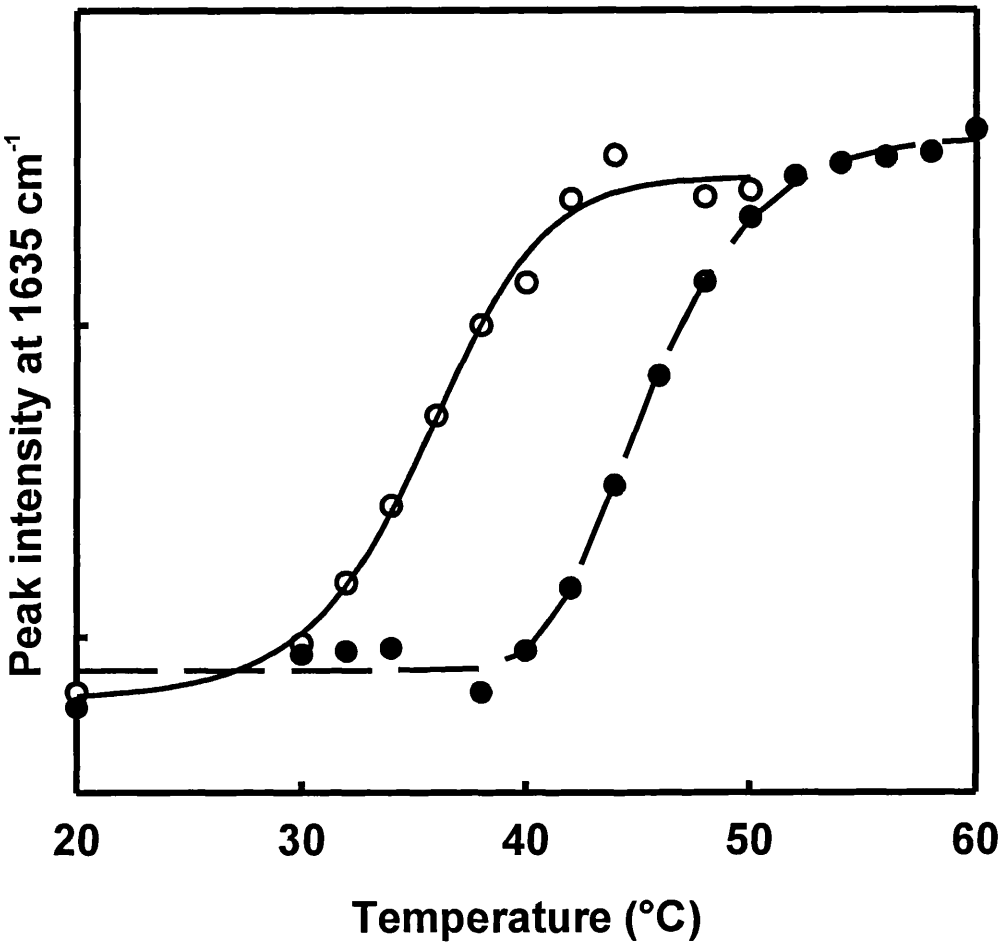


Figure 7.8 Dependence of the intensity of the second derivative band at 1635 cm^{-1} for the vWF-A-222 domain as a function of temperature in the presence of Mg^{2+} (●) and EDTA (○). The data points were fitted to a sigmoidal function represented by the continuous lines.

coil structures exhibited an increase in negative intensity with a reversed sigmoidal dependence as the temperature was increased. Thermal denaturation experiments that were also performed using CD spectroscopy confirmed that the Mg^{2+}-bound form was the more stable form (data not shown). It was concluded from both FT-IR and CD spectroscopy that the presence of Mg^{2+} significantly improved the stability of the vWF-A-222 domain.

7.3.4 Mg^{2+}-dependence of the vWF-A structure by ^1H NMR spectroscopy

^1H NMR spectra of both the vWF-A-218 and vWF-A-222 domains demonstrated significant NMR signal dispersion compared to the random coil spectra of the 20 amino acids (Figure 7.9). This meant that their folded structures in solution could be monitored by ^1H NMR. 500 MHz and 600 MHz ^1H NMR spectra were recorded at pH 7.50 at 30°C in the presence of 0.5 mM Mg^{2+} for both the vWF-A-218 and vWF-A-222 domains. Both domains presented two high field-shifted signals upfield of 0 ppm, one denoted as 2v at -0.29 ppm, and a broader one denoted as 1v. The 1v signal appeared at -0.93 ppm for the vWF-A-222 domain and -0.97 ppm for the vWF-A-218 domain (Figure 7.9a), and is approximately one-third the intensity of the 2v signal, so these were assigned as single-proton and single-methyl signals respectively. Exploratory 2D-NMR spectra to determine the spin systems to which these protons belonged were not productive. The appearance of two identical sets of several resonances immediately downfield of the residual water signal (not shown) was consistent with the presence of β-sheet structures in both vWF-A domains (Wishart *et al.*, 1992). The aromatic regions of the NMR spectra between 6 to 8 ppm for both vWF-A domains exhibited many overlapping signals, and in general both spectra were very similar (Figure 7.9b). However, at the left of Figure 7.9(b), an extra resonance at 7.86 ppm was present in the vWF-A-218 spectrum that was not present in that for vWF-A-222. The general similarity of both vWF-A NMR spectra was as expected from their sequences, where the two domains differ by the extension of the C-terminal α-helix A7 of the vWF-A-218 domain by the four hydrophilic residues DESQ in the vWF-A-222 domain and Cys267 was replaced by Ser267 in the vWF-A-222 domain (Williams *et al.*, 1999). If the packing of the α-helix A7 against the vWF-A domain was affected by the C-terminal sequence extension, this may cause localised structural perturbations that would be reflected in the two

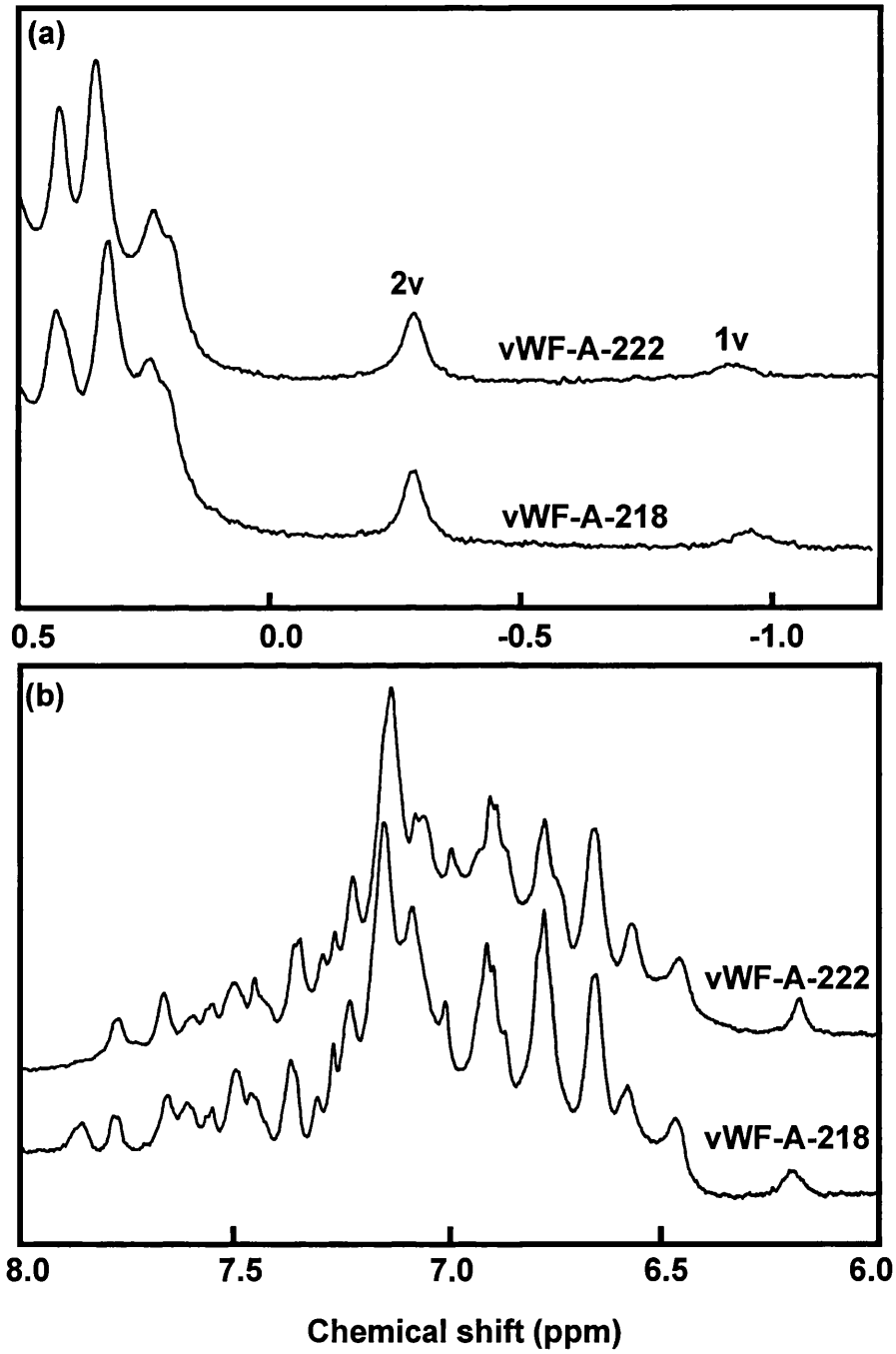


Figure 7.9 Comparison of the vWF-A-218 and vWF-A-222 domains by ^{1}H NMR spectroscopy. (a) and (b) show the upfield and the downfield spectral regions respectively of the vWF-A-218 and vWF-A-222 domains in the presence of 0.5 mM Mg^{2+}. The two highfield-shifted signals are labelled 2v and 1v in (a). A line broadening of 1 Hz was applied to the spectra before Fourier transformation.

minor NMR spectral differences involving the single-proton peak 1v and the extra aromatic signal at 7.86 ppm. It should be noted that Cys267 is on α-helix A1 (Figure 7.5), which is adjacent to α-helix A7, and its replacement may affect α-helix A7. Irrespective of how the shift changes were caused, the comparison of the vWF-A-218 and vWF-A-222 spectra suggested that the C-terminal α-helix A7 possessed a small degree of mobility within the vWF-A structure. This is consistent with the different conformations seen for α-helix A7 seen in different vWF-A crystal structures.

NMR titrations with Mg^{2+} and EDTA were performed to explore the effect of metal on the vWF-A structure (Figures 7.10a and 7.10b). When 1 mM EDTA was added to the vWF-A-218 domain, the high field signal 2v at -0.29 ppm was shifted downfield to -0.19 ppm, denoted as 2v' (not shown). The aromatic signals in the downfield region of the NMR spectrum also demonstrated minor perturbations on the addition of 1 mM EDTA (not shown). To avoid the effect of the strong ^1H NMR signals from EDTA on the spectrum, and to show that the spectral changes were not caused by interaction of EDTA with the vWF-A domain, the vWF-A-222 domain was prepared in the absence of Mg^{2+} ions, and the NMR spectrum was studied in the presence of $MgCl_2$ at concentrations between 0 mM to 5 mM (Figure 7.10a). The reverse chemical shift change was observed, in which signal 2v' (at -0.19 ppm) was initially observed in the absence of Mg^{2+}, then this was gradually replaced by signal 2v (at -0.29 ppm) on the addition of 0.1 mM and 0.2 mM Mg^{2+} (Figure 7.10a). The disappearance of signal 2v' occurred when the $MgCl_2$ concentration reached 0.5 mM, which was approximately double the vWF-A concentration of 0.16 mM. Other spectral changes occurred during the Mg^{2+} titration, as indicated by the two asterisked peaks in Figure 7.10(a) and the seven asterisked signals between 6.0 and 8.0 ppm in Figure 7.10(b), all of which were identified with the help of difference spectra. These differences demonstrated further that the vWF-A domain can exist in two conformations that depend on the presence or absence of Mg^{2+} ions. No change in the vWF-A-222 spectrum was observed when $CaCl_2$ was added to a final concentration of 5 mM.

In order to examine whether the same Mg^{2+}-induced conformational change could be

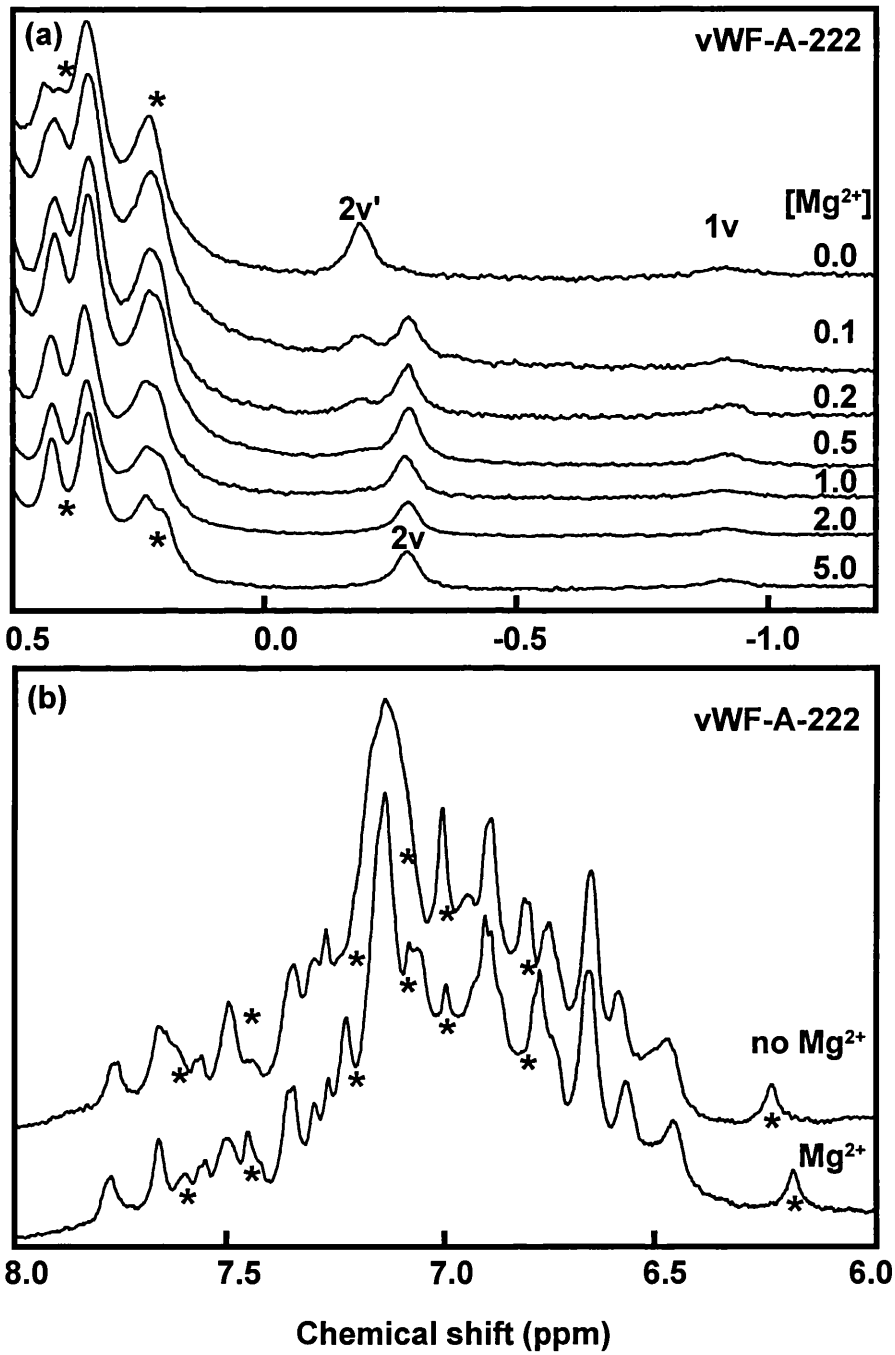


Figure 7.10 The effect of Mg^{2+} on the vWF-A-218 and vWF-A-222 domains by ^1H NMR spectroscopy. A line broadening of 1 Hz was applied to all the spectra before Fourier transformation.

(a) The upfield NMR spectrum of the vWF-A-222 sample, with the Mg^{2+} concentration in mM shown to the right. Apart from the shift change denoted by 2v and 2v', two other differences are highlighted by * symbols.

(b) The aromatic region of the vWF-A-222 domain spectrum in the Mg^{2+}-bound and Mg^{2+}-free states. Seven differences between the two spectra are asterisked for ease of visualisation.

seen in the Bb fragment prepared from native factor B, the NMR experiments were repeated using this. The upfield NMR spectrum of the Bb fragment showed broader signals that corresponded to the doubling of the molecular mass of the vWF-A domain, and a high degree of spectral overlap was present (Figure 7.11). It was nonetheless possible to identify both the 2v and 1v signals at similar chemical shifts to those for the vWF-A spectra. In particular, the 2v signal of the Bb fragment showed an identical conformational dependence on the presence of Mg^{2+} to that of the vWF-A-222 domain, in that the 2v peak at -0.29 ppm was shifted downfield to the 2v' position at -0.19 ppm when 1 mM EDTA was added. The 1v peak of the Bb spectrum was unaffected by the presence of Mg^{2+} or EDTA. The importance of these results is that they show that the conformational dependence of the vWF-A domain on the presence of bound Mg^{2+} is relevant to the vWF-A domain in the context of the native Bb fragment.

7.3.5 Temperature and pH stability of the vWF-A domain by ^{1}H NMR spectroscopy

The thermal stability of the vWF-A domain was studied by ^{1}H NMR spectra. In the presence of 5 mM Mg^{2+}, the signal 2v for the vWF-A-222 domain was unchanged in chemical shift and intensity between 30°C and 45°C, and disappeared at 50°C (data not shown). In contrast, the broad signal 1v moved downfield with increase of temperature, and also disappeared at 50°C. Similar NMR results were obtained for the vWF-A-218 domain (Figure 7.12a). When the metal-free vWF-A-222 domain was studied in the same manner, the high field-shifted methyl signals and the conformationally-dependent aromatic signals disappeared at 40°C and above (data not shown). These NMR results showed that the Mg^{2+}-bound form of the vWF-A-222 domain denatured at 45-50°C in the presence of Mg^{2+} and at 40°C in its absence. This difference is in full agreement with the CD and FT-IR spectroscopy data above.

The pH stability of the vWF-A-218 domain in 0.5 mM Mg^{2+} was studied by NMR. In Figure 7.12(b), only the main results are shown for clarity, even though spectra were recorded at smaller pH intervals than those shown. The vWF-A-218 signals 1v and 2v were

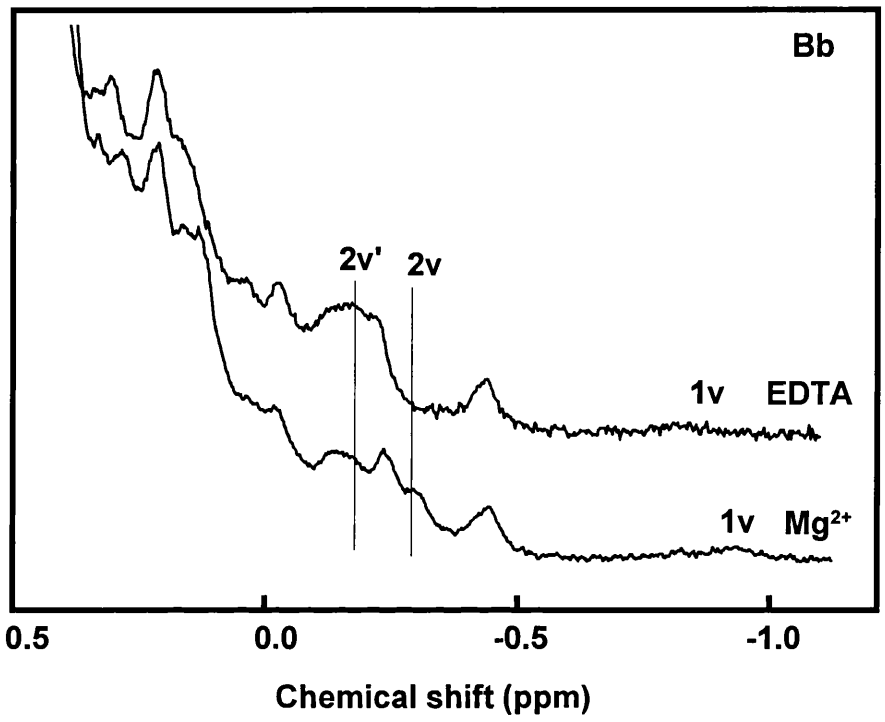


Figure 7.11 The dependence of the upfield spectral regions of the 500 MHz ^{1}H NMR spectra of the Bb fragment on the presence of Mg^{2+}. The sample concentration was 3 mg/ml. The lower spectrum was measured from a sample containing 1 mM $MgCl_2$ and the upper spectrum from the same sample containing 1 mM EDTA. The two highfield-shifted signals are labelled 2v and 1v in the upper spectrum and 2v' and 1v in the lower spectrum.

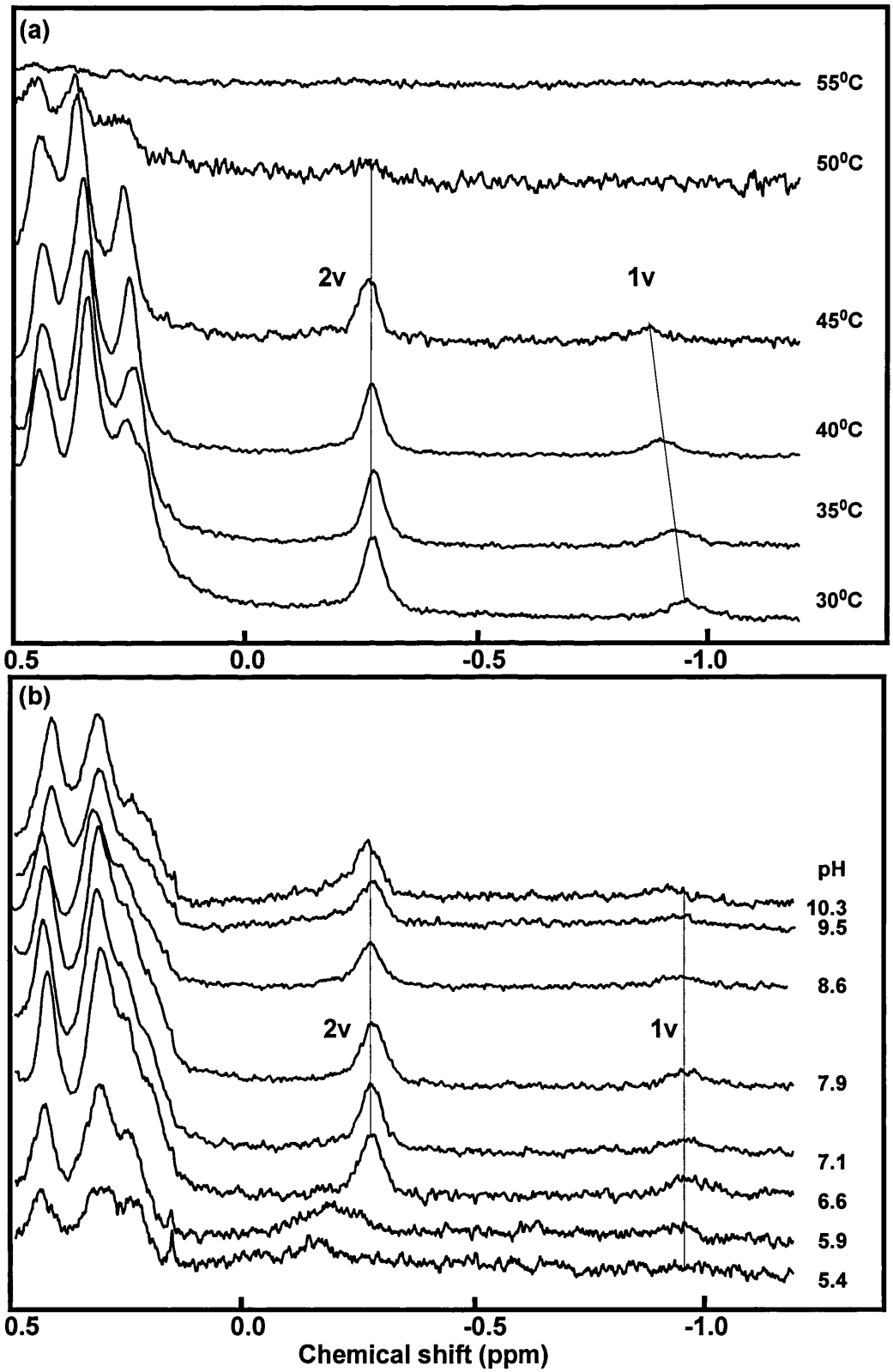


Figure 7.12 Conformational properties of the upfield region of the vWF-A-218 NMR spectrum. Chemical shifts were referenced using the H^2HO signal which had been previously calibrated against trimethylsilylpropionic acid. Lines showing the positions of signals 2v and 1v are added for clarity. (a) Temperature dependence at pH 7.5 in the presence of 0.5 mM Mg^{2+}. (b) pH dependence at 30°C in the presence of 0.5 mM Mg^{2+}.

unchanged throughout the pH range between 6.6 and 9.5, and likewise the aromatic region of the spectrum. The vWF-A-218 domain was poorly soluble below pH 5.0 when it precipitated. Despite this, it could be seen from Figure 7.12(b) that signal 1v was almost unchanged in its shift of -0.94 ppm at pH 5.9, while that for signal 2v underwent large downfield shifts at pH 5.9 and pH 5.4. This change is readily accounted for in terms of the protonation of a neighbouring Asp or Glu residue at low pH, and is further evidence of the conformational dependence of the 2v signal on bound metal, in the context of which two Asp residues are involved (Table 7.1).

7.3.6 Ring current calculations of the vWF-A homology model

Ring current calculations provide a molecular explanation for the highfield-shifted NMR signals of the vWF-A domain starting from a homology model for the vWF-A domain that was based on 23 sets of vWF-A coordinates (Hinshelwood *et al.*, 1999). The number of observed methyl signals in the upfield-shifted spectral region between 0.5 ppm and 0.0 ppm was estimated from signal areas. This gave an estimated total of about 7 methyl signals for both the vWF-A-218 and vWF-A-222 domains when compared in intensity with the methyl signal 2v and the single proton signal 1v (Figures 7.9, 7.10 and 7.12) observed upfield of 0 ppm. This total was compared with the calculated ring current shifts of all the vWF-A methyl groups as the result of their proximity to neighbouring aromatic sidechains, using the vWF-A homology model in conjunction with the program RCCAL (Perkins, 1982). This predicted 7 resonances to occur between 0.5-0.0 ppm, in good agreement with the estimated number of observed signals (Figures 7.9 and 7.10). In addition, RCCAL predicted that two ring-current shifted signals originating from Leu314 and Leu383 were present upfield of 0.0 ppm (Table 7.2). The positional accuracy of the aliphatic and aromatic side chains in the homology model was not considered sufficient to permit any firm assignments of the 2v signal to a given methyl group. This is illustrated by the prediction of the Leu383 methyl signal to occur at -0.42 ppm as the result of the location of its methyl group at 4.6 Å from the plane of the Tyr338 ring, while that for Leu314 was predicted to occur at -1.41 ppm as the result of the location of its methyl group at 4.0 Å from the plane of the Trp296 ring. This is a difference of only 0.6 Å in relative position from a given aromatic ring. Four candidate ring-

Table 7.2 Four ring-current interactions predicted to lead to the signal 2v in the factor B vWF-A ^{1}H NMR spectrum

Methyl group (secondary structure)	**Aromatic residue source of shift (secondary structure)**	**Predicted shift* (ppm from DSS)**
Leu314 (A3-A4 loop)	Trp296 (BC)	- 1.41
Leu383 (A5)	Tyr338 (A4)	- 0.42
Val439 (A7)	Tyr401 (BE), Phe428 (BF)	0.04
Thr355 (A4-BD loop)	Trp352 (A4-BD loop)	0.21

* The next predicted highest-field signal occurs at 0.43 ppm.

current interactions that may be responsible for signal 2v are summarised in Table 7.2 and Figure 7.13. All four corresponded to aliphatic and aromatic sidechains that are buried in the protein core, and all four involve the vWF-A α-helices. Their location in the vWF-A homology model (Figure 7.13) showed that two involved the upper face of the β-sheet, while the other two involved its lower face. It was of interest that two of the four predictions involved residues (Tyr338 and Thr355/Trp352) close to the conformationally-flexible side of the active site cleft at the two metal-binding residues Thr328 and Asp364 (Table 7.1), as signal 2v had shown metal-dependent and pH-dependent properties (Figures 7.10a, 7.11 and 7.12b). An analysis of ring-current shifted single protons upfield of 0 ppm indicated that two such signals were possible, where both protons were proximate to Cys267 on α-helix A1 in a hydrophobic pocket associated with the aromatic ring of Tyr284. This prediction is consistent with the small dependence of signal 1v on the length of α-helix A7 and the replacement of Cys267 in the vWF-A-218 and vWF-A-222 domains.

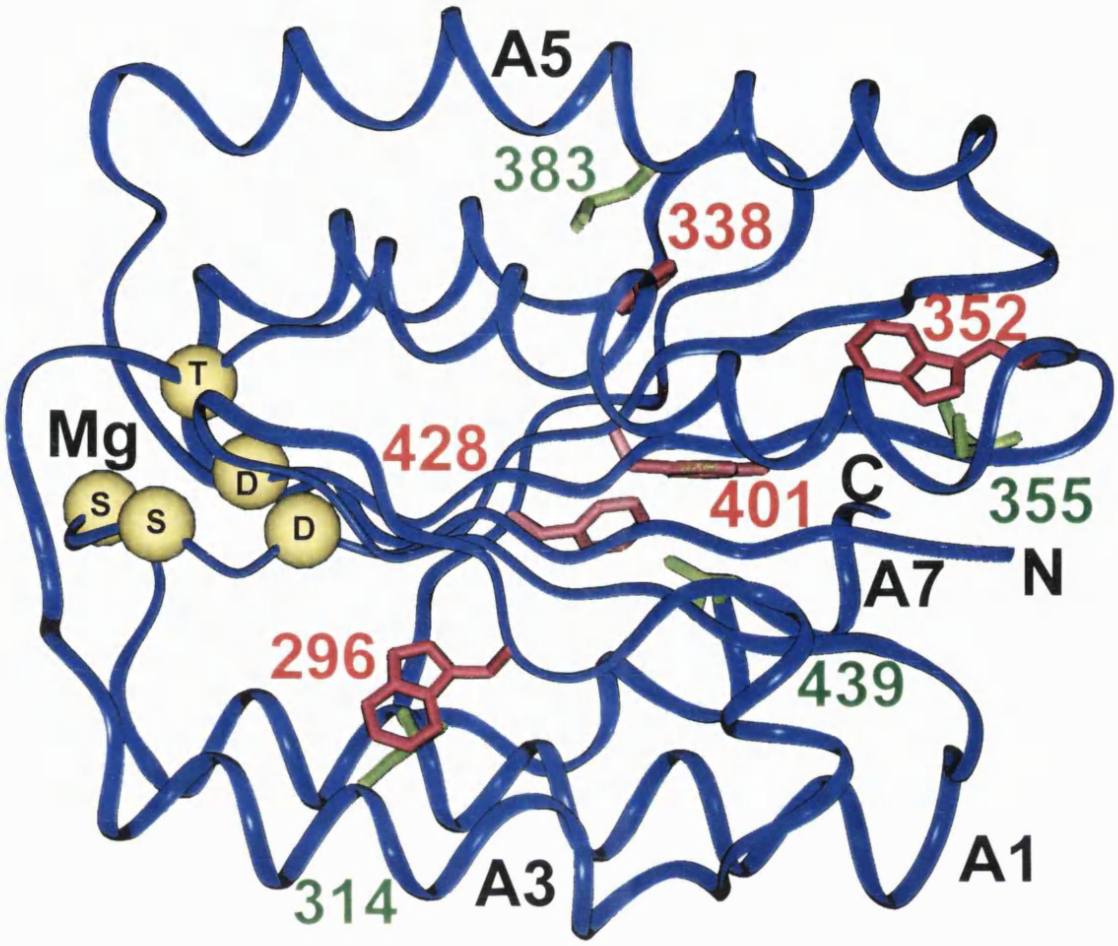


Figure 7.13 Methyl-aromatic ring interactions in the homology model for the vWF-A domain of factor B. Four potential interactions responsible for the methyl signals 1v and 2v/2v' in Figures 7.8, 7.9, 7.10 and 7.11 are shown. The sidechains of the methyl and aromatic residues are shown in green and red respectively, and are identified by residue numbering (Figure 7.5). The N- and C-termini are labelled as N and C respectively, and the α-helices A1, A3, A4, A5 and A6 are labelled as well. The five Cα atoms of the Mg^{2+} binding site at Asp251 and Asp364 in the DxSxS/D/T motif are denoted by yellow spheres.

7.4 Conclusions

The combined use of independent data sets from CD, FT-IR and ^{1}H NMR spectroscopy has provided the first unequivocal structural evidence that a vWF-A domain has two distinct conformations in the presence and absence of metal. The joint application of CD and FT-IR spectroscopy offers complementary views of the vWF-A secondary structure. Since all three methods are based on solution studies, no ambiguities were caused by the presence of a crystallographic lattice. In the CD study of the vWF-A domain from factor B, where precautions had been taken to ensure the presence of excess metal or EDTA as appropriate, the magnitude of the ellipticity change between 215 nm to 220 nm on the addition of metal to the vWF-A-218 and vWF-A-222 domains was about 20%. This difference is most easily attributed to an alteration of at least the α-helix content of the vWF-A domain, as this secondary structure type makes the largest contribution to CD spectra (Drake, 1994). The FT-IR spectra were able to follow separately each of the α-helix and β-sheet bands, and these identified a small but detectable change in the β-sheet content of the metal-free vWF-A domain. The combination of both spectroscopic methods indicated that both the α-helix and β-sheet vWF-A secondary structures were perturbed by Mg^{2+} binding. The use of NMR spectroscopy permitted the monitoring of conformation-dependent chemical shifts that arise from interactions between proximate aliphatic and aromatic sidechains. The large number of signal perturbations seen in the presence of Mg^{2+} showed again that the metal affected the vWF-A structure. The same result was found for the vWF-A domain within the Bb fragment. Through the observation of NMR signals for single groups, the NMR data provided the clearest evidence yet for two metal-dependent conformational states of the vWF-A domain in solution. In thermal denaturation experiments, all three spectroscopic methods indicated a significant increase in the stability of the factor B vWF-A domain in the presence of Mg^{2+}. A similar stabilisation effect was recently observed for the vWF-A domain from the α1β1 integrin (very late activation protein-1; VLA-1; CD49a) in the presence of Mn^{2+} (Gotwals *et al.*, 1999).

Up to now, there has been no strong structural evidence that such a metal-induced conformational change occurs in integrin vWF-A domains, even though it is well established

that they possess a Mg^{2+}-binding site. Only one crystal structure out of 23 had reported such a change for the domain found in CR3, but a subsequent crystal structure study of the same CR3 domain based on diffusion of Mg^{2+} or Cd^{2+} into the crystal had argued against this (Lee *et al.*, 1995a, 1995b; Baldwin *et al.*, 1998). It was argued that the original metal-induced conformational change may be a crystallographic artefact caused by the movement of the C-terminal α-helix A7 of CR3 to bring Glu314 close to the metal-binding site of an adjacent molecule in the crystal lattice. In spectroscopic studies of integrin vWF-A domains to date, a maximal intensity change of 0% to 10% has been seen in the CD spectrum for the vWF-A domain of CR3 in which Mg^{2+} or Ca^{2+} was either present or absent (Fairbanks *et al.*, 1995), while little change was detected in the CD spectrum of the vWF-A domain from the α1β1 integrin in the presence or absence of metal (Nolte *et al.*, 1999).

From this study and others, there is an accumulation of evidence that the conformation of vWF-A domains can be modified during ligand binding and activation (Loftus and Liddington, 1997). Crystallography originally indicated that the vWF-A domain of CR3 possessed different biologically-relevant conformations when bound to either Mg^{2+} or Mn^{2+} (Lee *et al.*, 1995a, 1995b). Evidence from mutational studies indicated that the vWF-A1 domain of von Willebrand factor existed in "on" and "off" conformations (Cooney and Ginsberg, 1996). The separate locations of the mutation sites on opposite sides of the vWF-A1 domain of von Willebrand factor that were associated with each of the Types 2B and 2M phenotypes in von Willebrand's disease indicated that the vWF-A1 domain was able to transmit ligand binding signals within its structure in order to upregulate or downregulate the binding of von Willebrand factor to platelets (Jenkins *et al.*, 1998).

In application to factor B, this Mg^{2+}-induced conformational variability is likely to be important for the activation and function of the C3 convertase when the Bb fragment is complexed with C3b. The Mg^{2+}-bound form will be involved in C3b binding, as the physiological concentration of Mg^{2+} in plasma is 0.7-1.1 mM (Kumar and Clark, 1990), of which 60% is present as free cations (Pryzdial and Isenman, 1986). This means that the metal site will be occupied in plasma factor B (Figure 7.10a). In this C3b.Bb complex, the vWF-A

active site cleft makes contact with C3b (Hinshelwood *et al.*, 1999), so it is likely that C3b will interact with bound Mg^{2+}. It is known that metal is trapped within the Bb.C3b complex after the complex is formed in the presence of metal (Fishelson *et al.*, 1983). The role of Mg^{2+} in the formation of this complex has not yet been defined, as the Bb.C3b complex can be formed in the presence and absence of Mg^{2+} (Pryzdial and Isenman, 1987; Williams and Sim, 1994). It is recognised from the present study that the vWF-A domain structure is stabilised by metal. As it has been shown here that the vWF-A domain can exist in one of two different conformations, it follows that this may provide the basis for an allosteric conformational signal for the activation of factor B, in which the binding of C3b will alter the Mg^{2+}-stabilised structure of the vWF-A domain. It can be proposed that a C3b-binding interaction at the active site cleft will alter the Mg^{2+} coordination in such a way that the conformations of surface α-helices directly connected to this cleft are modified (Perkins *et al.*, 1999). Evidence that this is plausible was shown by the greater conformational flexibility of one side of the cleft by crystallography (Table 7.1). In factor B, this involves Thr328 and Asp364 which are preceded by α-helix A3 and followed by α-helix A5. Topologically, one of these coincides with the region where the mutation sites on α-helix A3 cause the upregulation of von Willebrand factor in Type 2B von Willebrand's disease (Jenkins *et al.*, 1998). This α-helix A3 is adjacent to the α-helix A7 (Figure 7.13), which in factor B is joined to the serine protease domain of factor B by a short peptide link. The proposal that an allosteric signal can be transmitted from the active site cleft to the opposite surface of the vWF-A domain is therefore consistent with what is known for vWF-A domains to date, and may explain how the serine protease domain of factor B is activated in the C3bBb complex.

Chapter 8

Experimental:
Conformational changes during the assembly of factor B from its domains by ^{1}H NMR spectroscopy and molecular modelling

8.1 Introduction

All the enzymes participating in the major steps concerned with complement activation or control (C1r,C1s, MASPs, C2, factor B, factor D, and factor I) belong to the family of mammalian serine proteases (Johnson *et al.*, 1984; Perkins and Smith, 1993). On the basis of their primary structure, complement SP domains belong to the chymotrypsin family of proteins and, on the basis of their specificity for Arg residues, to the trypsin superfamily.

Serine proteases cleave a peptide bond within the polypeptide substrate in a two step reaction via nucleophillic attack on the carbonyl carbon (C_1) of the scissile bond by the hydroxyl oxygen of the reactive Ser (Ser195: chymotrypsin numbering) (Gerard *et al.*, 1998). The first step produces a covalent bond between the C_1 carbon atom of the substrate before the scissile bond and the hydroxyl group Ser195 of the enzyme. Production of this acyl-enzyme intermediate proceeds through a negatively charged transition state intermediate where the bonds of C_1 have a tetrahedral geometry in contrast to the planar triangular geometry in the peptide group. During this step the peptide bond is cleaved, one peptide product is attached to the enzyme and the other peptide product rapidly diffuses away. In the second step of the reaction, deacylation, the acyl-enzyme intermediate is hydrolysed by a water molecule to release the second product peptide with a complete C-terminus and to restore the Ser hydroxyl of the enzyme.

A typical mammalian SP domain contains two subdomains, each with a six-stranded antiparallel β-sheet barrel (Figure 8.1). The spatial relationships among the catalytic triad residues (His57, Asp102 and Ser195: chymotrypsin numbering) are crucial for the synergistic action and are invariable in the active conformation of all chymotrypsin-like serine proteases of known structure (Figure 8.2). All the complement enzymes possess these three catalytic residues and the surrounding highly conserved regions.

Serine proteases have four important structural features that facilitate the catalysis mechanism. (1) A His residue acts a general base that can accept a proton from the hydroxyl group of the reactive Ser to facilitate formation of the covalent tetrahedral transition state.

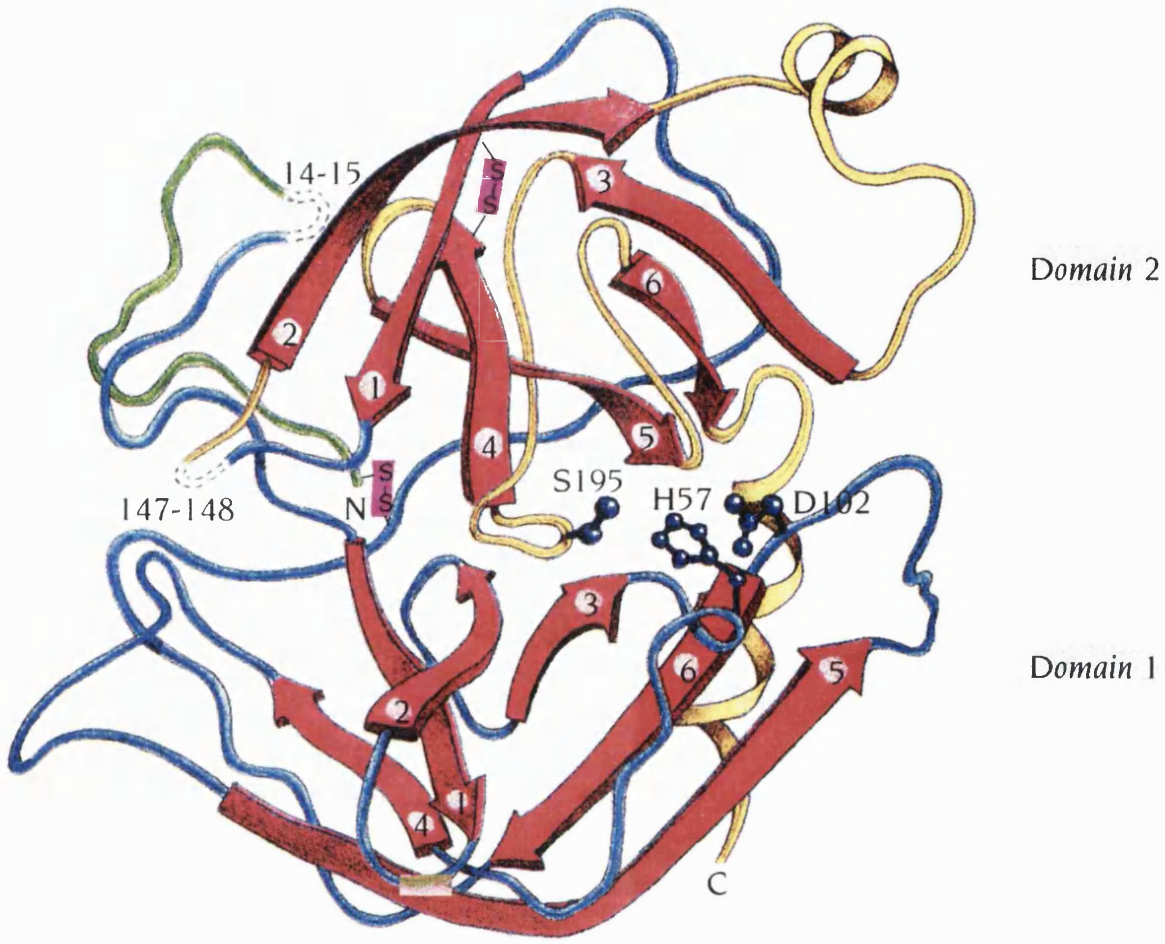


Figure 8.1 Schematic diagram of the structure or chymotrypsin, which is folded into two antiparallel β domains. The six β strands of each domain are red, the side chains of the catalytic triad are blue and the disulphide bridges that join the three polypeptide chains are marked in violet. Chain A (green residues 1-13) is linked to chain B (blue, residues 16-146) by a disulphide bridge between Cys1 and Cys122. Chain B is in turn linked to Chain C (yellow, residues 149-245) by a disulphide bridge between Cys136 and Cys201. Dotted lines indicate residues 14-15 and 147-148 in the inactive precursor, chymotrypsinogen. These residues are excised during the conversion to the active enzyme chymotrypsin (From Branden and Tooze, 1999).

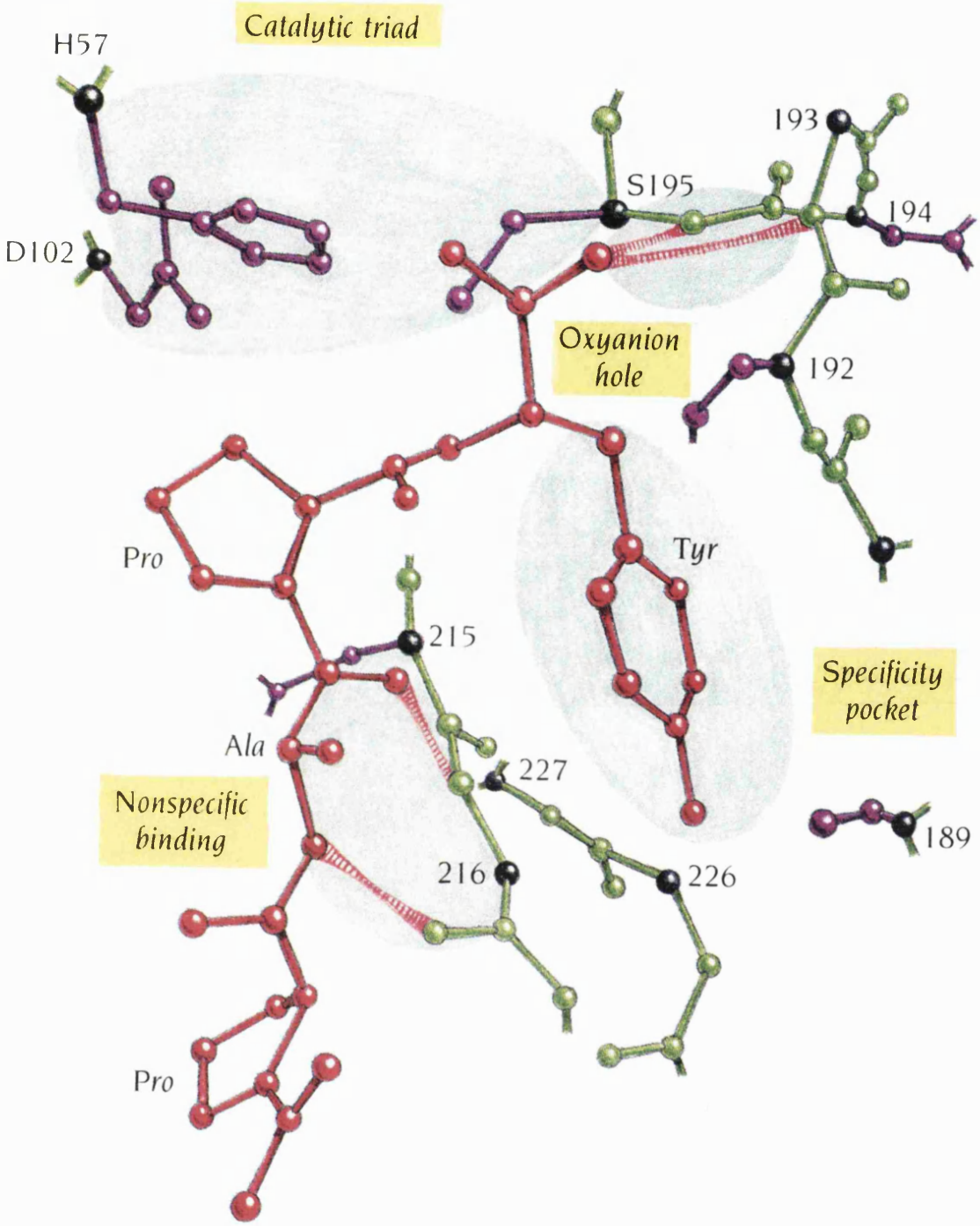


Figure 8.2 A diagram of the active site of chymotrypsin with a bound inhibitor, Ac-Pro-Ala-Pro-Tyr-COOH. The diagram illustrates how this inhibitor binds in relation to the catalytic triad, the substrate specificity pocket, the oxyanion hole and the non-specific substrate binding region. The inhibitor is red. The hydrogen bonds between the enzyme and inhibitor are striped. (Adapted from Branden and Tooze, 1999).

(2) Stabilisation of the transition state intermediate is accomplished by groups in a pocket called the oxyanion hole which hydrogen bond to the negatively charged oxygen atom attached to C_1. (3) Most serine proteases have no absolute substrate specificity. Polypeptide substrates exhibit non-specific binding to the enzyme through their main-chain atoms, which form hydrogen bonds with main chain atoms of a loop region of the enzyme to form a short antiparallel β-sheet. (4) The preferred side-chain before the scissile bond (large aromatic side-chains in the case of trypsin) is orientated so as to fit into the S1 specificity pocket.

The activation of serine proteases generally leads from an inactive disordered zymogen to a more rigid active structure and so they are allosteric proteins (Huber and Bode, 1978). The substrate-binding specificity and affinity of serine proteases are determined by the geometry and the chemical nature of the substrate-binding pockets, especially the S1 specificity pocket. In some cases the activation can require specific protein-protein and enzyme-substrate interactions and enzyme self assembly.

Despite their overall structural similarity to typical pancreatic serine proteases, the complement enzymes have unique structural features that allow them to carry out their highly specialised functions. These unique structural qualities endow complement serine proteases with their extremely restricted substrate specificity, which is necessary for their function and reflected in their low reactivity with synthetic substrates and active site inhibitors. Factor B is an atypical serine protease whose SP domain contains large sequence insertions and deletions which, together with the N-terminal vWF-A domain, endows it with a restricted substrate specificity (Perkins and Smith, 1993; Jing *et al.*, 2000). Unusually the activation of factor B involves a cleavage at the N-terminus of the vWF-A domain, not at the N-terminus of the SP domain.

Factor D consists of a single SP domain. Factor D circulates in a "zymogen-like" conformation, and its activation is induced by its substrate C3bB (Kam *et al.*, 1987; Kim *et al.*, 1994). The catalytic triad and substrate-binding residues of free factor D possess atypical conformations that preclude its catalytic activity (Narayana *et al.*, 1994). In its S1 specificity

pocket, Asp189 (chymotrypsinogen numbering) forms a salt bridge with Arg218 instead of performing its classical role at the base of the oxyanion pocket. In addition, Thr214-Arg218 act as a self-inhibitory loop that narrows the active site cleft and accounts for the low reactivity of factor D with synthetic esters (Volanakis and Narayana, 1996). The recognition of C3bB by factor D leads to the rearrangement of this self-inhibitory loop and the generation of an active catalytic triad and S1 pocket, and so there is no requirement for an additional inhibition mechanism (Jing *et al.*, 1998).

At present, little is known of how proteolytic activity is generated in the C3bBb complex (Arlaud *et al.*, 1998). As predicted from sequence comparisons with the archetypal serine proteases (Perkins and Smith, 1993), the crystal structure of a recombinant SP domain in factor B showed that large surface sequence insertions were present (Jing *et al.*, 2000). The crystal structure showed that large deletions were also present, and that the active site cleft was formed except for its oxyanion pocket. Recent crystal structures for β2-I glycoprotein and membrane cofactor protein (CD46) have yielded the most accurate models to date of the SCR domain (Bouma *et al.*, 1999; Casasnovas *et al.*, 1999), and enable the SCR domains in the Ba fragment to be modelled. A homology model for the vWF-A domain has been built using comparisons with five vWF-A crystal structures (Chapter 6) (Hinshelwood *et al.*, 1999). The combination of these structures with ^{1}H 1D-NMR spectroscopy enabled the solution structures of factor B and its fragments and domains to be studied (Evans, 1995; Perkins and Wüthrich, 1980). The conformational dependence of the Ba and Bb fragments and the isolated vWF-A and SP domains on pH and temperature was assessed and these were compared with factor B and factor D. The first structural evidence of detectable conformational changes during the assembly of factor B from its domains was demonstrated, and this was correlated with the proteolytic activity of factor B. From this it is proposed that both the Ba fragment and the vWF-A domain have roles in the allosteric activation of the SP domain of factor B.

8.2 Materials and Methods

8.2.1 Preparation of factor B, the Ba and Bb fragments, C3, factor D and the SP domain

The protein samples were prepared as described previously (Chapter 4). Samples were concentrated under pressure on an Amicon YM10 membrane (mass cut-off of 10,000 Da). Each preparation was dialysed into PBS in ^2H_2O with 1 mM $MgCl_2$ over 36 h with 4 buffer changes at 8°C.

8.2.2 ^{1}H NMR spectroscopy

^{1}H NMR spectra were recorded using 500 MHz and 600 MHz Varian spectrometers at the MRC NMR Centre at the National Institute of Medical Research, Mill Hill and analysed using VNMR software. Sample concentrations were in the range 2-10 mg/ml. pH meter readings in ^2H_2O were uncorrected values measured on a Corning 240 pH meter, and additions of stock solutions of NaO^2H or ^2HCl were used to adjust the pH. The H^2HO chemical shift was referenced with respect to the chemical shift of trimethylsilylpropionic acid (1 mg/ml) in the same buffer and temperature range as the protein samples, and the protein NMR spectra were then referenced to the H^2HO signal. A line broadening of 1 Hz was applied before Fourier transformation. The pK_a values of His residues were determined from their Cε proton (also designated as C(2)H) and the Cδ proton chemical shifts by fitting to a Henderson-Hasselbach reverse sigmoidal function (Bycroft and Fersht, 1988), based on the five parameter sigmoid regression function in SIGMAPLOT (Jandel).

8.2.3 Homology modelling of the domains of factor B

The human factor B sequence (SWISSPROT accession number P00751; Mole *et al.*, 1984) was aligned with the sequences of reference structures in the Protein Data Bank (PDB) database. The human C2 and mouse factor B and C2 sequences have accession codes of P06681, P04186 and P21180 respectively. The identity of each alignment of two sequences was defined as (the number of identical residues) × 100 / (the total number of topologically-equivalent positions in the alignment, excluding gaps in either sequence). Secondary structure elements were identified using DSSP (Kabsch and Sander, 1983).

Residue solvent accessibilities were calculated using a probe of 1.4 Å in the COMPARER program (Lee and Richards, 1971; Šali and Blundell, 1990). For alignment purposes, the DSSP and COMPARER analyses were supplemented by visual inspection using INSIGHT II software (MSI, San Diego, U.S.A.) on Silicon Graphics INDY Workstations in conjunction with Crystal Eyes stereo glasses. The rigid body fragment assembly method in HOMOLOGY together with BIOPOLYMER and DISCOVER (MSI) was used for modelling. A precalculated Cα distance matrix identified the PDB loops that best fitted the corresponding Cα distance matrix calculated from the structurally conserved regions that defined the start and end of each searched loop for a specified number of flanking and intervening residues. Sidechain atoms were automatically generated for both the structurally conserved regions and structurally variable regions using the template structures and general rules for residue exchanges. Models were refined using energy minimization, where 300 steps of steepest descent minimisation were performed, and assessed using PROCHECK (Laskowski *et al.*, 1993). The electrostatic surface of each model was calculated using a full Coulombic boundary condition in DELPHI (MSI). The internal and external relative permittivities (dielectric constants) were set as 2 and 80 respectively. The solvent accessible surface of each model was displayed using the Connolly algorithm with a solvent probe of diameter 0.14 nm. The surface was coloured red for potentials less than -5 k_BT (acidic), blue for potentials greater than +5 k_BT (basic) and white for neutral potentials of 0 k_BT. Colours between these values were produced using a linear interpolation.

In an earlier study (Hinshelwood and Perkins, 1998), the first SCR domain in the Ba fragment was referenced to the NMR structure of the fifteenth SCR domain of human complement factor H (PDB code: 1hfi; Barlow *et al.*, 1993), and the second and third SCR domains in the Ba fragment were referenced to the NMR structure of the sixteenth SCR domain of factor H (PDB code: 1hcc; Norman *et al.*, 1991). This required the modelling of 35 residues in 8 searched loops. Subsequently the modelling was much improved by referencing the first and third SCR domains of the Ba fragment to the second SCR domain in the crystal structure of β2-I glycoprotein at 2.7 Å resolution, and the second SCR domain of Ba to the first SCR domain of β2-I glycoprotein (PDB code: 1qub; Bouma *et al.*, 1999).

This required the modelling of 22 Ba residues in 3 searched loops, one of which was a large one at Thr62-Lys70 that could not be precisely modelled. The N-terminal and C-terminal residues Thr1-Arg7, Ala75-Ile76, Gly136-Gly138 and Asp195-Asp200 were added as arbitrarily extended structures. Likewise, in the earlier study (Hinshelwood and Perkins, 1998), the SP domain was modelled using bovine α-chymotrypsin as the reference (PDB code 2cha; Cohen *et al.*, 1981; Wang *et al.*, 1985; Tsukada and Blow, 1985) and the multiple sequence alignment of Perkins and Smith (1993). The three disulphide bridges at Cys486-Cys502 (the His loop), Cys631-Cys657 (the Met loop) and Cys670-Cys700 (the Ser loop) as well as that at Cys453-Cys571 were well conserved and readily modelled. Modelling of the disulphide bridge at Cys574-Cys590 required a sequence realignment in which Asp123 of α-chymotrypsin was assigned to the position of Cys590. As the result of the sequence insertions, a total of 136 residues were remodelled in 16 searched loops, which is 47% of the total sequence. Subsequently the homology model was replaced by the SP crystal structure at 2.1 Å resolution, kindly provided by Dr S. V. L. Narayana (Jing *et al.*, 2000).

8.2.4 Calculation of ring current shifts

In the proton NMR spectra of proteins, resonances upfield of 0.5 ppm correspond to ring current shifts of methyl groups or single protons of amino acid side-chains positioned within about 7 Å of the centre of an aromatic ring and affect only a small proportion of proton signals (Perkins, 1982). RCCAL (Perkins, 1982) was used on a Silicon Graphics INDY Workstation to calculate these from protein coordinates in PDB format. These included the SCR, vWF-A (Chapter 6) and SP homology models (Hinshelwood and Perkins, 1998; Hinshelwood *et al.*, 1999), the factor H SCR NMR structures (PDB codes 1hfi, 1hfh, 1hcc; Norman *et al.*, 1991; Barlow *et al.*, 1991, 1992, 1993), five factor D structures (PDB codes 1dsu, 1dst, 1dfp; Narayana *et al.*, 1994; Kim *et al.*, 1995; Cole *et al.*, 1997; Jing *et al.*, 1998), and the factor B SP domain (PDB code 1dle; Jing *et al.*, 2000). The correction of the random coil chemical shifts of the free amino acid residues (Bundi and Wüthrich, 1978) by the predicted ring current shifts resulted in the values reported in Table 8.1. The validity of these calculations has been established in calibration studies (Perkins, 1982).

Table 8.1 Ring-current interactions predicted to lead to upfield-shifted signals in the ^{1}H NMR spectrum of factor B and factor D

Domain	Methyl group (secondary structure)	Aromatic residue source of shift (secondary structure)	Predicted shift* (ppm from DSS)
SCR-21	Leu33 (B2)	Trp57 (B5)	-0.48
SCR-2	Ile102 (B2)	Trp126 (B5)	0.06
SCR-3	Val162 (B2)	Trp186 (B5)	0.09
SP**	Ala500 (DE loop)	Tyr534 (GH loop)	0.45 ± 0.01
	Val528 (G)	Phe503 (DE loop), Phe530 (G)	-0.74 ± 0.01
	Leu554 (H)	Trp726 (C-terminus helix)	-0.49 ± 0.03
	Ile570 (HI loop)	Trp470 (AB loop)	-1.36 ± 0.15
	Thr575 (IJ loop)	Phe738 (C-terminus helix)	0.04 ± 0.21
	Leu615 (JK loop)	Tyr465(AB loop)	0.34 ± 0.06
	Thr669 (LM loop)	Phe607 (J)	0.20 ± 0.04
	Leu678 (M)	His717 (C-terminus helix)	0.49 ± 0.07
	Ile718 (C-terminus helix)	Phe655 (L)	0.45 ± 0.21
	Leu727 (C-terminus helix)	Phe738 (C-terminus helix), Phe495 (D)	0.15***
	Leu731 (C-terminus helix)	Phe495 (D)	-0.17***
factor D**	Leu73 (FG loop)	Trp141 (JK loop)	0.43 ± 0.10, -0.74 ± 0.15
	Leu105 (H)	Trp237 (C-terminus helix)	-0.29 ± 0.12
	Leu123 (I)	Tyr29 (AB loop)	-0.13 ± 0.13
	Leu155 (JK loop)	Trp141 (JK loop)	-1.31 ± 0.22, 0.47 ± 0.02
	Leu199 (M)	Tyr228 (O)	0.50 ± 0.15
	Leu242 (C-terminus helix)	Trp51 (D), Trp237 (C-terminus helix)	0.08 ± 0.07 -0.41 ± 0.19

see overleaf for footnotes:

Table 8.1 continued

Footnotes:

† Only contributions of more than 0.2 ppm to the observed shifts are listed. Of all the methyl and aromatic residues shown, only two are in structurally homologous positions, i.e. Trp470 and Phe495 of factor B are equivalent to Tyr29 and Trp51 respectively in factor D.

* Only signals predicted to be upfield of 0.5 ppm are shown. If two values are shown, these correspond to the two methyl groups.

** The calculated ring current shift for the SP domain of factor B is the mean of those from two coordinate sets (PDB codes 1dle (two molecules)). The calculated ring current shift for factor D is the mean of those from five coordinate sets (PDB codes: 1dst, 1dsu (two molecules), 1dfp (two molecules)).

*** Either Leu727 or Leu731 but not both were predicted to be ring current shifted in the two coordinate sets.

8.3 Results and Discussion

8.3.1 Purification and activity of factor B and its fragments

The purifications of factor B and the Ba and Bb fragments (Chapter 4) resulted in single bands by SDS-PAGE and homogeneous peaks by gel filtration, but yielded multiple bands on isoelectric focusing gels. The latter was attributed to a minor polymorphism of factor B due to the occurrence of allotypes in the SCR-1 domain (Horiuchi *et al.*, 1993) and variable carbohydrate compositions at four sites in the SCR-2 and vWF-A domains (Campbell and Bentley, 1985). In contrast to this, both the recombinant vWF-A domain (Gly229-Ile444 of factor B with 218 residues; Williams *et al.*, 1999) and the elastase-cleaved SP domain each migrated as single bands on isoelectric focusing gels with isoelectric points at approximately pH 6.0 and pH 9.0 respectively (Chapter 4; Figures 4.11 and 4.12), in good agreement with values of pH 6.2 and pH 8.5 respectively that were predicted from their amino acid sequences. This outcome for the SP domain was attributed to the absence of glycosylation sites. Sequencing of the first fifteen residues in the SP domain by automated Edman degradation gave DESQSLSLxGMVWEH, where x was attributed to a derivatised cystine residue. This agreed with the N-terminal sequence for the SP domain (Lambris and Müller-Eberhard, 1984), and showed that it comprised Asp445-Leu739 of factor B. As a further confirmation of the integrity of the disulphide-rich SP domain, assays for possible free sulphydryl groups (Chapter 4) showed no appreciable increase in the absorbance of the 5,5'-dithiobis-2-nitrobenzoic acid stock at 412 nm in either the folded or unfolded states. This indicated that all its Cys residues were involved in disulphide bonds.

Assays of the isolated SP domain demonstrated detectable catalytic activity against C3 by SDS-PAGE and showed that this preparation was functionally active (Figure 8.3(a)). A low molecular mass band at 8 kDa that was assigned to the cleaved C3a fragment was visible in the incubations with C3 and the SP domain (Figure 8.3(a)). The SP domain showed activity between pH 6.0 to 11.0, with an optimum at around pH 8.0 (Figure 8.3(b)). It also cleaved the synthetic C3 analogue, Boc-Leu-Gly-Arg-7-amino-4-methyl coumarin (Caporale *et al*, 1981), as measured by an increase in the fluorescence of the reaction mixture. Here, no pH optimum could be determined, since both the starting level of fluorescence and its rate

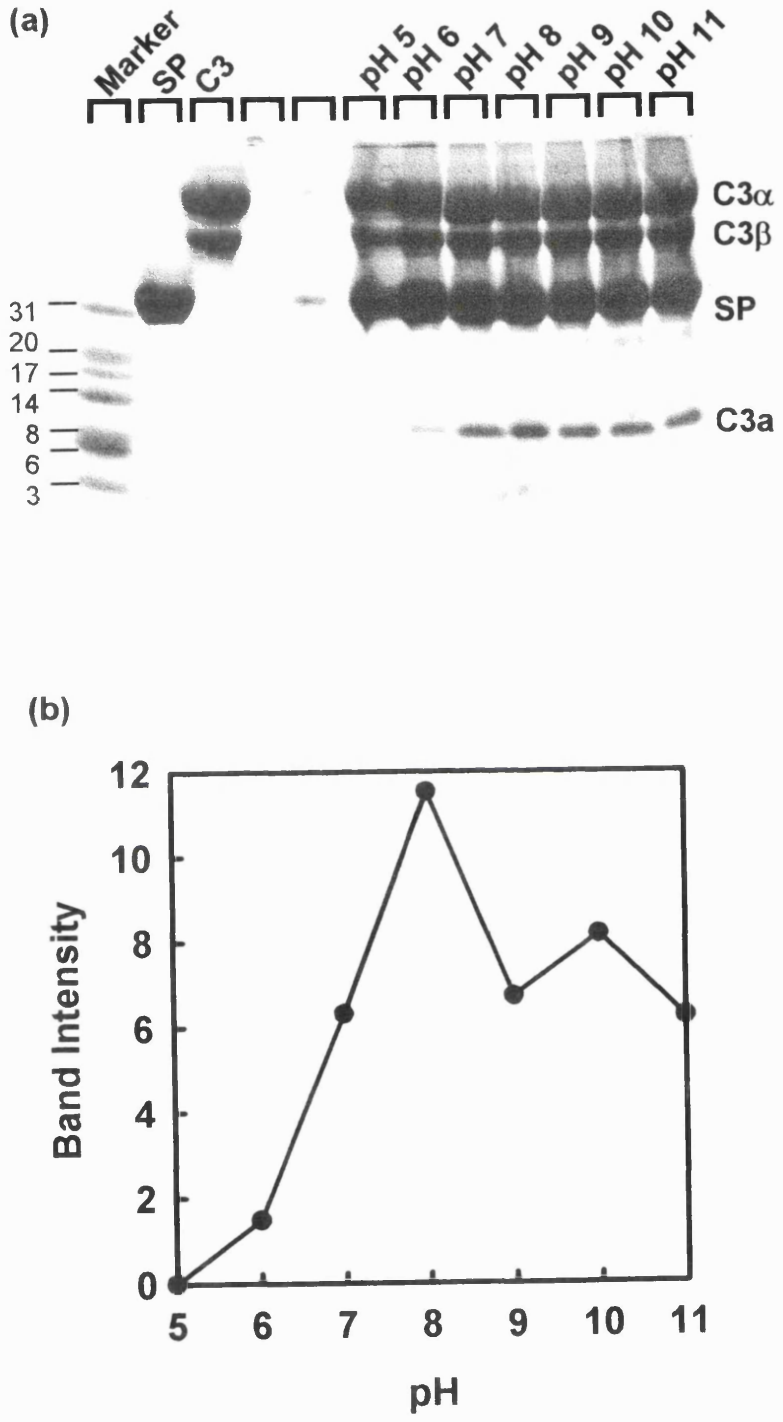


Figure 8.3 The pH dependence of the activity of the isolated SP domain from factor B.

(a) 16% Tricine gel stained with Coomassie blue following a two-hour incubation at 37°C of C3 with the SP domain. The reducing gel separated C3 into its α and β chains. Molecular mass markers are shown in kDa. The SP and C3 lanes are controls. The cleavage product C3a is observed at 8 kDa.

(b) Relative intensity of the C3a band in (a) as a function of pH.

of increase increased significantly above pH 7.5, although the activity was reduced below pH 6.0. This agrees with observations that both the Bb fragment and SP domain show low levels of activity compared to factor B (Fishelson and Müller-Eberhard, 1984; Lambris and Müller-Eberhard, 1984). It is inferred that the SCR domains regulate the activity of the SP domain of factor B.

Functional assays of recombinant factor D showed that this was active against native factor B in the presence of C3(H_2O), an equivalent of C3b, where SDS-PAGE analysis indicated the appearance of the 30 kDa Ba and 60 kDa Bb fragments (not shown). Activity assays between pH 4.0 and pH 12.0 by SDS-PAGE gels showed that the optimum activity for maximal cleavage was at about pH 8.0, and activity was significantly reduced below pH 6.0 and above pH 11.0. In negative controls based on the absence of either factor D or C3, the cleavage of factor B was inefficient, and this indicated the tight regulation of the factor D proteolytic activity.

8.3.2 pH stability of factor B by ^{1}H NMR spectroscopy

All five protein preparations demonstrated significant ^{1}H NMR signal dispersion in their spectra at pH 7.5 at 30°C in phosphate buffer saline with 1 mM $MgCl_2$. This indicated folded protein structures. In the downfield region, all five His C(2)H signals of the Ba fragment were visible between 7.0 and 8.3 ppm. Analysis of their chemical shifts as a function of pH (Figure 8.4a, Table 8.2) gave individual pK_a values that varied between pH 6.0 to pH 7.3, which is typical for a folded globular protein structure. There are nine His residues in the SP domain, of which the C(2)H signals of six could be identified in the pH titration, and yielded individual pK_a values that varied between pH 5.7 and pH 6.4 (Figure 8.4b, Table 8.2).

The upfield spectral region between 0.4 and -1.0 ppm is sensitive to the positions of methyl groups and single protons within about 7 Å of the centre of aromatic rings within the protein, and can be used as a sensitive conformational probe of hydrophobic structures (Perkins, 1982). The NMR spectra from factor B and the Bb fragment (not shown) showed poor resolution for reason of high molecular masses, and their spectral dispersion was

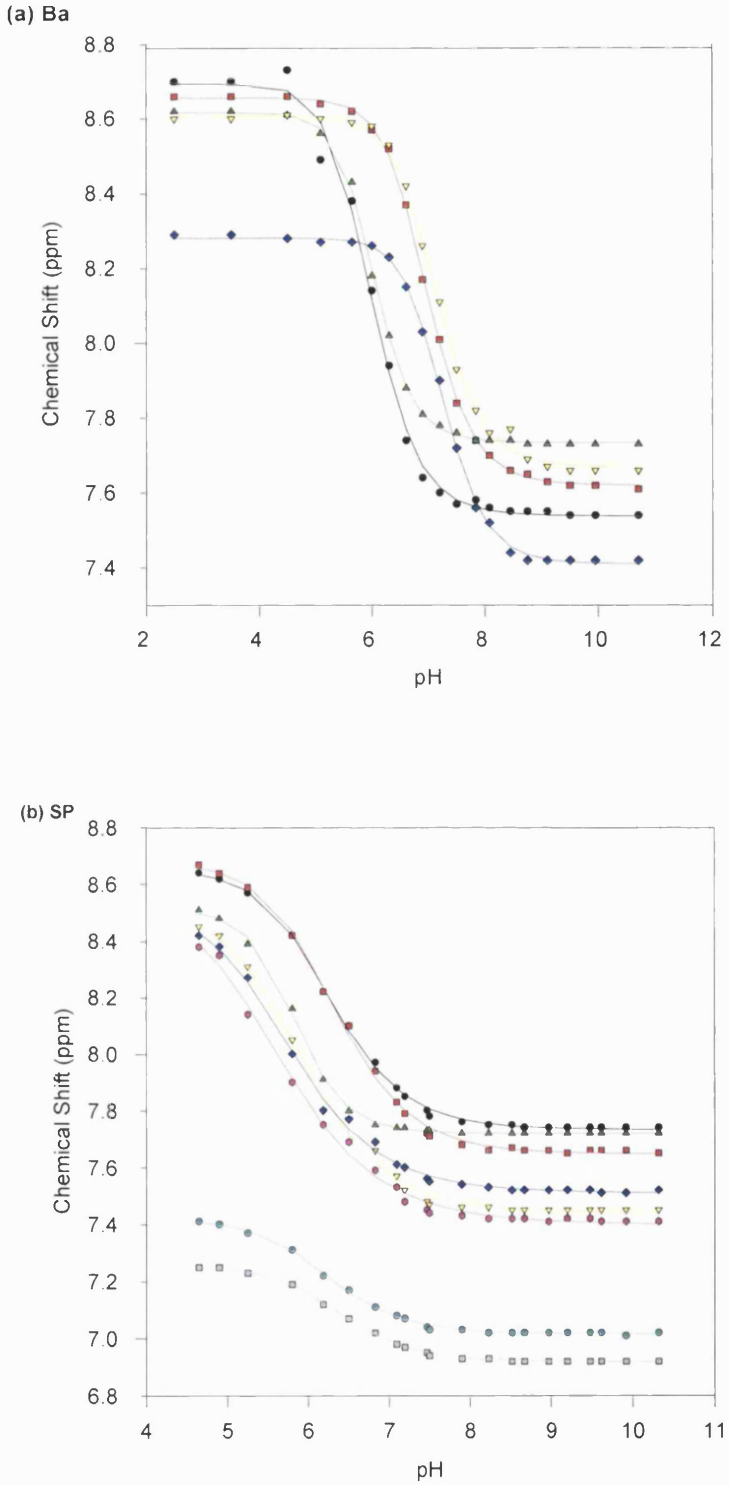


Figure 8.4 Plots of the chemical shift of the His proton resonances of the NMR spectra of (a) the Ba fragment and (b) the SP domain of factor B in the pH titration experiments as a function of pH. All five His C(2)H (Cε proton) resonances were visible in the Ba spectra and six of the nine His C(2)H (Cε proton) resonances and two Cδ proton resonances were visible in the SP spectra. These resonances were plotted as a function of pH and the experimental data were fitted to a Henderson-Hasselbalch reverse sigmoidal function (SigmaPlot). From these curves the pK_as of the five His of the Ba fragment and six of the nine His pK_as of the SP domain were calculated.

Table 8.2 The calculated pK_as of the Ba fragment and SP domain histidines. (a) The pK_as of the five histidines of the Ba fragment (calculated from the pH titrations of the observable histdyl C(2)H resonances). (b) The pK_as of six of the nine histidines of the SP domain (using the pH titrations of the observable C(2)H resonances). (c) The pK_as of two of the histidines of the SP domain (using the pH titrations of the observable Cδ proton resonances).

(a) Ba fragment	C(2)H
pK_a1	6.0
pK_a2	7.0
pK_a3	6.0
pK_a4	7.1
pK_a5	7.3

(b) SP domain	C(2)H
pK_a1	6.3
pK_a2	6.3
pK_a3	5.8
pK_a4	5.9
pK_a5	5.8
pK_a6	5.7

(c) SP domain	(Cδ proton)
pK_a1	6.3
pK_a2	6.4

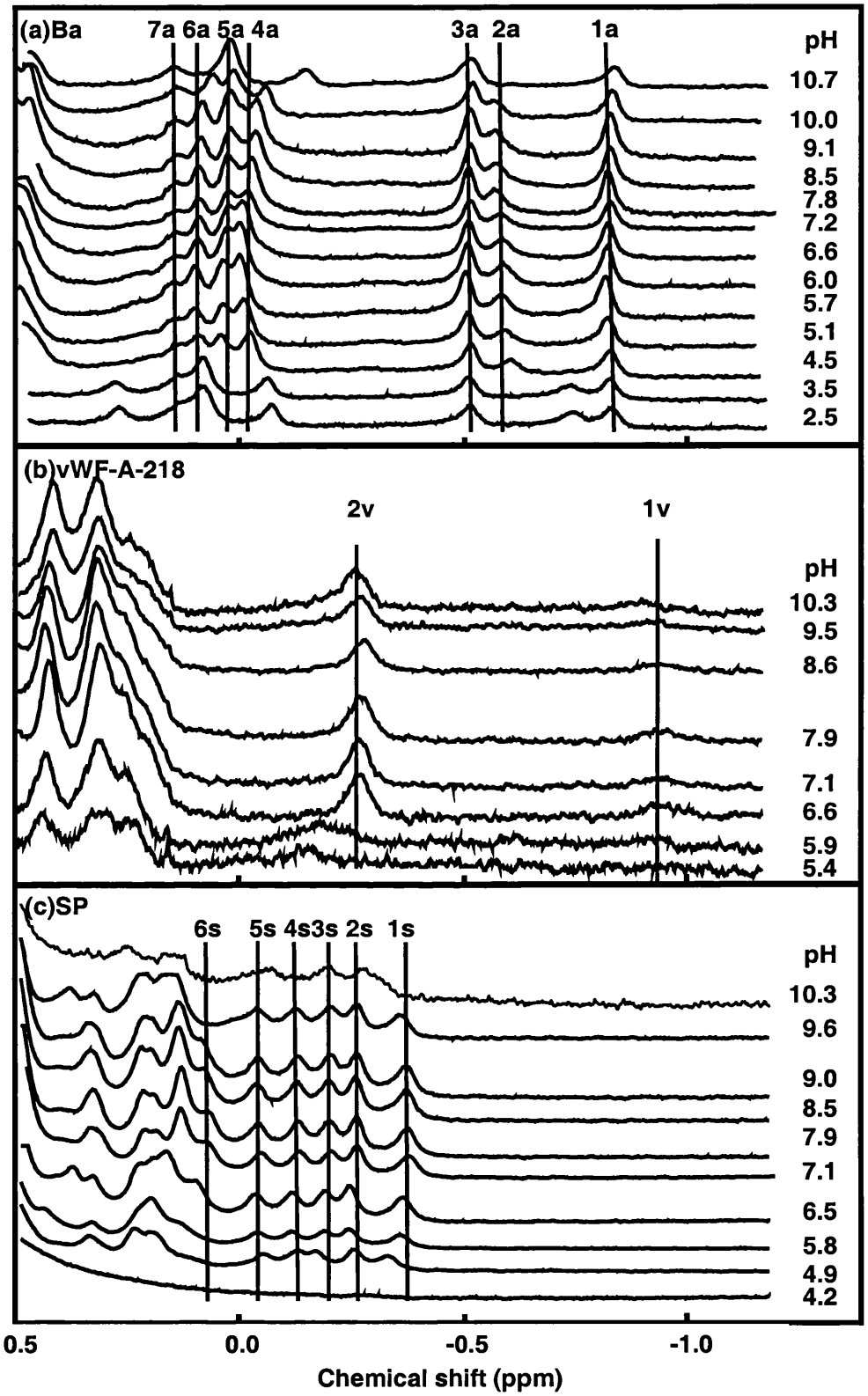


Figure 8.5 pH dependence of the upfield ^{1}H NMR spectra of (a) the Ba fragment, (b) the vWF-A-218 and (c) the SP domains. The signals of single methyl or one-proton signals are labelled in decreasing order of chemical shift to correspond to the Ba fragment [a], vWF-A-218 domain [v] and SP domain [s]. To permit the extent of variation of the chemical shift on pH to be visualised, the vertical lines represent the peak positions at pH 7.5.

reduced below pH 5.5 to indicate that both were structurally unstable at low pH. In contrast, Figure 8.5(a) showed that the upfield NMR spectrum of the Ba fragment (30 kDa) was well defined. Seven methyl signals (1a-7a) were detectable between pH 2.5 and 10.7 and showed that the Ba fragment was stable over this pH range. The four signals 4a-7a demonstrated pronounced pH dependences, which suggested a degree of conformational flexibility in the SCR domains. In the vWF-A spectrum (Hinshelwood and Perkins, 2000), two highfield signals 2v and 1v that correspond to a methyl group and a single proton were visible between pH 6.6 and 10.3 (Figure 8.5(b)). As these spectral details were lost at low pH, the denaturation of the vWF-A domain (which contains no disulphide bridges) was likely to contribute to the instability of factor B and the Bb fragment at low pH. In the SP spectrum, six highfield methyl signals (1s-6s) were observed between 0.0 ppm and -1.0 ppm (Figure 8.5(b)). While the five signals 1s to 5s showed little variation with pH, signal 6s and others between 0.0 and 0.5 ppm showed pH-dependent variations in chemical shifts, and this indicated conformational flexibility. Inspection of the full NMR spectra of the Ba fragment and the vWF-A and SP domains showed that signals in the SP spectra were best resolved between pH 7.5 and 9.5. This corresponded to the highest activity of the isolated SP domain between pH 6.0 to 11.0, with an optimum close to pH 8.0 as shown in Figure 8.3(b). That all three preparations exhibited pH-dependent conformational flexibility by NMR is consistent with an allosteric activation mechanism for factor B, for which it is necessary that different conformations can exist.

8.3.3 Interactions between the domains of factor B by ^{1}H NMR

In order to assess whether the structures of the Ba and Bb fragments were preserved within factor B, the upfield methyl NMR spectra of the Ba and Bb fragments were compared with factor B at pH 7.5 and 30°C (Figure 8.6(a)). Six signals in the upfield factor B spectrum could be identified in the spectra of the component Ba and Bb fragments. It was possible to determine that the signals 2a and 3a (from the Ba fragment) and the signals 1v, 2v and 2s (from the Bb fragment) were unshifted compared to intact factor B. While signal 1a from the Ba fragment was detectable in the factor B spectrum, it was either much broadened or had shifted to the position of signal 1v, suggesting a conformational movement had occurred when

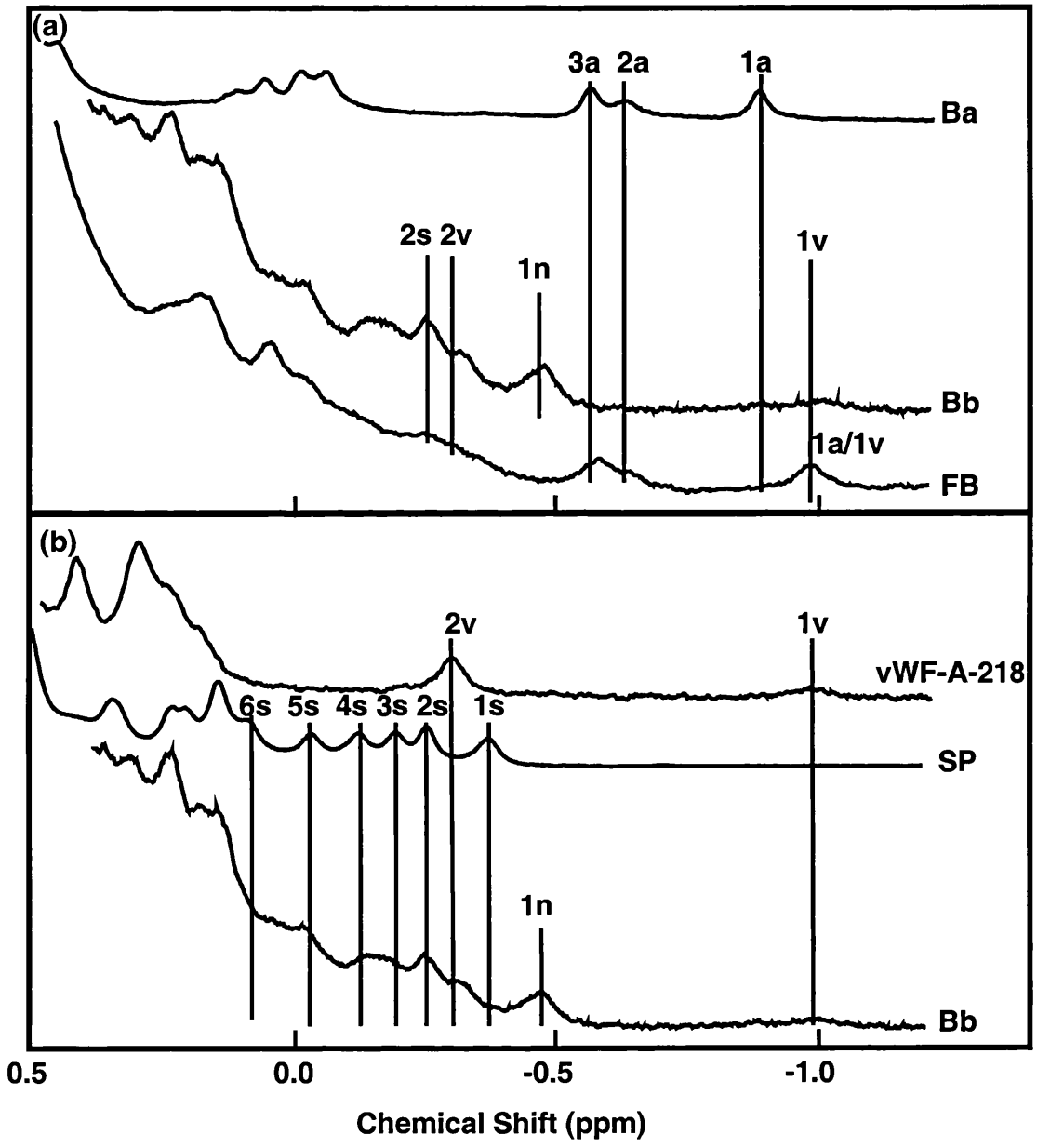


Figure 8.6 Comparisons of the ^{1}H NMR spectra of factor B and its fragments and domains. Vertical lines permit the visualisation of similar signals in the spectra. The signals are labelled according to the scheme of Figure 8.5.

(a) The upfield NMR spectra of the Ba and Bb fragments are compared with that of factor B. The new signal observed in the Bb fragment is denoted by 1n.

(b) The upfield NMR spectra of the vWF-A-218 and SP domains are compared with that of factor B.

the Ba fragment was incorporated into factor B. The Ba signals in the factor B spectrum were broader than those of the free Ba fragment. This showed that the Ba fragment had become less mobile on the NMR timescale when incorporated within factor B. Despite the presence of a 42-residue linker peptide Asp195-Arg234 between the end of the SCR-3 domain and the factor D cleavage site in factor B, there was no indication from the detection of possible narrow linewidths that the Ba fragment was freely attached by a mobile linker to the vWF-A domain. Secondary structure predictions using the program PHD (Rost and Sander, 1993) suggested that the linker possessed a loop structure without α-helix or β-strand, except for the residues K235-L239 immediately after the factor D cleavage site which were predicted to be β-strand. In agreement with this, no high similarity of the linker sequence with that of a known folded protein structure was identified. Figure 8.6(a) also showed that a new signal 1n was visible in the Bb spectrum that was not present in that of factor B.

In order to assess whether the structures of the vWF-A and SP domains were preserved within the Bb fragment, the upfield methyl NMR spectra from the vWF-A and SP domains were compared with that of the Bb fragment at pH 7.5 and 30°C (Figure 8.6(b)). The broadening of the signals in the Bb spectrum in comparison with those of the vWF-A and SP domains is attributed to the higher molecular mass of the Bb fragment. Eight signals in the Bb spectrum could be identified in the spectra of the two separated domains. Both the signals 1v (-0.95 ppm) and 2v (-0.29 ppm) were detected in the Bb and vWF-A spectra at the same shifts. While the five signals 2s to 6s (0.1 ppm to -0.3 ppm) were detected in both the Bb and SP spectra at the same shifts, that for 1s appeared broader in the Bb spectrum. A new signal 1n was identified in the spectrum of the Bb fragment at -0.48 ppm that was not present in the vWF-A or SP spectra. Even though the signal 1s may be present in the Bb spectrum, it was not possible to determine whether it had shifted to that of signal 1n in the Bb fragment. In general, the presence of many unchanged conformation-dependent shifts indicated that the vWF-A or SP domains did not undergo large conformational changes when joined to form the Bb fragment.

The computer summation of the Ba, vWF-A and SP spectra resulted in the NMR

spectrum of factor B in which signal 1n was absent (data not shown). This showed that the Bb fragment possessed a unique conformation different from those of the Ba, vWF-A and SP domains. The absence of signal 1n in the factor B spectrum is attributed to the SCR domains, and is further evidence that the Ba and Bb fragments interact with each other in factor B. There are three possible explanations for signal 1n in the Bb fragment, all of which postulate that localised conformational changes have occurred: (i) The SP domain exists in two allosteric conformational states, one of which is present in intact factor B or the free SP domain, and the other exists when only the vWF-A domain is present. The interaction between the vWF-A and SP domains would perturb the local structure responsible for signal 1s to generate signal 1n. (ii) Another possibility is that a new hydrophobic interaction is formed at the interface between the vWF-A and SP domains to result in signal 1n when both domains are present together in the isolated Bb fragment; in this case, the local environment of signal 1s is perturbed to result in a much broadened signal. This was discounted by an NMR study of a mixture of the vWF-A and SP domains, which showed that the signal 1n was not visible (spectra not shown), even though no sequence overlap existed between the two domains that would preclude their association. (iii) Another possibility is that the newly-cleaved extended Bb N-terminus KIVLD- becomes incorporated within the vWF-A or SP domain in proximity to an aromatic residue. This was discounted by titrations of the SP domain with an excess of the dipeptide Lys.Ile which is a mimic of the vWF-A N-terminus of the Bb fragment Lys235.Ile236 (see Huber and Bode (1978) for more details). The dipeptide had no effect on the NMR spectra and showed that the SP domain was unaffected by the availability of the free N-terminus of the Bb fragment after factor B cleavage.

Comparisons of the downfield spectra of the aromatic residues in Figure 8.7 supported the above results. Despite the spectral overlap in this region, the summation of individual spectra enabled spectral comparisons to be made. The sum of the downfield spectra from the Ba and Bb fragments was very similar to that of factor B (not shown). It was inferred that no major conformational change resulted in factor B on cleavage into the Ba and Bb fragments. The summation of the downfield spectra for the vWF-A and SP domains gave a result very similar to the Bb spectrum except for the appearance of three additional signals

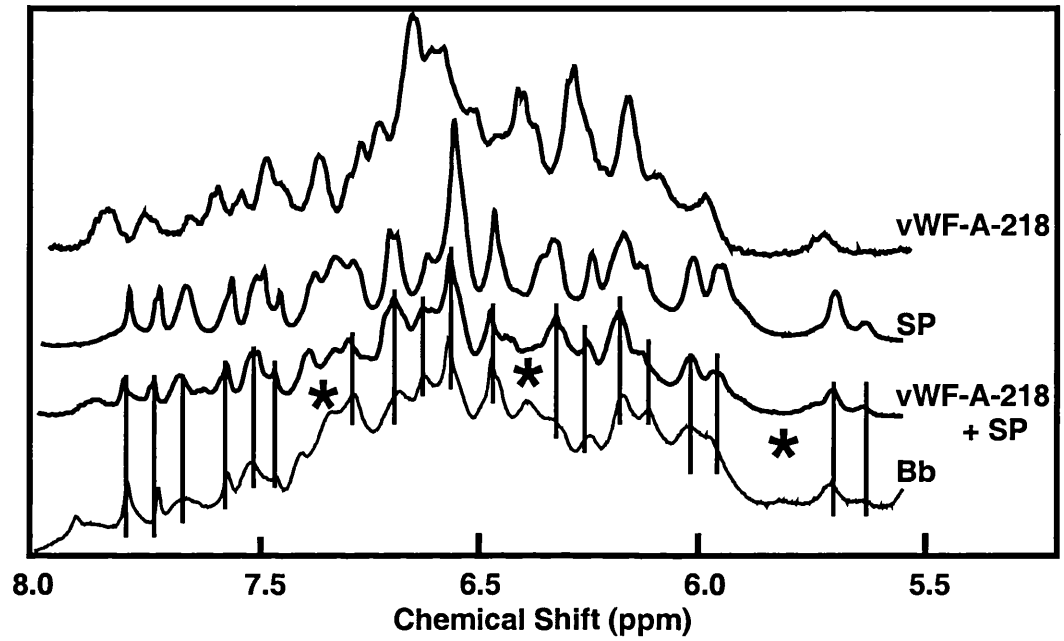


Figure 8.7 Comparisons of the downfield ^{1}H NMR spectra the vWF-A-218 and SP domains with their summation and the spectrum of the Bb fragment. Vertical lines permit the visualisation of similar signals in the spectra. Asterisks highlight the main differences between the summed vWF-A-218 and SP spectrum with that of the Bb fragment.

in the Bb fragment (marked by the asterisks in Figure 8.7).

8.3.4 Temperature stability of factor B by ^{1}H NMR

Temperature dependence studies of the upfield shifted NMR signals at pH 7.5 provided data on both the conformational stability of the methyl-aromatic interactions and the protein denaturation temperatures. In Figure 8.8(a), the Ba spectra showed noticeable temperature dependences in the upfield-shifted signals, and this again indicated conformational flexibility. Denaturation occurred above 60°C and no precipitation was seen even when the Ba fragment was heated to 80°C. In comparison, in Figure 8.8(b), the appearance of the vWF-A spectra was largely unchanged throughout the temperature range. The vWF-A denaturation temperature is metal-dependent and is 45-50°C in the presence of Mg^{2+} as shown, and at 40°C in its absence (Hinshelwood and Perkins, 2000). In Figure 8.8(c), the SP domain started to precipitate at 50°C, but the methyl signals were visible up to 60°C. As also observed in the pH titration above, the five signals 1s-5s were less affected than the remaining signals seen between 0.0-0.5 ppm.

The temperature dependence of the NMR spectrum for the Bb fragment in Figure 8.9(a) showed that the signals 2s, 2v, 1n and 1v could be seen in the Bb spectrum up to 60°C, above which the protein precipitated and the signals disappeared. Signal 1s was not detected (Figure 8.6(b)). Even though temperature-dependent upfield chemical shifts were observed in the Ba and vWF-A spectra, it was noteworthy that signal 1n displayed almost no shift dependence with temperature, much as observed for the 1s-5s signals of the free SP domain. This showed that the hydrophobic interaction responsible for signal 1n was structurally defined and stable. In the NMR spectra of factor B in Figure 8.9(b), despite overlap between signals 1a and 1v, the signals 3a, 2a and 1a/1v were visible up to above 50°C, and did not reappear when the protein was cooled back to 30°C. Factor B did not precipitate even when heated to 80°C. Similar results were obtained from analysis of the downfield spectra for all five proteins. It was of interest that the denaturation temperature for the Bb fragment (55°C) was intermediate between those of the vWF-A and SP domains (50°C and 60°C respectively). That for factor B (50°C) was however less than those for the

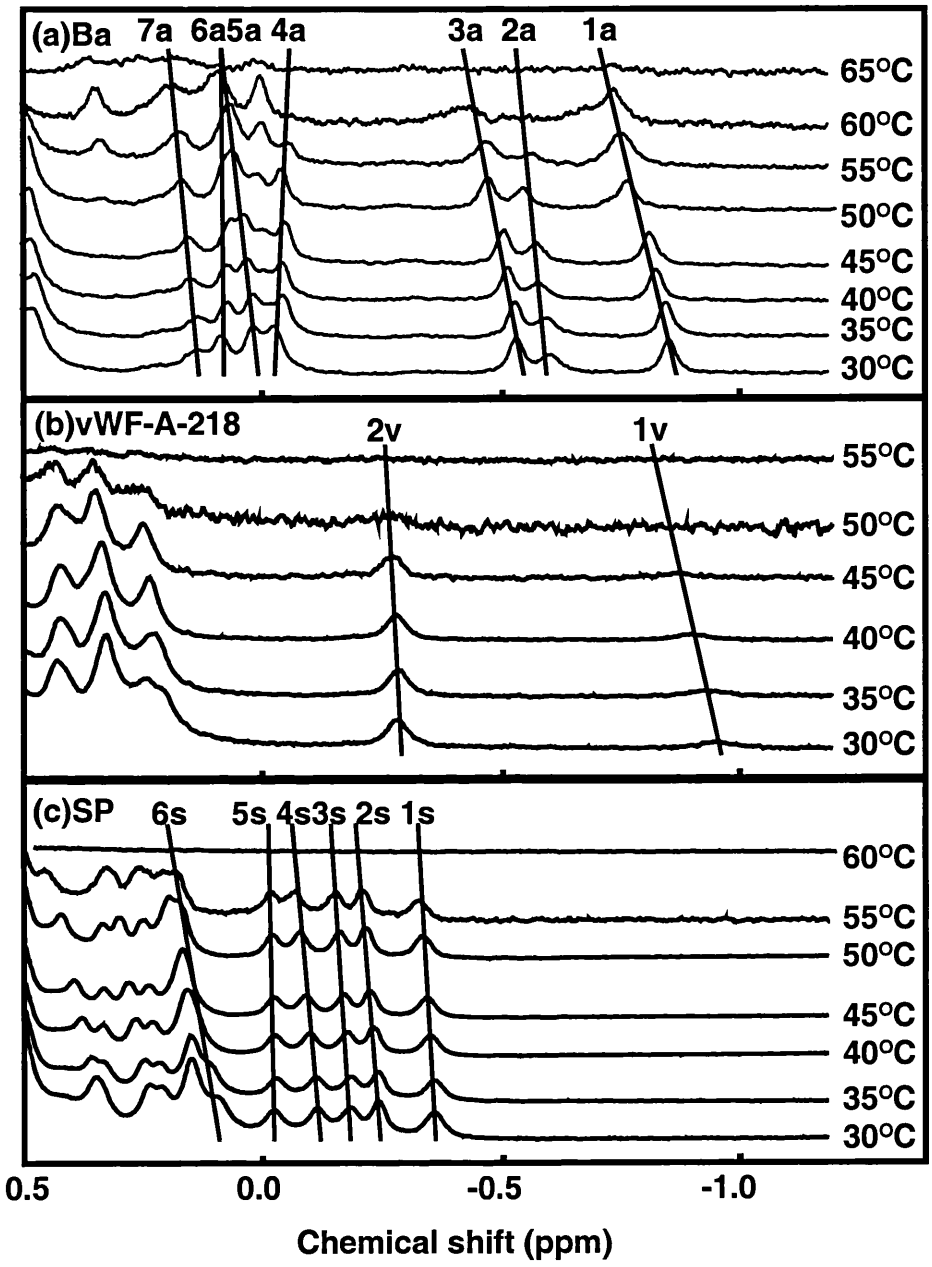


Figure 8.8 Temperature dependence of the upfield ^{1}H NMR spectra at pH 7.5 of (a) the Ba fragment, (b) the vWF-A-218 domain and (c) the SP domain of factor B. Lines permit the visualisation of the rate of shift change with temperature. The signals are labelled according to the scheme of Figure 8.5.

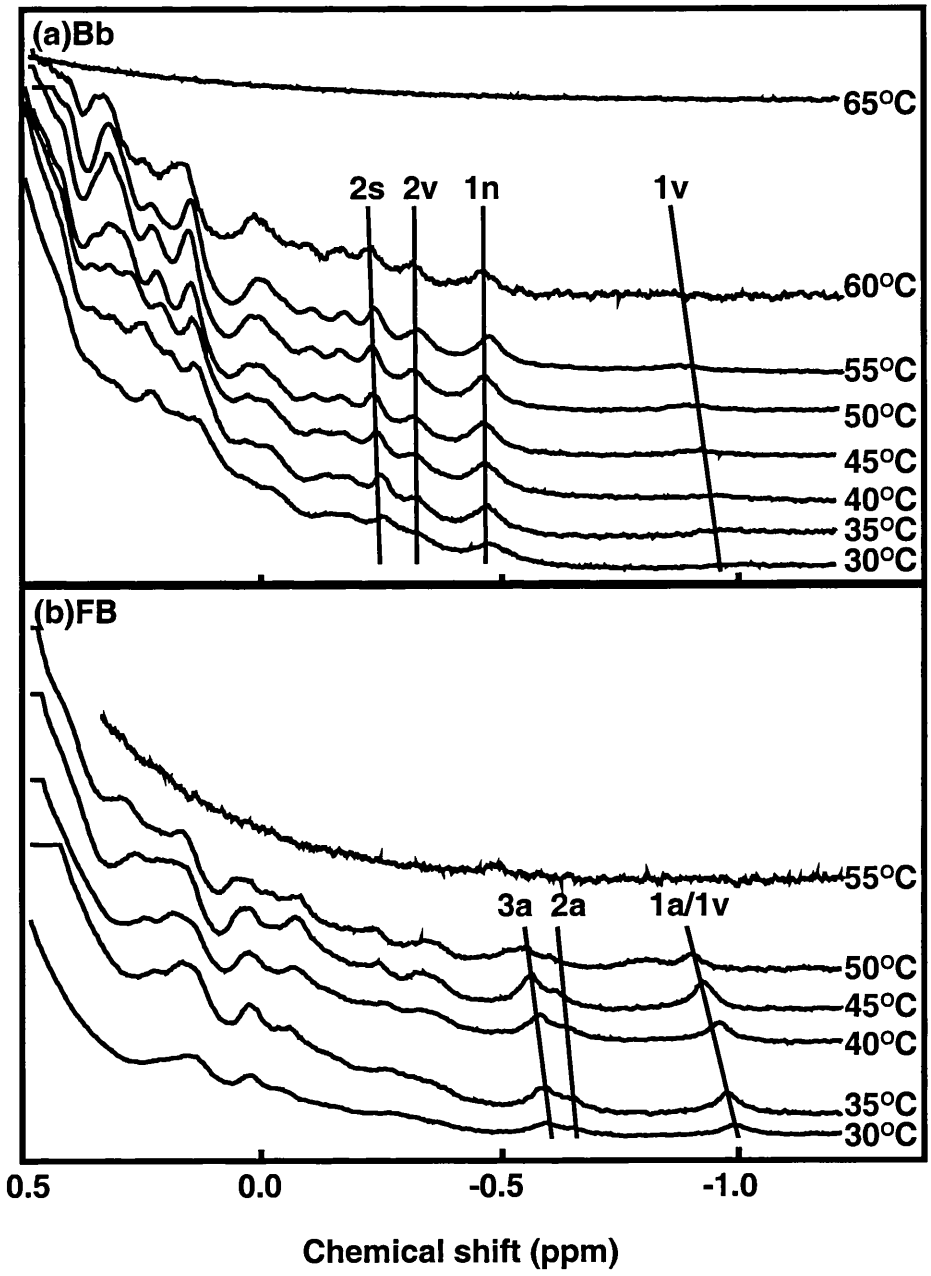


Figure 8.9 Temperature dependence of the upfield ^{1}H NMR spectra at pH 7.5 of (a) the Bb fragment and (b) factor B. Lines permit the visualisation of the rate of shift change with temperature. The signals are labelled according to the scheme of Figure 8.5.

Ba and Bb fragments (65°C and 55°C). This suggested that the Ba and Bb fragments interact with each other to partly destabilise the structure of intact factor B.

8.3.5 NMR spectra of factor D

As large sequence insertions and deletions were present in the SP domain of factor B when this is compared with the archetypal serine proteases such as β-trypsin and α-chymotrypsin (Perkins and Smith, 1993; Jing *et al.*, 2000), these may affect the structure of the SP domain. Accordingly comparative NMR studies of factor D were performed on the grounds that this is similar in size to β-trypsin and α-chymotrypsin, and corresponds to a complement SP domain with a known crystal structure. Three broad upfield-shifted methyl signals 1d-3d were well resolved (Figure 8.10). It was noticeable that their mean linewidth was 33 ± 5 Hz, which is higher than that of 22 ± 3 Hz for the signals 1s-5s from the SP domain of factor B or that of 26 ± 6 Hz for the signals 1a-6a from the Ba fragment, all of which are similar in size. As crystallography (Narayana *et al.*, 1994; Kim *et al.*, 1995; Cole *et al.*, 1997; Jing *et al.*, 1998) and biochemical assays (Figure 8.3: see above) show that factor D has a "zymogen-like" conformation, while the isolated SP domain from factor B possesses residual proteolytic activity, the broad linewidths for factor D may correspond to multiple conformations of an inactive factor D molecule, each one of which possessed distinct ring current shift perturbations, while the narrower linewidth for the SP domain of factor B corresponds to a better-defined single conformation. In support of this explanation, the NMR study of the activation of the zymogen trypsinogen to active β-trypsin did not reveal any new signals, but only differences in line broadening (Perkins and Wüthrich, 1980). Factor D is monomeric under the conditions studied by NMR (Perkins *et al.*, 1993), so the possibility of dimer formation cannot explain the broader linewidths.

In studies of the pH and temperature stability of factor D, the methyl signals 1d-3d were present between pH 5.1 and pH 11.0. At pH above 9.6, there was an sharp change in which signal 1d disappeared and a new signal 4d appeared, and no change in the linewidths was observed. In comparison with the SP domain of factor B, the NMR spectrum denatured above pH 9.6. At pH values below 7.0, heavy precipitation of factor D occurred, while this

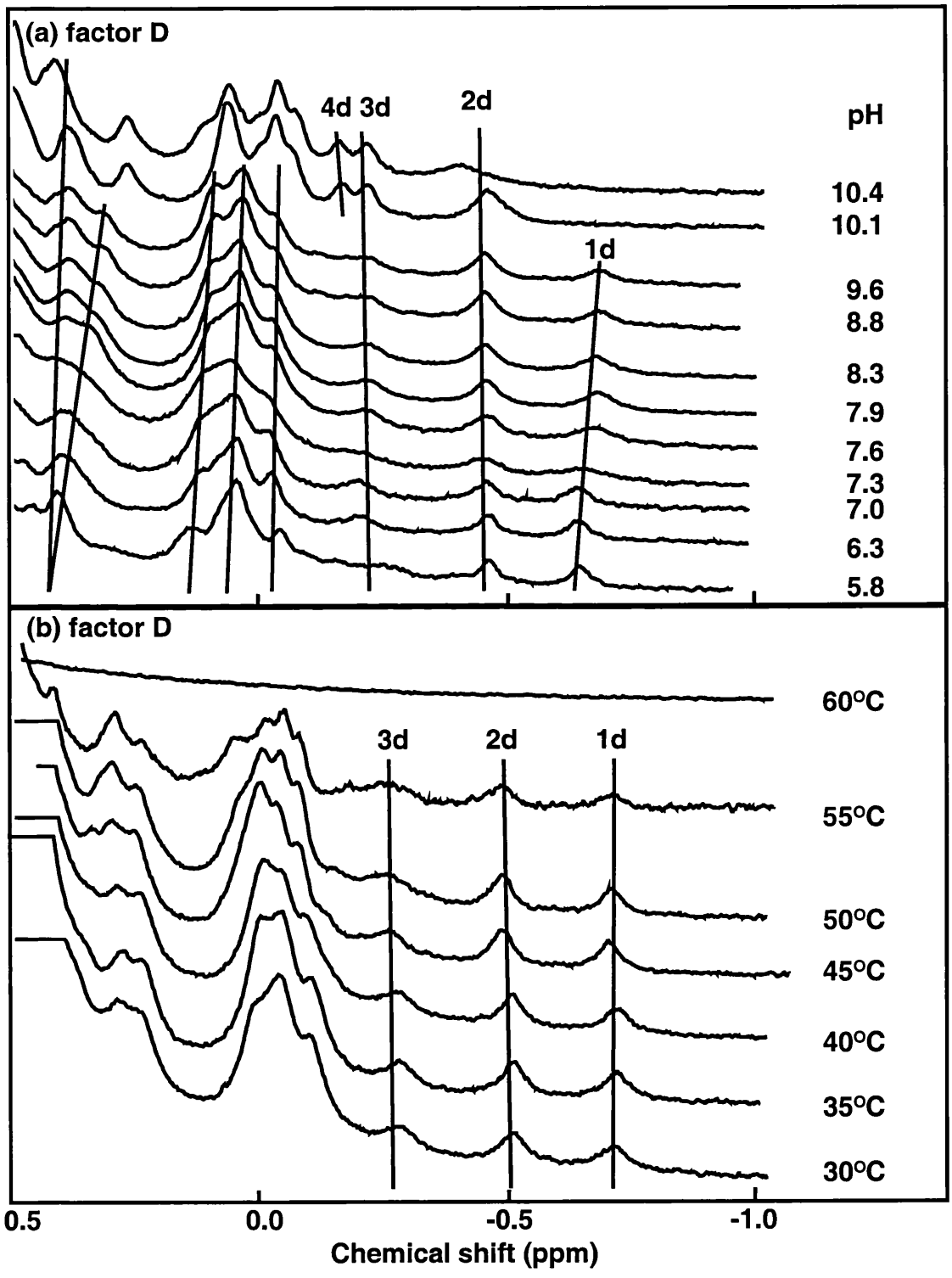


Figure 8.10 pH and temperature dependence of the upfield ^{1}H NMR spectra at pH 7.5 of factor D. The signal labelling (1d to 4d) corresponds to resolved signals in decreasing order of chemical shift. The vertical lines in (a) represent the peak positions at pH 7.5.

occurred only below pH 5.0 for the SP domain of factor B. These observations are consistent with activity assays which showed a maximum at pH 8.0 for both proteins (see above) and with previous studies which showed that the regulation of the alternative pathway was pH dependent (Fishelson *et al.*, 1987). Temperature studies showed that factor D began to precipitate above 50°C and the upfield shifted signals of factor D disappeared above 55°C. The denaturation temperature was less than that of 60°C seen for the SP domain of factor B, but was similar to that of 55°C for the Bb fragment, and indicated that factor D is less thermostable than the SP domain of factor B.

8.3.6 Ring current calculations based on homology models or crystal structures

Ring current calculations starting from crystal structures or homology models (Figure 8.11) provide a molecular explanation for highfield shifted NMR signals (Perkins, 1982). The computation of these in the earlier work for the Ba fragment (Hinshelwood and Perkins, 1998) was based on the construction of homology models for the SCR-1, SCR-2 and SCR-3 domains using NMR structures for the fifteenth and sixteenth SCR domains in factor H (Methods). This was handicapped by significant sequence insertions and deletions and the fair quality of the NMR-derived coordinates as assessed by PROCHECK (Laskowski *et al.*, 1993). In contrast, in the present study, the first and second SCR domains in the crystal structure of β2-I glycoprotein (Bouma *et al.*, 1999) were of high quality, and (apart from a large insertion at Thr62-Lys70) only two one-residue deletions were required to perform the modelling of the three Ba SCR domains. Good sequence identities of 21-31% were found, and the three models gave consistent ring current calculations (Table 8.1). The 48-residue linker Asp195-Gly242 between the SCR-3 and vWF-A domains was not modelled. The vWF-A homology model was based on the crystal structure of complement receptor type 3, and its ring current calculation is reported elsewhere (Hinshelwood *et al.*, 1999; Hinshelwood and Perkins, 2000). In the earlier work, the construction of a homology model for the SP domain based on the crystal structure of α-chymotrypsin (Hinshelwood and Perkins, 1998) had required the extensive remodelling of 136 residues in 16 searched loops. For reason of sequence insertions this structure could not be modelled well. In contrast, in the present study,

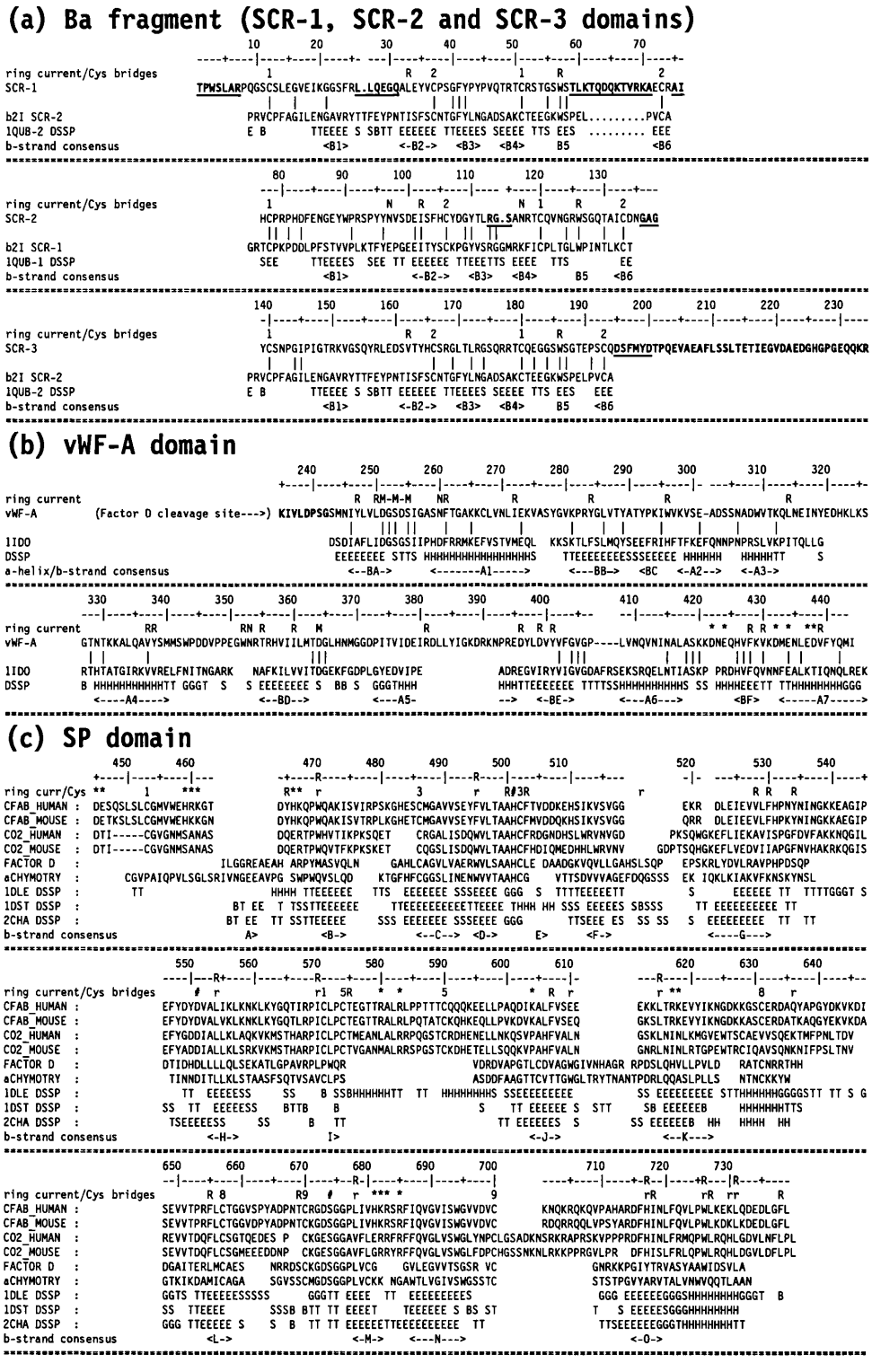


Figure 8.11 Alignment of the human factor B sequence with those of the structures used in homology modelling.

Continued overleaf.

Figure 8.11 Alignment of the human factor B sequence with those of the structures used in homology modelling. Strict residue conservation is denoted by vertical strokes. The secondary structure elements identified by DSSP are labelled as follows: E, β-strand; B single residue β-ladders; T, turn; G, 3_{10}-helix; H, α-helix. Residues involved in large predicted ring current shift interactions in factor B are denoted by R (Table 8.1).

(a) For the Ba fragment, the PDB code 1qub corresponds to the first and second SCR domains of human β2-I glycoprotein, with the secondary structure assignment taken from Molina *et al.* (1995). Residues that were remodelled as searched loops are embolded and underlined. The disulphide bridge connectivity is denoted by pairs of numbers. The linker peptide at Asp195-Gly242 is embolded.

(b) For the vWF-A domain, the PDB code 1ido corresponds to the vWF-A domain of human complement receptor type 3, with the secondary structure taken from Hinshelwood *et al.* (1999). The five residues important for Mg^{2+} coordination are denoted by M above the sequence (Lee *et al.*, 1995). The residues D422 and E424 on a surface loop and D432, D434, E437, D438, D445 and E446 on α-helix A7 (asterisked) form a prominent group of surface acidic residues.

(c) The human factor B SP sequence is compared with those found in mouse factor B and human and mouse C2 in that order from top to bottom. The PDB code 1dst corresponds to human factor D and the PDB code 2cha corresponds to bovine α-chymotrypsin. The structural alignment is taken from Jing *et al.* (2000). Residues involved in large predicted ring current shift interactions in factor D are denoted by r (Table 8.1). The labelling of the β-strands in the consensus secondary structure and the disulphide numbering is taken from Perkins & Smith (1993). The catalytic triad at His501-Asp551-Ser674 is denoted by hashes (#). The residues H459, R460 and K461 at the N-terminus, H466, K467 and R580, R583 on two surface loops, and K604, R617, K618, H681, K682, R683 and R685 in association with 4 β-strands (asterisked) form a prominent group of surface basic residues in proximity to the C-terminus of the vWF-A domain.

the two coordinate sets in the recent crystal structure of the SP domain of factor B (Jing *et al.*, 2000) resulted in satisfactory ring current calculations (Table 8.1; see below). In the case of factor D, five sets of crystal coordinates were available, from which mean ring current shifted signals were computed with a standard deviation of ± 0.19 ppm (Table 8.1).

The ring current calculations for factor B and factor D showed reasonable agreements between the total number of predicted and observed upfield shifted methyl signals (Table 8.3). Between four and five signals were predicted for the Ba fragment, to be compared with the observation of 8 signals in the same spectral region after spectral integration. Between 7 and 13 signals were predicted for the vWF-A domain, to be compared with the observation of 8 signals in the same spectral region. Between 10 and 14 signals were predicted for the SP domain of factor B, to be compared with the observation of 13-15 signals in the same spectral region. Between 10-16 signals were predicted for factor D, to be compared with the observation of 15-20 signals in the same spectral region. Despite these agreements, the positional accuracy of the relative separations between aliphatic and aromatic sidechains was not considered sufficient to permit any firm assignments of signals to a given methyl group.

For the Ba fragment, the most significant ring current shifts were predicted to arise from the structurally homologous interaction between a buried aliphatic residue in β-strand B2 and the conserved Trp residue in β-strand B5 in all three SCR domains (Figure 8.11(a)). These three pairs of residues were located in the hydrophobic core of each domain in Figure 8.12, and may correspond to the signals 1a-3a of Figure 8.5. All six residues had solvent accessibilities of 0-19%. The Ba assignment is consistent with the assignment of methyl signals in the NMR spectra of SCR domains from factor H and the ring current calculations based on the resulting factor H SCR coordinates (Norman *et al.*, 1991; Barlow *et al.*, 1991, 1992, 1993). The Ile27/Trp51, Leu32/Trp55 and Val29/Trp52 residue pairs in the individual SCR-5, SCR-15 and SCR-16 domains of factor H respectively, together with contributions from other aromatic residues, resulted in predicted upfield-shifted methyl signals that agreed with their observed NMR signals. Other observed high-field shifted signals arising from Ile9, Ile11 and Ile11 in the SCR-5, SCR-15 and SCR-16 domains respectively also agreed with

Table 8.3 Comparison between the predicted and observed upfield shifted methyl signals in the ^{1}H NMR spectrum of factor B

Protein		Signals between 0.5 and 0.0 ppm†	Signals upfield of 0.0 ppm
Ba fragment	Predicted	3 (+ up to 1 more)	1
	Observed	4	4 (signals 1a-4a)
vWF-A domain*	Predicted	5 (+ up to 6 more)	2
	Observed	7	1 (signal 2v)
SP domain	Predicted	6 (+ up to 4 more)	4
	Observed	36806	5 (signals 1s-5s)
factor D	Predicted	4 (+ up to 6 more)	6
	Observed	76259	5 (including signals 1d-3d)

Footnotes

† Since the average standard deviation of 16 predicted ring current shifted methyl signals based on five coordinate sets for factor D is 0.19 ppm, additional signals predicted to occur between 0.5 to 0.7 ppm are included in brackets as these are within error of the observed signals.

* From Chapter 7 and Hinshelwood and Perkins (2000)

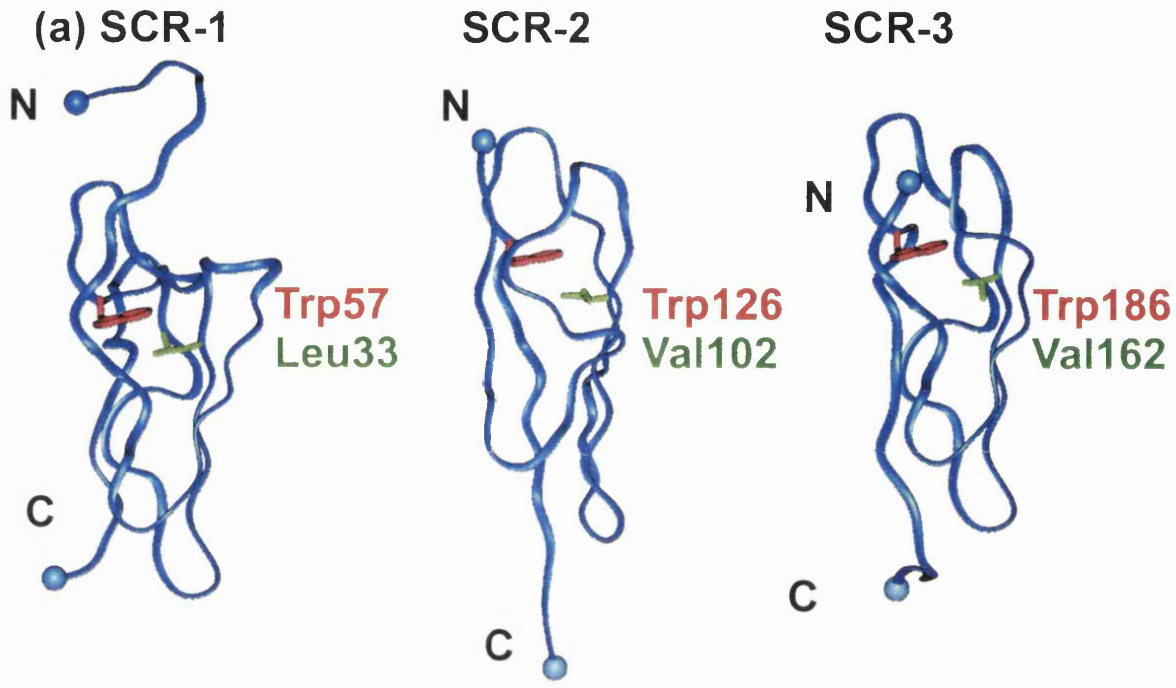


Figure 8.12 Methyl-aromatic ring interactions in homology models of the three SCR domains from the Ba fragment of factor B. All the structures are viewed in the same scale. Only residues involved with ^{1}H NMR signals predicted to be upfield of 0.5 ppm are highlighted (Table 8.1), with aliphatic sidechains in green and aromatic ones in red. The N-terminus and C-terminus of each structure are shown as blue spheres and labelled N and C respectively.

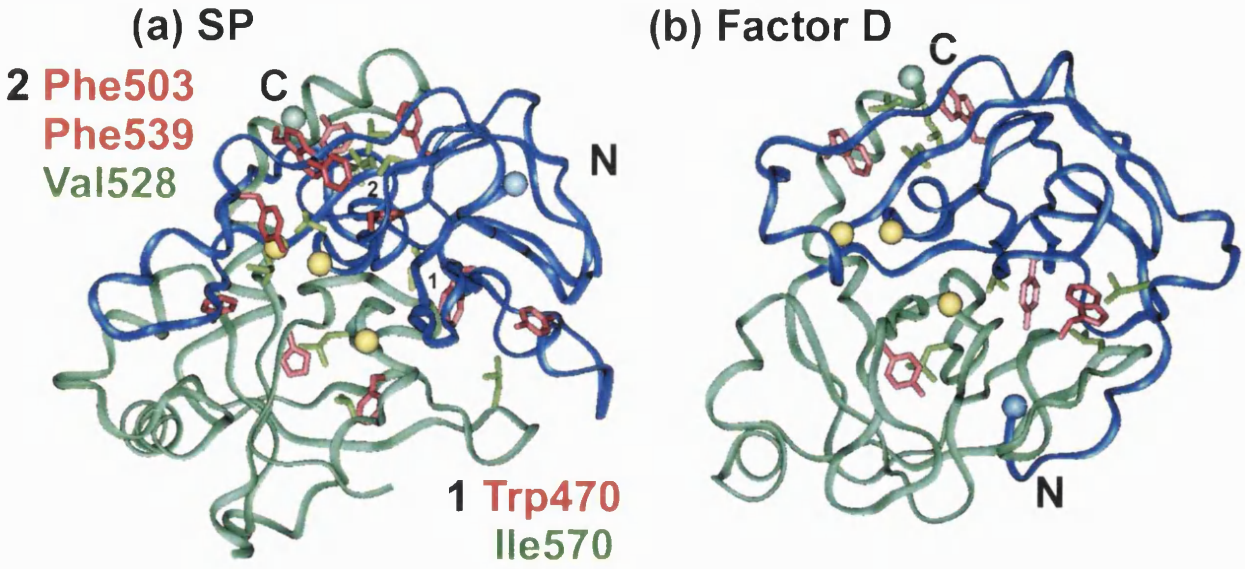


Figure 8.13 Methyl-aromatic ring interactions in homology models or crystal structures of the SP domains in factor B and factor D. Both the structures are viewed in the same scale. Only residues involved with ^{1}H NMR signals predicted to be upfield of 0.5 ppm are highlighted (Table 8.1), with aliphatic sidechains in green and aromatic ones in red. The N-terminus and C-terminus of each structure are shown as blue spheres and labelled N and C respectively.

(a) The SP domain of factor B, with the catalytic triad Asp551-His501-Ser674 shown as yellow spheres from left to right, and the N-terminal and C-terminal subdomains are shown in dark and light blue respectively. The two highest-field chemical shifts are predicted to arise from regions involving (1) Trp470-Ile570 and (2) Phe503/Phe530 and Val528, all five of which are in the upper N-terminal subdomain (dark blue) (Table 8.1). The His717-Leu678 and Phe607-Thr669 and Leu615 residues originate in the lower C-terminal domain (light blue) (Table 8.1). Note that the N-terminus lies within the N-terminal subdomain.

(b) The SP domain of factor D shown in the same view, with the catalytic triad shown as yellow spheres. The Trp141-Leu155 and Tyr228-Leu199 ring current interactions originate in the lower C-terminal subdomain. Note that the N-terminus now lies within the C-terminal subdomain.

the factor H ring current calculations. As none of these aliphatic residues occurs at homologous positions in the three Ba SCR domains, no other Ba residues giving rise to upfield-shifted signals could be consistently identified.

For the SP domain of factor B, it was of interest that 9 of the 22 most significant ring current residues occurred at structurally homologous positions with 9 of the 13 most significant ring current residues in factor D (Table 8.1). This reinforces the value of the comparison of the NMR spectra in Figures 8.5, 8.8 and 8.9 (factor B) with Figure 8.10 (factor D). The visualisation of these pairs of interactions in the two coordinate sets showed that 21 of the 22 ring current residues in factor B were buried with solvent accessibilities between 0 and 19%. Only Phe530 was solvent exposed. In factor D, all 12 ring current residues were buried with solvent accessibilities between 0-19%. In both SP structures, a cluster of these residues occurred at the N-terminal subdomain of the SP domain (dark blue in Figures 8.13(a) and 8.13(b)), and another cluster occurred in the C-terminal α-helix that flanked it (upper left corner as viewed in Figures 8.13(a) and 8.13(b)). The five highest-field shifted signals for the SP domain of factor B were predicted to arise from the N-terminal subdomain and its association with the C-terminal α-helix (Table 8.1).

8.3.7 Association of the vWF-A and SP domains

The models were examined for information on possible sources of signals 1s and 1n. The ring current and accessibility calculations and the temperature dependence studies suggest that these signals arise from fully buried residues, rather than from an interaction between two hydrophobic surfaces. Given the dependence of signals 1s/1n on the presence or absence of the vWF-A domain, the signals are likely to originate from a localised structure near the junction between the vWF-A and SP domains. Atypically for SP domains, the N-terminal peptide of the factor B SP domain that is linked with the C-terminus of the vWF-A domain remains wholly located within the N-terminal SP subdomain. This is the opposite of what is observed with the SP domain of factor D and the classic SP domains, in which the N-terminus is inserted into the C-terminal subdomain to form a salt-bridge with Asp194, which is adjacent to Ser195 of the catalytic triad (Figure 8.13). One possible ring current interaction is that

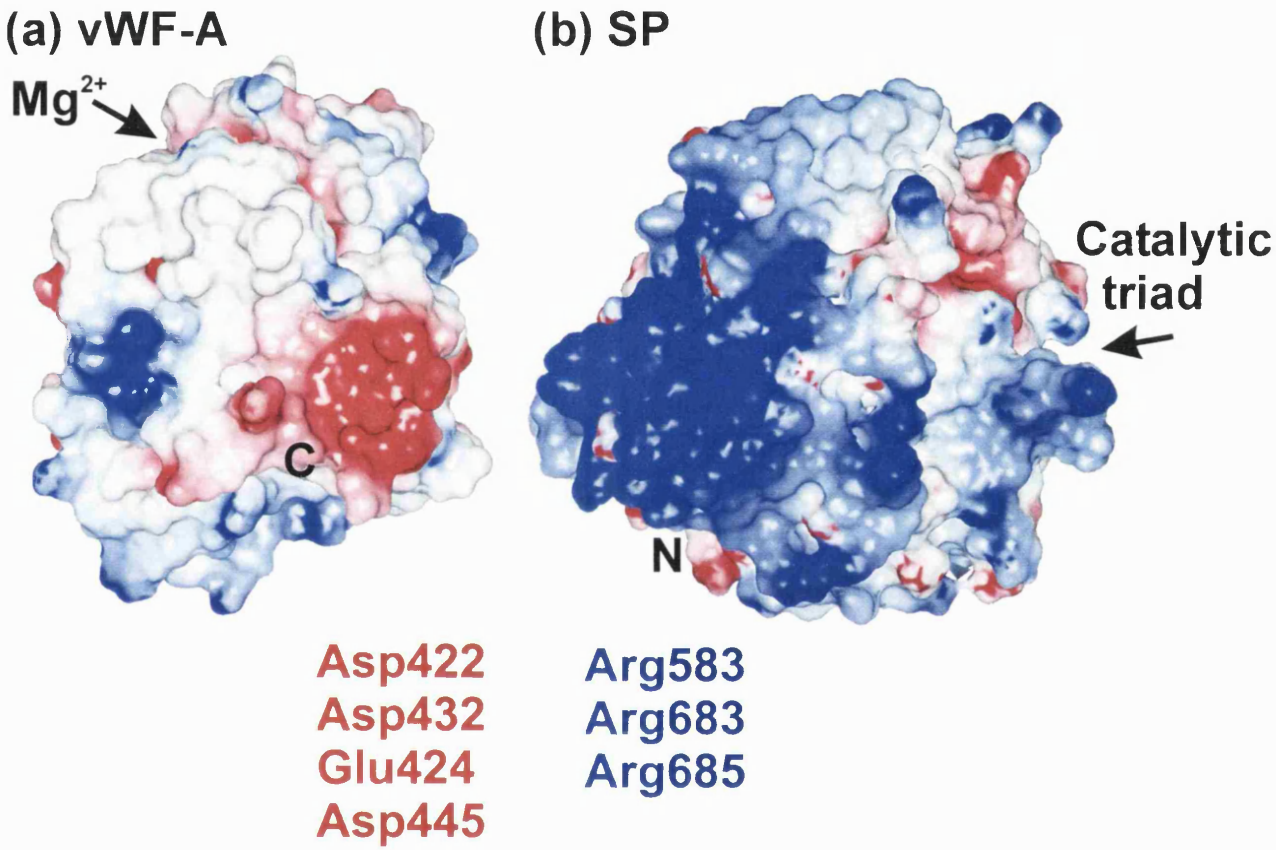


Figure 8.14 Electrostatic views of the C-terminus of the vWF-A domain and the N-terminus of the SP domain. The acidic region in the vWF-A domain and the basic region in the SP domain are highlighted, together with prominent conserved acidic and basic residues as listed. The locations of the Mg^{2+} cleft and catalytic triad are indicated.

between Trp470 and Ile570 in the N-terminal subdomain (labelled 1 in Figure 8.13(a)). This involves two loop residues in what may be a conformationally flexible region of the secondary structure, one of which is close to the N-terminus of the SP domain, and gives rise to the most highfield shifted signal for the SP domain of factor B (Table 8.1). Another involves Phe503, Phe530 and Val528 in the N-terminal subdomain, which lead to the second most highfield shifted signal, and which are close to the catalytic triad (labelled 2 in Figure 8.13(a)).

The locations of signals 1s and/or 1n to the N-terminal SP subdomain support a proposed activation mechanism for the SP domain in which its conformation is regulated by contacts with the vWF-A domain. The calculation of electrostatic surfaces for the SP domain revealed a large basic surface near the N-terminus which is defined by as many as 14 His, Lys and Arg residues (asterisked in Figure 8.11). This may complement a large acidic surface at or near the α-helix A7 on the vWF-A surface which is defined by 8 Asp and Glu residues (asterisked in Figure 8.11; see Figure 6.8 of Chapter 6). These two surfaces may provide interdomain contacts that would explain the NMR-observed interaction between the two domains in the Bb fragment (Figure 8.14). The significance of these residues is indicated by the full conservation of four of these acidic vWF-A residues (Asp422, Glu424, Asp432 and Asp445) in the human and mouse factor B and C2 sequences, but which are not conserved in five vWF-A crystal structures (Figures 6.5 and 6.6 of Chapter 6). Likewise, the basic SP residues Arg583, Arg683 and Arg685 are also fully conserved between human factor B and C2, together with three conservative replacements between Lys and Arg residues in these sequences (Figure 8.11), but these are absent from nine mammalian SP crystal structures (Figure 1 of Perkins and Smith, 1993). The presence of Arg580 and Arg583 on the largest insertion in the SP sequence (residues 575-594) and Arg683 and Arg685 on another insertion (residues 682-685) provides an explanation for the existence for these insertions in the SP domain (Figure 8.11). As α-helix A7 is conformationally mobile between the metal-bound and metal-free forms of the vWF-A domain, it can be seen that a movement in this α-helix may provide the means for the transmission of an allosteric signal between the vWF-A and SP domains (Perkins *et al.*, 1999).

8.4 Conclusions

The comparative study of the conformation-dependent upfield-shifted NMR spectra of factor B, its Ba and Bb fragments, and the individual vWF-A and SP domains has revealed the first structural evidence for an allosteric activation mechanism for the protease activity of factor B. NMR demonstrated the interaction of the Ba fragment with the vWF-A and SP domains within intact factor B, and another between the vWF-A and SP domains within the Bb fragment. It also revealed the existence of structural lability in the SP domain of both factor B and factor D, where this is higher in factor D. The proposed role of the domains in factor B in controlling its protease activity as deduced from this NMR study is summarised as follows. In factor B, the SCR domains fold against the vWF-A and SP domains to hold these apart and all 5 domains interact with each other; the active site of the SP domain is sterically masked by the SCR domains. On removal of the SCR domains by factor D cleavage, the vWF-A and SP domains form close contacts with each other in the Bb fragment, and there is a small gain in protease activity (Lambris and Müller-Eberhard, 1984). Studies of Mg^{2+}-dependent conformational changes in the vWF-A domain suggest that the vWF-A domain may undergo allosteric conformational change when bound to C3b (Hinshelwood and Perkins, 2000). The present NMR data show that the vWF-A interaction with C3b via its binding cleft can be structurally transmitted to the SP domain to regulate its proteolytic activity, possibly through electrostatic contacts (Figure 8.14).

The principal NMR observations that resulted in the deduction of allosteric interdomain interactions are summarised. The overall similarity of the NMR spectra for the Ba fragment and the vWF-A and SP domains, both when free and when associated into the Bb fragment and factor B, showed that each of the domains of factor B is independently folded. The exceptions are the changes in the signals 1s/1n in the Bb spectrum and in the signals 1a in the Ba spectrum, both of which show the occurrence of conformational changes during the assembly of factor B from its fragments (Figure 8.6). The ring current calculations showed that the changes in signals 1s/1n were most likely to arise from the region of the SP domain that is close to its covalent link with the vWF-A domain (Table 8.1; Figure 8.13). Further evidence that the five domains of factor B are associated with each other was indicated by the

reduced denaturation temperature of factor B at 50°C, while the Ba and Bb fragments denatured at 65°C and 60°C respectively. The conformational dependence of the factor B vWF-A domain on the presence and absence of metal has been demonstrated by multidisciplinary spectroscopic studies (Hinshelwood and Perkins, 2000). These observations are most simply explained by the proposed activation mechanism above. Other evidence for the interactions between the vWF-A and SP domains include observations that the linker between the vWF-A and SP domains is inaccessible to elastase proteolysis if the Bb fragment is used, but is cleaved when factor B is used (Lambris and Müller-Eberhard, 1984). The SP domain and the uncomplexed Bb fragment in the absence of C3b have low proteolytic activities towards C3 compared to that of factor B (Fishelson and Müller-Eberhard, 1984; Lambris and Müller-Eberhard, 1984; Sanchez-Corral *et al.*, 1990). These NMR and biochemical data would explain the necessity of an initial interaction between C3b and the Ba fragment during factor B activation, as it is necessary to remove the Ba-Bb contacts to permit activation to occur, and why the Bb fragment loses its catalytic activity when it dissociates from the C3b complex, as there are structural contacts between the vWF-A and SP active site clefts (Ueda *et al.*, 1987; Pryzdial and Isenman, 1988; Fishelson and Müller-Eberhard, 1984; Hourcade *et al.*, 1995).

It has been long known that the induction of catalytic activity of the serine proteases is associated with an N-terminal cleavage which induces conformational changes in and around the active site (Huber and Bode, 1978). For the complement serine proteases, the joint requirements of high specificity for the substrate and a high level of control of this activation has resulted in the development of more sophisticated and atypical activation mechanisms. The proposal of a cooperative allosteric mechanism for the activation of the SP domain of factor B falls into this category (and also for that of factor D which is activated by binding to its substrate). The recent crystal structure of the factor B SP domain is based on the same sequence used in this NMR study (Asp445-Leu739) and showed that the catalytic triad and non-specific substrate-binding site displays active conformations while the oxyanion pocket displays an inactive conformation (Jing *et al.*, 2000). The crystal structures of factor D displayed atypical conformations for all three of the catalytic triad, substrate binding site and

oxyanion pocket (Volanakis and Narayana, 1996). These crystal structures are consistent with the observation of broader linewidths in NMR spectra of factor D compared to those of the SP domain of factor B (Figures 8.5, 8.8 and 8.10). Accordingly the NMR data imply that substrate-induced conformational changes will be less important for the SP domain of factor B than for factor D. The most interesting crystallographic difference between the factor B and D SP structures is the location of all the factor B N-terminal SP residues within the N-terminal subdomain, as consequently this forms a direct molecular link with the vWF-A C-terminal α-helix A7. It has been proposed that it is difficult to visualize the activation of the SP domain as being entirely due to the translation of a cross-domain conformational change (Jing *et al.*, 2000). The NMR data however show that the transmission of conformational changes within the Bb fragment is observable. As the binding of the SP domain to C3b has been characterised, and the SP domain on its own has low proteolytic activity (Lambris and Müller-Eberhard, 1984), it is likely that C3b binding to the SP domain will complete the activation of the SP domain.

It is of interest to compare the role of the vWF-A domain in complement and integrins. In the evolution of the vWF-A domain, phylogenic analysis has shown that the factor B and C2 vWF-A domains have evolved separately from the integrin CR3 and CR4 vWF-A domains (Tuckwell, 1999). An N-terminal polypeptide cleavage occurs in the complement vWF-A domains of factor B and C2 during its physiologically-relevant cleavage, while the integrin vWF-A domains are inserted into a predicted β-propellor structure, and no cleavage occurs during its function (Leitinger and Hogg, 1999). As the function of the vWF-A domain is expected to be similar in the complement proteins and integrins, it is unlikely that the cleavage of factor B to form the Bb fragment directly leads to the activation of the SP domain. It is more likely to involve the ability of the vWF-A domain to transmit an allosteric signal (Perkins *et al.*, 1999).

Chapter 9

Epilogue

9.1 Assessment of the structural studies of factor B

The original aims of these studies were to investigate the three-dimensional structure of the vWF-A and SP domains of complement factor B. The experiments were to be primarily concerned with the production of protein crystals and determination of the molecular structures by X-ray diffraction. It was possible to express the vWF-A domain of factor B in recombinant systems and the isolated protein domain existed as a discrete folded unit. The observation of a compact solution structure was in agreement with homologous vWF-A domains (Lee *et al.*,1995a; Lee *et al.*,1995b) and was consistent with the observation of a three lobed structure for intact factor B by electron microscopy. The poor solubility of the vWF-A-218 domain preparation made it difficult to work with and sequence analysis and homology modelling allowed the rational development of a modified recombinant system. The mutation of the vWF-A-218 domain to the vWF-A-222 domain was performed which proved successful and resulted in a more soluble vWF-A domain. The SP domain was purified from native factor B pre-treated with elastase. The characterisation of these protein preparations suggested that they would be good candidates for structural studies and so a considerable amount of time and effort was devoted to the pursuit of obtaining protein crystals. Large quantities of the proteins were consumed in this process and achieved only limited information on various biochemical and physical properties of the protein preparations. Some crystals of the vWF-A-222 domain were obtained after four year's work, but these were not of suitable quality for X-ray diffraction (Chapter 4). This work did lead onto reveal useful information relating to the hydrodynamic properties and functional activity of the vWF-A domain (Chapter 5; Williams *et al.*, 1999). The crystallographic investigations remained frustratingly hindered, even though the results of this thesis provided many useful insights into published protein X-ray crystal and NMR structures.

Much of the material of this thesis has been focussed on the factor B vWF-A domain which has led to considerable insights into the activity of this domain. The vWF-A domain has been of great general biochemical interest by reason of its widespread occurrence and its varied functions in many different proteins of the immune system and the extracellular matrix as well as in blood coagulation (Colombatti and Bonaldo, 1991). During the past 5 years

significant progress has been reported on the understanding of vWF-A domain structures. When studies for this thesis began, no atomic structures were available for this domain type, nor was the topology of the vWF-A domain known. An interesting aspect of structural studies on the vWF-A domain was the success of prediction methods in identifying the vWF-A protein fold before the crystal structure had been determined (Perkins *et al.*, 1994; Edwards and Perkins, 1995; Lee *et al.*, 1995; Edwards and Perkins, 1996). By mid-1999, the crystal structures for five different vWF-A domains in CR3, LFA-1, VLA-2 and the A1 and A3 domains of vWF had been determined. The accuracy of the structure prediction results from significant progress in our understanding of protein structures, but the controversy over the metal-binding structural properties of the vWF-A domain illustrates the limitations imposed on the protein model by these structure determination techniques.

One part of this thesis describes the development of the SELDIAMS method, which coupled with homology modelling has provided a powerful technology used to identify one role of the vWF-A domain in factor B as being to bind to C3b via its active site cleft (Chapter 6; Hinshelwood *et al.*, 1999). These techniques add to repertoire of the structural biologist and can now be employed to assist in the analyses of other protein-protein interactions for the identification of molecular recognition sites.

The combined use of independent data sets from CD, FT-IR and ^{1}H NMR spectroscopy has provided the first unequivocal structural evidence that a vWF-A domain has two distinct conformations in the presence and absence of metal. The joint application of CD and FT-IR spectroscopy offers complementary views of the vWF-A secondary structure. Since all three methods are based on solution studies, no ambiguities were caused by the presence of a crystallographic lattice. In application to factor B, the Mg^{2+}-induced conformational variability is likely to be important for the activation and function of the C3 convertase when the Bb fragment is complexed with C3b. The proposal that an allosteric signal can be transmitted from the active site cleft to the opposite surface of the vWF-A domain is therefore consistent with what is known for vWF-A domains to date, and may explain how the serine protease domain of factor B is activated in the C3bBb complex.

(Chapter 7).

The comparative study of the conformation-dependent upfield-shifted NMR spectra of factor B, its Ba and Bb fragments, and the individual vWF-A and SP domains, has revealed structural evidence for an allosteric activation mechanism for the protease activity of factor B. The proposed role of the domains in factor B is summarised as follows. In factor B, the SCR domains fold against the vWF-A and SP domains to hold these apart and all 5 domains interact with each other; the active site of the SP domain is sterically masked by the SCR domains. On removal of the SCR domains by factor D cleavage, the vWF-A and SP domains form close contacts with each other in the Bb fragment, and there is a small gain in protease activity. The present NMR data show that the vWF-A interaction with C3b via its binding cleft can be structurally transmitted to the SP domain to regulate its proteolytic activity.

With the benefit of hindsight it is possible to evaluate this PhD project on the basis of the general costs balanced against the wealth of information which was derived. A disproportionate amount of time and resources were involved in the crystallisation phase of the project, which may be considered to have failed because no X-ray data were obtained. However, if an X-ray structure determination were to follow this work, this thesis would have presented necessary and worthwhile experiments. Much experience was gained during these experiments to provide the groundwork for a solid basis in protein biochemistry which laid the foundations for the experiments and data presented in this thesis. The biochemical studies of Chapters 5 and 6 could be considered good value because they were conservative in their consumption of protein, the proteins were readily available at the time of the experiments to expedite the data collection, and the data analysis and interpretation was straightforward. This was in contrast to the spectroscopic studies in Chapters 7 and 8 which were expensive in their use of protein and required a great deal of time for data collection, data processing and the interpretation of the results. Fortunately, each of Chapters 5, 6, 7 and 8 provided novel material suitable for publication in scientific journals which rendered them all good value in this context.

Protein structure is generally well conserved between proteins containing homologous domains, but the function of the individual domain is less well conserved (Tuckwell, 1999). However, there are some domain types that consistently have a function associated with their structure. The mammalian serine protease domain is a good example of this and structural studies have identified the basic mechanisms employed by these homologous domains. It seems likely that the vWF-A domain structure has associated conserved functional roles of ligand binding and acting as an allosteric regulator of protein function.

9.2 Future work

There is still much progress required to develop our knowledge of protein structure to gain a fuller understanding of the influence and role of vWF-A domains in the structure and function of their host. During these studies some time was spent on the development of other recombinant protein expression systems including those for the Ba fragment, the SP domain and factor B (data not presented). The use of these recombinant systems would be valuable to aid in the development of our understanding of complement and multi-domain proteins in general. An array of techniques has been presented in this thesis which could be exploited to characterise new recombinant proteins. With this in mind it is possible to design a variety of mutant proteins to examine the role of specific residues in ligand binding (SELDIAMS, BIACORE, analytical ultracentrifuge), and their influence on the structure/allostery by spectroscopy (NMR, CD, FTIR, fluorescence). Molecular graphics would be a significant aid to the design of mutants and the interpretation of results.

Through the studies presented in this thesis it has been shown that the two recombinant factor B vWF-A proteins exhibit complex behaviour. The protein domain is prone to aggregation at high concentration and the protein is fairly unstable in the absence of metal. This instability may be attributed to some extent to the truncation of the C-terminal helix (even in vWF-A-222) and a long N-terminal extension. Some improvement in the stability was achieved through the extension of the C-terminal helix. For the other vWF-A domains that have been crystallised to date, it was necessary to remove large N or C terminal extensions before protein crystals could be grown. A significant improvement in the stability

of the factor B vWF-A domain was achieved by the 4 amino acid extension. This could be developed further to try to produce material suitable for crystallisation or multi-dimensional NMR (increased solubility). These experiments could include:

(a) to truncate the vWF-A N-terminus back to the factor D cleavage site

(b) to extend the vWF-A C-terminal α-helix

(c) to replace specific methyl or aromatic residues for the purpose of NMR assignment or to determine a role in ligand binding/ allostery

NMR spectroscopy can often provide an alternative for the determination of protein structure which is in many ways complementary to the X-ray crystallographic technique. Typically, structural determination of proteins by NMR requires a suitable recombinant expression system to yield large amounts of highly soluble protein. Thus a full NMR structural determination of the factor B protein preparations described in this thesis was not feasible. It was hoped that the improved solubility achieved with the vWF-A-222 domain would make this more suitable for such a project but it would be sensible to explore the possibility of further mutation experiments to increase the solubility and stability of this domain preparation so that mM concentrations could be achieved. A combination of isotope enrichment techniques, expression in deuterium oxide and multidimensional NMR spectroscopy could then be used to determine the atomic structure of this domain. An NMR structure of this domain would be useful to allow the allosteric properties to be investigated further.

The contribution made to the stability of the Bb fragment and factor B by the metal could be investigated. It would be interesting to investigate whether a peptide ligand or peptide activator can be determined for the vWF-A domain. For example experiments could be conducted to examine whether the KI peptide has an effect on the structure or function of the vWF-A, Bb and factor B (NMR, CD, FTIR, SELDIAMS), or a peptide could be produced which would bind to the active site cleft of the vWF-A domain. It would be interesting to follow up these studies using similar analyses with other recombinant vWF-A domains to discover if what we have seen in our recombinant proteins is a general property of the vWF-A domain or a special case for complement factor B.

Reference List

Abath, F.G. and Simpson, A.J. (1991). A simple method for the recovery of purified recombinant peptides from glutathione-S-transferase-fusion proteins. *BioTechniques* **10**, 178.

Alberts, B, Bray, D, Lewis, J, Raff, M, Roberts, K, and Watson, JD (1989). *Molecular Biology of the Cell.* Second edition. Garland publishing, New York.

Alex, R., Sozeri, O., Meyer, S. and Dildrop, R. (1992). Determination of the DNA sequence recognized by the bHLH-zip domain of the N-myc protein. *Nucleic Acid Research* **20**, 2257-2263.

Allan, R., Rodrick, M., Knobel, H. R., and Isliker, H. (1979). Inhibition of the interaction between the complement component Clq and immune complexes. *International Archives of Allergy and Applied Immunology* **58,** 2,140-148.

Alper, C. A., Colten, H. R., Rosen, F. S., Rabson, A. R., Macnab, G. M., and Gear, J. S. (1972). Homozygous deficiency of C3 in a patient with repeated infections. *Lancet* **2,** 788,1179-1181.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215,** 403-410.

Anderson, C. M. (1986). Glycoprotein structure of components C2 and factor B of the human complement system. *D. Phil Thesis*, University of Oxford.

Amann, E. and Brosius, J. (1985). ATG vectors for regulated high-level expression of cloned genes in escherichia-coli. *Gene* **40,** 183-190.

Amann, E., Ochs, B., and Abel, K. J. (1988). Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in escherichia-coli. *Gene* **69,** 301-315.

Ambrus, J. L., Peters, M. G., Fauci, A. S., and Brown, E. J. (1990). The Ba fragment of complement factor B inhibits human lymphocyte-b proliferation. *Journal of Immunology.* **144,** 1549-1553.

Aoki, D., Appert, H.E., Johnson, D., Wong, S.S. and Fukuda, M.N. (1990). Analysis of the substrate binding sites of human galactosyltransferase by protein engineering. *EMBO Journal* **9**, 3171-3178.

Argraves, W. S., Tran, H., Burgess, W. H., and Dickerson, K. (1990). Fibulin is an extracellular matrix and plasma glycoprotein with repeated domain structure. *Journal of Cell Biology* **111,** 6 Pt 2,3155-3164.

Arlaud, G. J. and Gagnon, J. (1981). Clr and Cls subcomponents of human complement: two serine proteinases lacking the 'histidine-loop' disulphide bridge. *Bioscience Reports* **1,** 10,779-784.

Arlaud, G. J., Gagnon, J., Villiers, C. L., and Colomb, M. G. (1986). Molecular characterization of the catalytic domains of human complement serine protease C1r. *Biochemistry* **25,** 18,5177-5182.

Arlaud, G. J., Thielens, N. M., and Aude, C. A. (1989). Structure and function of C1r and C1s: current concepts. *Behring Institute Mitteilungen* 84,56-64.

Arlaud, G. J., Volanakis, J. E., Thielens, N. M., Narayana, S. V. L., Ross, V. and Xu, Y. (1998). The atypical serine proteases of the complement system. *Adv. Immunol.* **69**, 249-307.

Atkins, P. W. (1987). *Physical Chemistry*. Third Edition. Oxford University Press, Oxford and New York.

Atkinson, J. P. and Farries, T. (1987). Separation of self from non-self in the complement-system. *Immunology Today* **8,** 212-215.

Aude, C. A., Lacroix, M. B., Arlaud, G. J., Gagnon, J., and Colomb, M. G. (1988). Differential accessibility of the carbohydrate moieties of C1s-C1r-C1r-C1s, the catalytic subunit of human C1. *Biochemistry* **27,** 23,8641-8648.

Baines, J.D. and Roizman, B. (1993). The UL_{10} gene of herpes simplex virus 1 encodes a novel viral. *Journal of Virology.* **67**, 1441-1452.

Bairoch, A. (1991). Prosite - a dictionary of sites and patterns in proteins. *Nucleic Acids Research* **19,** 2241-2245.

Baldwin, E. T., Sarver, R. W., Bryant, G. L., Curry, K. A., Fairbanks, M. B., Finzel, B. C., Garlick, R. L., Heinrikson, R. L., Horton, N. C., Kelley, L. L. C., Mildner, A. M., Moon, J. B., Mott, J. E., Mutchler, V. T., Tomich, C. S. C., Watenpaugh, K. D., and Wiley, V. H. (1998). Cation binding to the integrin CD11b I domain and activation model assessment. *Structure* **6,** 923-935.

Barlow, P. N., Baron, M., Norman, D. G., Day, A. J., Willis, A. C., Sim, R. B., Campbell, and ID. (1991). Secondary structure of a complement control protein module by two-dimensional ^{1}H NMR. *Biochemistry* **30,** 4,997-1004.

Barlow, P. N., Norman, D. G., Steinkasserer, A., Horne, T. J., Pearce, J., Driscoll, PC, Sim, R. B., and Campbell, I. D. (1992). Solution structure of the fifth repeat of factor H: a second example of the complement control protein module. *Biochemistry* **31,** 14,3626-3634.

Beavis, R. C. and Chait, B. T. (1990). Rapid, sensitive analysis of protein mixtures by mass spectrometry. *Proceedings of The National Academy of Sciences of The United States of America* **87,** 17,6873-6877.

Beavis, R. C. and Chait, B. T. (1990). High-accuracy molecular mass determination of proteins using matrix-assisted laser desorption mass spectrometry. *Analytical Chemistry* **62,** 17,1836-1840.

Beavis, R. C. and Chait, B. T. (1996). Matrix-assisted laser desorption ionization mass-spectrometry of proteins. *Methods in Enzymology* **270,** 519-551.

Becherer, J. D. and Lambris, J. D. (1988). Identification of the C3b receptor-binding domain in third component of complement. *Journal of Biological Chemistry* **263,** 28,14586-14591.

Berger, M., Balow, J. E., Wilson, C. B., and Frank, M. M. (1983). Circulating immune complexes and glomerulonephritis in a patient with congenital absence of the third component of complement. *New England Journal of Medicine* **308,** 17,1009-1012.

Bhakdi, S. and Tranumjensen, J. (1991). Complement lysis - a hole is a hole. *Immunology Today* **12,** 318-320.

Bienkowska, J., Cruz, M. A., Handin, R. I., and Liddington, R. C. (1996). Crystal structure of the von Willebrand factor A3 domain. *Circulation* **94,** 1266.

Bienkowska, J., Cruz, M., Atiemo, A. and Liddington, R. (1997). The von Willebrand A3 domain does not contain a metal ion-dependent adhesion site motif. *Journal of Biological Chemistry* **272**, 25162-25167.

Birnboim, H. C. and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research* **7,** 6 ,1513-1523.

Bokisch, V. A. and Müller-Eberhard, H. J. (1970). Anaphylatoxin inactivator of human plasma: its isolation and characterization as a carboxypeptidase. *Journal of Clinical Investigation* **49,** 12,2427-2436.

Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C. Heyneker, H. L. and Boyer, H. W. (1977). Constuction and characterization of new cloning vehicles. II. a multipurpose cloning system. *Gene*, **2**, 95-113.

Bomstein, Y. and Fishelson, Z. (1997). Enhanced sensitivity of P-glycoprotein-positive multidrug resistant tumor cells to complement-mediated lysis. *European Journal of Immunology* **27,** 2204-2211.

Bork, P. (1991). Shuffled domains in extracellular proteins. *FEBS Letters* **286,** 1-2,47-54.

Bork, P. and Beckmann, G. (1993). The CUB domain. A widespread module in developmentally regulated proteins. *Journal of Molecular Biology* **231,** 2,539-545.

Bork, P., Downing, A. K., Kieffer, B., and Campbell, I. D. (1996). Structure and distribution of modules in extracellular proteins. *Quarterly Reviews of Biophysics* **29,** 2,119-167.

Borsos, T. (1989). Immune complex mediated activation of the classical complement pathway. *Behring Institute Mitteilungen* 84,93-101.

Boulay, F., Mery, L., Tardif, M., Brouchon, L., and Vignais, P. (1991). Expression cloning of a receptor for C5a anaphylatoxin on differentiated HL-60 cells. *Biochemistry* **30,** 12,2993-2999.

Bouma, B., de Groot, P. G., van den Elsen, J. M. H., Ravelli, R. B. G., Schouten, A., Simmelink, M. J. A., Derksen, R. H. W. M., Kroon, J. and Gros, P. (1999). Adnesion mechnasim of human β2-glycoprotein I to phospholipids based on its crystal structure. *EMBO Journal.* **18**, 5166-5174.

Bowie, J. U. and Eisenberg, D. (1993). Inverted protein structure prediction. *Current Opinioins in Structural Biology* **3**, 437-444.

Branden, C. and Tooze, T. (1999). Introduction to protein structure. Second edition. Garland publishing.

Brass, A., Kadler, K. E., Thomas, J. T., Grant, M. E., and Boot-Handford, R. P. (1992). The fibrillar collagens, collagen VIII, collagen X and the C1q complement proteins share a similar domain in their C-terminal non-collagenous regions. *FEBS Letters* **303,** 2-3,126-128.

Brodsky-Doyle, B., Leonard, K. R., and Reid, K. B. (1976). Circular-dichroism and electron-microscopy studies of human subcomponent C1q before and after limited proteolysis by pepsin. *Biochemical Journal* **159,** 2,279-286.

Brown, S. C., Mueller, L., and Jeffs, P. W. (1989). ^{1}H NMR assignment and secondary structural elements of human transforming growth factor alpha. *Biochemistry* **28,** 2,593-599.

Bundi, A. and Wüthrich, K. (1978). ^{1}H NMR parameters of the common amino acid residues measured in aqueous solutions of the linear tetrapeptides H-Gly-Gly-X-L-Ala-OH. *Biopolymers*, **18**, 285-297.

Bycroft, M. and Fersht, A. R. (1988). Assignment of histidine resonances in the 1H NMR

(500 MHz) spectrum of subtilisin BPN' using site-directed mutagenesis. *Biochemistry,* **27**, 7390-7394.

Campbell, R. D., Bentley, D. R., and Morley, B. J. (1984). The factor B and factor C2 genes. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **306,** 367-378.

Campbell, R. D. and Bentley, D. R. (1985). The structure and genetics of the C2 and factor B genes. *Immunological Reviews* **87,** 19-37.

Caporale, L. H., Gaber, S. S., Kell, W., and Gotze, O. (1981). A Fluorescent Assay for Complement Activation. *Journal of Immunology* **126,** 1963-1965.

Casasnovas, J. M., Larvie, M. and Stehle, T. (1999). Crystal structure of two CD46 domains reveals an extended measles virus-binding surface. *EMBO Journal.* **18,** 2911-2922.

Celikel, R., Varughese, K. I., Madhusan, Yoshioka, A., Ware, J. and Ruggeri, Z. M. (1998). Crystal structure of the von Willebrand factor A1 domain in complex with the function-blocking NMC-4 Fab. *Nature Structure Biology* **5**, 189-194.

Chattopadhyay, D., Barchue, J., Ma, Y., Volanakis, J. E., and Narayana, S. V. L. (1998). Structural study of the von Willebrand factor type A domain of human complement factor B. *Molecular Immunology* **35,** 58.

Christie, D. L. And Gagnon, J. (1982). Isolation, characterization and n-terminal sequences of the cnbr- cleavage peptides from human-complement factor B - localization of a free thiol-group and a sequence defining the site cleaved by factor D. *Biochemical Journal* **201,** 555-567.

Christie, D. L. and Gagnon, J. (1983). Amino acid sequence of the Bb fragment from complement factor B - Sequence of the major cyanogen bromide-cleavage peptide (CB-II) and completion of the sequence of the Bb fragment. *Biochemical Journal* **209,** 61-70.

Campbell, R. and Dwek, R. (1984). *Biological spectroscopy.* The Benjamin/Cummings publishing company.

Carrell, R. and Travis, J. (1985). Alpha-1-antitrypsin and the serpins - variation and countervariation. *Trends in Biochemical Sciences* **10,** 20-24.

Catterall, C. F., Lyons, A., Sim, R. B., Day, A. J., and Harris, T. J. (1987). Characterization of primary amino acid sequence of human complement control protein factor I from an analysis of cDNA clones. *Biochemical Journal* **242,** 3,849-856.

Celikel, R., Varughese, K. I., Madhusan, Yoshioka, A., Ware, J. and Ruggeri, Z. M. (1998). Crystal structure of the von Willebrand factor A1 domain in complex with the function-blocking NMC-4 Fab. *Nature Structure Biology* **5**, 189-194.

Cheng, Y.S., McGowan, M.H., Kettner, C.A., Schloss, J.V., Erickson-Viitanen, S. and Yin, F.H. (1990). High-level synthesis of recombinant HIV-1 protease and the recovery of active enzyme from inclusion bodies. *Gene* **87**, 243-248.

Chou, P. Y. and Fasman, G. D. (1978). Prediction of the secondary structure of proteins from their amino acid sequence. *Advan. Enzymol. Relat. Areas Mol. Biol.* **47**, 45-148.

Choi-Miura, N. H., Takahashi, Y., Nakano, Y., Tobe, T., and Tomita, M. (1992). Identification of the disulfide bonds in human plasma protein SP-40,40 (apolipoprotein-J). *Journal of Biochemistry* **112,** 4,557-561.

Choi, N. H., Mazda, T., and Tomita, M. (1989). A serum protein SP40,40 modulates the formation of membrane attack complex of complement on erythrocytes. *Molecular Immunology* **26,** 9,835-840.

Chothia, C. (1975). Structural invariants in protein folding. *Nature* **254,** 5498,304-308.

Chothia, C. and Janin, J. (1982). Orthogonal packing of beta-pleated sheets in proteins. *Biochemistry* **21,** 17,3955-3965.

Chu, M. L., Pan, T. C., Conway, D., Kuo, H. J., Glanville, R. W., Timpl, R., Mann, K., and Deutzmann, R. (1989). Sequence analysis of alpha 1(VI) and alpha 2(VI) chains of human type VI collagen reveals internal triplication of globular domains similar to the A domains of von Willebrand factor and two alpha 2(VI) chain variants that differ in the carboxy terminus. *EMBO Journal* **8,** 7,1939-1946.

Codd, R., Cox, G.B., Guss, J.M., Solomon, J.G. and Webb, D. (1992). The expressioin, purification and crystallization of the e subunit of the F_1 portion of the ATPase of *Escherichia coli. Journal of Molecular Biology* **228**, 306-309.

Cohen, G. H., Silverton, E. W. and Davies, D. R. (1981). Refined crystal structure of γ-chymotrypsin at 1.9 A resolution. Comparison with other pancreatic serine proteases. *Journal of Molecular Biology,* **148**, 449-479.

Colomb, M. G., Arlaud, G. J., and Chesne, S. (1989). Models for C1. Tools or toys? The real biological challenge. *Behring Institute Mitteilungen* 84,121-128.

Colombatti, A. and Bonaldo, P. (1991). The superfamily of proteins with von Willebrand factor type A-like domains: one theme common to components of extracellular matrix, hemostasis, cellular adhesion, and defense mechanisms. *Blood* **77,** 11,2305-2315.

Cole, L. B., Chu, N., Kilpatrick, J. M., Volanakis, J. E., Narayana, S. V. L. and Babu, Y. S. (1997). Structure of diisopropyl flurophosphate-inhibited factor D. *Acta Cryst.* **D53**, 143-150.

Comp, P. C., Nixon, R. R., Cooper, M. R., and Esmon, C. T. (1984). Familial protein S deficiency is associated with recurrent thrombosis. *Journal of Clinical Investigation* **74,** 6,2082-2088.

Cooney, K. A. and Ginsburg, D. (1996). Comparative analysis of type 2b von Willebrand disease mutations: Implications for the mechanism of von Willebrand factor binding to platelets. *Blood*, **87**, 2322-2328.

Cooke, R. M., Wilkinson, A. J., Baron, M., Pastore, A., Tappin, M. J., Campbell, I. D., Gregory, H., and Sheard, B. (1987). The solution structure of human epidermal growth factor. *Nature* **327,** 6120,339-341.

Cooper, N. R. (1985). The classical complement pathway: activation and regulation of the first complement component. *Advances in Immunology* **37,** 151-216.

Corbi, A. L., Miller, L. J., O'Connor, K., Larson, R. S., and Springer, T. A. (1987). cDNA cloning and complete primary structure of the alpha subunit of a leukocyte adhesion glycoprotein, p150,95. *EMBO Journal* **6,** 13,4023-4028.

Corbi, A. L., Kishimoto, T. K., Miller, L. J., and Springer, T. A. (1988). The human leukocyte adhesion glycoprotein Mac-1 (complement receptor type 3, CD11b) alpha subunit. Cloning, primary structure, and relation to the integrins, von Willebrand factor and factor B. *Journal of Biological Chemistry* **263,** 25,12403-12411.

Cowan, S. W., Newcomer, M. E., and Jones, T. A. (1990). Crystallographic refinement of

human serum retinol binding protein at 2A resolution. *Proteins* **8,** 1,44-61.

Creighton, T. E. (1993). *Proteins: structures and molecular properties*. Second edition. Freeman, New York.

Cruz, M. A., Emsley, J., Diacovo, T., Liddington, R., and Handin, R. I. (1997). Crystal structure of the von Willebrand factor A1 domain and mapping of its glycoprotein Ib/IX binding site. *Blood* **90,** 1907.

Daha, M. R., Fearon, D. T., and Austen, K. F. (1976). C3 requirements for formation of alternative pathway C5 convertase. *Journal of Immunology* **117,** 2,630-634.

Dahlback, B., Smith, C. A., and Müller-Eberhard, H. J. (1983). Visualization of human C4b-binding protein and its complexes with vitamin K-dependent protein S and complement protein C4b. *Proceedings of the National Academy of Sciences of the United States of America* **80,** 11,3461-3465.

Daly, N. L., Scanlon, M. J., Djordjevic, J. T., Kroon, P. A., and Smith, R. (1995). Three-dimensional structure of a cysteine-rich repeat from the low-density lipoprotein receptor. *Proceedings of the National Academy of Sciences of the United States of America* **92,** 14,6334-6338.

Damerau, B. (1987). Biological activities of complement-derived peptides. *Reviews of Physiology Biochemistry and Pharmacology* **108,** 151-206.

Dangott, L. J., Jordan, J. E., Bellet, R. A., and Garbers, D. L. (1989). Cloning of the mRNA for the protein that crosslinks to the egg peptide speract. *Proceedings of the National Academy of Sciences of the United States of America* **86,** 7,2128-2132.

Danley, D.E., Geoghegan, K.F., Scheld, K.G., Lee, S.E., Merson, J.R., Hawrylik, S.J., Rickett, G.A., Ammirati, M.J. and Hobart, P.M. (1989). Crystallizable HIV-1 protease derived from expression of the viral *pol* gene in *Escherichia coli*. *Biochem. Biophys. Res. Comm.* **165**, 1043-1050

Daoudaki, M. E., Becherer, J. D., And Lambris, J. D. (1988). A 34-amino acid peptide of the third component of complement mediates properdin binding. *Journal of Immunology* **140,** 1577-1580.

Davies, A., Simmons, D. L., Hale, G., Harrison, R. A., Tighe, H., Lachmann, P. J., and Waldmann, H. (1989). CD59, an ly-6-like protein expressed in human lymphoid-cells, regulates the action of the complement membrane attack complex on homologous cells. *Journal of Experimental Medicine* **170,** 637-654.

Dayhoff, M. O., Schwartz, R. M. and Orcutt, B. C. (1978). *Atlas of Protein Sequence and Structure.* National Biomedical Research Foundation, Washington DC.

Deboer, H. A., Comstock, L. J., and Vasser, M. (1983). The tac promoter - a functional hybrid derived from the trp and lac promoters. *Proceedings of The National Academy of Sciences of The United States of America-biological Sciences* **80,** 21-25.

de Bruijn, M. H. and Fey, G. H. (1985). Human complement component C3: cDNA coding sequence and derived primary structure. *Proceedings of the National Academy of Sciences of the United States of America* **82,** 3,708-712.

Derom, C., Gheysen, D. and Fiers, W. (1982). High-level synthesis in *Escherichia coli* of the SV40 small-t antigen under contro of the lambda p_L promoter. *Gene*, **17**, 45-54.

DiScipio, R. G. and Hugli, T. E. (1989). The molecular architecture of human complement

component C6. *Journal of Biological Chemistry* **264,** 27,16197-16206.
DiScipio, R. G. (1981). The binding of human-complement proteins C5,factor B, beta-1H and properdin to complement fragment C3b on zymosan. *Biochemical Journal* **199,** 485-496.
DiScipio, R. G. (1981). The conversion of human-complement component C5 into fragment C5b by the alternative-pathway C5 convertase. *Biochemical Journal* **199,** 497-504.
Dodds, A. W., Sim, R. B., Porter, R. R., and Kerr, M. A. (1978). Activation of the first component of human complement (C1) by antibody-antigen aggregates. *Biochemical Journal* **175,** 2,383-390.
Dodds, A. W. (1993). Small-scale preparation of complement components C3 and C4. *Methods in Enzymology* **223,** 46-61.
Dodds, A. W. and Petry, F. (1993). The phylogeny and evolution of the first component of complement, C1. *Behring Institute Mitteilungen* 93,87-102.
Dodds, A. W. and Day, A. J. (1993). The phylogeny and evolution of the complement system. Whaley, K., Loos, M., and Weiler, J. M. Complement in Health and Disease. **second,** 39-88. Dordrecht, Netherlands, Kluwer academic publishers.
Doolittle, R. F. (1995). The multiplicity of domains in proteins. *Annual Review of Biochemistry* **64,** 287-314.
Drake, A. F. (1994). Circular Dichroism in *Methods in Molecular Biology* (Jones, C. *et al.*, eds.) **22**, pp219-244. Humana Press, Totowa.
Drickamer, K. (1993). Evolution of Ca(2+)-dependent animal lectins. *Progress in Nucleic Acid Research and Molecular Biology* **45,** 207-232.
Driscoll, P.C., Cyster, J.G., Campbell, I.D. and Williams, A.F. (1991). Structure of domain 1 of rat T lymphocyte CD2 antigen. *Nature (London)* **353**, 762-765.
Duffaud, G.D., March, P.E. and Inouye, M. (1987). Expression and secretion of foreign proteins in *Escerichia coli. Methods Enzymol.* **153**, 492-507.
Edwards, Y. J. K. and Perkins, S. J. (1995). The protein fold of the von-Willebrand-factor type-A domain is predicted to be similar to the open twisted beta-sheet flanked by alpha-helices found in human ras-p21. *FEBS Letters* **358,** 283-286.
Edwards, Y. J. and Perkins, S. J. (1996). Assessment of protein fold predictions from sequence information: the predicted alpha/beta doubly wound fold of the von Willebrand factor type A domain is similar to its crystal structure. *Journal of Molecular Biology* **260,** 2,277-285.
Emsley, J., King, S. L., Bergelson, J. M., and Liddington, R. C. (1997). Crystal structure of the I domain from integrin alpha 2 beta 1. *Journal of Biological Chemistry* **272,** 28512-28517.
Emsley, J., Cruz, M., Handin, R., and Liddington, R. (1998). Crystal structure of the von Willebrand factor A1 domain and implications for the binding of platelet glycoprotein Ib. *Journal of Biological Chemistry* **273,** 10396-10401.
Engh, R. A. and Huber, R. (1991). Accurate bond and angle parameters for x-ray protein-structure refinement. *Acta Crystallographica Section a* **47,** 392-400.
Entwistle, R. A. and Furcht, L. T. (1988). C1q component of complement binds to fibrinogen and fibrin. *Biochemistry* **27,** 1,507-512.
Esser, A. F. (1991). Big mac attack - complement proteins cause leaky patches.

Immunology Today **12,** 316-318.

Evans, J. N. S. (1995) *Biomolecular NMR spectroscopy*. Oxford university press, Oxford and New York..

Fairbanks, G., Steck, T. L., and Wallach, D. F. (1971). Electrophoretic analysis of the major polypeptides of the human erythrocyte membrane. *Biochemistry* **10,** 13 ,2606-2617.

Fairbanks, M. B., Pollock, J. R., Prairie, M. D., Scahill, T. A., Baczynskyj, L., Heinrikson, R. L. and Stockman, B. J. (1995). Purification and structural charcterization of the CD11b/CD18 integrin α subunit I domain reveals a folded conformation in solution. *FEBS. Lett.* **369**, 197-201.

Farries, T. C., Finch, J. T., Lachmann, P. J., and Harrison, R. A. (1987). Resolution and analysis of 'native' and 'activated' properdin. *Biochemical Journal* **243,** 2,507-517.

Farries, T. C., Lachmann, P. J., and Harrison, R. A. (1988). Analysis of the interaction between properdin and factor B, components of the alternative-pathway C3 convertase of complement. *Biochemical Journal* **253,** 3,667-675.

Farries, T. C., Steuer, K. L., and Atkinson, J. P. (1990). The mechanism of activation of the alternative pathway of complement by cell-bound C4b. *Molecular Immunology* **27,** 11,1155-1161.

Fearon, D. T. and Austen, K. F. (1975). Properdin: binding to C3b and stabilization of the C3b-dependent C3 convertase. *Journal of Experimental Medicine* **142,** 4,856-863.

Fearon, D. T. (1978). Regulation by membrane sialic acid of beta1H-dependent decay-dissociation of amplification C3 convertase of the alternative complement pathway. *Proceedings of the National Academy of Sciences of the United States of America* **75,** 4,1971-1975.

Fearon, D. T. (1979). Regulation of the amplification C3 convertase of human complement by an inhibitory protein isolated from human erythrocyte membrane. *Proceedings of the National Academy of Sciences of the United States of America* **76,** 11,5867-5871.

Fearon, D. T. and Wong, W. W. (1983). Complement ligand-receptor interactions that mediate biological responses. *Annual Review of Immunology* **1,** 243-271.

Fearon, D. T., Klickstein, L. B., Wong, W. W., Wilson, J. G., Moore, F. D. Jr, Weis, J. J., Weis, J. H., Jack, R. M., Carter, R. H., and Ahearn, J. A. (1989). Immunoregulatory functions of complement: structural and functional studies of complement receptor type 1 (CR1; CD35) and type 2 (CR2; CD21). *Progress in Clinical and Biological Research* **297,** 211-220.

Fishelson, Z.V.I., Pangburn, M. K. and Muller-Eberhard, H.J. (1983). C3 convertase of the alternative complement pathway: Demonstration of an active, stable B3b,Bb(Ni) complex. *Journal of Biolical Chemistry* **258**, 7411-7415.

Fishelson, Z., Pangburn, M. K., and Müller-Eberhard, H. J. (1984). Characterization of the initial C-3 convertase of the alternative pathway of human-complement. *Journal of Immunology* **132,** 1430-1434.

Fishelson, Z., Horstmann, R. D. and and Müller-Eberhard, H. J. (1987). Regulation of the alternative pathway of complement by pH. *J. Immunol.* **138**, 3392-3395.

Flower, D. R., North, A. C., and Attwood, T. K. (1991). Mouse oncogene protein 24p3 is

a member of the lipocalin protein family. *Biochemical and Biophysical Research Communications* **180,** 1,69-74.

Flower, D. R., North, A. C., and Attwood, T. K. (1993). Structure and sequence relationships in the lipocalins and related proteins. *Protein Science* **2,** 5,753-761.

Frangioni, J. V. and Neel, B. G. (1993). Solubilization and purification of enzymatically active glutathione- s-transferase (pGEX) fusion proteins. *Analytical Biochemistry* **210,** 179-187.

Frankel, S., Sohn, R., and Leinwand, L. (1991). The use of sarkosyl in generating soluble-protein after bacterial expression. *Proceedings of The National Academy of Sciences of The United States of America* **88,** 1192-1196.

Freeman, M., Ashkenas, J., Rees, D. J., Kingsley, D. M., Copeland, N. G., Jenkins, N. A., and Krieger, M. (1990). An ancient, highly conserved family of cysteine-rich protein domains revealed by cloning type I and type II murine macrophage scavenger receptors. *Proceedings of the National Academy of Sciences of the United States of America* **87,** 22,8810-8814.

Fujita, T., Gigli, I., and Nussenzweig, V. (1978). Human C4-binding protein. II. Role in proteolysis of C4b by C3b- inactivator. *Journal of Experimental Medicine* **148,** 4,1044-1051.

Gaither, T. A., Vargas, I., Inada, S., and Frank, M. M. (1987). The complement fragment C3d facilitates phagocytosis by monocytes [published erratum appears in Immunology 1988 Mar;63(3):559]. *Immunology* **62,** 3,405-411.

Garnier, J., Osguthorpe, D. J. and Robson, B. (1978). Analysis of the accuracy and implications of simple methods for predicting the secondary structures of globular proteins. *Journal of Molecular Biology.* **120**, 97-120.

Ghendler, Y., Parizade, M., Arnon, R., Mckerrow, J. H., and Fishelson, Z. (1996). Schistosoma mansoni: evidence for a 28-6tda membrane-anchored protease on schistosomula. *Experimental Parasitology* **83,** 73-82.

Gerard, N. P. and Gerard, C. (1991). The chemotactic receptor for human C5a anaphylatoxin. *Nature* **349,** 6310,614-617.

Gewurz, H., Zhang, X. H., and Lint, T. F. (1995). Structure and function of the pentraxins. *Current Opinion in Immunology* **7,** 1,54-64.

Ghebrehiwet, B. (1989). Functions associated with the C1q receptor. *Behring Institute Mitteilungen* 84,204-215.

Gibrat, J. F., Garnier, J. and Robson B. (1987). Further developments of protein secondary structure predictions using information theory - new parameters and consideration of residue pairs. *Journal of Molecular Biology.* **198**, 425-443.

Gigli, I., Fujita, T., and Nussenzweig, V. (1979). Modulation of the classical pathway C3 convertase by plasma proteins C4 binding protein and C3b inactivator. *Proceedings of the National Academy of Sciences of the United States of America* **76,** 12, 6596-6600.

Godovac-Zimmermann, J. (1988). The structural motif of beta-lactoglobulin and retinol-binding protein: a basic framework for binding and transport of small hydrophobic molecules? *Trends in Biochemical Sciences* **13,** 2,64-66.

Goldberger, G., Bruns, G. A., Rits, M., Edge, M. D., and Kwiatkowski, D. J. (1987).

Human complement factor I: analysis of cDNA-derived primary structure and assignment of its gene to chromosome 4. *Journal of Biological Chemistry* **262,** 21,10065-10071.

Gotwals, P. J., Chi-Rosso, G., Ryan, S. T., Sizing, I., Zafari, M., Benjamin, C., Singh, J., Venyaminov, S. Y., Pepinsky, R. B. and Koteliansky, V. (1999). Divalent cations stabilise the α1β1 integrin I domain. *Biochemistry*, **38**, 8280-8288.

Gotze, O., Medicus, R. G., and Müller-Eberhard, H. J. (1977). Alternative pathway of complement: nonenzymatic, reversible transition of precursor to active properdin. *Journal of Immunology* **118,** 2,525-532.

Gotze, O., Bianco, C., and Cohn, Z. A. (1979). The induction of macrophage spreading by factor B of the properdin system. *Journal of Experimental Medicine* **149,** 2,372-386.

Goundis, D. and Reid, K. B. (1988). Properdin, the terminal complement components, thrombospondin and the circumsporozoite protein of malaria parasites contain similar sequence motifs. *Nature* **335,** 6185,82-85.

Greer, J. (1986). Comparative structural anatomy of the complement anaphylatoxin proteins C3a, C4a and C5a. *Enzyme* **36,** 1-2,150-163.

Gresham, H. D. and Volanakis, J. E. (1986). Structure and function of human complement receptors: 1985. *Year in Immunology* **2,** 177-186.

Gutierrez, J. C., Gotze, O., and Caporale, L. H. (1983). An improved fluorogenic substrate for the alternative complement pathway C3/5 converting enzyme-CVFBb. *Biochimica et Biophysica Acta* **744,** 276-280.

Haefliger, J. A., Tschopp, J., Nardelli, D., Wahli, W., Kocher, H. P., Tosi, M., and Stanley, K. K. (1987). Complementary DNA cloning of complement C8 beta and its sequence homology to C9. *Biochemistry* **26,** 12,3551-3556.

Haefliger, J. A., Peitsch, M. C., Jenne, D. E., and Tschopp, J. (1991). Structural and functional characterization of complement C8 gamma, a member of the lipocalin protein family. *Molecular Immunology* **28,** 1-2,123-131.

Hanson, D. C., Siegel, R. C., and Schumaker, V. N. (1985). Segmental flexibility of the C1q subcomponent of human complement and its possible role in the immune response. *Journal of Biological Chemistry* **260,** 6,3576-3583.

Haris, P. I., Lee, D. C. and Chapman, D. (1986). A Fourier transform infrared investigation of the structural differences between ribonuclease A and ribonuclease S. *Biochim. Biophys. Acta*, **874**, 255-265.

Haris, P. I., Chapman, D., Harrison, R. A., Smith, K. F., and Perkins, S. J. (1990). Conformational transition between native and reactive center cleaved forms of alpha 1-antitrypsin by Fourier transform infrared spectroscopy and small-angle neutron scattering. *Biochemistry* **29,** 6,1377-1380.

Haris, P. I. and Chapman, D. (1994) Analysis of polypeptide and protein structures using Fourier transform infrared spectroscopy. *Methods in Molecular Biology* (Jones, C. *et al.*, eds.) **22**, pp183-202. Humana Press, Totowa.

Harrison, R. A., Thomas, M. L., and Tack, B. F. (1981). Sequence determination of the thiolester site of the fourth component of human complement. *Proceedings of the National Academy of Sciences of the United States of America* **78,** 12,7388-

7392.
Hashimoto, N., Zhang, W.-R. and Goldstein, B.J. (1992). Insulin receptor and epidermal growth factor receptor dephosphorylation by three major rat liver protein-tyrosine phosphatases expressed in a recombinant bacterial system. *Biochemical Journal* **284**, 569-576.
Heiland, I. and Gething, M.J. (1981). Cloned copy of the haemagglutinin gene codes for human influenza antigenic determinants in *E. coli*. *Nature* **292**, 851-852.
Heinz, H. P. (1989). Biological functions of C1q expressed by conformational changes. *Behring Institute Mitteilungen* **84**, 20-31.
Henikoff, S. and Henikoff, J. G. (1992). Amino-acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences of the United States of America* **89,** 10915-10919.
Herz, J., Hamann, U., Rogne, S., Myklebost, O., Gausepohl, H., and Stanley, K. K. (1988). Surface location and high affinity for calcium of a 500-kd liver membrane protein closely related to the LDL-receptor suggest a physiological role as lipoprotein receptor. *EMBO Journal* **7,** 13,4119-4127.
Hessing, M. (1991). The interaction between complement component C4b-binding protein and the vitamin K-dependent protein S forms a link between blood coagulation and the complement system. *Biochemical Journal* **277,** Pt 3,581-592.
Hinshelwood, J. and Perkins, S. J. (1998). The solution structure of human factor B reveals interactions between its domain: a study by NMR and homology modelling. *Molecular Immunology*, **35**, 390-390 (abstract).
Hinshelwood, J. and Perkins, S. J. (2000). Metal-dependent conformational changes in a recombinant vWF-A domain from human factor B: A solution study by circular dichroism, Fourier transform infrared and ^{1}H NMR spectroscopy.*Journal of Molecular Biology* **298**, 135-147.
Hinshelwood, J., Williams, S. C., Sim, R. B. and Perkins, S. J. (1996) Structural and functional studies of the isolated vWF-A and SP domains of complement factor B. *Molecular Immunology* **33**, 80 (abstract)
Hinshelwood, J., Spencer, D., Edwards, Y. J. K. and Perkins, S. J. (1999). Identification of the C3b binding site in a recombinant vWF-A domain of complement factor B by surface enhanced laser deionisation spectrometry and homology modelling: implications for the activation of factor B. *Journal of Molecular Biology* **294**, 587-599.
Hobohm, U. and Sander, C. (1994). Enlarged representative set of protein structures. *Protein Science* **3,** 522-524.
Hoekzema, R., Martens, M., Brouwer, M. C., and Hack, C. E. (1988). The distortive mechanism for the activation of complement component C1 supported by studies with a monoclonal antibody against the "arms" of C1q. *Molecular Immunology* **25,** 5,485-494.
Hohenester, E., Maurer, P., Hohenadl, C., Timpl, R., Jansonius, J. N., and Engel, J. (1996). Structure of a novel extracellular Ca(2+)-binding module in BM-40. *Nature Structural Biology* **3,** 1,67-73.
Hohenester, E., Maurer, P., and Timpl, R. (1997). Crystal structure of a pair of follistatin-like

and EF-hand calcium-binding domains in BM-40. *EMBO Journal* **16,** 13,3778-3786.

Holguin, M. H., Wilcox, L. A., Bernshaw, N. J., Rosse, W. F., and Parker, C. J. (1989). Relationship between the membrane inhibitor of reactive lysis and the erythrocyte phenotypes of paroxysmal nocturnal hemoglobinuria. *Journal of Clinical Investigation* **84,** 5,1387-1394.

Holt, G. D., Pangburn, M. K., and Ginsburg, V. (1990). Properdin binds to sulfatide [Gal(3-SO4)beta 1-1 Cer] and has a sequence homology with other proteins that bind sulfated glycoconjugates. *Journal of Biological Chemistry* **265,** 5,2852-2855.

Honig, B. and Nicholls, A. (1995). Classical electrostatics in biology and chemistry. *Science* **268,** 1144-1149.

Horiuchi, T., Macon, K. J., Engler, J. A., and Volanakis, J. E. (1991). Site-directed mutagenesis of the region around Cys-241 of complement component C2. Evidence for a C4b binding site. *Journal of Immunology* **147,** 2,584-589.

Horiuchi, T., Kim, S., Matsumoto, M., Watanabe, I., Fujita, S. and Volanakis, J. E. (1993). Human complement factor B: cDNA cloning, nucleotide sequencing, phenotypic conversion by site-directed mutagenesis and expression. *Molecular Immunology* **30,** 1587-1592.

Hourcade, D., Holers, V. M., and Atkinson, J. P. (1989). The regulators of complement activation (RCA) gene cluster. *Advances in Immunology* **45,** 381-416.

Hourcade, D. E., Wagner, L. M., and Oglesby, T. J. (1995). Analysis of the short consensus repeats of human-complement factor B by site-directed mutagenesis. *Journal of Biological Chemistry* **270,** 19716-19722.

Hourcade, D. E., Mitchell, L. M., and Oglesby, T. J. (1998). A conserved element in the serine protease domain of complement factor B. *Journal of Biological Chemistry* **273,** 25996-26000.

Hourcade, D. E., Mitchell, L. M., and Oglesby, T. J. (1998). The factor B and C2 serine protease domains: A unique evolutionary divergence. *Molecular Immunology* **35,** 292.

Hourcade, D. E., Mitchell, L. M., and Oglesby, T. J. (1998). The type A domain in the assembly of the alternative pathway convertases. *Molecular Immunology* **35,** 291.

Hourcade, D. E., Mitchell, L. M., and Oglesby, T. J. (1999). Mutation of the type A domain of complement factor B that promote high-affinity C3b-binding. *Journal of Immunology* **162,** 2906-2911.

Huang, H. J., Jones, N. H., Strominger, J. L., and Herzenberg, L. A. (1987). Molecular cloning of Ly-1, a membrane glycoprotein of mouse T lymphocytes and a subset of B cells: molecular homology to its human counterpart Leu-1/T1 (CD5). *Proceedings of the National Academy of Sciences of the United States of America* **84,** 1,204-208.

Hubbard, T. J. P. and Blundell, T. L. (1987). Comparison of solvent-inaccessible cores of homologous proteins - definitions useful for protein modeling. *Protein Engineering* **1,** 159-171.

Huber, R. and Bode, W. (1978). Structural basis of the activation and action of trypsin. Acc. Chem. Res. **11,** 1114-1122.

Huber, R. and Bode, W. (1978). Structural basis of the activation and action of trypsin. *Journal of the American Chemidtry Society* **11**, 114-122.

Huber, R., Scholze, H., Paques, E. P., and Deisenhofer, J. (1980). Crystal structure analysis and molecular model of human C3a anaphylatoxin. *Hoppe-Seylers Zeitschrift fur Physiologische Chemie* **361,** 9,1389-1399.

Huey, R. and Hugli, T. E. (1985). Characterization of a C5a receptor on human polymorphonuclear leukocytes (PMN). *Journal of Immunology* **135,** 2063-2068.

Hugli, T. E., Vallota, E. H., and Müller-Eberhard, H. J. (1975). Purification and partial characterization of human and porcine C3a Anaphylatoxin. *Journal of Biological Chemistry* **250,** 4 ,1472-1478.

Hugli, T. E. (1975). Human anaphylatoxin (C3a) from the third component of complement. Primary structure. *Journal of Biological Chemistry* **250,** 21 ,8293-8301.

Huizinga, E. G., van der Plas, R. M., Kroon, J., Sixma, J. J. and Gros, P. (1997). Crystal structure of the A3 domain of human von Willebrand factor: implications for collagen binding. *Structure*, **5**, 1147-1156.

Hutchens, T. W., Nelson, R. W., Allen, M. H., Li, C. M., and Yip, T. T. (1992). Peptide metal-ion interactions in solution - detection by laser desorption time-of-flight mass-spectrometry and electrospray ionization mass-spectrometry. *Biological Mass Spectrometry* **21,** 151-159.

Hutchens, T. W. and Yip, T. T. (1993). New desorption strategies for the mass-spectrometric analysis of macromolecules. *Rapid Communications in Mass Spectrometry* **7,** 576-580. Igarashi, M., Nagata, A., Toh, H., Urade, Y., and Hayaishi, O. (1992). Structural organization of the gene for prostaglandin D synthase in the rat brain. *Proceedings of the National Academy of Sciences of the United States of America* **89,** 12,5376-5380.

Isenman, D. E., Kells, D. I., Cooper, N. R., Müller-Eberhard, H. J., and Pangburn, M. K. (1981). Nucleophilic modification of human complement protein C3: correlation of conformational changes with acquisition of C3b- like functional properties. *Biochemistry* **20,** 15,4458-4467.

Ito, S. and Tamura, N. (1983). Inhibition of classical C5 convertase in the complement system by factor H. *Immunology* **50,** 4,631-635.

Jackson, M., Haris, P. I. and Chapman, D. (1991). Fourier transform infrared spectroscopic studies of Ca^{2+}-binding proteins. *Biochemistry*, **30**, 9681-9686.

Jenkins, P. V., Pasi, K. J., and Perkins, S. J. (1998). Molecular modeling of ligand and mutation sites of the type a domains of human von Willebrand factor and their relevance to von Willebrand's disease. *Blood* **91,** 2032-2044.

Jenne, D. E. and Tschopp, J. (1992). Clusterin: the intriguing guises of a widely expressed glycoprotein. *Trends in Biochemical Sciences* **17,** 4,154-159.

Jing, H., Babu, Y. S., Moore, D., Kilpatrick, J. M., Liu, X.-Y., Volanakis, J. E. and Narayana, S. V. L. (1998). Structures of native and complexed complement factor D: implications of the atypical His57 conformation and self-inhibitory loop in the regulation of specific serine protease activity. *Journal of Molecular Biology* **282,** 1061-1081.

Jing, H., Xu, Y., Carson, M., Moore, D., Macon, K. J., Volanakis, J. E. and Narayana, S.

V. (2000). New structural motifs on the chymotrypsin fold and their potential roles in complement factor B. *EMBO Journal*, **19**, 164-173.

Johnson, D. M., Gagnon, J., and Reid, K. B. (1984). Amino acid sequence of human factor D of the complement system. Similarity in sequence between factor D and proteases of non-plasma origin. *FEBS Letters* **166,** 2,347-351.

Johnson, K.S., Harrison, G.B., Lightowlers, M.W., O'Hoy, K.L., Cougle, W.G., Lawrence, S.B., Vinton, J.G., Heath, D.D. and Rickard, M.D. (1989). Vaccination against ovine cysticercosis using a defined antigen. *Nature (London)* **338**, 585-587.

Johnson, R. J. and Chenoweth, D. E. (1985). Labeling the granulocyte C5a receptor with a unique photoreactive probe. *Journal of Biological Chemistry* **260,** 12,7161-7164.

Johnson, W. C. (1990). Protein secondary structure and circular-dichroism - a practical guide. *Proteins-structure Function and Genetics* **7,** 205-214.

Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992). A new approach to protein fold recognition. *Nature* **358,** 86-89.

Jurianz, K., Garciaschuler, H., Fishelson, Z., and Kirschfink, M. (1998). Modulation of membrane-associated complement regulators on tumor cells. *Molecular Immunology* **35,** 129.

Kabsch, W. and Sander, C. (1983). Dictionary of protein secondary structure - pattern-recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22,** 2577-2637.

Kaelin Jr., W.G., Pallas, D.C., DeCaprio, J.A., Kaye, F.J. and Livingston, D.M. (1991). Identification of cellular proteins that can interact the T/E1A-binding region of the retinoblastoma gene product. *Cell* **64**, 521-532.

Kalter, E. S., Daha, M. R., ten Cate, J. W., Verhoef, J., and Bouma, B. N. (1985). Activation and inhibition of Hageman factor-dependent pathways and the complement system in uncomplicated bacteremia or bacterial shock. *Journal of Infectious Diseases* **151,** 6,1019-1027.

Kam, C. M., Mcrae, B. J., Harper, J. W., Niemann, M. A., Volanakis, J. E., and Powers, J. C. (1987). Human-complement protein-D, protein-C2, and protein-B - active-site mapping with peptide thioester substrates. *Journal of Biological Chemistry* **262,** 3444-3451.

Kemp, D.J., Smith, D.B., Foote, S.J., Samaras, N. and Peterson, G. (1989). Colorimetric detection of specific DNA segments amplified by polymerase chain reactions. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2423-2427.

Kieffer, B., Driscoll, P. C., Campbell, I. D., Willis, A. C., Vandermerwe, P. A., and Davis, S. J. (1994). 3-dimensional solution structure of the extracellular region of the complement regulatory protein CD59, a new cell-surface protein domain related to snake-venom neurotoxins. *Biochemistry* **33,** 4471-4482.

Kim, S., Narayana, S. V. L. and Volanakis, J. E. (1994). Mutational analysis of the substrate binding site of human complement factor D. *Biochemistry*, **33**, 14393-14399.

Kim, S., Narayana, S. V. L. and Volanakis, J. E. (1995). Crystal structure of a complement factor D mutant expressing enhanced catalytic activity. *Journal of Biological Chemistry* **270**, 24399-24405.

King, R. D. (1996). Prediction of secondary structure. In: *Protein structure prediction.*

Editor, M. J. E. Sternberg. Oxford University Press, Oxford. pp 79-99.

Kirszbaum, L., Sharpe, J. A., Murphy, B., d'Apice, A. J., Classon, B., Hudson, P., and Walker, I. D. (1989). Molecular cloning and characterization of the novel, human complement-associated protein, SP-40,40: a link between the complement and reproductive systems. *EMBO Journal* **8,** 3,711-718.

Kishimoto, T. K., Larson, R. S., Corbi, A. L., Dustin, M. L., Staunton, D. E., Springer, and TA. (1989). The leukocyte integrins. *Advances in Immunology* **46,** 149-182.

Kiss, I., Deak, F., Holloway, R. G. Jr, Delius, H., Mebust, K. A., Frimberger, E., Argraves, W. S., Tsonis, P. A., Winterbottom, N., and Goetinck, P. F. (1989). Structure of the gene for cartilage matrix protein, a modular protein of the extracellular matrix. Exon/intron organization, unusual splice sites, and relation to alpha chains of beta 2 integrins, von Willebrand factor, complement factors B and C2, and epidermal growth factor. *Journal of Biological Chemistry* **264,** 14,8126-8134.

Klickstein, L. B., Bartow, T. J., Miletic, V, Rabson, L. D., Smith, J. A., and Fearon, D. T. (1988). Identification of distinct C3b and C4b recognition sites in the human C3b/C4b receptor (CR-1, CD35) by deletion mutagenesis. *Journal of Experimental Medicine* **168,** 1699-1717.

Kline, T. P., Brown, F. K., Brown, S. C., Jeffs, P. W., Kopple, K. D., and Mueller, L. (1990). Solution structures of human transforming growth factor alpha derived from ^{1}H NMR data. *Biochemistry* **29,** 34,7805-7813.

Kluin-Nelemans, H. C., van Velzen-Blad, H., van Helden, H. P., and Daha, M. R. (1984). Functional deficiency of complement factor D in a monozygous twin. *Clinical and Experimental Immunology* **58,** 3,724-730.

Knobel, H. R., Villiger, W., and Isliker, H. (1975). Chemical analysis and electron microscopy studies of human C1q prepared by different methods. *European Journal of Immunology* **5,** 1,78-82.

Knott, H. M., Berndt, M. C., Kralicek, A. V., O'Donoghue, S. I. and King, G. F.(1992). Determination of the solution structure of a platelet-adhesion peptide of von Willebrand Factor. *Biochemistry*, **31**, 11152-11158.

Kohda, D., Shimada, I., Miyake, T., Fuwa, T., and Inagaki, F. (1989). Polypeptide chain fold of human transforming growth factor alpha analogous to those of mouse and human epidermal growth factors as studied by two-dimensional ^{1}H NMR. *Biochemistry* **28,** 3,953-958.

Kristensen, T., Deustachio, P., Ogata, R. T., Chung, L. P., Reid, K. B. M., and Tack, B. F. (1987). The superfamily of C3b/C4b-binding proteins. *Federation Proceedings* **46,** 2463-2469.

Krone, J. R., Nelson, R. W., Dogruel, D., Williams, P., and Granzow, R. (1997). BIA/MS: Interfacing biomolecular interaction analysis with mass spectrometry. *Analytical Biochemistry* **244,** 124-132.

Kuby, J (1997). *Immunology*. Third edition. Freeman, New York.

Kumar, P. J. and Clark, M. L., Eds. (1990). *Clinical Medicine*, Second Edition, Bailliere-Tindall, London.

Kunapuli, S.P., DeLa Cadena, R.A. and Colman, R.W. (1992). Bacterial expression of biological active high molecular weight kininogen light chain. *Thrombosis and*

Haemostasis **67**, 428-433.

Kunnath-Muglia, L. M., Chang, G. H., Sim, R. B., Day, A. J., and Ezekowitz, R. A. (1993). Characterization of Xenopus laevis complement factor I structure--conservation of modular structure except for an unusual insert not present in human factor I. *Molecular Immunology* **30,** 14,1249-1256.

Kuwata, H., Yip, T.-T., Yip, C. L., Tomita, M. and Hutchens, T. W. (1998). Bactericidal domain of lactoferrin: Detection, quantitation and characterization of lactoferricin in serum by SELDI affinity mass spectrometry. *Biochem. Biophys. Res. Commun.* **245**, 764-773.

Lachmann, P. J. (1990). Biological functions of the complement system. *Biochemical Society Transactions* **18,** 6,1143-1145.

Lachmann, P. J. (1991). The control of homologous lysis. *Immunology Today* **12,** 9,312-315.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227,** 259 ,680-685.

Lambris, J. D. (1988). The multifunctional role of C3, the third component of complement. *Immunology Today* **9,** 12,387-393.

Lambris, J. D. and Müller-Eberhard, H. J. (1984). Isolation and characterization of a 33,000-dalton fragment of complement factor B with catalytic and C3b binding-activity. *Journal of Biological Chemistry* **259,** 12685-12690.

Lambris, J. D., Lao, Z., Oglesby, T. J., Atkinson, J. P., Hack, C. E., and Becherer, J. D. (1996). Dissection of CR1, factor H, membrane cofactor protein, and factor B binding and functional sites in the third complement component. *Journal of Immunology* **156,** 4821-4832.

Lankinen, H., McLauchlan, J., Weir, M., Furlong, J., Conner, J., McGarrity, A., Mistry, A., Clements, J.B. and Marsden, H.S. (1991). Purification and characterization of the herpes simplex virus type 1 ribonucleotide reductase small subunit following expression in *Escerichia coli. Journal of General Virology.***72**, 1383-1392.

Laskowski, R. A., Macarthur, M. W., Moss, D. S., and Thornton, J. M. (1993). Procheck - a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* **26,** 283-291.

Lattman, E. E. (1995). Protein-structure prediction - a special issue. *Proteins-structure Function and Genetics* **23,** R1.

Law, S. K., Lichtenberg, N. A., Holcombe, F. H., and Levine, R. P. (1980). Interaction between the labile binding sites of the fourth (C4) and fifth (C5) human complement proteins and erythrocyte cell membranes. *Journal of Immunology* **125,** 2,634-639.

Law, S. K. and Reid, K. B. (1995). *Complement.* Second edition. IRL press, Oxford and Washington.

Lawler, J. and Hynes, R. O. (1986). The structure of human thrombospondin, an adhesive glycoprotein with multiple calcium-binding sites and homologies with several different proteins. *Journal of Cell Biology* **103,** 5,1635-1648.

Lee, B. and Richards, F. M. (1971). An interpretation of protein structures: estimation of static accessibility. *Journal of Molecular Biology.* **55**, 379-400.

Lee, D. C., Haris, P. I., Chapman, D. and Mitchell, R. C. (1990). Determination of protein

secondary structure using factor analysis of infrared spectra. *Biochemistry*, **29**, 9185-9193.

Lee, J. O., Rieu, P., Arnaout, M. A., and Liddington, R. (1995a). Crystal structure of the A domain from the alpha subunit of integrin CR3 (CD11b/CD18). *Cell* **80,** 4,631-638.

Lee, J. O., Bankston, L. A., Arnaout, M. A., and Liddington, R. C. (1995b). Two conformations of the integrin A-domain (I-domain): A pathway for activation? *Structure* **3,** 1333-1340.

Lee, T. H., Wisniewski, H. G., and Vilcek, J. (1992). A novel secretory tumor necrosis factor-inducible protein (Tsg-6) is a member of the family of hyaluronate binding-proteins, closely related to the adhesion receptor CD44. *Journal of Cell Biology* **116,** 545-557.

Leitinger, B. and Hogg, N. (1999). Integrin I domains and their function. *Biochemical Society Transactions* **27**, 826-832.

Lesavre, P., Gaillard, M. H., and Halbwachsmecarelli, L. (1982). Inhibition of alternative pathway factor-D by factor B-related Synthetic hexapeptides. *European Journal of Immunology* **12,** 252-254.

Lesavre, P. H., Hugli, T. E., Esser, A. F., and Müller-Eberhard, H. J. (1979). The alternative pathway C3/C5 convertase: chemical basis of factor B activation. *Journal of Immunology* **123,** 2 ,529-534.

Leung, S. M., Zeyda, T., and Thatcher, B. (1998). A new and rapid method for phenotype screening using SELDI Proteinchip(TM) arrays demonstrated on serum from knockout and wildtype mice. *Molecular Biology of The Cell* **9,** 2040.

Lichtenheld, M. G. and Podack, E. R. (1989). Structure of the human perforin gene. A simple gene organization with interesting potential regulatory sequences. *Journal of Immunology* **143,** 12,4267-4274.

Liddington, R. and Bankston, L. (1998). The integrin I domain: crystals, metals and related artefacts. *Structure*, **6**, 937-938.

Liszewski, M. K., Post, T. W., and Atkinson, J. P. (1991). Membrane cofactor protein (MCP or CD46): newest member of the regulators of complement activation gene cluster. *Annual Review of Immunology* **9,** 431-455.

Loftus, J. C. and Liddington, R. C. (1997). New insights into integrin-ligand interaction. *J. Clin. Investig.* **99**, 2302-2306.

Löwenadler, B., Jansson, B., Paleus, S., Holmgren, E., Nilsson, B., Moks, T., Palm, G., Josephson, S., Philipson, L. and Uhlén, M. (1987). A gene fusion system for generating antibodies against short peptides. *Gene* **58**, 87-97.

Lopez-Trascasa, M., Bing, D. H., Rivard, M., and Nicholson-Weller, A. (1989). Factor J: isolation and characterization of a new polypeptide inhibitor of complement C1. *Journal of Biological Chemistry* **264,** 27,16214-16221.

Lublin, D. M. and Atkinson, J. P. (1989). Decay-accelerating factor: biochemistry, molecular biology, and function. *Annual Review of Immunology* **7,** 35-58.

Lundblad, R. L. (1995). *Techniques in Protein Modification,* CRC Press, Boca Raton.

Macarthur, M. W., Driscoll, P. C., and Thornton, J. M. (1994). NMR and crystallography - complementary approaches to structure determination. *Trends in Biotechnology* **12,** 149-153.

Malhotra, R. and Sim, R. B. (1989). Chemical and hydrodynamic characterization of the human leucocyte receptor for complement subcomponent C1q. *Biochemical Journal* **262,** 2,625-631.

Malhotra, R., Thiel, S., Reid, K. B., and Sim, R. B. (1990). Human leukocyte C1q receptor binds other soluble proteins with collagen domains. *Journal of Experimental Medicine* **172,** 3,955-959.

Malhotra, R., HAURUM, J., Thiel, S., and Sim, R. B. (1992). Interaction of C1q receptor with lung surfactant protein-a. *European Journal of Immunology* **22,** 1437-1445.

Malhotra, R., Lu, J., Holmskov, U., and Sim, R. B. (1994). Collectins, collectin receptors and the lectin pathway of complement activation. *Clinical and Experimental Immunology* **97 Suppl 2,** 4-9.

Malhotra, R., Wormald, M. R., Rudd, P. M., Fischer, P. B., Dwek, R. A., and Sim, R. B. (1995). Glycosylation changes of IgG associated with rheumatoid arthritis can activate complement via the mannose-binding protein [see comments] [published erratum appears in Nat Med 1995 Jun;1(6):599]. *Nature Medicine* **1,** 3,237-243.

Marston, F. A. O. (1986). The purification of eukaryotic polypeptides synthesized in escherichia-coli. *Biochemical Journal* **240,** 1-12.

Matsushita, T. and Sadler, J. E. (1995). Identification of amino acid residues essential for von Willebrand Factor binding to platelet glycoprotein Ib. *Journal of Biological Chemistry*. **270**, 13406-13414.

Medicus, R. G., Gotze, O., and Müller-Eberhard, H. J. (1976). Alternative pathway of complement: recruitment of precursor properdin by the labile C3/C5 convertase and the potentiation of the pathway. *Journal of Experimental Medicine* **144,** 4,1076-1093.

Micklem, K. J., Sim, R. B., and Sim, E. (1984). Analysis of C3-receptor activity on human B-lymphocytes and isolation of the complement receptor type 2 (CR2). *Biochemical Journal* **224,** 1,75-86.

Mole, J. E., Anderson, J. K., Abraham, N., and Bhown, A. S. (1981). Structure and function of factor B of the alternative complement pathway. *Federation Proceedings* **40,** 962.

Mole, J. E., Anderson, J. K., Davison, E. A., and Woods, D. E. (1984). Complete primary structure for the zymogen of human-complement factor B. *Journal of Biological Chemistry* **259,** 3407-3412.

Molina, H., Perkins, S. J., Guthridge, J., Gorka, J., Kinoshita, T. and Holers, V. M. (1995). Characterization of a complement receptor 2 (CR2, CD21) ligand binding site for C3. An initial model of ligand interaction with two linked short consensus repeat modules. *Journal of Immunology,* **154**, 5426-5435.

Mollnes, T. E. and Lachmann, P. J. (1988). Regulation of complement. *Scandinavian Journal of Immunology* **27,** 2,127-142.

Monahan, J. B. and Sodetz, J. M. (1981). Role of the beta subunit in interaction of the eighth component of human complement with the membrane-bound cytolytic complex. *Journal of Biological Chemistry* **256,** 7,3258-3262.

Monahan, J. B., Stewart, J. L., and Sodetz, J. M. (1983). Studies of the association of the eighth and ninth components of human complement within the membrane-bound

cytolytic complex. *Journal of Biological Chemistry* **258,** 8,5056-5062.

Montelione, G. T., Wuthrich, K., Nice, E. C., Burgess, A. W., and Scheraga, H. A. (1987). Solution structure of murine epidermal growth factor: determination of the polypeptide backbone chain-fold by nuclear magnetic resonance and distance geometry. *Proceedings of the National Academy of Sciences of the United States of America* **84,** 15,5226-5230.

Morgan, B. P., Luzio, J. P., and Campbell, A. K. (1986). Intracellular Ca2+ and cell injury: a paradoxical role of Ca2+ in complement membrane attack. *Cell Calcium* **7,** 5-6,399-411.

Morgan, B. P. and Walport, M. J. (1991). Complement deficiency and disease. *Immunology Today* **12,** 9,301-306.

Morgan, E. L., Thoman, M. L., Hoeprich, P. D., and Hugli, T. E. (1985). Bioactive complement fragments in immunoregulation. *Immunology Letters* **9,** 4,207-213.

Morley, B. J. and Campbell, R. D. (1984). Internal homologies of the Ba fragment from human-complement component factor B, a class-III MHC antigen. *Embo Journal* **3,** 153-157.

Morris, A. L., Macarthur, M. W., Hutchinson, E. G., and Thornton, J. M. (1992). Stereochemical quality of protein-structure coordinates. *Proteins-structure Function and Genetics* **12,** 345-364.

Moy, F. J., Scheraga, H. A., Liu, J. F., Wu, R., and Montelione, G. T. (1989). Conformational characterization of a single-site mutant of murine epidermal growth factor (EGF) by ^{1}H NMR provides evidence that leucine-47 is involved in the interactions with the EGF receptor. *Proceedings of the National Academy of Sciences of the United States of America* **86,** 24,9836-9840.

Müller-Eberhard, H. J., Polley, M. J., and Calcott, M. A. (1967). Formation and functional significance of a molecular complex derived from the second and the fourth component of human complement. *Journal of Experimental Medicine* **125,** 2,359-380.

Nagar, B., Jones, R. G., Diefenbach, R. J., Isenman, D. E., and Rini, J. M. (1998). X-ray crystal structure of C3d: a C3 fragment and ligand for complement receptor 2. *Science* **280,** 5367,1277-1281.

Narayana, S. V. L., Carson, M., El-Kabbani, O., Kilpatrick, J. M., Moore, D., Chen, X., Bugg, C. E., Volanakis, J. E. and DeLucas, L. J. (1994). Structure of human factor D: A complement serine protease at 2.0 Å resolution. *Journal of Molecular Biology* **235,** 695-708.

Narayana, S. V. L., Jing, H., Moore, D., Macon, K. J., Ma, Y. Y. X., and Volanakis, J. E. (1998). Structure of factor B serine protease domain. *Molecular Immunology* **35,** 272.

Nicholson-Weller, A., Burge, J., Fearon, D. T., Weller, P. F., and Austen, K. F. (1982). Isolation of a human erythrocyte membrane glycoprotein with decay-accelerating activity for C3 convertases of the complement system. *Journal of Immunology* **129,** 1,184-189.

Nilsson, U. R., Mandle, R. J., Jr., and McConnell-Mapes, J. A. (1975). Human C3 and C5: subunit structure and modifications by trypsin and C42-C423. *Journal of*

Immunology **114,** 2 pt 2,815-822.

Nolan, K. F., Schwaeble, W., Kaluz, S., Dierich, M. P., and Reid, K. B. (1991). Molecular cloning of the cDNA coding for properdin, a positive regulator of the alternative pathway of human complement. *European Journal of Immunology* **21,** 3,771-776.

Nolte, M., Pepinsky, R.B., Venyaminov, S. Y., Koteliansky, V., Gotwals, P. J. and Karpusas, M. (1999). Crystal structure of the α1β1 integrin I-domain: insights into integrin I-domain function. *FEBS Lett.* **452**, 379-385.

Norman, D. G., Barlow, P. N., Baron, M., Day, A. J., Sim, R. B., and Campbell, I. D. (1991). Three-dimensional structure of a complement control protein module in solution. *Journal of Molecular Biology* **219,** 4,717-725.

Nusinow, S. R., Zuraw, B. L., and Curd, J. G. (1985). The hereditary and acquired deficiencies of complement. *Medical Clinics of North America* **69,** 3,487-504.

Ohlin, A. K., Linse, S., and Stenflo, J. (1988). Calcium binding to the epidermal growth factor homology region of bovine protein C. *Journal of Biological Chemistry* **263,** 15,7411-7417.

Oka, T., Sakamoto, S., Miyoshi, K., Fuwa, T., Yoda, K., Yamasaki, M. and Tamura, G. (1985). Synthesis and secretion of human epidermal growth factor by *Escherichia coli. Proc. Natl. Acad. Sci. U.S.A.* **82**, 7212-7216.

Okada, H., Nagami, Y., Takahashi, K., Okada, N., Hideshima, T., Takizawa, H., and Kondo, J. (1989). 20 KDa homologous restriction factor of complement resembles T cell activating protein. *Biochemical and Biophysical Research Communications* **162,** 3,1553-1559.

Oppermann, M., Haubitz, M., Quentin, E., and Gotze, O. (1988). Complement activation in patients with renal-failure as detected through the quantitation of fragments of the complement protein-C3, protein-C5, and factor B. *Klinische Wochenschrift* **66,** 857-864.

Orren, A., Potter, P. C., Cooper, R. C., and du Toit, E. (1987). Deficiency of the sixth component of complement and susceptibility to Neisseria meningitidis infections: studies in 10 families and five isolated cases. *Immunology* **62,** 2,249-253.

Pan, T. C., Sasaki, T., Zhang, R. Z., Fassler, R., Timpl, R., and Chu, M. L. (1993). Structure and expression of fibulin-2, a novel extracellular matrix protein with multiple EGF-like repeats and consensus motifs for calcium binding. *Journal of Cell Biology* **123,** 5,1269-1277.

Panayotou, G., Gish, G., End, P., Truong, O., Gout, I., Dhand, R., Fry, M.J., Hiles, I., Pawson, T. and Waterfield, M.D. (1993). Interactions between SH2 domains and tyrosine-phosphorylated platelet-derived growth factor B-receptor sequences: analysis of kinetic parameters by a novel biosensor-based approach. *Molecular Cell Biology* **13**, 3567-3576.

Pangburn, M. K. and Müller-Eberhard, H. J. (1980). Relation of putative thioester bond in C3 to activation of the alternative pathway and the binding of C3b to biological targets of complement. *Journal of Experimental Medicine* **152,** 4,1102-1114.

Pangburn, M. K., Morrison, D. C., Schreiber, R. D., and Müller-Eberhard, H. J. (1980). Activation of the alternative complement pathway: recognition of surface structures

on activators by bound C3b. *Journal of Immunology* **124,** 2,977-982.

Pangburn, M. K., Schreiber, R. D., and Müller-Eberhard, H. J. (1981). Formation of the initial C3 convertase of the alternative complement pathway. Acquisition of C3b-like activities by spontaneous hydrolysis of the putative thioester in native C3. *Journal of Experimental Medicine* **154,** 3,856-867.

Pangburn, M. K. and Müller-Eberhard, H. J. (1983). Kinetic and thermodynamic analysis of the control of C3b by the complement regulatory proteins factors H and I. *Biochemistry* **22,** 1,178-185.

Pangburn, M. K. and Müller-Eberhard, H. J. (1984). The alternative pathway of complement. *Springer Seminars in Immunopathology* **7,** 2-3,163-192.

Pangburn, M. K. and Müller-Eberhard, H. J. (1986). The C3 convertase of the alternative pathway of human complement. Enzymic properties of the bimolecular proteinase. *Biochemical Journal* **235,** 3,723-730.

Parkes, C., Gagnon, J., and Kerr, M. A. (1983). The reaction of iodine and thiol-blocking reagents with human complement components C2 and factor B. Purification and N-terminal amino acid sequence of a peptide from C2a containing a free thiol group. *Biochemical Journal* **213,** 1 ,201-209.

Pearson, W. R. and Lipman, D. J. (1988). Improved tools for biological sequence comparison . *Proceedings of the National Academy of Sciences of the United States of America* **85,** 2444-2448.

Peitsch, M. C., Amiguet, P., Guy, R., Brunner, J., Maizel, J. V. Jr, and Tschopp, J. (1990). Localization and molecular modelling of the membrane-inserted domain of the ninth component of human complement and perforin. *Molecular Immunology* **27,** 7,589-602.

Pepys, M. B. and Baltz, M. L. (1983). Acute phase proteins with special reference to C-reactive protein and related proteins (pentaxins) and serum amyloid A protein. *Advances in Immunology* **34,** 141-212.

Perkins, S. J. (1982). Applications of ring current calculations to the proton NMR of proteins and transfer RNA. *Biological Magnetic Resonance*, **4**, 193-336.

Perkins, S. J. (1986). Protein volumes and hydration effects - the calculations of partial specific volumes, neutron-scattering matchpoints and 280-nm absorption-coefficients for proteins and glycoproteins from amino- acid-sequences. *European Journal of Biochemistry* **157,** 169-180.

Perkins, S. J. (1989). Models for the C1 complex determined by physical techniques. *Behring Institute Mitteilungen* **84**, 129-141.

Perkins, S. J. and Wuthrich, K. (1980). Conformational transition from trypsinogen to trypsin: ^{1}H Nuclear magnetic resonance at 360 MHz and ring current calculations. *Journal of Molecular Biology* **138**, 43-64.

Perkins, S. J. and Nealis, A. S. (1989). The quaternary structure in solution of human complement subcomponent C1r2C1s2. *Biochemical Journal* **263,** 2,463-469.

Perkins, S. J. and Smith, K. F. (1993). Identity of the putative serine-proteinase fold in proteins of the complement system with nine relevant crystal structures. *Biochemical Journal* **295,** Pt 1,109-114.

Perkins, S. J., Haris, P. I., Sim, R. B., and Chapman, D. (1988). A study of the structure of human complement component factor H by Fourier transform infrared spectroscopy and secondary structure averaging methods. *Biochemistry* **27,** 11,4004-4012.

Perkins, S. J., Nealis, A. S., Haris, P. I., Chapman, D., Goundis, D., and Reid, K. B. (1989). Secondary structure in properdin of the complement cascade and related proteins: a study by Fourier transform infrared spectroscopy. *Biochemistry* **28,** 18,7176-7182.

Perkins, S. J., Smith, K. F., and Nealis, A. S. (1990). Molecular modelling strategies in application to complement. *Biochemical Society Transactions* **18,** 6,1151-1154.

Perkins, S. J., Nealis, A. S., and Sim, R. B. (1991). Oligomeric domain structure of human complement factor H by X- ray and neutron solution scattering. *Biochemistry* **30,** 11,2847-2857.

Perkins, S. J., Smith, K. F., Nealis, A. S., Haris, P. I., Chapman, D., Bauer, C. J., and Harrison, R. A. (1992). Secondary structure changes stabilize the reactive-centre cleaved form of SERPINs. A study by ^{1}H nuclear magnetic resonance and Fourier transform infrared spectroscopy. *Journal of Molecular Biology* **228,** 4,1235-1254.

Perkins, S. J., Smith, K. F., Kilpatrick, J. M., Volanakis, J. E. and Sim, R. B. (1993). Modelling of the serine-proteinase fold by X-ray and neutron scattering and sedimentation analyses: occurrence of the fold in factor D of the complement system. *Biochemical Journal,* **295**, 87-99.

Perkins, S. J., Smith, K. F., Williams, S. C., Haris, P. I., Chapman, D., and Sim, R. B. (1994). The secondary structure of the von Willebrand factor type A domain in factor B of human complement by Fourier transform infrared spectroscopy. Its occurrence in collagen types VI, VII, XII and XIV, the integrins and other proteins by averaged structure predictions. *Journal of Molecular Biology* **238,** 1,104-119.

Perkins, S. J., Hinshelwood, J., Edwards, Y. J. K. and Jenkins, P.V. (1999). Structural and functional modelling of vWF-A domains in complement and coagulation. *Biochemical Society Transactions* **27**, 815-821.

Pervaiz, S. and Brew, K. (1987). Homology and structure-function correlations between alpha 1-acid glycoprotein and serum retinol-binding protein and its relatives. *FASEB Journal* **1,** 3,209-214.

Petry, F., Reid, K. B., and Loos, M. (1992). Isolation, sequence analysis and characterization of cDNA clones coding for the C chain of mouse C1q. Sequence similarity of complement subcomponent C1q, collagen type VIII and type X and precerebellin. *European Journal of Biochemistry* **209,** 1,129-134.

Plummer, T. H. Jr and Hurwitz, M. Y. (1978). Human plasma carboxypeptidase N. Isolation and characterization. *Journal of Biological Chemistry* **253,** 11,3907-3912.

Podack, E. R. and Müller-Eberhard, H. J. (1978). Binding of desoxycholate, phosphatidylcholine vesicles, lipoprotein and of the S-protein to complexes of terminal complement components. *Journal of Immunology* **121,** 3,1025-1030.

Poon, P. H., Schumaker, V. N., Phillips, M. L., and Strang, C. J. (1983). Conformation and restricted segmental flexibility of C1, the first component of human complement. *Journal of Molecular Biology* **168,** 3,563-577.

Prater, C. A., Plotkin, J., Jaye, D., and Frazier, W. A. (1991). The properdin-like type I repeats of human thrombospondin contain a cell attachment site. *Journal of Cell*

Biology **112,** 5,1031-1040.

Preissner, K. T., Podack, E. R., and Müller-Eberhard, H. J. (1985). The membrane attack complex of complement: relation of C7 to the metastable membrane binding site of the intermediate complex C5b-7. *Journal of Immunology* **135,** 1,445-451.

Provencher, S. W. (1982) A constrained regularization method for inverting data represented by linear algebraic or integral-equations. CONTIN - a general-purpose constrained regularization program for inverting noisy linear algebraic and integral-equations. *Comput. Phys. Commun.* **27,** 213-227 and 229-242.

Pryzdial, E. L. G. and Isenman, D. E. (1985). A re-examination of the role of Mg-2+ in the alternative pathway C-3 convertase. *Federation Proceedings* **44,** 989.

Pryzdial, E. L. G. and Isenman, D. E. (1986). A reexamination of the role of magnesium in the human alternative pathway of complement. *Molecular Immunology* **23,** 87-96.

Pryzdial, E. L. G. and Isenman, D. E. (1987). Alternative complement pathway activation fragment-Ba binds to C3b - evidence that formation of the factor B-C3b complex involves 2 discrete points of contact. *Journal of Biological Chemistry* **262,** 1519-1525.

Pryzdial, E. L. G. and Isenman, D. E. (1988). A thermodynamic study of the interaction between human-complement component-C3b or component-C3(H2O) and factor B in solution. *Journal of Biological Chemistry* **263,** 1733-1738.

Qu, A. and Leahy, D. J. (1995). Crystal structure of the I-domain from the CD11a/CD18 (LFA-1, alpha L beta 2) integrin. *Proceedings of the National Academy of Sciences of the United States of America* **92,** 22,10277-10281.

Qu, A. and Leahy, D. J. (1996). The role of the divalent cation in the structure of the I domain from the CD11a/CD18 integrin. *Structure* **4,** 8,931-942.

Rabin, D.U., Palmer-Crocker, R., Mierz, D.V. and Yeung, K.K. (1992). An ELISA sandwich capture assay for recombinant fusion proteins glutathione-S-transferase. *J. Immunol. Meth.* **156**, 101-105.

Ramachandran, G. N. and Sassiekharan, V. (1968). Conformation of polypeptides and proteins. *Advan. Protein Chem.* **23,** 283-437.

Ramm, L. E., Whitlow, M. B., and Mayer, M. M. (1985). The relationship between channel size and the number of C9 molecules in the C5b-9 complex. *Journal of Immunology* **134,** 4,2594-2599.

Rao, A. G., Howard, O. M., Ng, S. C., Whitehead, A. S., Colten, H. R., and Sodetz, J. M. (1987). Complementary DNA and derived amino acid sequence of the alpha subunit of human complement protein C8: evidence for the existence of a separate alpha subunit messenger RNA. *Biochemistry* **26,** 12,3556-3564.

Reid, K. B. (1986). Activation and control of the complement system. *Essays in Biochemistry* **22,** 27-68.

Reid, K. B. and Porter, R. R. (1976). Subunit composition and structure of subcomponent C1q of the first component of human complement. *Biochemical Journal* **155,** 1,19-23.

Reid, K. B. and Day, A. J. (1989). Structure-function relationships of the complement components [published erratum appears in Immunol Today 1989 Jul;10(7):following 216] [see comments]. *Immunology Today* **10,** 6,177-180.

Reiter, Y., Ciobotariu, A., Jones, J., Morgan, B. P., and Fishelson, Z. (1995). Complement membrane attack complex, perforin, and bacterial exotoxins induce in K562 cells calcium-dependent cross-protection from lysis. *Journal of Immunology* **155,** 2203-2210.

Remaut, E., Marmenout, A., Simons, G., and Fiers, W. (1987). Expression of heterologous unfused protein in escherichia-coli. *Methods in Enzymology* **153,** 416-431.

Richardson, J. and Richardson, D. C. (1989). Principles and patterns of protein conformation. In: *Prediction of protein structure and the principles of protein conformation.* Editor G. Fasman. Plenum, New York. pp1-98.

Richmond, T. J. and Richards, F. M. (1978). Packing of α-helices: geometrical constraints and contact areas. *Journal of Molecular Biology.* **119**, 537-555.

Robson, B. and Pain, R. H. (1971). Analysis of the code relating sequence to the conformation in protteins: possible implications for the mechanism of formation of helical regions. *Journal of Molecular Biology.* **58**, 237-259.

Robson, B. and Suzuki, E. (1976). Conformational properties of amino acids residues in globuar proteins. *Journal of Molecular Biology.* **107**, 327-356. Rollins, S. A. and Sims, P. J. (1990). The complement-inhibitory activity of CD59 resides in its capacity to block incorporation of C9 into membrane C5b-9. *Journal of Immunology* **144,** 9,3478-3483.

Ross, G. D. (1989). Complement and complement receptors. *Current Opinion in Immunology* **2,** 50-62.

Rost, B. and Sander, C., (1993*a*). Improved prediction of protein secondary structure by the use of sequence profiles and neural networks. *Proceedings of The National Academy of Sciences of The United States of America.* **90**, 7558-7562.

Rost, B. and Sander, C. (1993). Prediction of protein secondary structure at better than 70-percent accuracy. *Journal of Molecular Biology* **232,** 584-599.

Sadler, J. E., Sheltoninloes, B. B., Sorace, J. M., Harlan, J. M., Titani, K., and Davie, E. W. (1985). Cloning and characterization of 2 cDNAS coding for human von Willebrand Factor. *Proceedings of the National Academy of Sciences of the United States of America* **82,** 6394-6398.

Šali, A. and Blundell, T. L. (1990). Definition of general topological equivalence in protein structures: a procedure involving comparison of properties and relationships through simulated annealing and dynamic programming. *Journal of Molecular Biology.* **212,** 403-428.

Šali, A., Overington, J. P., Johnson, M. S., and Blundell, T. L. (1990). From comparisons of protein sequences and structures to protein modelling and design. *Trends in Biochemical Sciences* **15,** 6,235-240.

Sambrook, J, Fritsch, E. F., and Maniatis, T. (1989). Molecular cloning: a laboratory manual.

Sanchez-Corral, P., Anton, L. C., Alcolea, J. M., Marques, G., Sanchez, A., and Vivanco, F. (1990). Proteolytic activity of the different fragments of factor B on the 3rd component of complement (C3) - involvement of the N-terminal domain of Bb in magnesium binding. *Molecular Immunology* **27,** 891-900.

Schlesinger, M., Nave, Z., Levy, Y., Slater, P. E., and Fishelson, Z. (1990). Prevalence of hereditary properdin, C7 and C8 deficiencies in patients with meningococcal

infections. *Clinical and Experimental Immunology* **81,** 3,423-427.

Schonermark, S., Filsinger, S., Berger, B., and Hansch, G. M. (1988). The C8-binding protein of human erythrocytes: interaction with the components of the complement-attack phase. *Immunology* **63,** 4,585-590.

Schreiber, R. D., Pangburn, M. K., and Müller-Eberhard, H. J. (1981). C3 modified at the thiolester site: acquisition of reactivity with cellular C3b receptors. *Bioscience Reports* **1,** 11,873-880.

Schumaker, V. N., Hanson, D. C., Kilchherr, E., Phillips, M. L., and Poon, P. H. (1986). A molecular mechanism for the activation of the first component of complement by immune complexes. *Molecular Immunology* **23,** 5,557-565.

Schumaker, V. N., Zavodszky, P., and Poon, P. H. (1987). Activation of the first component of complement. *Annual Review of Immunology* **5,** 21-42.

Schwaeble, W., Luttig, B., Sokolowski, T., Estaller, C., Weiss, E. H., Zumbuschenfelde, K. H. M., Whaley, K., and Dippold, W. (1993). Human complement factor B - functional-properties of a recombinant zymogen of the alternative activation pathway convertase. *Immunobiology* **188,** 221-232.

Scott, J. (1989). Lipoprotein receptors. Unravelling atherosclerosis [news]. *Nature* **338,** 6211,118-119.

Selander, M., Persson, E., Stenflo, J., and Drakenberg, T. (1990). ^{1}H NMR assignment and secondary structure of the Ca2(+)-free form of the amino-terminal epidermal growth factor like domain in coagulation factor X. *Biochemistry* **29,** 35,8111-8118.

Seya, T., Turner, J. R., and Atkinson, J. P. (1986). Purification and characterization of a membrane protein (gp45- 70) that is a cofactor for cleavage of C3b and C4b. *Journal of Experimental Medicine* **163,** 4,837-855.

Seya, T., Okada, M., Matsumoto, M., Hong, K. S., Kinoshita, T., and Atkinson, J. P. (1991). Preferential inactivation of the C5 convertase of the alternative complement pathway by factor I and membrane cofactor protein (MCP). *Molecular Immunology* **28,** 10,1137-1147.

Siciliano, S. J., Rollins, T. E., and Springer, M. S. (1990). Interaction between the C5a receptor and Gi in both the membrane-bound and detergent-solubilized states. *Journal of Biological Chemistry* **265,** 32,19568-19574.

Siegel, L. M. and Monty, K. J. (1966). Determination of molecular weights and frictional ratios of proteins in impure systems by use of gel filtration and density gradient centrifugation. Application to crude preparations of sulfite and hydroxylamine reductases. *Biochimica et Biophysica Acta* **112,** 2 ,346-362.

Sim, E. and Sim, R. B. (1981). Binding of fluid-phase complement components-C3 and components-C3b to human-lymphocytes. *Biochemical Journal* **198,** 509-518.

Sim, E. and Sim, R. B. (1983). Enzymic assay of C3b receptor on intact-cells and solubilized cells. *Biochemical Journal* **210,** 567-576.

Sim, R. B. and Sim, E. (1983). Autolytic fragmentation of complement components C-3 and C-4 and its relationship to covalent binding-activity. *Annals of the New York Academy of Sciences* **421,** 259-276.

Sim, R. B., Malhotra, V., Ripoche, J., Day, A. J., Micklem, K. J., and Sim, E. (1986). Complement receptors and related complement control proteins. *Biochemical*

Society Symposia **51,** 83-96.

Sim, R. B., Malhotra, V., Day, A. J., and Erdei, A. (1987). Structure and specificity of complement receptors. *Immunology Letters* **14,** 3,183-190.

Sim, R. B., MALHOTRA, V, Day, A. J., and Erdei, A. (1987). Structure and specificity of complement receptors. *Immunology Letters* **14,** 183-190.

Sim, R. B. and Perkins, S. J. (1990). Molecular modelling of C3 and its ligands. *Current Topics in Microbiology and Immunology* **153,** 209-222.

Sim, R. B. and Reid, K. B. (1991). C1: molecular interactions with activating systems. *Immunology Today* **12,** 9,307-311.

Sim, R. B. and Malhotra, R. (1994). Interactions of carbohydrates and lectins with complement. *Biochemical Society Transactions* **22,** 1,106-111.

Sinha, D., Bakhshi, M., Kunapuli, S., Vora, R., Gabriel, J. L., Kirby, E. P. and Budzynski, A. Z. (1994). Asp514 within the A1 domain of bovine von Willebrand factor is required for interaction with platelet glycoprotein IB. *Biochem. Biophys. Res. Comm.* **203**, 881-888.

Smith, C. A., Pangburn, M. K., Vogel, C. W., and Müller-Eberhard, H. J. (1982). Ultrastructural studies of the human-complement components C-3 and factor B and of cobra venom factor (CVF). *Molecular Immunology* **19,** 1403.

Smith, C. A., Vogel, C. W., and Müller-Eberhard, H. J. (1982). Ultrastructure of cobra venom factor-dependent C3/C5 convertase and its zymogen, factor B of human-complement. *Journal of Biological Chemistry* **257,** 9879-9882.

Smith, C. A., Vogel, C. W., and Müller-Eberhard, H. J. (1984). MHC Class-III products - an electron-microscopic study of the C-3 convertases of human-complement. *Journal of Experimental Medicine* **159,** 324-329.

Smith, C. S., Diacovo, T., Estavillo, D., Emsley, J., Liddington, R., and Cruz, M. A. (1998). Identification of the collagen binding site in the alpha 2 (GPIA/IIA) I-domain by crystal structure and site specific mutagenesis. *Blood* **92,** 1435.

Smith, D. B. and Johnson, K. S. (1988). Single-step purification of polypeptides expressed in escherichia- coli as fusions with glutathione s-transferase. *Gene* **67,** 31-40.

Smith, D. B. and Corcoran, L. M. (1994). Expression and purification of glutathione-S-transferase fusion proteins. in *Current Protocols in Molecular Biology.* (Ausubel *et al.*, eds.), vol 2, John Wiley and Sons. USA.

Smith, K. F., Haris, P. I., Chapman, D., Reid, K. B., and Perkins, S. J. (1994). Beta-sheet secondary structure of the trimeric globular domain of C1q of complement and collagen types VIII and X by Fourier-transform infrared spectroscopy and averaged structure predictions. *Biochemical Journal* **301,** Pt 1,249-256.

Snouwaert, J.N., Leebeek, F.W.G. and Fowlkes, D.M. (1991). Role of disulphide bonds in biological activity of human interleukin-6. *Journal of Biological Chemistry.* **266**, 23097-23102.

Specks, U., Mayer, U., Nischt, R., Spissinger, T., Mann, K., Timpl, R., Engel, J., and Chu, M. L. (1992). Structure of recombinant N-terminal globule of type VI collagen alpha 3 chain and its binding to heparin and Hyaluronan. *Embo Journal* **11,** 12 ,4281-4290.

Spencer, D. I. R., Robson, L., Bhatia, J., Sharma, S. K., Michael, N. P., Whitelegg, N. R.,

Rees, A. R., Chester, K. A. and Begent, R. H. J. (1999). Identifying immunogenic sites on ADEPT enzyme CPG2 using a scFv library and SELDI-AMS. British Cancer Research Meeting, Edinburgh, 11th-14th July. *Brit. J. Cancer*, **80** (Suppl. 2), 52 (Abstract).

Stanley, K. K. (1986). Homology with hemopexin suggests a possible scavenging function for s-protein vitronectin. *FEBS Letters* **199,** 249-253.

Steinkasserer, A., Solari, R., Mott, H. R., Aplin, R. T., Robinson, C. C., Willis, A. C., and Sim, R. B. (1992). Human interleukin-1 receptor antagonist - high-yield expression in Escherichia-coli and examination of cysteine residues. *Febs Letters* **310,** 63-65

Stewart, J. L., Kolb, W. P., and Sodetz, J. M. (1987). Evidence that C5b recognizes and mediates C8 incorporation into the cytolytic complex of complement. *Journal of Immunology* **139,** 6,1960-1964.

Straus, D. and Gilbert, W. (1985). Chicken triosephosphate isomerase complements an *Escerichia coli* deficiency. *Proc. Natl. Acad. Sci. U.S.A.* **82,** 2014-2018.

Stryer, L. (1995). *Stryer*. New York, Freeman.

Sugita, Y., Nakano, Y., and Tomita, M. (1988). Isolation from human erythrocytes of a new membrane protein which inhibits the formation of complement transmembrane channels. *Journal of Biochemistry* **104,** 4,633-637.

Surewicz, W. K., Mantsch, H. H. and Chapman, D. (1993). Determination of protein secondary structure by Fourier transform infrared spectroscopy: A critical assessment. *Biochemistry*, **32**, 389-394.

Tack, B. D. and Prahl, J. W. (1976). Third component of human complement: purification from plasma and physicochemical characterization. *Biochemistry* **15,** 20 ,4513-4521.

Takada, Y. and Hemler, M. E. (1989). The primary structure of the VLA-2/collagen receptor alpha 2 subunit (platelet GPIa): homology to other integrins and the presence of a possible collagen-binding domain. *Journal of Cell Biology* **109,** 1,397-407.

Taniguchi-Sidle, A. and Isenman, D. E. (1992). Mutagenesis of the Arg-gly-Asp triplet in human-complement component- C3 does not abolish binding of iC3b to the leukocyte integrin complement receptor type-III (CR3, CD11b/CD18). *Journal of Biological Chemistry* **267,** 635-643.

Taniguchi-Sidle, A. and Isenman, D. E. (1994). Interactions of human-complement component C3 with factor B and with complement receptors type-1 (CR-1, CD35) and type-3 (CR3, CD11b/CD18) involve an acidic sequence at the N-terminus of C3 alpha'-chain. *Journal of Immunology* **153,** 5285-5302.

Taylor, M. A. J., Pratt, K. A., Revell, D. F., Baker, K. C., Sumner, I. G. and Goodenough, P. W. (1992). Active papain renatured and processed from insoluble recombinant propapain expressed in *Escherichia coli. Prot. Eng.* **5,** 455-459.

Thiel, S., Vorup-Jensen, T., Stover, C. M., Schwaeble, W., Laursen, S. B., Poulsen, Willis, A. C., Eggleton, P., Hansen, S., Holmskov, U., Reid, K. B., Jensenius, and JC (1997). A second serine protease associated with mannan-binding lectin that activates complement. *Nature* **386,** 6624,506-510.

Thielens, N. M., Villiers, C. L., Villiers, M. B., and Colomb, M. G. (1984). Comparative study of the fluid-phase proteolytic cleavage of human complement subcomponents

C4 and C2 by C1s and C1r2-C1s2. *FEBS Letters* **165,** 1,111-116.

Thielens, N. M., Aude, C. A., Lacroix, M. B., Gagnon, J., and Arlaud, G. J. (1990). Ca2+ binding properties and Ca2(+)-dependent interactions of the isolated NH2-terminal alpha fragments of human complement proteases C1-r and C1-s. *Journal of Biological Chemistry* **265,** 24,14469-14475.

Thompson, R. A. and Winterborn, M. H. (1981). Hypocomplementaemia due to a genetic deficiency of beta 1H globulin. *Clinical and Experimental Immunology* **46,** 1,110-119.

Tomlinson, M.G., Williams, A.F. and Wright, M.D. (1993). Epitope mapping of anti-rat CD53 monoclonal antibodies. Implications for the membrane orientation of the transmembrane 4 superfamily. *Eur. J. Immunol.* **23**, 136-140.

Tosi, M., Duponchel, C., Meo, T., and Julier, C. (1987). Complete cDNA sequence of human complement Cls and close physical linkage of the homologous genes Cls and Clr. *Biochemistry* **26,** 26,8516-8524.

Toumadje, A., Alcorn, S. W., and Johnson, W. C. (1992). Extending CD spectra of proteins to 168 nm improves the analysis for secondary structures. *Analytical Biochemistry* **200,** 321-331.

Tschopp, J., Masson, D., and Stanley, K. K. (1986). Structural/functional similarity between proteins involved in complement- and cytotoxic T-lymphocyte-mediated cytolysis. *Nature* **322,** 6082,831-834.

Tsukada, H. and Blow, D. M. (1985). Structure of α-chymotrypsin refined at 1.68 Å resolution. *Journal of Molecular Biology.* , **184,** 703-711.

Tuckwell, D. S., Xu, Y. Y., Newham, P., Humphries, M. J., and Volanakis, J. E. (1997). Surface loops adjacent to the cation-binding site of the complement factor B von Willebrand factor type A module determine C3b binding specificity. *Biochemistry* **36,** 6605-6613.

Tuckwell, D. (1999). Evolution of von Willebrand factor A (VWA) domains. *Biochemical Society Transactions* **27**, 835-840.

Ueda, A., Kearney, J. F., Roux, K. H., and Volanakis, J. E. (1987). Probing functional sites on complement protein-B with monoclonal- antibodies - evidence for C3b-binding sites on Ba. *Journal of Immunology* **138,** 1143-1149.

Udaka, S. and Yamagata, H. (1993). High-level secretion of heterologous proteins by *Bacillus-brevis*. *Methods in Enzymology* **217,** 23-33.

Ueda, T., Rieu, P., Brayer, J. and Arnaout, M. A. (1994). Identification of the complement iC3b binding site in the β2 integrin CR3 (CD11b/CD18). *Proc. Natl. Acad. Sci. U. S. A.* **91**, 10680-10684.

Uhlén, M. and Abrahmsen, L. (1989). Secretion of recombinant proteins into the culture medium by *Escherichia coli* and *Staphylococcus aureus*. *Biochemical Society Transactions* **17**, 340-341.

Uhlén, M. and Moks, T. (1990). Gene fusions for purpose of expression: an introduction. *Meths. Enzymol.* **185**, 129-143. Ullman, C. G. and Perkins, S. J. (1997). The Factor I and follistatin domain families: the return of a prodigal son [letter]. *Biochemical Journal* **326,** Pt 3,939-941.

Varsano, S., Frolkis, I, Rashkovsky, L., Ophir, D., and Fishelson, Z. (1996). Protection of

human nasal respiratory epithelium from complement- mediated lysis by cell-membrane regulators of complement activation. *American Journal of Respiratory Cell and Molecular Biology* **15,** 731-737.

Van de Velde, H., von, Hoegen, I, Luo, W., Parnes, J. R., and Thielemans, K. (1991). The B-cell surface protein CD72/Lyb-2 is the ligand for CD5 [see comments]. *Nature* **351,** 6328,662-665.

Vanscheeuwijck, P., Huang, Y., Schullery, D. and Regan, J.W. (1993). Antibodies to a human a2-C10 adrenergic receptor fusion confirm the cytoplasmic orientation of the V-VI loop. *Biochem. Biophys. Res. Comm.* **190**, 340-346.

Volanakis, J. E. (1990). Participation of C3 and its ligands in complement activation. *Current Topics in Microbiology and Immunology* **153,** 1-21.

Volanakis, J. E. and Narayana, S. V. L. (1996). Complement factor D, a novel serine protease. *Protein Science* **5**, 553-564.

Volpers, C., Sapp, M., Komly, C.A., Richalet-Secordel, P. and Streeck, R.E. (1993). Development of type-specific and cross-reactive serological probes for the minor capsid protein of human papillomavirus type 33. *J. Virol.* **67**, 1927-1935.

Von Zabern, I., Nolte, R., and Vogt, W. (1981). Treatment of human complement components C4 and C3 with amines or chaotropic ions. Evidence of a functional and structural change that provides uncleaved C4 and C3 with properties of their soluble activated forms C4b and C3b. *Scandinavian Journal of Immunology* **13**, 413-431.

Walport, M. J. and Lachmann, P. J. (1988). Erythrocyte complement receptor type 1, immune complexes, and the rheumatic diseases. *Arthritis and Rheumatism* **31,** 2,153-158.

Walport, M. J. and Lachmann, P. J. (1990). Complement deficiencies and abnormalities of the complement system in systemic lupus erythematosus and related disorders. *Current Opinion in Rheumatology* **2,** 4,661-663.

Wang, D., Bode, W. and Huber, R. (1985). Bovine chymotrypsinogen A. X-ray crystal structure analysis and refinement of a new crystal form at 1.8 Å resolution. *Journal of Molecular Biology,* **185**, 595-624.

Watanabe, H., Circolo, A., Wetsel, R., Hoters, V. M., Boackle, S. A., Colten, H. R., and Gilkeson, G. S. (1998). Amelioration of renal disease in lupus mice genetically deficient in complement factor B. *Arthritis and Rheumatism* **41,** 878.

Weiss, V., Fauser, C., and Engel, J. (1986). Functional model of subcomponent C1 of human complement. *Journal of Molecular Biology* **189,** 3,573-581.

Weiss, M.A. and Keutmann, H.T. (1990). Alternating zinc finger motifs in the male-associated protein ZFY: defining architectural rules by mutagenesis and design of an "aromatic swap" second- site revertant. *Biochemistry* **29**, 9808-9813.

Wetlaufer, D. B. (1981). Folding of protein fragments. *Advances in Protein Chemistry* **34,** 61-92.

Wetsel, R. A., Lemons, R. S., Le Beau, M. M., Barnum, S. R., Noack, D., and Tack, B. F. (1988). Molecular analysis of human complement component C5: localization of the structural gene to chromosome 9. *Biochemistry* **27,** 5,1474-1482.

Williams, S. C. and Sim, R. B. (1993). Dye-ligand affinity purification of human-complement factor B and beta-2 glycoprotein-I. *Journal of Immunological Methods* **157,** 25-

30.
Williams, S. C. and Sim, R. B. (1994). Binding-sites involved in the formation of the C3(H2O)-factor B- complex of pathway of complement. *Biochemical Society Transactions* **22,** S2.
Williams, S. C., Hinshelwood, J., Perkins, S. J. and Sim, R. B. (1999). Production and functional activity of a recombinant von Willebrand Factor (vWF-A) domain from human complement factor B. *Biochem. J.* **342**, 625-632.
Wishart, D. S., Sykes, B. D. and Richards, F. M. (1992). The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry,* **31**, 1647-1651.
Xu, Y. and Volanakis, J. E. (1995). Construction and expression of human-complement C2 factor B chimeras. *Faseb Journal* **9,** A489.
Xu, Y. Y. and Volanakis, J. E. (1997). Contribution of the complement control protein modules of C2 in C4b Binding assessed by analysis of C2/factor B chimeras. *Journal of Immunology* **158,** 5958-5965.
Yamamoto, T., Bishop, R. W., Brown, M. S., Goldstein, J. L., and Russell, D. W. (1986). Deletion in cysteine-rich region of LDL receptor impedes transport to cell surface in WHHL rabbit. *Science* **232,** 4755,1230-1237.
Yamauchi, Y., Stevens, J. W., Macon, K. J., and Volanakis, J. E. (1994). Recombinant and native zymogen forms of human complement factor D. *Journal of Immunology* **152,** 7 ,3645-3653.
Yang, N., Schule, R., Manelsdorf, D.J. and Evans, R.M. (1991). Characterization of DNA binding and retinoic acid binding properties of retinoic acid receptor. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 3559-3563.
Yang, J. T., Wu, C. S. C., and Martinez, H. M. (1986). Calculation of protein conformation from circular-dichroism. *Methods in Enzymology* **130,** 208-269.
Zalman, L. S., Brothers, M. A., and Müller-Eberhard, H. J. (1989). Isolation of homologous restriction factor from human urine. Immunochemical properties and biologic activity. *Journal of Immunology* **143,** 6,1943-1947.
Ziccardi, R. J. and Cooper, N. R. (1976). Activation of C1r by proteolytic cleavage. *Journal of Immunology* **116,** 2,504-509.
Ziccardi, R. J. and Cooper, N. R. (1979). Active disassembly of the first complement component, C-1, by C- 1 inactivator. *Journal of Immunology* **123,** 2,788-792.
Ziccardi, R. J. (1983). The first component of human complement (C1): activation and control. *Springer Seminars in Immunopathology* **6,** 2-3,213-230.

Abstracts

Hinshelwood, J., Chen, X., Sim, R.B., and Perkins, S. J. (1995). Cleavage of complement factor B by factor D. *Immunology* **86**, 162 (abstract).

Hinshelwood, J., Williams, S. C., Sim, R. B. and Perkins, S. J. (1996) Structural and functional studies of the isolated vWF-A and SP domains of complement factor B. *Molecular Immunology* **33**, 80 (abstract).

Chamberlain, D., Hinshelwood, J., Ullman, C. G., Smith, K.F., and Sim, R.B. (1996) Molecular modelling of the compact five-domain structure of complement factor B by X-ray and neutron scattering. *Molecular Immunology* **33**, 319 (abstract).

Chamberlain, D., Hinshelwood, J., and Perkins, S. J. (1998). The compact five-domain structure of factor B of human complement by X-ray and neutron scattering and homology modelling. *Molecular Immunology* **35**, 237 (abstract).

Hinshelwood, J. and Perkins, S. J. (1998). The solution structure of human factor B reveals interactions between its domain: a study by NMR and homology modelling. *Molecular Immunology*, **35**, 390 (abstract).

Hinshelwood, J., Spencer, D., Edwards, Y. J. K. and Perkins, S. J. (2000). Identification of the C3b binding site in the vWF-A domain of factor B by SELDI mass spectrometry and homology modelling (abstract #93). **Selected for an oral presentation at the XVIII th International Complement Workshop, Utah, USA.**

Hinshelwood, J. and Perkins, S. J. (2000). Metal-dependent conformational changes in the vWF-A domain from factor B: A study by CD, FT-IR and ^{1}H NMR spectroscopy (abstract #115).

Hinshelwood, J. and Perkins, S. J. (2000). Conformational changes during the assembly of factor B from its domains by ^{1}H NMR: the regulation of factor B activity (abstract #116).

Publications

Hinshelwood, J., Ullman, C. G., Chamberlain, D., and Perkins, S. J. (1998). Proposed structural similarity of the FIMAC domain of factor I of complement with the follistatin domain of sparc protein: A study by spectroscopy and homology modelling. *Molecular Immunology* **35**, 235 (abstract).

Williams, S. C., Hinshelwood, J., Perkins, S. J. and Sim, R. B. (1999). Production and functional activity of a recombinant von Willebrand Factor (vWF-A) domain from human complement factor B. *Biochemical Journal* **342**, 625-632.

Perkins, S. J., Hinshelwood, J., Edwards, Y.J.K., and Jenkins, P.V. (1999). Structural and functional modelling of vWF-A domains in complement and coagulation. *Biochemical Society Transactions*, **27**, 815-821.

Hinshelwood, J., Spencer, D., Edwards, Y. J. K. and Perkins, S. J. (1999). Identification of the C3b binding site in a recombinant vWF-A domain of complement factor B by surface enhanced laser deionisation spectrometry and homology modelling: implications for the activation of factor B. *Journal of Molecular Biology* **294**, 587-599.

Hinshelwood, J. and Perkins, S. J. (2000). Metal-dependent conformational changes in a recombinant vWF-A domain from human factor B: A solution study by circular

dichroism, Fourier transform infrared and ^{1}H NMR spectroscopy. *Journal of Molecular Biology* **298**, 135-147.

Hinshelwood, J. and Perkins, S. J. (2000). Conformational changes during the assembly of factor B from its domains by ^{1}H NMR spectroscopy and molecular modelling: their relevance to the regulation of factor B activity *Journal of Molecular Biology* (submitted).

Biochem. J. (1999) **342**, 625–632 (Printed in Great Britain)

Production and functional activity of a recombinant von Willebrand factor-A domain from human complement factor B

Samantha C. WILLIAMS*[1], Justin HINSHELWOOD†, Stephen J. PERKINS†[2] and Robert B. SIM*

*MRC Immunochemistry Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K., and †Department of Biochemistry and Molecular Biology, Royal Free Campus, Royal Free and University College Medical School, University College London, Rowland Hill Street, London NW3 2PF, U.K.

Factor B is a five-domain 90 kDa serine protease proenzyme which is part of the human serum complement system. It binds to other complement proteins C3b and properdin, and is activated by the protease factor D. The fourth domain of factor B is homologous to the type A domain of von Willebrand Factor (vWF-A). A full-length human factor B cDNA clone was used to amplify the region encoding the vWF-A domain (amino acids 229–444 of factor B). A fusion protein expression system was then used to generate it in high yield in *Escherichia coli*, where thrombin cleavage was used to separate the vWF-A domain from its fusion protein partner. A second vWF-A domain with improved stability and solubility was created using a $Cys^{267} \rightarrow$ Ser mutation and a four-residue C-terminal extension of the first vWF-A domain. The recombinant domains were investigated by analytical gel filtration, sucrose density centrifugation and analytical ultracentrifugation, in order to show that both domains were monomeric and possessed compact structures that were consistent with known vWF-A crystal structures. This expression system and its characterization permitted the first investigation of the function of the isolated vWF-A domain. It was able to inhibit substantially the binding of ^{125}I-labelled factor B to immobilized C3b. This demonstrated both the presence of a C3b binding site in this portion of factor B and a ligand-binding property of the vWF-A domain. The site at which factor D cleaves factor B is close to the N-terminus of both recombinant vWF-A domains. Factor D was shown to cleave the vWF-A domain in the presence or absence of C3b, whereas the cleavage of intact factor B under the same conditions occurs only in the presence of C3b.

Key words: factor B, factor D, expression, mass spectroscopy, vWF-A domain.

INTRODUCTION

The complement system consists of a large number of soluble (plasma) and membrane-bound proteins that have a major role in innate immunity. Its function is to recognize foreign material and to facilitate its phagocytosis. Human complement factor B is a 90 kDa serine protease (SP) proenzyme present in plasma. When the complement system is activated, the major opsonin of complement, C3, is cleaved to form C3b. C3b forms a complex with factor B, whereupon factor D, a serine protease, cleaves factor B in the C3b–factor B complex, to form two fragments, Ba and Bb. Fragment Bb remains transiently bound to C3b, and the C3b–Bb complex itself contains an SP in the Bb fragment that activates more C3 [1,2]. The amplification of C3 activation, which occurs by this mechanism, is essential for opsonization of complement-activating materials.

When examined by electron microscopy, factor B presents a three-lobed globular structure [3]. The Ba fragment (approx. 30 kDa) consists of three complement control protein domains (also known as short consensus repeats) [4], each approx. 60 amino acids in length. The Bb fragment (60 kDa) consists of a C-terminal SP domain [2,5] and an N-terminal von Willebrand factor type A (vWF-A) domain [6,7]. vWF-A domains also occur in proteins of the extracellular matrix and in other proteins of the immune system (including complement C2, the leucocyte integrins CR3, CR4, LFA-1, and the VLA-1 and VLA-2 integrins) (reviewed in [6,7]). As this domain was first identified in von Willebrand factor [8,9], it is most frequently known as the vWF-A domain, but it is also known as an A-domain or I-domain. A total of 75 sequences in 25 proteins have been analysed to show a relatively low degree of sequence conservation and no conserved disulphide bridges [7]. Its protein fold was predicted to resemble the doubly wound α/β fold of nucleotide binding proteins [10,11], and vWF-A crystal structures have been solved to confirm the prediction [12,13] (reviewed in [9]).

Factor B, C2, CR3 and CR4 all interact with degradation fragments of C3, known as C3b or iC3b, or with those of C4, a homologue of C3. Since the vWF-A domain is common to Factor B/C2 and CR3/CR4, it is possible that this domain mediates ligand binding [14]. The evidence for this is mostly indirect. Site-directed mutagenesis of C3 near the N-terminus of the α'-chain of C3b [15] resulted in mutant C3b proteins that were less able to support cleavage of factor B by factor D than normal C3b, implying a reduced binding with factor B, and mutant forms of iC3b likewise have reduced binding to CR3. As the only common region of CR3 and factor B is a vWF-A domain, this result implied that this domain is responsible for C3b and iC3b binding. Mutation of the divalent metal-ion-binding sites in the vWF-A domain of CR3 abolished binding to iC3b [16].

A model for the C3b–factor B complex involves interaction via two (or more) sites on factor B [17–22]. A C3b-binding site in the Ba fragment was identified using a monoclonal anti-Ba antibody that inhibited the binding of ^{125}I-labelled factor B to red-cell-

Abbreviations used: MALDI, matrix-assisted laser desorption ionization; SP, serine protease; vWF-A, von Willebrand factor type A.

[1] Present address: Medical Research Council, Technology Transfer Group, 20 Park Crescent, London W1N 4AL, U.K.

[2] To whom correspondence should be addressed (steve@rfhsm.ac.uk).

bound C3b by up to 50% [18], and by the use of a cross-linking agent [17]. A C3b-binding site in the SP domain of factor B was identified by the direct binding of the isolated ^{125}I-labelled SP domain to C3b bound to zymosan in the presence of Mg^{2+} ions [21], and the SP domain also has Mg^{2+}-ion-independent binding affinity for fluid-phase C3b [20]. A C3b-binding site on the vWF-A domain was indirectly identified by a comparison of the Bb fragment and the SP domain in that study [20], and mutagenesis of recombinant factor B suggested that surface loops adjacent to the Mg^{2+}-site in the vWF-A domain influenced its C3b binding specificity [22]. As C2 and C4 are close homologues of factor B and C3 respectively, it is relevant that, in recombinant C2, mutations at residues 240–244 in the vWF-A domain affected the binding of C2a to C4b [23].

Study of the molecular function of the vWF-A domain in factor B requires the use of a suitable recombinant protein. Here we describe the development and characterization of two expression systems for isolated vWF-A domains. One of these has previously been used to study the vWF-A secondary structure by Fourier-transform IR spectroscopy [7]. Here, further evidence is presented to show that the recombinant vWF-A domain is folded and monomeric, and was therefore suitable for functional studies. The functional activity of the two recombinant vWF-A domains from factor B provide new insights into the role of vWF-A domains in complement activation.

MATERIALS AND METHODS

Materials

Glutathione–agarose beads (G-4510) and human thrombin (T-3010) were from Sigma (Poole, Dorset, U.K.), as were the constituents of the buffer used to lyse *E. coli* cells [DNAse (D-7291), lysozyme (L-6876), Triton X-100 (T-6878) and benzamidine (B-6506)], as well as lactoperoxidase (L-8257) used for radiolabelling of factor B. Thiol-Sepharose 4B, T^7 DNA sequencing kits and the expression vector pGEX-2T were from Pharmacia (Milton Keynes, Bucks., U.K.). Fresh-frozen plasma for the purification of complement components came from the Oxford Regional Blood Transfusion Service (John Radcliffe Hospital, Oxford, U.K.). The SP inhibitor Pefabloc-SC [4-(2-aminoethyl)benzenesulphonyl fluoride] was obtained from Pentapharm AG (Basel, Switzerland).

Vector construction

General procedures were as described previously [24]. The region encoding the vWF-A domain was amplified from a full-length factor B cDNA clone named pFB-2 [25] by PCR. The factor B cDNA clone was provided by Dr W. Schwaeble (Department of Microbiology and Immunology, University of Leicester, U.K.). The middle five exons of the factor B gene encode the vWF-A domain of factor B [5]. The first exon begins with Gly^{229} and the last ends with Ile^{444}. Oligonucleotides were designed to flank the region of interest with the introduction of a *Bam*H1 restriction site at the 5′-end and an *Eco*R1 restriction site at the 3′-end to enable sub-cloning into the pGEX-2T vector in the correct orientation (5′ oligonucleotide CG GGA TCC GGG GAA CAA CAG AAG CGG AAG; 3′ oligonucleotide GGA ATT CTA GAT CAT TTG GTA GAA AAC ATC TTC). The amplified cDNA insert (663 bp) produced was digested with *Eco*R1 and *Bam*H1 and ligated into the pGEX-2T vector pre-cut with the same enzymes. The ligation mixture was used to transform MC1061 competent cells. Individual bacterial clones containing cloned cDNA were identified by the production of correct-sized PCR products using the oligonucleotides described above from DNA mini-preps [26]. Sub-cloned DNA from a large-scale plasmid purification [27] was sequenced using a 'T^7 Sequencing Kit' following the manufacturer's protocol. The sequence was found to be identical with that found in the pFB-2 clone from which the fragment was amplified. The modifications to the ends of the amplified product were found to be as intended. As this recombinant product contained 218 residues, it was denoted vWF-A-218.

Expression and purification of vWF-A-218

vWF-A-218 was expressed and purified from *E. coli* by a combination of the methods described in [24,28]. Overnight cultures of *E. coli* transformed with the pGEX-2T constructs were diluted 1:10 with fresh Luria broth/ampicillin (100 mg/l) and grown for a further 1 h with shaking. At this point isopropyl β-D-thiogalactoside (Calbiochem-Novabiochem, Nottingham, U.K.; A10007) was added to 1 mM. After a further 6 h of growth the cells were pelleted by centrifugation (20 min; 4000 ***g***) and frozen at −80 °C overnight. Samples were then thawed and resuspended in lysis buffer (1/50 of the starting volume) and left on ice for 1 h. This was followed by sonication for 3 × 20 s on ice using an MSE Ultrasonic Processor (375 W) and a 6.5 mm tapered microtip. The lysis buffer was a hypertonic PBS (MTPBS) (150 mM NaCl/16 mM Na_2HPO_4/54 mM NaH_2PO_4, pH 7.2) containing 0.2 mg/ml lysozyme, 1 μg/ml DNAse, 0.1% (v/v) Triton X-100, 1 mM Pefabloc-SC, 5 μg/ml iodoacetamide, 50 mM benzamidine and 5 mM EDTA. After centrifugation (10000 ***g***; 20 min) the supernatant was incubated with glutathione–agarose beads (10 ml per litre of culture) at 4 °C overnight using a rotary stirrer. The beads were recovered by centrifugation (5 min; 500 ***g***) and washed five times with 2 vol. of MTPBS. To allow separation of vWF-A-218 from the glutathione S-transferase carrier protein, the fusion protein was cleaved directly on the beads with human thrombin. The fusion protein bound to 10 ml of packed resin (suspended in two volumes of MTPBS in a 50 ml conical centrifuge tube) was digested with 50 μg of human thrombin for 4 h at 37 °C with gentle mixing. The beads were then spun down (5 min; 1000 rev./min) and the supernatant containing the soluble expressed domain was pooled with four subsequent washes of the resin, each with 10 ml of MTPBS. The expressed domain was then concentrated 3-fold by partial freeze-drying and separated from contaminating *E. coli* proteins by gel filtration on an FPLC Superose-12 HR10/30 column (Pharmacia). Gel filtration at a flow rate of 0.3 ml/min was done using PBS (137 mM NaCl/2.6 mM KCl/8.2 mM Na_2HPO_4/1.5 mM KH_2PO_4, pH 7.3) that was prepared from Dulbecco 'A' tablets (Oxoid, Unipath Ltd, Basingstoke, U.K.). The pooled fractions containing purified vWF-A-218 [as assessed by SDS/PAGE on 10% (w/v) polyacrylamide gels] were stored at −20 °C with the addition of iodoacetamide to 2 mM. The absorption coefficient at 280 nm (1 cm, 1%) was calculated from the sequence [29] as 15.6 in order to determine protein concentrations and yields.

SDS/PAGE and N-terminal sequence analysis

For SDS/PAGE [30], vWF-A-218 samples were prepared for electrophoresis under reduced or unreduced (alkylated) conditions, and gels were stained using Coomassie Blue [31]. Samples for N-terminal sequence analysis were run on SDS/PAGE under reducing conditions and then electroblotted onto a Problott membrane (Applied Biosystems) in a Bio-Rad mini Trans-Blot electrophoretic transfer cell. The blots were stained with Coomassie Brilliant Blue and the bands corresponding to the protein of interest were excised. The N-terminal amino acid sequences

Table 1 Primers used for the mutation of the vWF-A-218 cDNA to vWF-A-222 by the 'QuickChange' method

The mut1/antimut1 and mut2/antimut2 sequences correspond to the $Cys^{267} \rightarrow$ Ser mutation and the four-residue C-terminal extension respectively. Characters in bold represent alterations from the original vWF-A cDNA sequence. The underlined TAG motif represents the stop codon required for termination of translation. The melting temperature T_m was calculated according to the Stratagene QuickChange manual. $T_m = 81.5 + 0.41(\%GC) - 675/N - \%$ mismatch

Identification	Amino acid and DNA sequences	Direction	Length (N)	GC	T_m (°C)
mut1	T G A K K **S** L V N C ACA GGA GCC AAA AAG **TCA** CTA GTC AAC	Forward	28	13	74
antimut1	GTT GAC TAG **TGA** CTT TTT GGC TCC TGT G	Reverse	28	13	74
mut2	V F Y Q M I **D E S Q** Stop GTT TTC TAC CAA ATG ATC **GAC GAG AGC CAG** TAG AAT TCA TCG TGA CTG	Forward	48	21	60
antimut2	CAG TCA CGA TGA ATT CTA **CTG GCT CTC GTC** GAT CAT TTG GTA GAA AAC	Reverse	48	21	60

were obtained using an Applied Biosystems 470A protein sequencer and Applied Biosystems 120A analyser.

Expression and purification of vWF-A-222

A second recombinant vWF-A domain, denoted vWF-A-222, was prepared by site-directed mutagenesis of the pGEX-2T plasmid containing the vWF-A-218 clone using the quick-change site-directed mutagenesis kit (QuickChange 200518, Stratagene, Cambridge, U.K.). Two cycles of mutagenesis were employed (Table 1). In the first, Cys^{267} was replaced by Ser and an *Spe*I restriction site (A/CTAGT) was introduced in the mut1 sequence. In the second, the C-terminus was extended by the next four residues of the factor B sequence (DESQ), and one *Bfa*I restriction site (C/TAG) was lost in the mut2 sequence. Oligonucleotide primers were synthesized at Perkin–Elmer Applied Biosystems (Warrington, Cheshire, U.K.). The mutant plasmids were ligated into Epicurian Coli XL1-Blue supercompetent cells for selection. These were identified by the modified restriction enzyme sites which gave unique digestion fragments. The restriction enzyme *Bfa*I was from New England Biolabs (Hitchin, Herts., U.K.), and the enzyme *Spe*I was from Promega (Southampton, U.K.). The double mutant plasmid was then transformed into *E. coli* MC1061/P3-competent cells (Invitrogen, Groningen, The Netherlands) for protein expression. vWF-A-222 was purified in a similar manner to vWF-A-218, except that iodoacetamide was not now required. The calculated vWF-A-222 absorption coefficient at 280 nm (1 cm, 1%) was 15.2 [29].

Hydrodynamic analyses of vWF-A-218 and vWF-A-222

In order to estimate the diffusion coefficient of vWF-A-218, its elution relative to those for standards was studied on the Superose 12 column described above. Diffusion coefficients of these standard proteins were plotted against their elution volumes (human ovalbumin, 7.8×10^{-7} cm$^2 \cdot$s^{-1}; soya bean trypsin inhibitor, 9.05×10^{-7} cm$^2 \cdot$s^{-1}; hen egg lysozyme, 11.2×10^{-7} cm$^2 \cdot$s^{-1} [32]). The diffusion coefficient yields the Stokes' radius a [33]:

$$a = kT/6\pi\eta D,$$

where k is the Boltzmann constant (1.4×10^{-16} erg/ °C), T is the temperature (293 K), η is the viscosity of water at 20 °C (0.01 centipoise) and D is the diffusion coefficient at 20 °C in water.

The sedimentation coefficient of vWF-A-218 was estimated by linear sucrose-density gradients (12 ml) of 5–40% sucrose in 10 mM Tris/HCl, pH 7.4 [32]. vWF-A-218 (50 μl; 300 μg/ml) in PBS was loaded onto the gradient. Centrifugation was carried out in a Beckman SW40 Ti rotor at 37000 rev./min for 12 h at 4 °C. Gradients were fractionated by peristaltic pumping from the base of the gradient, and fractions were analysed by measuring the absorbance at 280 nm and also by SDS/PAGE on 10% (w/v) polyacrylamide gels. The sedimentation coefficient was calculated by comparing the mobility of vWF-A-218 with those of standard proteins run under identical conditions (thyroglobulin, 19.2 S; bovine liver catalase, 11.2 S; BSA, 4.22 S; myoglobin, 2.04 S) [34]. Knowledge of the Stokes' radius a and the sedimentation coefficient gives the relative molecular mass M_r [35]:

$$M_r = (6\pi\eta N a\, s^\circ_{20,w})/(1 - \bar{v}\rho)$$

where N is Avogadro's number, $\bar{v}$ is the partial specific volume (calculated to be 0.739 ml/g from [29]) and ρ is the density of water at 20 °C. The frictional ratio f/f_0 was calculated from the ratio of Stokes radii a/a_0, where a_0 is the radius of the sphere corresponding in volume to the hydrated volume of vWF-A-218 in which the hydration was 0.3 g of H_2O/g of protein.

Analytical ultracentrifugation of vWF-A-218 and vWF-A-222 was performed on a Beckman XLi instrument in which the sample was monitored using its absorbance at 280 nm and its refractive index measured by interferometry. Sedimentation equilibrium data were acquired over 48 h in six-sector cells with column heights of 3 mm at rotor speeds of 6000, 8500, 11000 and 15000 rev./min until equilibrium had been reached at each speed. The data were analysed on the basis of a single species within the Beckman software provided as an add-on to Origin Version 4.1 (Microcal Inc.), where $\bar{v}$ for vWF-A-222 was calculated to be 0.738 ml/g [29]. Sedimentation velocity data were acquired over 8 h at rotor speeds of 42000 rev./min in two-sector cells with column heights of 12 cm, where scans were recorded at 15 min intervals. Sedimentation coefficients were derived using the transport method, also provided as an add-on within the Origin program. To calculate the sedimentation coefficient of vWF-A from the crystal structure of the homologous vWF-A domain in CR3 (PDB code 1ido [12]), this structure was converted into spheres of the same total volume as vWF-A from factor B, using a grid with cubes of side 0.43 nm, then the model was hydrated by the addition of a surface layer of spheres using the HYPRO program [36] to give a total of 535–545 spheres, from which the frictional coefficient was calculated using the GENDIA method [36,37].

Assay of functional (C3b-binding) activity of vWF-A-218

An assay system was developed to study factor B binding to C3b, using C3b immobilized on thiol-Sepharose via the free SH group in C3b [38]. C3 was purified and converted into C3b [39]. Factor B was purified by dye–ligand affinity chromatography from

fresh–frozen human plasma [40]. Aliquots of purified factor B (100 μg in 0.2–1 ml of PBS) were radioiodinated, using 0.5 mCi of $Na^{125}I$ (Amersham International, Bucks., U.K.), by lactoperoxidase-catalysed iodination using the method of Marchalonis [41] to a specific activity of 1×10^6 c.p.m./μg.

C3b–thiol-Sepharose was produced as follows. Briefly, the resin was reduced by adding dithiothreitol (final concentration, 50 mM) to a 1:1 (v/v) slurry (i.e. 1 vol. of packed resin:1 vol. of buffer) in PBS. The mixture was then incubated for 30 min at 37 °C. The resin was then washed with at least five vol. of PBS before incubation with C3b. Purified C3b in 25 mM Tris/HCl/140 mM NaCl, pH 8.2, was incubated with pre-swollen thiol-Sepharose (0.5 mg C3b/ml of resin) for 2 h at room temperature using a rotary stirrer. The percentage of C3b bound was estimated by measuring the A_{280} of the original solution and three subsequent washes with PBS. Approx. 20–25 % of the C3b bound to the resin, in which the final content of C3b was routinely 100–125 μg/ml of resin. Remaining free SH groups on the thiol-Sepharose resin were blocked by incubation with 5 mg of iodoacetamide per ml of resin for 1 h at room temperature. Non-specific binding sites on the C3b–thiol-Sepharose were blocked by incubation of the resin with 1 mg/ml BSA in PBS for 1 h at room temperature.

The effect of vWF-A-218 on the binding of radioiodinated factor B to C3b–thiol-Sepharose was assessed. In each assay, 50 μl of C3b–thiol-Sepharose suspended in 50 μl of binding buffer [10 mM Pipes (sodium salt), 30 mM NaCl, 1 mg/ml BSA, pH 7.0, containing 0.2 mM $MgCl_2$] was pre-incubated with serial two-fold dilutions of vWF-A-218 (200 μl; maximum concentration 125 μg/ml) for 2 h at 4 °C. ^{125}I-labelled factor B (4×10^5 c.p.m.) was then added and radioactivity bound to the C3b–thiol-Sepharose resin was measured after 1 h incubation at room temperature and five washes, each with 0.85 ml of binding buffer.

MS analyses

To measure the mass of vWF-A-218, matrix-assisted laser desorption ionization (MALDI) MS [42] was performed on a VG time-of-flight spectrometer. The sample was in 20 mM acetic acid and loaded in a 1:1 ratio with the matrix sinapinic acid (*trans*-3,5-dimethoxy-4-hydroxycinnamic acid) to aid in desorption. The instrument was calibrated externally using cytochrome *c* and insulin. To measure the mass of vWF-A-222, electrospray ionization MS was performed on a Finnigan Navigator spectrometer with a Waters Alliance liquid-chromatography system for sample delivery. The sample was in deionized water and was loaded at a 1:1 ratio with 50 % acetonitrile. The instrument was calibrated using water clusters. MALDI mass analysis was also performed with a PBS-1 mass analyser using a normal-phase chip surface (Ciphergen, Camberley, Surrey, U.K.). The vWF-A-218 and vWF-A-222 preparations (at 1 mg/ml in PBS containing 0.5 mM $MgCl_2$; Sigma, D-5773) were incubated with 2 % (w/w) recombinant factor D overnight at 37 °C and dialysed against distilled water before MS analysis. Each sample was mixed with an equal volume of matrix solution (50 % acetonitrile/1 % trifluoroacetic acid with sinapinic acid) before application to a normal-phase chip used in the mass analyser. The instrument was calibrated internally against bovine β-lactoglobulin A as standard. Recombinant factor D was a gift from Professor J. E. Volanakis (Department of Medicine, University of Alabama at Birmingham, AL, U.S.A.) [43]. Molecular masses were calculated from their amino acid sequences using averaged isotopic atomic masses in the PAWS MS program (http://www.proteometrics.com).

RESULTS

Expression and purification of the vWF-A-218 domain of factor B

The 218-residue vWF-A domain (denoted as vWF-A-218) was expressed as a fusion protein with glutathione S-transferase. The expected molecular masses from amino acid sequences were 51 kDa for the fusion protein, 26 kDa for the fusion partner glutathione S-transferase and 25 kDa for the vWF-A domain. Figure 1 shows from SDS/PAGE analysis that a major protein product was present with an apparant molecular mass of about 43 kDa in cell lysates, which has therefore migrated anomalously compared with its expected value. The fusion protein was stable even after 6 h of induction and showed no signs of degradation. Cleavage of the fusion protein bound to glutathione–agarose beads with thrombin resulted in the release of the recombinant domain into the supernatant, as confirmed by the 25 kDa band in Figure 1. The glutathione–agarose beads were suspended in 2 vol. of MTPBS before cleavage with thrombin, as smaller volumes resulted in protein precipitation in the supernatant. The yield was estimated as 7–10 mg/l of culture from the absorbance at 280 nm (see Materials and methods section). N-terminal sequence analysis of the recombinant product gave a homogeneous sequence NH_2-GS^{229}GEQQKRKIV..... The residues GS preceding the vWF-A domain sequence were derived from the thrombin-recognition site LVPRGS in the vector sequence. This sequence also contained the site at which factor D cleaves intact factor B between Arg-234 and Lys-235 [44]. MS of the product gave a molecular mass in agreement with the expected value for the vWF-A-218 domain (see below). On storage of the purified domain, a protein band was occasionally apparent at approx. twice the molecular mass of the recombinant domain when assessed by non-reducing SDS/PAGE on a 10 % (w/v) polyacrylamide gel. This band was not seen under reducing conditions

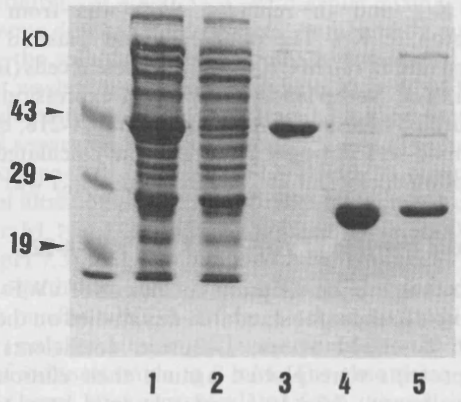


Figure 1 SDS/PAGE analysis on a 10% (w/v) polyacrylamide gel of stages in the expression and purification of the vWF-A-218 domain under reducing conditions

Aliquots of 10 μl were run on SDS/10%-PAGE and samples were stained with Coomassie Blue. Lane 1, total *E. coli* cell lysate after 6 h of induction with isopropyl β-D-thiogalactoside. The major band close to an apparent molecular mass of 43 kDa represents the glutathione S-transferase–vWF-A fusion protein, which has migrated anomalously compared with its expected value of about 50 kDa. Lane 2, total *E. coli* cell lysate after incubation with glutathione–agarose overnight at 4 °C to show that the fusion protein had been removed. Lane 3, fusion protein attached to the glutathione–agarose beads. Lane 4, material remaining bound to the glutathione beads after thrombin cleavage, where the major band at ~ 26 kDa is glutathione S-transferase. Lane 5, supernatant after thrombin cleavage that contains the soluble vWF-A-218 domain (~ 25 kDa).

and was therefore attributed to a disulphide-linked dimer of vWF-A-218, as there is a single free cysteine at position 267 [45]. Treatment with 2 mM iodoacetamide before storage prevented dimer formation.

For the large-scale production of vWF-A-218 for structural studies, the glutathione–agarose purification protocol was efficient, and over 90% purity was routinely obtained by thrombin cleavage. Since the solubility of the fusion protein was impaired if the *E. coli* cells were stored at −20 °C between harvest and lysis, maximal yields were achieved by the growth of cells, their lysis and the binding to glutathione–agarose within 1 day. The use of 5 mM dithiothreitol as a reducing agent during lysis inhibited dimer formation. After binding the fusion protein to glutathione–agarose, the dithiothreitol was washed out using degassed buffer before the thrombin cleavage step. Since freeze–drying promoted vWF-A aggregation, protein concentration was achieved instead by using Amicon stirred cells under nitrogen pressure. To prevent dimer formation during storage at 4 °C, the vWF-A-218 sample was exposed to 10 mM iodoacetamide for 1 min at 37 °C. Despite this, minor aggregation and precipitation occurred during storage, and further purification to better than 95% by size-exclusion chromatography on a HiLoad 16/60 Superdex 75 column was successfully used to separate the monomer from other forms. The occurrence of these aggregates reduced the net yield of purified vWF-A-218 to 3 mg/l of culture during storage, and its solubility did not exceed 3 mg/ml.

Development of the vWF-A-222 expression system

To improve the yield and solubility of vWF-A, two modifications were made. (1) Since dimerization was attributable to the single Cys^{267} residue within vWF-A-218, the sequence alignment of the vWF-A superfamily [7] was examined to investigate the significance of Cys^{267}. This was found not to be conserved in the superfamily, even though another free Cys residue is found in the vWF-A domain of complement C2 [7]. Whereas this thiol group in C2 was essential for the normal assembly and decay of the classical pathway C3 convertase, the function of Cys^{267} in native factor B was uncertain [45]. Thus a $Cys^{267} \rightarrow$ Ser mutation should eliminate dimerization with only a minor perturbation of the vWF-A structure, as both residues are polar and small. (2) The examination of vWF-A crystal structures showed that the two C-terminal residues of the vWF-A-218 domain were Met and Ile, both of which are hydrophobic and solvent-exposed. In addition, the C-terminal α-helix A12 was found to have been truncated (see Figure 1 of [11]). Since these may account for the reduced structural stability of vWF-A-218, the addition of the next four residues in factor B (DESQ) to vWF-A-218, all of which are hydrophilic, should assist α-helix formation and the stability and solubility of the vWF-A domain.

Primers were designed to adapt the vWF-A-218 domain to incorporate these two modifications in two steps, the final product being denoted vWF-A-222 (Table 1). The first and second mutations were verified by experimental observation of the expected changes in the restriction digestion pattern of the plasmid by *Eco*R1 and *Spe*1 (first mutation) and by *Bfa*1 (second mutation). In SDS/PAGE analysis of the vWF-A-222 cell culture, the cell lysate contained a major band at around 43 kDa. This fusion product was positive in a Western blot using factor B polyclonal antiserum. After purification (see Materials and methods section), both the cleaved vWF-A-218 and vWF-A-222 domains were again recognized in Western blot analyses using this polyclonal antiserum. By MS, both vWF-A-218 and vWF-A-222 showed a prominent single peak in clean spectra with masses of 24548 and 25002 respectively in MALDI and electrospray experiments. These masses were in good agreement with predicted masses of 24562 and 25005 respectively (see Materials and methods section). In large-scale preparations of vWF-A-222, yields were improved to 10 mg/l of culture, no dimerization or precipitation was observed during storage at 4 °C and the solubility was increased to about 8 mg/ml in PBS.

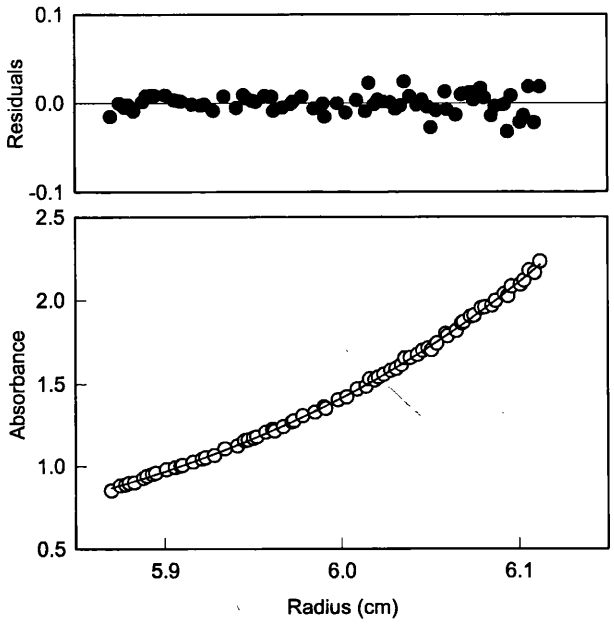


Figure 2 Sedimentation equilibrium data for the vWF-A-222 domain

The absorbance values at 280 nm are shown as a function of radial distribution at equilibrium for a loading concentration of 0.8 mg/ml at 20 °C and a rotor speed of 15000 rev./min. The data are fitted to a monomer model which is shown as a line through the experimental data points (○). The corresponding distribution of the residuals in the upper panel (●) is small and random, indicating that a good fit solution had been obtained.

Hydrodynamic properties of the recombinant vWF-A-218 and vWF-A-222 domains

Given the occurrence of aggregation and precipitation during storage, it was necessary to demonstrate for the functional analyses below that the vWF-A domain was monomeric in solution. Based on the elution positions of three standard proteins in gel filtration (Materials and methods section), the diffusion coefficient for vWF-A-218 was found to be between 8.4×10^{-7} and 9.9×10^{-7} $cm^2 \cdot s^{-1}$. From this, the Stokes' radius a of vWF-A-218 was found to be 2.4 ± 0.2 nm. Using sucrose-density-gradient centrifugation (see Materials and methods section) [32], the sedimentation coefficient was found to be 2.4 ± 0.2 S, based on comparison with four standard proteins. These data gave a vWF-A-218 molecular mass between 22 and 30 kDa, where the range reflects the errors in the determinations of a and $s^o_{20,w}$. The comparison with the expected molecular mass of 25 kDa from the sequence showed that vWF-A-218 was monomeric. The frictional ratio f/f_0 was calculated as 1.1 (see Materials and methods section); since this was close to 1.0 (the value for a sphere), this showed that vWF-A-218 possessed a compact structure.

To confirm quantitatively that the vWF-A-218 and vWF-A-222 domains were monomeric over the concentration range used for functional and structural studies, both were subjected to sedimentation equilibrium analysis (Figure 2). Curve fits on the

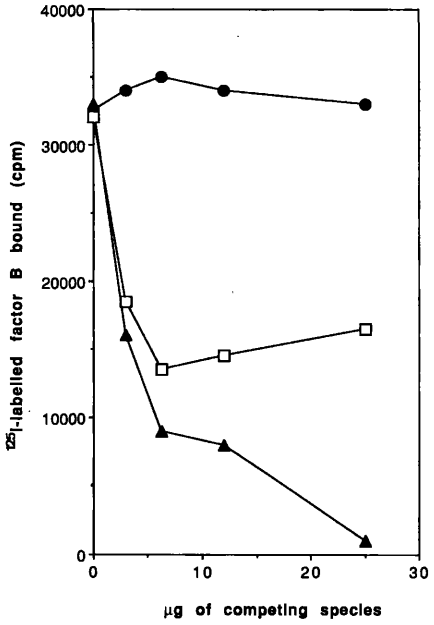


Figure 3 Effect of the recombinant vWF-A domain on the binding of ^{125}I-labelled factor B to C3b–thiol-Sepharose

In each tube 50 μl of C3b–thiol-Sepharose (containing 5 μg of C3b) was pre-incubated with serial 2-fold dilutions of intact unlabelled factor B (▲), vWF-A domain (□) or ovalbumin (●) (200 μl; maximum concentration 125 μg/ml) for 2 h at room temperature. The assays were all carried out in 10 mM Pipes (sodium salt)/0.2 mM $MgCl_2$/30 mM NaCl/1 mg/ml BSA, pH 7.0. ^{125}I-labelled factor B (4×10^5 c.p.m.) was then added to each tube and the bound radioactivity was measured after a 1 h incubation at room temperature and five washes with the binding buffer.

basis of a single monomeric species in a concentration range between 0.3–0.9 mg/ml gave molecular masses of 24000±2000 for vWF-A-218 and 27000±2000 for vWF-A-222, both of which agree with the expected values of 24500 and 25000 respectively. Sedimentation velocity experiments gave an $s^o_{20,w}$ value of 2.5±0.3 S for the vWF-A-218 and vWF-A-222 domains. Calculations of the sedimentation coefficient from the homologous vWF-A crystal structure from CR3 gave values of between 2.5 and 2.6 S. This comparison with the experimental values showed that the recombinant domains possessed monomeric structures that were similar to the CR3 crystal structure.

CD, Fourier-transform IR spectroscopy and NMR spectroscopy on the vWF-A-218 and vWF-A-222 domains demonstrated nearly identical results, which showed that both domains exhibited folded protein structures with α-helix and β-sheet [7]. This showed that the three-dimensional structure of vWF-A had not been perturbed in modifying the original expression vector.

Functional binding of the vWF-A-218 domain to C3b

The availability of a structurally characterized recombinant vWF-A-218 domain provided a unique opportunity to study its functional activity, as work on the isolated vWF-A domain had not previously been possible. Direct binding of ^{125}I-labelled vWF-A-218 to C3b–thiol-Sepharose was first investigated. Approx. 100 μg of purified vWF-A domain was radiolabelled to a specific activity of 1.2×10^6 c.p.m./μg using the lactoperoxidase method, and a test assay was set up to investigate whether direct binding of the vWF-A domain to C3b–thiol-Sepharose could be demonstrated. No direct binding of ^{125}I-labelled vWF-A was noted. Binding to C3b–thiol-Sepharose was no higher than to a control preparation of thiol-Sepharose which contained no C3b.

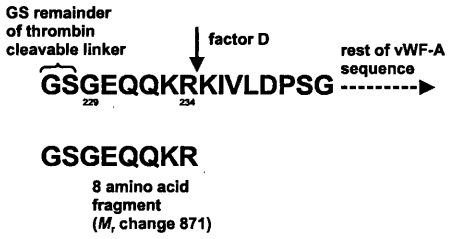


Figure 4 Representation of the factor D cleavage site in the recombinant vWF-A domain sequence

The potential interaction of the vWF-A-218 domain with C3b was then studied by a more sensitive indirect binding assay, based on the assessment of its ability to compete with ^{125}I-labelled factor B for binding to C3b attached to thiol-Sepharose. Using this, direct binding of ^{125}I-labelled factor B to C3b–thiol-Sepharose was observed in the presence of Mg^{2+} ions. Competition binding studies were carried out (Figure 3). The assay was validated using unlabelled factor B as a positive control and an irrelevant protein (ovalbumin) as a negative control. The maximum inhibition obtained with unlabelled factor B was close to 100% as expected. The recombinant vWF-A-218 domain was able to inhibit the binding of ^{125}I-labelled factor B to C3b–thiol-Sepharose to a maximum of 60%. This was achieved with 6.25 μg of vWF-A domain that was approx. an 8-fold molar excess over C3b. The same level of inhibition was achieved with approx. 4 μg of Factor B, a 1.7-fold molar excess over C3b.

Effect of factor D on vWF-A-218 and vWF-A-222

To test whether factor D could cleave the vWF-A-218 domain in the presence of C3b in the fluid phase, conditions were used in which factor D cleaves intact factor B. A mixture of vWF-A-218, C3b and factor D (in 10 mM Pipes/30 mM NaCl/0.2 mM $MgCl_2$, pH 7.0) was placed in a 1.5 ml tube at a 50:25:1 ratio by mass in a final volume of 200 μl [vWF-A (100 μg)/C3b (50 μg)/factor D (2 μg)] and incubated for 4 h at 37 °C. Cleavage of vWF-A-218 by factor D would result in the removal of an eight amino acid peptide (GSGEQQKR; Figure 4), so that the cleaved domain would have the new N-terminus KIVLDPSG. N-terminal sequence analysis (see Materials and methods section) was carried out on vWF-A-218 which had been incubated with C3b and factor D. Comparative sequencing was also carried out on vWF-A-218 which had been incubated with C3b only. The results obtained indicated that 10% of the vWF-A domain had been cleaved in the sample incubated with both C3b and factor D, whereas no cleavage was detected in the control in which factor D was omitted. These results suggest that factor D is capable of cleaving the vWF-A domain in the presence of C3b.

To test whether the vWF-A domain could be cleaved by factor D in the absence of C3b, vWF-A-218 and vWF-A-222 (50 μl at 1 mg/ml) were incubated with recombinant factor D (1 μl of factor D at 1 mg/ml) in PBS at 37 °C and pH 7.5 for 24 h. The products were dialysed into water, then mixed with matrix solution and applied to a mass analyser (see Materials and methods section). In both cases, the incubation resulted in the slow production of a cleaved vWF-A domain with a mass

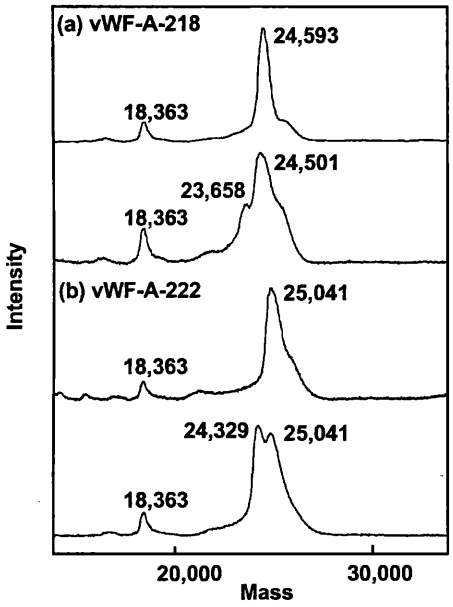


Figure 5 Effect of factor D on vWF-A-218 and vWF-A-222 by mass analysis

Mass spectra from (**a**) the vWF-A-218 and (**b**) the vWF-A-222 domains are shown. Molecular mass values are reported for each peak, based on internal calibrations against bovine β-lactoglobulin A (18363 Da). In each of (**a**) and (**b**), the upper trace corresponds to the purified domain and the lower trace to that of the domain after treatment with factor D for 24 h. In both lower traces, a second peak with a reduced molecular mass is visible and corresponds to the cleaved vWF-A domain when each is compared with its respective upper trace.

reduction of 796 ± 73 (mean $\pm$ S.D.; three measurements) (Figure 5). This agreed well with the expected mass reduction of 871 based on the removal of the N-terminal eight-residue peptide in the recombinant domains (Figure 4). In both cases, peaks corresponding to the uncleaved domain were also visible (Figure 5). Even if it was assumed that the efficiency of desorption was equal for cleaved and uncleaved vWF-A-218 and vWF-A-222, this result suggested that cleavage had occurred, and that factor D was active against the free vWF-A domain in the absence of C3b.

DISCUSSION

In this study, a full-length cDNA clone for factor B was used to amplify the 218 residues encoding the vWF-A domain (residues 229–444) by PCR, and this resulted in expression with high yield. Expression was validated from SDS/PAGE molecular-mass determination and amino acid sequencing. This recombinant system corresponded to the exon structure of the gene and gave monomeric protein in gel filtration and sucrose-density centrifugation, but led to slow disulphide-linked dimerization and aggregation and precipitation during storage. The inspection of our alignment for 75 vWF-A sequences and our analysis of the vWF-A crystal structure [7,11–13] enabled us to develop rationally a modified recombinant system in which a $Cys^{287} \rightarrow Ser$ mutation had been performed and the C-terminus of the vWF-A-218 domain had been lengthened by four residues to incorporate a C-terminal α-helix. With these changes, the protein stability was improved, and monomeric protein in solution was obtained (Figure 2). In illustration of the improvement, crystallization trials with vWF-A-222 have yielded larger better-formed crystals compared with the large number of small crystals obtained with vWF-A-218 (J. H. and S. J. P., unpublished work), and structural studies have been performed [46]. The observation of a compact solution structure is in agreement with the crystal structures of homologous vWF-A domains [12,13]. It is also consistent with the observation of three- and two-lobed structures for intact factor B and the Bb fragment by electron microscopy, since it can be assumed that the vWF-A domain would form one of these lobes. The production of eight contiguous vWF-A domains from the α-3 chain of collagen VI has been achieved using a eukaryotic expression system [47]. There, the structure of the expressed fragment was visualized to contain eight small globules using electron microscopy, which is consistent with the results on vWF-A domains and factor B.

The production of the vWF-A-218 domain enabled its functional binding to C3b to be studied. The studies carried out here agree with earlier suggestions for the role of the vWF-A domain in factor B [20]. Since the present work shows that the vWF-A domain can specifically inhibit the binding of factor B to C3b, it contains a binding site for C3b. Although the domain was unable to bind to C3b–thiol-Sepharose directly, it was able to compete with ^{125}I-labelled factor B for binding, the inhibition being maximally $\sim 60\%$ in the presence of 0.2 mM Mg^{2+}. The negative results from the direct binding studies are best explained by postulating that the binding of the vWF-A domain to C3b is weak and that it dissociates during the washing steps. Interaction of Ba with C3b was detected by studying inhibition, by Ba, of formation of the convertase C3bBb, but direct binding of unmodified Ba to C3b was not observed [17]. These observations are characteristic of a complex formed by multiple weak binding interactions, not one formed by a single-site high-affinity interaction.

The availability of the recombinant vWF-A domain also provided insight into the cleavage of factor B by factor D. Factor D circulates in its active form as there is an apparent absence of a zymogen form in blood. The enzyme is also highly specific, as it cleaves its only known substrate, factor B, only when factor B is bound to C3b or C3(H_2O) [44]. Several hexapeptides corresponding to the factor B sequence surrounding the bond that is cleaved by factor D were studied [48]. The peptides were assessed for their ability to inhibit factor D enzymic activity and for their susceptibility to cleavage by several SPs including factor D. The peptides were all able to inhibit factor B cleavage by factor D, but were not substrates for factor D. Active-site mapping of factor D with peptide thioesters revealed some interesting features [19,49]. Factor D was able to express esterolytic activity against some Arg thioesters, but its catalytic efficiency was found to be three or four orders of magnitude below that of trypsin and C1s. It was suggested on the basis of these results that the active site of factor D as it exists in serum has a zymogen-like conformation, which has an obstructed binding site. This view is supported by recent crystal structures for factor D which revealed a self-inhibitory loop in its structure [50]. The active conformation is then induced by its substrate the C3b–factor B complex.

Since the recombinant vWF-A domains contain factor D cleavage sites, the effect of factor D on vWF-A in the presence and absence of C3b was studied by two independent methods. It is significant that factor D was observed to cleave vWF-A slowly on its own in both cases, since this indicates that the free vWF-A domain is an adequate substrate for factor D. In contrast, if intact factor B was incubated with factor D in the presence of C3b, more than 95% of the factor B was efficiently converted into Ba and Bb. One possible explanation for the different extents of cleavage of free vWF-A domain and factor B is that the affinity of vWF-A for C3b is much less than that of intact

factor B, because of the existence of fewer contact points (Figure 3). This would mean that the concentration of the C3b–vWF-A complex is much lower than that of the C3b–factor B complex, so the cleavage of the C3b–vWF-A complex would be slower. Another possible explanation is based on the existence of two alternative conformations of the vWF-A domains. General evidence for these in the vWF-A superfamily is found from the distinct locations of mutation sites on the same vWF-A domain that corresponds to type 2B and 2M von Willebrand's disease [9], and several crystallographic studies of different structures in the presence and absence of metal in the vWF-A active site [13,51]. Thus, if the vWF-A domain of factor B has two conformations that depend on whether it is bound to C3b or to the remainder of the factor B structure, and only the C3b-bound vWF-A conformation is efficiently cleaved by factor D, the less efficient cleavage of the isolated vWF-A domain may correspond to the existence of both conformations of the isolated vWF-A domain when this is free in solution, only one of which is cleavable by factor D. Further kinetic and structural studies are required to resolve these questions.

The production of the isolated vWF-A domain from factor B has provided interesting insights into its previously uncharacterized function. Proteins that are members of the vWF-A domain superfamily have a wide variety of ligand-binding roles [6]. The vWF-A domain from factor B can now be included in the list of vWF-A domains that have ligand-binding functions.

We thank Dr. G. R. Stuart for assistance with the sucrose-density-gradient centrifugation, Mr. A. C. Willis for the protein sequence analysis, Dr. W. Schwaeble for supplying the factor B cDNA clone, Dr. A. Steinkasserer for help in constructing the vWF-A-218 expression vector, Professor J. E. Volanakis for the provision of factor D, the School of Pharmacy for access to MALDI and electrospray apparatus, and Dr. D. I. R. Spencer and Professor R. H. J. Begent for access to MS apparatus. S. C. W. held a Medical Research Council studentship, and R. B. S. thanks the Medical Research Council for support. J. H. and S. J. P. thank the Wellcome Trust for project grant support and support for the purchase of the Beckman XLi ultracentrifuge.

REFERENCES

1 Law, S. K. A. and Reid, K. B. M. (1995) Complement. 2nd. Edn., IRL Press, Oxford
2 Christie, D. L. and Gagnon, J. (1983) Biochem. J. **209**, 61–70
3 Smith, C. A., Vogel, C.-W. and Müller-Eberhard, H. J. (1984) J. Exp. Med. **159**, 324–329
4 Morley, B. J. and Campbell, R. D. (1984) EMBO J. **3**, 153–157
5 Campbell R. D., Bentley D. R. and Morley, B. J. (1984) Phil. Trans. R. Soc. Lond. (Biol.) **306**, 367–378
6 Colombatti, A. and Bonaldo, P. (1991) Blood **77**, 2305–2315
7 Perkins, S. J., Smith, K. F., Williams, S. C., Haris, P. I., Chapman, D. and Sim, R. B. (1994) J. Mol. Biol. **238**, 104–119
8 Sadler, J. E., Shelton-Inloes, B. B., Sorace, J. M., Harlan, J. M., Titani, K. and Davie, E. W. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 6394–6398
9 Jenkins, P. V., Pasi, K. J. and Perkins, S. J. (1998) Blood, **91**, 2032–2044
10 Edwards, Y. J. K. and Perkins, S. J. (1995) FEBS Lett. **358**, 283–286
11 Edwards, Y. J. K. and Perkins, S. J. (1996) J. Mol. Biol. **260**, 277–285
12 Lee, J.-O., Rieu, P., Arnaout, M. A. and Liddington, R. (1995) Cell **80**, 631–638
13 Lee, J.-O., Bankston, L. A., Arnaout, M. A. and Liddington, R. C. (1995) Structure **3**, 1333–1340
14 Sim, R. B. and Perkins, S. J. (1989) Curr. Topics Microbiol. Immunol. **153**, 209–222
15 Taniguchi-Sidle, A. and Isenman, D. E. (1994) J. Immunol. **153**, 5285–5302
16 Michishita, M., Videm, V. and Arnaout, M. A. (1993) Cell **72**, 857–867
17 Pryzdial, E. L. G. and Isenman, D. E. (1987) J. Biol. Chem. **262**, 1519–1525
18 Ueda, A., Kearney, J. F., Roux, K. H. and Volanakis, J. E. (1987) J. Immunol. **138**, 1143–1149
19 Volanakis, J. E. (1989) Curr. Topics Microbiol. Immunol. **153**, 1–21
20 Sanchez-Corral, P., Anton, L. C., Alcolea, J. M., Marques, G., Sanchez, A. and Vivanco, F. (1990) Mol. Immunol. **27**, 891–900
21 Lambris, J. D. and Müller-Eberhard, H. J. (1984) J. Biol. Chem. **259**, 12685–12690
22 Tuckwell, D. S., Xu, Y., Newham, P., Humphries, M. J. and Volanakis, J. E. (1997) Biochemistry **36**, 6605–6613
23 Horiuchi, T., Macon, K. J., Engler, J. A. and Volanakis, J. E. (1991) J. Immunol. **147**, 584–589
24 Steinkasserer, A., Solari, R., Mott, H. R., Aplin, R. T., Robinson, C. C., Willis, A. C. and Sim, R. B. (1992) FEBS Lett. **310**, 63–65
25 Schwaeble, W., Luttig, B., Sokolowski, T., Estaller, C., Weiss, E. H., Meyer zum Buschenfelde, K. H., Whaley, K. and Dippold, W. (1993) Immunobiology **188**, 221–232
26 Birnboim, H. C. and Doly, J. (1979) Nucleic Acids Res. **7**, 1513–1523
27 Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) In Molecular Cloning: A Laboratory Manual, Cold Spring Harbour Laboratory Press, NY
28 Smith, D. B. and Johnson, K. S. (1988) Gene **67**, 31–40
29 Perkins, S. J. (1986) Eur. J. Biochem. **157**, 169–180
30 Laemmli, U. K. (1970) Nature (London) **227**, 680–685
31 Fairbanks, G., Steck, T. L. and Wallach, D. F. (1971) Biochemistry **10**, 2606–2617
32 Martin, R. G. and Ames, B. N. (1961) J. Biol. Chem. **236**, 1372–1379
33 Ackers, G. K. (1964) Biochemistry, **3**, 723–730
34 Smith, M. H. (1970) in Handbook of Biochemistry and Selected Data for Molecular Biology, CRC Press, Boca Raton, FL
35 Siegel, L. M. and Monty, K. J. (1966) Biochim. Biophys. Acta **112**, 346–362
36 Ashton, A. W., Boehm, M. K., Gallimore, J. R., Pepys, M. B. and Perkins, S. J. (1997) J. Mol. Biol. **272**, 408–422
37 Ashton, A. W., Boehm, M. K., Johnson, D. J. D., Kemball-Cook, G. and Perkins, S. J. (1998) Biochemistry, **37**, 8208–8217
38 Williams, S. C. and Sim, R. B. (1994) Biochem. Soc. Trans. **22**, 2s
39 Dodds, A. W. (1993) Methods Enzymol. **223**, 46–61
40 Williams, S. C. and Sim, R. B. (1993) J. Immunol. Methods **157**, 25–30
41 Marchalonis, J. (1969) Biochem. J. **113**, 299–305
42 Beavis, R. C. and Chait, B. T. (1996) Methods Enzymol. **270**, 519–551
43 Yamauchi, Y., Stevens, J. W., Macon, K. J. and Volanakis, J. E. (1994) J. Immunol. **152**, 3645–3653
44 Lesavre, P. H., Hugli, T. E., Esser, A. F. and Müller-Eberhard, H.-J. (1979) J. Immunol. **123**, 529–534
45 Parkes, C., Gagnon, J. and Kerr, M. A. (1983) Biochem. J. **213**, 201–209
46 Hinshelwood, J., Williams, S. C., Sim, R. B. and Perkins, S. J. (1996) Molecular Immunol. **33**, 80 (abstract)
47 Specks, U., Mayer, U., Nischt, R., Spissinger, T., Mann, K., Timpl, R., Engel, J. and Chu, M. L. (1992) EMBO J. **11**, 4281–4290
48 Lesavre, P. H., Gaillard, M. H. and Halbwachs-Mecarelli, L. (1982) Eur. J. Immunol. **12**, 252–254
49 Kam, C. M., McRae, B. J., Harper, J. W., Niemann, M. A., Volanakis, J. E. and Powers, J. C. (1987) J. Biol. Chem. **262**, 3444–3451
50 Jing, H., Babu, Y. S., Moore, D., Kilpatrick, J. M., Liu, X.-Y., Volanakis, J. E. and Narayana, S. V. L. (1998) J. Mol. Biol. **282**, 1061–1081
51 Baldwin, E. T., Sarver, R. W., Bryant, G. L., Curry, K. A., Fairbanks, M. B., Finzel, B. C., Garlick, R. L., Heinrikson, R. L., Horton, N. C., Kelley, L. L. C. et al. (1998) Structure **6**, 923–935

Received 10 February 1999/21 June 1999; accepted 8 July 1999

Structure and Function of A-Domains

Society Special Colloquium Organized by D. Tuckwell and C. Kielty (School of Biological Sciences, University of Manchester) and Edited by D. Tuckwell. 669th Meeting held at the University of Keele, 20–22 July 1999.

Structural and functional modelling of von Willebrand factor type A domains in complement and coagulation

S. J. Perkins*[1], J. Hinshelwood*, Y. J. K. Edwards* and P. V. Jenkins†

*Department of Biochemistry and Molecular Biology, Royal Free Campus, Royal Free and University College Medical School, Rowland Hill Street, London NW3 2PF, U.K., and †Department of Haematology, Royal Free Campus, Royal Free and University College Medical School, Rowland Hill Street, London NW3 2PF, U.K.

Introduction

The von Willebrand factor type A (vWF-A) domain has been of great interest for reason of its widespread occurrence and its varied functions in many different proteins of the immune system and the extracellular matrix, as well as in blood coagulation [1]. In complement, it occurs once in each of the plasma proteins factor B and C2, and it is also found in the membrane proteins complement receptor types 3 and 4 (CR3 and CR4 respectively). Typically the vWF-A domain contains approximately 206 residues, and 75 vWF-A domains had already been sequenced by 1994 [2]. It was termed the type A domain since it corresponded to one of the four different domain types A–D that were characterized in the sequence of human von Willebrand factor (vWF), a protein important in coagulation [3]. For this reason, it became designated as the vWF-A domain in the SWISSPROT and PROSITE databases [4]. Other groups have termed this domain as an A-domain to correspond to its 'adhesive' properties, while those who work with the integrins have termed it as an 'inserted' or I-domain, as it occurs in some integrins but not in others [5].

During the past 5 years, significant progress on vWF-A domain structures has been reported, although there is still much to be done. One interesting aspect of structural studies on the vWF-A domain was the success of prediction methods in identifying the vWF-A secondary structure even before the first vWF-A crystal structure was determined [2,6], and likewise the successful identification of the protein fold before this crystal structure was published [7–9]. The latter stimulated much interest in new protein fold recognition methods, in which there is little similarity between a given sequence for an unknown protein structure and the sequence of a known protein fold, and yet features of that known protein fold are compatible with that starting sequence [10]. Both the use of modelling and its application to functional studies are discussed here. At present, modelling studies in combination with recent crystal structures raise the question of whether or not the vWF-A domain exhibits allosteric properties. The analysis of naturally occurring genetic mutations in the first vWF-A domain of vWF and the structural and functional properties of a recombinant vWF-A domain from complement factor B provided relevant information.

Structural studies of the vWF-A domain from factor B

Factor B is a five-domain 90 kDa serine protease proenzyme which is part of the human serum complement system [11]. It binds to the complement protein C3b and is activated by the protease factor D whereupon it becomes cleaved into the Ba

Abbreviations used: CR3, complement receptor type 3; GpIb, glycoprotein Ib; LFA-1, lymphocyte function-associated antigen-1; MIDAS, metal-ion-dependent adhesion site; SCR, short consensus/complement repeat; SP domain, serine protease domain; vWF-A, von Willebrand factor type A.

[1]To whom correspondence and requests for reprints should be addressed.

and Bb fragments. The C3b·Bb complex constitutes the alternative pathway C3 convertase, which is able to convert C3 into C3b and C3a in a rapid amplification loop. The Ba fragment contains three short consensus/complement repeat (SCR) domains (also known as complement control protein domains), while the Bb fragment contains a vWF-A domain and a serine protease (SP) domain. A recombinant vWF-A domain was successfully expressed in high yield as the middle five exons of human complement factor B, and a modified form of this exhibited improved stability [12,13]. Its functional activity was demonstrated by its ability to compete with radioiodinated factor B for binding to C3b, which showed that this vWF-A domain has a ligand-binding role. As the overall binding of factor B to C3b involves two or more weak interactions, which when combined give a higher affinity, the binding of recombinant vWF-A to C3b is typical of a system with multiple weak interactions, in which the SCR and SP domains in factor B also contribute to C3b binding [14–17].

Structural studies of this recombinant vWF-A domain showed that it behaved as a monomer in solution and had a compact folded structure [13]. Fourier transform infrared spectroscopy of the vWF-A domain in buffers based on both H_2O and ^2H_2O solvents revealed vibrational bands at frequencies that clearly corresponded to solvent-exposed α-helices and solvent-inaccessible buried β-strands [2]. The use of standard secondary structure predictions based on the classical Chou–Fasman and GOR approaches, but averaging the results over 75 vWF-A sequences, indicated an almost complete alternation of α-helix and β-strand secondary structures throughout the sequence alignment. The prediction and infrared data were therefore in full agreement with each other.

Comparison of the identified protein fold of the vWF-A domain with its crystal structure

The predictive and infrared work in [2] implied that the chances of identifying a known protein fold that resembled the vWF-A structure were high, so further predictive analyses were performed [7]. Sequence database searches for homologous proteins failed to reveal significant sequence similarity at the $> 30\%$ level. Other secondary structure prediction algorithms (PHD and SAPIENS) were therefore used to supplement the Chou–Fasman and GOR algorithms. Their consensus showed that six α-helices (A) and six β-strands (B) were predicted in the order BABBA-ABABABA. The presence of BAB repeats is typical of either an open twisted β-sheet structure flanked by α-helices above and below a buried parallel β-sheet, or a closed β-sheet barrel in which a buried core of parallel β-strands is flanked by a surface layer of α-helices [18]. There are no known closed β-barrels with less than eight β-strands, so the predicted presence of six β-strands meant that this general family of folds could be eliminated as a candidate vWF-A structure. Of the known doubly-wound α/β folds, the observed secondary structure for the GTP-binding domain of ras-p21, with five α-helices and six β-strands, came close to the predicted one for vWF-A. Protein fold recognition programs (which are based on matching side-chain pairwise potentials and solvation energies of a given sequence with a given three-dimensional topology) suggested that the ras-p21 fold gave a good match to the vWF-A sequence. The ras-p21 fold is a typical dinucleotide-binding protein. The active site occurs at a crevice at the centre of the carboxyl-edge of the β-sheet structure, from which two parallel β-strands lead towards opposite ends of the central β-sheet. Two well-conserved Asp residues in vWF-A had previously been shown to be important for Mg^{2+} binding [19]. When the position of these Asp residues was compared with the fold prediction, it was highly significant that both Asp residues were found adjacent to each other in the ras-p21 structure, and this observation completed the fold identification. Despite the lack of sequence or functional similarity between the vWF-A domain and the dinucleotide-binding fold, the position of several internal disulphide bridges in the vWF-A superfamily could also be explained on the basis of the ras-p21 protein fold, and this increased the confidence of the prediction.

When the vWF-A domain crystal structure from CR3 (CD11b) was subsequently published [6], this raised the question of the accuracy of the fold prediction. We found that the vWF-A secondary structure was predicted with 62–75% accuracy on a residue-by-residue basis, and 12 of the 13 secondary structure elements BABBAA-ABABABA were identified correctly [8,9]. The solvent accessibility predictions for individual residues were 69–71% accurate. It was also found that the fold recognition program analysis was much improved by averaging the results from 70 complete vWF-A sequences, and this showed that the ras-p21 fold, and also that for flavodoxin,

scored highly. Many other features of the fold prediction were largely accurate, in particular the predicted Mg^{2+}-binding site in the active-site cleft, which appeared within a five-residue MIDAS (metal-ion-dependent adhesion site) motif consisting of the residues Asp-Xaa-Ser-Xaa-Ser together with Thr and Asp residues from elsewhere in the structure [6]. The main difference from the original prediction is the reversal in direction of a β-hairpin at one end of the central β-sheet. This observation is important as it distinguishes the vWF-A protein fold from those of closely related dinucleotide-binding proteins such as ras-p21 or the flavin mononucleotide-binding protein flavodoxin, where the β-hairpin is absent (see Figure 3 of [8]). This difference is never mentioned by vWF-A crystallographers, yet is relevant to type 2B von Willebrand's disease!

Molecular basis of von Willebrand's disease

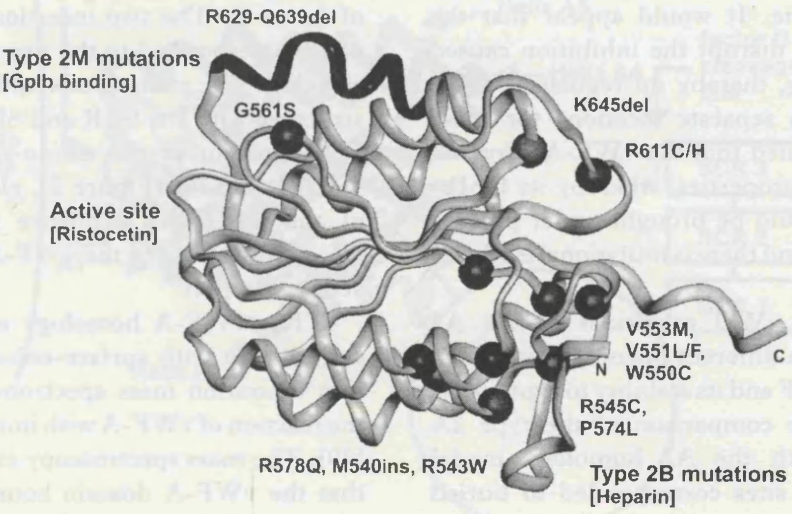
The vWF-A structure provided insights into the molecular basis of von Willebrand's disease, the most common inherited bleeding disorder [20]. vWF is a large multimeric, multidomain glycoprotein found in platelets, endothelial cells and plasma. Three vWF-A domains occur in vWF (termed A1, A2 and A3). These mediate binding to glycoprotein Ib (GpIb), ristocetin, botrocetin, collagen, sulphatides and heparin, and the A2 domain provides a physiologically relevant metalloprotease site involved in the cleavage of large vWF multimers. Genetic mutations causing types 2B, 2M and 2A von Willebrand's disease are located in the A1 and A2 domains. In the light of crystal structures for CR3 and the lymphocyte function-associated antigen-1 (LFA-1; CD11a) that were available at the time [6,21–23], homology modelling of all three vWF-A domains in vWF was undertaken in order to provide a molecular interpretation of vWF function and mutation sites. These were supplemented using the alignment of 75 vWF-A sequences [2] and the availability of 28 new sequences for the A1 and A2 domains from a wide range of mammalian species [24]. This detailed sequence data enabled residue conservation to be assessed in terms of their functional significance.

One key result from this work [25] was the absence of one of the two metal-binding Asp residues in the vWF-A active-site cleft in all three type A domains of vWF, so this implied that the

Figure 1

Location of the mutation sites in the crystal structure of the vWF A1 domain ([25,27]; PDB code 1auq) that are involved in type 2M and type 2B von Willebrand's disease

The central β-sheet is seen horizontally and edge-on, with the flanking α-helices above and below the β-sheet. The carboxyl-edge of the β-sheet is at the left side at which the active-site cleft is located, and the amino-edge is to the right. The positions of four type 2M mutations associated with GpIb and ristocetin binding are shown to the upper left as spheres or a dark ribbon, while those of eight type 2B mutations associated with a predicted heparin-binding site are shown to the lower right as spheres. The N-terminus and C-terminus are denoted by N and C respectively.



vWF-A active site did not bind metal. Instead, comparison with previous residue- and peptide-mapping studies showed that the active-site cleft at the crossover point at the carboxyl-edge of the central β-sheet of vWF-A corresponded to the ristocetin-binding site in the A1 domain (ristocetin is an antibiotic that stimulates vWF binding to platelets) and to the metalloprotease cleavage site in the A2 domain. The other key result from the homology modelling was the conservation of the vWF residues Arg-571, Lys-572, Arg-578, Arg-579 and Lys-585 in all 28 A1 mammalian sequences. These basic residues were not conserved in 72 other vWF-A sequences [2]. The location of these residues corresponded to the α-helices A2 and A3 at the amino-edge of the vWF-A fold. The spatial separation of the charged groups on these basic residues was consistent with a prediction that these constituted the vWF heparin-binding site in the A1 domain [25]; heparin is known to inhibit the interaction of vWF with GpIb.

The A1 homology model was able to provide molecular explanations of the mutations causing type 2B and type 2M von Willebrand's disease [25]. The type 2M mutations are involved in the down-regulation of GpIb binding to vWF. Both in the A1 model and in two subsequent crystal structures [26,27], these mutations were spatially clustered at the carboxyl-edge of the β-sheet and above it (Figure 1). The ristocetin-binding site is located at the active-site cleft, and the GpIb-binding site is located in the vicinity of this cleft. Distinct from this, the type 2B vWD mutations are involved in up-regulation of GpIb binding to vWF. Their mutation sites were spatially clustered at the amino-edge of this β-sheet and below it, on the opposite side of the A1 domain from the type 2M mutation sites (Figure 1). The type 2B mutations are comparatively close to the predicted heparin-binding site. It would appear that the type 2B mutations disrupt the inhibition caused by heparin binding, thereby up-regulating vWF function. The two separate locations for these mutation sites implied that the vWF-A1 domain displays allosteric properties, whereby its GpIb-binding activity could be brought under physiological regulation, and there is mutational evidence for this [28].

The type 2A vWD mutations in the A2 domain presented a different phenotype in which the secretion of vWF and its stability to proteolysis were affected. The comparison of the type 2A mutation sites with the A2 homology model showed that these sites corresponded to buried residues that were otherwise 100% conserved across all 28 mammalian species [25]. These residues are likely to be important for the protein folding of the A2 domain and the conformation of the surface metalloprotease site in the active-site cleft.

Structure–function studies of the vWF-A domain from factor B

By mid-1999, the crystal structures for five different vWF-A domains in CR3, LFA-1, very late activation protein-2 receptor (VLA-2; CD49b) and the A1 and A3 domains of vWF had been determined. A comparison of the MIDAS metal-binding motif in these structures shows that the Mg^{2+} co-ordination site with the two Asp residues shows slight differences between them. Metal binding is absent from the vWF A1 and A3 structures as predicted from sequences [2,24,25], although the α-helices A2 and A3 in these structures are combined into one. In addition, the C-terminal α-helix A7 displays conformational flexibility in the five structures (Figure 2). As there is presently no crystal structure for the vWF-A domain of factor B, its structure was modelled by standard homology methods [29]. Comparisons with the five different vWF-A crystal structures showed that the vWF-A domain from factor B (and its homologue C2 in the classical complement pathway) is significantly larger than any of these. All six hydrophobic β-strands in the vWF-A domain of factor B are well conserved, together with the α-helices A1 and A4, and could be readily identified in the five crystal structures. Together with the location of the five conserved metal-binding residues in the factor B and C2 sequences, this permitted the identification of four insertions in surface loop regions at the carboxyl- and amino-edges of the central β-sheet in the vWF-A domain of factor B. The two insertions at the carboxyl-edge corresponded to the two sides of the active-site cleft. The connections of the factor B vWF-A structure with the SCR and SP domains in factor B both occur at the amino-edge of the central vWF-A β-sheet (Figure 2), close to the position of the other two sequence insertions and on the opposite side of the vWF-A domain from the active-site cleft.

The vWF-A homology model was used in conjunction with surface-enhanced laser desorption ionization mass spectrometry to investigate the reaction of vWF-A with immobilized $C3(NH_3)$ [29]. The mass spectroscopy experiments showed that the vWF-A domain bound to $C3(NH_3)$, in

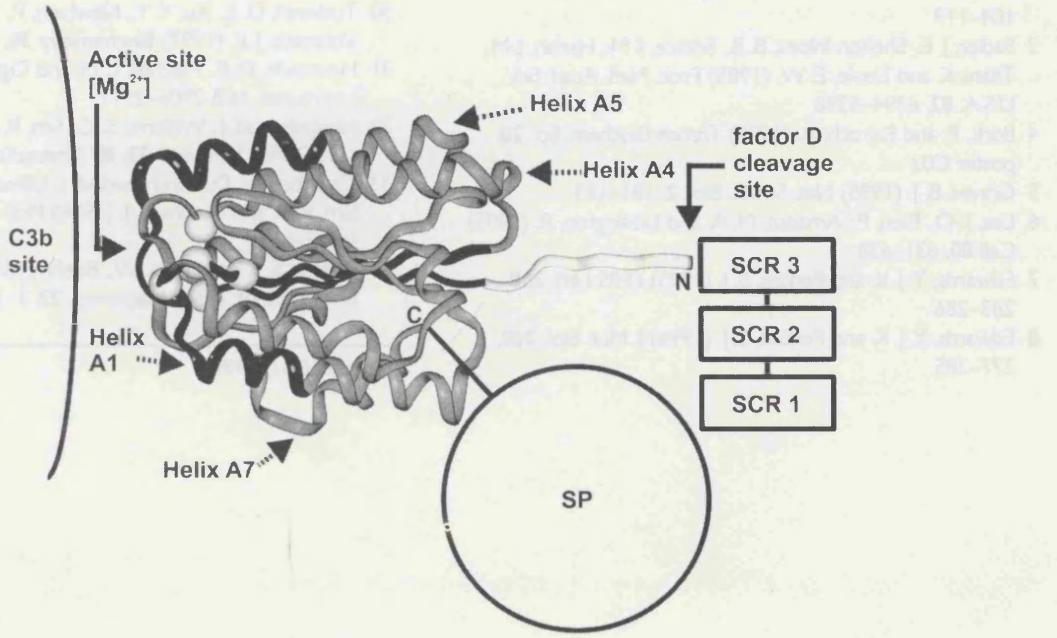
agreement with the C3b competition experiments [13]. The vWF-A domain was then subjected to tryptic proteolysis when it was bound to immobilized C3(NH_3). After washing, mass spectroscopy identified two vWF-A peptides (residues Gly-229–Lys-265 and Thr-355–Arg-381 involving the α-helices A1 and A5: black ribbons in Figure 2) that were still bound to C3(NH_3). The comparison of these peptides with the occurrence of surface-exposed residues and the elimination of buried residues in the vWF-A homology model showed unequivocally that a major function of the vWF-A domain is to bind to activated C3 at its active-site cleft (Figure 2). This is consistent with the results of site-specific mutagenesis studies of single residues or short peptides that are located in the active-site cleft [30,31].

The vWF-A homology model showed that the factor D cleavage site is eight residues away from its N-terminus (Figure 2). The recombinant vWF-A domain also included an extra eight residues that are N-terminal to the factor D cleavage site. It was possible for factor D to cleave this recombinant vWF-A domain both in the presence and in the absence of C3b, although this cleavage was slow in both cases [13]. In contrast to this, the cleavage of intact factor B by factor D occurs efficiently in the presence of C3b, but does not occur in the absence of C3b. These observations show that the free vWF-A domain is an adequate substrate for factor D. This means that this factor D cleavage site either is sterically concealed and/or it possesses an inactive conformation in intact factor B until it comes into contact with C3b. If the factor B vWF-A domain has allosteric properties by analogy with the vWF A1 domain, one explanation of the cleavage results for intact factor B is that the binding of the active site of vWF-A to C3b would induce a conformational change at its opposite face to up-regulate and render the factor D cleavage site accessible for cleavage (Figure 2). Another possible explanation would be that the independent binding and re-orientation of the SCR domains of intact factor B to another region of C3b would cause a conformational change in factor B to make this peptide available for cleavage by factor D. Irrespective of which explanation is followed, these experiments

Figure 2

Participation of the vWF-A domain of factor B in complement activation

The homology model is shown in the same view as in Figure 1. The active-site cleft at the carboxyl-edge of the central β-sheet with its five putative Mg^{2+}-binding residues (light spheres) is shown to the left, in proximity to the C3b-binding site identified by mass spectroscopy at the α-helices A1 and A5 (black ribbons). The N-terminus of the model (N) is connected to the three SCR domains of factor B by eight residues up to the factor D cleavage site (arbitrarily depicted as an extended light-coloured ribbon), and the C-terminus of the model (C) at the end of α-helix A7 is connected to the SP domain. The schematic outlines of the SCR and SP domains are not drawn to scale.



show that the role of the vWF-A domain in factor B is to facilitate the cleavage of the latter by factor D after binding to C3b. The release of the three SCR domains from factor B or an allosteric structural change in the vWF-A domain then brings about the activation of the SP domain. It is interesting that the proximity relationship of the SP domain of factor B which is connected to α-helix A7 of the vWF-A domain is similar to that for the predicted inhibitory heparin-binding site to the A1 domain in vWF on the nearby α-helices A2 and A3 (Figures 1 and 2), as this may suggest a common functional property of the same α-helical face in both domains.

Structural studies of the domains of factor B by NMR and by constrained solution scattering [32–34] are currently being completed, and will provide further details on the molecular mechanism of factor B activation. Many of the structural aspects of the vWF-A domain in factor B are expected to be applicable to the vWF-A domain of C2, the homologous protein found in the classical pathway of complement activation [11], since the C2 vWF-A sequence shows similarities with that of factor B.

We thank the Wellcome Trust for project grant support, and gratefully acknowledge the support of our collaborators Dr. R. B. Sim (MRC Immunochemistry Unit, University of Oxford, U.K.) and Dr. J. Pasi (Department of Haematology, Royal Free and University College Medical School).

1 Colombatti, A. and Bonaldo, P. (1991) Blood **77**, 2305–2315
2 Perkins, S. J., Smith, K. F., Williams, S. C., Haris, P. I., Chapman, D. and Sim R. B. (1994) J. Mol. Biol. **238**, 104–119
3 Sadler, J. E., Shelton-Inloes, B. B., Sorace, J. M., Harlan, J. M., Titani, K. and Davie, E. W. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 6394–6398
4 Bork, P. and Bairoch, A. (1995) Trends Biochem. Sci. **20**, poster C02
5 Graves, B. J. (1995) Nat. Struct. Biol. **2**, 181–183
6 Lee, J.-O., Rieu, P., Arnaout, M. A. and Liddington, R. (1995) Cell **80**, 631–638
7 Edwards, Y. J. K. and Perkins, S. J. (1995) FEBS Lett. **358**, 283–286
8 Edwards, Y. J. K. and Perkins, S. J. (1996) J. Mol. Biol. **260**, 277–285
9 Russell, R. B. and Sternberg, M. J. E. (1995) Curr. Biol. **5**, 488–490
10 Perkins, S. J., Ullman, C. G., Brissett, N. C., Chamberlain, D. and Boehm, M. K. (1998) Immunol. Rev. **163**, 237–250
11 Law, S. K. A. and Reid, K. B. M. (1995) Complement, 2nd edn., IRL Press, Oxford
12 Williams, S. C. and Sim, R. B. (1994) Biochem. Soc. Trans. **22**, 2S
13 Williams, S. C., Hinshelwood, J., Perkins, S. J. and Sim, R. B. (1999) Biochem. J. **342**, 625–632
14 Pryzdial, E. L. G. and Isenman, D. E. (1988) J. Biol. Chem. **263**, 1733–1738
15 Ueda, A., Kearney, J. F., Roux, K. H. and Volanakis, J. E. (1987) J. Immunol. **138**, 1143–1149
16 Sanchez-Corral, P., Anton, L. C., Alcolea, J. M., Marques, G., Sanchez, A. and Vivanco, F. (1990) Mol. Immunol. **27**, 891–900
17 Lambris, J. D. and Müller-Eberhard, H. J. (1984) J. Biol. Chem. **259**, 12685–12690
18 Branden, C. and Tooze, J. (1991) Introduction to Protein Structure, Garland Publishing Inc., New York
19 Michishita, M., Videm, V. and Arnaout, M. A. (1993) Cell **72**, 857–867
20 Sadler, J. E., Matsushita, T., Dong, Z., Tuley, E. A. and Westfield, L. A. (1995) Thromb. Haemostasis **74**, 161–166
21 Lee, J. O., Bankston, L. A., Arnaout, M. A. and Liddington, R. C. (1995) Structure **3**, 1333–1340
22 Qu, A. and Leahy, D. J. (1995) Proc. Natl. Acad. Sci. U.S.A. **92**, 10277–10281
23 Qu, A. and Leahy, D. J. (1996) Structure **4**, 931–942
24 Porter, C. A., Goodman, M. and Stanhope, M. J. (1996) Mol. Phylogenet. Evol. **5**, 89–101
25 Jenkins, P. V., Pasi, K. J. and Perkins, S. J. (1998) Blood **91**, 2032–2044
26 Celikel, R., Varughese, K. I., Madhusudan, Yoshioka, A., Ware, J. and Ruggeri, Z. M. (1998) Nat. Struct. Biol. **5**, 189–194
27 Emsley, J., Cruz, M., Handin, R. and Liddington, R. (1998) J. Biol. Chem. **273**, 10396–10401
28 Cooney, K. A. and Ginsburg, D. (1996) Blood **87**, 2322–2328
29 Hinshelwood, J., Spencer, D. I. R., Edwards, Y. J. K. and Perkins, S. J. (1999) Biochem. Soc. Trans. **27**, A144
30 Tuckwell, D. S., Xu, Y. Y., Newham, P., Humphries, M. J. and Volanakis, J. E. (1997) Biochemistry **36**, 6605–6613
31 Hourcade, D. E., Mitchell, L. M. and Oglesby, T. J. (1999) J. Immunol. **162**, 2906–2911
32 Hinshelwood, J., Williams, S. C., Sim, R. B. and Perkins, S. J. (1996) Mol. Immunol. **33**, 80 (abstract)
33 Chamberlain, D., Hinshelwood, J., Ullman, C. G., Smith, K. F., Sim, R. B. and Perkins, S. J. (1996) Mol. Immunol. **33**, 80 (abstract)
34 Perkins, S. J., Ashton, A. W., Boehm, M. K. and Chamberlain, D. (1998) Int. J. Biol. Macromol. **22**, 1–16

Received 6 July 1999

show that the role of the vWF-A domain in factor B is to facilitate the cleavage of the latter by factor D after binding to C3b. The release of the three SCR domains from factor B or an allosteric structural change in the vWF-A domain then brings about the activation of the SP domain. It is interesting that the proximity relationship of the SP domain of factor B which is connected to α-helix A7 of the vWF-A domain is similar to that for the predicted inhibitory heparin-binding site to the A1 domain in vWF on the nearby α-helices A2 and A3 (Figures 1 and 2), as this may suggest a common functional property of the same α-helical face in both domains.

Structural studies of the domains of factor B by NMR and by constrained solution scattering [32–34] are currently being completed, and will provide further details on the molecular mechanism of factor B activation. Many of the structural aspects of the vWF-A domain in factor B are expected to be applicable to the vWF-A domain of C2, the homologous protein found in the classical pathway of complement activation [11], since the C2 vWF-A sequence shows similarities with that of factor B.

We thank the Wellcome Trust for project grant support, and gratefully acknowledge the support of our collaborators Dr. R. B. Sim (MRC Immunochemistry Unit, University of Oxford, U.K.) and Dr. J. Pasi (Department of Haematology, Royal Free and University College Medical School).

1 Colombatti, A. and Bonaldo, P. (1991) Blood **77**, 2305–2315
2 Perkins, S. J., Smith, K. F., Williams, S. C., Haris, P. I., Chapman, D. and Sim R. B. (1994) J. Mol. Biol. **238**, 104–119
3 Sadler, J. E., Shelton-Inloes, B. B., Sorace, J. M., Harlan, J. M., Titani, K. and Davie, E. W. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 6394–6398
4 Bork, P. and Bairoch, A. (1995) Trends Biochem. Sci. **20**, poster C02
5 Graves, B. J. (1995) Nat. Struct. Biol. **2**, 181–183
6 Lee, J.-O., Rieu, P., Arnaout, M. A. and Liddington, R. (1995) Cell **80**, 631–638
7 Edwards, Y. J. K. and Perkins, S. J. (1995) FEBS Lett. **358**, 283–286
8 Edwards, Y. J. K. and Perkins, S. J. (1996) J. Mol. Biol. **260**, 277–285
9 Russell, R. B. and Sternberg, M. J. E. (1995) Curr. Biol. **5**, 488–490
10 Perkins, S. J., Ullman, C. G., Brissett, N. C., Chamberlain, D. and Boehm, M. K. (1998) Immunol. Rev. **163**, 237–250
11 Law, S. K. A. and Reid, K. B. M. (1995) Complement, 2nd edn., IRL Press, Oxford
12 Williams, S. C. and Sim, R. B. (1994) Biochem. Soc. Trans. **22**, 2S
13 Williams, S. C., Hinshelwood, J., Perkins, S. J. and Sim, R. B. (1999) Biochem. J. **342**, 625–632
14 Pryzdial, E. L. G. and Isenman, D. E. (1988) J. Biol. Chem. **263**, 1733–1738
15 Ueda, A., Kearney, J. F., Roux, K. H. and Volanakis, J. E. (1987) J. Immunol. **138**, 1143–1149
16 Sanchez-Corral, P., Anton, L. C., Alcolea, J. M., Marques, G., Sanchez, A. and Vivanco, F. (1990) Mol. Immunol. **27**, 891–900
17 Lambris, J. D. and Müller-Eberhard, H. J. (1984) J. Biol. Chem. **259**, 12685–12690
18 Branden, C. and Tooze, J. (1991) Introduction to Protein Structure, Garland Publishing Inc., New York
19 Michishita, M., Videm, V. and Arnaout, M. A. (1993) Cell **72**, 857–867
20 Sadler, J. E., Matsushita, T., Dong, Z., Tuley, E. A. and Westfield, L. A. (1995) Thromb. Haemostasis **74**, 161–166
21 Lee, J. O., Bankston, L. A., Arnaout, M. A. and Liddington, R. C. (1995) Structure **3**, 1333–1340
22 Qu, A. and Leahy, D. J. (1995) Proc. Natl. Acad. Sci. U.S.A. **92**, 10277–10281
23 Qu, A. and Leahy, D. J. (1996) Structure **4**, 931–942
24 Porter, C. A., Goodman, M. and Stanhope, M. J. (1996) Mol. Phylogenet. Evol. **5**, 89–101
25 Jenkins, P. V., Pasi, K. J. and Perkins, S. J. (1998) Blood **91**, 2032–2044
26 Celikel, R., Varughese, K. I., Madhusudan, Yoshioka, A., Ware, J. and Ruggeri, Z. M. (1998) Nat. Struct. Biol. **5**, 189–194
27 Emsley, J., Cruz, M., Handin, R. and Liddington, R. (1998) J. Biol. Chem. **273**, 10396–10401
28 Cooney, K. A. and Ginsburg, D. (1996) Blood **87**, 2322–2328
29 Hinshelwood, J., Spencer, D. I. R., Edwards, Y. J. K. and Perkins, S. J. (1999) Biochem. Soc. Trans. **27**, A144
30 Tuckwell, D. S., Xu, Y. Y., Newham, P., Humphries, M. J. and Volanakis, J. E. (1997) Biochemistry **36**, 6605–6613
31 Hourcade, D. E., Mitchell, L. M. and Oglesby, T. J. (1999) J. Immunol. **162**, 2906–2911
32 Hinshelwood, J., Williams, S. C., Sim, R. B. and Perkins, S. J. (1996) Mol. Immunol. **33**, 80 (abstract)
33 Chamberlain, D., Hinshelwood, J., Ullman, C. G., Smith, K. F., Sim, R. B. and Perkins, S. J. (1996) Mol. Immunol. **33**, 80 (abstract)
34 Perkins, S. J., Ashton, A. W., Boehm, M. K. and Chamberlain, D. (1998) Int. J. Biol. Macromol. **22**, 1–16

Received 6 July 1999

Reprinted from *J. Mol. Biol.* (1999) **294**, 587–599

JMB

Identification of the C3b Binding Site in a Recombinant vWF-A Domain of Complement Factor B by Surface-enhanced Laser Desorption-ionisation Affinity Mass Spectrometry and Homology Modelling: Implications for the Activity of Factor B

Justin Hinshelwood, Daniel I. R. Spencer, Yvonne J. K. Edwards and Stephen J. Perkins

Article No. jmbi.1999.3223 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (1999) 294, 587–599

JMB

Identification of the C3b Binding Site in a Recombinant vWF-A Domain of Complement Factor B by Surface-enhanced Laser Desorption-ionisation Affinity Mass Spectrometry and Homology Modelling: Implications for the Activity of Factor B

Justin Hinshelwood[1], Daniel I. R. Spencer[2], Yvonne J. K. Edwards[1] and Stephen J. Perkins[1]*

[1]*Department of Biochemistry and Molecular Biology, and*

[2]*Department of Oncology Royal Free Campus, Royal Free and University College Medical School, Rowland Hill Street London NW3 2PF, UK*

**Corresponding author*

Factor B is a key component of the alternative pathway of the complement system. During complement activation, factor B complexed with activated C3 is cleaved into the Ba and Bb fragments by the protease factor D to form the C3 convertase from the complex between C3b and Bb. The Ba fragment contains three short consensus/complement repeat (SCR) domains, and the Bb fragment contains a von Willebrand factor type A (vWF-A) domain and a serine protease (SP) domain. Surface-enhanced laser desorption-ionization affinity mass spectrometry (SELDIAMS) was used to investigate the reaction of factor B with immobilised activated C3(NH_3) in the presence of Mg^{2+}. A recombinant vWF-A domain (residues G229-Q448), the native Ba and Bb fragments and native factor B all demonstrated specific interactions with C3(NH_3), while no interactions were detected using bovine serum albumin as a control. A mass analysis of the proteolysis of the vWF-A domain when this was bound to immobilised C3(NH_3) identified two peptides (residues G229-K265 and T355-R381) that were involved with vWF-A binding to C3(NH_3). A homology model for the vWF-A domain was constructed using the vWF-A crystal structure in complement receptor type 3. Comparisons with five different vWF-A crystal structures showed that large surface insertions were present close to the carboxyl and amino edges of the central β-sheet of the factor B vWF-A structure. The peptides G229-K265 and T355-R381 corresponded to the two sides of the active site cleft at the carboxyl edge of the vWF-A structure. The vWF-A connections with the SCR and SP domains were close to the amino edge of this vWF-A β-sheet, and shows that the vWF-A domain can be involved in both C3b binding and the regulation of factor B activity. These results show that (i) a major function of the vWF-A domain is to bind to activated C3 during the formation of the C3 convertase, which it does at its active site cleft; and that (ii) SELDIAMS provides an efficient means of identifying residues involved in protein-protein interactions.

Keywords: complement C3; factor B; vWF-A domain; SELDIAMS; homology modelling

Present address: Y. J. K. Edwards, UK HGMP Resource Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SB, UK.

Abbreviations used: SCR, short consensus/complement repeat; SELDIAMS, surface-enhanced laser desorption-ionisation affinity mass spectrometry; SP, serine protease domain; vWF-A, von Willebrand factor type A domain.

E-mail address of the corresponding author: steve@rfhsm.ac.uk

0022-2836/99/470587–13 $30.00/0

Introduction

The complement system is an important component of immune defence against infection, comprising over 30 proteins which function as enzymes, binding proteins, regulators or membrane-bound receptors. It is part of the innate (non-adaptive) immune system which responds to challenges by micro-organisms before an adaptive response has developed (Law & Reid, 1995). Factor B is a key component of the alternative pathway of complement activation. It is cleaved into the Ba and Bb fragments in the presence of C3b, Mg^{2+} and factor D. This leads to the formation of the alternative pathway C3 convertase, the complex between C3b and Bb, that converts C3 into C3b and C3a. The formation of C3b leads to the exposure of the thiolester group which is buried in native C3, and its deposition on an appropriate surface initiates a rapid positive amplification loop. C3u is another activated form of C3 in which the internal thiolester is hydrolysed by water, and C3u undergoes a conformational change like that found in C3b.

Factor B contains five domains. The N-terminal three domains are short consensus/complement repeat (SCR; also known as complement control protein) domains that correspond to the Ba fragment, and the remaining two are a von Willebrand factor type A (vWF-A) domain and a serine protease (SP) domain in the Bb fragment. vWF-A domains (also known as "adhesive" or A-domains and "inserted" or I-domains) occur in many cell surface receptor proteins and proteins of the extracellular matrix. The vWF-A protein fold was first identified by prediction analyses that revealed that it was similar to that of a dinucleotide-binding fold with amphiphatic α-helices surrounding a central hydrophobic, largely parallel six-stranded β-sheet (Edwards & Perkins, 1995, 1996). Since then, 14 crystal structures have been solved for five different vWF-A domains, namely complement receptor type 3 (CR3; CD11b), lymphocyte function-associated antigen-1 (LFA-1; CD11a), very late activation protein receptor-2 (VLA-2; CD49b), and the A1 and A3 domains of von Willebrand factor (Lee *et al.*, 1995a,b; Baldwin *et al.*, 1998; Qu & Leahy, 1995, 1996; Emsley *et al.*, 1997, 1998; Celikel *et al.*, 1998; Huizinga *et al.*, 1997; Bienkowska *et al.*, 1997). Factor B, C2, CR3, and complement receptor type 4 are four complement proteins that each contain a vWF-A domain. As all four interact with various forms of C3 or C4 (a homologue of C3), in which C3b and C3u are the activated forms of C3, and iC3b is a cleaved form of C3b, it is likely that the vWF-A domain mediates this ligand binding with C3 or C4. In confirmation of this, the vWF-A domain of factor B binds to C3b (Williams *et al.*, 1999), and mutagenesis of residues in the vWF-A domain of intact factor B affects its binding to C3b (Tuckwell *et al.*, 1997; Hourcade *et al.*, 1999). Competition binding studies have demonstrated that both Ba and Bb fragments can separately inhibit the binding of factor B to immobilised C3u (Williams & Sim, 1994). The SCR domains in the Ba fragment interact with C3b or C3u (Hourcade *et al.*, 1995; Pryzdial & Isenman, 1988; Ueda *et al.*, 1987). The SP domain also demonstrates C3b binding activity (Lambris & Müller-Eberhard, 1984; Sanchez-Corral *et al.*, 1990). Despite the evidence for these protein-protein interactions, structural knowledge of the residues responsible for these is presently incomplete.

The SP domain of factor B is the site of catalytic action of the C3 convertase of the alternative pathway. This SP domain differs from the digestive and blood coagulation serine proteases by possessing a sizeable vWF-A domain at its N terminus. The cleavage leading to the activation of factor B occurs at the N terminus of the vWF-A domain, in contrast to the cleavage of small N-terminal activation peptides in these other SP domains. The clarification of the relationship between the vWF-A and SP domains in factor B is therefore of great interest.

Matrix-assisted laser desorption ionisation mass spectrometry is a powerful method for the investigation of protein-protein interactions (Beavis & Chait, 1996; Krone *et al.*, 1997). In application to proteins and peptides, a matrix compound such as sinapinic acid is used to assist the vapourization of these from a chip surface, and the mass spectrometer uses the difference in mass-to-charge ratio (m/z) to separate these and determine their masses to high accuracies (Krone *et al.*, 1997). This accuracy is sufficient to identify the peptide sequences in question. Matrix-assisted laser desorption ionisation mass spectrometry has enabled proteolytic fragmentation patterns to be identified in protein-protein complexes (Beavis & Chait, 1996), and this provides structural information on the residues that participate at the interface between the two proteins in the complex. The use of an activated chip surface to which one of the two proteins can be bound covalently simplifies the preparation and identification of these proteolytic fragments (Spencer *et al.*, 1999). This method is known as surface-enhanced laser desorption ionisation affinity mass spectrometry (SELDIAMS) (Hutchens & Yip, 1993; Kuwata *et al.*, 1998; Leung *et al.*, 1998). Here, we describe the application of SELDIAMS to study the interactions between C3b and factor B based on a recombinant vWF-A domain, the native Ba and Bb fragments and native factor B. The interpretation of the peptide data using a homology model for the factor B vWF-A domain based on the knowledge of 14 homologous crystal structures for five different vWF-A domains enabled us to demonstrate that the active site crevice at the carboxyl edge of the β-sheet structure of the vWF-A domain binds directly to C3b, and it was inferred that the amino edge of this β-sheet provides the connections with the SCR and SP domains of factor B. The implications of these observations are discussed.

Results and Discussion

Characterisation of the protein preparations

The preparations of the 222-residue recombinant vWF-A domain, the native Ba and Bb fragments, and native factor B and C3 each eluted as a single clean peak from the HiLoad 16/60 Superdex-75 gel filtration column. Except for C3, each migrated as a single band on SDS-PAGE under both reducing and non-reducing conditions (data not shown), and corresponded to the following expected approximate sizes: vWF-A, 25 kDa; Ba, 30 kDa; Bb, 60 kDa; and factor B, 90 kDa. The vWF-A domain was shown to be folded by Fourier transform infrared, circular dichroism and NMR spectroscopy (Perkins *et al.*, 1994; Hinshelwood *et al.*, 1996). The vWF-A domain was shown to be functionally active through its ability to inhibit the binding of factor B to C3b (Williams *et al.*, 1999). C3 migrated as a single band in SDS-PAGE under non-reducing conditions, and as two bands in reducing conditions, which corresponded to the expected sizes of 180 kDa for C3, and 110 kDa and 70 kDa for its α and β chains.

Mass spectrometry of immobilised C3(NH_3) with factor B and its components

C3(NH_3) was used as an equivalent of C3b, the activated form of C3 in which nucleophilic attack by free ammonia at the thiolester site converted it into the C3b conformation, which binds to factor B (Sim & Sim, 1983; von Zabern *et al.*, 1981; Pryzdial & Isenman, 1988). C3(NH_3) was covalently immobilised to the preactivated chip (see Materials and Methods). When this was probed with the recombinant carbohydrate-free vWF-A domain, a peak at a mass of 25,021 Da was observed in the resulting mass spectrum (Figure 1(a)). This peak was present even if the activated chip surface was washed with 50 mM urea (data not shown), and its mass value agreed well with a calculated value of 25,005 Da from the PAWS program (see Materials and Methods). In control experiments, no peak was observed in this region of the spectrum when the chip was probed with the vWF-A domain in the absence of immobilised C3(NH_3) (Figure 1(b)) or when the immobilised C3(NH_3) was run alone (Figure 1(c)). It was concluded that this vWF-A domain exhibited a specific affinity for C3(NH_3), in agreement with the competition experiments by Williams *et al.* (1999).

The immobilised C3(NH_3) was likewise probed with the glycosylated native Ba and Bb fragments and factor B. Broad peaks of masses 26,786 Da, 61,215 Da and 90,273 Da, respectively, were observed in the mass spectra with estimated errors of ±800 Da (Figure 1(d), (e) and (f), respectively). These are in approximate agreement with the expected masses of the three glycoproteins. The calculated masses of the protein component were 25,983 Da, 57,036 Da and 83,001 Da, respectively,

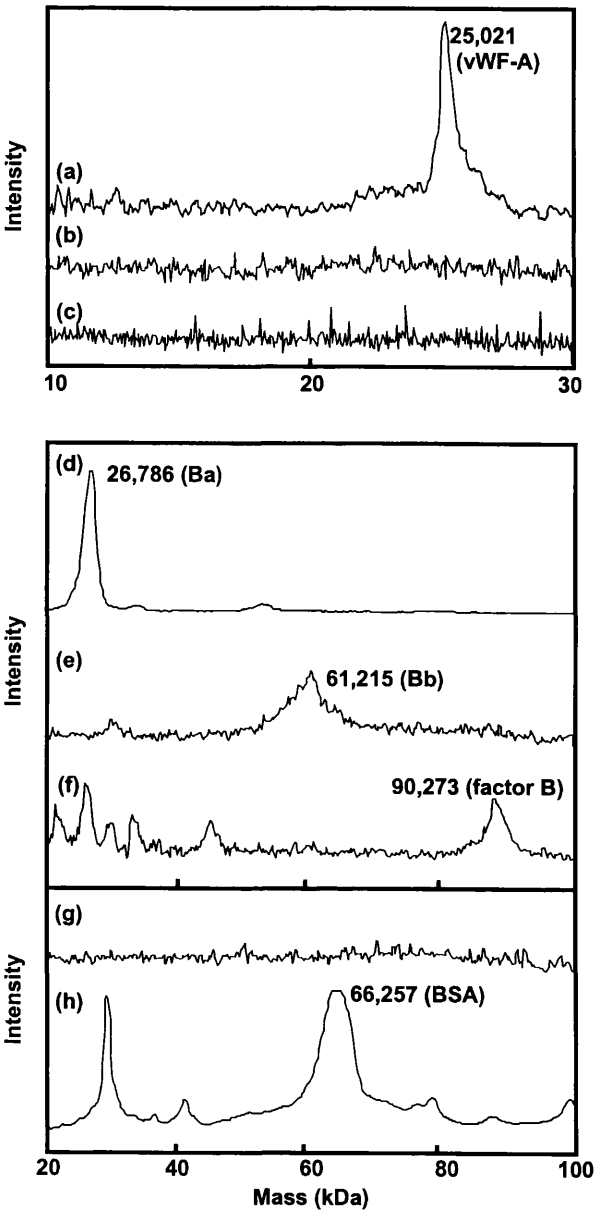


Figure 1. Mass analysis of the interaction of the vWF-A domain, the Ba and Bb fragment, factor B and bovine serum albumin with immobilised C3(NH_3). (a) C3(NH_3) was immobilised on the surface of the activated chip, then probed with vWF-A; (b) no C3(NH_3) was present on the surface, but the chip was probed with vWF-A; (c) C3(NH_3) was immobilised on the chip, but not probed with vWF-A. The same experiment as in (a) was performed for (d) the Ba fragment, (e) the Bb fragment, and (f) native factor B, and also for (g) bovine serum albumin (BSA) as a negative control. The instrument was calibrated externally using bovine IgG, and the protein masses in Da are shown to the right of the peak. The effect of studying a BSA sample on a normal phase chip surface is shown in (h).

which were increased to 28-30 kDa, 61 kDa and 89-91 kDa, respectively if the masses of one/two, two and three/four biantennary complex-type

oligosaccharide chains, respectively, (each of approximate mass 2100 Da) were added to these. The appearance of broad peaks was attributed to the presence of two N-linked glycosylation sites in the second SCR domain of the Ba fragment, and two more in the vWF-A domain of the Bb fragment, each of which contain a heterologous population of primarily biantennary complex-type oligosaccharides, together with minor amounts of tri- and tetraantennary types and variable sialic acid contents (Anderson, 1986). When bovine serum albumin was used to probe the immobilised $C3(NH_3)$, no peak was observed at the appropriate mass of about 66 kDa (Figure 1(g)). If bovine serum albumin was non-covalently coated onto a normal phase surface, a peak at a mass of 66,257 Da was observed (Figure 1(h)). It was concluded that the observed peaks in Figure 1(d), (e) and (f) corresponded to specific interactions of these proteins with activated C3. In particular, this is direct evidence that both Ba and Bb fragments have binding sites for activated C3, in accordance with the multisite model for the assembly of the C3bBb convertase (Lambris & Müller-Eberhard, 1984; Pryzdial & Isenman, 1988; Ueda *et al.*, 1987; Sanchez-Corral *et al.*, 1990).

Mass spectrometry of proteolysed vWF-A domain after interaction with $C3(NH_3)$

The interaction between the recombinant carbohydrate-free vWF-A domain and immobilised $C3(NH_3)$ was investigated in two experiments using proteolysis with trypsin with and without washes with 50 mM urea (see Materials and Methods). Trypsin was employed for reason of the accurate prediction of cleavage sites at Arg or Lys residues, which meant that the proteolytic fragments could be identified by determination of their masses and comparing these with the factor B sequence. Two sets of mass spectra are shown from these experiments, in which four well-resolved peaks with averaged masses of 3019 Da, 4048 Da, 4174 Da and 6027 Da were observed (Figure 2(a) and (b)). There were additional peaks

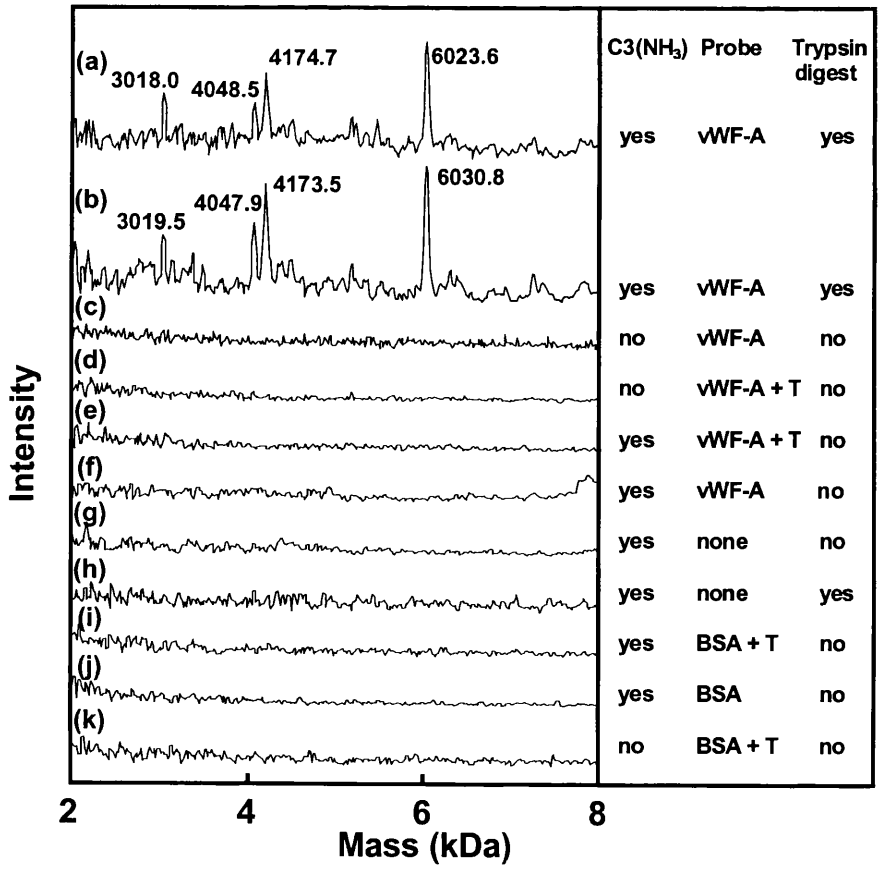


Figure 2. Mass analysis of peptides derived from the trypsin digest of the vWF-A domain bound to immobilised $C3(NH_3)$ on the activated chip. Only the spectrum between 2 and 8 kDa is shown here, in distinction to Figure 1. The four strongest peaks are labelled with their corresponding mass values (Table 1). The instrument was calibrated internally against bovine ubiquitin (mass 8564.8 Da). (a) The complex of vWF-A and $C3(NH_3)$ was digested with trypsin, then washed with PBS, then with water. (b) The experiment in (a) was repeated after washing with PBS that contained 50 mM urea to act as a more stringent washing agent. Other experiments are summarised in the right-hand panel which defined the components that were present, including the control with bovine serum albumin (BSA) and the pretreatment with trypsin (T).

that were also observed at mass values above 8000 Da (not shown), and these were attributed to larger-sized partial digest fragments. It was inferred from these two experiments that four peptides from the vWF-A domain had remained associated with the immobilised C3(NH_3) after proteolysis. As controls, a range of different conditions were tested using the permutations summarised to the right in Figure 2, including the presence or absence of C3(NH_3), the use of either the vWF-A domain or bovine serum albumin, and treatment of the vWF-A domain or bovine serum albumin with trypsin prior to or after adding the probe to the chip surface. The outcome of 64 runs (including duplicates to confirm reproducibility) showed that the four peaks in Figure 2(a) and (b) were absent in the nine experiments outlined in Figure 2(c) to (k). The possibility of non-specific interactions of the vWF-A peptide fragments with immobilised C3(NH_3) or the chip surface was ruled out with the experiments shown Figure 2(d) and (e), respectively. The possibility that the tryptic peptides originated from C3(NH_3) or the self-degradation of trypsin was eliminated using Figure 2(g). The controls confirmed that the four peaks did indeed originate from vWF-A peptides with C3b-binding affinity, and indicated that the four peptides remain bound with sufficient affinity to C3(NH_3) during the final post-trypsin wash stage.

In order to assign the four peptide mass peaks to specific residues in the vWF-A domain, the vWF-A sequence was analysed for tryptic fragments corresponding to these masses using the PAWS program (see Materials and Methods). A single segment of the vWF-A sequence could be unambiguously identified for each mass peak to within 0.2% (Table 1). This is compatible with the accuracy of matrix-assisted laser desorption ionisation time-of-flight spectrometry (Krone *et al.*, 1997). Thus the 6027 Da, 4174 Da, 4048 Da and 3019 Da mass signals corresponded to the vWF-A residues G229-R283 (+GS), G229-K266 (+GS), G229-K265 (+GS) and T355-R381, respectively, in factor B. The N-terminal GS dipeptide results from the remnant of the fusion protein thrombin cleavage site (Smith & Johnson, 1988; Williams *et al.*, 1999). The nearest approach to these mass values by other possible tryptic peptides involved larger mass differences of 150 Da, 22 Da, 20 Da and 24 Da in comparison to the small mass differences of 1 to 7 Da in Table 1, and could be ruled out. It was concluded that the residues G229-K265 (peptide 1) and T355-R381 (peptide 2) of factor B were the minimum sizes of peptides that were involved in the interaction of the vWF-A domain with C3(NH_3), although residues in peptide 1 extending up to R283 may also participate in the interaction with C3(NH_3).

Factor B is polymorphic due to the existence of allotypes and variable carbohydrate compositions (Campbell & Bentley, 1985; Anderson, 1986). No peptide signals were observable in SELDIAMS experiments when the immobilised C3 was probed with the native glycosylated factor B or the Ba or Bb fragments, then subjected to tryptic proteolysis. In native factor B, the two vWF-A peptides identified in the SELDIAMS experiments are associated with the N-linked glycosylation sites at N260 and N353 (Figure 3(b)), and the difficulty of identification of any peptide signals from the Bb fragment is partly attributed to the presence of heterologous carbohydrate glycoforms.

Homology model for the vWF-A domain in factor B

The interpretation of the peptide data for the proteolysed vWF-A domain required a structural analysis of this superfamily, as no experimentally determined structure was available for vWF-A from factor B. A total of 23 sets of coordinates from 14 crystal structures for five different homologous vWF-A domains were analysed (see Materials and Methods), from which a sequence alignment could be derived from the identification of their secondary structures (Figure 3(a)). The six hydrophobic β-strands BA, BB, BC, BD, BE and BF were well conserved in position and length to within one residue in the 23 structures (except for BC in LFA-1 which was shifted by two residues). While the amphipathic surface-exposed α-helices A1 and A4 were well conserved to ±2 or ±3 residues in length, the other α-helices A2, A3, A5, A6 and A7 showed significant variation in terms of length and conformation between the five structures, especially towards the C terminus of the protein fold. Thus A2 and A3 were merged into a single α-helix in the VLA-2, vWF-A1 and vWF-A3 struc-

Table 1. Observed masses of the peptides that remain associated with C3(NH_3) and their assignment to vWF-A peptides

Observed mass (Da)[a]	Calculated mass of peptide (Da)	Mean difference (Da)	Sequence of factor B
6023.6, 6030.8	6024.9	2.3	G229-R283[b]
4174.7, 4173.5	4169.7	4.4	G229-K266[b]
4048.5, 4047.9	4041.5	6.7	G229-K265[b]
3018.0, 3019.5	3017.5	1.3	T355-R381

[a] The two mass values correspond to the experiments shown in Figure 2(a) and (b).
[b] In the recombinant vWF-A domain, there is an additional N-terminal dipeptide Gly-Ser which is derived from the thrombin cleavage site of the fusion protein (Perkins *et al.*, 1994; Williams *et al.*, 1999).

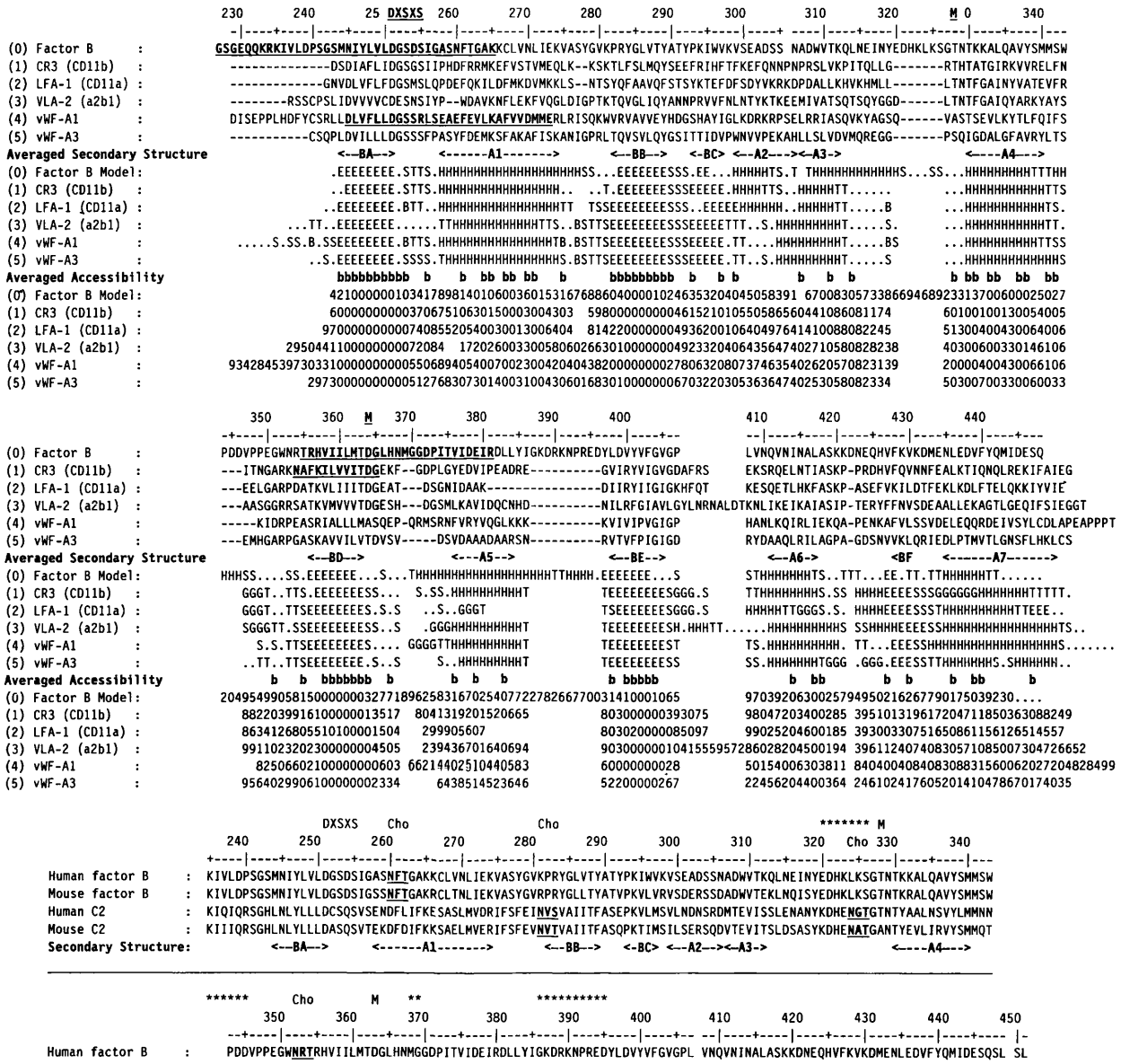


Figure 3. Sequence alignments of vWF-A domains. (a) The consensus alignment of known crystal structures used for the homology model of factor B. The SELDIAMS-identified peptides 1 and 2 in factor B are in bold and underlined. The GpIb receptor site in the von Willebrand factor A1 domain and the iC3b site in CR3 are likewise bold and underlined. The location of the metal binding residues is denoted by DXSXS and M above the sequences. The secondary structure elements identified by DSSP are denoted as: H, α-helix; G, 3_{10}-helix; E, β-strand; B, single residue β-ladders; T, turns; S, bends, and their positions were labelled A1 to A7 and BA to BF to follow Lee *et al.* (1995a) and Edwards & Perkins (1996). The solvent accessibilities calculated using COMPARER are denoted by values of 0 to 9 to correspond to accessibilities of 0-9 %, 10-19 %, and so on up to 90-99 %, and are denoted by b if the mean crystallographic value is 0 or 1. For each of the five independent vWF-A crystal structure sets (see Materials and Methods), the secondary structures and accessibilities were averaged from the individual analyses for between two and eight coordinate sets. (b) The human and mouse factor B and C2 sequences. The four sequence insertions in factor B (see the text) are denoted by asterisks. The location of the carbohydrate sites is denoted by Cho above the sequences. The sequences are shown in comparison with the consensus secondary structure from (a) for ease of reference.

tures, A5 was missing in LFA-1 for reason of a sequence deletion, A6 was sometimes preceded or followed by additional 3_{10}-helices or α-helices, and A7 was variable in its length and position (or was divided into two in the vWF-A3 structure). These large structural variations in α-helices between the five different vWF-A sequences is attributable to their surface location in the vWF-A domain at

which packing constraints are absent. This exposure leads, for example, to the noticeable solvent dependence of the α-helix band in Fourier transform infrared spectra of the vWF-A domain (Perkins *et al.*, 1994).

The comparison of the vWF-A domain of factor B with the five different crystal structures in Figure 3(a) supplements that shown in Figure 1 by Edwards & Perkins (1996), which was based on comparisons with predicted structures. The sequence identity (defined in Materials and Methods) was 23.8% for the comparison of factor B and CR3 using the alignment of Figure 3(a), which was higher than those derived for the other four sequences of Figure 3(a). These were 20.6% for the comparison with LFA-1, 19.8% with VLA-2, 13.7% with vWF-A1 and 15.3% with vWF-A3. The CR3 structure also contained a Mg^{2+} site, unlike the vWF-A1 and vWF-A3 structures, and possessed an iC3b site which may or may not resemble the C3b site of the vWF-A domain of factor B. The CR3 vWF-A crystal structure was thus used as a template for the homology modelling of residues S243-M443 in the vWF-A domain of factor B (see Materials and Methods), although the C-terminal α-helix A7 has exhibited variable conformations in the 8 CR3 coordinate sets available to date. Evidence in support of the vWF-A homology model was provided by the conservation of the six β-strands and the α-helices A1 and A4, the location of the conserved metal-binding residues DXSXS, T and D in their expected positions, and by the high solvent accessibilities of N260 (80%) and N353 (50%), on α-helix A1 and between α-helix A4 and β-strand BD, respectively, which agrees with their putative glycosylation (Figure 3(a)). Likewise, the high accessibilities of 60% for the two different glycosylation sites in the homologous C2 vWF-A structure supported this factor B vWF-A model (Figure 3(a) *versus* (b)). The use of PROCHECK to test the model showed that only E423 and Q424 were minor outliers in a Ramachandran plot (while residue 232 in CR3 which was a minor outlier in its crystal structure was no longer so in the model). That E423 and Q424 were outliers was attributed to their position next to an insertion that was not present in any of the vWF-A crystal structures (Figure 3(a)), which did not facilitate their modelling.

The most notable result from the vWF-A homology model was that four sequence insertions at about E320-S326, P344-D346, M369-G370 and G387-D396 could be identified relative to the positions of conserved sequence/structure features in the five vWF-A crystal structures and were not present in any of these crystal structures (Figure 4). Two of these (M369-G370 and G387-D396) occur at the carboxyl edge of the vWF-A β-sheet near the C3b binding site, while the other two (P344-D346, E320-S326) occur at the amino edge of this β-sheet, close to the linkers that joins the vWF-A domain to the three SCRs in the Ba fragment and the SP domain in the Bb fragment. The four insertions

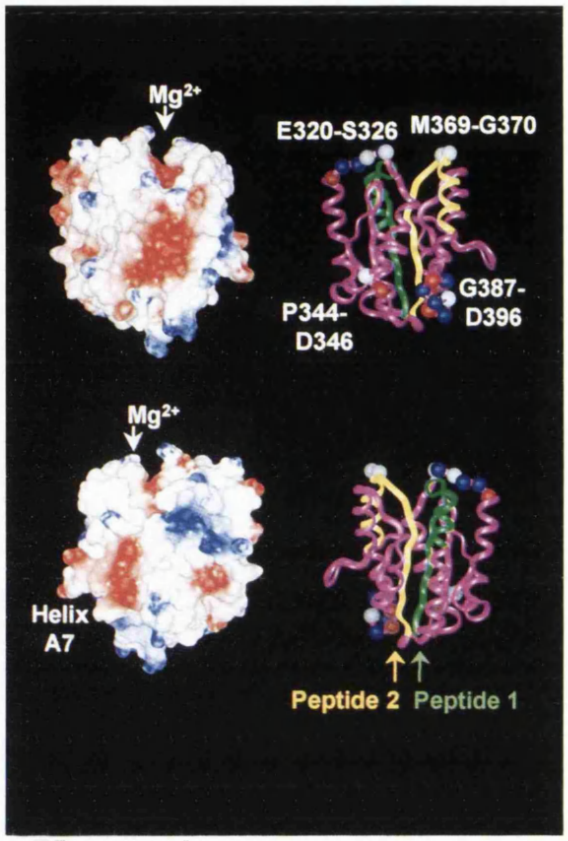


Figure 4. Two views rotated by 180° that depict the surface of the factor B vWF-A homology model. Electrostatic and ribbon representations are shown for each view. Peptides 1 and 2 are shown in green (G229-K265) and yellow (T355-R381), respectively. The Mg^{2+} binding site is in a cleft (arrowed) and lined at its base by acidic residues. The N terminus and C terminus are at the base of the structure. The positions of four insertions (E320-S326 and M369-G370 next to the Mg^{2+} site; P344-D346 and G387-D396 near the N terminus and C terminus) that are not seen in other crystal structures in the vWF-A superfamily are indicated by spheres (blue, basic; red, acidic; white, other). The electrostatic views show several other acidic and basic regions, of which a prominent group of acidic residues corresponds to a surface loop (D422 and E424) and the surface of helix A7 (D432, D434, E437, D438, D445 and E446) to the left in the lower view.

occurred just before and after the α-helices A4 and A5, and have the effect of deepening one side of the active site cleft in the vWF-A domain. It is significant that the four insertions are present also in the homologous vWF-A C2 sequence, although two insertions are different in size (asterisked in Figure 3(b)). In factor B, the three larger insertions possess remarkably high proportions of charged residues compared to the rest of the vWF-A domain, although none of them are involved with the C3b binding site identified by SELDIAMS experiments. In human factor B, the ten-residue insertion G387-D396 possesses three acidic and four basic residues, while the three-residue inser-

tion P344-D346 possesses two acidic residues, and the seven-residue insertion E320-S326 possesses two acidic and three basic residues. Similar totals are present in the insertions of mouse factor B and human and mouse C2 (Figure 3(b)). Electrostatic views of the carboxyl edge face of the model indicates that there is a prominently acidic active site cleft at this side where Mg^{2+} is bound, while the amino edge face shows no prominent acidic or basic regions (Figure 4). There is also a cluster of acidic residues near to α-helix A7 in both factor B and C2. In factor B, the oligosaccharide at N260 is located at the carboxyl edge face of the β-sheet, while that at N353 is located at its amino edge (Figure 5). In C2, the oligosaccharide at N281 is located at the amino edge face of the β-sheet, while that at N324 is located on the opposite face at the carboxyl edge. Both occur in topologically similar locations to those observed in factor B (Figure 5).

Location of peptides 1 and 2 in the vWF-A homology model

Both the vWF-A peptides identified from SELDIAMS occurred at structurally well-defined positions in the alignment and in the vWF-A homology model. Peptide 1 corresponded to a region from the N terminus that included the BA β-strand to half way along the α-helix A1. This region contained the conserved DXSXS motif found in many vWF-A sequences, which corresponds to part of the Mg^{2+} binding site (Perkins *et al.*, 1994). The possible extension of peptide 1 up to R283 (Table 1) includes the remainder of the α-helix A1 up to the start of the β-strand BB. Peptide 2 included the BD β-strand, a two-residue insertion, and the start of the A5 α-helix. It also contained D364 that provided part of the Mg^{2+} binding site (Figure 3(a)).

Topologically, peptides 1 and 2 are seen to be proximate to each other in the factor B vWF-A homology model (yellow and green ribbons, respectively; Figure 4), and include the crossover point in the parallel β-sheet from where the two central β-strands wind to the opposite ends of the β-sheet. This crossover point forms a cleft at the carboxyl edge of doubly wound α/β folds, and contains the Mg^{2+} site in many members of the vWF-A superfamily. Only a limited number of the 39 and 27 residues in peptides 1 and 2, respectively, can be available for protein-protein interactions with C3b. For instance, since the peptide G229-R234 is cleaved from the vWF-A domain on activation by factor D, these residues are ruled out from a C3b binding role. The adjacent residues K235-S243 at the N terminus of the vWF-A domain are unlikely to bind to C3b, since they will form part of the factor D binding site which has to be accessible to factor D after the vWF-A domain is bound to C3b (Williams *et al.*, 1999). The residues M244-S253 in peptide 1 and V358-D364 in peptide 2 correspond to buried side-chains in the β-strands BA and BD, respectively, some of which contribute to the Mg^{2+} site, so cannot be available for direct interactions with C3b. From Figure 3, other predicted buried residues in the vWF-A model can be discounted also (I256, F261, A264 and K265 in peptide 1; V376 and E379 in peptide 2). This leaves eight and 17 residues in peptides 1 and 2, respectively, that have surface-accessible side-chains that can potentially form contacts with C3b (Figure 3). Of these 25 residues, four are acidic, three are basic, 11 are neutral hydrophilic, and seven are hydrophobic, and the majority of these are located

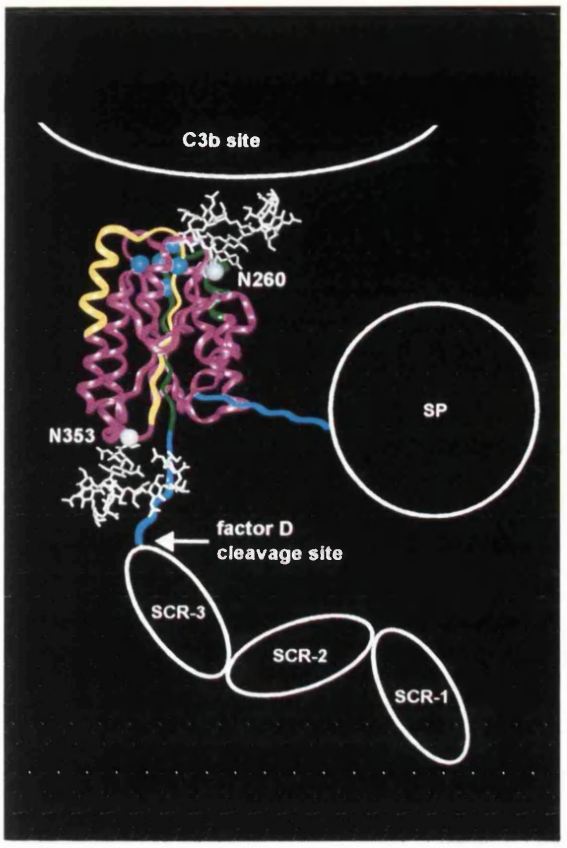


Figure 5. Schematic view of the vWF-A structure in relation to C3b and the other domains of factor B. The SCR and SP domains are of approximately correct sizes but their positions are not known. In the vWF-A domain, the Mg^{2+} cleft is at the top in proximity to the C3b binding site, its N terminus is connected to the SCR domains, and its C terminus is connected to the SP domain. This view is rotated by 90° compared to those shown in Figure 4. The five Mg^{2+} binding residues D251, S253 and S255 in peptide 1 (green ribbon), T328 (purple ribbon) and D364 in peptide 2 (yellow ribbon) are shown as blue spheres. The N-terminal vWF-A residues K235-S243 (Figure 3) are attached to the homology model in an extended conformation (blue ribbon) up to the factor D cleavage site (arrowed). The C-terminal vWF-A residues I444-L452 up to the first Cys residue of the SP domain are attached as an extended conformation (blue ribbon), to which the SP domain is joined. Biantennary complex-type oligosaccharide chains are shown at N260 (near the Mg^{2+} site) and N353 (near the N terminus).

at the carboxyl edge of the β-sheet in the vWF-A domain. It is thus concluded by elimination that the amino acids that correspond to the C3b binding site are located at the active site crossover point of the vWF-A domain.

Conclusions

The combination of experiments using a new development of SELDIAMS technology with a vWF-A homology model constructed from five different crystal structures has provided novel insight into the functional activity of the vWF-A domain in factor B, i.e. that the function of its active site cleft is to bind to C3b. Mass spectrometry analysis of tryptic digests of a functionally active 222-residue recombinant vWF-A domain from factor B (Williams *et al.*, 1999) showed that two peptides G229-K265 and T355-R381 were directly implicated in ligand binding to C3b. This experimental approach is based on a recombinant vWF-A domain that is of high purity, being correctly folded and functionally active. The primary advantage of SELDIAMS is that it corresponds to the biochemical analysis of the whole protein surface. The comparison with the vWF-A homology model showed that both peptides flanked its active site cleft at the carboxyl edge of the central hydrophobic β-sheet. The vWF-A homology model was used to narrow down the specific residues that participate in C3b binding. This showed that potential C3b binding residues in both peptides could be eliminated to leave only about 25 that are primarily located at the exposed surface of this active site cleft. Of interest is that the first of these factor B vWF-A peptides coincides in position with the GpIb platelet receptor binding site at residues 514-542 in the A1 domain of von Willebrand factor (Knott *et al.*, 1992; Sinha *et al.*, 1994; Matsushita & Sadler, 1995), numbered as residues 245-273 in Figure 3(a). The second of these peptides coincides in position with the iC3b active site peptide in the vWF-A domain of CR3 identified by Ueda *et al.* (1994), numbered as residues 354-365 in Figure 3(a).

Additional information on the C3b binding site of the vWF-A domain was obtained from its homology model. Residues in the metal binding site of the integrin vWF-A domains CR3, LFA-1 and VLA-2 are conserved in factor B. This suggests that this Mg^{2+} binding site is conserved in the vWF-A domain of factor B, however this observation is complicated by the observation that factor B will bind to C3b in the absence of Mg^{2+} (Pryzdial & Isenman, 1987; Williams & Sim, 1994), the absence of one of these metal binding residues in mouse C2 (Figure 3(b)), and the non-conservation of the metal binding site in all three vWF-A domains of von Willebrand factor (Jenkins *et al.*, 1998). Both hydrophobic and ionic interactions have been implicated in C3b-factor B complex formation, where low ionic strengths favour complex formation (Pryzdial & Isenman, 1988; DiScipio, 1981). This is consistent with the 25 potential C3b-contact residues which include both hydrophobic and hydrophilic residues, as well as seven charged residues. The presence of other charged residues at three of four significant insertions identified in the vWF-A homology model and on α-helix A7 may also be relevant. Of interest is that E736 and E737 in C3 have been implicated in the interaction of C3b with factor B (Taniguichi-Sidle & Isenman, 1994). It can be speculated that these charged groups may complete the Mg^{2+} coordination by analogy with the observed intermolecular lattice contacts in the crystal structure of CR3 where a Glu residue behaves as a ligand mimic (Lee *et al.*, 1995a).

This application of SELDIAMS is complementary to site-specific mutagenesis experiments in the sense that an increase or decrease in activity following a single-residue mutagenesis change does not unambiguously identify the role of that specific residue. It is necessary to show that the protein folding is unaffected by the mutation, so a systematic study would require mutagenesis experiments throughout the protein. Consequently, SELDIAMS offers much economy of effort through its ability to analyse the whole protein surface, and is potentially applicable to other protein-protein complexes (Spencer *et al.*, 1999). The results of the present SELDIAMS analysis agree with a mutagenesis study of single-residue changes in factor B (Hourcade *et al.*, 1999) that demonstrated that D254, N260 and D364 were involved for C3b binding (Figure 5). The SELDIAMS experiments also agreed with the outcome of experiments with chimeras 1 and 3 of intact factor B by Tuckwell *et al.* (1997). Chimera 1 corresponded to part of peptide 1, in which G252-S259 of factor B was mutated to C252-N259 of human C2, and chimera 3 corresponded to part of peptide 2, in which L366-D372 of factor B was mutated to K366-S372 of human C2, both of which resulted in loss of C3b binding activity. As chimera 1 contained a free C252 residue, which could become disulphide-linked with the free C267 residue of factor B, that study had not been conclusive in showing that the formation of a potential Cys-Cys link might have caused the lack of interaction between chimera 1 and C3b.

The role of factor B in binding to C3b and the resulting activation of the alternative pathway of complement activation has been clarified in the present studies. Early crystallographic studies on the vWF-A domain in CR3 suggested that it could exist in two conformations depending on the presence of the bound cation (Lee *et al.*, 1995a,b), although the biological significance of this finding is unclear as the crystal lattice packing contacts may be responsible for some of these observations (Baldwin *et al.*, 1998; Liddington & Bankston, 1998). Nonetheless, there is evidence from mutational studies indicating "on" and "off" conformations, as well as the separate locations of the mutation sites associated with the types 2B and

2 M phenotypes in von Willebrand's disease, that the vWF-A1 domain in von Willebrand factor is able to transmit ligand binding signals within its structure (Cooney & Ginsberg, 1996; Jenkins *et al.*, 1998). The vWF-A homology model showed that the factor D cleavage site is eight residues away from its N terminus (Figure 5). The recombinant vWF-A domain also included an extra eight residues that are N terminal of the factor D cleavage site. It was possible for factor D to cleave this recombinant vWF-A domain both in the presence or absence of C3b, although this cleavage is slow in both cases (Williams *et al.*, 1999). In contrast to this, the cleavage of intact factor B by factor D in the presence of C3b occurs efficiently, but does not occur in the absence of C3b. This means that this factor D cleavage site is either sterically concealed, and/or it possesses an inactive conformation in intact factor B until it comes into contact with C3b.

If the factor B vWF-A domain is able to transmit conformational signals by analogy with the von Willebrand factor A1 domain (Cooney & Ginsberg, 1996; Jenkins *et al.*, 1998), one possible explanation of the factor D cleavage properties of intact factor B is that the binding of the active site of vWF-A to C3b would induce a conformational change at its opposite face to upregulate factor D cleavage (Figure 5). Another possible explanation would be that the independent binding and reorientation of the SCR domains of intact factor B to another region of C3b would cause a conformational change in factor B to make this peptide available for cleavage by factor D (Figure 5). Irrespective of which explanation is followed, these experiments show that the role of the vWF-A domain in factor B is to facilitate its cleavage by factor D after binding to C3b, and can mediate the regulation of the activities of factor B. The release of the three SCR domains from factor B or an allosteric structural change in the vWF-A domain then brings about the activation of the SP domain. It is interesting that the SP domain of factor B is attached to the α-helix A7 of the vWF-A domain (Figure 5), which is in topological proximity at the same α-helix face to the predicted inhibitory binding site at α-helices A2 and A3 of the A1 domain in von Willebrand factor (Perkins *et al.*, 1999). This may suggest a common functional property of both domains.

Materials and Methods

Preparation of recombinant vWF-A, factor B, the Bb and Ba fragments, and C3

A 222-residue recombinant vWF-A domain of factor B (residues G229-Q448) from *Escherichia coli* was prepared by thrombin cleavage of a vWF-A glutathione-*S*-transferase fusion protein as described by Williams *et al.* (1999). The vWF-A domain incorporated a C267S mutation to prevent its dimerisation. To ensure the removal of thrombin proteolytic activity, the vWF-A domain was passed through a 50 ml arginine/agarose column (Affinity Chromatography Ltd, Ballasalla, Isle of Man, UK) prior to the final gel filtration stage of its purification. To prepare native factor B, fresh frozen plasma was obtained from the Royal Free Hospital Blood Bank and stored at −20 °C. Purification was achieved using Cibrachrom Blue dye-ligand affinity chromatography of human serum (Williams & Sim, 1993), followed by ion-exchange chromatography on a monoQ anion exchange column (dimensions 5 mm × 50 mm) (Pharmacia, Milton Keynes, Bucks, UK). The native Ba and Bb fragments were prepared by the proteolytic cleavage of factor B (Lambris & Müller-Eberhard, 1984). C3 was purified from human plasma (Dodds, 1993). C3(NH_3) was prepared by adding ammonium bicarbonate to a final concentration of 100 mM at pH 8.0 to purified C3, then incubating this at 37 °C for one hour, then dialysing overnight into phosphate buffered saline (PBS) (137 mM NaCl, 2.7 mM KCl, 8.1 mM Na_2HPO_4, 1.5 mM KH_2PO_4, 0.5 mM $MgCl_2$, pH 7.3).

All protein preparations were concentrated under nitrogen gas using an Amicon ultrafiltration stirred cell with a YM-10 membrane before being loaded onto a HiLoad 16/60 Superdex-75 gel filtration column (Pharmacia). The protein preparations were characterised by SDS-PAGE using 10% gels (except for 8% gels used for factor B) in an ATTO mini gel apparatus (Genetic Research Instruments, Dunmow, Essex, UK) to determine molecular masses and purities. Samples were denatured in 8 M urea in the presence of either 20 mM dithiothreitol (reducing gels) or 40 mM iodoacetamide (non-reducing gels) prior to loading onto the gel. Rainbow protein molecular mass markers were used as reference (Amersham). Protein concentrations were determined spectrometrically at 280 nm, using absorption coefficients (1 cm, 1%) calculated from the sequence as follows (Perkins, 1986): vWF-A domain, 15.2; factor B, 13.8; Ba fragment, 16.7; Bb fragment, 12.5; C3, 9.8.

SELDIAMS analyses

For the SELDIAMS experiments, the covalent binding of C3(NH_3) to preactivated chips (Ciphergen Biosystems Ltd., Camberley, Surrey, UK) was achieved *via* a primary amine-based coupling to the N-terminal amino acid and lysine residues. To prevent cross-contamination over the SELDI ProteinChip™ surface, a circle was drawn around each spot on the chip using a hydrophobic marker (Dako, Bucks, UK). The chip surface was prepared by adding 1 μl of acetonitrile and allowing it to dry. A 2 μl sample of C3(NH_3) at 0.1 mg/ml in PBS was then added to the spots on the chip and kept in a humid environment overnight to allow covalent attachment of the protein. Unbound C3(NH_3) was rinsed off the activated chip surface with 4 μl of PBS, and any remaining sites were blocked using 4 μl of 1 M ethanolamine (pH 7.5) for 30 minutes. After the spots had been rinsed three times with 4 μl of PBS, they were then probed with 2 μl of protein ligand (either the vWF-A domain, the Ba or Bb fragment, or factor B) at 0.1 mg/ml in PBS. This was left at room temperature in a humid environment for one hour, after which the spots were rinsed with PBS. The protease incubations were conducted using 2 μl of sequencing-grade trypsin (Boehringer Mannheim, East Sussex, UK) at 5 μg/ml in PBS at room temperature for one hour, after which the spots were rinsed three times with PBS, then three times with water.

Mass analysis was performed with a PBS-1 mass analyser (Ciphergen Biosystems Ltd.). A 1 μl sample of matrix solution (5 mg/ml) was added to the prepared chip; this was allowed to dry. The spots were analysed

by matrix assisted laser desorption ionisation time of flight mass spectrometry. The protein mass analyses were performed using sinapinic acid (*trans*-3,5-dimethoxy-4-hydroxycinnamic acid) as the matrix in 50% acetonitrile and 0.05% trifluoroacetic acid in distilled water. The peptide mass analyses employed a matrix solution consisting of α-cyano-4-hydroxycinnamic acid in 50% (v/v) acetonitrile, 50% (v/v) deionized water and 0.01% (v/v) trifluoroacetic acid. To determine masses, the instrument was calibrated externally against bovine IgG (mass 147.3 kDa) and internally against bovine ubiquitin (mass 8546.4 Da). The masses from amino acid sequences were calculated using averaged isotopic atomic masses with the Protein Analysis Worksheet (PAWS) program (http://www.proteometrics.com).

Homology modelling of the vWF-A domain in factor B

In order to model the vWF-A domain structure of factor B (residues S243 to M443), a total of 23 sets of coordinates from the Protein Data Bank for 14 vWF-A crystal structures for CR3, LFA-1, VLA-2 and the A1 and A3 domains of von Willebrand factor were analysed in order to assess the highest similarity with the vWF-A sequence of factor B. The identity of each alignment of two sequences was defined as (the number of identical residues) × 100/(the total number of topologically equivalent positions in the alignment, excluding gaps in either sequence). Secondary structure elements were identified using DSSP (Kabsch & Sander, 1983). Residue solvent accessibilities were calculated using a probe of 1.4 Å in the COMPARER program (Lee & Richards, 1971; Šali & Blundell, 1990). For alignment purposes, the DSSP and COMPARER analyses were supplemented by visual inspection using INSIGHT II software (Biosym/MSI, San Diego, USA) on Silicon Graphics INDY Workstations in conjunction with Crystal Eyes stereo glasses and the superfamily alignment of 75 vWF-A sequences reported by Perkins *et al.* (1994). Coordinates for eight CR3 structures were taken from the PDB entries 1ido, 1bho, 1bhq, 1jlm and 1idn. Those for 7 LFA-1 structures were taken from the PDB entries 1zoo, 1zop, 1lfa and 1zon. Those for two VLA-2 structures were taken from the PDB entry 1aox. Those for two vWF-A1 structures were taken from the PDB entries 1auq and 1oak. Those for four vWF-A3 structures were taken from the entries 1ao3 and 1atz.

The homology model was built by aligning the factor B vWF-A sequence with that of the CR3 vWF-A domain (PDB code 1ido; Lee *et al.*, 1995a). This alignment was initially extracted from Perkins *et al.* (1994), and was modified by eye to optimize residue and topological equivalences and minimize the number of gaps in the alignment. Regions without gaps were defined as structurally conserved regions that formed the secondary structure framework. The insertions and deletions were mainly found in loop positions and were defined as structurally variable regions. This conserved framework that was used for the core of the vWF-A model contained nine peptide fragments from the CR3 structure and comprised 133 residues. Eight peptide fragments of the correct length from known protein structures in the Protein Data Bank were selected for modelling the structurally variable regions in the vWF-A model (factor B residues V275-P282, S305-N306, N315-S326, S342-T355, G365-I374, L384-Y397, V405-L408 and K421-E424). A precalculated C^{α} atom distance matrix identified the PDB loops that best fitted the corresponding C^{α} atom distance matrix calculated from the structurally conserved regions of the vWF-A framework that defined the starts and ends of these searched loops for a specified number of flanking and intervening residues. These remodelled 68 loop residues comprise 34% of the vWF-A model. Typically for homology models, these loop residues will be less well-defined in position than the 133 framework residues. Side-chain atoms were automatically generated for both the structurally conserved regions and structurally variable regions using the template structures and general rules for residue exchanges. The vWF-A homology model was refined using energy minimization, where 300 steps of steepest descent minimisation was performed with HOMOLOGY and DISCOVER software (MSI/Biosym, San Diego, USA). The energy refinements significantly improved the covalent geometry at the splice junctions with the searched loops and the bad contacts of the vWF-A model as assessed by the PROCHECK program (Laskowski *et al.*, 1993).

Acknowledgements

We thank the Wellcome Trust for project grant support, Dr R. B. Sim (MRC Immunochemistry Unit, Oxford) for useful discussions, Professor J. E. Volanakis (Department of Medicine, University of Alabama at Birmingham, USA) for the provision of factor D, and Professor R. H. J. Begent (Department of Oncology, Royal Free Hospital) for access to the PBS-1 mass analyser.

References

Anderson, C. M. (1986). Glycoprotein structure of components C2 and factor B of the human complement system. D Phil thesis, University of Oxford.

Baldwin, E. T., Sarver, R. W., Bryant, G. L., Jr, Curry, K. A., Fairbanks, M. B., Finzel, B. C., Garlick, R. L., Heinrikson, R. L., Horton, N. C., Kelley, L.-L. C., Mildner, A. M., Moon, J. B., Mott, J. E., Mutchler, V. T. & Tomich, C.-S. C., *et al.* (1998). Cation binding to the integrin CD11b I domain and activation model assessment. *Structure*, **6**, 923-935.

Beavis, R. C. & Chait, B. T. (1996). Matrix-assisted laser desorption ionization mass-spectrometry of proteins. *Methods Enzymol.* **270**, 519-551.

Bienkowska, J., Cruz, M., Atiemo, A. & Liddington, R. (1997). The von Willebrand A3 domain does not contain a metal ion-dependent adhesion site motif. *J. Biol. Chem.* **272**, 25162-25167.

Campbell, R. D. & Bentley, D. R. (1985). The structure and genetics of the C2 and factor B genes. *Immunol. Rev.* **87**, 19-37.

Celikel, R., Varughese, K. I., Madhusan, Yoshioka, A., Ware, J. & Ruggeri, Z. M. (1998). Crystal structure of the von Willebrand factor A1 domain in complex with the function-blocking NMC-4 Fab. *Nature Struct. Biol.* **5**, 189-194.

Cooney, K. A. & Ginsburg, D. (1996). Comparative analysis of type 2b von Willebrand disease mutations: implications for the mechanism of von Willebrand factor binding to platelets. *Blood*, **87**, 2322-2328.

DiScipio, R. G. (1981). The binding of human complement proteins C5, factor B, β1H and properdin to complement fragment C3b on zymosan. *Biochem. J.* **199**, 485-496.

Dodds, A. W. (1993). Small-scale preparation of complement components C3 and C4. *Methods Enzymol,* **223**, 46-61.

Edwards, Y. J. K. & Perkins, S. J. (1995). The protein fold of the von Willebrand Factor type A domain is predicted to be similar to the open twisted β-sheet flanked by α-helices found in human ras-p21. *FEBS Letters,* **358**, 283-286.

Edwards, Y. J. K. & Perkins, S. J. (1996). Assessment of protein fold predictions from sequence information: the predicted α/β doubly wound fold of the von Willebrand Factor Type A domain is similar to its crystal structure. *J. Mol. Biol.* **260**, 277-285.

Emsley, J., King, S. L., Bergelson, J. M. & Liddington, R. C. (1997). Crystal structure of the I domain from integrin α2β1. *J. Biol. Chem.* **272**, 28512-28517.

Emsley, J., Cruz, M., Handin, R. & Liddington, R. (1998). Crystal structure of the von Willebrand factor A1 domain and implications for the binding of platelet glycoprotein Ib. *J. Biol. Chem.* **273**, 10396-10401.

Hinshelwood, J., Williams, S. C., Sim, R. B. & Perkins, S. J. (1996). Structural and functional studies of the isolated vWF-A and SP domains of complement factor B. *Mol. Immunol.* **33**, 80.

Hourcade, D. E., Wagner, L. M. & Oglesby, T. J. (1995). Analysis of the short consensus repeats of human complement factor B by site-directed mutagenesis. *J. Biol. Chem.* **270**, 19716-19722.

Hourcade, D. E., Mitchell, L. M. & Oglesby, T. J. (1999). Mutations of the type A domain of complement factor B that promote high-affinity C3b-binding. *J. Immunol.* **162**, 2906-2911.

Huizinga, E. G., van der Plas, R. M., Kroon, J., Sixma, J. J. & Gros, P. (1997). Crystal structure of the A3 domain of human von Willebrand factor: implications for collagen binding. *Structure,* **5**, 1147-1156.

Hutchens, T. W. & Yip, T.-T. (1993). New desorption strategies for the mass spectrometric analysis of macromolecules. *Rapid Commun. Mass Spectrom.* **7**, 576-580.

Jenkins, P. V., Pasi, K. J. & Perkins, S. J. (1998). Molecular modelling of ligand and mutation sites of the Type A domains of human von Willebrand factor and their relevance to von Willebrand's disease. *Blood,* **91**, 2032-2044.

Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers,* **22**, 2577-2637.

Knott, H. M., Berndt, M. C., Kralicek, A. V., O'Donoghue, S. I. & King, G. F. (1992). Determination of the solution structure of a platelet-adhesion peptide of von Willebrand factor. *Biochemistry,* **31**, 11152-11158.

Krone, J. R., Nelson, R. W., Dogruel, D., Williams, P. & Granzow, R. (1997). BIA/MS: Interfacing biomolecular interaction analysis with mass spectrometry. *Anal. Biochem.* **244**, 124-132.

Kuwata, H., Yip, T.-T., Yip, C. L., Tomita, M. & Hutchens, T. W. (1998). Bactericidal domain of lactoferrin: detection, quantitation and characterization of lactoferricin in serum by SELDI affinity mass spectrometry. *Biochem. Biophys. Res. Commun.* **245**, 764-773.

Lambris, J. D. & Müller-Eberhard, H. J. (1984). Isolation and characterization of a 33,000-dalton fragment of complement factor B with catalytic and C3b binding activity. *J. Biol. Chem.* **259**, 12685-12690.

Laskowski, R. A., Macarthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK - a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283-291.

Law, S. K. & Reid, K. B. M. (1995). *Complement,* 2nd edit., IRL Press, Oxford.

Lee, B. & Richards, F. M. (1971). An interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**, 379-400.

Lee, J. O., Rieu, P., Arnaout, M. A. & Liddington, R. (1995a). Crystal structure of the A domain from the α subunit of integrin CR3 (CD11b/CD18). *Cell,* **80**, 631-638.

Lee, J. O., Bankston, L. A., Arnaout, M. A. & Liddington, R. C. (1995b). Two conformations of the integrin A-domain (I-domain): a pathway for activation? *Structure,* **3**, 1333-1340.

Leung, S. M., Zeyda, T. & Thatcher, B. (1998). A new and rapid method for phenotype screening using SELDI Proteinchip (TM) arrays demonstrated on serum from knockout and wild type mice. *Mol. Biol. Cell,* **9**, 2040.

Liddington, R. & Bankston, L. (1998). The integrin I domain: crystals, metals and related artefacts. *Structure,* **6**, 937-938.

Matsushita, T. & Sadler, J. E. (1995). Identification of amino acid residues essential for von Willebrand factor binding to platelet glycoprotein Ib. *J. Biol. Chem.* **270**, 13406-13414.

Perkins, S. J. (1986). Protein volumes and hydration effects: the calculation of partial specific volumes, neutron scattering matchpoints and 280 nm absorption coefficients for proteins and glycoproteins from amino acid sequences. *Eur. J. Biochem.* **157**, 169-180.

Perkins, S. J., Smith, K. F., Williams, S. C., Haris, P. I., Chapman, D. & Sim, R. B. (1994). The secondary structure of the von Willebrand domain in factor B of human complement by Fourier transform infrared spectroscopy: its occurrence in collagen types VI, VII, XII and XIV, the integrins and other proteins by averaged secondary structure predictions. *J. Mol. Biol.* **238**, 104-119.

Perkins, S. J., Hinshelwood, J., Edwards, Y. J. K. & Jenkins, P. V. (1999). Structural and functional modelling of vWF-A domains in complement and coagulation. *Biochem. Soc. Trans.* In the press.

Pryzdial, E. L. & Isenman, D. E. (1987). Alternative complement pathway activation fragment Ba binds to C3b. Evidence that formation of the factor B-C3b complex involves two discrete points of contact. *J. Biol. Chem.* **262**, 1519-1525.

Pryzdial, E. L. G. & Isenman, D. E. (1988). A thermodynamic study of the interaction between human complement component C3b or component C3(H_2O) and factor B in solution. *J. Biol. Chem.* **263**, 1733-1738.

Qu, A. & Leahy, D. J. (1995). Crystal structure of the I-domain from the CD11a/CD18 (LFA-1, αLβ2) integrin. *Proc. Natl Acad. Sci. USA,* **92**, 10277-10281.

Qu, A. & Leahy, D. J. (1996). The role of the divalent cation in the structure of the I domain from the CD11a/CD18 integrin. *Structure,* **4**, 931-942.

Šali, A. & Blundell, T. L. (1990). Definition of general topological equivalence in protein structures: a procedure involving comparison of properties and relationships through simulated annealing and dynamic programming. *J. Mol. Biol.* **212**, 403-428.

Sanchez-Corral, P., Anton, L. C., Alcolea, J. M., Marques, G., Sanchez, A. & Vivanco, F. (1990). Pro-

teolytic activity of the different fragments of factor B on the third component of complement C3 - involvement of the N-terminal domain of Bb in magnesium binding. *Mol. Immunol.* **27**, 891-900.

Sim, E. & Sim, R. B. (1983). Enzymic assay of C3b receptor on intact cells and solubilized cells. *Biochem. J.* **210**, 567-576.

Sinha, D., Bakhshi, M., Kunapuli, S., Vora, R., Gabriel, J. L., Kirby, E. P. & Budzynski, A. Z. (1994). Asp514 within the A1 domain of bovine von Willebrand factor is required for interaction with platelet glycoprotein IB. *Biochem. Biophys. Res. Commun.* **203**, 881-888.

Smith, D. B. & Johnson, K. S. (1988). Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione-*S*-transferase. *Gene*, **67**, 31-40.

Spencer, D. I. R., Robson, L., Bhatia, J., Sharma, S. K., Michael, N. P., Whitelegg, N. R., Rees, A. R., Chester, K. A. & Begent, R. H. J. (1999). Identifying immunogenic sites on ADEPT enzyme CPG2 using a scFv library and SELDI-AMS. British Cancer Research Meeting, Edinburgh, 11th-14th July. *Brit. J. Cancer*, **80**, **(Suppl. 2)**, 52.

Taniguchi-Sidle, A. & Isenman, D. E. (1994). Interactions of human complement component C3 with factor B and with complement receptors type 1 (CR1, CD35) and type 3 (CR3, CD11b/CD18) involve an acidic sequence at the N-terminus of C3 α'-chain. *J. Immunol.* **153**, 5285-5302.

Tuckwell, D. S., Xu, Y. Y., Newham, P., Humphries, M. J. & Volanakis, J. E. (1997). Surface loops adjacent to the cation-binding site of the complement factor B von Willebrand Factor type A module determine C3b binding specificity. *Biochemistry*, **36**, 6605-6613.

Ueda, A., Kearney, J. F., Roux, K. H. & Volanakis, J. E. (1987). Probing functional sites on complement protein B with monoclonal antibodies. Evidence for C3b-binding sites on Ba. *J. Immunol.* **138**, 1143-1149.

Ueda, T., Rieu, P., Brayer, J. & Arnaout, M. A. (1994). Identification of the complement iC3b binding site in the β2 integrin CR3 (CD11b/CD18). *Proc. Natl Acad. Sci. USA*, **91**, 10680-10684.

von Zabern, I., Nolte, R. & Vogt, W. (1981). Treatment of human complement components C4 and C3 with amines or chaotropic ions. Evidence of a functional and structural change that provides uncleaved C4 and C3 with properties of their soluble activated forms C4b and C3b. *Scand. J. Immunol.*, **13**, 413-431.

Williams, S. C. & Sim, R. B. (1993). Dye-ligand affinity purification of human complement factor B and β-2 glycoprotein I. *J. Immunol. Methods*, **157**, 25-30.

Williams, S. C. & Sim, R. B. (1994). Binding sites involved in the formation of the C3 (H2O)-factor B complex of the alternative pathway of complement. *Biochem. Soc. Trans.* **22**, 2S.

Williams, S. C., Hinshelwood, J., Perkins, S. J. & Sim, R. B. (1999). Production and functional activity of a recombinant von Willebrand Factor (vWF-A) domain from human complement factor B. *Biochem. J.* **342**, 625-632.

Edited by R. Huber

(Received 9 August 1999; received in revised form 15 September 1999; accepted 15 September 1999)

doi:10.1006/jmbi.2000.3632 available online at http://www.idealibrary.com on IDEAL® *J. Mol. Biol.* (2000) **298**, 135–147

JMB

Metal-dependent Conformational Changes in a Recombinant vWF-A Domain from Human Factor B: A Solution Study by Circular Dichroism, Fourier Transform Infrared and ^{1}H NMR Spectroscopy

Justin Hinshelwood and Stephen J. Perkins*

Department of Biochemistry and Molecular Biology, Royal Free and University College Medical School, University College London, Rowland Hill Street, London, NW3 2PF, UK

Corresponding author

Factor B is a key component of the alternative pathway of complement and is cleaved by factor D into the Ba and Bb fragments when complexed with the activated form of C3, namely C3b. The Bb fragment contains a von Willebrand factor type A (vWF-A) domain, which is composed of an open twisted almost-parallel β-sheet flanked on both sides by seven α-helices A1 to A7, with a metal coordination site at its active-site cleft. Homology modelling of this vWF-A domain shows that the metal-binding site was present. Two recombinant vWF-A domains (Gly229-Ile444 and Gly229-Gln448) were examined by circular dichroism and Fourier transform infrared spectroscopy and indicated a significant conformational transition in the presence and absence of Mg^{2+}. Two upfield-shifted signals in the ^{1}H NMR spectrum were used as sensitive probes of the vWF-A protein structure, one of which was assigned to a methyl group and demonstrated metal- and pH-dependent properties between two distinct conformations. Temperature denaturation studies followed by spectroscopy showed that metal-binding caused the vWF-A structure to become significantly more stable. Ring current calculations based on a homology model for the vWF-A structure correlated one upfield-shifted signal with a methyl group on the α-helices in the vWF-A structure and the other one with individual single protons. An allosteric property of the vWF-A domain has thus been identified, and its implications for factor B activation were examined. Since the vWF-A domain after α-helix A7 is connected by a short link to the catalytic serine protease domain in the Bb fragment, the identification of a metal-free and a more stable metal-bound conformation for the vWF-A domain implies that the vWF-A interaction with C3b may alter its Mg^{2+}-bound coordination in such a way as to induce conformational changes that may regulate the proteolytic activity of factor B.

Keywords: vWF-A domain; complement factor B; ^{1}H NMR; circular dichroism; Fourier transform infrared

Introduction

Factor B is the zymogen of the protease component of the C3 and C5 convertases in the alternative pathway of the complement system (Law & Reid, 1995). The activated form of C3, C3b, forms a complex with factor B both in the presence and absence of Mg^{2+} (Pryzdial & Isenman, 1987; Williams & Sim, 1994), whereupon factor B is cleaved by factor D to give the Ba and Bb fragments. The dissociation of the Ba fragment from the complex leaves the Bb fragment attached to

Abbreviations used: CD, circular dichroism; FB, factor B; Ba and Bb, Ba and Bb fragments of factor B; FT-IR, Fourier transform infrared; PBS, phosphate buffered saline; vWF-A, von Willebrand factor type A domain; SCR, short consensus/complement repeat; SP, serine protease.

E-mail address of the corresponding author: steve@rfhsm.ac.uk

0022-2836/00/010135–13 $35.00/0

C3b to form the active C3 convertase enzyme C3bBb. Factor B is essential for the opsonisation and cell killing functions of the alternative pathway.

Human factor B is a five domain glycoprotein (Mole *et al.*, 1984). The N-terminal Ba fragment contains three short consensus/complement repeat (SCR) domains (also known as complement control protein domains), each of size about 60 amino acid residues. The C-terminal Bb fragment contains two domains, the N-terminal one being a von Willebrand factor type A (vWF-A) domain (also known as I-domains or A-domains), and the C-terminal one being a serine protease (SP) domain. All three domains are implicated in multi-site interactions with C3b. Of particular interest is the vWF-A domain of factor B, which is responsible for binding to C3b at its active site cleft (Tuckwell *et al.*, 1997; Hourcade *et al.*, 1999; Hinshelwood *et al.*, 1999). The presence of this domain N-terminal to that of the SP domain is distinct to that of the multidomain serine proteases of blood coagulation. The SP domain of factor B does not follow the classic activation mechanism in which a polypeptide cleavage immediately prior to its N terminus creates a new positive charge, which is then inserted into the SP domain and causes it to adopt an active SP conformation (Huber & Bode, 1978; Perkins & Wuthrich, 1980). In factor B, it is most likely that the vWF-A domain is involved in the activation of the SP domain, however the molecular mechanism for this is presently unknown.

The vWF-A domains are widespread and occur in cell surface receptors (integrins) and in collagen Types VI, VII, XII and XIV as well as in other complement proteins (Perkins *et al.*, 1999). Its structure is an open twisted mostly-parallel β-sheet structure flanked by α-helices above and below the β-sheet (Figure 1), in which an active site cleft at the crossover point where the β-sheet winds towards opposite ends of the structure frequently binds metal (Edwards & Perkins, 1996). Crystallography has yielded 23 sets of vWF-A coordinates for 14 vWF-A crystal structures for five different

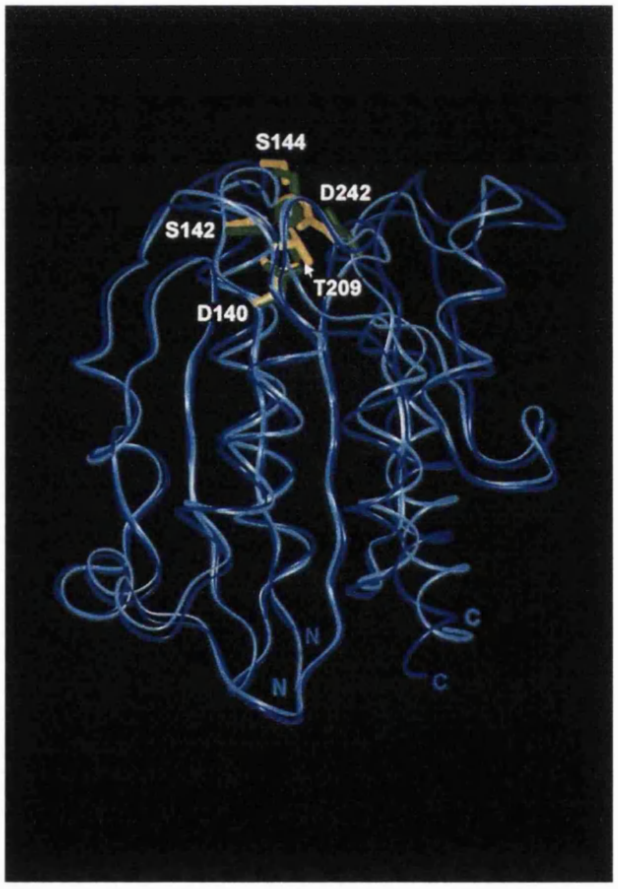


Figure 1. Conformational change seen between two superimposed crystal structures for complement receptor type 3 (CR3) in the presence of Mg^{2+}. The central six-stranded β-sheet is viewed face-on. The structure corresponding to the PDB code 1ido is shown as a light blue ribbon with the active-site residues depicted with side-chains (green) and metal atom (Table 1). That corresponding to the PDB code 1bho (dark blue) with side-chains (yellow) and metal atom. A large movement in Asp242 and the α-helices relative to an almost invariant β-sheet is seen between the two forms.

vWF-A proteins. Metal-binding sites were observed in the integrin vWF-A structures of complement receptor type 3 (CR3; CD11b), leucocyte-function associated antigen-1 (LFA-1; CD11a) and very late activation protein-2 (VLA-2; platelet glycoprotein Ia; CD49b), but not in the A1 and A3 domains of von Willebrand factor (Lee *et al.*, 1995a,b; Baldwin *et al.*, 1998; Qu & Leahy, 1995, 1996; Emsley *et al.*, 1997, 1998; Celikel *et al.*, 1998; Huizinga *et al.*, 1997; Bienkowska *et al.*, 1997). The effect of metal binding to the vWF-A domain of integrins is presently unclear, since the Mg^{2+} binding site in the vWF-A crystal structure for CR3 differs from those in others, namely LFA-1, VLA-2 and others for CR3 (Figure 1). It is not known whether the integrin vWF-A structure is affected by metal binding, and this is currently an area of controversy (Lee *et al.*, 1995a; Baldwin *et al.*, 1998; Liddington & Bankston, 1998). Further uncertainty results from the absence of a Mg^{2+} site in the vWF-A1 and vWF-A3 domains of von Willebrand factor (Celikel *et al.*, 1998; Emsley *et al.*, 1998; Jenkins *et al.*, 1998; Huizinga *et al.*, 1997; Bienkowska *et al.*, 1997). There is also uncertainty in the role of the Mg^{2+} in factor B as the metal is not absolutely required for assembly of the C3bBb complex nor for its activation (Pryzdial & Isenman, 1986).

A clarification of the structural role of bound Mg^{2+} to the vWF-A domain is essential for an understanding of its metal-binding properties. Given the large difference reported between the metal-bound and metal-free forms of vWF-A by crystallography (Figure 1), it was surprising that initial circular dichroism (CD) and Fourier transform infrared (FT-IR) spectroscopic studies of vWF-A domains were unable to detect any significant conformational rearrangement upon metal binding (Perkins *et al.*, 1994; Fairbanks *et al.*, 1995; Nolte *et al.*, 1999). Here, we describe the application of a multidisciplinary spectroscopic approach to a recombinant vWF-A domain from factor B in the presence of excess metal or EDTA. Two recombinant vWF-A domains were used, one with 218 residues (molecular weight 24,500) based on the factor B vWF-A sequence Gly229-Ile444, and a more stable one with 222 residues (molecular weight 25,000) based on the sequence Gly229-Gln448 (Williams *et al.*, 1999). A homology model for the vWF-A domain enabled the data to be interpreted (Hinshelwood *et al.*, 1999). We present the first unequivocal evidence for a detectable conformational change between the metal-bound and metal-free forms of a vWF-A domain, and discuss the functional implications of this result for factor B.

Results and Discussion

The metal-binding site in the vWF-A superfamily

The sequence alignment of the vWF-A superfamily (Perkins *et al.*, 1994) showed that the metal coordination site corresponded to a conserved DxSxS motif on one side of the active-site cleft, together with conserved Thr and Asp residues on the other side of the cleft. This site was examined in 14 sets of CR3, LFA-1 and VLA-2 coordinates by molecular graphics. The molecular coordination to Mg^{2+} or to other divalent cations such as Mn^{2+} or Cd^{2+} differed only between the first-determined CR3-Mg^{2+} structure (PDB code 1ido) and the remaining 13 vWF-A-metal structures (Table 1). The metal coordination with DxSxS is the same in all 14 structures, and involves a water molecule bridge between the metal and a bidentate linkage with an Asp carboxyl group (distances of 3.9 Å to 4.3 Å), and two direct metal bonds with Ser hydroxyl groups (distances of 1.7 Å to 2.3 Å). In distinction to this, that with the conserved Thr and Asp residues involves a shift of the water molecule bridge from Asp in the first vWF-A structure (Lee *et al.*, 1995a; PDB code 1ido) to Thr in the remaining 13 structures (Figure 1). In the latter, the Thr hydroxyl-metal distance ranges between 3.8 Å to 4.6 Å, and the Asp carboxyl-metal distance ranges between 1.9 Å to 2.3 Å. The two vWF-A structures that did not bind metal are missing three (A1 domain) or two (A3 domain) residues in this DxSxS-D-T motif, thus explaining the lack of metal binding (Perkins *et al.*, 1994; Jenkins *et al.*, 1998). These comparisons show that two types of metal-vWF-A structures occur, one found in 13 of the 14 available metal-bound structures, and the other involving a structural change at one side of the active site cleft, which is hereinafter termed the conformationally-flexible side of this cleft. In the light of the recent controversy over CR3 integrin-metal structures (Baldwin *et al.*, 1998; Liddington & Bankston, 1998), in which it was unclear if metal affected the vWF-A structure, a reinvestigation of the binding of Mg^{2+} to the vWF-A domain by independent approaches based on solution structure methods was carried out.

CD studies of factor B and its fragments

CD spectroscopy is a monitor of the overall protein secondary structure and is sensitive to conformational changes (Drake, 1994). For CD studies, the recombinant vWF-A-218 and vWF-A-222 domains, the Bb fragment and factor B were prepared, each of which was eluted as a single peak from the HiLoad 16/60 Superdex-75 gel filtration column. Each migrated as a single band on SDS-PAGE under both reducing and non-reducing conditions. From these, the approximate masses of these proteins were estimated as: vWF-A-218 and vWF-A-222, 25 kDa; Bb fragment, 60 kDa; factor B, 90 kDa, all of which were as expected from their sequences. All the preparations were recognised in Western blots using a polyclonal antiserum to factor B, and were determined to be monomeric by mass spectrometry or analytical ultracentrifugation (Williams *et al.*, 1999; Hinshelwood *et al.*, 1999). Both the vWF-A-218 and vWF-A-222 domains

Table 1. Survey of distances between the metal ion and its coordinating atoms in homologous vWF-A crystal structures

vWF-A[a] (PDB code)	Metal	Coordination residue[b] DxSxS			Thr	Asp
CR3 (CD11b)		Asp140	Ser142	Ser144	Thr209	Asp242
1ido	Mg^{2+}	4.2, 4.2	2.2	2.2	2.1	4.2, 4.2
1jlm	Mn^{2+}	4.0, 4.4	2.3	2.1	4.0	2.3
1bho-1	Mg^{2+}	4.0, 4.1	2.0	2.0	4.2	2.1
1bho-2	Mg^{2+}	3.9, 4.6	2.3	1.7	4.6	2.1
1bhq-1	Cd^{2+}	3.9, 4.4	2.3	2.6	3.8	2.3
1bhq-2	Cd^{2+}	4.1, 4.3	2.2	2.2	3.8	1.9
LFA-1 (CD11a)		Asp137	Ser139	Ser141	Thr206	Asp239
1zoo-1	Mg^{2+}	3.9, 4.0	2.1	2.1	4.2	2.0
1zoo-2	Mg^{2+}	4.0, 4.1	2.1	2.1	4.2	2.1
1zop-1	Mn^{2+}	3.9, 4.0	2.1	2.1	4.2	2.0
1zop-2	Mn^{2+}	4.0, 4.1	2.1	2.1	4.0	2.1
1lfa-1	Mn^{2+}	4.0, 4.0	2.1	2.1	4.0	2.1
1lfa-2	Mn^{2+}	4.0, 4.0	2.1	2.1	4.0	2.1
VLA-2 (CD49b)		Asp151	Ser153	Ser155	Thr221	Asp254
1aox-1	Mg^{2+}	4.1, 4.0	2.2	2.0	4.2	2.1
1aox-2	Mg^{2+}	4.1, 4.0	2.3	2.0	4.2	2.0

Distances are in Å. The ligation site is to either the OD1 and/or OD2 atoms in D residues (*via* a water bridge in the case of the first D residue and the T residue in 13 of the 14 structures), the OG atom in S residues, and the OG1 atom in T residues. In the LFA-1 structures, chloride ions and/or water molecules form additional ligands with metal (not shown).

[a] Where two molecules appear in the unit cell, they are described by a suffix of −1 or −2.

[b] The crystallographic resolutions range between 1.7 Å to 3.0 Å.

were functionally active (Williams & Sim, 1994; Williams *et al.*, 1999). The vWF-A-222 domain was more soluble than the vWF-A-218 domain, and was therefore better suited for CD studies.

The secondary structures of the above four proteins were examined using their CD spectra. Each CD spectrum showed a large negative differential molar extinction coefficient $\Delta\varepsilon$ between 210 nm and 220 nm, with a small trough between these wavelengths, as expected for proteins with α-helix and β-sheet contents (Drake, 1994). The vWF-A-218 and vWF-A-222 spectra in the presence of 0.5 mM $MgCl_2$ were very similar (Figure 2), and indicated that these two recombinant vWF-A domains had similar folded structures. They were readily distinguishable from the factor B and Bb spectra by showing a higher negative $\Delta\varepsilon$. The observed secondary structures for each of the metal-bound and metal-free vWF-A domains were quantified using CONTIN (Provencher, 1982; Drake, 1994). This indicated a secondary structure with 26% α-helix and 31% β-sheet that agreed well with previous estimates by Fourier transform infrared spectroscopy (Perkins *et al.*, 1994), circular dichroism (Fairbanks *et al.*, 1995; Specks *et al.*, 1992) and the consensus secondary structure content from 23 vWF-A crystal structures (Hinshelwood *et al.*, 1999). The CD data agreed with the known presence of both α-helix and β-sheet in vWF-A domains.

On the addition of 1 mM EDTA to the vWF-A-218 or vWF-A-222 domains, a significant reduction of 15-24% in $\Delta\varepsilon$ at 215 nm from approximately $-3.5\ M^{-1}\ cm^{-1}$ to $-2.9\ M^{-1}\ cm^{-1}$ was observed (Figure 2). Similar experiments on the Bb fragment in the presence of Mg^{2+} and EDTA revealed a 4% reduction of $\Delta\varepsilon$ when EDTA was added. The direction of this change is the same as that seen for the recombinant vWF-A domains. No change was reproducibly detected within error for factor B under these conditions. These results were obtained in three CD sessions with up to three independent preparations. The same outcome was also obtained in the reversed experiment in which the CD spectrum of the vWF-A-222 domain prepared in the absence of Mg^{2+} was compared with the CD spectrum of the same sample containing 2 mM $MgCl_2$. In control experiments, no changes occurred in the CD spectrum of the vWF-A-222 sample when this was titrated with $CaCl_2$ instead of $MgCl_2$.

The changes in $\Delta\varepsilon$ in Figure 2 are best explained by noting that the SP domain is common to both the Bb fragment and factor B, but is absent from the vWF-A samples. The presence of the SP and SCR domains which contain dominantly β-sheet structures will reduce the proportion of α-helix by a factor of about 2 for the Bb fragment and about 3 for intact factor B when compared to the vWF-A domain, and would result in smaller Mg^{2+}-induced effects as observed. The reduced effect of Mg^{2+} on the CD spectra of the Bb fragment and factor B may also be affected by contacts between the SP domain and/or the Ba fragment with the vWF-A domain. Since α-helices have large effects on CD spectra, the increase in negative $\Delta\varepsilon$ for the recombinant vWF-A domain in the presence of Mg^{2+} implies that the bound metal has stabilised a small

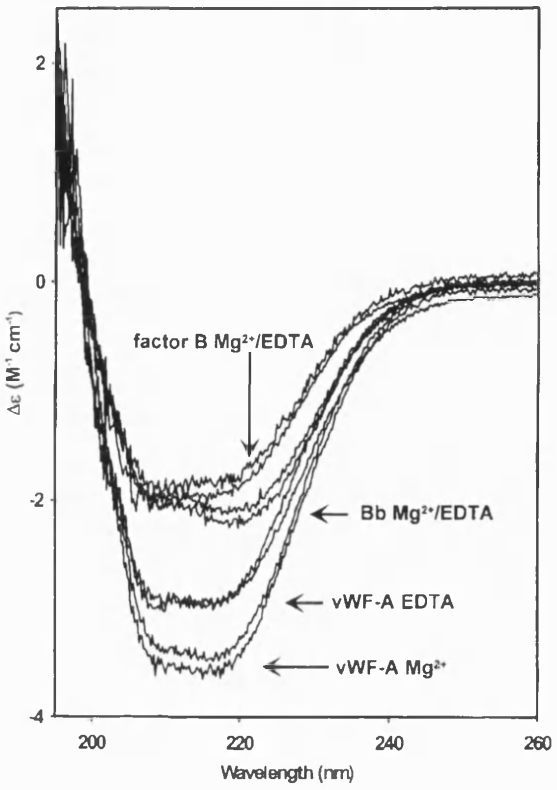


Figure 2. CD spectra at 20 °C of the vWF-A-218 and vWF-A-222 domains, the Bb fragment and factor B. The differential molar extinction coefficient $\Delta\varepsilon$ is shown as a function of wavelength. Sample concentrations ranged between 0.2 to 1.5 mg/ml. The buffer was 5 mM Tris-HCl (pH 7.5), 0.5 mM $MgCl_2$, to which 1 mM EDTA was added to remove protein-bound metal. The pairs of spectra for the vWF-A-218 and vWF-A-222 domains correspond to the presence of Mg^{2+} (lowest pair at 215 nm, as arrowed) and EDTA (next lowest pair as arrowed). The pairs of spectra for each of the Bb fragment and factor B correspond to the presence of Mg^{2+} or EDTA.

amount of ordered α-helix. In this context, it is of interest that the occurrence of α-helices is variable in crystal structures for the vWF-A superfamily (Figure 3 of Hinshelwood *et al.*, 1999), even within different crystal forms of the same molecule (Lee *et al.*, 1995a,b). For example, α-helix A5 becomes a short 3_{10}-helix in the LFA-1 (CD11a) structure, unlike in the other vWF-A structures. In the case of factor B, the vWF-A homology model showed that significant insertions were predicted just before and after the α-helices A4 and A5, the effect of which was to deepen the conformationally-flexible side of the metal-binding active site cleft (Figure 3).

Mg^{2+}-dependence of the vWF-A structure by FT-IR spectroscopy

FT-IR spectroscopy monitors the protein secondary structure through the separate observation of signals for α-helix, β-strand and other secondary structures (Haris & Chapman, 1994). In the absorbance spectrum of the vWF-A-222 domain in a ^2H_2O buffer with 5 mM $MgCl_2$, the amide I band appears as a broad signal centered at 1641-1642 cm^{-1} that corresponded principally to main-chain C=O stretching motions (Figure 4(a)). After calculation of the second derivative spectrum of the vWF-A-222 domain (a procedure that inverts the spectral intensities of the bands), two major bands at 1650 cm^{-1} and 1636 cm^{-1} were well resolved (Figure 4(b)). These two bands were assigned to α-helix and β-sheet structures, respectively. There were also small sub-components of the amide I band at 1659 cm^{-1}, 1670 cm^{-1} and 1692 cm^{-1}. The band at 1659 cm^{-1} may be attributed to either the presence of 3_{10}-helix or β-turns. The bands at 1670 cm^{-1} and 1692 cm^{-1} are likely to correspond to β-turn or anti-parallel β-sheet structures (Haris *et al.*, 1986; Surewicz *et al.*, 1993). These observations agree with the predicted structure of the factor B vWF-A domain and its homology model and the previous FT-IR study of the vWF-A-218 domain (Perkins *et al.*, 1994; Edwards & Perkins, 1995, 1996; Hinshelwood *et al.*, 1999).

Here, the greater stability of the vWF-A-222 domain compared to the vWF-A-218 domain permitted the effect of Mg^{2+} on its FT-IR spectrum to be studied in more detail. The FT-IR spectra were recorded using ^2H_2O buffers both in the presence and absence of metal. The removal of Mg^{2+} caused the β-sheet band at 1636 cm^{-1} to shift reproducibly to a slightly lower frequency of 1634 cm^{-1} with a minor reduction in its intensity (Figure 4(b)). This shift was attributable to a greater ^1H-^2H exchange of the main-chain amide protons of the metal-free vWF-A-222 domain as the result of a more flexible structure (Haris *et al.*, 1986, 1990; Jackson *et al.*, 1991; Perkins *et al.*, 1992). This explanation was based on observations of the extent of ^1H-^2H exchange using the intensity of the amide II band at both 1536 cm^{-1} and 1546 cm^{-1}, where the amide II band arises principally from N-H bending vibrations, which can be used to monitor ^1H-^2H exchange. These intensities were normalised by comparison with that of the well-resolved 1516 cm^{-1} Tyr band. The intensities of the residual amide II bands at both 1536 cm^{-1} and 1546 cm^{-1} were found to be approximately four times higher for the metal-bound vWF-A domain compared to that measured in the presence of EDTA after similar periods of dialysis into ^2H_2O buffers for both samples. In distinction to the changes seen in the CD spectrum, the α-helix band at 1650 cm^{-1} showed less dependence on the presence of Mg^{2+}, where no significant perturbation was observed on chelation of the metal (Figure 4(b)). The possibility that the observed spectral changes resulted from bands from EDTA was eliminated for reason of the use of (sample-buffer) difference spectra throughout and the lack of FT-IR bands from EDTA itself in the spectral range between 1630 cm^{-1} and 1700 cm^{-1}.

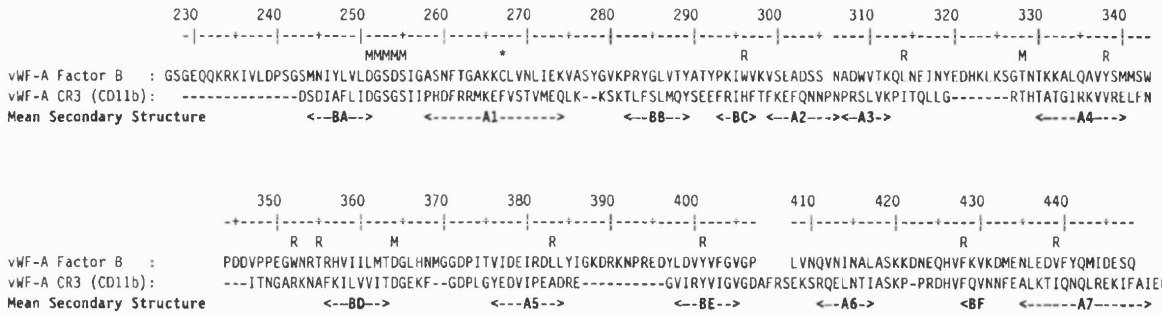


Figure 3. Sequence and numbering of the vWF-A domain in factor B. This is aligned with the sequence of CR3 and the mean secondary structure of the vWF-A superfamily (Hinshelwood *et al.*, 1999). The metal binding site DxSxS-D-T is indicated by M symbols above the sequences. Residues potentially involved in methyl ring current shift interactions are denoted by R for factor B. The vWF-A-218 sequence is shown up to Ile444, and vWF-A-222 corresponds to the replacement of Cys267 (asterisked) with Ser267 and the addition of four C-terminal residues.

The stability of the Mg^{2+}-bound and Mg^{2+}-free forms of the vWF-A-222 domain was investigated using thermal denaturation experiments. The 1634-1636 cm^{-1} band (β-sheet) exhibited a sigmoidal dependence on temperature, where its negative intensity was much diminished above 50 °C (Figure 5). The melting temperatures of the vWF-A domain were determined from the midpoints of the two graphs to be 46 °C for the Mg^{2+}-bound form and 36 °C for the Mg^{2+}-free form. Analysis of the negative intensity of the 1650 cm^{-1} band (α-helix) also demonstrated a sigmoidal temperature dependence with a similar denaturation temperature (data not shown). In contrast, a band at 1628 cm^{-1} that corresponded to random coil structures exhibited an increase in negative intensity with a reversed sigmoidal dependence as the temperature was increased. Thermal denaturation experiments that were also performed using CD spectroscopy confirmed that the Mg^{2+}-bound form was the more stable form (data not shown). It was concluded from both FT-IR and CD spectroscopy that the presence of Mg^{2+} significantly improved the stability of the vWF-A-222 domain.

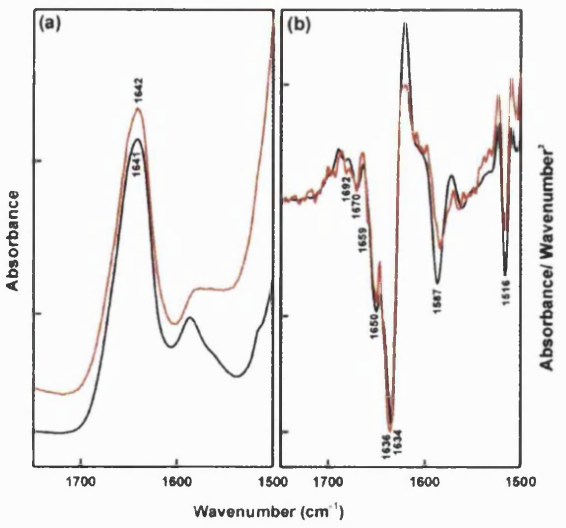


Figure 4. FT-IR spectroscopy of the amide I band of vWF-A-222 domain studied in PBS in ^2H_2O with either 5 mM $MgCl_2$ (red) or 5 mM EDTA (black). The concentration of the vWF-A-222 sample was 5 mg/ml for both samples. (a) The absorbance spectra; (b) the second derivative spectra. Band positions are identified by wavenumber labels.

Mg^{2+}-dependence of the vWF-A structure by ^1H NMR spectroscopy

The ^1H NMR spectra of both the vWF-A-218 and vWF-A-222 domains demonstrated significant NMR signal dispersion compared to the random coil spectra of the 20 amino acids (Figure 6). This meant that their folded structures in solution could be monitored by ^1H NMR. 500 MHz and 600 MHz ^1H NMR spectra were recorded at pH 7.50 at 30 °C in the presence of 0.5 mM Mg^{2+} for both the vWF-A-218 and vWF-A-222 domains. Both domains presented two high field-shifted signals upfield of 0 ppm, one denoted as 2v at −0.29 ppm, and a broader one denoted as 1v. The 1v signal appeared at −0.93 ppm for the vWF-A-222 domain and −0.97 ppm for the vWF-A-218 domain (Figure 6(a)), and is approximately one-third the intensity of the 2v signal, so these were assigned as single-proton and single-methyl signals, respectively. The appearance of two identical sets of several resonances immediately downfield of the residual water signal (not shown) was consistent with the presence of β-sheet structures in both vWF-A domains (Wishart *et al.*, 1992). The aromatic regions of the NMR spectra between 6 to 8 ppm for both vWF-A domains exhibited many overlapping signals, and in general both spectra were very similar (Figure 6(b)). However, at the left of Figure 6(b), an extra resonance at 7.86 ppm was present in the vWF-A-218 spectrum that was not present in that for vWF-A-222. The general

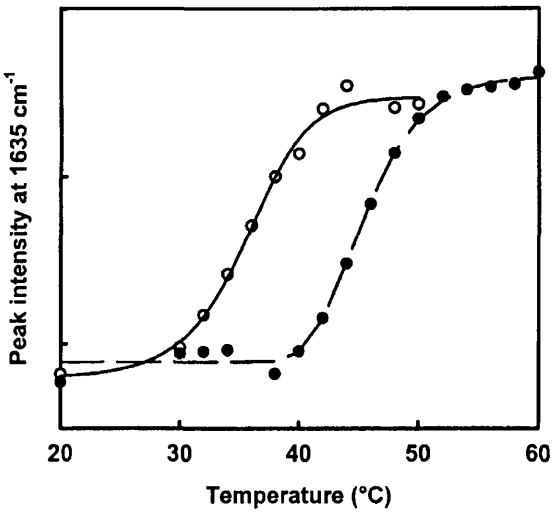


Figure 5. Dependence of the intensity of the second derivative band at 1635 cm^{-1} for the vWF-A-222 domain as a function of temperature in the presence of Mg^{2+} (●) and EDTA (○). The data points were fitted to a sigmoidal function represented by the continuous lines.

similarity of both vWF-A NMR spectra was as expected from their sequences (Figure 3). If the packing of the α-helix A7 against the vWF-A domain was affected by the C-terminal sequence extension in vWF-A-222, this may cause localised structural perturbations that would be reflected in the two minor NMR spectral differences involving the single-proton peak 1v and the extra aromatic signal at 7.86 ppm. It should be noted that Cys267 is on α-helix A1 (Figure 3), which is adjacent to α-helix A7, and its replacement may affect α-helix A7. Irrespective of how the shift changes were caused, the comparison of the vWF-A-218 and vWF-A-222 spectra suggested that the C-terminal α-helix A7 possessed a small degree of mobility within the vWF-A structure. This is consistent with the different conformations seen for α-helix A7 seen in different vWF-A crystal structures (Figure 1).

NMR titrations with Mg^{2+} and EDTA were performed to explore the effect of metal on the vWF-A structure (Figure 6(c) and (d)). When 1 mM EDTA was added to the vWF-A-218 domain, the high field signal 2v at −0.29 ppm was shifted downfield to −0.19 ppm, denoted as 2v′ (not shown). The aromatic signals in the downfield region of the NMR spectrum also demonstrated minor perturbations on the addition of 1 mM EDTA (not shown). To avoid the effect of the strong ^1H NMR signals from EDTA on the spectrum, and to show that the spectral changes were not caused by interaction of EDTA with the vWF-A domain, the vWF-A-222 domain was prepared in the absence of Mg^{2+}, and the NMR spectrum was studied in the presence of $MgCl_2$ at concentrations between 0 mM to 5 mM (Figure 6(c)). The reverse chemical shift change was observed, in which signal 2v′ (at −0.19 ppm) was initially observed in the absence of Mg^{2+}, then this was gradually replaced by signal 2v (at −0.29 ppm) on the addition of 0.1 mM and 0.2 mM Mg^{2+} (Figure 6(c)). This corresponded to slow exchange on the NMR timescale between the metal-free and metal-bound forms. The disappearance of signal 2v′ occurred when the $MgCl_2$ concentration reached 0.5 mM, which was approximately double the vWF-A concentration of 0.16 mM. The apparent excess of Mg^{2+} over the vWF-A domain needed to cause signal 2v′ to disappear is ascribed to the weak binding of metal to the protein. Other spectral changes occurred during the Mg^{2+} titration, as indicated by the two asterisked peaks in Figure 6(c) and the seven asterisked signals between 6.0 and 8.0 ppm in Figure 6(d), all of which were identified with the help of difference spectra. These differences demonstrated further that the vWF-A domain can exist in two conformations that depend on the presence or absence of Mg^{2+}. In a control experiment, no change in the vWF-A-222 spectrum was observed when $CaCl_2$ was added to a final concentration of 5 mM.

In order to examine whether the same Mg^{2+}-induced conformational change could be seen in the Bb fragment prepared from native factor B, the NMR experiments were repeated using this. The upfield NMR spectrum of the Bb fragment showed broader signals that corresponded to the doubling of the molecular weight of the vWF-A domain, and a high degree of spectral overlap was present (Figure 7). It was nonetheless possible to identify both the 2v and 1v signals at similar chemical shifts to those for the vWF-A spectra. In particular, the 2v signal of the Bb fragment showed an identical conformational dependence on the presence of Mg^{2+} to that of the vWF-A-222 domain, in that the 2v peak at −0.29 ppm was shifted downfield to the 2v′ position at −0.19 ppm when 1 mM EDTA was added. The 1v peak of the Bb spectrum was unaffected by the presence of Mg^{2+} or EDTA. These results show that the conformational dependence of the vWF-A domain on the presence of bound Mg^{2+} is relevant to the vWF-A domain in the context of the native Bb fragment.

Temperature and pH stability of the vWF-A domain by ^1H NMR spectroscopy

The thermal stability of the vWF-A domain was studied by ^1H NMR spectra. In the presence of 5 mM Mg^{2+}, the signal 2v for the vWF-A-222 domain was unchanged in chemical shift and intensity between 30 °C and 45 °C, and disappeared at 50 °C (data not shown). In contrast, the broad signal 1v moved downfield with increase of temperature, and also disappeared at 50 °C. Similar NMR results were obtained for the vWF-A-218 domain. When the metal-free vWF-A-222 domain was studied in the same manner, the high field-

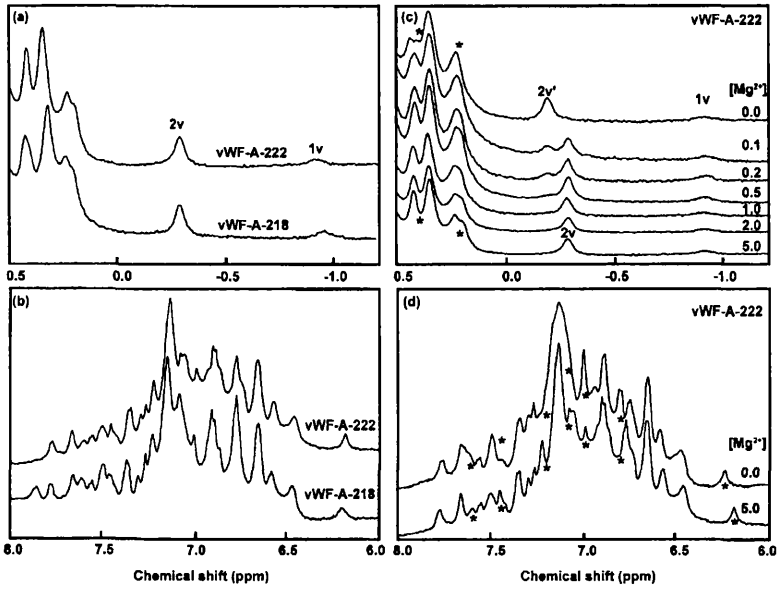


Figure 6. Comparison of the vWF-A-218 and vWF-A-222 domains by NMR and the effect of Mg^{2+}. (a) and (b) Comparison of the upfield and downfield spectral regions, respectively of the vWF-A-218 and vWF-A-222 domains (4 mg/ml; 0.16 mM) in the presence of 0.5 mM Mg^{2+}. The two highfield-shifted signals are labelled 2v and 1v. (c) Effect of Mg^{2+} on the upfield NMR spectrum of the vWF-A-222 sample, with the Mg^{2+} concentration in mM shown within square brackets. In addition to the difference between 2v and 2v', two more are highlighted by * symbols. (d) The aromatic region of the vWF-A-222 domain spectrum in the Mg^{2+}-bound and Mg^{2+}-free states. Seven differences between the two spectra are shown by an asterisk, for ease of visualisation.

shifted methyl signals and the conformationally-dependent aromatic signals disappeared at 40 °C and above. These NMR results showed that the Mg^{2+}-bound form of the vWF-A-222 domain denatured at 45-50 °C in the presence of Mg^{2+} and at 40 °C in its absence. This difference is in full agreement with the CD and FT-IR spectroscopy data above.

The pH stability of the vWF-A-218 domain in 0.5 mM Mg^{2+} was studied by NMR (data not shown). The vWF-A-218 signals 1v and 2v were unchanged throughout the pH range between 6.6 and 9.5, and likewise the aromatic region of the spectrum. The vWF-A-218 domain was poorly soluble below pH 5.0 when it precipitated. Despite this, the signal 2v underwent large downfield shifts at pH 5.9 and pH 5.4. This change is explained in terms of the general protonation of a nearby Asp or Glu residue at low pH.

Ring current calculations of the vWF-A homology model

Ring current calculations provide a molecular explanation for the highfield-shifted NMR signals of the vWF-A domain starting from a homology model for the vWF-A domain that was based on 23 sets of vWF-A coordinates (Hinshelwood *et al.*, 1999). The number of observed methyl signals in the upfield-shifted spectral region between 0.5 ppm and 0.0 ppm was estimated from signal areas. This gave an estimated total of about seven methyl signals for both the vWF-A-218 and vWF-A-222 domains when compared in intensity with the methyl signal 2v and the single proton signal 1v (Figure 6) observed upfield of 0 ppm. This total was compared with the calculated ring current shifts of all the vWF-A methyl groups as the result of their proximity to neighbouring aromatic side-chains, using the vWF-A homology model in conjunction with the program RCCAL (Perkins, 1982). This predicted seven resonances to occur between 0.5-0.0 ppm, in good agreement with the estimated number of observed signals (Figure 6). In addition, RCCAL predicted that two ring-current shifted signals originating from Leu314 and Leu383 were present upfield of 0.0 ppm (Table 2). The positional accuracy of the aliphatic and aromatic side-chains in the homology model was not considered sufficient to permit any firm assignments of the 2v sig-

Table 2. Four candidate ring-current interactions one of which is predicted to lead to the signal 2v in the vWF-A ^1H NMR spectrum of factor B

Methyl group (secondary structure)	Aromatic residue source of shift (secondary structure)	Predicted shift[a] (ppm from DSS)	Location in Figure 8
Leu314 (A3-A4 loop)	Trp296 (BC)	−1.41	Lower
Leu383 (A5)	Tyr338 (A4)	−0.42	Upper
Val439 (A7)	Tyr401 (BE), Phe428 (BF)	0.04	Lower
Thr355 (A4-BD loop)	Trp352 (A4-BD loop)	0.21	Upper

[a] The next predicted highest-field signal occurs at 0.43 ppm.

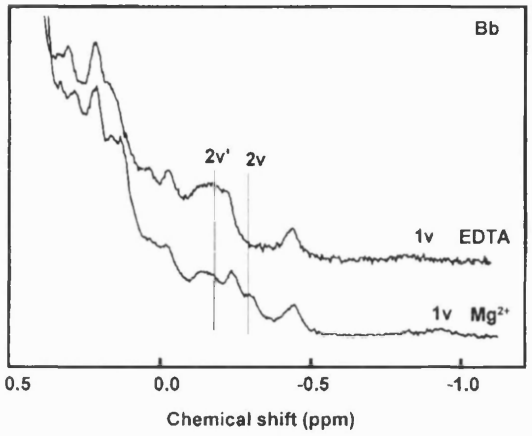


Figure 7. The dependence of the upfield spectral regions of the 500 MHz ^{1}H NMR spectra of the Bb fragment on the presence of Mg^{2+}. The sample concentration was 3 mg/ml. The lower spectrum was measured from a sample containing 1 mM $MgCl_2$ and the upper spectrum from the same sample containing 1 mM EDTA. The two highfield-shifted signals are labelled 2v and 1v in the upper spectrum and 2v′ and 1v in the lower spectrum.

nal to a given methyl group. This is illustrated by the prediction of the Leu383 methyl signal to occur at −0.42 ppm as the result of the location of its methyl group at 4.6 Å from the plane of the Tyr338 ring, while that for Leu314 was predicted to occur at −1.41 ppm as the result of the location of its methyl group at 4.0 Å from the plane of the Trp296 ring. This is a difference of only 0.6 Å in relative position from a given aromatic ring. In turn, this indicates that the shift change between the signals 2v and 2v′ may correspond to only a small movement between the methyl group and its proximate aromatic ring.

Four candidate ring-current interactions one of which may be responsible for signal 2v are summarised in Table 2 and Figure 8. All four corresponded to aliphatic and aromatic side-chains that are buried in the protein core, and all four involve the vWF-A α-helices. Their location in the vWF-A homology model (Figure 8) showed that two involved the upper face of the β-sheet, while the other two involved its lower face. It was of interest that two of the four predictions involved residues (Tyr338 and Thr355/Trp352) close to the conformationally-flexible side of the active-site cleft at the two metal-binding residues Thr328 and Asp364 (Table 1), as signal 2v had showed metal-dependent and pH-dependent properties (Figures 6(c) and 7). For signal 1v, an analysis of ring-current shifted single protons upfield of 0 ppm indicated that two such signals were possible, where both protons were proximate to Cys267 on α-helix A1 in a hydrophobic pocket associated with the aromatic ring of Tyr284. This prediction is consistent with the small dependence of signal 1v on the length of α-helix A7 and the replacement of Cys267 in the vWF-A-218 and vWF-A-222 domains.

Conclusions

The combined use of independent data sets from CD, FT-IR and ^{1}H NMR spectroscopy has provided the first unequivocal structural evidence that a vWF-A domain has two distinct conformations in the presence and absence of metal. Our joint application of CD and FT-IR spectroscopy offers complementary views of the vWF-A secondary structure. Since all three methods are based on solution studies, no ambiguities were caused by the presence of a crystallographic lattice. In our CD

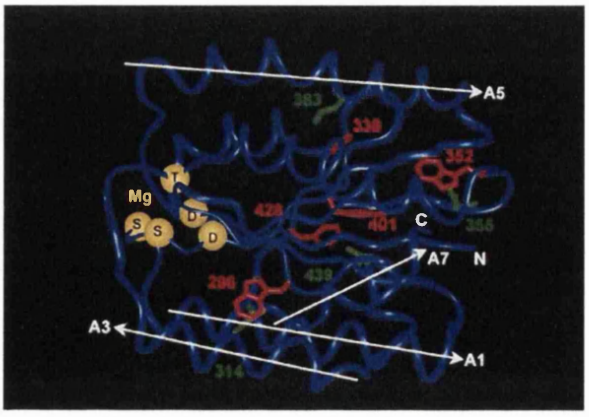


Figure 8. Methyl-aromatic ring interactions in the vWF-A homology model for factor B. The six-stranded β-sheet is viewed edge-on. Four potential interactions that may lead to the methyl signals 1v and 2v/2v′ are shown. The methyl and aromatic side-chains (green and red, respectively) are identified by residue numbering (Figure 3). The N and C termini are labelled as N and C, respectively, and the α-helices A1, A3, A4, A5 and A6 are labelled. The five C^{α} atoms of the Mg^{2+} binding site in the DxSxS-D-T motif are shown as yellow spheres.

study of the vWF-A domain from factor B in the presence of excess metal or EDTA, the magnitude of the $\Delta\varepsilon$ change between 215 nm to 220 nm on the addition of metal to the vWF-A-218 and vWF-A-222 domains was about 20%. This difference is most easily attributed to an alteration of at least the α-helix content of the vWF-A domain, as this secondary structure type makes the largest contribution to CD spectra (Drake, 1994). The FT-IR spectra were able to follow separately each of the α-helix and β-sheet bands, and these identified a small but detectable change in the β-sheet content of the metal-free vWF-A domain. The combination of both spectroscopic methods indicated that both the α-helix and β-sheet vWF-A secondary structures were perturbed by Mg^{2+} binding. The use of NMR spectroscopy permitted the monitoring of conformation-dependent chemical shifts that arise from interactions between proximate aliphatic and aromatic side-chains. The large number of signal perturbations seen in the presence of Mg^{2+} showed again that the metal affected the vWF-A structure. The same result was found for the vWF-A domain within the Bb fragment. Through the observation of NMR signals for single groups, the NMR data provided the clearest evidence yet for two metal-dependent conformational states of the vWF-A domain in solution. In thermal denaturation experiments, all three spectroscopic methods indicated a significant increase in the stability of the factor B vWF-A domain in the presence of Mg^{2+}. A similar stabilisation effect was recently observed for the vWF-A domain from the α1β1 integrin (very late activation protein-1; VLA-1; CD49a) in the presence of Mn^{2+} (Gotwals *et al.*, 1999).

Up to now, there has been no strong structural evidence that such a metal-induced conformational change occurs in integrin vWF-A domains, even though it is well established that they possess a Mg^{2+}-binding site. Only one crystal structure out of 23 had reported such a change for the domain found in CR3, but a subsequent crystal structure study of the same CR3 domain based on diffusion of Mg^{2+} or Cd^{2+} into the crystal had argued against this (Lee *et al.*, 1995a,b; Baldwin *et al.*, 1998). It was argued that the original metal-induced conformational change may be a crystallographic artefact caused by the movement of the C-terminal α-helix A7 of CR3 to bring Glu314 close to the metal-binding site of an adjacent molecule in the crystal lattice. In spectroscopic studies of integrin vWF-A domains to date, a maximal intensity change of 0 to 10% has been seen in the CD spectrum for the vWF-A domain of CR3 in which Mg^{2+} or Ca^{2+} was either present or absent (Fairbanks *et al.*, 1995), while little change was detected in the CD spectrum of the vWF-A domain from the α1β1 integrin in the presence or absence of metal (Nolte *et al.*, 1999).

From this study and others, there is an accumulation of evidence that the conformation of vWF-A domains can be modified during ligand binding and activation (Loftus & Liddington, 1997). Crystallography originally indicated that the vWF-A domain of CR3 possessed different biologically-relevant conformations when bound to either Mg^{2+} or Mn^{2+} (Lee *et al.*, 1995a,b). Evidence from mutational studies indicated that the vWF-A1 domain of von Willebrand factor existed in "on" and "off" conformations (Cooney & Ginsberg, 1996). The separate locations of the mutation sites on opposite sides of the vWF-A1 domain of von Willebrand factor that were associated with each of the Types 2B and 2M phenotypes in von Willebrand's disease indicated that the vWF-A1 domain was able to transmit ligand binding signals within its structure in order to upregulate or downregulate the binding of von Willebrand factor to platelets (Jenkins *et al.*, 1998).

In application to factor B, this Mg^{2+}-induced conformational variability is likely to be important for the activation and function of the C3 convertase when the Bb fragment is complexed with C3b. The Mg^{2+}-bound form will be involved in C3b binding, as the physiological concentration of Mg^{2+} in plasma is 0.7-1.1 mM (Kumar & Clark, 1990), of which 60% is present as free cations (Pryzdial & Isenman, 1986). This means that the metal site will be occupied in plasma factor B (Figure 6(c)). In this C3b.Bb complex, the vWF-A active-site cleft makes contact with C3b (Hinshelwood *et al.*, 1999), so it is likely that C3b will interact with bound Mg^{2+}. It is known that metal is trapped within the Bb.C3b complex after the complex is formed in the presence of metal (Fishelson *et al.*, 1983). The role of Mg^{2+} in the formation of this complex has not yet been defined, as the Bb.C3b complex can be formed in the presence and absence of Mg^{2+} (Pryzdial & Isenman, 1987; Williams & Sim, 1994). It is recognised from the present study that the vWF-A domain structure is stabilised by metal (Figure 5). As it has been shown here that the vWF-A domain can exist in one of two different conformations, it follows that this may provide the basis for an allosteric conformational signal for the activation of factor B, in which the binding of C3b will alter the Mg^{2+}-stabilised structure of the vWF-A domain. It can be proposed that a C3b-binding interaction at the active site cleft will alter the Mg^{2+} coordination in such a way that the conformations of surface α-helices directly connected to this cleft are modified (Perkins *et al.*, 1999). Evidence that this is plausible was shown by the greater conformational flexibility of one side of the cleft by crystallography (Table 1). In factor B, this involves Thr328 and Asp364 which are preceded by α-helix A3 and followed by α-helix A5. Topologically, one of these coincides with the region where the mutation sites on α-helix A3 cause the upregulation of von Willebrand factor in Type 2B von Willebrand's disease (Jenkins *et al.*, 1998). This α-helix A3 is adjacent to the α-helix A7 (Figure 8), which in factor B is joined to the serine protease domain of factor B by a short peptide link. The proposal that an allosteric signal can be transmitted from the active site cleft to the opposite sur-

face of the vWF-A domain is therefore consistent with what is known for vWF-A domains to date, and may explain how the serine protease domain of factor B is activated in the C3b.Bb complex.

Materials and Methods

Molecular graphics analyses of vWF-A domains

The metal binding site in the vWF-A structure was analysed using 23 sets of coordinates from the Protein Data Bank for 14 crystal structures for five vWF-A domains in the following proteins, namely the three integrins CR3, LFA-1 and VLA-2 as well as the A1 and A3 domains of von Willebrand factor (Lee *et al.*, 1995a,b; Baldwin *et al.*, 1998; Qu & Leahy, 1995, 1996; Emsley *et al.*, 1997, 1998; Celikel *et al.*, 1998; Huizinga *et al.*, 1997; Bienkowska *et al.*, 1997). The Protein Data Bank codes are identified in Table 1. The structures were subjected to visual inspection on Silicon Graphics INDY Workstations using INSIGHT II software (Biosym/MSI, San Diego) in conjunction with Crystal Eyes stereo glasses and the superfamily alignment of 75 vWF-A sequences in Perkins *et al.* (1994). The consensus secondary structure was identified using DSSP (Kabsch & Sander, 1983) and taken as the average across the crystal structures (Hinshelwood *et al.*, 1999). The construction of the homology model was based on the CR3 vWF-A domain (PDB code 1ido; Lee *et al.*, 1995a; Hinshelwood *et al.*, 1999).

Preparations of factor B, its Bb fragment and the recombinant vWF-A domain

Fresh frozen plasma was obtained from the Royal Free Hospital blood bank and stored at −20 °C before use. From this, factor B and the Bb fragment were purified following the procedures of Williams & Sim (1993), Lambris & Müller-Eberhard (1984) and Hinshelwood *et al.* (1999). Two recombinant vWF-A domains of factor B (denoted as vWF-A-218 and vWF-A-222) were prepared (Williams *et al.*, 1999), where vWF-A-218 contained the factor B sequence G229-I444 and was preceded by the residues GS derived from the thrombin recognition site LVPRGS (Figure 3). The vWF-A-222 sequence contained the factor B sequence G229-Q448, but differed in that a Cys267Ser mutation was made in order to eliminate a free Cys residue, as well as containing the N-terminal residues GS. To assess purity, the protein preparations were analysed by SDS-PAGE and their masses were determined by matrix-assisted laser desorption mass spectrometry (Hinshelwood *et al.*, 1999).

CD spectroscopy

CD spectroscopy was performed at 20 °C using a Jobin-Yvon CD6 spectropolarimeter with quartz cells of path length 0.2 mm and 0.1 mm. The instrument was calibrated with an aqueous solution of recrystallised D10-camphosulphonic acid ($\theta^{0.1\%,\,1\,cm} = 0.308$ at 290 nm). The vWF-A-222 domain, the Bb fragment and native factor B were dialysed into 5 mM Tris-HCl (pH 7.5), 0.5 mM $MgCl_2$, while the vWF-A-218 sample was dialysed into 100 mM $NaClO_4$, 5 mM Tris-HCl (pH 7.5), 0.5 mM $MgCl_2$, because of its lower solubility. Sample concentrations used to derive the differential molar extinction coefficient $\Delta\varepsilon$ were calculated from the absorbance at 280 nm (Drake, 1994) and were in the range of 0.2 to 1.5 mg/ml. The absorption coefficients A_{280} (1 cm, 1%) were calculated from the total of Trp, Tyr and Cys residues in the sequence to be 15.6 for vWF-A-218, 15.2 for vWF-A-222, 12.5 for the Bb fragment and 13.8 for factor B (Perkins, 1986). Experiments were performed in the presence of $MgCl_2$ to which EDTA was subsequently added to investigate the effect of Mg^{2+}. Alternatively, the CD spectra were recorded in the absence and presence of Mg^{2+} without the use of EDTA. Spectral quantification to obtain secondary structure contents was performed using CONTIN (Provencher, 1982). In each case the background spectrum from the sample buffer was subtracted from the protein sample spectrum.

FT-IR spectroscopy

FT-IR spectroscopy was performed using the vWF-A-222 domain using a 1750 Perkin-Elmer FT-IR spectrometer continuously purged with N_2 gas to reduce water vapour absorption in the spectral region of interest (Haris & Chapman, 1994). The sample was contained between CaF_2 windows fitted with a Teflon spacer to give a path length of 50 μm in a thermostatically-controlled cell housing. vWF-A samples at 5 mg/ml were in PBS buffer in 99.8% ^2H_2O with either 5 mM $MgCl_2$ or 5 mM EDTA. FT-IR spectra were obtained from both the protein sample and buffer at each temperature. 1000 scans were recorded at 20 °C and 100 scans were recorded for each temperature increment thereafter at a resolution of 4 cm^{-1}. The scans were signal averaged and the absorbance spectrum was obtained by digital subtraction of the ^2H_2O buffer spectrum from the sample spectrum (Haris *et al.*, 1986). Detailed analysis of the spectra was carried out using a second derivative procedure provided by the GRAMS software (Perkin-Elmer) with 13 data point Savitzy-Golay smoothing.

^1H NMR spectroscopy

Each protein preparation was dialysed into PBS in 99.8% ^2H_2O with 1.5 mM $MgCl_2$ over 36 hours with four buffer changes at 8 °C, except for the vWF-A-222 studies for which Mg-free PBS was used. NMR spectra were recorded at 30 °C using 500 MHz and 600 MHz Varian spectrometers and analysed using Varian VNMR software at the MRC NMR Centre at the National Institute of Medical Research, Mill Hill. Sample concentrations were in the range 2-10 mg/ml. pH meter readings in ^2H_2O were uncorrected values measured on a Corning 240 pH meter, and additions of stock solutions of NaO^2H or ^2HCl were used to adjust the pH. The H^2HO chemical shift was referenced with respect to the chemical shift of trimethylsilylpropionic acid (1 mg/ml) in the same buffer and temperature range as the protein samples, and the protein NMR spectra were then referenced with respect to the H^2HO signal. A line broadening of 1 Hz was applied before Fourier transformation. The effect of Mg^{2+} was investigated by adding up to 5 mM EDTA to some of the preparations before measuring the spectra. Alternatively spectra were recorded in the absence and presence of Mg^{2+}.

Calculation of ring current shifts

In the proton NMR spectra of proteins, resonances upfield of 0.5 ppm correspond to ring current shifts of methyl groups of amino acid side-chains positioned

close to one or several aromatic rings (Perkins, 1982). The ring current shifts are localised to within about 7 Å of the centre of the ring, and affect only a small proportion of proton signals. The program RCCAL (Perkins, 1982) was used on a Silicon Graphics INDY Workstation to calculate the ring current shifts from protein coordinates in Protein Data Bank format. The appearance of the NMR spectra were then predicted by correction of the random coil chemical shifts of the free amino acid residues by these calculated ring current shifts (Bundi & Wüthrich, 1978). The validity of these calculations has been established in calibration studies (Perkins, 1982).

Acknowledgements

We thank the Wellcome Trust for support, Dr P. I. Haris (De Montfort University, Leicester) and Dr R. B. Sim (MRC Immunochemistry Unit, Oxford) for useful discussions, Mr J. Gor for expert technical assistance, Dr A. J. Beavil (King's College, London) for access to a CD spectrometer and useful discussions, and Dr J. Feeney, Dr T. Frenkiel, Dr M. Gradwell and Dr G. Kelly for useful discussions and the provision of ^{1}H NMR facilities at the MRC National Institute of Medical Research, Mill Hill, London.

References

Baldwin, E. T., Sarver, R. W., Bryant, G. L., Jr, Curry, K. A., Fairbanks, M. B., Finzel, B. C., Garlick, R. L., Heinrikson, R. L., Horton, N. C., Kelley, L.-L. C., Mildner, A. M., Moon, J. B., Mott, J. E., Mutchler, V. T., Tomich, C.-S. C., *et al.* (1998). Cation binding to the integrin CD11b I domain and activation model assessment. *Structure*, **6**, 923-935.

Bienkowska, J., Cruz, M., Atiemo, A. & Liddington, R. (1997). The von Willebrand A3 domain does not contain a metal ion-dependent adhesion site motif. *J. Biol. Chem.* **272**, 25162-25167.

Bundi, A. & Wüthrich, K. (1978). ^{1}H NMR parameters of the common amino acid residues measured in aqueous solutions of the linear tetrapeptides H-Gly-Gly-X-L-Ala-OH. *Biopolymers*, **18**, 285-297.

Celikel, R., Varughese, K. I., Madhusan, Yoshioka, A., Ware, J. & Ruggeri, Z. M. (1998). Crystal structure of the von Willebrand factor A1 domain in complex with the function-blocking NMC-4 Fab. *Nature Struct. Biol.* **5**, 189-194.

Cooney, K. A. & Ginsburg, D. (1996). Comparative analysis of type 2b von Willebrand disease mutations: implications for the mechanism of von Willebrand factor binding to platelets. *Blood*, **87**, 2322-2328.

Drake, A. F. (1994). Circular dichroism. *Methods Mol. Biol.* **22**, 219-244.

Edwards, Y. J. K. & Perkins, S. J. (1995). The protein fold of the von Willebrand Factor type A domain is predicted to be similar to the open twisted β-sheet flanked by α-helices found in human ras-p21. *FEBS Letters*, **358**, 283-286.

Edwards, Y. J. K. & Perkins, S. J. (1996). Assessment of protein fold predictions from sequence information: the predicted α/β doubly wound fold of the von Willebrand Factor Type A domain is similar to its crystal structure. *J. Mol. Biol.* **260**, 277-285.

Emsley, J., King, S. L., Bergelson, J. M. & Liddington, R. C. (1997). Crystal structure of the I domain from integrin α2β1. *J. Biol. Chem.* **272**, 28512-28517.

Emsley, J., Cruz, M., Handin, R. & Liddington, R. (1998). Crystal structure of the von Willebrand factor A1 domain and implications for the binding of platelet glycoprotein Ib. *J. Biol. Chem.* **273**, 10396-10401.

Fairbanks, M. B., Pollock, J. R., Prairie, M. D., Scahill, T. A., Baczynskyj, L., Heinrikson, R. L. & Stockman, B. J. (1995). Purification and structural charcterization of the CD11b/CD18 integrin α subunit I domain reveals a folded conformation in solution. *FEBS Letters*, **369**, 197-201.

Fishelson, Z. V. I., Pangburn, M. K. & Muller-Eberhard, H. J. (1983). C3 convertase of the alternative complement pathway: demonstration of an active, stable B3b,Bb(Ni) complex. *J. Biol. Chem.* **258**, 7411-7415.

Gotwals, P. J., Chi-Rosso, G., Ryan, S. T., Sizing, I., Zafari, M., Benjamin, C., Singh, J., Venyaminov, S. Y., Pepinsky, R. B. & Koteliansky, V. (1999). Divalent cations stabilise the α1β1 integrin I domain. *Biochemistry*, **38**, 8280-8288.

Haris, P. I. & Chapman, D. (1994). Analysis of polypeptide and protein structures using Fourier transform infrared spectroscopy. *Methods Mol. Biol.* **22**, 183-202.

Haris, P. I., Lee, D. C. & Chapman, D. (1986). A Fourier transform infrared investigation of the structural differences between ribonuclease A and ribonuclease S. *Biochim. Biophys. Acta*, **874**, 255-265.

Haris, P. I., Chapman, D., Harrison, R. A., Smith, K. F. & Perkins, S. J. (1990). Conformational transition between the native and reactive-centre cleaved forms of α_1-antitrypsin by Fourier transform infrared spectroscopy and neutron scattering. *Biochemistry*, **29**, 1377-1380.

Hinshelwood, J., Spencer, D., Edwards, Y. J. K. & Perkins, S. J. (1999). Identification of the C3b binding site in a recombinant vWF-A domain of complement factor B by surface enhanced laser deionisation spectrometry and homology modelling: implications for the activation of factor B. *J. Mol. Biol.* **294**, 587-599.

Hourcade, D. E., Mitchell, L. M. & Oglesby, T. J. (1999). Mutations of the Type A domain of complement factor B that promote high-affinity C3b-binding. *J. Immunol.* **162**, 2906-2911.

Huber, R. & Bode, W. (1978). Structural basis of the activation and action of trypsin. *Acc. Chem. Res.* **11**, 1114-1122.

Huizinga, E. G., van der Plas, R. M., Kroon, J., Sixma, J. J. & Gros, P. (1997). Crystal structure of the A3 domain of human von Willebrand factor: implications for collagen binding. *Structure*, **5**, 1147-1156.

Jackson, M., Haris, P. I. & Chapman, D. (1991). Fourier transform infrared spectroscopic studies of Ca^{2+}-binding proteins. *Biochemistry*, **30**, 9681-9686.

Jenkins, P. V., Pasi, K. J. & Perkins, S. J. (1998). Molecular modelling of ligand and mutation sites of the Type A domains of human von Willebrand factor and their relevance to von Willebrand's disease. *Blood*, **91**, 2032-2044.

Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577-2637.

Kumar, P. J. & Clark, M. L. (1990). *Clinical Medicine*, Second edit., pp. 490-510, Bailliere-Tindall, London.

Lambris, J. D. & Müller-Eberhard, H. J. (1984). Isolation and characterization of a 33,000-dalton fragment of complement Factor B with catalytic and C3b binding activity. *J. Biol. Chem.* **259**, 12685-12690.

Law, S. K. & Reid, K. B. M. (1995). *Complement*, Second edit., IRL Press, Oxford.

Lee, J. O., Rieu, P., Arnaout, M. A. & Liddington, R. (1995a). Crystal structure of the A domain from the α subunit of integrin CR3 (CD11b/CD18). *Cell*, **80**, 631-638.

Lee, J. O., Bankston, L. A., Arnaout, M. A. & Liddington, R. C. (1995b). Two conformations of the integrin A-domain (I-domain): a pathway for activation? *Structure*, **3**, 1333-1340.

Liddington, R. & Bankston, L. (1998). The integrin I domain: crystals, metals and related artefacts. *Structure*, **6**, 937-938.

Loftus, J. C. & Liddington, R. C. (1997). New insights into integrin-ligand interaction. *J. Clin. Investig.* **99**, 2302-2306.

Mole, J. E., Anderson, J. K., Davison, E. A. & Woods, D. E. (1984). Complete primary structure for the zymogen of human complement factor B. *J. Biol. Chem.* **259**, 3407-3412.

Nolte, M., Pepinsky, R. B., Venyaminov, S. Y., Koteliansky, V., Gotwals, P. J. & Karpusas, M. (1999). Crystal structure of the α1β1 integrin I-domain: insights into integrin I-domain function. *FEBS Letters*, **452**, 379-385.

Perkins, S. J. (1982). Applications of ring current calculations to the proton NMR of proteins and transfer RNA. *Biol. Magn. Reson.* **4**, 193-336.

Perkins, S. J. (1986). Protein volumes and hydration effects: the calculation of partial specific volumes, neutron scattering matchpoints and 280 nm absorption coefficients for proteins and glycoproteins from amino acid sequences. *Eur. J. Biochem.* **157**, 169-180.

Perkins, S. J. & Wuthrich, K. (1980). Conformational transition from trypsinogen to trypsin: ^{1}H nuclear magnetic resonance at 360 MHz and ring current calculations. *J. Mol. Biol.* **138**, 43-64.

Perkins, S. J., Smith, K. F., Nealis, A. S., Haris, P. I., Chapman, D., Bauer, C. J. & Harrison, R. A. (1992). Secondary structure changes stabilise the reactive-centre cleaved form of SERPINs: a study by ^{1}H NMR and Fourier transform infrared spectroscopy. *J. Mol. Biol.* **228**, 1235-1254.

Perkins, S. J., Smith, K. F., Williams, S. C., Haris, P. I., Chapman, D. & Sim, R. B. (1994). The secondary structure of the von Willebrand domain in factor B of human complement by Fourier transform infrared spectroscopy: its occurrence in collagen types VI, VII, XII and XIV, the integrins and other proteins by averaged secondary structure predictions. *J. Mol. Biol.* **238**, 104-119.

Perkins, S. J., Hinshelwood, J., Edwards, Y. J. K. & Jenkins, P. V. (1999). Structural and functional modelling of vWF-A domains in complement and coagulation. *Biochem. Soc. Trans.* **27**, 815-821.

Provencher, S. W. (1982). A constrained regularization method for inverting data represented by linear algebraic or integral-equations. CONTIN: a general-purpose constrained regularization program for inverting noisy linear algebraic and integral-equations. *Comput. Phys. Commun.* **27**, 213-227, 229-242.

Pryzdial, E. L. & Isenman, D. E. (1986). A reexamination of the role of magnesium in the human alternative pathway of complement. *Mol. Immunol.* **23**, 87-96.

Pryzdial, E. L. & Isenman, D. E. (1987). Alternative complement pathway activation fragment Ba binds to C3b. Evidence that formation of the factor B-C3b complex involves two discrete points of contact. *J. Biol. Chem.* **262**, 1519-1525.

Qu, A. & Leahy, D. J. (1995). Crystal structure of the I-domain from the CD11a/CD18 (LFA-1, αLβ2) integrin. *Proc. Natl Acad. Sci. USA*, **92**, 10277-10281.

Qu, A. & Leahy, D. J. (1996). The role of the divalent cation in the structure of the I domain from the CD11a/CD18 integrin. *Structure*, **4**, 931-942.

Specks, U., Mayer, U., Nischt, R., Spissinger, T., Mann, K., Timpl, R., Engel, J. & Chu, M.-L. (1992). Structure of recombinant N-terminal globule of type VI collagen α3 chain and its binding to heparin and hyaluronan. *EMBO J.* **11**, 4281-4290.

Surewicz, W. K., Mantsch, H. H. & Chapman, D. (1993). Determination of protein secondary structure by Fourier transform infrared spectroscopy: A critical assessment. *Biochemistry*, **32**, 389-394.

Tuckwell, D. S., Xu, Y. Y., Newham, P., Humphries, M. J. & Volanakis, J. E. (1997). Surface loops adjacent to the cation-binding site of the complement factor B von Willebrand Factor type A module determine C3b binding specificity. *Biochemistry*, **36**, 6605-6613.

Williams, S. C. & Sim, R. B. (1993). Dye-ligand affinity purification of human complement factor B and β-2 glycoprotein I. *J. Immunol. Methods*, **157**, 25-30.

Williams, S. C. & Sim, R. B. (1994). Binding sites involved in the formation of the C3 (H2O)-factor B complex of the alternative pathway of complement. *Biochem. Soc. Trans.* **22**, 2S.

Williams, S. C., Hinshelwood, J., Perkins, S. J. & Sim, R. B. (1999). Production and functional activity of a recombinant von Willebrand Factor (vWF-A) domain from human complement factor B. *Biochem. J.* **342**, 625-632.

Wishart, D. S., Sykes, B. D. & Richards, F. M. (1992). The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry*, **31**, 1647-1651.

Edited by R. Huber

(Received 4 January 2000; received in revised form 14 February 2000; accepted 14 February 2000)