Proteasome genes in the MHC: implications for antigen processing
by
Richard James Glynne

A thesis submitted for the degree of Doctor of Philosophy, May 1994

Imperial Cancer Research Fund
Lincoln's Inn Fields
London WC2A 3PX
and
Department of Genetics and Biometry
University College London
ACKNOWLEDGEMENTS

My thanks go to the following for their support. The members of the lab over the past years have made my time in it very enjoyable. They are, in no particular order, Isabel Hanson, Ruth Lovering, Jiannis Ragoussis, Ian Mockridge, Ian Campbell, Will Foulkes, Stephen Powis, Gordon Allan, Lesley-Anne Kerr, Adrian Kelly, Philippe Sanséau, Amanda Jackson, Louise Hosking, Jethro Herburg, Frances Sanderson, Cynthia Dixon, Pat Miller and Mônica Belich. Thanks are due also to John Trowsdale for his guidance and supervision. Mum, Dad, Bob and Fid have been a great family throughout my life. Gwyn has excelled at being a friend, girlfriend and flat-mate and consistently laughed at my jokes, no matter how unfunny.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS ........................................................................................................ 2

TABLE OF CONTENTS ......................................................................................................... 3

TABLE OF FIGURES ............................................................................................................. 8

ABSTRACT .......................................................................................................................... 12

Chapter 1: INTRODUCTION ............................................................................................ 14

1. A history of immunogenetics and the major histocompatibility complex ......................... 14
2. MHC-encoded proteins present antigens to T cells ......................................................... 17
3. Processing of antigen for presentation through MHC class I and class II ......................... 17
4. Elution of peptides from class I molecules ....................................................................... 19
5. Processing of proteins for presentation through MHC class I ........................................... 21
6. Mutant cell lines unable to process antigen for presentation through class I ................... 22
7. Assembly of class I, β2m and peptide ............................................................................. 24
8. Towards a genetic explanation of the defect in the mutant cell lines - cloning of ATP-binding cassette (ABC) transporter genes in the MHC ................................................................. 26
9. The function of the TAP gene products defined by transfection ...................................... 28
10. The function of the TAPs: biochemical assays .............................................................. 31
11. The degradation machinery for the production of peptides to be presented through class I - cytosolic proteolysis and the proteasome ........................................................................ 35
    11a. Characterisation of the proteasome ......................................................................... 35
    11b. Non-proteolytic activities of the proteasome ......................................................... 38
11c. Mechanism of proteolysis by the proteasome ..........39
11d. Ubiquitin and the 26S complex .................................................43
11e. Regulation of the proteasome .............................................51
11f. Structure of the proteasome ................................................56
11g. Primary sequences of proteasome subunits ......................61

12. The LMPs ...................................................................................69

13. Is there a role for the class I molecule in the processing
of its own antigens? ........................................................................72

14. Summary of antigen processing for presentation
through class I ..................................................................................75

Chapter 2: MATERIALS AND METHODS ..................................................80

1. Solutions ..................................................................................80

2. DNA preparation and manipulation ........................................82

   Labelling of DNA fragments using random hexamer
   priming ..................................................................................82

   Screening of cDNA libraries .....................................................82

   Hybridisation of filters ..........................................................83

   Southern blotting .....................................................................84

   Plasmid minipreps ..................................................................84

   Plasmid maxipreps ..................................................................85

   Sequencing of double stranded plasmid DNA .......................87

   Shotgun sequencing ..................................................................88

       M13 vector ........................................................................88

       Insert ................................................................................88

       Ligation .............................................................................89

       Preparation of JM109 bacteria .............................................89

       Single strand M13 minipreps for sequencing .......................90

   Solid phase sequencing of PCR products ..............................91

   Strand specific labelling of DNA fragments .........................91
Chapter 3: CLONING OF LMP7, A PROTEASOME-RELATED GENE IN THE CLASS II REGION OF THE MHC

INTRODUCTION ........................................................................................................... 104

RESULTS and DISCUSSION ................................................................................... 105
Chapter 4: AN INVESTIGATION OF THE LMP GENE PRODUCTS

SUMMARY ........................................................................................................124

INTRODUCTION .............................................................................................124

RESULTS and DISCUSSION ..........................................................................126

Production and specificity of the BNFl antiserum ..........................126
LMP7 is a component of the proteasome complex .........................132
LMP2 is a component of the proteasome ..........................................134
The LMPs are synthesised with an N-terminal leader sequence .................................................................138
Assembly of the LMPs into the proteasome .....................................145
The mature forms of the LMPs are upregulated by interferon-γ .................................................................152
LMP7-containing proteasomes are proteolytically active ..................158

CONCLUSIONS ...............................................................................................160

ACKNOWLEDGEMENTS ..............................................................................162
Chapter 5: UNUSUAL GENOMIC ORGANISATION OF THE LMP7 LOCUS

SUMMARY ........................................................................................................163
INTRODUCTION .............................................................................................164
RESULTS and DISCUSSION ......................................................................168
  Isolation of two different LMP7 transcripts........................................168
  Messenger RNA expression levels of the alternative splice products.......................................................................................................174
  A coding polymorphism within the first exon of LMP7b.............................177
  Three types of transcripts from the opposite strand to LMP7 and TAP1 ..........................................................180
  RING9 RNA is upregulated by interferon-γ...........................................186
  A comparison of the human and mouse genomic DNA sequences at the TAP1/ LMP7 locus.........................................................188
CONCLUSIONS ...............................................................................................196

Chapter 6: CONCLUDING COMMENTS ...........................................................199
  Do the LMPs have a role in antigen processing? ......................................199
  Future prospects .............................................................................................206

Chapter 7: REFERENCES .................................................................................210
TABLE OF FIGURES

Figure 1.1
A model for the structure of the eukaryotic proteasome ........................................... 59

Figure 1.2
A dendogram of aligned proteasome sequences ............................................................ 64

Figure 1.3a
A similarity plot of the alpha-type proteasome sequences ........................................... 67

Figure 1.3b
A similarity plot of the beta-type proteasome sequences ............................................. 68

Figure 1.4
Potential processes involved in the processing of antigen for class I presentation ............. 76

Figure 3.1
The genomic position of LMP2 and LMP7 in relation to the TAP genes ......................... 107

Figure 3.2
The nucleotide sequence of the LMP7 cDNA ............................................................... 109

Figure 3.3
Alignment of LMP7 with proteasome sequences .......................................................... 110

Figure 3.4
LMP7 compared to proteasome sequences by diagon plots ......................................... 113

Figure 3.5
Comparison of LMP7 with protease active sites ......................................................... 114

Figure 3.6
Northern analysis of LMP7 mRNA expression ............................................................ 118

Figure 4.1
Activity of the anti-LMP7 serum assessed in an ELISA assay ..................................... 128
Figure 4.2
Specificity of the anti-LMP7 serum determined by western blot ................................................................. 129

Figure 4.3
Specificity of the anti-LMP7 serum determined by immunoprecipitation ....................................................... 129

Figure 4.4
Immunoprecipitation of proteins by the BNF1 antiserum can be inhibited by R10C15 peptide .................................. 131

Figure 4.5
Immunoprecipitations with the BNF1 antiserum resemble those of an anti-proteasome serum .................................. 133

Figure 4.6
Western analysis of proteasome precipitates ..................................................................................... 135

Figure 4.7
The anti-LMP2 serum specifically recognises a 22 kD protein ..................................................................... 136

Figure 4.8
Specificity of the anti-LMP2 serum determined by immunoprecipitation .......................................................... 137

Figure 4.9
Precipitation of LMP7 after dissociation of the proteasome ........................................................................ 139

Figure 4.10
LMP7 is processed from a 28 kD precursor ..................................................................................................... 141

Figure 4.11
Precipitation of LMP2 after dissociation of the proteasome ........................................................................ 143

Figure 4.12
LMP2 is processed from a 26 kD precursor ..................................................................................................... 144

Figure 4.13
Schematic of the reprecipitation experiment of figure 4.14..........................146

Figure 4.14
The mature forms of the LMPs are assembled into the proteasome approximately coincidentally with their processing ..........................................................................................................147

Figure 4.15
The LMPs are upregulated by interferon-γ ...................................................153

Figure 4.16
Surface expression of class I parallels expression of the LMPs..........................155

Figure 4.17
The LMP7-containing complex has proteolytic activity .............................159

Figure 5.1a
Genomic organisation of LMP7 ......................................................................169

Figure 5.1b
Genomic sequence of the alternative first exons of LMP7 ............................170

Figure 5.1c
Homology between the interferon response element of the GBP promoter and genomic sequence 5' of the LMP7b gene ..........171

Figure 5.2
A comparison of the derived protein sequences of LMP7a, LMP7b and MC13 by diagon analysis ...........................................................173

Figure 5.3
Northern analysis of LMP7a and LMP7b ..........................................................175

Figure 5.4
Schematic representation of the RING9 splice patterns ................................181

Figure 5.5
Sequences of the RING9 clones and the exons of TAP1, LMP7a and LMP7b. ............................................................................................182
Figure 5.6
Northern analysis of RING9 expression ......................................................187

Figure 5.7
A comparison of human and mouse RING9 genomic sequence by diagonal analysis ...........................................................................191

Figure 5.8
A similarity plot of the RING9 genomic sequence of human and mouse .........................................................................................................192

Figure 5.9
Alignments of conserved, non-coding regions from the human and mouse RING9 genomic sequence ..............................................................193

Figure 6.1
Redundancy in the proteasome subunits ........................................................201
ABSTRACT

The aim of the work described in this thesis was to clone new genes from the class II region of the human major histocompatibility complex (MHC) and to investigate their function.

The first section describes the cloning of LMP2 and LMP7, two genes with a proposed role in antigen processing. Both genes have homology to components of the proteasome, a 700 kD cytoplasmic complex, containing 10-20 subunits. The proteasome has well defined protease activities, suggesting that LMP2 and LMP7 may function in the degradation of cytoplasmically synthesised proteins to form antigenic peptides. This hypothesis is strengthened by the fact that the LMPs are closely linked to, and coordinately regulated with, another pair of genes, called TAP1 and TAP2. Studies with mutant cell lines have shown that the TAP gene products function to mediate the transport of antigenic peptides into the endoplasmic reticulum.

Using antisera raised against LMP2 and LMP7, I showed that both proteins are synthesised as precursors. These precursors are cleaved at the N terminus to form mature proteins which are incorporated into the proteasome. The expression of both LMP2 and LMP7 proteins was up-regulated by interferon-γ, a cytokine which also up-regulates MHC proteins at the cell surface.

Further examination of the LMP7 locus revealed an unusual genomic structure. Two LMP7 transcripts with alternative first exons
were isolated from cDNA libraries. In addition, three splice variants transcribed from the opposite strand to \textit{TAP1} and \textit{LMP7} were cloned. These did not contain an obvious open reading frame. A comparison of this locus with that of the mouse showed that only one of the alternative \textit{LMP7} exons was conserved and that the opposite strand transcripts were poorly conserved.
CHAPTER 1: INTRODUCTION

1. A history of immunogenetics and the major histocompatibility complex

The major histocompatibility complex, or MHC, is possibly the best characterised region of the human genome. The phenomenon of histocompatibility came from work on the house mouse. E. E. Tyzzer, in 1916, showed that tumours from one strain of inbred mouse were unable to grow when grafted onto other strains. When unrelated strains were crossed, first generation mice were susceptible to tumours grafted from a parent. However, a variable number of the second generation were susceptible to such grafted tumours. It was concluded that the trait of susceptibility to a grafted tumour was polygenic (Little and Tyzzer, 1916).

The field of blood group antigens was linked to tumour rejection by P. A. Gorer. There was a correlation between the agglutination of erythrocytes from a range of mouse strains by a rabbit anti-mouse serum and the rejection of a tumour grafted onto the same mouse strains. It seemed that a blood group antigen, called antigen II, and the tumour rejection antigen may be related and it was concluded that "antigen II must be present in the tissues of the host, otherwise the tumour will regress" (Gorer, 1937).

Observations by P. B. Medawar demonstrated that graft rejection had features in common with immunological phenomena. Skin grafts onto human burns patients were accepted for some time before eventual rejection. However, a subsequent graft from the same donor onto the
same patient was rapidly rejected. A similar process occurred when grafting between mice strains was used as an experimental system: sensitisation of graft rejection was donor specific and was independent of where the second graft was transplanted (Medawar, 1946).

To separate the loci involved in tumour rejection, G. D. Snell embarked on a programme of back-crossing progeny from a cross between mice of different tumour rejection phenotypes. At each cross the animals were selected for the ability to reject a given tumour. The result was a series of congenic lines that differed in some of the resistance loci but not all. Use of the congenic lines showed that histocompatibility antigens could be divided into major antigens, which produced a strong rejection response, and minor, which produced only a weak effect, often overcome by the tumour (reviewed in Snell, 1981 and Bailey, 1975). One set of crosses showed that a tumour resistance phenotype was genetically linked to a fused tail mutation present in one of the parents (Snell, 1948). The absence of antigen II from these mice led to the suggestion that antigen II and the tumour resistance phenotype were also linked (Gorer, Lyman et al, 1948). This locus was named histocompatibility-2, or H-2.

The discovery of the HLA complex in humans was prompted by two observations. First, patients who had undergone multiple rounds of blood transfusions raised antibodies against the transfused lymphocytes (Dausset, 1958). Second, multiple pregnancies led to the production of antibodies in the mother’s blood raised against paternally derived antigens on fetal lymphocytes (Payne and Rolfe, 1958, van Rood, Earnisse et al, 1958). Family studies confirmed that, as in the mouse, these lymphocyte antigens were encoded at a single locus.
A second type of lymphocyte antigen was defined using techniques other than serology and grafting. Mixing lymphocytes from unrelated individuals induced the cells to transform and proliferate. This became the basis of the mixed lymphocyte reaction or MLR (Bach and Amos, 1967). Typing of lymphocytes using MLR or serology gave results which were broadly consistent. However, rare recombination events were able to distinguish the MLR locus from the serology locus (Yunis and Amos, 1971).

A further typing assay was devised in the guinea pig and the mouse which relied on the response of the animal to injections of simple polypeptides. The response to such challenges was controlled by what became known as the immune response gene, or Ir. Again there was a strong correlation between the response to the polypeptides and the conventional serological typing. More detailed typing of mouse strains recombinant over the H-2 region showed that the Ir locus was linked to, but distinct from, H-2 (McDevitt, Deak et al, 1972).

The Ir and MLR-defined loci encode what have now become called the class II MHC antigens. The antigens defined by serology and grafting are known as class I MHC antigens. Class II proteins form heterodimers of MHC-encoded α and β chains (Walsh and Crumpton, 1977). Class I antigens associate with a low molecular weight protein known as β2 microglobulin which is necessary for their expression. In the Daudi cell line, which has a mutation in the initiation codon of β2m, class I proteins do not reach the cell surface (Rosa, Berissi et al, 1983). Crystal structures for class I/β2 microglobulin and class II α/β complexes have now been solved (Bjorkman, Saper et al, 1987, Madden, Gorga et al, 1991, Brown, Jardetzky et al, 1993).
2. MHC-encoded proteins present antigens to T cells

Matching of MHC haplotypes is now routinely used in organ transplant procedures. However, a more physiological role for the MHC was detailed by Zinkernagel and Doherty. Sensitised mouse cytotoxic T cells (CTL) were able to lyse target cells when infected by the lymphocytic choriomeningitis virus. This lysis only occurred if the CTL and the target cells shared alleles at class I of the MHC. The theory that CTL recognised viral proteins in the context of particular MHC-encoded proteins became known as MHC restriction (Zinkernagel and Doherty, 1974, Blanden, Doherty et al, 1975). A similar phenomenon was seen when human cells were used (McMichael, Ting et al, 1977).

Two subclasses of T cells could be defined on the basis of the differential expression of cell surface markers. On the Ly marking system, T cells differentiated from Ly $123^+$ to be either Ly $1^+$ or Ly $23^+$. Only those T cells which were Ly $23^+$ (CD$8^+$, CTL) were able to develop cytolytic activity. This cytolytic activity could be amplified by the presence of Ly $1^+$ helper T cells (Th). Both CTL and Th are restricted through allelic products of the MHC. However, CTL are restricted through MHC class I and Th through MHC class II (Cantor and Boyse, 1975).

3. Processing of antigen for presentation through MHC class I and class II

The presentation of $^{125}$I-labelled *Listeria monocytogenes* through class II of macrophages to Th was used as a model system for class II
restricted antigen processing (Ziegler and Unanue, 1982). The catabolism and presentation of this antigen to Th were inhibited by the lysosomotropic agents chloroquine and ammonia, although uptake and ingestion were unaffected. These data suggested that presentation through class II required degradation of the antigen in an acidified intracellular compartment.

CTL lysis of infected cells through class I was thought to be initiated by the recognition of viral encoded surface glycoprotein next to MHC class I in the cell membrane. However, experiments using defined reassortant influenza A viruses showed that a subset of anti-viral CTL were specific for non-membrane proteins such as viral polymerase and nucleoprotein (Bennink, Yewdell et al, 1982, Townsend and Skehel, 1982, Kees and Krammer, 1984). These results were confirmed by transfection of the individual viral nucleoprotein gene into L cells. The transfected cells are targets for CTL lysis despite the lack of detectable nucleoprotein at the cell surface (Townsend, 1984). Furthermore, transfection only of fragments of nucleoprotein could allow CTL lysis (Townsend, Gotch et al, 1985). Subsequently, synthetic peptides were also used to sensitize cells to CTL lysis and to define the epitopes within an antigenic protein (Townsend, Rothbard et al, 1986).

Both class I and class II pathways of antigen processing appeared to necessitate degradation of the antigen. The discovery of CD4+ class II restricted mouse T cell clones with cytolysis activity (Lukacher, Morrison et al, 1985) allowed the direct comparison of class I and class II presentation in the same assay, that is, in vitro cellular cytotoxicity. Influenza haemagglutinin (HA) specific class I and class II specific
cytotoxic T cells were used with either a B cell lymphoma line or class II transfected L cells as the presenting cells.

Several differences between the class I and class II presentation pathways were described. Only the class II restricted clones could recognise target cells treated with non-infectious virions. Class I restricted clones required de novo protein synthesis in the target cell for cytotoxicity whereas inhibition of protein synthesis had no effect on the presentation of HA to class II restricted clones. Conversely, only the class II pathway of presentation was sensitive to chloroquine. Finally, expression of HA from a recombinant vaccinia virus sensitised the target cells to lysis by the class I restricted T cell clones but not those restricted through class II. The conclusion from these experiments was that presentation through class I occurred through an "endogenous" pathway necessitating cytoplasmic protein synthesis. Class II presentation occurred through an "exogenous" pathway - able to utilise only protein added externally and degrading this protein in a chloroquine-sensitive step (Morrison, Lukacher et al, 1986, Braciale, Morrison et al, 1987). The work described in this thesis is likely to have more bearing on the mechanics of processing antigen for presentation through MHC class I. The production and presentation of cytosolic proteins by class I molecules is, therefore, described in more detail below.

4. Elution of peptides from class I molecules

The first crystal structure of a class I molecule was consistent with presentation of peptides as opposed to intact proteins. There was electron dense material in a groove made up of the α1 and α2 domains which could not be accounted for by the predicted class I protein
sequence (Bjorkman, Saper et al, 1987). In the light of the above data (Townsend, Gotch et al, 1985, Townsend, Rothbard et al, 1986) this was interpreted as being due to epitopes bound in one of two conformations: either ~20 amino acid (aa) peptides bound as alpha-helices or 8-9 aa peptides bound in an extended conformation.

To study the naturally processed peptides in more detail, a preparation of the total cellular peptide pool was made from influenza infected tumour cells. This procedure relies upon the fact that most peptides are soluble in trifluoroacetic acid (TFA) while most higher molecular weight material can be removed by denaturation and precipitation. The acid soluble material was fractionated by HPLC and the fractions assayed for antigenic peptides by specific CTL. Influenza specific CTL restricted through two different class I molecules recognised different peptides from flu nucleoprotein (Rotzchke, Falk et al, 1990). These peptides turned out to be similar to the epitopes defined using synthetic peptides: an analysis of the synthetic peptides by HPLC showed that lower molecular weight by-products of the synthesis were recognised by the CTL more effectively than the main product. For both class I molecules, one of these by-products corresponded to the peptide present in the acid soluble extracts. Both epitopes were nonamers, consistent with Bjorkman's "extra" electron density as peptides bound in an extended conformation. A more refined class I crystal structure, of the B27 molecule, confirmed the presence of nonamer peptides in the $\alpha1/\alpha2$ groove (Madden, Gorga et al, 1991).

A more detailed analysis became possible with the elution of bound peptides, by acid extraction, from precipitated class I molecules. The eluted material was separated by high performance liquid
chromatography (HPLC) and the more abundant peptides were sequenced. This approach revealed that the peptides bound to a particular class I molecule are of a specific length (8 or 9 aa) and fit into consensus sequences. These consensus sequences consist of one or two well conserved "anchor" positions: the carboxy-terminal residue is hydrophobic (for example H-2-Kd, Db, Kb and HLA-A2, Falk, Rötzschke et al, 1991) or basic (for example, HLA-B27, Jardetzky, Lane et al, 1991); other anchor positions occur at different positions dependent on the class I molecule (for example, a preference for arginine at position 2 in HLA-B27, but for asparagine at position 5 in H-2-Db).

5. Processing of proteins for presentation through MHC class I

The data summarised in sections 1-4 led to a number of predictions (Townsend and Bodmer, 1989). As epitopes are presented as short peptides, but presumably synthesised as intact proteins, then there should be some protease within the cell responsible for this degradation. Also, because class I restricted epitopes are derived from cytosolically synthesised proteins there is a need for a transport system for the movement of the protein (or an antigenic derivative thereof) into the exocytic pathway. Another possibility, which would obviate the need for a protease, has been proposed - the "pepton hypothesis". This hypothesis was formulated from experiments on tum- antigens. These antigens were created by mutagenesis of tumour cell lines. The resultant mutant lines were unable to form progressive tumours when transplanted into mice because of T cell mediated rejection. Cloning of the tum- antigens was achieved by transfection of cosmid libraries from the tum- mutant cell lines. Strikingly, the tum- genes conferred antigenicity on tumour cell lines even if present on cosmids containing only 3' fragments of the
genes. It was proposed that antigenic peptides could be produced by translation of short genetic regions, rather than by degradation of the intact protein (Boon and van Pel, 1989).

6. Mutant cell lines unable to process antigen for presentation through class I

The use of mutant cell lines deficient in class I surface expression has allowed the genes involved in the processing of antigen through class I to be characterised. The human B cell line LCL721 was mutated with γ-irradiation and selected with antibody and complement for loss of the class I allele B8. The resultant cell line, LCL721.45 (.45), had lost the entire B8-containing MHC haplotype. LCL721.45 was further mutated and selected with antibodies against class I A2, producing the line LCL721.134 (.134). The cell line LCL721.174 (.174) was produced by the selection of mutagenised LCL721.45 with an anti-class II DR and DQ antibody.

Both .174 and .134 contained class I genes B5 and A2 of the remaining MHC haplotype, although .174 had a deletion over the class II region. MRNA for B5 and A2 could also be detected. However, FACS analysis showed that surface B5 levels of the mutants were not detectable and surface A2 levels were only 50-60% that of the parent .45. Fusion of the mutant cell lines with a B cell line with an intact copy of the MHC restored expression of B5 and A2 to .45 levels. Therefore, fully functional genes for B5 and A2 are present in .174 and .134 - the defect in surface expression of B5 and A2 must be post-transcriptional and trans -acting. It seemed likely that the mutations creating .174 and .134 were within the region of hemizygosity of the parent cell line .45, that is, the MHC. Also,
mapping of the deletion of .174 suggested that there was a gene or genes within the class II region that affected class I expression (DeMars, Chang et al, 1984, DeMars, Rudersdorf et al, 1985).

This interpretation was supported when .174 was fused to a T cell line CEM to create T1, which expressed high levels of A2 and B5. A derivative of T1, called T2, which had lost both copies of the CEM derived chromosome 6, had a phenotype similar to .174. This suggests that the trans-acting factor responsible for class I expression is encoded on chromosome 6, and is presumably within the homozygous deletion in the class II region of .174. Preclearing of .45 lysates with anti-β2m serum led to an 80% reduction in the subsequent precipitation of class I heavy chain. In contrast, less than 20% of class I in the .174 line was associated with β2m by the same analysis. A pulse chase experiment showed that immunoprecipitated class I was Endoglycosidase H (Endo H) resistant after 60 min for .45. In .174, Endo H sensitivity was seen after 3 hours chase. Thus class I molecules in .174 neither assemble with β2m nor reach the medial Golgi (Salter and Cresswell, 1986).

Two other mutant cell lines have been of use in defining the class I processing pathway. A mouse cell line, RMA-S, has a defect in class I surface expression. This line was derived from the virus-induced lymphoma RBL-5. Mutagenesis of RBL-5 produced the subline RMA, and RMA-S is an anti-H-2-K and D selected variant of RMA. As with T2, the H-2 heavy chains are Endo H resistant and are not coprecipitated with an anti-β2m serum (Ljunggren, Pääbo et al, 1989). The human B cell line BM36.1 has a similar phenotype to RMA-S. Class I molecules are unstable, do not reach the cell surface and are unable to present influenza proteins to CTL. BM36.1 was derived from a hemizygous derivative of
7. Assembly of class I, β2m and peptide

RMA-S and .174 were used to examine the biochemistry of class I assembly with peptide and β2m, and for their ability to present antigen to CTL. RMA-S was resistant to lysis by CTL specific for influenza nucleoprotein (NP) after infection with the virus, whereas RMA was efficiently recognised. However, treatment of the two cell lines with a synthetic peptide NP epitope led to recognition of RMA-S by CTL at levels similar to RMA. This was accompanied by an increase in surface expression of H-2-D^b, associated with β2m, in RMA-S. The restoration of D^b surface expression was peptide specific and there was a correlation between those peptides able to inhibit CTL recognition of RMA-S treated with the NP epitope and those peptides able to induce D^b surface expression. Additionally, the same peptides were able to promote the assembly of D^b with β2m, to promote the folding of D^b as assayed by a conformation sensitive antibody, and to promote Endo H resistance of D^b glycosylation. Treatment of RMA-S with brefeldin A, to prevent egress of class I from the endoplasmic reticulum (ER), prevented the peptide-induced surface expression of D^b. β2m association of the heavy chain was not inhibited, implying that this process is induced by peptide in the ER (Townsend, Ohlen et al, 1989).

In a cell-free system, peptides are also able to induce the assembly of class I heavy chain with β2m in RMA-S lysates. The effect was specific for the length and the sequence of the added peptide. The assembly and
folding of class I molecules could also be driven by the addition of high concentrations of β2m to the lysate, without the addition of peptides. In contrast to the effect of added peptides which were specific for different class I molecules, excess β2m was able to stabilise both D^b and K^b (Townsend, Elliott et al, 1990; Elliott, Cerundolo et al, 1991).

Experiments on the cell line .174 gave similar results to RMA-S. .174 was unable to present antigen from infecting viruses, but was able to present synthetic peptide epitopes when added exogenously. In fact, .174 could present exogenous peptide better than the wild type cell line 721. Surface class I expression was also induced by peptide epitope (Cerundolo, Alexander et al, 1990). Similarly, the derivative of .174, T2 (Salter and Cresswell, 1986), was unable to present endogenously synthesised antigen but could present added peptide epitope (Hosken and Bevan, 1990).

Class I molecules can be stabilised at the cell surface of RMA-S by a drop in temperature. The class I is associated with β2m but is unable to present endogenous antigens. However, RMA-S was able to present exogenous peptide epitope better at 26° C than the parent cell RMA was at 37° C. The addition of peptides to RMA-S at 26° C was able to stabilise class I when the cell line was shifted to 37° C (Ljunggren, Stam et al, 1990). Direct binding of peptide to the surface of RMA-S cultured at 26° C was demonstrated using a radio-labelled peptide epitope (Schumacher, Heemels et al, 1990).

The data summarised in this section is consistent with a defect in the supply of peptide to the ER for class I assembly in these mutant lines. MHC class I molecules accumulate in
the ER, probably by association with the chaperonin p88 (Degen and Williams, 1991 and Jackson, Cohen-Doyle et al, 1994). However, addition of exogenous peptide allows exit to the cell surface. This may occur by stabilising "empty" class I molecules which have "leaked" to the cell surface from the ER without peptide. These empty class I molecules would be unstable and be degraded so that steady state levels of class I would be low without the addition of peptide. Stabilisation of class I by a drop in temperature or addition of peptide would then allow significant amounts of class I to accumulate at the cell surface (Ljunggren, Stam et al, 1990). Presumably, the increased ability of the class I molecules of RMA-S to present exogenous peptide in this experiment reflected a lack of endogenous peptide in the class I groove.

In the case of .174 and T2, the gene or genes responsible for this defect could be mapped to the .174 deletion in the class II region of the MHC. Fusion of RMA-S (H-2^b haplotype) with a wild-type cell line (H-2^k haplotype) restored the wild-type phenotype. However, when the fusion clones were selected for loss of the H-2^k haplotype derived from the wild-type cell line, the resultant cell populations had the RMA-S phenotype of low surface H-2^b. This experiment suggested that the RMA-S defect was linked to the H-2 locus on chromosome 17 (Hosken and Bevan, 1992).

8. Towards a genetic explanation of the defect in the mutant cell lines - cloning of ATP-binding cassette (ABC) transporter genes in the MHC

Genes with homology to the ABC family of transporters have been cloned and mapped within the class II region of the MHC of human, rat
and mouse. The circumstances by which these were cloned are described below.

Antisera raised between congeneric mouse strains were used in immunoprecipitation studies. A group of low molecular weight proteins (LMPs) were precipitated by such sera (Monaco and McDevitt, 1982, Monaco and McDevitt, 1984, Monaco and McDevitt, 1986, this work is further discussed in section 12). Recombinations within the H-2 region mapped the LMP genes to between Pb and Ob, the murine equivalents of DPB and DOB in the human MHC (Hanson and Trowsdale, 1991). Cosmids over this region had previously been isolated (Steinmetz, Stephan et al, 1986) and were used to screen cDNA libraries for new genes. Two genes, called HAM1 and HAM2 (Histocompatibility antigen modifier) were isolated but did not code for the LMPs (Monaco, Cho et al, 1990). However, they did have homology to the ATP-binding cassette family of transporters. This family of genes includes members known to have a role in the transport of molecules across cell membranes. In particular, the oligopeptide permease operon of bacteria transports small peptides (Hiles, Gallagher et al, 1987).

A locus in the rat, called cim for class I modification, had been described that altered the properties of the rat class I molecule RT1.A\textsuperscript{a}. Mapping of the cim locus placed it in the class II region of the rat MHC, outside the RT1.A class I locus (Livingstone, Powis et al, 1989). The class I allelic gene product RT1.A\textsuperscript{a} differs in its reactivity to a monoclonal antibody JY3/84 depending on its association with different alleles at the cim locus. Also affected by cim are the presentation of certain minor histocompatibility antigens through RT1.A\textsuperscript{a} to CTL lines and the intracellular transport and assembly of RT1.A\textsuperscript{a}. Fast processing (cim\textsuperscript{a})
was dominant over slow processing (\textit{cim}^b). The speed of processing of another class I allelic product, RT1.A^c, was not affected by the \textit{cim} locus (Livingstone, Powis et al, 1989, Powis, Howard et al, 1991). The \textit{cim} locus was further mapped by typing a panel of F1 hybrid rats from MHC recombinant strains (Livingstone, Powis et al, 1991). Cosmids over this region detected two clones mtp1 and mtp2 (Deverson, Gow et al, 1990). These were highly homologous to HAM1 and HAM2 and prompted speculation that they were responsible for the \textit{cim} phenotype.

Finally two groups cloned homologous genes within the human MHC by screening cDNA libraries with cosmids that mapped across the region defined by the .174 deletion (Blanck and Strominger, 1988), in particular close to CpG islands (Bird, 1987). Two genes were cloned, named \textit{RING4} and \textit{RING11} by one group (Trowsdale, Hanson et al, 1990, Powis, Mockridge et al, 1991) and \textit{PSF1} and \textit{PSF2} (Peptide Supply Factor) by the other (Spies, Bresnahan et al, 1990, Colonna, Bresnahan et al, 1992). These genes have now been renamed \textit{TAPI} and \textit{TAP2}.

9. The function of the TAP gene products defined by transfection

The \textit{mtp} genes in the rat mapped within the region defined for the \textit{cim} locus. In addition, the \textit{TAP} genes in human mapped within the .174 deletion. The genes have homology to ABC transporters, and both the .174 and \textit{cim} phenotypes could be explained by a defect in the supply of peptides for class I assembly. Evidence to support the hypothesis that the TAP gene products transport antigenic peptides across the ER membrane is outlined below.
An antiserum raised against TAP1 specifically precipitated two bands of ~80 kD. One of these showed altered mobility on SDS-PAGE consistent with polymorphisms in the TAP2 gene product. Thus, the two TAP genes encode proteins that associate with each other, possibly as a heterodimeric complex (Kelly, Powis et al, 1991). This would fit with a model in which the TAPs function in a similar way to other members of the ABC family - a dimer of TAPI and TAP2 would form a complex similar in structure to the membrane proteins encoded in the opp operon. The same serum, when used for electron micrography of frozen cell sections, showed that TAPI was located in the ER and cis Golgi membrane - consistent with a role in peptide transport into the ER (Kleijmeer, Kelly et al, 1992). Direct evidence to support a role of the TAPs in antigen processing came from transfection experiments.

Anti-TAP1 precipitations from the mutant cell line BM36.1 showed that the upper band of the doublet was fainter and of a higher molecular weight than that of the parent cell line BM28.7. Sequencing of TAP2 cDNA from BM36.1 revealed a frame-shift mutation that resulted in replacement of the C-terminal 52 amino-acids of the ABC domain and an extension of the open reading frame by 51 amino acids, consistent with the immunoprecipitation result. To show that the TAP2 mutation was causative for the BM36.1 phenotype, wild-type TAP2 cDNA was transfected into BM36.1 and caused an increase in surface class I to 50% that of the parental line BM28.7. Presentation of intracellular antigen was restored to that of the parent line (Kelly, Powis et al, 1991).

Similar results were obtained in other mutants. LCL721.134 has a similar phenotype to BM36.1. In this case, transfection of TAP1 restored wild-type class I levels, class I assembly and presentation of antigen...

As the mtp genes mapped to the cim locus in the rat, it was reasoned that the cim phenotype might be caused by different mtp alleles. The mtp genes from a cim\(^a\) dominant haplotype were transfected into a cell line with a cim\(^b\) haplotype. Transfection of mtp1 cDNA had no effect on the phenotype of the recipient cell line. However, transfection of the mtp2 allele transferred to the transfected cell line a cim\(^a\)-like phenotype. RT1.A\(^{a+}\) was now expressed, as determined by serology or CTL, and was rapidly processed. Sequencing of mtp2 from several haplotypes with defined cim phenotypes revealed 29 amino acid differences between the alleles, of which 25 were correlated with cim phenotype. These were predominantly in the membrane spanning region. Peptides recovered by acid elution from RT1.A\(^{a}\) in either cim\(^a\) or cim\(^b\) backgrounds were separated by HPLC. RT1.A\(^{a+}\) had peptides eluting from HPLC as a broad peak. RT1.A\(^{a+}\), on the other hand, had a profile showing more hydrophobic, late-eluting peptides. The peptide-elution phenotype of RT1.A\(^{a-}\) could be shifted by transfection of mtp2 from the dominant cim\(^a\) haplotype (Powis, Deverson et al, 1992). Thus, the mtp2 allele had a direct affect on the peptides that were presented in the RT1.A\(^{a}\) molecule, consistent with the mtp gene products forming a peptide pump to provide the substrates for class I presentation.

The relevance of the TAPs for the whole animal has been demonstrated in a TAP1-/- knockout mouse. Spleen cells from this
mouse have a phenotype similar to the mutant cell lines described in section 8 - they are able to present exogenously added peptide to CTL, but not endogenously synthesised proteins, and have low class I surface expression. An analysis of the T cells from various organs showed that CD4/CD8 double positive cells are present at wild type levels, as are CD4 single positives. However, CD8 positive CTL T cells are not found in the spleen, blood or lymph nodes of the mutant mice. This is consistent with a role for class I in the positive selection of CD8 positive cells in the thymus (van Kaer, Ashton-Rickardt et al, 1992).

10. The function of the TAPs: biochemical assays

Reconstitution of mutant cell lines using the TAP genes and an effect of the rat TAP alleles on the peptides eluted from class I both suggest a role for the TAP complex in antigen processing through class I. The widely held view was that this role would be in the ATP-dependent transport of peptides, and direct biochemical evidence is now available for this. However, three reports concluded that peptides could be transported into microsomes without ATP and, in one case, without a need for the TAPs.

The first assay studied peptide transport across microsomes. This used a TAP deficient cell line (T2) as a control for TAP function. The assay depends on the assembly of in vitro translated class I with β2m and a biotinylated peptide epitope. A conformation-dependent antibody or streptavidin were used to precipitate the class I complex. Peptide associated with class I and induced assembly in an ATP-dependent manner. Treatment of the microsomes with proteinase K showed that this assembly was not affected by the digestion of cytoplasmic portions of
ER membrane proteins, which would probably include the ATP-binding domain of the TAPs (Kleijmeer, Kelly et al, 1992). When peptide transport across the ER membrane was measured by binding to BiP, rather than class I, it seemed that this process was independent of ATP. In addition, a comparison of microsomes from T1 and T2 revealed no difference in peptide transport by this assay, although class I assembly was lower in T2. The authors concluded that peptide transport into the ER occurred in an ATP- and TAP-independent manner. However, an ATP-dependent step in class I assembly occurred downstream of peptide transport and was affected by the T2 deletion (Levy, Gabathuler et al, 1991).

Two assays to investigate transport of peptide across the ER membrane have been developed for dog pancreas microsomes. The assay depends on the assertion that transport of a peptide containing a glycosylation signal into the microsomes would deplete the pool of dolichol high mannose oligosaccharides. Subsequent glycosylation of a marker protein in the presence of the microsome fractions would then be inhibited. Using this inhibition as a measure of peptide transport, the assay showed no effect of ATP depletion, nor was there an effect on transport by an inhibitor of ATPases, oligomycin. However, the transport was temperature sensitive - there was no peptide transport at 4°C but transport did occur at 25°C (Koppelman, Zimmerman et al, 1992).

A third assay also questioned the necessity for ATP in peptide transport. In vitro translation of class I and β2m by a rabbit reticulocyte lysate in the presence of dog pancreas microsomes could support the assembly of class I. However, the process was dependent on the presence of both peptide and oxidised glutathione (for disulphide bond
formation). An iodinated peptide containing a glycosylation site was transported into the microsomes in an ATP-independent way - glycosylation of the peptide indicating transport into the ER. In this system, class I stabilisation by peptide occurred after the microsomes had been solubilised (Bijlmakers, Neefjes et al, 1993).

There is some evidence for TAP-independent presentation of certain peptides to CTL. Infection of RMA-S with VSV, or transfection with VSV nucleoprotein gene, allowed efficient recognition by a VSV-restricted CTL (Hosken and Bevan, 1992). Whether this was due to partial function of the remaining HAM gene product, perhaps acting as a homodimer to transport peptide into the ER, or due to a TAP-independent mechanism, is unclear. The TAP deficient cell line .174 expresses HLA-A2 at ~50% wild-type levels. Peptides eluted from these A2 molecules are derived from signal peptides (Wei and Cresswell, 1992). These would be formed in the ER lumen by the action of signal sequence peptidase on secretory or membrane-bound proteins. It is unlikely that this pathway explains the ATP-independent transport of peptides in the assays described above as there is no reason to assume that the peptides used in these assays would be good substrates for the translocation machinery.

Data which is consistent with the TAPs transporting peptides into the ER has recently been described. In an assay which uses permeabilised cells, rather than microsome preparations, a difference in peptide transport was seen between T2 and T2 transfected with the rat TAP cDNAs. Transport of the peptide in the T1 permeabilised cells was ATP-dependent (Neefjes, Momburg et al, 1993). Transfection of alleles of the rat mtp2 gene has shown that differences in peptides eluted from
RT1.A\textsuperscript{a} molecules reflect mtp-mediated peptide transport into the ER (Heemels, Schumacher et al, 1993, Momburg, Roelse et al, 1994). This result suggests that the specificity of the TAP complex is restrictive for antigen presentation, at least in the rat.

Additional evidence for a role of the TAPs in peptide transport came from experiments using TAP knockout mice. Microsome preparations from wild type and TAP\textsuperscript{1-/-} mice were both able to assemble exogenously added peptide with class I. However, when the temperature of the reaction was lowered from 37° C to 23° C the rate of accumulation of peptides in the TAP\textsuperscript{1+/-} microsomes was 4 fold higher than that of TAP\textsuperscript{1-/-}. This difference in rates, presumably due to the TAP complex, was ATP-dependent (Shepherd, Schumacher et al, 1993).

The discrepancy between experiments showing ATP-dependent and ATP-independent peptide transport can probably be explained by differences in the experimental conditions. For example, peptide concentration, temperature and time of incubation are all likely to be relevant. Another factor may be the microsome preparation - if the microsomes are "leaky" for small molecules this could mask any TAP-dependent transport. Finally, the concentration of TAP complexes in dog pancreas microsomes may be too low for TAP-dependent transport to be seen.

That the TAPs are intimately involved in the transport of peptides across the ER membrane now seems reasonably certain. The only real proof will be the biochemical purification of recombinant TAPs, the incorporation of the complex into a synthetic bilayer, and the specific transport of peptides across this membrane. Until this can be done, there
will always be a possibility that the TAPs have an indirect effect on peptide transport, perhaps through the transport of another molecule or ion.

11. The degradation machinery for the production of peptides to be presented through class I - cytosolic proteolysis and the proteasome

11a. Characterisation of the proteasome

As discussed in section 3, the degradation of antigen for class I presentation was not sensitive to lysosomotropic agents (Morrison, Lukacher et al, 1986). Non-lysosomal protein degradation was first mooted after a study on the degradation of metabolically labelled proteins by pulse-chase analysis. Proteolysis of normal cellular proteins could be separated from that of abnormal proteins (containing the amino-acid analogue canavanine) by the addition of various agents to the medium. Two pathways for protein degradation were defined from this work. One pathway, sensitive to cathepsin B inhibitors, was lysosomal and showed little specificity for protein substrate. The other pathway was non-lysosomal and preferentially degraded proteins that denatured easily, such as those containing canavanine (Knowles and Ballard, 1976).

A non-lysosomal cellular protease was further characterised by several investigators working in different fields. One experiment followed the degradation of abnormal haemoglobin in reticulocyte lysates. The protein was rapidly degraded by a process that was sensitive to manipulations that depleted the ATP pool. The proteolytic activity could be assayed in a cell-free system and was activated by ATP but not non-hydrolysable analogues. The optimal pH for proteolysis was
7.8, ruling out any role for the lysosomal acid proteases in this process (Etlinger and Goldberg, 1977). ATP-dependent proteolytic systems were then described in rat and mouse liver. In both cases the pH optimum was slightly alkaline. Fractionation of either rat or mouse homogenates showed that the protease had a molecular weight of ~550 kD (DeMartino and Goldberg, 1979, Rose, Warms et al, 1979).

The enzyme was also isolated from pituitary as the complex that was active in the production of opioid peptides from precursors. The complex had a molecular weight of ~700 kD but could be dissociated into subunits of 24-28 kD. It had three proteolytic activities that could be defined using model tripeptide substrates coupled to chromogenic C-terminal leaving groups. Cleavage occurred at the C terminus of hydrophobic residues, basic residues and glutamate. These activities were termed chymotrypsin-like, trypsin-like and peptidyl-glutamyl peptide (PGP) respectively. Treatment of the complex with various protease inhibitors, or with SDS, produced differential inhibition/activation of the three activities, showing that they were distinct and that the enzyme was multicatalytic (Orlowski and Wilk, 1981).

Research into the degradation of α and β crystallin led to the isolation of a multicatalytic protease from bovine lens that was similar to that purified from pituitary (Ray and Harris, 1985). The enzyme was probably responsible for the degradation of α-crystallin at neutral pH by partially purified extracts of bovine lens (Blow, van Heyningen et al, 1975). The isolation of multicatalytic proteases from both pituitary and lens suggested that the enzyme might have a widespread tissue distribution (Ray and Harris, 1985). This prediction was validated by the
subsequent isolation of large proteolytic complexes with neutral or alkaline pH optima from rat liver (Rivett, 1985) and muscle (Dahlmann, Kuehn et al, 1985) and human lung (Zolfaghari, Baker et al, 1987) and kidney (Zolfaghari, Baker et al, 1987). The enzymes had similar properties to the ATP-dependent proteolytic systems isolated earlier (DeMartino and Goldberg, 1979, Rose, Warms et al, 1979).

The rat liver enzyme was then re-isolated as one of three enzymes that was able to degrade oxidised glutamine synthetase. Similar to the bovine enzyme, the molecular weight of this complex was 650 kD but it could be dissociated into subunits of 22-34 kD. Proteolysis of oxidised glutamine synthetase occurred without the detection of degradation intermediates. However, when the oxidised B chain of insulin (Insulin B\textsubscript{OX}) was used as a substrate, the products could be separated and characterised by HPLC. Cleavage of this 30 amino acid peptide occurred at five positions, which were different from those seen with other types of protease (lysosomal cathepsins, calpain I and II, papain, insulin protease, metalloproteases, serine proteases, chymotrypsin, cathepsin c and elastase) (Rivett, 1985). Subsequent analysis showed that both enolase and pyruvate kinase were also better substrates when oxidised than the native enzymes. The data suggest that this protease may be involved in the degradation of abnormal proteins that was described by Knowles and Ballard (1976).

The metabolism of vasoactive peptides was another area of research that resulted in the isolation of a multicatalytic protease. A neutral protease had been isolated from rabbit brain by its ability to inactivate bradykinin (Camargo, Shapanka et al, 1973). An enzyme with matching activity from human lung had properties that were broadly the
same as the rat neutral peptidase: high molecular weight when undenatured (650 kD); dissociation into lower molecular weight subunits and inhibition by thiol binding reagents (Zolfaghari, Baker et al, 1987). The same enzyme was purified from human kidney. In this case a demonstration of the multicatalytic nature of the human enzyme was made. Cleavage of the three model substrates for trypsin, chymotrypsin and PGP activities had different pH optima and were affected differentially by treatment of the enzyme with SDS, protease inhibitors, cations and albumin (Zolfaghari, Baker et al, 1987).

To summarise, many diverse fields contributed to the characterisation of a multicatalytic protease from several tissue types. The degradation of lens proteins in cataract research, vasoactive and opioid peptide metabolism and ATP-dependent protein degradation all involved a proteolytic complex with high molecular weight that had a neutral pH optimum.

11b. Non-proteolytic activities of the proteasome

A complex with similar physical properties to the multicatalytic protease had been assigned several activities other than proteolysis. For example, a complex with a sedimentation coefficient of 19S on sucrose density fractionation co-purified with aminoacyl transferase I activity (Shelton, Kuff et al, 1970). The complex was also implicated in the repression of mRNA translation - a 19S particle made up of subunits of 20-30 kD and small cytoplasmic RNA (ScRNA) was found associated with globin and other "repressed" mRNAs in HeLa cells. The complex was named "prosome" (Schmid, Akhayat et al, 1984). Finally, an enzyme activity that processes the 5' leader of the primary transcript of a tRNA
molecule was isolated from *Xenopus laevis* ovaries. The activity copurified with a prosome-like particle (Castaño, Ornberg et al, 1986).

The complexes that repressed mRNA translation were identical to the multicatalytic protease. The *Drosophila* 19S particle appeared similar to the rat multicatalytic protease complex on electron micrographs. The two complexes shared epitopes defined by cross-reacting antibodies, contained RNA molecules of 80-100 nt and both had proteolytic activity (Falkenburg, Haass et al, 1988). A similar analysis compared the 19S particle from HeLa cells with the multicatalytic protease isolated from either human red blood cells or rat liver and reached the same conclusion (Arrigo, Tanaka et al, 1988). The multicatalytic protease, as purified from rabbit skeletal muscle, inhibited translation of mRNA in a rabbit reticulocyte system (Kuehn, Dahlmann et al, 1990). This effect was not specific for the mRNA, although a specific effect of prosome on the repression of viral, but not cellular, mRNA has also been shown (Horsch, Martins de Sa et al, 1989). The name "proteasome" was suggested for the prosome/multicatalytic protease complex. This name is now popular in the literature and will be used throughout this thesis.

11c. Mechanism of proteolysis by the proteasome

Many groups have investigated the catalytic action of the proteasome using model oligopeptide substrates and protease inhibitors. The results from these studies have not been entirely consistent, and the proteasome does not behave like any of the well-defined groups of proteases. For example, the susceptibility of the rat liver enzyme to thiol blocking reagents and resistance to DFP and PMSF suggested a mechanism similar to the cysteine proteases (Rivett, 1985). Conversely,
each of the trypsin, chymotrypsin and peptidylglutamyl sites of bovine pituitary proteasome were sensitive to isocoumarin derivatives (which react with the active site of serine proteases), suggesting that the proteasome belonged to the class of serine proteases (Orlowski and Michaud, 1989).

A more recent study on proteasome catalysis used a wider range of peptide substrates and serine-protease inhibitors. 3,4-dichloroisocoumarin (DCI) inhibited activity with high concentrations of a PGP substrate but stimulated activity with low concentrations, suggesting two sites able to digest these substrates. Additionally, activity against substrates with either phenylalanine or tyrosine at the "P1" position could be distinguished by the action of chymostatin analogues, guanidine/HCl or casein. This provided evidence for two chymotrypsin-like sites. Taking into account the trypsin-like site, the authors of this study concluded that the proteasome may have five active sites (Djaballah, Harness et al, 1992).

The digestion of longer peptides and proteins, which are perhaps more physiological substrates, has allowed the proteasome to be characterised in more detail. Evidence for further active sites came when it was noted that acetylation of the bovine pituitary enzyme with N-acetylimidazole (NAI) inhibited cleavage of peptide substrates, but stimulated hydrolysis of casein to a 21 kD intermediate. Digestion of the 21 kD protein could be achieved with the native (unacetylated) enzyme, but not with the acetylated enzyme. Thus, the proteasome may act sequentially - the acetylation-sensitive site would catalyse the initial cleavage and the subsequent proteolysis would be achieved by the trypsin-like, chymotrypsin-like and PGP sites (Yu, Pereira et al, 1991).
The same group showed that the caseinolytic activity of the proteasome was also increased by DCI at concentrations which inhibited the activity against the chromogenic peptides. In contrast to treatment with NAI, DCI-treated proteasome did not digest casein to a detectable intermediate. In fact, the 21 kD casein intermediate resistant to degradation by NAI-proteasome could be digested by DCI-proteasome (Pereira, Nguyen et al, 1992). The DCI-resistant component was investigated in more detail using neurotensin and proinsulin as substrates. Proinsulin produced similar degradation products when treated with either native or DCI-proteasome, suggesting that proinsulin is degraded by the DCI-resistant component of proteasome. Neurotensin, a 13 amino acid peptide, showed different cleavage patterns after treatment with DCI-proteasome when compared to the native enzyme. The DCI-resistant activity seemed to favour cleavage after branched amino acid side chains (Cardozo, Vinitsky et al, 1992).

One product of cleavage of insulin \( B_{\text{ox}} \) required that the enzyme cleaved twice within the substrate. Addition of potential intermediates in this reaction, that is the products after either of the sites alone had been cleaved, did not lead to rapid production of the final product. It was suggested that the proteasome may produce this peptide by a "bread slicer" type mechanism whereby the two cleavages would be made without any intermediate leaving the enzyme complex (Dick, Moomaw et al, 1991).

The work described in this section illustrates that the proteasome has at least three and possibly greater than five definable active sites. The nature of each of these is unclear as none of the inhibitor profiles fits any
of the well characterised protease families. Additionally, there is no/little homology in any of the primary sequences of proteasome components to other protease sequences. An explanation of the chemistry of proteasome activity will require a detailed knowledge of those residues that are part of the active sites.

Two different approaches have come close to defining the active sites. Treatment of bovine pituitary proteasome with N-ethylmaleimide (NEM) inhibited the trypsin-like activity. The enzyme was protected from NEM in the presence of leupeptin. A cysteine residue that was labelled by NEM only in the absence of leupeptin was identified in subunit 13. Presumably, this residue is part of the active site for the trypsin-like activity as leupeptin is a competitive inhibitor of this activity (Dick, Moomaw et al, 1992). The other approach takes advantage of the conservation of the proteasome in yeast. The yeast (Saccharomyces cerevisiae) proteasome, called YscE, was discovered in a strain deficient in the major vacuolar proteases. To isolate mutants in the proteasome, this yeast strain was further mutated and the colonies screened for loss of chymotryptic activity using a chromogenic peptide. These colonies fell into two complementation groups. The defective genes were cloned by transfection of a yeast genomic library and had homology to proteasome components. The phenotype of the mutant colonies was such that only the chymotrypsin-like activity of YscE was affected. Therefore, the mutated residues in the yeast colonies from this experiment are likely to be involved in the chymotrypsin active site (Heinemeyer, Kleinschmidt et al, 1991, Heinemeyer, Gruhler et al, 1993). Another yeast proteasome subunit has been cloned by a similar protocol but with selection for loss of the PGP hydrolysing activity (Hilt, Enenkel et al, 1993).
11d. Ubiquitin and the 26S complex

Further components necessary for the ATP-dependent degradation pathway were characterised by fractionation. First, a small protein, called APF-1 for ATP-dependent proteolysis factor, was isolated from reticulocyte lysates. This had no intrinsic proteolytic activity, but stimulated ATP-dependent proteolysis by another fraction of the lysate. APF-1 became covalently bound to proteins in an ATP-dependent manner and it was suggested that these covalent conjugates could be intermediates in the proteolysis reaction (Ciechanover, Heller et al, 1980). APF-1 was subsequently found to be identical to ubiquitin on the basis of physical properties, amino acid composition and by functional complementation: ubiquitin could substitute for APF-1 in stimulating ATP-dependent proteolysis and could also form covalent conjugates with reticulocyte proteins (Wilkinson, Urban et al, 1980).

That ubiquitin-conjugation was a prerequisite for degradation in this pathway seemed more likely when an enzyme complex was discovered that was able to degrade ubiquitinated proteins, but not non-ubiquitinated proteins. Using ubiquitin-lysozyme conjugates as a substrate, a high molecular weight ATP-dependent protease was purified from rabbit reticulocytes (Hough, Pratt et al, 1986). The pH optimum for the enzyme was 7.8, the same as that for the whole lysate (Etlinger and Goldberg, 1977). The sedimentation coefficient of the enzyme was 26S. Further purification of this enzyme showed that it was made up of components of 34 to 100 kD and 21-32 kD. The 21-32 kD components of the 26S complex comigrated with the subunits of the 20S proteasome complex (Hough, Pratt et al, 1987).
A reticulocyte lysate that had been depleted for ATP was fractionated and the fractions were assayed for proteolytic activity. Three necessary components were found. These were named CF1, CF2 and CF3 and all were necessary in the degradation of ubiquitin-lysozyme conjugates. The kinetics of this degradation showed that there was a time lag which could be abolished if the three factors were preincubated with ATP before addition of the substrate. During this preincubation, CF1, 2 and 3 assembled to form a large complex with similar properties to the 26S complex (Ganoth, Leshinsky et al, 1988).

The relationship of the 20S complex, which does not degrade ubiquitin conjugates, to the 26S complex, which does, has been somewhat controversial. Although there appeared to be low molecular weight components of the 20S complex in the 26S fraction (Hough, Pratt et al, 1987), there are two reports that these proteins did not react with antibodies against the 20S subunits (Seelig, Kloetzel et al, 1991, Kuehn, Dahlmann et al, 1992).

On the other hand several approaches have concluded that the 20S complex is incorporated into the 26S complex in a process which is dependent on ATP. The 20S complex could substitute for the CF3 fraction in the assembly of a 26S complex which was active in the degradation of ubiquitinated substrates (Eytan, Ganoth et al, 1989, Driscoll and Goldberg, 1990). This was directly shown by prelabelling the 20S complex with fluorescein. The incorporation of the labelled 20S enzyme into a 26S complex was coincident with peptide cleaving activity of the heavier complex (Hoffman, Pratt et
al, 1992). The same study also showed that the 26S complex was reactive with a serum raised against the 20S complex. Orino et al have shown that a monoclonal antibody against a 20S component coprecipitated higher molecular weight proteins similar to those found in the 26S complex (Orino, Tanaka et al, 1991). Finally, mutations in 20S proteasome components in yeast have a phenotype that includes increased stability of ubiquitinated proteins (Richter-Ruoff, Heinemeyer et al, 1992, Seufert and Jentsch, 1992).

Whether ubiquitination is a prerequisite for the degradation of proteins for presentation through class I is unknown as yet. However, two experiments suggest that ubiquitination is necessary for the presentation of at least some antigens. The first of these studied the presentation of epitopes contained within genes in vaccinia virus constructs. Infection by vaccinia virus inhibits the presentation of certain epitopes in the late phase of infection. The inhibition was overcome by manipulations of the inserted gene that led to a more rapid ubiquitin-dependent degradation of the encoded antigen. Deletion of the signal sequence from haemagglutinin created a short lived cytosolic protein and completely restored presentation of a haemagglutinin epitope. For influenza nucleoprotein as the antigen, use was made of the "N-end rule" of ubiquitin-mediated protein degradation (Bachmair, Finley et al, 1986). A construct of Ub-Met-NP was not recognised any more efficiently than NP alone when expressed from the late promoter in vaccinia constructs. However, the less stable Ub-Arg-NP protein, if expressed from a similar construct, was presented well (Townsend, Bastin et al, 1988).

The second experiment depended on a temperature sensitive mutant in the ubiquitin pathway. The phenotype of the mutant cell
line was a decrease in the conjugation of ubiquitin to proteins and in protein degradation. Sensitisation of cells for CTL lysis was achieved by the introduction of exogenous ovalbumin by pinosome lysis. Presentation of ovalbumin through the mouse class I molecule H-2-K\textsuperscript{b} was greatly reduced at the non-permissive temperature. A possible criticism of this experiment is that the non-permissive temperature, and the associated build up of malfolded and ubiquitinated proteins, may prevent presentation through some non-specific affect. However, presentation of the peptide epitope, when encoded by a minigene on a vaccinia construct, was not affected by the ts mutation (Michalek, Grant et al, 1993).

Although this experiment added to the evidence that the ubiquitin pathway was necessary for antigen presentation, it was similar to a previously published experiment that had come to the opposite conclusion. Methylation of ovalbumin, which inhibits ubiquitin dependent degradation \textit{in vitro}, prior to pinosome-mediated uptake had no effect on presentation (Carbone, Hosken et al, 1989). The epitope in this experiment mapped to aa 258-276 and was presented through H-2-K\textsuperscript{b}. Similarly, the ovalbumin epitope studied in the ts ubiquitin mutant was H-2-K\textsuperscript{b}-restricted and mapped to aa 257-264. Therefore, inhibition of ubiquitination either by methylation or by a ts mutation in E1 caused different effects on the presentation of the same or similar epitope.

A possible explanation for this would be as follows. The ts mutation may be preventing presentation through an indirect effect on the ubiquitination of a protein other than ovalbumin. In the methylation protocol, only the ubiquitination of exogenous
ovalbumin should be affected, whereas the ts mutation would be expected to cause a general inhibition of the ubiquitin pathway. When analysed in this way, the methylation experiment is easier to interpret as it can be concluded that ubiquitination of the antigen is not necessary for presentation, whereas the ts mutation can only lead to the conclusion that ubiquitination of some protein(s) is necessary for presentation, but provides no information as to what that protein(s) might be. A more trivial explanation would be that the ts mutant line is from a hamster background transfected with mouse class I and ICAM, whereas the earlier experiment considered class I presentation in an isogenic background. Therefore, differences between hamster and mouse in other molecules involved in processing or presentation of antigen may explain the discrepancies between the two experiments.

If ubiquitination is a marker for cytosolic degradation then those proteins which are ubiquitinated will be a source of antigentic peptides. The mechanics of the ubiquitination process have been delineated in some detail. Three enzymes, E1, 2 and 3, were purified from a ubiquitin affinity column. All were necessary for the conjugation of ubiquitin to lysozyme. In the first step a high energy thiol ester bond between the C-terminal glycine of ubiquitin and E1 is formed with the concomitant production of PPI from ATP (Haas, Warms et al, 1982). The E1-ubiquitin molecule can donate ubiquitin to an E2 molecule to form covalent E2-ubiquitin adducts. Finally, mixing of these adducts with E3 and a protein substrate (lysozyme) produced stable ubiquitin-substrate conjugates (Hershko, Heller et al, 1983). In vitro experiments have shown that ubiquitination of histones could proceed directly from certain E2 enzymes in the
absence of E3 (Haas, Bright et al, 1988). The significance of ubiquitination pathways that are not dependent on an E3 molecule is unclear.

Specificity of ubiquitination is likely to be due both to the substrate itself and to the expression of different types of ubiquitinating enzymes in the cell. There is evidence for diversity and possible specialisation in each of the E1, 2 and 3 enzymes. Three different E1 enzymes have been isolated from wheat (Hatfield and Vierstra, 1989) and a testis-specific E1 encoding gene has been cloned with a role in spermatogenesis (Kay, Ashworth et al, 1991, Mitchell, Woods et al, 1991).

More diversity has been noted in the E2 enzymes. E2 enzymes differed in their ability to ubiquitinate histones (Haas, Bright et al, 1988). There are at least 10 E2 enzymes, called UBC proteins (Jentsch, 1992). All these have a conserved domain surrounding the cysteine residue that binds to ubiquitin. Some have C-terminal extensions that are thought to mediate substrate specificity. The UBC family have diverse functions, for example UBC2 was first characterised as RAD6, a protein involved in DNA repair. UBC6 has been implicated in the degradation of ER membrane proteins (Sommer and Jentsch, 1993). The yeast transcriptional regulator MATα2 is ubiquitinated (Hochstrasser, Ellison et al, 1991) and is degraded through at least two pathways which need either UBC4 and 5 or UBC6 and 7 (Chen, Johnson et al, 1993).
Two E3 enzymes have been isolated from rabbit reticulocyte lysates with different substrate specificity (Heller and Hershko, 1990) and it is possible that further E3s will be discovered. There are likely to be many pathways of ubiquitination, which may overlap in their specificity. There is some scope for control of degradation of sets of proteins by the differential expression of combinations of the ubiquitinating enzyme isoforms.

A complementary level of control could act at the level of the substrate. Natural substrates of ubiquitin degradation have been few in number. However, an important signal for degradation has been delineated using ubiquitin-\(\alpha\)galactosidase constructs. Ubiquitin is cleaved from the hybrid protein by ubiquitin C-terminal hydrolases. The N-terminal residue of \(\alpha\)galactosidase then determined the half-life and this was termed the "N-end rule" (Bachmair, Finley et al, 1986). A second determinant of the half life was the presence of a lysine residue, in the correct context, near the N-terminal end of the protein (Bachmair and Varshavsky, 1989). This residue acts as an acceptor for the ubiquitin chain (Chau, Tobias et al, 1989).

Another sequence-specific marker for ubiquitination and destruction of a substrate has been defined by a variety of fusion constructs between cyclin and protein A. This work showed that the ubiquitin-mediated destruction of cyclin B in the metaphase stage of cell cycle in Xenopus eggs was dependent on a "destruction box" at the N terminus (Glotzer, Murray et al, 1991). However, this sequence is not conserved in other natural substrates for ubiquitination. For example, neither plant phytochrome nor MAT\(\alpha\)2 contain this
Deletional mutagenesis of MATα2 showed that it contained two degradation signals (Hochstrasser and Varshavsky, 1990). One of these signals was not effective in a strain mutated for UBC6, although the other signal was not affected by this mutation (Chen, Johnson et al, 1993). It is possible that different ubiquitinating enzymes, for example within the E2 series, have different recognition moieties. This recognition may be affected by modifications to the primary amino acid sequence: although degradation of Mos in maturing Xenopus oocytes is dependent largely on the penultimate amino acid residue, phosphorylation of a neighbouring serine residue may give some stabilisation (Nishizawa, Okazaki et al, 1992).

There is a large amount of evidence that denatured or abnormal proteins are good substrates for ubiquitination. Proteins synthesised in the presence of an amino acid analogue are preferentially conjugated to ubiquitin (Hershko, Eytan et al, 1982). Similarly, denaturation of haemoglobin by phenylhydrazine led to increased ubiquitination and degradation when the protein was microinjected into cells (Chin, Kuehl et al, 1982). One experiment suggests that ubiquitin can increase the susceptibility of a protein to proteolysis \textit{in vitro} without covalent conjugation to the substrate, possibly by exerting some chaotropic effect (Wenzel and Baumeister, 1993).

In summary, there are several factors that determine the likelihood of a given protein being degraded in the cytoplasm by the
ubiquitin pathway. The expression of the various ubiquitinating enzymes (E1, 2 and 3), the presence of primary sequence determinants, post-translational modifications and the protein conformation could all have an effect. Which of these are relevant for the degradation of viral proteins for class I presentation is unknown at present. Of possible relevance to this argument is the discovery of a ubiquitin homologue which is upregulated by IFN-α, β and, to a lesser extent, γ (Loeb and Haas, 1992). It should also be pointed out that there is evidence for the degradation of cytoplasmic proteins without prior ubiquitination (Murakami, Matsufuji et al., 1992). The importance of ubiquitin-independent degradation pathways to antigen processing is also unknown.

11e. Regulation of the proteasome

Since the discovery of the proteasome, almost every published study on its multicatalytic nature has looked at the effect of a variety of inorganic agents and inhibitors of "classical-proteases" on the activities. While these studies have added weight to the idea of multiple active sites within the complex, it is unclear if the compounds used have any parallel in proteasome regulation in vivo. A number of reports have concluded that the proteasome can be isolated in a latent form. This form can be activated by treatments which include low concentrations of detergent, heat and dialysis against water (McGuire, McCullough et al., 1989, Mykles and Haire, 1991). A structural difference in the subunit composition between latent and active proteasomes has been noted (Weitman and Etlinger, 1992) but it is unknown how this change occurs, or whether it is a cause or an effect of activation.
The vacuolar or lysosomal proteases are enclosed by a membrane. Control of proteolysis can then be achieved by the transport of condemned proteins into vesicles. The proteasome is generally thought to be "free" in the cytoplasm. However, proteasomes have been purified with membranes and organellar fractions, association with the cytoskeleton has been described (reviewed in Rivett and Knecht, 1993) and certain proteasome subunits have a conserved basic motif which may mediate nuclear translocation of the complex (Tanaka, Yoshimura et al, 1990). Changes in the localisation of proteasomes have been shown to occur during the cell cycle (eg Amsterdam, Pitzer et al, 1993) and cell differentiation (eg Grossi de Sa, Martins de Sa et al, 1988, Shimbara, Orino et al, 1992). If differential localisation of proteasome occurs, then the targeting of substrates or proteasomes to specific cellular locations may be one mechanism by which cytosolic degradation is regulated.

The absolute level of proteasome within a cell appears to be regulated in some circumstances. Proteasomes were 10 fold more abundant in lymphocytic leukaemias than normal lymphocytes and were also observed at increased levels following activation of resting lymphocytes with phytohaemagglutinin (Kumatori, Tanaka et al, 1990). The level of proteasomes decreases in lymphocytic cells following differentiation (Shimbara, Orino et al, 1992). Activation of T cells with mitogen led to an increase in the transcription and translation of some subunits without any change in the steady state levels of these subunits (Shimbara, Orino et al, 1992). This suggests that synthesis of subunits can be regulated to compensate for a higher rate of proteasome turnover. Regulation of total proteasome amounts within the cell may then reflect the burden of protein degradation.
Immunoprecipitation studies have shown heterogeneity within the proteasome population, especially with reference to the LMP subunits (Brown, Driscoll et al, 1993, Patel, Monaco et al, 1994). Proteasome composition can be regulated by the cytokine interferon-γ, which up-regulates the LMPs at the expense of other subunits (Yang, Waters et al, 1992, Aki, Shimbara et al, 1994 and Belich, Glynne et al, submitted). The LMP containing and LMP lacking proteasomes have different activities against peptide substrates (Driscoll, Brown et al, 1993, Gaczynska, Rock et al, 1993). The LMPs may be unusual in that they have counterparts elsewhere in the genome (Belich, Glynne et al, submitted, and see below) whereas the other components appear to be single-copy. It is not clear whether proteasome composition is altered in differentiation, cell cycle or otherwise. However, at least for the LMPs, it seems clear that proteasome composition can be altered by an external signal (for example, IFN-γ).

It is likely that the marking of particular proteins for destruction with a ubiquitin tag is a major factor in the control of cytoplasmic protein degradation. However, a number of endogenous factors have been isolated that may be activators or inhibitors of the proteasome in vivo. Potential inhibitors of the proteasome include a hexamer of a 40 kD subunit (Murakami and Etlinger, 1986); a tetramer of 50 kD subunits (Li, Gu et al, 1991) and multimers of a 31 kD protein (Chu-Ping, Slaughter et al, 1992).

The 240 kD hexamer was subsequently shown to be equivalent to the CF2 component. While this fraction inhibited proteolysis of 125I-lysozyme or Suc-LLVY-MCA, it was essential for the degradation of ubiquitin-conjugated lysozyme (Driscoll, Frydman et al, 1992). The 40 kD
subunits appear to be ubiquitinated when incorporated into the 26S complex (Li and Etlinger, 1992). It has been suggested that ubiquitination of the CF2 subunits regulates the assembly of CF2 with CF1 and CF3 (20S proteasome) (Li and Etlinger, 1992) and CF2 is essential for breakdown of ubiquitin conjugates (Ganoth, Leshinsky et al, 1988). Therefore, the breakdown of ubiquitin conjugates may be partly regulated through the ubiquitination of the CF2 subunit.

An 11S activator of the proteasome has been isolated from rabbit reticulocytes. This is made up of 30 kDa subunits which reversibly associate with the 20S proteasome and activate the LLVY cleaving activity up to 60 fold. Cleavage of proteins, either free or as ubiquitin conjugates, was not affected (Dubiel, Pratt et al, 1992). A molecule with similar physical properties (native Mr ~180 kD, subunit Mr ~28 kD) also activated 20S proteasome cleavage of peptide, but not protein, substrates (Ma, Slaughter et al, 1992). The isolation of this activator in an inactive form raises the possibility that further levels of regulation may occur (Ma, Willy et al, 1993). Finally, DeMartino's group have isolated a 700 kD activator, PA700, which associates with the 20S complex in an ATP-dependent manner and which may be equivalent to CF1. PA700 is itself made up of a number of different subunits (Ma, Vu et al, 1994). They concluded that three forms of the proteasome exist in the cell: unassociated 20S complexes, which can be activated by exogenous PA28; 20S and endogenous PA28 complexes and a 1.75 MDa complex that probably contains PA700 associated with the 20S complex. The relationship of these forms to the 20S and 26S complexes described previously and particularly to the slow and fast migrating forms of the proteasome described by Hoffman et al is still to be described (Hoffman, Pratt et al, 1992).
There are therefore several potential mechanisms for proteasome regulation. Changes in subcellular location of the complex, expression patterns of the subunits or levels of activators or inhibitors would all be predicted to affect the quality or quantity of proteins that are degraded by this pathway. Removal or inactivation of an inhibitor may be the equivalent of the latent to active change seen \textit{in vitro}. These changes are likely to be relatively non-specific. For example, an increase in the expression of a proteasome activator may be appropriate under conditions of stress or heat shock when total cellular protein turnover would be expected to increase. Degradation of particular proteins is more likely to be regulated by differential ubiquitination.

Antigen presenting cells may use one or more of the mechanisms described above to bias proteasome degradation towards viral proteins. Alternatively, it is possible that the peptides presented at the cell surface represent a random sampling of the proteins present within the cytoplasm. This may be sufficient to elicit a T cell response as these are thought to be sensitive to low amounts of peptide-MHC complexes at the cell surface. However, the suppression of presentation of self-peptides on viral infection may be advantageous to prevent an autoimmune reaction.
11f. Structure of the proteasome

In 1970, particles from animal and plant cells were described which had a sedimentation coefficient of 19S. Electron microscopy showed that these were rectangular in shape and had dimensions of 10 x 14.5 nm (Shelton, Kuff et al, 1970). The structure of the proteasome is now being resolved in greater detail. Studies on the complex isolated from the archaeabacterium, Thermoplasma acidophilum, have been particularly informative. This complex has only two subunits, called α and β (Zwickl, Lottspeich et al, 1991, Zwickl, Grziwa et al, 1992) and expression of these subunits as recombinant proteins in E. coli is sufficient for the assembly of functional proteasomes (Zwickl, Lottspeich et al, 1992).

The three dimensional structure of the complex was derived from electron microscopy using random conical tilting (Hegerl, Pfeifer et al, 1991). In two dimensions, two projections were seen. One was a ring shape with a diameter of 11 nm, the other a rectangle measuring 15 x 11 nm. 3D reconstruction showed that the complex was a cylinder made up of four stacked rings. The outer subunits, which were made up of α subunits (Grziwa, Baumeister et al, 1991), appeared to be disc shaped. The inner subunits, made up of β subunits (Grziwa, Baumeister et al, 1991), appeared as rings. This structure was consistent with electron micrographs of proteasome crystals (Pühler, Weinkauf et al, 1992). In this case, the stoichiometry, molecular weight and symmetry could be discerned. A model for the T. acidophilum proteasome was proposed as α7β7β7α7. A preliminary report of X-ray diffraction data from the archaeabacterium proteasome has been published which confirms the seven fold symmetry (Jap, Pühler et al, 1993).
The eukaryotic proteasome is made up of more subunits (probably 12-14 per particle) and this increased complexity has hindered a detailed explanation of the structure. Additionally, RNA has sometimes been found associated with the proteasome when isolated from eukaryotic organisms. A RNA molecule of 80 nucleotides has been copurified with proteasome (Skilton, Eperon et al, 1991, Coux, Nothwang et al, 1992) and this has been identified as tRNA\textsuperscript{lys,3}. The significance of this is unknown and the RNA is present in very low quantities (less than 1 RNA to 1000 proteasome particles) (Skilton, Eperon et al, 1991). It is possible that RNA associated with proteasome particles is a substrate for translation repression (Schmid, Akhayat et al, 1984). Alternatively, there may be a role for associated tRNA molecules in the transfer of amino acids to the N terminus of protein substrates to alter their susceptibility to degradation via the N end rule pathway (Varshavsky, 1992). Copurification of a 19S particle with amino-acyl transferase activity may be of relevance to this argument (Shelton, Kuff et al, 1970).

The structure of the proteinaceous part of the proteasome has been studied from a range of tissues and organisms. The rat muscle complex has been described as a 11 x 16 nm cylinder made up of four stacked rings. At a resolution of 2-3 nm this was, like the \textit{T. acidophilum} proteasome, closed at the end but with a hollow interior (Kopp, Steiner et al, 1986). By analysis of space constraints, it was proposed that the rings each contained 6 subunits. A different conclusion was reached by Tanaka et al (Tanaka, Yoshimura et al, 1988). The X-ray diffraction pattern obtained in this report was interpreted as being due to prolate ellipsoid particles, although it was also consistent with the model
proposed by Pühler et al for the T. acidophilum proteasome (Pühler, Weinkauf et al, 1992).

The symmetry of the end-on projection of the human proteasome is still controversial. Six and 8 fold symmetry have been proposed (Baumeister, Dahlmann et al, 1988, Kopp, Steiner et al, 1986, Arrigo, Tanaka et al, 1988). Conversely, because of the high degree of conservation of the proteasome throughout evolution, it might be expected that the eukaryotic proteasome would have a similar architecture to the archaebacterium proteasome, that is, seven fold symmetry. The importance of this argument is that the number of subunits per particle is likely to be four times the symmetry number (as a reflection of the four rings or discs). The yeast proteasome has fourteen subunits when analysed by 2D electrophoresis (Heinemeyer, Kleinschmidt et al, 1991) and there is some evidence that each subunit is found twice in each human proteasome particle (Kopp, Dahlmann et al, 1993). Taking this data into account it seems likely that the eukaryotic proteasome will have pseudo-sevenfold symmetry with a "complex dimer" structure. A model for the human proteasome structure is shown in figure 1.1.

A structure for the 26S complex has been described in which the 20S proteasome cylinder is attached at one or both ends to a "ball" component to form either a "mushroom" or "barbell" shaped structure (Ikai, Nishigai et al, 1991). The ball component is likely to be equivalent to a factor isolated by Hoffman et al which can combine with the 20S complex to form a 26S complex (Hoffman, Pratt et al, 1992), and may be equivalent to CF1/CF2 (Ganoth, Leshinsky et al, 1988).
Figure 1.1: A model for the structure of the eukaryotic proteasome, based on the structure of the *Thermoplasma acidophilum* proteasome (Grziwa, Baumeister et al, 1991, Hegerl, Pfeifer et al, 1991, Pühler, Weinkauf et al, 1992, Jap, Pühler et al, 1993). The left hand figure shows the assembled complex. The right-hand figure is an expanded view of the complex showing four rings of seven subunits each. The outer two rings are shown as α subunits (solid boundaries) and the inner two rings are shown as β subunits (hatched boundaries). An antiparallel ordering of the subunits in the two αβ7 halves of the complex is shown and each subunit is present twice per complex (Kopp, Dahlmann et al, 1993, Peters, Cejka et al, 1993). The pseudo-seven fold symmetry is based on analogy with the *Thermoplasma* complex and the sequence analysis which shows fourteen sub-families of subunits (Figure 1.2). Also, the yeast proteasome has 14 subunits when analysed by 2D SDS PAGE (Heinemeyer, Kleinschmidt et al, 1991). A pseudo-helical arrangement of the subunits from one ring to another is shown, based on the conclusions of Djaballah et al (Djaballah, Rowe et al, 1993). The alpha subunits may form a disc rather than a ring (Kopp, Dahlmann et al, 1986, Grziwa, Baumeister et al, 1991). The black line running through the centre of the discs represents the axis of pseudo-rotational symmetry.
Peters et al have shown that the 26S complex of Drosophila is made up of two 19S particles attached in a *trans* configuration at either end of the 20S complex (Peters, Cejka et al, 1993). They interpret this to be consistent with Kopp's data: the 20S complex has a C2 symmetry in which the arrangement of the subunits in each of the α and β rings is reversed in the two halves of the complex. An attempt has been made to incorporate such an arrangement into figure 1.1. Additionally, the low variation between images is consistent with a highly specific interaction between the two 19S particles and the α rings. The argument is extended to include the β subunits as no rotation between the 19S particles is observed between complexes - they are always observed in a *trans* configuration. Thus, it is likely that each 20S subunit has a particular position within the complex that is defined by its interactions with the other subunits.

It has not yet been possible to relate the structure of either the 20S or 26S complexes to their function due to the relatively low resolution obtained. However, an attractive theory would be that the active sites of the enzyme are occluded within a channel inside the enzyme. This would allow for regulation of access to the proteolytic sites by a process, possibly energy dependent, that allows the substrate to enter the enzyme. In this way a type of "internal compartmentalisation" may be achieved. Work on the yeast proteasome has shown that mutations in some subunits lead to defects in the proteasome activities (Heinemeyer, Kleinschmidt et al, 1991, Heinemeyer, Gruhler et al, 1993, Hilt, Enenkel et al, 1993). In all cases these have been β-type subunits (more related to the *Thermoplasma* β sequence than the α sequence). Recently, the primary structure of the rat RC10-II subunit has been determined and is of the β-type (Nishimura, Tamura et al, 1993). The sequence has strong similarity to
bovine subunit 13 which was protected from NEM labelling by leupeptin and may be involved in the trypsin-like activity (Dick, Moomaw et al., 1992). It is likely that the eukaryotic proteasome resembles that from *T. acidophilum* in the α7β7β7α7 structure. Therefore, the β subunits would be in the interior of the complex, although the active sites may be towards the inside or outside of the cylinder.

Whether the "hole" sometimes observed in end-on perspectives runs throughout the complex is unclear. The outer α subunits appear to form discs at a resolution of 2-3 nm, although this does not preclude the presence of a smaller hole. The structure of the 26S complex allows for the CF2 inhibitor component to act by regulating the access of proteins to the active sites. Finally, one study shows that treatment of the 20S complex with some inorganic compounds can alter both the sedimentation and the activity of the complex (Djaballah, Rowe et al., 1993). This is suggestive of a mechanism of proteasome regulation whereby interaction with other proteins leads to a conformational change in the enzyme which would be causative for the change in activity.

11g. Primary sequences of proteasome subunits

The proteasome complex has been well conserved throughout evolution. Primary sequences of many of the proteasome subunits have now been published and these allow an analysis of the evolution of the proteasome through several species. The archaebacterium *T. acidophilum* proteasome has only two components and is the simplest that has been described as yet (Dahlmann, Kopp et al., 1989). It is often described as an ancestral proteasome. Ubiquitin has also been found in this archaebacterium (Wolf, Lottspeich et al., 1993) and the proteasome is able
to degrade ubiquitin associated proteins in vitro (Wenzel and Baumeister, 1993). The archaebacterial proteasome has only one defined active site (Dahlmann, Kuehn et al, 1992), as opposed to the multiple active sites of eukaryotic proteasome. This is probably a reflection of the lower complexity of the more ancient enzyme.

Curiously there is no evidence for a ubiquitin pathway of degradation or for a proteasome complex in eubacteria. However, structural similarity has been noted between a groEL-like chaperonin complex from the eubacterium Comamonas acidovorans and the proteasome of T. acidophilum (Zwickl, Pfeiffer et al, 1990). If eubacteria preceded archaebacteria in evolution, then the groEL complex may be the forerunner of the proteasome. Alternatively, Rechsteiner et al have proposed that the 26S proteasome may be related to the Clp and Lon ATP-dependent protease complexes in E. coli (Rechsteiner, Hoffman et al, 1993).

Primary sequences from the 20S proteasome components from a variety of species have been arranged in a dendogram (Fig. 1.2). The sequences can be broadly divided into two families dependent on their similarity to the α or β chains of the T. acidophilum proteasome (Zwickl, Lottspeich et al, 1991; Zwickl, Grziwa et al, 1992). The α and β families can each be divided into seven sub-families as shown (Fig. 1.2), suggesting a seven fold symmetry. In general, each subfamily contains only one sequence for any given species. The two exceptions to this are that LMP7 and MB1 are in the same sub-family, as are LMP2 and delta (Belich, Glynne et al, submitted). This analysis is suggestive of a degree of redundancy in the proteasome of humans. The LMPs have a quite restricted expression pattern (Monaco and McDevitt, 1986) and gene
disruption of yeast subunits are, with one exception, lethal (Emori, Tsukahara et al, 1991; Heinemeyer, Kleinschmidt et al, 1991; Lee, Tanaka et al, 1992; Heinemeyer, Gruhler et al, 1993; Hilt, Enenkel et al, 1993). Therefore, the presence in humans of subunits homologous to the LMPs, but with a reciprocal expression pattern (Belich, Glynne et al, submitted), should not be surprising.
Figure 1.2: A dendogram of aligned proteasome sequences. All sequences are derived from cDNAs, with the exception of Mecl4 Human and YR51 Yeast. Mecl4 was translated from a database submission of genomic sequence covering five tightly linked genes on chromosome 16 (Larsen, Solheim et al, 1993). YR51 was translated from an open reading frame closely linked to the RAD51 gene (Basile, Aker et al, 1992). The remaining sequences were downloaded from the SwissProt database. The unpublished sequence, MB1, is also included (Belich, Glynne et al, submitted). The sequences were aligned with the Pileup programme in GCG (gap weight=3, gap length weight=0.1) and plotted using the Figure programme. The sequences were grouped into 14 subfamilies, seven each within the alpha and beta families of sequences. The subfamilies are coloured and the archetypal T. acidophilum sequences are shown in black.
An alignment of the \( \alpha \) and \( \beta \) groups of sequences followed by a consensus calculation allowed regions of conservation and divergence to be determined. The similarity across the sequences is plotted graphically in figure 1.3 and conserved amino acids are shown as a consensus.

As described in the preceding section on regulation, there have been many reports of the proteasome being found in the nucleus, and redistribution of the complex between the cytoplasmic and nuclear compartments has also been noted. Clusters of basic residues, which resemble nuclear localisation signals (NLS), have been found in several components. In addition, clusters of acidic residues have also been found within the amino acid primary sequence. A third consensus, for tyrosine phosphorylation, has also been described.

A mechanism of action for nuclear translocation has been proposed (Tanaka, Yoshimura et al, 1990). In this model, the cytoplasmic form of the enzyme has the acidic residues masking the basic residues. A conformational change, possibly induced by alteration of the phosphorylation state, would then expose the potential NLS and allow redistribution of the enzyme to the nucleus. This model is largely speculative. Mutagenesis of the respective consensus sequences would allow a definitive assignment of function to these sequences. For example, it is also possible that the positive and negative charged stretches are involved in the assembly of the complex by promoting electrostatic interactions between subunits.

The \( \beta \) subunits have a highly conserved box near the N terminus. This aligns with a number of sequences that have been obtained from the \( \beta \) subunits by N-terminal Edman degradation, but is some way into the
open reading frame of the cDNA sequences. Many, if not all, of the β subunits are likely to be processed by cleavage at the N terminus, explaining the discrepancy between the cDNA and protein sequences. The likely cleavage point is marked on the Plotsimilarity diagram (Fig. 1.3b). It is probable that the consensus for the cleavage of the β subunits is contained within 2-3 residues N-terminal to the cleavage point and ~20 residues C-terminal. The cleavage is likely to be autocatalytic as the *T. acidophilum* β subunit is cleaved in *E. coli* only when the α subunit is coexpressed (Zwickl, Lottspeich et al, 1992).

As described in section 11a, the proteasome does not behave biochemically like any of the well defined classes of extracellular or lysosomal proteases. There has been only one report of a consensus sequence for a protease active site (see chapter 3). The localisation of active sites by yeast genetics and/or biochemical methods, coupled with a detailed structure of the *T. acidophilum* complex, should elucidate the proteolytic mechanism more fully.

The α and β subunits were aligned as two groups and a profile calculated to give a weight for each amino acid at each position in the alignment. The profile was used to search the SwissProt database. When the α profile was used in the search, the top scores were α proteasome subunits and the next highest scores were β subunits. Similarly, the β derived profile also found matches with β and then α proteasome subunits. However, a yeast sequence called YR-51 matched well with this profile. This sequence is from an open reading frame near the RAD51 gene (Basile, Aker et al, 1992) and is likely to be a yeast proteasome subunit.
Figure 1.3a: The alpha family of sequences were aligned as for figure 1.2. The alignment and a consensus (plurality=20) were displayed with the Pretty programme (GCG) and a running average of the similarity plotted with the Plotsimilarity programme (window=10). Conserved residues and their position within the gapped alignment are shown beneath the plot. 100% conserved residues are shown in red.
Figure 1.3b: The beta family sequences were aligned as for figure 1.2 and displayed as for figure 1.3a. A consensus was calculated with plurality=16 and is shown beneath the plot. Absolutely conserved residues are shown in red. The likely cleavage site is shown with an arrow at the C-terminal side of the conserved glycine at amino acid 75.
At the time of writing, only two primary sequences from components specific to the 26S complex have been described, neither of which have any homology to the components of the 20S particle. The same "reverse genetics" approach used for cloning the first 20S components was used. Subunit 4 is a member of a family of ATPases, and is especially related to a group of transcription factors TBP1, MSS1 and SUG1 (Dubiel, Ferrell et al, 1992). Potential protein kinase C and calmodulin kinase phosphorylation sites were found within the derived protein sequence, as was a nuclear localisation consensus. The authors point out that there is weak homology between S4, TBP1, MSS1, SUG1 and the ATP-binding protein of the E. coli Clp protease, ClpA. They argue that TBP1, MSS1 and SUG1 may, in fact, be other subunits of the 26S complex, and that there could be some evolutionary relationship between Clp and the 26S complex. Amino acid sequencing of fragments of another 26S subunit, S7, showed that it was probably identical to MSS1 (Dubiel, Ferrell et al, 1993). It is possible that the affect of MSS1 on transcription is mediated through the degradation of other proteins.

12. The LMPs

Antisera were raised between congenic mouse strains against different H-2-encoded antigens. An anti-H-2^d_1 serum was used to immunoprecipitate proteins from H-2^d_1 cell lines. The proteins were separated by 2D electrophoresis. 45 kD H-2-K and -D products were seen, as were a group of 16 proteins of molecular weight 15-30 kD with a range of isoelectric points (pI 4->8). These were called Low Molecular weight Proteins, or LMPs. The LMP determinant was expressed by monocyte and macrophage cell lines, as well as normal macrophages and spleen cells (Monaco and McDevitt, 1982).
Mouse lines were assayed for the presence of the LMP complex determinant by testing the ability of the spleen cells to inhibit precipitation with the anti-H-2\textsuperscript{d} serum by preabsorption and by direct precipitation. Using mice with recombinations within the H-2 region, the polymorphic determinant responsible for the precipitation of the LMPs was mapped to between \textit{H-2-K} and \textit{I-A}. Cosmids between \textit{H-2-K} and \textit{A\beta3} (now called \textit{Pb}) did not detect any new genes in this region and it was concluded that the LMP determinant mapped between \textit{A\beta3} (\textit{Pb}) and \textit{A\beta2} (\textit{Ob}) (Monaco and McDevitt, 1986).

Precipitation of LMPs from anti-H-2\textsuperscript{d} immuno-reactive mouse haplotypes was carried out. Comparison of LMP profiles from mice of different haplotypes showed that there were two polymorphic proteins in the complex. LMP7 was found to differ in mobility even in a comparison between immuno-reactive haplotypes. This polymorphism cannot, therefore, be responsible for the antigen recognised by the anti-H-2\textsuperscript{d} serum (Monaco and McDevitt, 1982). The LMP complex can be purified by gel filtration from non-immuno-reactive mice. These complexes have a shift in the migration of LMP2 when compared to those from the immuno-reactive haplotypes. The other LMP spots had identical mobility from all haplotypes (Monaco and McDevitt, 1986). This suggests that some or all of the other proteins were coprecipitating in a complex because of the polymorphism of one component, LMP2, that mapped to the H-2 region and that a second polymorphic, but not antigenic, protein, LMP7, also mapped to the H-2 region.

Fractionation of the lysate prior to precipitation with the anti-H-2\textsuperscript{d} serum showed that all the LMPs were precipitated from a peak of Mr
~580 kD. The ratio of intensity of the LMP spots was identical from all active fractions. These results strengthened the idea that the LMPs were associated with each other as a high molecular weight complex. When a lysate from a human cell line was fractionated in a similar way to a mouse cell lysate, the human fraction corresponding to the LMP active mouse fraction also contained proteins that migrated similarly to the mouse LMP complex on 2D gels. The LMP complex seemed therefore to be conserved in humans, but this experiment does not map any human LMPs as it does not depend on precipitation (Monaco and McDevitt, 1984).

Biochemical characteristion of the LMPs suggested that they were cytoplasmic proteins: treatment of cells with an inhibitor of glycosylation, tunicamycin, did not affect the migration of the LMPs on 2D gels; the LMPs are not labelled by surface iodination; they are not precipitated by surface immunoprecipitation and an anti-LMP serum does not give a surface FACS signal. Therefore, the LMP complex mapped within the class II region of the MHC and was inducible by IFN-γ (Monaco and McDevitt, 1986), yet seemed distinct from classical class II molecules by physical, biochemical and serological characterisation. This early data led to speculation that the LMP complex may "play some role in the 'processing' of foreign antigen for presentation to the T cell" (Monaco and McDevitt, 1986).

Monaco et al then showed that the LMP complex was structurally related to the proteasome. Both complexes contain 15-20 components of 20-35 kD with similar isoelectric points. An antiserum raised against rat liver proteasome cross reacted with the mouse complex. When such an immunoprecipitate was run with an anti H-2d immunoprecipitate on the
same 2D gel, 16 of the subunits comigrated. In particular, the two LMP subunits that mapped to the H-2 region were also found in the proteasome. That the LMP and proteasome complexes were not identical was demonstrated by two criteria: four proteasome subunits were not found in the anti H-2d precipitate and proteasome could still be precipitated from lysate that had been precleared with the anti H-2d serum (Brown, Driscoll et al, 1991).

13. Is there a role for the class I molecule in the processing of its own antigens?

Rammensee and coworkers have suggested that peptides are protected from degradation by binding to the groove of class I (Falk, Rötzschke et al, 1990, Rammensee, Falk et al, 1993). This idea is based on a number of experiments. The initial observation was that the range of peptides that could be isolated from total cell lysates was dependent on the class I molecules that were expressed by that cell (Falk, Rötzschke et al, 1990). In fact, it has been shown in one case that expression of a particular class I allele is absolutely required for the detection of a minor histocompatibility peptide (Wallny, Rötzschke et al, 1992). The hypothesis forwarded by Rammensee's group was that peptide precursors would be produced in the cytoplasm but that the final degradation step would occur within the class I assembly compartment in a MHC-dependent manner. One possibility would be that longer peptide precursors would bind within the class I binding groove and that amino acid residues that extended at the N terminus would be removed by exopeptidases (epitope trimming). Recently, a TAP-dependent peptide translocation assay showed that competition for translocation of the marker peptide by test peptides was haplotype specific. This
phenomenon was not seen when microsomes from β2m knockout mice were used. Thus, \textit{in vitro}, class I molecules can have an affect on the peptides retained or transported into microsomes (Schumacher, Kantesaria et al, 1994).

Consistent with this model is the observation that (signal sequence-derived) peptides of up to 13 amino acids could be eluted from HLA-A2 in T2 cells and that 9 mers within these sequences were also eluted (Henderson, Michel et al, 1992, Wei and Cresswell, 1992). Additionally, it has been well documented that many different cell lines are able to process a particular epitope for presentation through a transfected class I molecule (for example, see Wallny, Rötzschke et al, 1992) - the potential to present an antigen therefore seems not to be restricted by any \textit{trans}-acting gene but only through the class I gene itself, whatever the background processing machinery. A caveat to this is that a different array of peptides are presented through rat RT1.A^ in the context of different \textit{cim} alleles (Powis, Deverson et al, 1992).

The trimming model necessitates that the TAP complex would allow the transport of peptides longer than the 8-9 mers commonly eluted from class I molecules. TAP-dependent translocation assays have provided some evidence for this. For example, the assay using β2m-negative microsomes to reduce/ eliminate any contribution from class I binding showed a length preference for peptides longer than 8 aa. Only minor differences in transport were seen between peptides of 9-13 aa (Schumacher, Kantesaria et al, 1994). It is possible that other TAP assays are measuring class I binding as well as TAP transport and that these two parameters cannot be distinguished - peptides unable to bind to available
class I within the ER may then appear to be insignificantly transported by the TAP complex.

Another prediction of the trimming model is that the degradation machinery would be able to produce peptides that were longer than 9 mers. The most likely candidate for the degradation step is the proteasome complex. The amount of information on the degradation products of proteasome digestion of whole proteins is limited. However, the recent work of Boes, Hengel et al (1994) has shown that proteasomes from IFN-γ treated cells produce an 11 aa peptide containing a 9 aa epitope from a 25 mer substrate.

That MHC class I molecules can have an effect on the array of peptides present in a cell at steady state seems clear (Falk, Rötzschke et al, 1990; Wallny, Rötzschke et al, 1992). However, the available data could be explained by two models: a selection of 9 mers may be available for class I binding in the ER and those that do not bind would be rapidly degraded such that they were undetectable at steady state; alternatively a smaller pool of longer peptides may be subsequently cleaved after class I binding - in this case a given 9 mer epitope would not be found in the absence of its cognate class I molecule. Without more definitive evidence for a trimming step, the speculation of Rammensee remains only an attractive hypothesis.
14. Summary of antigen processing for presentation through class I

The aspects of class I antigen processing that have been discussed above are shown as a cartoon in figure 1.4. Cytosolic proteins may be degraded by either ubiquitin-dependent or ubiquitin-independent pathways by the 26S or 20S proteasome complexes respectively. In the former case, the expression of different ubiquitinating enzymes E1, 2 and 3 may have some role, as may the various inhibitors or activators of the proteasome. Peptide products would then traverse the ER membrane via the TAP complex. At least one TAP-independent processing route has been delineated which achieves translocation of signal sequences across the ER membrane. Within the ER lumen, peptides and β2m assemble with class I, allowing its release from the chaperone p88. The trimeric complex then proceeds to the cell membrane. As the precision of proteasome cleavage is unknown with respect to epitopes, it is possible that the peptides are trimmed by an exoprotease to form 8-9 mers. This is shown in the figure as occurring on the class I molecule itself, such that the class I binding groove acts as a template for the final stages of epitope proteolysis.
Figure 1.4: Potential processes involved in the processing of antigen for class I presentation. The most likely candidate for antigen degradation is the proteasome. This is shown as either a 26S form degrading ubiquitinated protein or a 20S form degrading non-ubiquitinated protein. Ubiquitination may be regulated by differential expression of the enzymes E1, E2 and E3 and degradation may be regulated by inhibitors and activators of the proteasome. The peptide products cross the ER membrane in a process that is dependent on both the TAP complex and ATP. A second source of peptides is signal sequences produced by the action of signal sequence peptidase. Assembly of peptide and β2 microglobulin with class I allows release from p88 (calnexin) and egress of the trimeric complex from the ER. The possible role of the class I molecule itself acting as a template for peptide production is shown as a trimming step within the ER lumen.
ubiquitin o cytosolic protein

E1, E2, E3

proteasome activators/inhibitors?

peptide products

ribosome

transport

signal sequence peptidase

26S proteasome

ATP degradation

ADP + Pi

peptide

ADP + Pi

β2m

MHC class I

p88

and then to cell surface

to

TAP complex

ER membrane

ER lumen

Cytosol
15. Aims of the work described in this thesis

The aim of this work was to isolate new genes in the MHC which might encode proteins with a role in the class I processing route, or be of interest with regard to the many diseases associated with MHC haplotypes. The TAP1 gene had recently been cloned and sequence homology with ABC transporter genes led to the suggestion that the TAP1 protein provided peptides for assembly with class I. The LMP complex had some resemblance in its physical properties to the 20S proteasome and the TAP1 gene was within the region where the LMP genes were expected to map. It was an attractive idea that genes encoding the mechanisms for degradation and transport of class I antigens might be genetically linked. Additionally, from homology with other ABC transporter genes, it was possible that there could be a second transporter gene closely linked to TAP1 which would encode a polypeptide that associated with TAP1 to form a heteromeric complex. Therefore, I set out to look for other genes close to TAP1 and within the region in which the LMPs had been mapped in mouse.

In chapter 3 of this thesis, data is presented which describe the cloning of LMP7 which mapped less than 1 kb away from TAP1. The translated protein sequence had homology to proteasome components, adding to the evidence that the proteasome degraded proteins to form antigenic peptides. Another gene with proteasome homology, LMP2, was cloned shortly afterwards and was also tightly linked to TAP1. In chapter 4, I show biochemically that both LMP genes encode proteasome subunits and that the LMP proteins are synthesised with N-terminal leader sequences. The final results chapter contains data on the genomic organisation of the LMP7 locus. These data include the isolation of two
forms of LMP7 cDNA; a study of polymorphisms within the LMP7 open reading frame and a description of opposite strand transcripts isolated from cDNA libraries.
CHAPTER 2: MATERIALS AND METHODS

1. Solutions

Solutions not described in sections 2, 3 and 4 are listed:

GTE: 50 mM glucose
      25 Tris-Cl pH 8.0
      10 mM EDTA pH 8.0

PCI: 25 parts phenol pH 8.0
     24 parts chloroform
     1 part isoamyl alcohol

TE: 10 mM Tris-Cl pH 7.6
    1 mM EDTA

20X SSC: 3 M NaCl
         0.3 M sodium citrate pH 7.0

Southern
hybridisation solution: 6X SSC
                        5X Denhardt's solution
                        0.5% w/v SDS
                        10% v/v dextran sulphate
                        40 µg/ml salmon sperm DNA
100X Denhardt's:  
10g PVP40  
10 g BSA  
10 g Ficoll 400  
to 1 litre in distilled water

Salmon sperm DNA:  
500 mg salmon sperm DNA was cut into small pieces with scissors. The DNA was dissolved in 50 ml water (stirred overnight). The solution was sheared by three passes through a 19g needle and boiled for 10 minutes.

2YT:  
10 g Bacto tryptone  
10 g yeast extract  
5 g NaCl  
to 1 litre in distilled water

10X TBE:  
108 g Tris base  
54 g Boric acid  
40 ml 0.5 M EDTA  
to 1 litre in distilled water

GTC:  
4 M guanidinium thiocyanate  
0.1 M Tris-Cl pH 7.5

Northern hybridisation mix:  
20 ml 50% dextran sulphate  
50 ml formamide  
20 ml 5M NaCl  
5 ml 20% SDS  
5 ml water
4X random prime
labelling mix:  100 μl 10X hexanucleotide mix
               (Boehringer Mannheim)
               37.5 μl 3N mix
               (2mM dATP, dGTP, dTTP)
               112.5 μl distilled water

2. DNA preparation and manipulation

Standard protocols are based on those described in Sambrook, Fritsch et al, 1989.

Labelling of DNA fragments using random hexamer priming

20-50 ng of DNA were diluted to a volume of 16 μl with distilled water, boiled for 3 minutes and then cooled on ice. 7.5 μl of 4X mix, 0.5-5 μl of ^32P dCTP (10 μCi/μl, Amersham) and 1.5 μl Klenow enzyme (5 units/μl, NBL) were added and the reaction incubated at 37° C for 30-60 minutes. The unincorporated radioactivity was removed by centrifugation through a Sepharose G50 column poured in a 1 ml syringe and pre-washed with TE. The labelled probe was boiled for 5 minutes and then cooled on ice before being added to the hybridisation solution.

Screening of cDNA libraries

All cDNA libraries referred to in this thesis were prepared by the method of Seed (Seed, 1987) by David Simmons (ICRF, Oxford). The cDNA inserts were cloned into the CDM8 vector and electroporated into MC1061/P3 bacteria. The efficiency of transfection was determined by
plating on LB agar containing ampicillin and tetracycline and the library was then plated onto 20 x 20 cm Biodyne filters at a density of ~250,000 colonies on each of four master filters. Colony lifts were made by placing a second filter on top of the colonies on the master, pressing between two perspex plates and allowing the colonies to grow overnight. Two such filters were made for each master. The colonies on the replica filter were lysed by placing the filter onto 3MM filter soaked in 1.5 M NaCl/0.5M NaOH for five minutes. Excess solution was blotted on dry 3MM paper and the colonies were neutralised on 3MM paper soaked in 3 M potassium acetate pH 5.5. The filters were air dried and then baked at 80° C under vacuum for 2 hours.

Hybridisation of filters

Filters were prehybridised for 1-6 hours in hybridisation solution. For genomic cosmid probes, the probe was competed with 4 μg human carrier DNA in 1 ml 6X SSC. The mixture was boiled for 5 minutes and then incubated at 65° C for two hours. The labelled, competed probe was added to fresh hybridisation mix and incubated with the filters overnight at 65° C. Filters were washed in 0.1X SSC/1% SDS for high stringency hybridisation, but at 2X or 6X SSC if less stringency was required. The filters were placed against Kodak XAR-5 film backed with an intensifying screen and left overnight at -70° C.

Colonies that were positive on both filters were picked either as mixed minipreps or were plated out as secondary colonies on LB agar. Mixed miniprep cultures were prepared as described below (plasmid minipreps), cut with XhoI to excise the inserts and the digests were separated by agarose gel electrophoresis. The gels were capillary blotted
in 20X SSC onto HybondN or N+ filters and the filters were fixed by baking at 80° C under vacuum for 2 hours. The filters were then probed with the original cosmid fragment as above. Bacterial cultures giving positive mixed minipreps were plated out as secondary colonies. Secondary colonies were denatured and fixed onto circular Hybond membranes and reprobed. Individual positive clones were end-sequenced using primers derived from CDM8 sequence flanking the cloning site.

**Southern blotting**

Restriction enzyme digestions were carried out with a 5-10 fold excess of enzyme in the recommended buffer. DNA fragments were separated on 0.6-2% agarose gels in 0.5 or 1X TBE. High molecular weight DNA was depurinated by soaking the gel in 0.15 M HCl for 10 minutes. The gel was then rinsed in distilled water and soaked in 0.2 M NaOH, 0.6 M NaCl for 30 minutes to denature the DNA. The gel was washed in distilled water once more and then neutralised with two 20 minute incubations in 0.5 M Tris pH 7.6, 0.5 M NaCl. The DNA was capillary blotted onto nitrocellulose filter(s) (Hybond) in 20X SSC and the DNA fixed by baking the filters at 80° C for 2 hours. The filters were briefly rinsed in 2X SSC before prehybridisation. Hybridisation was as described above (Hybridisation of filters).

**Plasmid minipreps**

5 ml of LB broth containing the appropriate antibiotic (ampicillin 50 µg/µl and/or tetracycline 12.5 µg/ml) were inoculated with a single bacterial colony containing the required plasmid construct. The culture
was grown overnight at 37° C with shaking. 1.5 ml of the culture were spun at 13,000 rpm for 30 seconds and the pellet resuspended in 100 µl of ice cold GTE. (For CDM8 constructs 10 mM EDTA was used instead of GTE). 200 µl of 0.2 M NaOH (freshly diluted from 5 M stock), 1% SDS were added to lyse the bacteria and the contents mixed by inversion. Genomic DNA was precipitated with 150 µl of ice cold 3 M potassium, 5 M acetate and the contents mixed by vigorous shaking. The tubes were placed on ice for 5 minutes, shaken once more and spun at 13,000 rpm for 5-15 minutes. The supernatant was extracted with an equal volume of PCI and the phases separated with a short centrifugation. The plasmid DNA was precipitated from the upper (aqueous) layer by the addition of 1/10th volume 3 M NaAc pH 5.2 and 2-2.5 volumes absolute ethanol. The DNA was left to precipitate for 2-30 minutes at 4° C or room temperature. Centrifugation at 13,000 rpm for 5-30 minutes produced a pellet which was washed in 1 ml of 70% ethanol. The pellet was then dissolved in 50 µl of TE containing 20 µg/ml DNase-free RNase.

**Plasmid maxipreps**

5 ml of LB/antibiotic were inoculated with a single colony of bacteria and grown for 6-8 hours, 37° C with shaking. 1 ml of this culture was used to inoculate 400 ml of LB/antibiotic and the culture grown overnight. The culture was split into 50 ml aliquots and bacteria pelleted by centrifugation at 4,000 rpm in a Heraeus Minifuge T. The pellets were resuspended in a total volume of 10 ml ice cold GTE and combined. (For CDM8 constructs 10 mM EDTA was used instead of GTE). Bacteria were lysed by the addition of 20 ml freshly prepared 0.2 M NaOH, 0.1% SDS, mixed gently and incubated on ice for 5 minutes. 10 ml of ice cold 5 M potassium, 3M acetate were added and the mixture left on ice for a
further 15 minutes. The genomic DNA was removed by centrifugation at 3, 000 rpm, room temperature for 15 minutes in a Heraeus Minifuge T followed by straining the supernatant through gauze. 0.6 volumes of isopropanol were added to the supernatant and the plasmid DNA allowed to precipitate for 15 minutes to 2 hours at room temperature. The precipitate was collected by centrifugation at 2, 000 rpm, room temperature for 15 minutes in a Beckman J-6B centrifuge. The pellet was air-dried for 5 minutes and dissolved in 2.5 ml TE. This solution was added to 4.4 g CsCl on a balance. Further TE was added until 4 g liquid in total had been added. 400 μl of Ethidium bromide (10 mg/ml) were added and, if necessary, the density of the resulting solution adjusted with CsCl or TE to be 1.5 -1.6 g/ml. A pellicle of bacterial debris was removed after centrifugation at 3, 000 rpm, room temperature for 10 minutes in a Heraeus Minifuge T and the clear supernatant added to Beckman polyallomer Quick Seal centrifuge tubes using a Pasteur pipette. The tube was filled with CsCl in TE (1.5-1.6 g/ml). The tubes were centrifuged at 55, 000 rpm, 15° C for 6-18 hours.

The band of plasmid DNA in the resulting CsCl gradient was visualised under UV light. The tube was punctured with a 25g needle at the top and the plasmid DNA extracted with a 19 or 21g needle. The ethidium bromide was removed by repeated extraction against water-saturated butan-2-ol. The volume was increased 3 fold with TE, two volumes of absolute ethanol were added and the mixture allowed to stand for five minutes before centrifugation at 4, 000 rpm, room temperature for 10-15 minutes in a Heraeus Minifuge T. The pellet was resuspended in 400 μl of TE and transferred to a 1.5 ml tube. 40 μl of 3 M NaAc pH 5.2 and 1 ml of absolute ethanol were added, mixed and the precipitate collected by centrifugation at 13, 000, room temperature, 5
minutes. Two further precipitations were carried out before the pellet was washed in 1 ml 70% ethanol and resuspended in 400 µl of TE. The yield was determined by reading the absorbance of an aliquot of the DNA solution at 260 nm.

Sequencing of double stranded plasmid DNA

2-5 µg of high quality plasmid DNA (either CsCl purified or prepared on a Qiagen column) in 18 µl water or TE were denatured by the addition of 2 µl 2 M NaOH, 20 mM EDTA and left at room temperature for 5 minutes. The solution was neutralised with 3 µl of 3 M NaAc pH 5.2, 7 µl distilled water added and the DNA precipitated with 75 µl ethanol. The mixture was cooled in dry ice, the DNA collected by centrifugation and washed in 70% alcohol.

Sequencing was carried out using the Sequenase kit from USB based on the chain termination method of Sanger (Sanger, Nicklen et al, 1977). The pellet was dissolved in 7 µl distilled water and 2 µl Sequenase buffer. 1 ng of the sequencing primer was added, the solution heated to 65° C for two minutes and then cooled slowly to <35° C for the primer and template to anneal. 1 µl of 100 mM DTT, 2 µl of 5 fold diluted labelling mix, 0.5 µl of 35S dATP (10 µCi/µl, Amersham) and 2 µl of 8 fold diluted Sequenase enzyme were added and the labelling reaction incubated at room temperature for 2-5 minutes. 3.5 µl aliquots were added to 2.5 µl of each of the ddG, A, T, C termination mixes (pre-warmed to 37° C) and the reactions were incubated at 37° C for 5 minutes. The reactions were terminated by the addition of 4 µl stop solution. The sequence reactions were heated to 75° C for 3 minutes and then cooled in an ice/water bath prior to loading on a urea denaturing
6% acrylamide gel in 1X TBE. Gels were dried at 80° C under vacuum. Autoradiography was for 16-32 hours at room temperature with Kodak XAR-5 film.

**Shotgun sequencing**

Inserts from the cDNA clones were sequenced across their length by the shotgun method of Bankier and Barrell (Bankier, Weston et al, 1987). The DNA was self-ligated, sonicated to small fragments, end-repaired and blunt-end ligated into SmaI cut M13 mp18.

**M13 vector**

5 μg M13 mp18 replicative form was digested with 5 units SmaI and 20 units calf intestinal phosphatase in 50 μl 15 mM Tris-Cl pH 8.0, 5 mM MgCl2, 15 mM KCl for 1 hour at 37° C. Agarose gel electrophoresis was used to assay for complete digestion. The solution was extracted with phenol and the DNA precipitated with 0.3 M NaAc pH 5.2, 2.5 volumes ethanol. The vector was washed in 70% alcohol and resuspended in TE at 20 ng/μl.

**Insert**

~50 μgs of the cDNA clone was cut with XhoI in H buffer (Boehringer Mannheim) at 37° C. Sufficient enzyme and incubation time was used for complete digestion. The insert was removed from the CDM8 vector by electrophoresis on an agarose gel and purified from the gel by electrophoresis into 1X TBE in a dialysis bag. The solution was phenol extracted twice, ethanol precipitated and resuspended in 20 μl TE.
The fragment was self ligated by 400,000 units T4 DNA ligase in ligation buffer (NEB) at 15° C for 2-3 hours. The self-ligated insert was diluted into 250 µl of TE and sonicated sufficiently to produce fragments of size ~500 bp. The fragments were ethanol precipitated and resuspended in 30 µl TE. The fragment ends were repaired with 10 units each of Klenow and T4 DNA polymerase, 2 µl of 0.5 mM stocks of dATP, dCTP, dGTP and dTTP and 5 mM MgCl$_2$. The reaction was incubated for 45 minutes at room temperature. The DNA was then size-fractionated on an agarose gel and fragments between 350-700 bp were gel-purified by electroelution into a trough cut in the gel containing 5X TBE. The DNA was ethanol precipitated and resuspended in 50 µl of TE.

**Ligation**

Three ligation reactions were set up for each insert containing 1, 2 and 4 µl of insert DNA and 20 ng of vector DNA. Ligations were done in 20 µl of ligase buffer with 2 x10$^6$ units of T4 DNA ligase, at 15° C for 15 hours. Control ligations were also set up which had either the insert, the vector or the enzyme missing. Ligations were transfected into competent JM109 bacteria.

**Preparation of JM109 bacteria**

An overnight culture of JM109 in 2YT was diluted 100 fold into fresh 2YT and grown until the OD$_{600}$ reached 0.4-0.6. Bacteria were harvested by centrifugation at 2,000 rpm in a Heraeus minifuge T and, after removal of as much of the medium as possible, resuspended in 2.5 ml of ice cold 10 mM MES, 100 mM RbCl, 45 mM MnCl$_2$, 10 mM CaCl$_2$ and 3 mM hexaminecobaltic chloride. The bacteria were incubated on
ice: after 15 minutes 100 µl of DMSO were added; after a further 5 minutes, 100 µl of 2.25M DTT, 40 mM potassium acetate pH 6.0 were added; after 10 more minutes 100 µl of DTT were added and the bacteria left on ice until ready for transfection. 200 µl of competent cells were used per transfection. 20 µl of the ligation reaction were added to the cells in sterile glass tubes and incubated on ice for 45 minutes. The cells were then heat shocked at 42°C for 2-5 minutes and 3 ml top agar (10g Bacto tryptone, 8g sodium chloride, 8g Bacto agar in 1 litre), 25 µl 2% X gal, 25 µl 2.5% IPTG, pre-heated to 45°C, was added. The bacteria in the top agar were poured onto LB agar plates. The top agar was allowed to set at room temperature and the plates incubated overnight at 37°C.

**Single strand M13 minipreps for sequencing**

A single colony of JM109 was picked and grown as an overnight culture in 2YT. This was diluted 100 fold in 2YT. Individual white plaques of M13 in JM109 were picked into 1.5 ml aliquots of the diluted JM109 culture and grown at 37°C with shaking for 4.5-5.5 hours. The bacterial cells were pelleted by centrifugation and the phage were precipitated from the supernatant by the addition of 150 µl 20% PEG 8000, 2.5 M NaCl. The tubes were incubated for 20 minutes at room temperature and the particles collected by centrifugation for 5 minutes at 13,000 rpm, room temperature. Care was taken to remove as much of the supernatant as possible. The pellet was resuspended in 100 µl of TE, extracted with phenol to remove the phage coats, and the DNA ethanol precipitated from the aqueous layer. The DNA was resuspended in 30 µl of TE and 5 µl of this was used directly for sequencing as described above (page 87).
Solid phase sequencing of PCR products

PCR reactions were carried out with one primer biotinylated. Conditions were optimised for specificity of product and to exhaust the primers. The product was immobilised by incubation at room temperature for 30 minutes with 200 µg of streptavidin conjugated to magnetic beads (Dynal). The PCR buffer and reagents were removed using the magnetic particle concentrator (Dynal) and the DNA denatured by 5 minute incubations in 0.15 M NaOH and subsequently 0.15 M NaOH, 0.1 M NaCl. The beads were then resuspended in 7 µl water, 2 µl Sequenase reaction buffer and 5 ng sequencing primer. Sequencing of the immobilised biotinylated strand was performed as described (page 87).

Strand specific labelling of DNA fragments

Single strand labelling of DNA was performed as described in the literature (Espelund, Prentice Stacy et al, 1990). A PCR product was made with one of the two primers biotinylated. PCR conditions were optimised to exhaust all the primers and the specificity of the reaction was tested by agarose gel electrophoresis. 200-300 ng of DNA were bound to 25 µl of Dynal streptavidin magnetic beads in 100 µl 6X SSC, room temperature for 30 minutes with periodic agitation. The beads were washed twice in 6X SSC using the magnetic particle concentrator (Dynal). The DNA was denatured by two incubations of the beads in 0.15 M NaOH, 0.1 M NaCl for five minutes each. The beads were washed twice in 1X SSC, 0.1% SDS and twice with distilled water. All wash volumes were 100 µl.
Primining of the DNA was either random or primer specific. In the latter case, a 4 fold molar excess of the unbiotinylated primer used in the PCR reaction was added to the beads. The primer and template were annealed by cooling slowly from 65° C. 10 μl OLB, 3μl 32P dCTP and 2 μl Klenow were added and the reaction was incubated at 37° C. The extent of the labelling was measured by following incorporation of radioactivity into the bead fraction. When ~70% activity was in the bead fraction the supernatant was discarded and the second strand was eluted by heating the beads to 95° C in 1X SSC.

Subcloning of plasmid inserts

Vector preparation

5-10 μg of vector were cut to completion at the cloning site with an excess of restriction enzyme. 0.1 units of calf intestinal phosphatase (Boehringer Mannheim) were added and the reaction incubated for a further 30 minutes at 37° C. To stop the reaction, trinitriloacetic acid was added to 10 mM, EDTA added to 15 mM and the solution heated to 68° C for 30 minutes. The proteins were removed by 2-3 extractions with Strataclean resin (Stratagene) and the DNA precipitated with ethanol and NaAc. The vector was dissolved in TE at 100 ng/μl.

Insert preparation

Plasmid DNA was cut to completion with an excess of restriction enzyme. Enough plasmid was cut to produce 5-10 μg of insert. The insert was gel purified by one of several methods: electroelution into buffer in dialysis tubing; electroelution into a trough cut in the gel and
filled with 5X gel running buffer or Geneclean (BIO 101) following the manufacturer's instructions. The former two methods necessitate ethanol precipitation of the DNA.

**Ligations**

Ligation reactions were set up in 1X ligase buffer (NBL) with approximately equal molar ratios of vector (100 ng) and insert. Reactions were incubated either at 15° C overnight or at 37° C for 2 hours. In some cases the ligation reactions were then precipitated with ethanol and resuspended in a small volume of distilled water. At least two concentrations of insert were used for each ligation. Controls were set up without insert, vector, or enzyme respectively and a positive control with supercoiled plasmid was also included. Reactions were electroporated into competent bacteria and plated onto LB agar containing the appropriate antibiotic(s). Colonies were screened for insert containing constructs by hybridisation to colony lifts (as for cDNA libraries).

**Subcloning of PCR products**

Two methods were used for subcloning PCR products. In the first method, the PCR primers were designed so as to incorporate restriction enzyme sites compatible with those of the vector. The PCR product was gel-purified, cut with the restriction enzymes, phenol extracted (or Stratacleaned) and ligated as for plasmid DNA inserts.

The efficiency of this method is somewhat lower than the CloneAmp kit (Gibco BRL). The manufacturer's instructions were followed. Briefly, one of the PCR primers was tailed with the sequence
CAUCAUCAUCAU, the other with CUACUACUACUA. 10-50 ng of the PCR product was mixed with 2 μl (50 ng) of pAmp vector and 1 μl of uracil DNA glycosylase (UDG, 1 unit/μl) and incubated at 37° C for 30 minutes. The enzyme UDG modifies the uracil bases in the primer tails preventing them from base pairing. The denatured ends of the PCR product then anneal to the pAmp vector, which has been prepared with compatible ends. After the incubation the reaction was placed on ice and 1 μl was used for transfection into bacteria. In this case (construction of CAX1) 100% (n=6) of the resulting clones contained recombinant plasmid.

Preparation of electrocompetent bacteria

An overnight culture of bacteria was diluted 200 fold into fresh LB and the culture grown until the OD₆₀₀ reached 0.4-0.6. The bacteria were centrifuged at 4000 rpm for 15 minutes at 4° C in a Heraeus minifuge T and washed twice with an equal volume (to the culture volume) of ice cold water. The bacterial pellets were resuspended in a total volume of 50 ml ice cold 10% glycerol and combined, spun as above and the pellet resuspended in an equal volume (to the pellet) of ice cold 10% glycerol. The bacteria were aliquoted into cold 0.5 ml tubes, frozen in dry ice and stored at -70° C.

1 μl of ligation reactions, or 5-50 pg of closed circle plasmid were added to 50 μl aliquots of the frozen bacteria, mixed and added to ice cold electroporation cuvettes (0.2 cm gap). The bacteria and DNA were pulsed at 2.5 kV, 2.5 μF, 200 Ω. 1 ml of LB broth was added to the cuvette immediately after the pulse, the sample transferred to a clean culture
tube and incubated at 37°C with shaking for 30 - 60 minutes. The bacteria were plated out on LB agar containing the appropriate antibiotic.

**Single strand conformational polymorphism (SSCP)**

SSCP protocol was based on that of Hayashi (Hayashi, 1991). 1 µg of DNA from homozygous typing cell lines was added to a PCR reaction containing the following: 5 µl PCR buffer (Promega), 2 units Taq polymerase (Promega), 1 µCi ³²P dCTP, 0.7 µl 4N mix (10mM dATP, dCTP, dGTP, dTTP) and 300 ng each of primers derived from the introns of the LMP7 gene (one primer from each pair was biotinylated) in 50 µl water. PCR conditions were optimised to ensure that the primers were exhausted and the product was specific.

Two 2 µl aliquots of the PCR product were cut in 20 µl of the appropriate restriction buffer with AluI and Hinfl in separate reactions. These digests were diluted one to one with stop solution (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol). Each digest was denatured by boiling for 5 minutes and quickly cooled in an ice water bath. (For the first exon of LMP7b, the PCR fragment was only 300 bp. This fragment, being relatively short, was not digested, but was diluted 20 fold in 0.5X stop solution prior to loading).

8 µl aliquots of the samples were run on 6% acrylamide gels (0.15% bis-acrylamide) in 1X TBE at both room temperature and at 4°C. The room temperature gels included 10% glycerol and were run overnight at 200V, the cold gels were run for 4-6 hours at 600V. The gel was dried at 80°C under vacuum and visualised by autoradiography with Kodak XAR-5 film and an intensifying screen at -70°C.
3. RNA preparation and manipulation

RNA preparations

Cells were collected from a 50 ml culture volume by centrifugation. The pellet was resuspended in 10 ml GTC and sodium lauryl sarcosinate added to 0.5% to lyse the cells. 0.1 volumes 2 M NaAc pH 4 was added followed by 1 volume of water saturated phenol and 0.2 volumes of chloroform/isoamylalcohol (49: 1). The contents were mixed after each addition. The tube was incubated on ice for 15 minutes and then centrifuged at 4,000 rpm, 4° C for 10 minutes in a Heraeus minifuge T. The genomic DNA precipitated at the interface of the two phases. RNA was precipitated from the aqueous layer by the addition of an equal volume of isopropanol and the tube left at -20° C 1 hour / overnight. The RNA was then collected by centrifugation as before, resuspended in 300 μl of GTC and transferred to a 1.5 ml tube. A second precipitation was carried out with an equal volume of isopropanol. The RNA pellet was washed in 75% ethanol and resuspended in 50 μl 0.5% SDS by heating to 65° C for 10 minutes. The RNA concentration was assessed by absorbance: OD$_{260}$=1 equivalent to ~ 40 μg RNA/ ml.

Northern blotting

Preparation of gel for northern blotting

3 g agarose was melted in 255 ml water, 30 ml 10X MOPS. After cooling to 50° C, 16.2 ml of 37% formaldehyde was added before the gel was poured. 15 μg total RNA in 9 μl water were added to 7 μl formaldehyde, 20 μl formamide and 4 μl 10X MOPS and heated to 60° C
for 30 minutes. The RNA was cooled on ice, 2.5 μl 50% glycerol, 0.1 mg/ml bromophenol blue added and the RNA loaded onto the gel. The gel was run in recirculating 1X MOPS buffer at 100-150 volts. At the end of the run, the gel was stained in ~5 μg/ml ethidium bromide and destained in distilled water until the 28S and 18S ribosomal RNA bands were visible.

**Blotting of the RNA onto nitrocellulose and hybridisation**

RNA was capillary blotted onto HybondN nitrocellulose and fixed by baking at 80° C for two hours under vacuum. The filter was rinsed briefly in 5X SSC and prehybridised in northern hybridisation mix for 30 minutes at 42° C. The labelled probe was boiled with 300 μl of 10 mg/ml salmon sperm DNA for 5 minutes and then added to the hybridisation mix. Hybridisation was at 42° C overnight and the filters were washed to 0.2X SSC, 0.1% SDS. Autoradiography was at -70° C with an intensifying screen.

4. Preparation of antisera, immunoprecipitations and western blotting

**Production of BNF1 antiserum**

A peptide corresponding to the C-terminal 15 amino acids of the predicted open reading frame of LMP7 was synthesised. This was conjugated to keyhole limpet haemocyanin (KLH). Equal molar ratios of peptide and KLH were dissolved in 0.1 M NaHCO3 at a KLH concentration of 2 mg/ml. A fresh vial of 25% ultra pure grade glutaraldehyde was thawed and added to the solution to a concentration
of 0.05%. The mixture was stirred at room temperature overnight. Glycine ethyl ester, pH 8.0, was added to 0.1 M followed by a 30 minute room temperature incubation. The coupled peptide KLH was precipitated by the addition of 4-5 volumes ice cold acetone and incubation at -70° C for 30 minutes. The mixture was allowed to warm slowly to room temperature. The coupled protein was pelleted at 10,000g for 10 minutes. The pellet was resuspended in 0.9% NaCl at 1 mg/ml.

500 µg peptide/KLH and Freund's adjuvant were injected into a Sandy half-lop rabbit approximately once a month for 6 months. Test bleeds were taken 1 week after each injection.

ELISA assay

The peptide which was used as the epitope for the antiserum was conjugated to BSA as described above. 50 µl of the peptide/BSA conjugate was added to each well of a Immulon 2 flat bottomed 96 well plate (Dynatech) and left overnight 4° C. The protein solution was replaced with 100 µl gelatin (0.02%) and the plate again left overnight at 4° C. The gelatin solution was removed. The antiserum was serially diluted in PBSA, 10% FCS with 10 fold steps. 100 µl of the diluted antiserum solutions were added to the wells of the plate. The plate was left at room temperature for 1 hour and then washed three times with PBSA/Tween 20 (1 drop Tween 20 for 400 ml PBSA). The second layer antibody, an anti-rabbit immunoglobulin/ horse radish peroxidase conjugate, was added to each well at a 1: 100 dilution in PBSA, 10% FCS. The plate was left at room temperature for a further hour. The antibody was removed and the plate washed as before. 20 mg OPD (ortho phenylene diamine) was dissolved in 10 ml PBSA and 20 µl H2O2. 100 µl
of this solution was added to each well and OD readings were taken at 450 nm 2-15 minutes afterwards. The readings for each bleed of the serum were measured twice (two wells) for each dilution.

**Immunoprecipitation**

Immunoprecipitations were carried out essentially as described (Townsend, Ohlen et al, 1989). 10^7 cells were washed in PBSA and preincubated in 1 ml methionine-free medium for 1 hour. 75 μCi of ^35S methionine were added and incubation continued for 2 1/2 hours. The cells were washed twice in PBSA, lysed in 1 ml lysis buffer (0.5% NP40, 0.5% nonanoyl-N methylglucamide, 150 mM NaCl, 5 mM EDTA, 50 mM Tris pH 7.5) with the addition of 2 mM PMSF and 5 mM iodoacetamide for 20 minutes on ice and the nuclei pelleted. All subsequent steps were carried out at 4°C and the lysates were rotated during all incubations.

The supernatant was precleared with 200 μl 10% washed staphylococcus A overnight. The staph A was pelleted and 500 μl aliquots of supernatant were precipitated with the appropriate antiserum for 90 minutes. In some cases, SDS was added to 2% and the lysates heated to 95°C for 4 minutes to dissociate the proteasome prior to precipitation. RIA grade BSA was added to 1% and 100 μl of 5% protein-A sepharose w/w in lysis buffer were added for 45 minutes. The sepharose beads were washed four times in 1 ml lysis buffer, transferring the sepharose suspension to fresh tubes at the fourth wash. The lysis buffer was removed and the proteins eluted from the beads in 50 μl of sample buffer (62mM Tris pH6.8, 10% glycerol, 2% SDS, 5% 2-mercaptoethanol, 0.00125% bromophenol blue) at 95°C for 4 minutes.
Samples were run on 12 or 15% denaturing acrylamide gels. The gels were fixed (7% acetic acid, 40% methanol), Coomassie stained, treated with Amplify (Amersham) and dried on 3MM paper. Kodak XAR-5 film was exposed to the dried gel for an appropriate length of time at either room temperature or -70°C with an intensifying screen.

Proteolysis assay

Immunoprecipitations were carried out as described above, with the exception that protease inhibitors were omitted from the lysis buffer and precipitates were washed in detergent-free lysis buffer. The beads were resuspended in 50 mM Hepes buffer pH 8. Substrates used were Ala-Ala-Phe-7-amino-4-methyl-coumarin (AAF); N-t-Butoxycarbonyl-Leu-Ser-Thr-Arg-7-amino-4-methyl-coumarin (LSTR) and succinyl-Leu-Leu-Val-Tyr-7-amino-4-methyl-coumarin (LLVY). Substrates were added to 20 μM in a final volume of 400 μl and the reaction incubated at 37°C for 30 minutes. The reaction was stopped by the addition of 1.4 ml of 0.2 M sodium carbonate. Proteolysis was assayed by reading the fluorescence of the leaving group (excitation 340 nm, emission 425 nm). Blank reactions, in which the substrates were incubated without any precipitate, were included as a control.

Pulse chase analysis

10^8 cells were washed in PBSA and incubated in 3 ml methionine-free RPMI/10% FCS for 1 hour. 1 mCi of 35S methionine was added and incubation continued for a further 30 minutes. The cells were washed in PBSA and resuspended in 230 ml RPMI/10%FCS. Aliquots of 30 ml cells were removed after 0 hr, 1/2 hr, 2 hr, 5 hr and 22 hr of chase and lysed

These reagents were obtained from Sigma.
(as above). All of these lysates were heated to 95° C for 4 minutes in 2% SDS prior to precipitation with BNFl. After 0 hr and 22 hr chase additional 30 ml aliquots of cells were lysed and precipitated with BNFl without prior SDS treatment. The samples were run on a denaturing 15% acrylamide gel and the gel fixed, stained, treated with Amplify and dried as above. Kodak XAR-5 film was exposed to the dried gel at -70° C with an intensifying screen.

Reprecipitation

8 x 10^7 LCL721 cells were cleared for methionine and cysteine for 1 hour in 4 ml Met/Cys-free E4 medium. The cells were then labelled with 1.6 mCi of labelling mix (Amersham SJQ0079) for a further hour. After chase periods of 0, 2, 4 and 21 hours in conditioned medium 2 x 10^7 cells were lysed in 2 mls of lysis buffer, the nuclei removed and the lysates precleared with Staph A (as for "Immunoprécipitation").

80 μg of MCP21 and 3 μl of rabbit anti-mouse antibody were added to the precleared lysates and the precipitations collected as above. 50 μl of lysis buffer containing 2% SDS were added to the washed precipitates on protein A sepharose, mixed and heated to 95° C for 5 minutes. The beads were pelleted and the supernatant collected. A further aliquot of lysis buffer/2% SDS was added to the beads, heated and collected and the two aliquots were combined. 25 μl of the eluted proteins (1/4) were removed and added to 25 μl of sample buffer.

The remaining 75 μl were diluted to 1.5 ml in lysis buffer (without SDS) and precleared with Staph A to remove any MCP21 antibody. The sample was then divided into three and second precipitations were
carried out with BNF1 (αLMP7), AK14 (αLMP2) or a mock precipitation with no second antibody. The precipitations were carried out as described and the samples were run on a 15% acrylamide gel.

**Western blotting of immunoprecipitates**

Immunoprecipitates were prepared from labelled cells and run on a 15% denaturing gel as described above. The gel was blotted onto a nitrocellulose membrane. For all subsequent steps the membrane was on a rocking platform at room temperature. The membrane was blocked in 3% BSA in PBSA for 2 hours followed by a 1 hour incubation in 200 mM Tris pH 7.4, 10% FCS, 1% BNF1 serum. The membrane was washed several times in PBSA and then incubated with an anti-rabbit peroxidase conjugated antibody (P217, DAKO) as for BNF1. After several more washes in PBSA the membrane was incubated in 20 ml methanol containing 50 mg 4-chloro-1-naphthol for 5 minutes. The membrane was rinsed in water, dried on 3MM paper and photographed.

**Western blotting of cell lysates**

Cells were lysed in lysis buffer (as above). The lysate from $5 \times 10^4$ cells was mixed with an equal volume of sample buffer (as above) and heated to 95°C for 4 minutes. Samples were run on a denaturing 15% acrylamide gel. The gel was blotted onto a HybondC membrane. For all subsequent steps the membrane was on a rocking platform at room temperature. The membrane was placed in blocking solution (5% Marvel and 0.2% Tween 20 in PBSA) overnight. BNF1 antiserum was diluted 1:1000 in blocking solution and incubated with the membrane for 1 hr. The membrane was then washed extensively in large volumes of 0.2%
Tween in PBSA (2 brief washes, 1 wash of 15 minutes and two washes of 5 minutes). A second 1 hr incubation of the membrane in a 1:1000 dilution of P217 in blocking solution followed. The membrane was washed as before and treated with ECL reagents (Amersham) as described by the manufacturer. Kodak XAR-5 film was exposed to the membrane for 15 seconds to twenty minutes.

**Denaturing polyacrylamide protein gel electrophoresis**

A solution of 15% separating gel was prepared to the following recipe. 10 ml 1.5 M Tris-Cl, pH 8.8, 0.2 ml 20% SDS, 20 ml acrylamide solution (30% acrylamide, 2.6% C), 9.6 ml water. The solution was degassed under vacuum for 15 minutes and polymerisation initiated by the addition of 0.2 ml fresh 10% ammonium persulphate and 20 ml of TEMED. The separating gel was allowed to set for 2 hours - overnight. A stacking gel was then poured with composition 2.5 ml 1.5 M Tris-Cl, pH 8.8, 50 µl 20% SDS, 1.3 ml acrylamide solution, 6.1 ml water, polymerised with 50 µl 10% APS, 10 µl TEMED. Gels were run either at 70 V overnight or 100 V until the dye front reaches the separating gel followed by 200 V for 5 hours.
CHAPTER 3: CLONING OF LMP7, A PROTEASOME-RELATED GENE IN THE CLASS II REGION OF THE MHC

INTRODUCTION

Monaco and McDevitt had shown that there were one or more polymorphic proteins mapping within the H-2 region of the mouse which were part of a large complex (Monaco and McDevitt, 1982, Monaco and McDevitt, 1984, Monaco and McDevitt, 1986). There was a striking similarity between this "LMP" complex and the proteasome, a good candidate for the degradation of cytoplasmically synthesised proteins to produce antigenic peptides for presentation through class I. To further characterise the LMP complex and the MHC-encoded components, I set out to clone the LMP genes.

The human and mouse MHC and H-2 regions are syntenic throughout the region in which the LMP genes were mapped in the mouse (Hanson and Trowsdale, 1991). Therefore, we expected the LMP loci to map between DPB and DOB, the homologues of Pb and Ob in the mouse H-2 region. The TAP genes had recently been cloned and mapped to the U10 and U15 cosmids from the contig constructed by Blanck and Strominger (Blanck and Strominger, 1988, Deverson, Gow et al, 1990, Monaco, Cho et al, 1990, Spies, Bresnahan et al, 1990, Trowsdale, Hanson et al, 1990, Powis, Mockridge et al, 1991, Colonna, Bresnahan et al, 1992). The TAP gene products have been implicated in antigen processing for MHC class I presentation (Kelly, Powis et al, 1991, Powis, Townsend et al, 1991, Spies and DeMars, 1991, Kleijmeer, Kelly et al, 1992, Powis,
Deverson et al, 1992, Spies, Cerundolo et al, 1992) and this has been discussed in chapter 1.

The TAP genes mapped between DPB and DOB, in the same region that the LMPs had been mapped in the mouse, and there was speculation that the LMP gene products might have a related function to the TAPs. I therefore reasoned that the LMP and TAP loci could be linked genetically. There is a NotI restriction site in the region of overlap of U10 and U15 just telomeric of the TAPI gene (Fig. 3.1). As NotI sites are often found associated with CpG islands and, by extension, with the 5' end of genes (Bird, 1987) we used a fragment of the U15 cosmid on the other side of the NotI site to TAPI (Fig. 3.1). This fragment of U15 led to the cloning of LMP7. Another fragment of U15 contained LMP2. These two genes had homology to proteasome components.

RESULTS and DISCUSSION

Cloning and sequence analysis of LMP7

A 3 kb fragment (NS1) was excised from the telomeric end of the U15 cosmid, gel purified and used to probe cDNA libraries derived from a T cell line, CEM, and an interferon-γ stimulated macrophage line, U937. Positive clones were characterised as described in chapter 2. A clone was obtained and sequenced which matched the genomic sequence over this region. This clone, called NS1C1.1 and described in more detail in Chapter 5, did not have an open reading frame and we reasoned that this may be due to aberrant splicing. I used a 650 bp probe (called S6 and extending ~600 bp telomeric of the NS1 probe) from the 5' end of

(1) (Beck, Kelly et al, 1992)

(2) From the CEM T cell cDNA library.
NS1C.1 to screen for related clones from the CEM cDNA library. The clones from this screen were end-sequenced.

One clone, called RING10, had sequence at one end which was in an intron of NS1C.1. The other end of the insert had a poly A tail followed by sequence which matched genomic DNA sequence. It could be concluded that RING10 and NS1C.1 were transcribed from different strands of DNA. NS1C.1 will be further described in Chapter 5, and the relationship between RING10 and NS1C.1 is shown in figure 5.4.

The 1.3 kb insert from RING10 was gel purified and sequenced by the shotgun method (Bankier, Weston et al, 1987). There was an open reading frame of 816 bp, encoding a protein of 272 amino acids, $M_r$ 30 kD and $pI$ 5.4. The gene has since been referred to in the literature as LMP7 which will be used in much of the following text. To determine the genetic relationship between LMP7 and TAP1, the LMP7 locus was mapped by Southern blotting of cosmid digests (not shown) and by comparison of the cDNA sequence with the genomic sequence over this region (Beck, Kelly et al, 1992). LMP7 mapped between TAP1 and another ABC transporter gene called TAP2 (see figures 3.1 and 5.5).
Figure 3.1: The genomic position of LMP2 and LMP7 in relation to the TAP genes. The upper panel shows the main genes of the class II region of the human MHC. The lower panel shows the TAP/LMP locus in more detail. NotI sites are marked as N. The LMPs are shown in red and the TAPs in blue. RING9 is marked in light blue and is discussed in chapter 5. Class II genes are shown in yellow. The NS1 probe that was used for the isolation of clone NS1C1.1 is shown in pink.
The sequence of the LMP7 (RING10) clone, together with the predicted protein sequence, is shown in figure 3.2. Given the similarity in the physical properties of the LMP and proteasome complexes, it was of interest to see whether the LMP7 open reading frame had homology to any previously published proteasome sequences. The LMP7 protein sequence was checked against the PIR 27, Swiss Prot and OWL databases. Matches were found with components of human proteasome. The most striking match was with human ε component which shared 15 of 22 amino acids (Fig. 3.3).

The sequence from Hum ε was derived from N-terminal amino acid sequencing. However, the match with the LMP7 sequence started 69 amino acids into the predicted open reading frame of the LMP7 cDNA. There were a number of possible explanations for this. First, the N-terminal sequence of Hum ε could have been from an artefactual degradation product. This seemed unlikely as several other proteasome subunits had been end-sequenced at the protein level and also matched the LMP7 sequence at the same position (Fig. 3.3). Second, translation of LMP7 mRNA may begin from the second in-frame methionine, only 4 amino acids N-terminal of the block of conserved sequence. In fact, this initiation codon matches the Kozak consensus sequence more closely than the start site shown as the beginning of the open reading frame in figure 3.2 (Kozak, 1981). Third, it is possible that LMP7 mRNA is translated as shown and the N-terminal 68 amino acids are then removed. This would result in a mature protein which had an N terminus consistent with the sequence homology. Biochemical data which supports this explanation is presented in chapter 4.
Figure 3.2: The nucleotide sequence of the LMP7 cDNA. The derived protein sequence of the longest open reading frame is shown below. The initiating methionine is marked as an arrow, the putative cleavage point for processing of the immature protein is marked with an arrow. Potential active site residues are shown in bold and the regions of homology with subtilisin-like proteases are coloured. The green sequence was found by the program PROMOT, the yellow sequences were found by inspection. Alignments of these sequences with subtilisin-like proteases are shown in figure 3.5. The sequence is truncated at the 3' end and contains a further 124 nucleotides before the poly A tail.
Figure 3.3: Alignment of LMP7 with the N-terminal protein sequences (HUM ε, RAT 6, HUM δ, RAT 5, RAT 7, HUM β) and cDNA sequences (HUM C3, Dm 35, ScC1, TAα) from proteasome subunits. Dm 35 is from Drosophila melanogaster, ScC1 is from Saccharomyces cerevisiae and TAα is from Thermoplasma acidophilum. Dashes represent positions at which there is identity with LMP7, lower case represents conservative changes (Argos, 1987). X is a residue that was uncertain in the N-terminal sequencing. Numbering is according to the LMP7 derived protein sequence shown in figure 3.2.
The relationship between LMP7 and other cDNA sequences that had been published at the time was less striking than with the N-terminal amino acid sequences. However, alignments are shown both at the N-terminus of the predicted open reading frame (Fig. 3.3), by diagonal analysis (Fig. 3.4) and as part of a dendogram (Fig. 1.2). The degree of identity between LMP7 and derived protein sequence from published proteasome cDNAs (at the time of cloning LMP7) varied between approximately 10 and 20%, whereas that between human C3 and the α chain of *Thermoplasma acidophilum*, for example, is approximately 40%. For comparison, a diagonal of human C3 against the α chain of *Thermoplasma acidophilum* is shown in figure 3.4. However, it was clear that LMP7 had some homology to proteasome components throughout the length of the open reading frame and not just at the N-terminus.

The proteasome cDNA sequences that were published at the time of cloning LMP7 were largely of the α class of proteasome genes, whereas LMP7 is a member of the β family. The N-terminal amino acid sequences which showed the highest homology with LMP7 were from β-related subunits. The relationship between LMP7 and more recently published proteasome cDNA-derived sequences can be most clearly seen in the form of a dendogram (Fig. 1.2).

**Potential active site residues within the LMP7 sequence**

As proteasomes are thought to be multi-proteinase complexes with broad specificity, they should have active sites able to catalyse several proteolytic reactions (Rivett, 1993). Whilst inhibition has been reported with both serine and cysteine protease inhibitors (Rivett, 1985, Tanaka, Ii et al, 1986), homology to known consensus sequences for
protease active sites had not been found in proteasome cDNAs. Therefore, I screened the LMP7 predicted protein sequence against the PROMOT database of protein sequence motifs. This revealed a significant match between LMP7 and a subtilisin family serine protease consensus for an active site histidine at aa 185 (Fig. 3.5). The probability of such a match occurring randomly in a sequence of this length is close to 1 in 7,000. It is possible that LMP7 utilises this histidine as a proton acceptor in a proteolytic reaction.

If this were the case, then it would be expected that these amino acids would be well conserved through evolution. The mouse homologue of LMP7, MC13 shares over 90% identity with its human counterpart over much of the open reading frame (Frentzel, Graf et al, 1992). However, this histidine is not conserved - an asparagine residue is found at the equivalent position. The same is true of the rat sequence, RC1 (Aki, Tamura et al, 1992). Asparagine has an amino group in the side chain and it is conceivable that this is involved in a charge-transfer type of catalysis. Within the cluster of LMP7-related sequences shown in the dendogram (Fig. 1.2), the MHC-encoded genes have a basic residue at this position, (LMP7, MC13 and RC1) whereas the non-MHC encoded genes have an acidic residue (MB1, Pre2 and Pts1). It is tempting to speculate that this difference might be related to the difference in activity between LMP-containing and LMP-lacking proteasomes (Driscoll, Brown et al, 1993, Gaczynska, Rock et al, 1993).
Figure 3.4: Diagon plots of cDNA-derived proteasome sequences from human (chains 3 and 5), *Drosophila melanogaster* (Dm 35), *Saccharomyces cerevisiae* (ScCl) and *Thermoplasma acidophilum* (TAα) compared to LMP7. A plot of two α family sequences is shown for comparison (Hum C3 and TAα). The Similarity Investigation Program in the Staden package was used with a window of 31, a proportional score of 31 and the default score matrix. Extrapolation of the plots to the LMP7 axis in the first five cases suggests that the LMP7 protein starts some way into the open reading frame.
Figure 3.5: Comparison of LMP7 with serine protease active sites. Symbols and numbering are as in figure 3.3, with the exception that X represents "any amino acid" in the consensus. Potential active site catalytic triad residues are marked with an asterisk. The histidine consensus sequence was identified by screening the LMP7 protein sequence with the PROMOT database (Sternberg, 1991) and shows an exact match at amino acids 185-194. Alternative consensus residues are also shown. Suggested alignments, found by inspection, of LMP7 with the linear sequences surrounding the active site serine and aspartate residues of known subtilisin-like serine proteases are also shown. Subtilisin BPN' is from *Bacillus amyloliquefaciens*, subtilisin S from *B. subtilis* var. *amylosacchariticus*, furin is a human protease and KEX2 is from *S. cerevisiae*.

<table>
<thead>
<tr>
<th></th>
<th>LMP7</th>
<th>Subtilisin BPN'</th>
<th>Subtilisin S</th>
<th>Furin/PACE</th>
<th>KEX2</th>
</tr>
</thead>
<tbody>
<tr>
<td>162</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>L</td>
<td>M</td>
<td>G</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>Subtilisin BPN'</td>
<td>T</td>
<td>-</td>
<td>A</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Subtilisin S</td>
<td>T</td>
<td>-</td>
<td>A</td>
<td>t</td>
</tr>
<tr>
<td></td>
<td>Furin/PACE</td>
<td>T</td>
<td>a</td>
<td>S</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>KEX2</td>
<td>T</td>
<td>a</td>
<td>A</td>
<td>A</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>LMP7</th>
<th>Consensus</th>
</tr>
</thead>
<tbody>
<tr>
<td>185</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>t</td>
<td>a</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>LMP7</th>
<th>Subtilisin BPN'</th>
<th>Subtilisin S</th>
<th>Furin/PACE</th>
<th>KEX2</th>
</tr>
</thead>
<tbody>
<tr>
<td>206</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>V</td>
<td>M</td>
<td>D</td>
<td>S</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>Subtilisin BPN'</td>
<td>i</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Subtilisin S</td>
<td>i</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Furin/PACE</td>
<td>i</td>
<td>l</td>
<td>D</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>KEX2</td>
<td>i</td>
<td>v</td>
<td>D</td>
<td>-</td>
</tr>
</tbody>
</table>
Known serine protease active sites have three stretches of primary sequence each containing a residue involved in the catalytic mechanism. Typical consensus sites for the other two residues of the serine protease catalytic triad, aspartate and serine, were not identified by the motif search. Aspartate and serine residues which have some match with known active site residues with respect to their surrounding sequence are also shown in figure 3.5. It may be that these residues are in a more atypical linear sequence, or that LMP7 uses different active site residues in its reaction mechanism. It should also be noted that the order of the putative catalytic residues in LMP7 (serine, histidine, aspartate) differs from that of both subtilisin-like proteases (aspartate, histidine, serine) and trypsin-like proteases (histidine, aspartate, serine). The putative motifs for aspartate and serine are conserved in the LMP7-related sequences (MB1, RC1, MC13, Pre2 and Pts1). A definitive identification of these residues in catalysis would require a detailed biochemical analysis of proteasome reactions following site directed mutagenesis of the candidate residues.

The identification of possible residues involved in proteasome catalysis by sequence comparison is subject to the criticism that the biochemistry of proteasome cleavage differs from that of the well-characterised families of proteases. Active site residues identified by biochemical methods (Dick, Moomaw et al, 1992) or genetic methods (Haas, Warms et al, 1982, Heinemeyer, Kleinschmidt et al, 1991, Heinemeyer, Gruhler et al, 1993, Hilt, Enenkel et al, 1993) do not correspond with the residues described above from sequence analysis. Both the biochemical and genetic techniques would detect residues that are spatially near the active site but which may not, per se, be part of the catalytic mechanism. The residues which take part in catalysis may be
far away in the primary sequence and conceivably could even be on
different subunits. The use of substrate analogues which are predicted to
react covalently with an active site side chain should further define the
catalytic mechanism(s) and delineate the amino acids that are involved in
proteolysis. This approach is being followed by Rivett and co-workers
(Savory, Djaballah et al, 1993).

Cloning and sequence analysis of LMP2

When fragments of the cosmid U15 were used to probe a B
lymphoblastoid cell cDNA library, another clone was isolated that was
named RING12 in our lab nomenclature (Kelly, Powis et al, 1991). This
clone has subsequently become known as LMP2 and will be referred to
as such. The genomic position of LMP2 was determined by hybridisation
to Southern blots of U15 digests and then by comparison to the genomic
sequence when this became available (Beck, Kelly et al, 1992) (Fig. 3.1).

The LMP2 clone had a 715 bp insert which contained an open
reading frame expected to encode a protein of 219 amino acids, Mr 23.2
kD and pI 4.64. When this predicted protein sequence was run against
the protein databases, matches were found with proteasome sequences.
Homology between LMP2 and N-terminal amino acid sequence from
other proteasome subunits was found 20 amino acids into the LMP2
predicted open reading frame. Processing of LMP2 from an immature
protein with an N-terminal extension is a possible explanation for this.
(see chapter 4). The relationship of LMP2 to other proteasome sequences
is shown in the dendogram (Fig. 1.2). No homology was found with
protease active sites.
LMP2 maps very close to TAP1, TAP2 and LMP7. In fact, the 5' ends of the LMP2 and TAP1 cDNAs are only ~500 bp apart and are on opposite strands. Thus LMP2 and TAP1 could be transcribed from one bidirectional promoter. When the LMP2 cDNA sequence was compared to the genomic sequence over this region, one polymorphic change was seen. This resulted in a histidine to arginine dimorphism at amino acid 60 (Kelly, Powis et al, 1991).

**MRNA expression of the LMPs**

The speculation that the LMP gene products have a role in antigen processing would predict that their expression would parallel class I. To test this, labelled probes were prepared from the inserts of the LMP2 and LMP7 cDNAs. These were used to probe northern blots of total RNA from a range of cell lines. In addition, the effect of IFN-γ on the mRNA levels of these two genes was also studied. It had previously been shown that the polymorphic antigen of the LMP complex (which mapped to the H-2 region of the mouse) was upregulated by IFN-γ (Monaco and McDevitt, 1986). Similarly, the mRNA levels of the TAPs and the class I MHC antigens are increased by treatment with IFN-γ, and this cytokine is also known as immune interferon.

The northern data for the LMPs is consistent with a role in presentation of antigen. Thus, their mRNAs are constitutively expressed in B and T lymphoblastoid lines and are up-regulated by IFN-γ in colon carcinoma and fibroblast cell lines (for LMP2, see Kelly, Powis et al, 1991) and normal human keratinocytes (for LMP7, see figure 3.6). Upregulation of LMP7 mRNA occurs within 6 hours after treatment with the cytokine (not shown) and reaches a maximum at 72 hours (Fig. 3.6).

Normal (untransformed) keratinocytes were taken from a healthy individual undergoing a breast reduction operation. They were grown in RPMI/10% FCS containing 200 units/ml IFN-gamma. RNA was isolated at the indicated time points as described in Materials and Methods.
Figure 3.6: Northern analysis of LMP7 mRNA expression. Upper panel, lane 1, T-cell line (J6), lane 2, B-cell line (MANN) and lane 3, monocyte cell line (U937). Lower panel, normal human keratinocytes were treated with 200 units recombinant human interferon-γ for 2, 48, 72 and 96 hours before RNA extraction. 20 µg of RNA was run per track. The first lane shows RNA from untreated cells (lane marked “-”). The cDNA insert shown in figure 3.2 was used as a probe.
A more detailed analysis of LMP7 mRNA expression is presented in chapter 5.

**Relationship between the LMPs and other proteasome cDNA sequences**

As discussed in the introduction, the proteasome is well conserved and has been studied in organisms as diverse as the archaebacterium *Thermoplasma acidophilum*, yeast *Saccharomyces cerevisiae*, plant *Spinacea oleracea*, insect *Drosophila melanogaster*, mouse, rat and human (reviewed in Goldberg and Rock, 1992, Tanaka, Tamura et al, 1992, Rivett, 1993, Rivett and Knecht, 1993). The *T. acidophilum* proteasome has only two components, named α or β. There is striking sequence conservation between the subunits and this has allowed the cDNA sequences to be assigned as either α or β, dependent on their homology to the ancestral proteasome subunits. In such an analysis, the LMPs are both β subunits. Two other genes with high homology to the LMPs have recently been cloned (DeMartino, Orth et al, 1991 and Belich, Glynne et al, submitted). These genes, called *MB1* and δ, are approximately 65% identical to *LMP7* and *LMP2* respectively.

Inspection of the dendogram allows a division of the sequences into 14 subfamilies, seven each for α and β. If this corresponds to an incorporation of two subunits per particle (suggested by Kopp, Dahlmann et al, 1993) to produce an αβγβγαβγ stoichiometry then it is to be expected that there would be one sequence per sub-family for a given species. Whilst this seems to be true for the yeast *S. cerevisiae*, there is another human sequence in each of the subfamilies that contain LMP7 and LMP2, namely MB1 and delta respectively. The LMPs have a limited
tissue distribution but MB1 and delta are likely to be constitutive subunits that are replaced by the LMPs in certain cell types or conditions. This prediction has now been confirmed with antisera raised against each of the four subunits (Belich, Glynne et al, submitted).

**Evolution of the TAP/LMP gene cluster and the LMP-related genes**

The *LMP* and *TAP* genes form a tightly linked cluster of genes within the class II region of the MHC (Figure 3.1). The evolutionary relationship of these four genes is of some interest, as it may allow some insight into the function of the gene products. The genomic sequence of this cluster of genes has been published (Beck, Kelly et al, 1992). The *TAP* genes have the same number of exons and identical intron/exon boundary classes (that is, whether the 3' end of each exon is at position 1, 2 or 3 within a codon). Additionally, eight of the eleven exons are of identical sizes. The conclusion drawn from these data was that the *TAP* genes are the result of a duplication event. Similar analysis of the *LMP* genes showed that they both have six exons, share 4 out of 5 intron/exon boundaries and have one exon the same size. It was proposed that the four genes arose by a duplication of an initial *TAP* and *LMP* gene pair, followed by inversion of one of the *LMP* genes (Beck, Kelly et al, 1992).

The *LMP*-related genes, *MB1* and *delta*, map elsewhere in the genome (Belich, Glynne et al, submitted). The homology of MB1 to the yeast subunit Pre2 is slightly higher than the homology between LMP7 and Pre2, suggesting that *MB1* is the more ancient of the two human genes. Recently, sequence of the yeast proteasome subunit Pre3, with high homology to LMP2 and delta has been published (Enenkel, Lehmann et al, 1994). Homology between Pre3 and delta (55% identical
amino acids) is significantly greater than between Pre3 and LMP2 (44% identity). The protein expression patterns of MB1 and delta are more consistent with a constitutive "housekeeping" role in protein breakdown (Belich, Glynne et al, submitted). Conversely, the LMP gene products are expressed predominantly in those cells that process and present antigen. It is possible, then, that the LMPs represent specialised components that have evolved with a particular function and that this function is likely to be in antigen degradation.

As the homology between MB1 and LMP7 or delta and LMP2 is significantly higher than between LMP2 and LMP7, it is likely that MB1 / LMP7 and delta / LMP2 are the result of duplication events that are more recent than LMP2 / LMP7. As MB1 and delta are more related to the yeast homologues than to the LMPs and because they have constitutive expression patterns, it seems likely that MB1 and delta are the ancestral pair of genes that duplicated to form the LMPs.

Taking this into account, a theory for the evolution of the TAP, LMP, MB1 and delta genes would be as follows. Duplication of a proteasome-related gene would form MB1 and delta. This pair of genes would then duplicate to form another pair of linked proteasome genes, LMP2 and LMP7. As the TAP genes are closely linked to the LMPs, they may have been a part of this duplicated region and could have homologues near to MB1 and delta. After this duplication event, the TAP/ LMP gene cluster would have diverged from the ancestral genes and become more specialised in function. MB1 and delta have now been mapped to different chromosomes (Belich, Glynne et al, submitted) and this model would require that another event occurred after their duplication that separated the two constitutive proteasome genes.
The question of when the TAP/LMP cluster became part of the MHC is also relevant to the function of the LMPs. If a proteasome/ABC transporter arrangement of genes were found either outside of the MHC or in an organism that doesn’t have an MHC, then it would hint at a shared function for the gene products. The function of the proteasome genes in such a situation would then provide clues for a role of the LMPs in antigen processing. Duplication of MB1 and delta to form the LMPs presumably preceded acquisition of the limited expression pattern of the LMP gene products because gene knockouts of proteasome subunits in yeast have a lethal phenotype (Heinemeyer, Kleinschmidt et al, 1991, Heinemeyer, Gruhler et al, 1993, Hilt, Enenkel et al, 1993). In other words, redundancy within the proteasome subunits would be a prerequisite for specialisation of components to occur.

CONCLUSIONS

This chapter has described the cloning of LMP7, a gene in the class II region of the MHC which has homology to proteasome subunits. Subsequently, a second proteasome-related gene, LMP2, was cloned and mapped nearby to LMP7. The LMPs form a cluster of four genes with the TAP transporters. The mRNA levels of all four genes appear to be coordinately regulated with MHC class I, and the TAP gene products are necessary for efficient class I antigen presentation. Therefore, three pieces of circumstantial evidence support the hypothesis that the LMP gene products have a role in antigen degradation for class I: homology between the LMPs and a protease, the genomic position of the LMP genes close to the TAP genes and the LMP mRNA expression pattern.
ACKNOWLEDGEMENTS

The LMP2 gene was cloned and sequenced by Adrian Kelly (Kelly, Powis et al, 1991). The northern blots shown in figure 3.6 were kindly provided by Ruth Lovering (top panel) and Lesley-Anne Kerr (bottom panel).
CHAPTER 4: AN INVESTIGATION OF THE LMP GENE PRODUCTS

SUMMARY

This chapter describes the use of specific antisera raised against the C termini of LMP2 and LMP7. Experiments to show that the LMPs are bona fide proteasome subunits are presented. Both LMP proteins are processed from an immature form by cleavage at the N terminus resulting in loss of ~4 kD from LMP2 and ~6 kD from LMP7. The resulting mature proteins are likely to have an N terminus consistent with that derived from other subunits by N-terminal amino acid sequencing. The cleavage occurs before or shortly after the LMPs become associated with another proteasome subunit (defined by the monoclonal antibody MCP21). The steady state levels of the proteins are upregulated by the cytokine interferon-γ, as would be expected from the northern data (chapter 3 and Kelly, Powis et al, 1991). Finally, data are presented which show that the proteasome complex that contains the LMPs is proteolytically active against oligopeptide substrates.

INTRODUCTION

As discussed in chapter 3, two LMP genes within the human MHC had homology to proteasome components (Glynne, Powis et al, 1991, Kelly, Powis et al, 1991). Other proteasome subunits had previously been cloned by a "reverse genetics" approach: N-terminal amino acid sequence from rat proteasome subunits, or fragments thereof, was used to design oligonucleotide probes. These probes were then used to clone the cDNAs
The high conservation of proteasome sequences allowed homologues to be cloned from other species (for example Tamura, Tanaka et al, 1990). The LMPs were not cloned by this method: initial characterisation of the LMP gene products as proteasome components (Glynne, Powis et al, 1991, Kelly, Powis et al, 1991) rested on sequence homology and the results of Monaco and coworkers.

The mouse anti-H2d serum recognised 16 proteins that comigrated with proteins present in an anti-proteasome precipitate (Brown, Driscoll et al, 1991). Anti-proteasome precipitates contained a protein with polymorphic mobility which segregated with MHC markers in congenic mouse lines (Ortiz-Navarrete, Seelig et al, 1991). Preclearing experiments with the anti-H2d serum showed that the LMP complex was a subset of the total proteasome pool (Brown, Driscoll et al, 1991) and subsequent work has shown that the LMPs are present in ~70% of the proteasome population of B cells and interferon-γ induced hepatoma cells (Brown, Driscoll et al, 1993). Using antisera raised against the C terminus of the LMP predicted proteins, I was able to investigate the relationship of the human LMP proteins to the proteasome complex.

The discrepancy between cDNA-derived protein sequences and N-terminal amino acid sequences hinted at post-translational processing of the β type proteasome subunits and this has been discussed in the introduction. Synthesis and cleavage of leader sequences at the N terminus of extracellular proteases is well documented (Price and Stevens, 1989). In this case, cleavage of a leader sequence is used to protect the organism from aberrant intracellular proteolysis and is coincident with induction of proteolytic
activity of the enzyme. N-terminal leader sequences have also been shown to be important in intracellular protein trafficking (von Heijne, 1988). The proteasome has been found in several intracellular compartments (discussed in the introduction), and a possible role for leader sequences of proteasome components would be in the targeting of the complex within the cell. Finally, the N-terminal extensions of some proteasome subunits may be necessary for the formation of intermediates in the assembly of the proteasome complex.

RESULTS and DISCUSSION

Production and specificity of the BNF1 antiserum

To investigate the biochemical relationship of LMP7 to the proteasome, I raised an antiserum against the predicted protein sequence. The C and N termini of proteins are often good candidates for epitopes exposed to the solvent and accessible to antibody, as they are rarely buried in the interior of the protein. This is especially true if the sequence is hydrophilic. The position of the N terminus of the LMP7 protein was uncertain, so it was decided to raise an antiserum against the C terminus of the predicted open reading frame of the LMP7 cDNA clone. The \( \beta \) proteasome subunits share little homology over their C-terminal sequences (Fig. 1.3b) and we reasoned that there would be little chance of cross-reactivity with other subunits.

A peptide which corresponded to the C-terminal 15 amino acids of the predicted protein sequence was synthesised chemically. This peptide,
called R10C15 and with amino acid sequence TDVSDLLHQYREANQ, was coupled to keyhole limpet haemocyanin and the conjugate was injected into a rabbit at monthly intervals. Test bleeds were taken one week after each injection and assayed for anti-peptide activity against a R10C15: BSA conjugate in an ELISA assay. The titre of the first test bleed was not significantly higher than the pre-bleed. However, there was a large increase in the activity of the second bleed, as expected for a classic memory immune response. Bleed 2 (BNF1.2) had activity above background to a dilution of 10^-4 (Fig 4.1). Subsequent bleeds had approximately half the activity of BNF1.2. Accordingly, BNF1.2 was used in initial experiments.

To assay the specificity of the antiserum, lysates were made from the wild type B cell line LCL721 and the mutant cell line LCL721.174 which contains a deletion encompassing the LMP7 locus. Western blots of these lysates with BNF1.2 gave a band of ~23 kD specific to LCL721 (Fig. 4.2).

The specificity of BNF1 was also tested by immunoprecipitation from two metabolically labelled cell lines: BM16, a B cell line, and T2, a mutant cell line related to LCL721.174 (Salter, Alexander et al, 1985). The antiserum precipitated a group of proteins from BM16 with molecular weights between 20 and 30 kD. These proteins were not precipitated when BNF1 was used with a lysate from T2, nor when the prebleed was used with a BM16 lysate (Fig. 4.3). Two weak bands were precipitated by BNF1 from T2, presumably due to cross reaction, but these were not visible when BNF1 was used at a lower concentration (data not shown).
Figure 4.1: The bleeds of the BNFl antiserum were titrated by ten fold dilutions in an ELISA assay against R10C15 conjugated to BSA. The assay was zeroed against the prebleed. Dilutions of the bleeds by a factor of 10, 100 or a 1000 fold show that bleed 1 has very low anti-R10C15 activity, whereas bleed 2 has activity at least 100 fold higher: when bleed 2 was diluted 1000 fold, the activity was greater than for bleed 1 at a ten fold dilution. Dilution of all bleeds greater than 100,000 fold gave background activity.
Figure 4.2: The anti-LMP7 serum BNFl specifically recognises a 23 kD protein. Western blots of lysates from LCL721.174 (lanes 2 and 4) and LCL721 (lanes 3 and 5) were probed with BNFl at a concentration of 1:500 (lanes 2 and 3) or 1:1000 (lanes 4 and 5). Markers are shown in lane 1.

Figure 4.3: The BNFl serum, at a concentration of 1:100, specifically immunoprecipitates proteins of 22-30 kD from a wild-type B cell line BM16 (lane 1), but not from the mutant line T2 (lane 3). The prebleed does not precipitate any proteins in that size range from BM16 (lane 2). Proteins were run on a 12% denaturing acrylamide gel. There is some cross reaction with two proteins present in T2, but this is not seen when the antiserum is used at a lower concentration.
Immunoprecipitation was also carried out with increasing concentrations of peptide R10C15 to compete with the labelled LMP7. The precipitation of a stack of proteins between 20 and 30 kD could be inhibited with this peptide at a concentration of 1 μg/ml for ~50% inhibition (Fig. 4.4). No inhibition of precipitation was observed using a control peptide (from the N terminus of TAP2) at 50 μg/ml (the highest concentration used). The precipitation of higher molecular weight bands from the mutant cell line .174 was not inhibited by the addition of the R10C15 peptide. This suggests that these bands are not due to cross reaction of anti-R10C15 antibody with epitopes on other proteins. Instead, they are likely to be precipitated by immunoglobulin raised against epitopes in either the carrier (KLH) or the adjuvant (Freund's) or by natural antibodies.

Thus, anti-R10C15 antibodies within the BNF1 serum immunoprecipitated a group of proteins between 20 and 30 kD from native lysates. When lysates were denatured and separated on SDS PAGE, BNF1 recognised a protein of 23 kD, which was not present in LCL721.174 cells. The simplest interpretation of these results is that BNF1 recognises a 23 kD protein that is part of a larger complex. This complex contains several other proteins of similar molecular weight. The 23 kD protein is almost certainly encoded at the LMP7 locus as the R10C15 peptide sequence was derived from the LMP7 cDNA and because the protein was not present in LCL721.174 cell lysates.
Figure 4.4: Immunoprecipitation of proteins by the BNF1 antiserum can be inhibited by R10C15 peptide. Immunoprecipitations were carried out as described in Materials and Methods with the exception that one of two peptides were added to the lysates at varying concentrations before addition of the antiserum. In lanes 1-5, decreasing concentrations of R10C15, the peptide against which the BNF1 antiserum was raised, are added to LCL721 lysates: 50 μg/ml, lane 1; 10 μg/ml, lane 2; 1.0 μg/ml, lane 3; 0.1 μg/ml, lane 4; 0.01 μg/ml, lane 5. The precipitation of proteins in the size range 22-30 kD is inhibited by this peptide. The precipitation of higher molecular weight proteins is not affected. In lanes 6-8 decreasing concentrations of an "irrelevant" peptide, TAP2NTER are added to LCL721 lysates: 50 μg/ml, lane 6; 1.0 μg/ml, lane 7; 0.01 μg/ml, lane 8. There is no obvious effect of this peptide on immunoprecipitation by BNF1. Lanes 9 and 10 are immunoprecipitation of LCL721 and .174 cells respectively without the addition of peptide.
LMP7 is a component of the proteasome complex

Sequence homology between LMP7 and proteasome subunits was suggestive that LMP7 was itself part of the proteasome complex. The data presented in the last section showed that LMP7 was associated in a complex with a number of other proteins. A comparison between the BNFl immunoprecipitate run on SDS-PAGE with published proteasome preparations (eg Lee, Moomaw et al, 1990) suggested that this complex could be the proteasome. Stronger evidence depended on the use of an antiserum raised against human proteasome (antiserum #4802) which was obtained from K. Hendil, August Krogh Institute, University of Copenhagen.

A direct comparison between anti-LMP7 and anti-proteasome precipitates was made. Proteins immunoprecipitated from the B cell line BM28.7 using either the anti-proteasome serum or the anti-LMP7 serum had indistinguishable migration on SDS-PAGE (Fig. 4.5). This is similar to results in the mouse: an antiserum raised in congenic strains recombinant over the LMP7 locus precipitated a complex with components which comigrated on 2D electrophoresis with proteins precipitated by an anti-proteasome antiserum (Brown, Driscoll et al. 1991).

In order to determine which of the bands in the proteasome stack was LMP7, anti-proteasome precipitates from metabolically labelled cell lines were probed with BNFl on a western blot. The cell lines used were: LCL721, LCL721.45.1, which is hemizygous over the MHC class II region,
Figure 4.5: Immunoprecipitations with the BNF1 antiserum resemble those of an anti-proteasome serum (kind gift of K. Hendil). The cell line BM28.7, which is hemizygous for the LMP7 locus, was used. An anti-LMP7 precipitate (lane 1), using BNF1 at a dilution of 1: 250, was compared to an anti-proteasome precipitate (lane 2), with the anti-proteasome serum at a dilution of 1: 500. A similar pattern of bands was seen for each precipitation.
and LCL721.174 (DeMars, Chang et al, 1984). A specific LMP7 band of ~23 kD was present in anti-proteasome precipitates from LCL721 and LCL45.1, but not from LCL721.174. Autoradiography of the radiolabelled proteins on the membrane allowed an assignment of the LMP7 protein within the proteasome stack (Fig. 4.6).

There are a number of proteins of 30-110 kD that are characteristic of the 26S proteasome complex when biochemically purified. There was no evidence for any association of LMP7 with proteins of 30-110 kD. In particular, addition of R10C15 peptide to immunoprecipitations by BNF1 did not compete for precipitation of proteins in this size range (Fig. 4.4). Either LMP7 is predominantly in the 20S complex; the R10C15 epitope can only be recognised in the 20S but not the 26S complex or the immunoprecipitation procedure disassociated the 20S complex from the CF1 and CF2 components.

LMP2 is a component of the proteasome

A similar analysis was then carried out for the other MHC-encoded proteasome gene product, LMP2. An anti-LMP2 antiserum, called AK14, was raised against the C-terminal 15 amino acids of the LMP2 cDNA-derived protein sequence. Western blots of lysates from LCL721, LCL721.174 and T2 showed that this antiserum recognised a protein of ~22 kD that mapped to the .174/T2 deletion. This protein is almost certainly LMP2 (Fig. 4.7). There is a fainter band of ~26 kD also visible. This is probably due to an immature form of LMP2 and is discussed on page 140. Immunoprecipitation using AK14 showed specific bands again in the 20-30
Figure 4.6: Western analysis of anti-proteasome precipitates. Metabolically labelled immunoprecipitates using the anti-proteasome serum no. 4802 were run on a 15% denaturing acrylamide gel. The cell lines used were LCL721 (lane 1); LCL21.45.1, which is hemizygous for the class II region (lane 2) and LCL721.174, which has a homozygous deletion including the LMP7 locus (lane 3). The proteins were blotted onto nitrocellulose and probed with the BNFI serum on a western blot. The proteasome components were visualised by autoradiography (left hand panel) and aligned with the LMP7 band detected on the western blot (right hand panel). The LMP7 band within the proteasome stack is marked with an arrow. The two bands above the LMP7 band on the western blot were also present in a control immunoprecipitate from lysis buffer, rather than lysate. They are probably due to reaction of the second layer antibody in the western protocol with the light chain of the anti-proteasome serum. The difference in the relative intensity of some of the bands when anti-proteasome precipitates from LCL721 and LCL721.45.1 are compared to LCL721.174 may be the result of changes in the subunit composition of the proteasome in the absence of the LMPs.
Figure 4.7: The anti-LMP2 serum AK14 specifically recognises a 22 kD protein. Western blots of lysates from LCL721 (lane 1), LCL721.174 (lane 2) and T2 (lane 3) were probed with AK14 at a concentration of 1:1000. A faint band can be seen in lane 1 which migrates slightly slower than the 22 kD protein. This may be the precursor of LMP2.
Figure 4.8: The anti-LMP2 serum AK14 specifically precipitates proteins in the 20-30 kD size range. AK14 was used at concentrations of 1:100 (lanes 2 and 3) and 1: 20 (lanes 4 and 5) to immunoprecipitate proteins from LCL721 cells (lane 2 and 4) and LCL721.174 (lanes 3 and 5). An immunoprecipitate with the anti-LMP7 serum BNF1 from LCL721 cells is shown for comparison (lane 1). Most of the bands between 20 and 30 kD in anti-LMP2 precipitates comigrate with those in the anti-LMP7 precipitate. The dark band of ~26 kD in lanes 2 and 4 comigrates with the LMP2 precursor (see figures 4.11 and 4.12).
kD size range. As was the case for BNFl precipitates, the pattern of bands in the 20-30 kD range resembled that for anti-proteasome precipitates (Fig. 4.8). However, a dark band of ~26 kD was seen in anti-LMP2 precipitates. This band is of the same molecular weight as the faint band seen with the AK14 serum on a western blot and is likely to represent immature LMP2, possibly free in the cytoplasm. Additional evidence that LMP2 is part of the proteasome complex is given on page 145.

The LMPs are synthesised with an N-terminal leader sequence

The sizes of the proteins recognised by either the AK14 or BNFl sera on western blots were less than would be predicted from the cDNA sequences of the LMPs. The LMP2 cDNA encodes a protein of 23 kD although the protein recognised by the antiserum was only 22 kD. The LMP7 cDNA encodes a protein of 30 kD whereas a protein of 23 kD was seen on western blots. There were two possibilities that would explain these discrepancies. First, translation of the LMPs may start at an internal methionine. Second, there may be some post-translational processing step which results in faster migration on SDS-PAGE. To distinguish between these, the LMPs were precipitated in the absence of any associated proteins from denatured lysates. Two proteins, of apparent molecular weights 28 and 23 kD were specifically precipitated by BNFl from lysates which had been boiled in SDS to disrupt non-covalent protein:protein interactions (Fig. 4.9). The 23 kD band comigrated with the band that was recognised by BNFl on western blots of anti-proteasome precipitates (compare figures 4.6 and 4.9).
Figure 4.9: Precipitation of LMP7 after dissociation of the proteasome. LCL721.174 (lanes 1 and 2) and LCL721 (lanes 3 and 4) were precipitated with BNF1 both with (lanes 2 and 3) and without (lanes 1 and 4) prior dissociation of the proteasome by heating to 95°C for 5 minutes in 2% SDS. BNF1 was used at a concentration of 1:500. Two bands of 28 and 23 kD are specifically precipitated by BNF1 after SDS treatment. The 23 kD band comigrates with the LMP7 band in the proteasome stack.
To find out whether there was a precursor/product relationship between these two proteins, a pulse-chase experiment was performed. This showed that the 28 kD protein was processed into the 23 kD protein with a half-life of ~2 hours (Fig. 4.10). The 23 kD protein was stable after 48 hours chase time (not shown).

A similar analysis was carried out with the AK14 serum to investigate whether LMP2 was also subject to processing. Two bands of 26 and 22 kD were specifically immunoprecipitated by AK14 from denatured lysates (Fig. 4.11). A pulse chase experiment showed that the 26 kD precursor was converted to the long-lived 22 kD protein with comparable kinetics to the processing of LMP7 (Fig. 4.12). However, comparison of anti-LMP2 and anti-LMP7 precipitates from denatured lysates after 2 hour of metabolic labelling showed that LMP7 was largely in the mature form whereas LMP2 was largely in the precursor form (Fig. 4.11). Thus, the half-life of the LMP2 precursor is slightly longer than that of LMP7. This is consistent with the steady state data obtained by western analysis: if the LMP2 precursor had a longer half-life than the LMP7 precursor and if their rates of synthesis were similar, then the steady state level of immature LMP2 would be greater than that of immature LMP7. This would be an explanation for the detection of a protein of the same molecular weight as the LMP2 precursor by western blot with AK14 (Fig. 4.7), whereas it has not been possible to detect the LMP7 precursor at steady state (Fig. 4.2).
Figure 4.10: LMP7 is processed from a 28 kD precursor. The central five lanes show precipitates that were dissociated before precipitation, the first and last lanes show precipitations from native lysates. BNFl was used at a concentration of 1: 100. The chase times after the 30 minute labelling time are shown above each lane in hours.
Both the LMP2 and LMP7 mature forms are precipitated by the AK14 and BNF1 antisera. As these antisera were raised against C-terminal peptides, it is likely that the processing seen on pulse-chase reflects cleavage of amino acids from the N-terminal end of the proteins. It cannot be discerned from this experiment exactly where within the proteins processing has occurred and N-terminal sequencing of the LMPs would be necessary to determine this. However, there is homology between sequence derived from the N termini of other β proteasome subunits, LMP2 sequence from amino acid 21 (Kelly, Powis et al, 1991) and LMP7 sequence from amino acid 69 (chapter 3). Therefore, it is likely that LMP2 is cleaved at the C-terminal side of amino acid 20 and that LMP7 is cleaved at the C-terminal side of amino acid 68.

Post-translational processing could occur for all subunits of the β type. Aside from the data for the LMPs given above, N-terminal amino acid sequence from *T. acidophilum* β subunit corresponded to amino acids 9-37 of the derived protein sequence from the cDNA (Zwickl, Grziwa et al, 1992). Similarly, N-terminal amino acid sequence from the rat proteasome subunit C5, another β type subunit, corresponded to amino acids 28-47 of the cDNA derived protein sequence (Tamura, Tanaka et al, 1990).
Figure 4.11: Precipitation of LMP2 after dissociation of the proteasome. LCL721 (lanes 1 and 2) and LCL721.174 (lane 3) lysates were denatured by heating to 95° C for 5 minutes in 2% SDS before precipitation. AK14 precipitates proteins of 26 and 22 kD from LCL721 denatured lysates (lane 2). An anti-LMP7 precipitate is shown for comparison in lane 1.
Figure 4.12: LMP2 is processed from a 26 kD precursor to form a 22 kD protein. Lysates from LCL721 cells that had been labelled for 1 hour and then chased for 0, 2, 4 or 21 hours (lanes 1-4) were denatured by heating in SDS and then precipitated with AK14 at a dilution of 1:50. Times after the chase period are shown above each lane in hours. Lane 5 shows an AK14 precipitation from LCL721.174 cells as a negative control.
Assembly of the LMPs into the proteasome

To investigate the function of the leader sequences and proteasome assembly further, I next asked whether the LMPs were incorporated into the proteasome as precursors or in the mature form. The steady state levels of the immature forms of the LMPs were very low relative to the mature forms (Fig. 4.2 and 4.7). For LMP7, western blots of total lysates detected only the 23 kD protein and the immature form could only be seen after immunoprecipitation from labelled cell lysates. Therefore, it was not valid to conclude from the results of figure 4.6 that the immature form of LMP7 was not a part of the proteasome complex. For LMP2, a specific 26 kD band could be seen on western blots but was much fainter than the mature 22 kD protein (Fig. 4.7).

Therefore, metabolic labelling was used as a way of visualising the immature forms of both LMPs. Labelled cell lysates from a pulse chase experiment were first precipitated with a monoclonal antibody, MCP21, against a constitutive proteasome component. The proteins from this precipitation were eluted from the protein A sepharose beads by heating to 95 °C in 2% SDS and the eluate divided into four aliquots. One of these aliquots was mixed directly with sample buffer. The other three aliquots were precleared with Staph A to remove any undenatured MCP21 antibody.

1 MCP21 has been used as a means of purifying the proteasome by one group (K. Hendil and coworkers, University of Copenhagen, who kindly provided the hybridoma for our use).
To each of two aliquots were added BNFl or AK14 to investigate whether any labelled LMP proteins were present in the MCP21 fraction. The third aliquot was mock precipitated without the addition of a second antibody. A schematic representation of the experiment is given in figure 4.13, and the results are presented in figure 4.14.

Figure 4.13: Schematic of the reprecipitation experiment of figure 4.14
Figure 4.14: The mature forms of the LMPs are assembled into the proteasome approximately coincidentally with their processing. Panel 1: proteins eluted from MCP21 precipitates from native lysates. Lane 1 LCL721.174, no chase; lanes 2-5 LCL721, chase times are given above the lanes. Panel 2-4: proteins eluted from MCP21 precipitates were precleared with Staphylococcus A, split into three aliquots and reprecipitated with AK14 (panel 2), BNFl (panel 3) or mock precipitated with no second antibody (panel 4). Panel 2 shows reprecipitations with the anti-LMP2 serum AK14. Lanes are as for panel 1. The mature 22 kD LMP2 protein is incorporated into the MCP21 fraction in ~2 hours. Panel 3 shows reprecipitation with anti-LMP7 serum BNFl. The mature 23 kD LMP7 protein is incorporated within the labelling time (0 hours chase) and incorporation is complete by 2 hours chase. Panel 4 shows that any proteins that were precipitated in panels 2 and 3 were due to the added antiserum and were not artefacts caused, for example, by antibody remaining from the first precipitation.

A part of the lysates was not precipitated with MCP21 but was denatured and precipitated with anti-LMP2 and then anti-LMP7 to show the LMP proteins present in the total lysates. The LMP2 precipitates are shown in figure 4.12 as a demonstration of LMP2 processing. After the LMP2 precipitation, the supernatants were precleared and precipitated with anti-LMP7. This is shown in panel 5 (lanes as in panel 1). Figure 4.12 and panel 5 show that, at the first time point, the precursors of the LMPs were present in the lysates. However, these were not incorporated into the MCP21 fraction (panels 2 and 3).
Panel 1
MCP21 ppt

<table>
<thead>
<tr>
<th>cell line</th>
<th>.174</th>
<th>721</th>
</tr>
</thead>
<tbody>
<tr>
<td>chase time</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>5</td>
</tr>
<tr>
<td>lane no.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

- 30 kD
- 21.5 kD

Panel 2
MCP21 ppt, anti-LMP2 reppt

<table>
<thead>
<tr>
<th>cell line</th>
<th>.174</th>
<th>721</th>
</tr>
</thead>
<tbody>
<tr>
<td>chase time</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>5</td>
</tr>
<tr>
<td>lane no.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

- 30 kD
- 21.5 kD
Panel 3
MCP21 ppt, anti-LMP7 reppt

<table>
<thead>
<tr>
<th>cell line</th>
<th>.174</th>
<th>721</th>
</tr>
</thead>
<tbody>
<tr>
<td>chase time</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>lane no.</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

- 30 kD
- 21.5 kD

Panel 4
MCP21 ppt, no second antibody

<table>
<thead>
<tr>
<th>cell line</th>
<th>.174</th>
<th>721</th>
</tr>
</thead>
<tbody>
<tr>
<td>chase time</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>lane no.</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

- 30 kD
- 21.5 kD
Panel 5
anti-LMP7 ppts from denatured lysates

<table>
<thead>
<tr>
<th>cell line</th>
<th>.174</th>
<th>721</th>
</tr>
</thead>
<tbody>
<tr>
<td>chase time</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>lane no.</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

- 30 kD
- 21.5 kD
The experiment shows that only the mature forms of LMP2 and LMP7 are associated with the MCP21 subunit. There was no evidence for the immature forms of either of the LMP proteins in the MCP21 fraction even after prolonged exposure of the gel\(^2\). Thus, cleavage of the leader sequences occurs either before, or shortly after, association with the MCP21 subunit. As the experiment was done as a pulse chase, it can be seen that the association of the LMPs with the MCP21 subunit occurs (approximately) completely within 2 hours after LMP synthesis. The processing of the LMPs was also followed in this experiment (Fig. 4.12 for LMP2, Fig. 4.14 for LMP7). The appearance of the mature form of the LMPs parallels the appearance of LMPs in the MCP21 fraction, demonstrating that processing and incorporation of the LMPs into the proteasome are (near) coincident events.

The function of the leader sequences is not known. Many proteases are secreted in an inactive form and are activated in the gut by the proteolytic removal of a leader sequence (reviewed in Price and Stevens, 1989). It is possible that the proteasome subunits are synthesised with a leader sequence to prevent activation before assembly into the proteasome complex. The supposition that the \(\beta\) subunits make up the active sites of the proteasome would be consistent with this idea (Zwickl, Grziwa et al, 1992). Second, leader sequences are often important in protein targeting (reviewed in von Heijne, 1988). The proteasome has been found in many locations.

\(^2\) Recently, it has been shown that the LMP precursors assemble in a 13-16S complex before assembly into the 20S proteasome (Frentzel, Pesoldhurst et al, 1994). It is not known whether the MCP21 subunit is part of the 13-16S complex.
within the cell, including the cytoplasm, the nucleus and associated with the microsomal fraction (Rivett, 1989, Tanaka, Tamura et al, 1992, Yang, Waters et al, 1992). There is evidence that LMP7 may have two alternative leader sequences which would have different charges (see chapter 5). Expression of one or the other of these two leader sequences may direct LMP7 and associated subunits to different sub-cellular localisations. Alternatively, the leader sequences may determine with which proteins the LMPs associate, perhaps during the assembly of the proteasome.

**The mature forms of the LMPs are upregulated by interferon-γ**

More LMP complex could be precipitated (with a serum raised made between congenic mouse strains) after cells had been treated with IFN-γ (Monaco and McDevitt, 1986). The levels of mRNA for LMP2 and LMP7 are also increased by this cytokine (chapter 3). To determine whether the individual LMP subunits encoded by the human MHC were upregulated by interferon-γ, several cell lines were treated with the cytokine for 3 days. Western blots of the resulting cell lysates showed that both LMPs were upregulated at the protein level in the following cell lines: J6 (T cell), HeLa (cervical carcinoma), MOLT4 (T cell), U937 (macrophage), Bristol8 (B cell), HT1080 (fibrosarcoma) and JY (B cell). The induction of the LMPs in the B cell lines Bristol8 and JY was small due to the high constitutive levels in these cell lines (Fig. 4.15).
Figure 4.15: The LMPs are upregulated by interferon-γ. In even numbered lanes, cells were treated with 200 units/ml for 48 hours before lysis. Cells used are J6 (lanes 1 and 2), HeLa (lanes 3 and 4), MOLT4 (lanes 5 and 6), U937 (lanes 7 and 8), Bristol 8 (lanes 9 and 10), HT1080 (lanes 11 and 12) and JY (lanes 13 and 14). 20 μg of protein was loaded per lane as quantified with a protein assay kit (BioRad). The upper panel shows a western blot with the BNF1 anti-LMP7 serum at a dilution of 1: 500. There is a ~46 kD protein that is detected by the BNF1 serum in the T cell lines J6 and Molt4. The significance of this is unclear. The lower panel shows a western blot with the AK14 anti-LMP2 serum at a dilution of 1: 1000.
If the LMPs are involved in antigen processing for class I presentation, it would be predicted that the expression pattern of the LMPs might parallel surface expression of class I. To assess this, several cell lines used in the experiment shown in figure 4.15 were also surface stained for class I and, for comparison, class II. Cell surface expression of MHC class I was induced by interferon-γ treatment in HT1080, U937, MOLT4 and J6. The surface class I levels of Bristol8 and JY were constitutively high and were not induced after interferon-γ treatment. Class II (DRα and DRβ) was induced in HT1080 and U937 and was constitutively high in the two B cell lines. However, no DR expression was seen for MOLT4 or J6 with or without interferon-γ treatment (Fig. 4.16). The expression pattern of the LMPs is, therefore, more similar to class I surface expression than class II, despite their genomic position within the class II region. Additionally, X and Y boxes, found in the promoters of classical class II genes, are not present upstream of the TAP or LMP genes. This analysis is, therefore, consistent with a role for the LMPs in class I antigen processing.
Figure 4.16: Surface expression of class I parallels expression of the LMPs. HT1080, U937, MOLT4, J6, Bristol8 and JY from the experiment in figure 4.15 were surface stained for DRα (L243), DRβ (TAL14.1) and class I (W6/32).
Data from other groups suggests that upregulation of the LMPs has functional consequences. Treatment of HeLa cells with interferon-\(\gamma\) altered the composition of the proteasome as determined by 2D gel electrophoresis. Two of the spots that increased in intensity were identified as the LMPs (Yang, Waters et al, 1992). Thus, the increase in steady state LMP protein levels in the total lysates (Fig. 4.15) was accompanied by an increase in the incorporation of the LMPs into the proteasome complex. Additionally, the same quality of differences in proteasome activity were seen between LCL721 and .174 cells as were seen between resting and interferon-\(\gamma\) induced cells (Driscoll, Brown et al, 1993, Gaczynska, Rock et al, 1993). This implies that interferon-\(\gamma\) alters the proteasome activity by upregulation of the LMPs.

**LMP7-containing proteasomes are proteolytically active**

Immunoprecipitation experiments showed that the LMP7 antiserum recognised a protein that was part of the proteasome and that other subunits were coprecipitated. To determine whether these precipitates had proteolytic activity, we used the anti-LMP7 serum to precipitate the proteasome, collected the precipitate on protein A sepharose beads and washed the beads with detergent-free buffer. Chromogenic substrates were added to these precipitates and cleavage was assayed by fluorescence of the leaving group after cleavage. A positive control for the experiment was precipitation with MCP21, to assay activity in the total proteasome pool. A negative control was precipitation using the anti-LMP7 serum with LCL721.174 lysates. The results are shown in figure 4.17.
Figure 4.17: The LMP7-containing complex has proteolytic activity. Precipitates with both MCP21, for total proteasome, and BNFl, for LMP7 proteasome, were assayed for cleavage at the C terminus of phenylalanine (top panel), arginine (middle panel) and tyrosine (lower panel). In the top panel two concentrations of BNFl were used in the precipitation: 15 µl serum / 500 µl lysate (BNFl-15) and 25 µl serum / 500 µl lysate (BNFl-25). BNFl precipitations from the mutant line LCL721.174 had only background activity (blank values represent incubation of the substrate in reaction buffer but in the absence of a proteasome precipitate).
For all three substrates the anti-LMP7 precipitates from LCL721 have higher levels of fluorescence than the negative control (LCL721.174), showing that the LMP7-containing complex has proteolytic activity. The MCP21 precipitates have approximately half the activity from LCL721.174 lysates than from LCL721. We originally thought that this was due to an artefact of less proteasome in the LCL721.174 precipitations. However, it has now been shown that LCL721.174 proteasome has lower activity against hydrophobic and basic substrates than proteasome from the wild-type cell line LCL721, but higher activity against acidic substrates (Driscoll, Brown et al, 1993, Gaczynska, Rock et al, 1993). Our results are consistent with these data, but the lack of an acidic substrate in our assay meant that we could not conclude that the LMPs had an effect on the quality of peptide cleavage.

CONCLUSIONS

In this chapter I have described the characterisation of antisera raised against the MHC-encoded LMPs. Using these antisera in conjunction with anti-proteasome antibodies, it was possible to show that the LMPs are incorporated into the proteasome complex and that this complex has proteolytic activity. The absence of proteins of greater than ~30 kD from anti-LMP7 immunoprecipitates suggests that LMP7 is in the 20S complex.

---

3 The data shown here are the results of one experiment. The experiment has been repeated several times as part of a collaboration with Fred Goldberg and Maria Gaczynska (not shown) and the results presented here are representative of several sets of data.
and not the 26S complex, but this is not definitive and may be an artefact of immunoprecipitation.

Both the sequence comparisons presented in chapter 3 and the sizes of the LMPs recognised by the antisera suggested that the N termini of the proteins were not as shown in the cDNA-derived protein sequences. These discrepancies were resolved when the LMPs were precipitated from denatured lysates. I noted that both proteins were synthesised with N-terminal extensions which were not present in their mature forms. The temporal relationship of LMP leader sequence removal to incorporation into the proteasome was investigated by a reprecipitation experiment with a proteasome monoclonal antibody. This showed that cleavage of the leader sequences and association with the subunit defined by MCP21 occur approximately coincidentally. The function of the leader sequences is unknown, but they may be required in the regulation of proteasome assembly.

Coincident expression of the LMPs, TAPs and surface MHC class I was seen across a panel of cell lines and after IFN-γ treatment and this is suggestive evidence for all three classes of genes having related function. It would appear that incorporation of the LMPs into the proteasome after IFN-γ treatment may bias proteasome activity towards degradation products suitable for class I presentation (Gaczynska, Rock 1993, Driscoll, Brown et al, 1993 and Boes, Hengel et al, 1994).
ACKNOWLEDGEMENTS

The peptides used in the production of the anti-LMP sera were synthesised in the laboratory of Gerard Evan, ICRF, London. The AK14 antiserum was made by Adrian Kelly. The experiment shown in figure 4.7 to check the specificity of the serum on a western blot was also done by Adrian Kelly. MCP21 and the anti-proteasome serum were provided by Klaus Hendil, University of Copenhagen. The FACS profiles were done by Lesley-Anne Kerr. Proteolysis assays were done in collaboration with Enzo Cerondolo, Oxford.
CHAPTER 5: UNUSUAL GENOMIC ORGANISATION OF THE LMP7 LOCUS

SUMMARY

In the first part of this chapter, I describe the cloning of a second type of LMP7 cDNA. LMP7a (RING10) and LMP7b differ in their 5' exon. Both first exons have open reading frames that are in frame with the rest of the cDNA and would encode leader sequences with different charge. However, there was no evidence for differential regulation of LMP7a and LMP7b at the level of mRNA expression. A coding dimorphism was found in the first exon of LMP7b, although no coding polymorphisms were detected in the rest of the LMP7 (a or b) open reading frame.

The second part of the chapter describes the isolation of three transcripts from cDNA libraries that are transcribed from the opposite strand to LMP7 and TAP1. These are collectively known as RING9 and are alternative splice variants which do not have obvious open reading frames. All three RING9 clones are antisense to TAP1. One of these clones is also antisense to LMP7a.

In the final part of the chapter, I present sequence derived from the mouse genome at the equivalent of the TAP1/LMP7 loci. The mouse genomic sequence of the last intron of HAM1 (TAP1 homologue) and the HAM1: MC13 (LMP7b homologue) intergenic region was cloned and sequenced. Comparison of this sequence with the human sequence showed
that only one of the two *LMP7* first exons was conserved in mouse. There was limited conservation of sequence in the intronic or intergenic regions that code for the RING9 transcripts in human and no potential open reading frames are conserved. It is, therefore, unlikely that RING9 encodes a protein. A possible role for RING9 transcription in the regulation of LMP7 or TAP1 expression is discussed.

**INTRODUCTION**

Homologues to LMP7 have now been cloned in other mammals (Aki, Tamura et al, 1992, Frentzel, Graf et al, 1992). These are highly homologous to the human gene over much of their length. However, the first exon of the human LMP7 clone that was described in chapter 3 (RING10) was not related to the 5' end of either the mouse homologue (MC13) or the rat homologue (RC1). Inspection of the human genomic sequence showed that the first intron of RING10 contained sequence that was homologous to the 5' end of the mouse and rat cDNAs. I looked for further LMP7 cDNAs to determine whether this sequence was transcribed in human cells. It was apparent from the genomic sequence that the probe used to isolate RING10 (S6) would not have detected such transcripts were they to exist. Therefore, a probe derived from the 3' end of RING10 was used to screen for related clones.

Characterisation of the genomic structure of the *LMP7* locus, with reference to Beck, Kelly et al, 1992, facilitated a search for polymorphisms in the coding region using SSCP (single stranded conformational polymorphism). We considered it worthwhile to look for variation in the
LMPs for the following reasons. There are a large number of diseases that are associated with alleles of genes within the class II region. Whether the association of diseases with classical MHC genes is due to an effect of that gene product or of a genetically linked gene is unclear in most cases: for example, there is evidence of a link between resistance to type I diabetes and an aspartate residue at position 57 of DQB1 (Todd, Bell et al, 1987) but the identity of amino acid 57 of DQB1 does not explain every MHC haplotype association of type I diabetes (reviewed by Todd, 1990). Of all the MHC disease associations, only one has been assigned to a single gene: congenital adrenal hyperplasia is caused by a defective cytochrome P-450 gene called CYP21. The gene codes for steroid 21 hydroxylase and maps within the MHC class II region (White, New et al, 1984).

If the gene products involved in antigen processing were polymorphic, then the epitopes presented from a given antigen may differ, even through the same class I allele. The human MHC-linked transporters have already been shown to be polymorphic in their DNA sequence (Powis, Mockridge et al, 1991, Colonna, Bresnahan et al, 1992). There is strong evidence that different alleles, \(cim^a\) and \(cim^b\), of one of the rat MHC transporter genes alter the antigenicity of the rat class I RT1.A\(^a\) molecules as assayed by antibody binding and T cell reactivity (Livingstone, Powis et al, 1989, Livingstone, Powis et al, 1991). The profile of peptides eluted from RT1.A\(^a\) also differed according to whether the cell carried the \(cim^a\) or \(cim^b\) alleles (Powis, Deverson et al, 1992). The \(cim\) transporter functional polymorphism acts in an \textit{in vitro} assay at the level of peptide transport into the endoplasmic reticulum (Momburg, Roelse et al, 1994). In addition, the ability to express fragments of flu nucleoprotein in human cell lines through
normal B2702 alleles was genetically determined and MHC-linked in a family study, reminiscent of the cim phenotypes in rat (Pazmany, Rowland-Jones et al, 1992). However, restricted peptide transport for any of the human TAP allelic gene products has not been shown.

The LMPs were mapped in mouse on the basis of an antigenic polymorphism: that is, antisera raised in mice recombinant only in their MHC could precipitate the LMP complex (Monaco and McDevitt, 1982). Thus, there was evidence for both sequence polymorphism within the mouse LMPs and, for B27 presentation of flu epitopes, functional polymorphism mapping to the MHC in humans. To investigate whether the LMPs were polymorphic in humans and whether this could explain any MHC disease associations or class I presentation polymorphisms, I looked for polymorphism within the LMP7 gene using SSCP and solid phase sequencing.

In the course of screening the U10 and U15 cosmids for new genes, I isolated a clone called NS1C1.1, which was mentioned in chapter 3. Sequence from NS1C1.1 overlapped with both LMP7 (RING10) and TAP1 cDNA sequences. In a search for other clones from this locus we isolated a total of three types of cDNAs that were transcribed from the opposite strand to LMP7 and TAP1. There is some precedent for transcription of overlapping mRNA molecules in the MHC at the CYP21B (steroid 21 hydroxylase) locus within the class III region (Morel, Bristow et al, 1989). Other examples of antisense transcription have been found elsewhere in the genome: the c-erbAα locus may be regulated by transcription of Rev-ErbAα from the other strand and it is possible that this mechanism determines
alternative splicing events that lead to the translation of either a thyroid hormone receptor (α1) or a variant that does not bind thyroid hormone (α2) (Lazar, Hodin et al, 1990).

Although the CYP21B and Rev-ErbAα opposite strand genes contain open reading frames, other examples of antisense RNAs exist in eukaryotic organisms which have no open reading frame but still have a physiological effect through their effect on the translation of the complementary RNA. The EB4-PSV gene in Dictyostelium discoideum encodes a protein, but also gives rise to an opposite strand transcript with no long coding sequence. The two transcripts are expressed in an antiparallel manner. Inhibition of RNA synthesis stabilises the sense RNA molecule, implying that its normal degradation is dependent on another RNA molecule, probably that encoded on the other strand (Hildebrandt and Nellen, 1992). The production of mRNA from the C. elegans lin-14 locus is inhibited by transcription of small RNA molecules from the lin-4 locus. The lin-4 RNA molecules are transcribed in trans and act to destabilise lin-14 RNA through the (lin-14) 3' UTR (Lee, Feinbaum et al, 1993).

Finally, there are a number of examples of RNA molecules having functions which require neither their previous translation into protein nor an effect on other RNA molecules. These include the tRNA family and components of the ribosome. Recently, a gene called H19 has been shown to have tumor suppressing activity upon transfection. However, H19 has only short open reading frames that are not conserved between species. Presumably, therefore, H19 acts as an RNA molecule rather than a protein (Hao, Crenshaw et al, 1993).
The RING9 clones described in this chapter do not have an obvious open reading frame. To investigate if any of the small open reading frames were conserved we decided to sequence the mouse cosmid 10.13 over this region (Steinmetz, Stephan et al, 1986). There is striking synteny between the human and mouse genomes in the class II region (Hanson and Trowsdale, 1991). This suggests that any stretch of sequence, either coding or regulatory, should be conserved between the species. In addition, this approach was used to investigate whether the first exon of \textit{LMP7a} was conserved in the mouse and whether any regulatory regions of DNA sequence could be defined through conservation of non-coding sequence.

\textbf{RESULTS and DISCUSSION}

\textbf{Isolation of two different LMP7 transcripts}

A probe consisting of exons two through six of the LMP7 cDNA (chapter 3) was used to probe a B cell cDNA library. The sequence of one clone matched the 3' end of the LMP7 sequence. The other end of the clone had no homology with the LMP7 cDNA sequence but matched genomic sequence in the first intron (Beck, Kelly et al, 1992). The RING10 clone will be referred to as LMP7a and the related clone as LMP7b. Comparison of the LMP7a and LMP7b cDNAs with genomic sequence showed that both cDNAs have six exons (Fig. 5.1a). They are identical over exons 2-6, but differ in their first exons. Each of the alternative first exons has a potential initiating methionine codon followed by an open reading frame. The first exon of \textit{LMP7a} encodes a protein sequence of 45 amino acids.
Figure 5.1a: Genomic organisation of LMP7. Exons 2 through 6 are shown as thick, solid lines; the alternative exons of LMP7a and b are shown as thick, dashed lines; the introns are shown as thin lines between exons. Numbering is given as bp from the Sall site (see Fig. 5.1b). The positions of the potential initiating methionine codons in the first exons of LMP7a and LMP7b are shown, as is the second in-frame methionine codon in the shared second exon (bracketed). The stop codon at position 3784 bp is shown as *.

The homology with N-terminal amino acid sequence from proteasome components, marked as ---, corresponds to bp 1725-1802. The start and end points of the exons are numbered. Two possible interferon-stimulated response elements (ISREs) within this sequence are marked by arrows. The most 5' is shown in figure 5.1b and the sequence of the ISRE within the fourth intron is AGTTTGTTTTTCC at bp 2833-2845.

<table>
<thead>
<tr>
<th>Sall</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>112</td>
<td>332</td>
<td>466</td>
<td>751</td>
<td>740</td>
</tr>
<tr>
<td>1657</td>
<td>1715</td>
<td>1805</td>
<td>2075</td>
<td>1664</td>
</tr>
<tr>
<td>2484</td>
<td>2613</td>
<td>3216</td>
<td>3700</td>
<td>4029</td>
</tr>
</tbody>
</table>

---

LMP7a

LMP7b

M

(M)

ISRE

ISRE

*
Figure 5.1b: Genomic sequence 900 bp telomeric of the SaII site between TAP1 and LMP7. The cDNA sequences of the LMP7a and b alternative first exons are underlined with a dashed line. The ends of both exons are marked ">" at the splice junction. The derived protein sequences from each exon are shown in one letter code above the DNA sequence. A possible ISRE and a possible J box upstream of the start of the LMP7b cDNA are underlined (solid line) and a GAS consensus sequence is marked with dots. The polymorphic nucleotides within the first exon of LMP7b are shown below the sequence, the resulting amino acid dimorphism is also shown (Q->K).

LMP7a

1 GTCGACAGTTGCTGGGTAGATGAGGCCAACACAGGTTGCAAGAAGAGGCGGGGTTTAGAGGCGTGAAACTCCGCAGTGCTCAGCCAAGCAGGGAGCAACG

201 CAGAGACTGCGCCCTCTCCCTGGCGCCGCCTTCCCACACGCGGCGGGTATATTCTGTTGCAGTTGGCCCAGGACCTGTTTCCAAGACTCTGCCCTC

LMP7b

101 CTAGGAAGGCCGGCAAGAAAAGGCCACGCTCTTGTGGGTGACTACAGGTTAGGAGACCGTTGAACCTGGAGGGGCCCTAGGATGGACCCCGTGGAAAGATT

301 GCACCTCCGTCCCTTGGTCTTTTGTAAAGTGATGCTCATAGGAACCCCCACCCCGCGTGACACTACTCCCAGCTCCTGGCTGACTTCTAGTCTTCTGGTT

J box

401 GAAGCTGCGCCTTTAGATGACACGACCCTACCCACCCCTGTTTCCAGCGGATGCCCGGGCCTGGAGGTACCTCTTACTGTAACCCATCGCCAAGTGGCT

501 TTTGAAGGCCCGCTTCTCTTTCTCGTCTAAGCCGAGCCGCCGCCGCCGCCACACCCGACAAGAGGCCAGAGAAGGCCGGCTCCCAAGTG

ISRE

601 TGATGGCTCTGGTCAGGCATACCTGTCTTTTCTCGGAAAAGCGCAGGGGATGTGGAAAAGAGTCTTGTTCCCTCCCCTTCGATCTGTGGCTTTCGCT

GAS

701 TTCACTCTCTCTGAGCCAGGACACATCTCTGGGTGCTGGGCGGTCATGGCGCTACTAGATGTATGCGGAGCCCCCCGAGGGCAGCGGCCGGAATCGG

LMPTb

K

801 CTCTCCCGGTTGCGGGAAGCGGGCGTCGCTCGGACCCAGGACACTACAGTTCTTCTATCGATCTCGCAAGCTCGCTTTTACCAGGGGATGACCGTCTGG
Figure 5.1c: The genomic sequence 5' of the start of the LMP7b cDNA contains a region with homology to the interferon-γ promoter region of the guanylate binding protein (GBP) gene. Conserved regions are shown as bold characters, gaps in the alignment are shown by dots. The ISRE region was defined by homology with interferon-α inducible genes and the GAS sequence (for gamma-interferon activation site) in the GBP promoter was defined by exonuclease III protection, competition for DNA binding proteins with a range of mutated oligonucleotides, and CAT reporter constructs (Decker et al., 1991). The extent of the ISRE and GAS regions is shown by dashes above the GBP sequence.

| GAS | - - - - - - - - - - |
| ISRE | - - - - - - - - - - |
| GBP promoter | C T T T C A G T T T C A T A T T A C T C T |
| human LMP7 promoter | C T T T C G C T T T C A C . T T C C T C C |
| mouse LMP7 promoter | C T T T C G C T T T C A C . T T C C T T G |
with a pI of 3.3 and Mr 4.7 kD. The protein encoded by the first exon of LMP7b is similar in molecular weight (49 amino acids, Mr 5.2kD), but is very much more basic, with a pI of 10.5 (Fig. 5.1b). For both transcripts, the open reading frame extends through the five shared exons until the stop codon in exon 6.

The derived protein sequences of LMP7a, LMP7b and the mouse homologue MC13 were analysed by sequence alignment (not shown) and diagon plots (Fig. 5.2). The protein sequence encoded by the first exon of LMP7a (amino acids 1-45) had no homology to MC13, even at low stringency (Fig. 5.2a and Frentzel, Graf et al, 1992). The derived protein sequence from the LMP7b open reading frame was similar to MC13 throughout its length (Fig. 5.2c). However, a diagon plot of MC13 and LMP7b at higher stringency showed that the homology between the two derived protein sequences was weaker for the N-terminal ~70 amino acids (~70% identity) than for the rest of the sequence (~90% identity) (Fig. 5.2d).

If both the LMP7a and LMP7b first exons are translated then the two LMP7 proteins would differ by either an acidic (LMP7a) or basic (LMP7b) N-terminal leader sequence. There is no published mouse or rat equivalent of the LMP7a cDNA, but a comparison between MC13 and LMP7b showed that the N-terminal leader sequence has not been as conserved as the rest of the sequence. Perhaps evolutionary conservation of secondary structure, charge distribution and hydrophobicity at the N terminus is of importance rather than the primary sequence. It is possible that the different charges of the two leader sequences target the proteasome to different localisations, or to interact with different proteins. If this were the case, then it might be
Figure 5.2: A comparison of the derived protein sequences of LMP7a, LMP7b and MC13 by diagonal analysis. The MC13 sequence is compared to both LMP7a (panels a and b) and LMP7b (panels c and d) at low (panels a and c) and high (panels b and d) stringency. LMP7a has no homology to MC13 at the N-terminal end (panel a). LMP7b and MC13 share homology over the entire protein sequence (panel c) but this homology is weaker over the N-terminal 70 amino acids (panel d).
expected that LMP7a and LMP7b mRNA molecules would have different expression patterns.

**Messenger RNA expression levels of the alternative splice products**

Our original experiments with the RING10 cDNA (chapter 3) showed that LMP7 mRNA is expressed in a B cell line, a T cell line and is upregulated by interferon-γ. However, the probe used for this experiment would not have discriminated between LMP7a and LMP7b expression. Using probes specific to each of the alternative first exons, the expression levels of LMP7a and LMP7b were investigated in several cell lines. Both transcripts were expressed in the T cell line CEM, the B cell line JY, the mature T cell lymphoma line HUT78 and the cervical carcinoma cell line HeLa (data not shown). The effect of IFN-γ on the expression of each of the transcripts was also studied (Fig. 5.3). Both the LMP7a and LMP7b mRNAs were upregulated by IFN-γ treatment in the B cell line Bristol8, the macrophage cell line U937 and the fibrosarcoma cell line HT1080. In this experiment and in others (not shown) I found no conclusive evidence for differential regulation of the two LMP7 transcripts. However, there is some evidence that the steady state levels of proteasome subunits do not mirror the levels of their respective mRNAs (Shimbara, Orino et al, 1992 and Belich, Glynne et al, submitted). Thus, the two predicted LMP7 proteins might have differing expression which is regulated at some post-transcriptional step. The BNFl antiserum would recognise both LMP7a and LMP7b proteins and these could probably be distinguished by isoelectric focusing. Such an analysis was not done because of the lack of conservation of the LMP7a open reading frame in the mouse (see below).
Figure 5.3: Northern analysis of LMP7a and LMP7b. Total RNA samples from the B cell line Bristol8 (lanes 1 and 2), the macrophage cell line U937 (lanes 3 and 4) and the fibrosarcoma cell line HT1080 (lanes 5 and 6) both before (lanes 1, 3 and 5) and after (lanes 2, 4 and 6) treatment with 200 U/ml interferon-γ for 48 hours were probed with LMP7a, LMP7b, DQB and GAPDH. The DQB panel shows IFN-γ induction of U937 and HT1080 and the GAPDH panel shows approximately even loading in all tracks. The low signal in lane 5, panel 2 was an artefact that was not observed when the experiment was repeated: LMP7b is constitutively expressed in HT1080 cells with little induction by IFN-γ.

Exposure times for northern blots probed with LMP7a were generally 4–5 times longer than for LMP7b for comparable signals.
Figure 5.3: Northern analysis of LMP7a and LMP7b
Two sequences with homology to a consensus for an interferon-stimulated response element (ISRE) (Beck, Kelly et al, 1992, Friedman and Stark, 1985) may mediate interferon induction of transcription from this locus. One of these is 40 bp upstream of the start of the LMP7b cDNA, the other in the fourth intron (Fig. 5.1a and b). A GAS sequence (interferon-γ activating sequence) has been defined in the promoter of the GBP gene (a guanylate binding protein) and overlapped with the ISRE (Decker, Lew et al, 1991). A stretch of sequence was found within 50 bp of the 5' end of the LMP7b cDNA which had homology to this ISRE / GAS element from the GBP promoter and is shown as an alignment in figure 5.1c. A sequence which matches a consensus for a J box, shown to be important for the upregulation of MHC class II genes by IFN-γ (Sugawara, Ponath et al, 1991) was found 190 bp upstream of the start of the LMP7b cDNA (Fig. 5.1b). Neither ISRE, GAS nor J box consensus sequences were found within 1kb 5' of the LMP7a cDNA start.

A coding polymorphism within the first exon of LMP7b

Single stranded conformational polymorphism (SSCP) was used to determine exons with possible coding differences. The technique depends on the fact that denatured strands of DNA will run independently on a "renaturing" polyacrylamide gel with mobilities extremely sensitive to DNA sequence. Any SSCP shifts were followed up by solid-phase sequencing of the relevant exons. Primers were used to amplify the RING10 exons from the genomic DNA of a panel of 24 homozygous typing cells (HTCs). The HTCs were chosen to reflect a wide range of ethnic backgrounds and HLA haplotypes. PCR products were 500-1000 bp in length, and were made
radioactive by the inclusion of $^{32}$P dCTP in the reaction mixture. Each PCR product was digested, in separate reactions, with two frequently cutting enzymes and each of the products were run under two electrophoretic conditions. Such a protocol should detect almost all DNA sequence differences in the 24 samples (Hayashi, 1991).

The SSCP procedure showed several possible DNA sequence differences, which were investigated further. Sequencing of these exons was carried out in solid phase using a biotinylated PCR primer and streptavidin-coated magnetic beads (DYNAL). On average, sequences of approximately six DNA samples were obtained from each exon. Only one exonic difference was found in one Japanese cell line (KOSE) and this was non-coding (G -> A at bp 691, numbering of figure 3.2). In this study, no other DNA sequence polymorphism of the RING10 gene was detected.

After the LMPTb cDNA was cloned, I carried out a similar analysis on the alternative first exon. Two more changes within the coding region were found. A T -> C homozygous change at bp 877 (numbering of figure 5.1b) was found in one of eleven cell lines sequenced (Madura-T) and was found as a heterozygous change in one other (CBL). This change is non-coding. A second change, C -> A at bp 893 (numbering of figure 5.1b) was found in 3 cell lines as a homozygotic change and as a heterozygotic change in 3 others from the eleven cell lines that were sequenced$^1$. This change alters the last

---

1 It should be noted that the cell lines selected for sequencing were biased towards those which showed SSCP band shifts. Together with the fact that the HTC panel is a rather small and distorted collection of different ethnic groups, the frequencies given above for the
codon of the first exon of LMP7b from glutamine to lysine. Thus, only limited polymorphic changes within the LMP7 open reading frame were found. The only coding change was found within the leader sequence of LMP7b. This is unlikely to have any functional effect because this region of the protein is cleaved off before or shortly after incorporation of LMP7 into the proteasome (chapter 4).

Several studies of disease populations have been made with reference to the human TAP2 polymorphisms (for example Powis, Rosenberg et al, 1992, Pile, Burney et al, 1993). No linkage with disease has been found with the TAP2 alleles over and above that for the already characterised association with classical class II antigens. If changes in the LMP7 gene product were disease factors, then association of disease to the closely linked TAP2 alleles may have been observed in these studies. As LMP7 is closely linked to TAP2 and has only limited polymorphism, it is unlikely that hereditary changes in the LMP7 gene affect the relative risk of developing these MHC-linked diseases. It is now known that the polymorphism that gives rise to the antigenicity of the LMP complex in the mouse is due to an asparagine to aspartate change at amino acid 177 in LMP2 (Zhou, Cao et al, 1993). The human LMP2 gene shows one polymorphism, resulting in a histidine to arginine change at aa 60. For the same arguments, it is also unlikely that occurrence of polymorphisms are not representative. In order to assess the relative frequencies of the polymorphisms described in the first exon of LMP7b, a population study of a wider panel of genomic DNA samples is now in progress (M. Goldsworthy and S. H. Powis).
LMP2 polymorphisms predispose to any of the diseases studied for linkage with TAP2.

Three types of transcripts from the opposite strand to LMP7 and TAP1

When the NS1 probe was used to identify cDNAs from the CEM library, clone NS1C1.1 was isolated. This was sequenced and found to have an insert of 1812 bp, a poly A addition consensus sequence (AATAAA) and a poly A tail. The sequence of NS1C1.1 is shown in figure 5.4 as a part of the genomic sequence of this region. Translation of the insert revealed two potential open reading frames. The 5' end of the cDNA had an open reading frame of 153 amino acids. This did not start with a methionine codon, but the genomic sequence showed that there was a potential initiation codon ~150 bp upstream of the cDNA start, so that there is an open reading frame between bases -140 to +459 (numbering from the start of the cDNA). Translation would give rise to a protein of 200 amino acids. At the end of this open reading frame were stop codons in all three frames (between base pairs 460 and 530), followed by another open reading frame coding for a further 167 amino acids. Comparison with the genomic sequence showed that there was one intron in the NS1C1.1 clone, between base pairs 735 and 736. The splice sequence fitted with the splice consensus sequence - exon:GT...intron...AG:exon. The NS1C1.1 clone was transcribed from the other strand to both TAP1 and LMP7 and overlapped with them (Fig. 5.4, 5.5). Therefore, there was some potential for the control of TAP1 and LMP7 gene expression by an antisense mechanism.
Figure 5.4: Schematic representation of the RING9 splice patterns. The three clones, NS1C1.1, S6C4.1 and S6J8.2, are shown relative to the alternative first exons of \textit{LMP7a} and \textit{LMP7b} and the last exon of \textit{TAPI}. Exons are marked in horizontal solid lines, introns in diagonal dashed lines. The approximate positions of the \textit{NotI} and \textit{SalI} sites in the genomic DNA are also shown.
Figure 3.5. Human sequence from the penultimate exon of TAP1 through to the first exon of LMP7b. Coding exons are shaded grey. RING9 cDNA sequences are underlined and are transcribed from the other DNA strand: NSC1.1 is shown in red, S6C1.2 is shown in green and S6C4.1 is shown in blue. Poly A addition consensus sequences are shown as yellow boxes (AATAAA), the stop codon of TAP1 is marked with an asterisk and coloured red and the initiation codons of LMP7a and LMP7b are marked with M.
It was possible that NS1C.1 was derived from an aberrantly spliced or partially spliced transcript and that there were other transcripts in which the stop codons between the two open reading frames were spliced out. I looked for other cDNAs with a probe, called S6, consisting of the 5' 604 bp of the NS1C1.1 cDNA which was excised with a SalI / XhoI double digest. With the S6 probe, I isolated several more clones from the CEM library. One of these, S6C1.1, was the RING10 clone described in chapter 3. Another clone, S6C4.1, was similar to NS1C1.1. However, there was a different splice pattern (Fig. 5.4 and 5.5). S6C4.1 did not have any longer open reading frames than NS1C1.1. There were two introns within the S6C4.1 sequence, the 3' intron was identical to that within the NS1C1.1 clone. The 5' intron was unique to S6C4.1 and contained the first exon of LMP7a, with only a few base pairs on either side. The 3' end of S6C4.1 overlapped with the 3' end of TAP1. If complementary RNA molecules did inhibit the expression of TAP1 and LMP7, then differential splicing to form RNA molecules of NS1C1.1-type would inhibit both TAP1 and LMP7a expression, whereas S6C4.1-type RNA would inhibit only TAP1 expression.

To exclude the possibility that the NS1C1.1 and S6C4.1 clones were artefacts of the CEM library, we probed a JY library with the S6 probe. A series of clones were isolated. Two of these had the same splicing pattern as S6C4.1, one had the same splicing pattern as NS1C1.1 and one, S6J8.2, had a new splicing pattern (Fig. 5.4 and 5.5) with one intron. The 5' end of this intron differed from the 5' end of the introns in either NS1C1.1 or S6C4.1, but the 3' end was identical to the intron of NS1C1.1 and the second intron of S6C4.1. There was no obvious open reading frame within S6J8.2.
In summary, three splice variants transcribed from the opposite strand to TAP1 and LMP7 were isolated which were typified by NS1C1.1, S6C4.1 and S6J8.2 and collectively called RING9. RING9 clones have been isolated from two cDNA libraries, CEM and JY. NS1C1.1 and S6J8.2 were polyadenylated at the same residue, all clones had similar 5' ends and all shared at least one splice junction. From the number of positives obtained in cDNA library screens, it can be estimated that RING9 clones are present at a frequency of ~1 x 10^-4 cDNA clones in lymphoblastoid cell lines. All introns in all RING9 clones found begin with the bases AG and end with GT. Additionally, there is a poly A addition consensus site within 20 bp of the 3' end and most RING9 clones are polyadenylated. Finally, RING9 is not flanked by repeats in the genomic DNA. The above arguments mean that it is unlikely that RING9 is a processed pseudogene.

None of the RING9 clones have a convincing open reading frame. Instead, the splicing pattern of S6C4.1, which splices around the first intron of LMP7a, suggests a regulatory role. Alternatively, it is possible that RING9 clones are the product of aberrant transcription from the LMP7 promoter, which may have bidirectional transcription elements. To attempt to distinguish between these possibilities, a single stranded unique RING9 probe was used to examine the RNA expression pattern and the mouse sequence was obtained from this region to look for sequence conservation.
RING9 RNA is upregulated by interferon-γ

A region of the RING9 sequence that is common to all three splice patterns was amplified by PCR. The 5' primer was biotinylated so that the "sense" strand of the PCR product could be separated as single stranded DNA with streptavidin magnetic beads. The complementary strand was labelled using a specific oligonucleotide for the priming reaction.

To show that the probe recognised a specific RNA, a northern blot with total RNA from the cell lines LCL721 and LCL721.174 (with a deletion over the RING9 locus) was used. Two bands of ~3.5 and 1.5 kb were seen which were specific to the LCL721 lanes (Fig. 5.6, left hand panel). It is unlikely that these bands are caused by contamination of the probe with sequence that would hybridise to either TAP1 or LMP7 mRNA for two reasons: the probe is from an intron of TAP1 and is not repeated within the LMP7 sequence; the RING9 coding strand was separated from the TAP1 strand and primed with a specific primer, therefore the labelled DNA would not hybridise to unspliced TAP1 RNA. It may be of note that the RING9-specific RNA sizes are not consistent with the sizes of the cDNA inserts described above. Especially, the presence of a 3.5 kb band raises the possibility that the RING9 clones isolated are not full length and may be truncated at the 5' end. An alternative explanation for the presence of this band would be that it represents RING9 transcripts which have not been spliced. The genomic sequence that the RING9 clones cover is only 2.3 kb, making this explanation unlikely. However, accurate sizing of the RNAs was not possible due to low resolution of the gel and a lack of suitable size markers.
Figure 5.6: Northern analysis of RING9 expression. Left hand panel: a single-stranded, RING9 specific probe detected two RNAs of approximate sizes 3.5 and 1.5 kb that were specific to LCL721 (lanes 1 and 3) and not seen in LCL721.174 (lanes 2 and 4). The steady state levels of these RNAs were increased by treatment of colon carcinoma cells with 250 units/ml interferon-γ for two days (right hand panel). Cell lines were CC20 (lanes 1 and 2) and SW620 (lanes 3 and 4) with interferon-γ treatment (lanes 1 and 3) and without (lanes 2 and 4).
If RING9 was acting to downregulate TAP1 or LMP7 expression, it would be expected that RING9 RNA would have a reciprocal expression pattern to TAP1 or LMP7 RNA. For example, such a reciprocal relationship is seen for the lin-4 and lin-14 RNA molecules in C. elegans (Lee, Feinbaum et al, 1993). To test this, northern blots with RNA from interferon-γ induced cell lines were used. RING9 RNA was upregulated by this cytokine (Fig. 5.6, left hand panel) as was TAP1 (Trowsdale, Hanson et al, 1990) and LMP7 RNA (Glynne, Powis et al, 1991). This is not consistent with an antisense mechanism for RING9 regulation of the genes on the opposite strand.

A comparison of the human and mouse genomic DNA sequences at the TAP1/ LMP7 locus

I analysed the corresponding region of the mouse genome to investigate whether any open reading frames of the RING9 locus were conserved. Initially, the probe used for RING9 RNA expression on RNA blots was used on Southern blots of mouse cosmid DNA from this region. A specific EcoRI band of ~3 kb was detected from a digest of cosmid 5.9 (Steinmetz, Stephan et al, 1986). Southern blotting of further restriction digests of this cosmid with a range of restriction enzymes showed that the RING9 conserved sequence mapped between cross-reacting sequences for TAP1 exon 10 and LMP7b exon 1 (not shown).

To examine the extent of this sequence conservation, I then cloned and sequenced the genomic DNA between HAM1 and MC13 (the mouse TAP1 and LMP7 homologues respectively) and between the penultimate exons of HAM1. Assuming that the position of the last intron would be
conserved between mouse and human TAP1 genes, primers were designed to match the HAM1 sequence (Monaco, Cho et al, 1990) which would be predicted to amplify the last HAM1 intron in a PCR reaction. Primers between the 3' end of HAM1 and the 5' end of the mouse MC13 gene were also synthesised. All the primers were designed to contain restriction enzyme sites. The PCR products were then cleaved and cloned into Bluescript. The intergenic region between HAM1 and MC13 was cloned in this way (BSHM1). Cloning of the last intron of HAM1 was complicated by the presence of a BamH1 site within the PCR product. Part of this intron was cloned into Bluescript (BSHH1). The other part (CAX1) was cloned using CAU and CUA tailed primers with the ClonAmp kit (Gibco BRL).

The inserts of BSHM1, BSHH1 and CAX1 were sequenced by progressive walking with synthesised oligonucleotides. Most of the sequence was determined from both strands. The sequences were assembled with sequence from the last two (presumed) exons of HAM1 and the first (presumed) exon of MC13 to give a final stretch of mouse sequence spanning 3.7 kb. This was compared to sequence from the equivalent human region (Beck, Kelly et al, 1992 and accompanying database submission).

To determine which sequences were conserved, the GCG and Staden packages were used and the sequences compared with a diag plot (Fig. 5.7) and a similarity plot (Fig. 5.8). The former detects both on- and off-diagonal homology, the latter shows the extent of homology against position in the optimal alignment of the two sequences. Both methods of sequence comparison detected the conserved coding sequences of TAP1 exons 10 and
Examination of the mouse sequence showed that the potential open reading frames of human RING9 are not conserved between the species. This result means that it is very unlikely that RING9 mRNA is translated. As discussed above, a regulatory role for RING9 in TAP1/LMP7a expression is unlikely as the three transcripts have similar expression patterns. The data presented here would be consistent with aberrant transcription from the LMP7b promoter, which may be bi-directional. Alternatively, the RING9 gene may once have encoded a protein but is no longer active and has diverged through speciation between human and mouse with the accumulation of stop codons.

The first exon of LMP7a was conserved to ~60%, but there was not an open reading frame that would be in frame with exons 2-6 of MC13. This has been previously noted (Meinhardt, Graf et al, 1993). In fact, there was greater conservation between 100 and 400 bp 5' of the LMP7a open reading frame (72% identity) than the open reading frame itself (sequence 6, figure 5.9). It has been shown that, in a mouse cell line, LMP7 is synthesised as a precursor with a pI consistent with translation of the open reading frame of MC13 as published (equivalent of human LMP7b).
Figure 5.7: The genomic sequences from mouse and human were compared with the Compare programme in the GCG package with window=21 and stringency=16. Exons and coding regions of the human and mouse transporter genes, \textit{TAPI} and \textit{HAM1}, together with those of the proteasome genes, \textit{LMP7a} /\textit{b} and \textit{MC13} are shown with yellow lines. Sequences that are not coding in the mouse, but are conserved with the human sequence, are boxed in red and numbers refer to the alignments in figure 5.9.
Figure 5.8: The genomic sequence of human TAP1 exon 10 to LMP7b exon 1 was aligned with the genomic sequence of mouse HAM1 exon 10 to MC13 exon 1 using the Pileup programme in the GCG package with gap weight=5 and gap length weight=0.3. The similarity between the gapped sequences was determined with the plotsimilarity programme, window=100. The positions of the TAP1, HAM1, LMP7 and MC13 exons and coding regions are shown by dashed lines and conserved coding regions are shaded.
Figure 5.9: Alignments of conserved, non-coding regions in the mouse (top) and human (bottom) genomic sequences between TAPI / HAM1 exon 10 and LMP7b / MC13 exon 1. Alignments were made with the GAP programme, GCG package with gap weight=5 and gap length=0.3. Homologous regions were determined from the diagonal and similarity plots (figures 5.7 and 5.8).

Conserved sequence 1, percent identity: 100
178 ATCTTGCT 183
176 ATCTTGCT 183

Conserved sequence 2, percent identity: 67
293 TCTCCTGAGTCTCATTTCCATGAGGTGTCTCTGCTCTCCTCTCT 342
151 .CTTCCTAAATGCTATTCTCCACCACCACCCGCTTTTATACATCTT 198
343 GTGTAGTATACCAACACAGAATCTGTTCAAACTGTCTTGGAACTCT 392
199 CTGTTAGTTTTACTAATCTATAATTATACAAACGAGCTCTGAGTTCT 248
393 TGTTCCA 400
249 CAGTCC 255

Conserved sequence 3, percent identity: 76
541 ATCTCTAGATCGACAAAAATGAGGCTAGGAGAGGTTGATTAACCTTCT.. 588
963 ATCTCATGAGATCCAAAATGAGGTTTAAAGAAGTTGATATACCTGCTCA 1012
589 ....AAAGTGCAACGC.TTTGAGCAGTGGAGCTGGGTCC 624
1013 TATAGGATCATAATAGCTTTTGTGAGGGAGCGACCTTGACCC 1053

Conserved sequence 4, percent identity: 92
1667 GTCCCCAGCCCTGGAAAGCAGGGCT.CCCTGACTG 1702
1342 GTCCCCAGCCCTGGAAACACAGGTGTCTCCCTGGGCTG 1378

Conserved sequence 5, percent identity: 75
2305 TAACCAAAGAGCATAGCGCTTTGCTTCTATGGGTTTATATT 2354
1924 TAAGCTAGACATATCTGCGCATTAATAACCCGCTTGAGTTCTGATATT 1973
2355 TGTAATAAATCTGTGTCTTCTGCCTCGGGTTC 2396
1974 TATAAATAAATGTTGTCTTCTGATATTGATTTTCTATGGTTTC 2015
The lack of conservation of LMP7a in the mouse could mean either that the LMP7a gene product has a human specific function or that LMP7a maps elsewhere in the mouse genome. Otherwise, it would suggest that LMP7a is not translated, or has no function, in humans. LMP7a clones have been isolated from three cDNA libraries from different tissues and are unlikely to be library artefacts. Possible functions for LMP7a were discussed in chapter 4 but the lack of conservation of this exon in mouse casts some doubt on its physiological relevance. For this reason, no further attempt was made to characterise the expression patterns of LMP7a and LMP7b at the protein level.

The diagon and similarity plots show conservation of coding exons clearly. In addition, there were several stretches of sequence that were conserved between mouse and human, but which were not coding in the mouse, and these are shown as alignments in figure 5.9. The sequences correspond to the areas shown as red boxes in the diagon plot of figure 5.7. These regions may be functional, for example in transcription control, or they could be the remnants of a once expressed gene. The MacPattern programme was used to search human sequences 1-8 (Fig. 5.9) for transcription consensus sites. The results of this search were difficult to interpret as the redundancy of the consensus sites was generally rather high, leading to a large number of "hits" of uncertain significance. However, a conserved interferon-γ response element was found within sequence 7, which may be relevant in the upreglation of LMP7b and RING9 mRNA by interferon-γ.
CONCLUSIONS

In this chapter I have described a number of cDNAs that are transcribed from the \textit{LMP7} locus, but which differ from the original LMP7 clone described in chapter 3. At least in humans, there appears to be two alternative transcriptional start sites for LMP7, but the RNA is then spliced to identical second exons. To investigate any functional significance of this, I looked for differential expression of these transcripts but found that both transcripts had similar expression patterns. Furthermore, sequencing of a mouse cosmid showed only low sequence conservation of the first exon of \textit{LMP7a}, although the first exon of \textit{LMP7b} was well conserved. \textit{LMP7b} specific probes consistently gave stronger signals on northern blots than \textit{LMP7a}, suggesting that \textit{LMP7a} is only a minor form of \textit{LMP7} mRNA.

Whether \textit{LMP7a} is translated in human cells cannot be decided from the data presented in chapter 4 as the anti-\textit{LMP7} serum would be predicted to recognise both proteins and they would have similar molecular weights. Isoelectric focusing would resolve this, and this has been done for a mouse cell line (Fruh, Yang et al, 1992). In this case, there was no evidence for \textit{LMP7a} translation. However, it is not possible to extrapolate this result to human cells as there is no open reading frame for an \textit{LMP7a} protein at the mouse \textit{LMP7} locus (although \textit{LMP7a} may be coded elsewhere in the mouse genome).

To extend the analysis of the \textit{LMP7} locus, I used DNA from a wide range of genetic backgrounds to look for polymorphisms within the open reading frame. Two MHC-encoded LMP subunits have allelic migration on
2D PAGE in the mouse. However, human LMP7 predicted protein sequences were largely identical. One coding dimorphism in the N-terminal leader sequence of LMP7b was detected. A similar low level of polymorphism was observed for LMP2 (S. H. Powis, personal communication). Many diseases are linked to the MHC and the LMP gene products have a potential role in antigen processing. Despite this, the low level of polymorphism within the LMP genes makes it unlikely that variation within the LMPs is a mechanism for generating diversity of immune recognition within the human population.

Further transcripts were described from the opposite DNA strand to TAP1 and LMP7. These were collectively called RING9 as they had similar 5' and 3' ends and shared some intron: exon splice junctions. RING9 mRNA expression was similar to that of TAP1 and LMP7, so that antisense control by RING9 transcription was unlikely. A probe from a region of RING9 that was in the last intron of TAP1 cross-hybridised with cosmids from the mouse genome. In general, coding regions are well conserved between species while introns and other regions of DNA are more diverged. Therefore, this cross-hybridisation implied some functional role for RING9, despite the lack of a long open reading frame. Sequencing of the mouse genome over this region showed that relatively little of the RING9 sequence was conserved, and that the original cross hybridisation signal was due to a small stretch of ~40 bp. In the light of this data, a functional role for RING9 is improbable. The presence of RING9 cDNA clones in B cell libraries is more likely due to aberrant transcription from the LMP7b promoter.
The comparison of human and mouse genomic sequences highlighted other regions of DNA that were well conserved but which were not in the TAP1 or LMP7 open reading frames. These may represent remnants of a gene that is no longer functional and is diverging over time. Alternatively, they may be transcription elements or other control sequences that have been conserved because of some selective pressure. Without functional data it is not possible to distinguish between these hypotheses. However, it would be interesting to see whether these regions of DNA are bound to transcription factors, for example in DNAse protection assays.
DO THE LMPS HAVE A ROLE IN ANTIGEN PROCESSING?

This thesis described the cloning of two genes in the MHC with homology to proteasome components. The circumstantial evidence for a role of the proteasome in the processing of antigens for presentation by MHC class I has also been discussed. However, the results of Arnold, Driscoll et al, 1992 and Momburg, Ortiz-Navarrete et al, 1992 questioned this supposition. Both experiments showed that rescue of class I surface expression in mutant cell lines which were lacking in both the TAP and LMP genes could be achieved by transfection of the two TAP genes, in the absence of the LMP genes. The elution profile of A2 peptides isolated from .174 transfected with human TAP cDNAs resembles more that of the parent line .45 than that of untransfected .174 (Arnold, Driscoll et al, 1992). While this is suggestive that the same peptides are presented through A2 in the presence or absence of the LMPs, a better assay would look at the effect of the LMPs on presentation of several individual epitopes.

This approach was taken by Momburg, Ortiz-Navarrete et al, 1992. This study considered the transfection of rat TAP cDNAs into the T/B cell fusion mutant line, T2. Again, class I levels were increased to near wild type levels. Presentation of a minor histocompatibility antigen, HA-2, to a CTL clone was also restored. Interestingly, the ability of the transfected cell line to present viral peptides was less than that of the parent line T1 (50% presentation for influenza infection, 30% for an influenza protein expressed
from a recombinant vaccinia construct). This decrease in presentation may reflect the absence of the LMPs in these cells. Neither study rules out a role for the LMPs in antigen processing, although both show that presentation of antigen in the absence of the LMPs is possible.

Before the publication of these papers we had reasoned that the LMPs might act either to mediate an interaction between the proteasome and the TAP complex, or to affect the activity of the proteasome so as to bias the degradation products towards those which would be most able to bind to class I. The first of these hypotheses was tested by looking for association between the TAPs and LMPs in immunoprecipitation and western blotting experiments. No evidence for such an association was found using several protocols (G. Allan, personal communication).

There is now some evidence for an effect on proteasome activity by the LMPs. Two groups have shown that the presence of the LMPs in the proteasome complex leads to increased activity after basic and hydrophobic residues relative to cleavage after acidic residues (Driscoll, Brown et al, 1993, Gaczynska, Rock et al, 1993) and see figure 6.1. This is true whether a wild type B cell is compared to an LMP- mutant B cell, or IFN-γ induced cell lines are compared to resting cell lines. The relevance of this shift in activity is that peptides eluted from class I molecules invariably contain basic (Jardetzky, Lane et al, 1991) or hydrophobic (Falk, Rotzschke et al, 1991) residues at the C-terminal end. The authors propose that the role of the LMPs is to alter the activity of the proteasome in such a way that more of the peptides produced will be able to bind to the class I groove at the C terminal end as an "anchor" position.
Figure 6.1: Redundancy in the proteasome subunits may allow specialisation of the complex toward the production of class I binding peptides. The figure is a synthesis of the results of Gaczynska, Rock et al, 1993, Driscoll, Brown et al, 1993 and Belich, Glynne et al, submitted.

Although this would be a convenient explanation for LMP function, another contradictory study has been published (Boes, Hengel et al). In this case, the effect of IFN-γ on mouse fibroblasts was to suppress proteasome activity after hydrophobic amino acids, with no effect on the trypsin-like activity. A possible explanation for the discrepancy was given as differences in interferon-γ induction protocols.

The authors go on to ask what is the effect of IFN-γ on substrates longer than the 3 or 4 mer peptides usually used in these assays. A 25 mer peptide was synthesised which contains a known 9 mer epitope. This...
epitope was a minor component of digestion after treatment of the 25 mer with proteasomes from either resting or IFN-\(\gamma\) induced cells. However, proteasomes isolated from cells which had been treated with IFN-\(\gamma\) showed increased cleavage at the C terminus of the epitope and two amino acids N-terminal of the N terminus. In other words, although IFN-\(\gamma\) did not lead to increased production of the epitope itself, it did increase the production of a possible precursor. This precursor would bind to class I at the C terminus, but would require trimming of two amino acids from the N terminus. Such a trimming mechanism has been proposed by Ramensee (Rammensee, Falk et al, 1993). The products of digestion of the 25 mer by proteasome isolated from interferon-\(\gamma\) treated cells were no more able to stimulate target cells for CTL lysis than the products of 25 mer digestion by proteasome from untreated cells.

IFN-\(\gamma\) produces changes in the amounts/activities of more than the LMPs. In fact, Yang and coworkers found that IFN-\(\gamma\) increased the amounts of five subunits and decreased the amounts of four other subunits in proteasome from HeLa cells. Therefore, any changes caused by IFN-\(\gamma\) treatment cannot be definitively attributed to the LMPs. The LMP- cell lines (.174 or T2) are better reagents in this regard. Data on the cleavage of the 25 mer by the LMP- cell lines, as compared to the wild type lines, would therefore be of interest.

It is now known that results from the cleavage of the tripeptide fluorogenic reagents are not predictive for the cleavage patterns of oligopeptides (Dick, Moomaw et al, 1991, Rivett, 1985 and Boes, Hengel et al, 1994). The 25 mer peptide substrate used in this experiment is likely to have
more similarity to the *in vitro* substrates encountered within a cell than the fluorogenic substrates that have been traditionally used to define the active sites. However, to extend the argument, it is not clear that results from the cleavage of oligopeptides will be a fair reflection of the reaction of the proteasome with whole proteins.

Attempts to degrade intact proteins *in vitro* using the proteasome have often required that the substrate be denatured in some way. For example, methylated casein is often used as a model protein substrate for the proteasome. Oxidation of proteins is another modification which can increase sensitivity to the proteasome: glutamine synthetase and insulin B chain are degraded more efficiently after oxidation (Rivett, 1985). Recently, Pacifici et al have shown that oxidative damage of haemoglobin by hydroxyl damage leads to its degradation by the proteasome. Furthermore, they show that the hydroxylation and degradation of the protein are correlated with an increase in the hydrophobicity of the protein. The suggestion made is that antigen presenting cells may use oxidation to mark antigens for proteasome degradation (Pacifici, Kono et al, 1993). Finally, proteasome from the archaeabacterium *Thermoplasma acidophilus*, will degrade haemoglobin and α-lactalbumin when they have been ubiquitinated (Wenzel and Baumeister, 1993). The authors suggest that ubiquitin may have some chaotropic effect on the substrate and note that the effect of ubiquitination is similar to the effect of oxidation on proteasome substrates.

One theory for LMP function which has not yet been tested is the following. Incorporation of the LMPs into the proteasome may allow the complex to degrade proteins as they are synthesised from the ribosome.
This would be advantageous for the cell because it would allow a bias towards degradation of viral proteins and away from self proteins. At least in the early stages of viral infection the steady state levels of viral proteins in the cytoplasm would be lower than those of self proteins, although the synthesis of viral proteins may be greater than that of self proteins. As the infection proceeded, the concentration of viral proteins would increase, but it is at the early stage of infection that cytolytic killing of the infected cell is likely to be more effective in limiting the spread of virus. In other words, the advantage would be to kill the cell before the assembly of more viral particles.

Many viruses have mechanisms to shut off host protein synthesis and to divert or replace the translation machinery towards the synthesis of viral proteins. This would afford the infected cell an opportunity in which to distinguish foreign from self proteins. Thus, although the steady state levels of proteins would consist predominantly of self, the newly synthesised proteins would be predominantly of viral origin. If the proteasome were directed, by the LMPs, to degrade proteins as they exited the ribosome then the resultant pool of peptides available for class I binding would favour presentation of viral epitopes. A function of the proteasome to degrade newly synthesised proteins may explain the observations of proteasome mediated repression of translation (Horsch, Martins de Sa et al, 1989) and RNA association (Skilton, Eperon et al, 1991, Coux, Nothwang et al, 1992, Nothwang, Coux et al, 1992). Finally, this model would be consistent with the data of Momburg et al and Arnold et al, 1992. The culture time between transfection of the TAP genes into the mutant cells and assays of class I surface expression or presentation would mask any effect the LMPs might
have on the kinetics of processing. However, an investigation of the presentation of a newly synthesised protein(s) may reveal differences between LMP- and LMP+ cell lines.

For T helper cells, a few hundred peptide-MHC class II complexes are sufficient for stimulation of IL2 secretion (Demotz, Grey et al, 1990, Harding and Unanue, 1990). Similarly, class I-peptide complexes at a density of ~200 per presenting cell can be recognised by a CTL clone or mixed population of CTLs (Christinck, Luscher et al, 1991). The mechanism proposed above would then allow for immune recognition of virally infected cells almost as soon as the virus genome begins to be translated. This model could be tested by infecting cells with virus and pulse labelling at the beginning of the infection. Precipitation of a viral protein at chase intervals would show a faster degradation of the viral protein in LMP+ cells as compared to LMP- cells. Alternatively, a pulse chase analysis may show a faster accumulation of radiolabelled peptides eluted from class I molecules when the LMPs are present than in their absence. The advantage of this type of experiment is that it avoids the need for purification of the proteasome. The proteasome complex is poorly understood with relation to its mechanism of catalysis, substrates and regulation. Purification protocols run the risk of either contamination by other proteins or, and more likely to be important, the removal of normal cellular regulators. In fact there is some doubt as to whether the 20S proteasome complex has any role at all other than as a component of the 26S particle.

The evolution of the LMP genes with respect to the TAPs, MB1 and delta may also be informative with respect to LMP function. Speculation as
to a possible scenario for the evolution of these genes has been given in chapter 3. The possibility that the TAP / LMP cluster predates the classical MHC genes should not be discounted. Indeed, it might even be expected as digestion and transport of small peptides is a mechanism that is present in simple unicellular organisms (for example, mating factor secretion in yeast), whereas antigen presentation with an immune function is necessarily a property of multicellular organisms. If this speculation has any truth in it then it may be possible to find a cluster of genes in a lower organism that has homology to the TAP / LMP cluster found in mammals. As yet, there is no published evidence for ABC transporter genes in close linkage to proteasome genes in yeast or T. acidophilum, but this may be worth investigating.

FUTURE PROSPECTS

At the time of writing, the evidence that the LMP gene products are involved in antigen processing remains circumstantial (Brown, Driscoll et al, 1991, Glynne, Powis et al, 1991, Kelly, Powis et al, 1991, Driscoll, Brown et al, 1993, Gaczynska, Rock et al, 1993). In fact, published experiments that have directly addressed this question point to the LMPs being unnecessary for either class I surface expression or antigen presentation (Arnold, Driscoll et al, 1992, Momburg, Ortiz-Navarrete et al, 1992, Yewdell, Lapham et al, 1994). These experiments do not preclude a role for the LMPs when they are expressed, but only show that (some) antigens can be presented in vitro when the LMPs are not present.
Evidence that there is some redundancy within the proteasome of higher eukaryotes suggests that any effect of the LMPs may be masked by expression of the homologous subunits (Yang, Waters et al, 1992, Aki, Shimbara et al, 1994). This might be especially true within the limitations of in vitro experiments - it is possible that a phenotype for the LMPs will be found using whole animal models, for example by gene targetting. Alternatively, redundancy within the subunits could possibly be overcome by expression of the LMPs, MBl or delta with engineered dominant negative mutations. It seems likely that the mechanisms of proteasome catalysis will soon be defined and the identification of active site residues should make possible the construction of mutations with a dominant negative phenotype.

The speculation that the LMPs affect antigen processing through an alteration in proteasome activity would not answer the apparent necessity for redundancy in the subunits. A scenario whereby the LMPs were expressed constitutively and MBl and delta were not present would allow both efficient antigen processing and might still satisfy the requirements of the proteasome for house-keeping protein breakdown. In fact, such a situation exists in B cells which have constitutively low expression of MBl and delta (Belich, Glynne et al, submitted). To put it another way, if the LMPs can be incorporated into the proteasome to form a proteolytically active complex, then what is the necessity for MBl and delta? Gene disruption of either the LMPs, MBl or delta may be informative in this respect. It is possible that the presence of "extra" proteasome subunits is entirely fortuitous and that deletion of their genes will not have a phenotype. However, the presence of LMP genes in mouse, rat and human suggests a function specific to higher eukaryotes. It would be of interest to
see whether a cluster of proteasome and transporter genes could be found in lower organisms which predates the MHC complex.

The mechanisms of regulation of proteasome subunit expression are also intriguing. Initial results suggest that there may be some post-transcriptional mechanism that regulates the relative amounts of the different subunits (Belich, Glynne et al, submitted). It would be an elegant mechanism if the switch between MB1/ delta and LMP expression was effected by the proteasome complex itself, either by proteolysis of unincorporated subunits or repression of mRNA translation. Initial experiments are now in progress to further examine the kinetics of synthesis and degradation of proteasome subunits both in interferon-γ treated cells and in 721 and .174 cells.

Further levels of complexity within proteasome regulation are now being uncovered. Various activators and inhibitors have been characterised biochemically and may soon be cloned by reverse genetics. The same is true for subunits of the 26S complex. Additionally, different pathways of ubiquitination are being further delineated. A greater understanding of cytoplasmic protein breakdown in general, and proteasome-mediated breakdown in particular, should then lead to elucidation of the mechanisms which create peptide epitopes.

Are there any clinical implications of the discovery of proteasome genes within the MHC? Both LMP genes show limited polymorphism and are unlikely to be the basis for the major MHC-linked diseases. However, if firm evidence shows a role for the LMP gene products in antigen
degradation, then the proteasome would become a potential target for immunosuppressants. This is especially true as many inhibitors (both physiological and non-physiological) of proteasome activity have been characterised by proteasome biochemists. The challenge will be to specifically inhibit steps in the generation of reactive epitopes without disrupting the "house-keeping" function of general protein turnover. Such a step might be the switch between MB1/ delta and LMP expression, if this is in any way causative for antigen presentation.

In conclusion, the discovery of the LMP genes within the MHC has strengthened the suggestion that the proteasome is able to provide peptide substrates for the TAP transporter complex and thence for binding to class I. The role that the LMPs have in this process, or that they are involved at all, is still not clear, but it seems that the ubiquitin pathway of protein breakdown is able to provide at least some peptides for class I presentation (Townsend, Bastin et al, 1988, Michalek, Grant et al, 1993). Given the long established connection between the proteasome and ubiquitin-dependent protein breakdown (Hershko, Heller et al, 1983, Ganoth, Leshinsky et al, 1988, Eytan, Ganoth et al, 1989), the proteasome complex remains the best candidate for the degradation of cytosolic proteins to antigenic peptides.
CHAPTER 7: REFERENCES


Arnold D., Driscoll J., Androlewicz M., Hughes E., Cresswell P. and Spies T., Proteasome subunits encoded in the MHC are not generally required for the processing of peptides bound by MHC class I molecules, Nature (1992), 360: 171-174
Arrigo A.-P., Tanaka K., Goldberg A.L. and Welch W.J., Identity of the 19S 'prosome' particle with the large multifunctional protease complex of mammalian cells (the proteasome), Nature (1988), 331: 192-194


Bailey D.W., Genetics of histocompatibility in mice I: New loci and congenic lines, Immunogenet. (1975), 2: 249-256


Boon T. and van Pel A., T cell-recognised antigenic peptides derived from the cellular genome are not protein degradation products but can be generated directly by transcription of short subgenic regions: A hypothesis, Immunogenet. (1989), 29: 75-79


Cantor H. and Boyse E.A., Functional subclasses of T lymphocytes bearing different Ly antigens II: Cooperation between subclasses of Ly+ cells in the generation of killer activity, J. Exp. Med. (1975), 141: 1390-1399


Cardozo C., Vinitsky A., Hidalgo M.C., Michaud C. and Orlowski M., 3,4-dichloroisocoumarin-resistant component of the multicatalytic proteinase complex, Biochemistry (1992), 31: 7373-7380


Chen P., Johnson P., Sommer T., Jentsch S. and Hochstrasser M., Multiple ubiquitin-conjugating enzymes participate in the in vivo degradation of the yeast MATα2 repressor, Cell (1993), 74: 357-369

Chin D.T., Kuehl L. and Rechsteiner M., Conjugation of ubiquitin to denatured hemoglobin is proportional to the rate of hemoglobin degradation in HeLa cells, Proc. natl. Acad. Sci. USA (1982), 79: 5857-5861


DeMars R., Chang C.C., Shaw S., Reitnauer P.J. and Sondel P.M., Homozygous deletions that simultaneously eliminate expressions of class I and class II antigens of EBV-transformed B-lymphoblastoid cells I: Reduced proliferative responses of autologous and allogeneic T cells to mutant cells that have decreased expression of class II antigens, Hum. Immunol. (1984), 11: 77-97


of the human, high-molecular-weight proteinase, macropain (proteasome), are distinct but homologous, Biochim. Biophys. Acta (1991), 1079: 29-38

Demotz S., Grey H.M. and Sette A., The minimal number of class II MHC-antigen complexes needed for T cell activation, Science (1990), 249: 1028-1030

Deverson E.V., Gow I.R., Coadwell W.J., Monaco J.J., Butcher G.W. and Howard J.C., MHC class II region encoding proteins related to the multidrug resistance family of transmembrane transporters, Nature (1990), 348: 738-741


Dick L.R., Moomaw C.R., Pramanik B.C., Demartino G.N. and Slaughter C.A., Identification and localization of a cysteiny1 residue critical for the trypsin-like catalytic activity of the proteasome, Biochemistry (1992), 31: 7347-7355


Djaballah H., Rowe A.J., Harding S.E. and Rivett A.J., The multicatalytic proteinase complex (proteasome) - structure and conformational-changes


Driscoll J. and Goldberg A.L., The proteasome (multicatalytic protease) is a component of the 1500-kDa proteolytic complex which degrades ubiquitin-conjugated proteins, J. Biol. Chem. (1990), 265: 4789-4792

Dubiel W., Ferrell K., Pratt G. and Rechsteiner M., Subunit 4 of the 26S protease is a member of a novel eukaryotic ATPase family, J. Biol. Chem. (1992), 267: 22699-22702

Dubiel W., Ferrell K. and Rechsteiner M., Peptide sequencing identifies MSS1, a modulator of HIV Tat-mediated transactivation, as subunit 7 of the 26S protease, FEBS Lett. (1993), 3: 276-278


Hanson I.M. and Trowsdale J., Colinearity of novel genes in the class II regions of the MHC in mouse and human, Immunogenet. (1991), 34: 5-11


Heinemeyer W., Kleinschmidt J.A., Saidowsky J., Escher C. and Wolf D.H., Proteinase yscE, the yeast proteasome/multicatalytic-multifunctional proteinase: mutants unravel its function in stress induced proteolysis and uncover its necessity for survival, EMBO (1991), 10: 555-562


Hilt W., Enenkel C., Gruhler A., Singer T. and Wolf D.H., The pre4 gene codes for a subunit of the yeast proteasome necessary for peptidylglutamyl-peptide-hydrolyzing activity - mutations link the proteasome to stress-
dependent and ubiquitin-dependent proteolysis, J. Biol. Chem. (1993), 268: 3479-3486


Hosken N.A. and Bevan M.J., Defective presentation of endogenous antigen by a cell line expressing class I molecules, Science (1990), 248: 367-370


Knowles S.E. and Ballard J.F., Selective control of degradation of normal and aberrant proteins in Reuber H35 hepatoma cells, Biochem. J. (1976), 156: 609-617

Kopp F., Dahlmann B. and Hendil K.B., Evidence indicating that the human proteasome is a complex dimer, J. Mol. Biol. (1993), 229: 14-19


Kozak M., Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes, Nucleic Acids Res. (1981), 9: 5233-5252


Kuehn L., Dahlmann B. and Reinauer H., Evidence indicating that the multicatalytic proteinase of rabbit reticulocytes is not incorporated as a core enzyme into a 26S proteinase complex, Arch. Biochem. Biophys. (1992), 295: 55-60


Li X. and Etlinger J.D., Ubiquitinated proteasome inhibitor is a component of the 26S proteasome complex, Biochemistry (1992), 31: 11963-11967

Li X.C., Gu M.Z. and Etlinger J.D., Isolation and characterization of a novel endogenous inhibitor of the proteasome, Biochemistry (1991), 30: 9709-9715


Medawar P., Immunity to homologous grafted skin II: the relationship between the antigens of blood and skin, Brit. J. Exp. Path. (1946), 27: 15-24


Murakami Y., Matsufuji S., Kameji T., Hayashi S., Igarashi K., Tamura T., Tanaka K. and Ichihara A., Ornithine decarboxylase is degraded by the 26S proteasome without ubiquitination, Nature (1992), 360: 597-599


Nishizawa M., Okazaki K., Furuno N., Watanabe N. and Sagata N., The 'second-codon rule' and autophosphorylation govern the stability and activity of Mos during the meiotic cell cycle in *Xenopus* oocytes, EMBO J. (1992), 11: 2433-2446


Pacifici R.E., Kono Y. and Davies K.J.A., Hydrophobicity as the signal for selective degradation of hydroxyl radical-modified hemoglobin by the multicatalytic proteinase complex, proteasome, J. Biol. Chem. (1993), 268: 15405-15411

Parham P., Transporters of delight, Nature (1990), 348: 674-675


Pile K.D., Burney R.O., Gibson K., Calin A., Kennedy G., Sinnott P.J., Powis S.H. and Wordsworth P., No additional role for the MHC class II encoded components of the HLA class I pathway of antigen presentation (LMP2, TAP1 and 2) in the etiology of ankylosing spondylitis (AS), Br. J. Rheumatol. (1993), 32: 25-25


Salter R.D. and Cresswell P., Impaired assembly and transport of HLA-A and -B antigens in a mutant TxB cell hybrid, EMBO J. (1986), 5: 943-949


Seufert W. and Jentsch S., In vivo function of the proteasome in the ubiquitin pathway, EMBO J. (1992), 11: 3077-3080


Snell G.D., Methods for the study of histocompatibility genes, J. Genet. (1948), 49: 87-108

Sommer T. and Jentsch S., A protein translocation defect linked to ubiquitin conjugation at the endoplasmic reticulum, Nature (1993), 365: 176-179

Spies T., Bresnahan M., Bahram S., Arnold D., Blanck G., Mellins E., Pious D. and DeMars R., A gene in the human major histocompatibility complex class II region controlling the class I antigen presentation pathway, Nature (1990), 348: 744-747

Spies T., Cerundolo V., Colonna M., Cresswell P., Townsend A. and DeMars R., Presentation of viral antigen by MHC class I molecules is dependent on a putative peptide transporter heterodimer, Nature (1992), 355: 644-646

Spies T., Cerundolo V., Colonna M., Cresswell P., Townsend A. and Demars R., Presentation of viral-antigen by mhc class-i molecules is dependent on a putative peptide transporter heterodimer, Nature (1992), 355: 644-646


Steinmetz M., Stephan D. and Fischer-Lindahl K., Gene organisation and recombinational hotspots in the murine major histocompatibility complex, Cell (1986), 44: 895-904
Sternberg M.J.E., PROMOT: a FORTRAN program to scan protein sequences against a library of known motifs, Computer Applications in the Biosciences (1991), 7: 257-260


Todd J., Bell J.I. and McDevitt H.O., HLA-DQβ gene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus, Nature (1987), 329: 599-604


Townsend A., Bastin J., Gould K., Brownlee G., Andrew M., Coupar B. and Boyle D., Defective presentation to class I restricted cytotoxic T lymphocytes is overcome by enhanced degradation of antigen, J. Exp. Med. (1988), 168: 1211-1224


van Kaer L., Ashton-Rickardt P.G., Ploegh H.L. and Tonegawa S., TAP1 mutant mice are deficient in antigen presentation, surface class I molecules, and CD4-8+ T cells., Cell (1992), 71: 1205-1214


Weitman D. and Etlinger J.D., A monoclonal antibody that distinguishes latent and active forms of the proteasome (multicatalytic proteinase complex), J. Biol. Chem. (1992), 267: 6977-6982


Wolf S., Lottspeich F. and Baumeister W., Ubiquitin found in the archaebacterium Thermoplasma acidophilum, FEBS Lett. (1993), 326: 42-44


Yu B., Pereira M.E. and Wilk S., Chemical modification of the bovine pituitary multicatalytic proteinase complex by N-acetylimidazole, J. Biol. Chem. (1991), 266: 17396-17400


Ziegler A., Muller C., Heinig J., Radka S.F., Kompf J. and Fonatsch C., Monosomy 6 in a human lymphoma line induced by selection with a monoclonal antibody, Immunobiology (1985), 169: 455-460


Zwickl P., Grziwa A., Puhler G., Dahlmann B., Lottspeich F. and Baumeister W., Primary structure of the thermoplasma proteasome and its implications
for the structure, function, and evolution of the multicatalytic proteinase, Biochemistry (1992), 31: 964-972


Zwickl P., Lottspeich F., Dahlmann B. and Baumeister W., Cloning and sequencing of the gene encoding the large (α-) subunit of the proteasome from Thermoplasma acidophilum, FEBS Lett. (1991), 278: 217-221