REFERENCE ONLY

UNIVERSITY OF LONDON THESIS

Degree PhD Year 2006 Name of Author AUCAMA J. K

COPYRIGHT
This is a thesis accepted for a Higher Degree of the University of London. It is an unpublished typescript and the copyright is held by the author. All persons consulting the thesis must read and abide by the Copyright Declaration below.

COPYRIGHT DECLARATION
I recognise that the copyright of the above-described thesis rests with the author and that no quotation from it or information derived from it may be published without the prior written consent of the author.

LOANS
Theses may not be lent to individuals, but the Senate House Library may lend a copy to approved libraries within the United Kingdom, for consultation solely on the premises of those libraries. Application should be made to: Inter-Library Loans, Senate House Library, Senate House, Malet Street, London WC1E 7HU.

REPRODUCTION
University of London theses may not be reproduced without explicit written permission from the Senate House Library. Enquiries should be addressed to the Theses Section of the Library. Regulations concerning reproduction vary according to the date of acceptance of the thesis and are listed below as guidelines.

A. Before 1962. Permission granted only upon the prior written consent of the author. (The Senate House Library will provide addresses where possible).

B. 1962 - 1974. In many cases the author has agreed to permit copying upon completion of a Copyright Declaration.

C. 1975 - 1988. Most theses may be copied upon completion of a Copyright Declaration.

D. 1989 onwards. Most theses may be copied.

This thesis comes within category D.

☑ This copy has been deposited in the Library of UCL

☐ This copy has been deposited in the Senate House Library, Senate House, Malet Street, London WC1E 7HU.

C:\Documents and Settings\lproctor\Local Settings\Temporary Internet Files\OLK8\Copyright - thesis (2).doc

Engineering a novel high-throughput screen for directed evolution of enzyme stability.

A thesis submitted to the University of London

for the degree of

Doctor of Philosophy

by

Jean Pieter Aucamp

February 2006

Advanced Centre for Biochemical Engineering

Department Chemical and Biochemical Engineering

University College London

Torrington Place

London

WC1E 7JE

UMI Number: U592617

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.

UMI U592617
Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.
All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.

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

Abstract

Directed evolution techniques have revolutionised strategies for enzyme development and improvement. Large libraries, containing typically > 1000 protein mutants, are produced and screened for desired properties (e.g. novel activity or improved stability). The correct application of selection pressure depends on the accuracy and robustness of the screen; hence direct screens are preferred to indirect screens.

Protein stability is often screened indirectly by monitoring protein aggregation or residual activity after incubation at elevated temperatures. Both screens rely on irreversible inactivation of the protein upon unfolding. The assumption is made that positive 'hits' result from variants resistant to denaturation under the test conditions and do not take spontaneous protein refolding into account.

We have established an affordable, high-throughput, direct screen for protein stability that can be automated. Building on previous work [Edgell *et. al.* (2003) Biochemistry, 42, 7587-7593] we have extended and optimised the technique for application entirely in the higher throughput microplate format. Pure target proteins are unfolded in microwells by serial addition of chemical denaturants such as urea or guanidine hydrochloride. The unfolding curves are recorded by measuring the changes in tryptophan fluorescence under equilibrium conditions. The midpoint of unfolding determined for each protein variant ($C_{1/2}$) is an indication of its stability. This screen was developed and tested using bovine and equine cytochrome *c* as well as preparations of fat-stabilised BSA.

This screen requires high-throughput protein purification and suitable buffer conditions for stability analysis of each mutant. A multi-step process route, amenable to automation, has been developed that involves cell culturing, protein purification with metal affinity resin, buffer exchange by dialysis and unfolding analysis, all in a high-throughput format. Factors taken into account are maximum cell density of micro-scale cell-culturing, target protein titre, resin binding capacity and minimum final protein concentration required for stability screening.

This route was developed using wild-type His_6-tagged transketolase (*E. coli tkt* gene) over-expressed in *E. coli* JM107 pQR791. Transketolase mutants, D381A and Y440A, with decreased dimer stability were also designed as model systems for the stability analysis. The reduced stability was clearly observed. The route needs more optimisation to yield sufficient transketolase required for accurate screening. Proper implementation of the process route will allow for screening of a thousand mutants per day for improved stability. The stabilities of proteins with 'hard to screen activities' and reversible folding can be improved more readily using this route.

Acknowledgments

I am grateful to the following:

Dr. Paul Dalby, my supervisor, for support and guidance throughout the course of this research. His enthusiasm provided motivation when needed most.
Dr. Gary Lye, my academic advisor, for providing an industrial perspective on the research work conducted.

The Association of Commonwealth Universities for arranging the scholarship, travelling and financial support. The British Council for financial support over the three year tenure. The UCL Graduate School for partial funding of conference fees.

My fellow Foster Court researchers for their support and friendship. In particular: the 'day-brigade' for interesting converse and good-humour; the 'night-brigade' for encouragement and comradeship. Ana Cosme for her contribution to Chapter 3.

And last but not least, my family for their love and understanding.

Contents

List of Figures

List of Tables

Abbreviations

aa seq	Amino acid sequence
Ala, A	Alanine
ASA	Accessible surface area
Asn, N	Asparagine
Asp, D	Aspartic acid
AU	Absorbance units
BCA	Bicinchoninic acid
β-ME	β-mercaptoethanol
BSA	Bovine serum albumin
CBP	Calmodulin binding peptide
cDNA	complementary Deoxyribonucleic Acid
CLEC	Cross-linked enzyme crystals
CV	Column volume
Cyt c	Cytochrome *c*
DNA	Deoxyribonucleic Acid
dNTP	Deoxynucleoside triphosphates
DR5P	D-ribose-5-phosphate
DR5P3E	D-ribulose-5-phosphate-3-epimerase
DSC	Differential scanning calorimetry
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
epPCR	Error-prone PCR
ESR	Estimated substitution rate
FACS	Fluorescence Activated Cell Sorting

FepPCR	Focused error-prone PCR
FLINT	Fluorescence intensity
GdnHCl	Guanidine hydrochloride
GDTPI	α-glycerophosphate dehydrogenase-triosephosphate isomerase
Gln. E	Glutamine
Gly. G	Glycine
GSH	Glutathione
GST	Glutathione-S-transferase
H/D	Hydrogen/Deuterium
His. H	Histidine
HPLC	High Performance Liquid Chromatography
ITCHY	Incremental Truncation for the Creation of Hybrid enzYmes
LB	Luria Bertani
MALDI-MS	Matrix-Assisted Laser Desorption/Ionisation Mass Spectrometry
MBP	Maltose binding protein
mRNA	messenger Ribonucleic Acid
MSA	Multiple sequence alignment
MWCO	Molecular weight cut-off
NaCl	Sodium chloride
NADH	Nicotinamide adenine dinucleotide (reduced)
Ni-NTA	Nickel-nitrilotriacetic acid
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PRI	Phosphoriboisomerase
RACHITT	RAdom CHImeragenesis on Transient Templates
SDS	Sodium dodecylsulfate

SEC	Size-exclusion chromatography
SOE	Strand Overlap Extension
SSDM	Saturation site-directed mutagenesis
StEP	Staggered extension process
TFA	Trifluoroacetic acid
TPP	Thiamine pyrophosphate
Tyr, Y	Tyrosine
UF	Ultra-filtration
UV	Ultraviolet
wt	Wild-type
β-HPA	Beta-Hydroxypyruvate

Chapter 1

Introduction

1.1 Transketolase

1.1.1 Structure and function

Transketolase connects the non-oxidative branch of the pentose phosphate pathway and the Calvin cycle providing pentose sugars as essential building blocks for various biochemical reactions. As a member of the family of thiamine-dependent enzymes, transketolase along with other members, such as pyruvate decarboxylase and pyruvate oxidase, require TPP (thiamine pyrophosphate) for activity. All the thiamine-dependent enzymes share a conserved sequence motif required for binding of the TPP cofactor (Schenk *et al.*, 1998). Comparison of X-ray crystallographic data from several thiamine-dependent enzymes revealed an identical three-dimensional structure for the TPP binding site, despite significant differences in the overall protein sequences.

A wealth of structural data is available since the three-dimensional structures of transketolases from *S. cerevisiae* (Lindqvist *et al.*, 1992; Nikkola *et al.*, 1994) and *E. coli* (Littlechild, 1995) have been elucidated. TPP binds in a deep cleft at the interface between two transketolase subunits (Figure 1.1). Transketolase is an obligatory homodimeric protein with both subunits contributing to the formation of the TPP binding and active sites. Apart from various hydrogen and hydrophobic interactions with transketolase, TPP is also

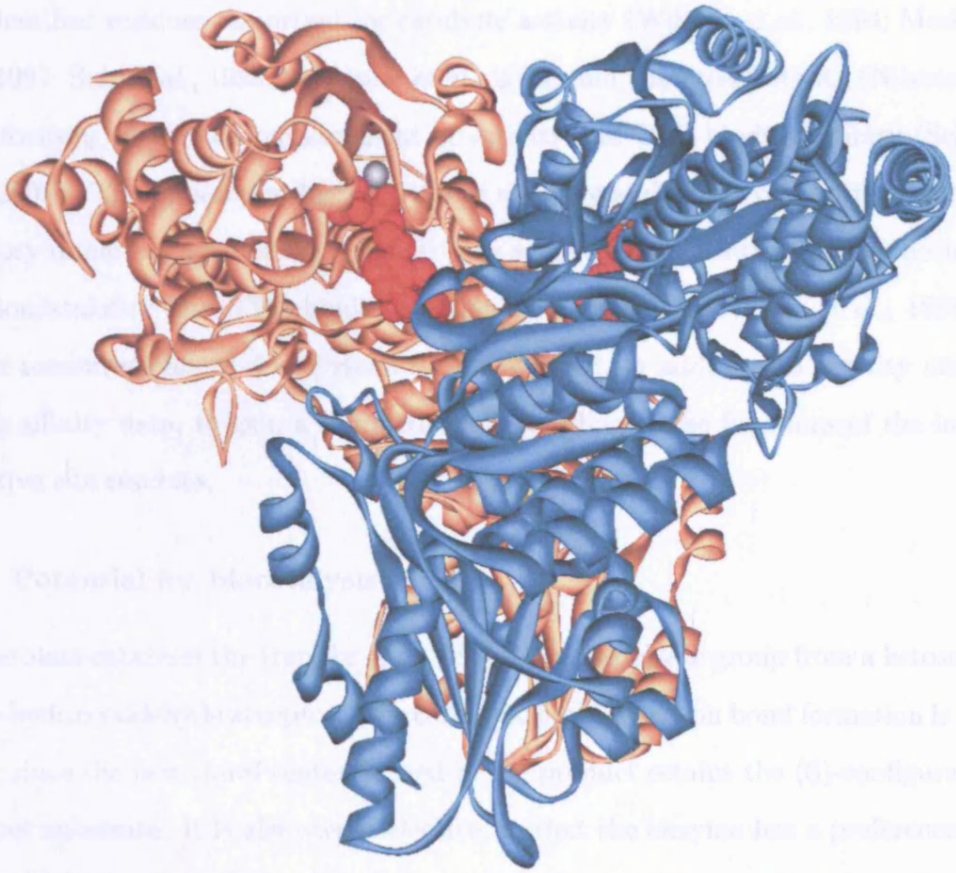


Figure 1.1: Schematic presentation of the crystal structure of *E. coli* transketolase. The structure (PDB ID: 1qgd), solved by J. Littlechild (University of Exeter) and obtained from the Protein Data Bank (http://www.pdb.org/) (Berman *et al.*, 2000). Transketolase displays a homodimeric structure with subunits coloured blue and brown respectively. TPP in red is bound in a cleft between the two subunits. A divalent metal ion (Ca^{2+}), coloured in gray, interacts with both a subunit and TPP cofactor.

anchored to the enzyme indirectly by means of a coordination complex with a divalent metal ion (Mg^{2+} or Ca^{2+}). The metal ion in turn is sequestered in a metal-binding motif conserved in all thiamine-dependent enzymes (Hawkins *et al.*, 1989).

The conformation of TPP is contorted into a 'V' shape, most likely a higher energy state needed to facilitate catalysis (Jordan, 2003). Thus far protein engineering studies have identified residues important for catalytic activity (Wikner *et al.*, 1994; Meshalkina *et al.*, 1997; Soh *et al.*, 1998; Selivanov *et al.*, 2004) and enantioselectivity (Nilsson *et al.*, 1998), focusing mainly on measurement of activity and TPP binding affinity (Selivanov *et al.*, 2003). The interaction between dimer stability and activity is pronounced in the obligatory dimeric nature of the enzyme, with some residues contributing to both dimer formation/stability and TPP binding (Meshalkina *et al.*, 1997; Soh *et al.*, 1998). As a result measurements of dimer stability are required, in addition to activity and TPP binding affinity data, to gain a complete understanding of the functions of the interface and active site residues.

1.1.2 Potential for biocatalysis

Transketolase catalyses the transfer of an active glycolaldehyde group from a ketose donor to an α-hydroxyaldehyde acceptor. The catalysed carbon-carbon bond formation is stereospecific since the new chiral center formed in the product retains the (S)-configuration of the donor substrate. It is also stereoselective, in that the enzyme has a preference for α-hydroxyaldehydes with (R)-configuration at C-2. Transketolase mediated condensations produce enantiomerically pure chiral compounds with [3S,4R] or D-threo stereochemistry. Non-phosphorylated, as well as phosphorylated, substrates are accepted by the enzyme; the best substrates for transketolase are α-hydroxyaldehydes of (R)-configuration, though α-oxo and α-unsubstituted aldehydes are also accepted (Turner, 2000).

The production of chiral synthons using transketolase have already found several industrial applications. Transketolase produces carbohydrate precursors required for the shikimic acid pathway mediated biosynthesis of the aromatic amino acids L-phenylalanine,

L-tryptophan and L-tyrosine (Ganem, 1978). These amino acids are used as precursors for the organic synthesis of various products; e.g. L-phenylalanine can be transformed into the artificial sweetener aspartame, L-tryptophan into the dye indigo and L-tyrosine into eumelanin, a UV-absorbing substance (Schenk *et al.*, 1998).

Several examples of the synthesis of the unnatural sugars have been reported. Transketolase from spinach was used to synthesise 6-deoxy-D-fructose and 6-deoxy-L-sorbose (Hecquet *et al.*, 1994a) which can serve as precursors of furaneol, an important industrial aromatic product with caramel-like flavour used in the food industry (Hecquet *et al.*, 1996). Yeast transketolase was used to synthesise 4-deoxy-D-fructose 6-phosphate α-unsubstituted aldehyde acceptor (Guérard *et al.*, 1999) an inhibitor useful for enzyme mechanistic studies and sugar metabolism. Analogues of L-threose and D-erythrose have been prepared in which the C-2 position has been modified (i.e. 2-O-methyl, 2-deoxy) (André *et al.*, 1998a). These modified aldehydes can serve as substrates for transketolase and was used to make novel heptulose analogues (André *et al.*, 1998b).

One of the major problems for the large scale industrial use of transketolase biotransformations is that high concentrations of substrate strongly decrease the turnover number. Using β-hydroxypyruvate (β-HPA) and glycolaldehyde as substrates it was found that transketolase was inactivated at high concentrations of the acceptor substrate (glycolaldehyde) (Mitra *et al.*, 1998). *In situ* product removal (Chauhan *et al.*, 1996) as well as the use of enzyme-membrane reactors (Bongs *et al.*, 1997) are engineering solutions ensuring the concentrations of substrates and products remain low, thus greatly increasing the half-life and productivity of transketolase biotransformation reactions. These problems can potentially also be solved by protein engineering methods.

1.2 Approaches in protein engineering

Protein engineering studies commenced in earnest in the early 1980's with first reports of the successful creation of site-directed mutations (Ulmer, 1983). In general, protein engineering can be described as the use of site-directed or random mutagenesis to alter the

primary amino acid sequence and subsequently the tertiary structure and properties of a protein or enzyme. Protein engineering has progressed significantly and has developed into a discipline in its own right, complete with designing and engineering approaches that will be discussed in this section.

1.2.1 Rational Design

A rational approach to protein engineering involves predicting protein structure, function and interaction from first principles. Search algorithms and energy functions use, as input, the primary amino acid sequence or three-dimensional structure and the vast number of residue and solvent interactions to estimate theoretical solutions to design problems. Rational design still relies heavily on the availability of three-dimensional structures. The greatest challenge of *de novo* protein design is selecting an optimal solution from the available sequence and structural possibilities. The large number of conformations provide an enormous computational burden. It was proposed that protein design can be simplified by adopting a strategy where the backbone is fixed (Ponder and Richards, 1987). The side-chains are assumed to exclusively adopt statistically preferred conformations (rotamers). Essentially the protein design problem is reduced to finding the lowest energy combination of side-chain rotamers for any given backbone template and can be determined with search algorithms linked to energy functions. Solutions to design problems are not always guaranteed with rational design as there are many more computational challenges remaining that impede the potential of this approach, including the adoption of dynamic protein structure and backbone flexibility in energy minimisations rather than reliance on static structures (Dalby, 2003).

1.2.1.1 Search algorithms

Search algorithms can be either stochastic or semi-random in nature. Stochastic algorithms such as Monte Carlo (Dahiyat and Mayo, 1996; Godzik, 1995) methods or Genetic algorithms (Jones, 1994; Lazar *et al.*, 1997) are good for dealing with problems of sig-

nificant combinatorial complexity since an exhaustive search is not required. There is however no guarantee that these methods will provide a global minimum or even produce the same solution when executed consecutively (Voigt, Gordon and Mayo, 2000). Two different deterministic methods are used widely. The first is the Self-Consistent Mean-Field (SCMF) which optimises arbitrary Boltzmann probabilities, assigned to each rotamer, using a weighted energy average for each rotamer that is also position dependent (Koehl and Delarue, 1994). A solution is obtained when the Boltzmann probabilities converge. SCMF always produces the same solution but it is not guaranteed to be the global minimum. The second and most powerful search algorithm to date is the dead-end elimination (DEE) method (Desmet *et al.*, 1992; Goldstein, 1994). Unlike the Monte Carlo or Genetic Algorithm that attempt to optimise a single solution, DEE eliminates bad rotamers or rotamer combinations until a single solution remains. It is also deterministic in nature, providing the global optimum as solution each time (Voigt, Gordon and Mayo, 2000).

1.2.1.2 Force field functions

Molecular mechanistic force field descriptors normally consist of van der Waals, electrostatic, dihedral angle, bond-angle and bond stretch components. In protein design the last two terms are eliminated since only ideal geometries are considered (Pokala and Handel, 2001). The function describing van der Waals interaction forms the dominant term of the force field and is used to optimise the hydrophobic core of the protein. Solvation energies are difficult to calculate accurately, mainly due to the poorly defined hydrophobic effect (Edinger *et al.*, 1997). The solvent entropy is reduced where hydrophobic rotamers are exposed and come into contact with water. The behaviour of water is extremely difficult to model and to overcome this problem, solvation energies are calculated based on the solvent accessible areas (Eisenberg and McLachlan, 1986). Electrostatic interactions such as H-bonds and salt bridges define the protein structure specificity and function. Solvation of surface polarised groups and the inter-molecule competition depends greatly on the local micro-environment, making it hard to estimate accurately (Bashford and

Case, 2000). In addition, the environment of charged residues has a marked effect on their ionisation energies and states, requiring environment dependent ionisation models (Lazaridis and Karplus, 1999). H-bonds are important in stabilising secondary structures (α-helices and β-sheets) and can be incorporated into the force field function as explicit terms or implicitly defined by electrostatic and hydrophobic interactions. Explicit terms take into account the distance between donor and acceptor as well as orientation and bond vectors (Mayo *et al.*, 1990). Additional terms that may be included are secondary structure propensity factors that can be used to place additional constraints on protein design or to aid prediction secondary structures stability (Dahiyat and Mayo, 1997; Strop *et al.*, 2000).

Despite the limitations of the fixed backbone approach, successes pertaining to rational design of core stabilisation, surface and interface interactions have been reported. Ultimately the aim of the rationale approach is *de novo* design of proteins without the aid of fixed backbone scaffolds. Introducing backbone flexibility will allow this, but requires more complex energy factors and functions. A possible way of circumventing the problem is by designing backbone families and selecting the one providing the best solution after energy optimisation (Harbury *et al.*, 1998). The 'gold standard' search algorithms and force field equations used in rational approaches are still too computationally demanding to be applied to design problems of large proteins (Pokala and Handel, 2001).

1.2.2 Directed Evolution

A random approach to protein engineering involves *in vitro* modification of protein sequences followed by characterisation of the perturbations. Directed evolution techniques are performed as iterative cycles of gene sequence modification and high-throughput assaying for desirable traits. While sequence modifications evolve the gene and produce genotype variation, rapid screening or selection procedures provide the artificial selective pressure and a means of finding the fittest phenotypes. Selected candidates are reintroduced into the next cycle to create stepwise fitness improvements. The success of this

approach is reliant on sufficient initial genetic diversity, suitable screening or selection methods and an understanding of population fitness principles.

1.2.2.1 Strategies for library creation

Gene sequence diversity can be created by several methods, including physical or chemical DNA damage, DNA replication with compromised fidelity or repair mechanisms, random or site-directed mutagenesis and gene fragmentation and recombination.

Physical and chemical damage

Classical strain improvement strategies often induce mutagenesis with ultraviolet (UV) radiation or chemical reagents. Mutant libraries have been created by inducing chemical damage using alkylating agents (Horsfall *et al.*, 1990), hydroxylamine (Taguchi *et al.*, 1998) and superoxide (Ono *et al.*, 1995). These methods are rather inefficient since biased mutations are created. For instance, alkylating agents induce G:C transversion with a higher frequency than any other base substitution.

Random mutagenesis

Mutator strains: Gene sequences can be mutated using cell strains with deficient DNA repair mechanisms, such as the XL1-Red strain from Stratagene (Greener *et al.*, 1997). Plasmid DNA is propagated in the mutator strain for a number of cell divisions to allow accumulation of point mutations. Mutated plasmids are purified and transformed into a regular expression strain for screening/selection purposes (Bornscheuer *et al.*, 1999). Mutator strains are highly strained by their deficiency as well as the burden of protein overexpression. Mutations that reduce protein expression often prevail allowing their host cells to out compete others with higher metabolic burdens. Slow growth, low mutation rates and low mortality of strains normally lead to DNA constructs having to pass through the strain several times. Mutations occur throughout the plasmid, including the selection gene, potentially giving rise to artifacts during the selection process. In addition, the

biased DNA repair mechanisms of mutator strains are generally regarded as an inefficient way of creating diversity.

Error-prone PCR (epPCR): DNA polymerase enzymes that lack proofreading activity, such as Taq, are used to introduce point mutations during DNA amplification. The fidelity of Taq is nonetheless very high and is reduced by adding small amounts Mn^{2+} to the reaction mixture (Lingen *et al.*, 2002; Bessler *et al.*, 2003). Biased concentrations of deoxynucleoside triphosphates (dNTP's) or non-natural dNTP analogues in the reaction mixture provide alternative methods for increasing the mutation rates (Zaccolo *et al.*, 1996). Phenotype diversity created with random single mutations methods is impeded by the natural codon bias. Depending on the exact codon, possible amino acid variation after a single mutation is limited to between two and six. Sequence space coverage may be increased by permitting more mutations per gene.

Site-directed mutagenesis

Since random point mutations access on average less than six amino acids (Jespers *et al.*, 1997), methods are required that produce double- and triple-base mutations. Saturation mutagenesis methods ensure that all amino acid combinations are accessed. The classical way of introducing variation is by cassette mutagenesis using degenerate oligonucleotides (Reidhaar-Olson *et al.*, 1991). The use of degenerate oligonucleotides in methods such as strand overlap extension (SOE) (Ho *et al.*, 1989; Juillerat *et al.*, 2003) or whole plasmid amplification (Miyazaki and Takenouchi, 2002; Zheng *et al.*, 2004) permit introduction of mutations at multiple positions. This still does not solve the mutation bias since some amino acids are coded for by more codons than others. For instance six codons translate for leucine and only one for tryptophan. Saturation site-directed mutagenesis (SSDM) provides more control over the amount and type of mutations that can be produced, but can only be applied on small parts of a protein.

Recombination

In nature, progeny obtain traits from both parents by the process of DNA recombination. During recombination homologous DNA regions line up and exchange genes creating diversity. *In vitro* homology recombination, known as DNA shuffling, is accomplished by fragmenting several homologous parental genes followed by reassembly using primerless DNA amplification (Stemmer, 1994). An alternative method of DNA shuffling, the staggered extension process (StEP) (Zhao *et al.*, 1998), employs short, iterative PCR cycles to create diversity. The templates consist of homologous parent DNA that are recombined in the amplification processes when the elongating strands anneal to different parental templates at each thermocycle. DNA shuffling is also used to recombine single point mutations from epPCR or mutator strain strategies (Zhang *et al.*, 1997; Giver *et al.*, 1998). As with the natural process there is a tendency to retain beneficial mutations whilst removing deleterious ones.

DNA shuffling methods have also been extended to enable recombination of parental genes with low homology. Random chimeragenesis on transient templates (RACHITT) (Coco *et al.*, 2001) relies on the ordering, trimming, and joining of randomly cleaved parental DNA fragments by annealing to a transient DNA scaffold. A large crossover number for recombination is achieved using this method, producing novel chimeric libraries with increased exploration of sequence space. In a method known as ITCHY, several parental genes are truncated incrementally and the end products linked to create fusion libraries (Ostermeier *et al.*, 1999a,b). This method essentially mimics domain swapping, used in nature to increase diversity. Incrementally truncated libraries can be recombined with DNA shuffling (SCRATCHY) to increase diversity further (Lutz *et al.*, 2001). Recombining parents from more diverse sources can lead to larger jumps in sequence space useful for finding novel enzyme function.

1.2.2.2 Screening and selection

High-throughput assays form an integral part of directed evolution, identifying variants with improved or novel traits. Assays can be classified as either selection or screening methods. The nature of selection methods is such that only protein with the desired trait is observed and carried forward to the next cycle of evolution. Well known examples of selection methods include antibiotic resistance and binding affinity selection. The fittest variants are gradually enriched by using increased selection pressure linked to higher growth rates or tighter affinity. Screening methods analyse and rank individual protein variants relative to each other using an arbitrary point of reference (i.e. the point of origin). Biochemical assays for thermostability and enzyme activity are examples of screening methods. Selection methods are very powerful, capable of assaying libraries with up to 10^{10} variants. The applications of selection assays are however limited since not all properties can be evaluated in this manner. Screening assays are capable of evaluating libraries containing 10^3-10^6 variants, depending on the method of assay.

For protein optimisation using directed evolution, it is important to retain association between the gene and its product. Several strategies have been implemented to link the genotype and phenotype.

Direct linking

With ribosomal display methods the genotype is provided in the form of mRNA and the protein is produced with an *in vitro* expression system. The mRNA and protein are fused covalently with puromycin, a translation inhibitor, and remain complexed with the ribosome (Roberts and Szostak, 1997). Selection is done using binding affinity properties of the fusion protein. The genotype is retrieved by reverse transcription producing cDNA libraries for analysis.

Another direct-link technique is the peptide on plasmid display method (Cull *et al.*, 1992). The target sequence is fused with the C-terminus of the *lac*-repressor protein. The construct is expressed on a plasmid containing two *lacO* DNA sequences. The

lac-repressor fusion protein bind non-covalently to the *lacO* sequences of the plasmid, effectively linking genotype and phenotype. Binding affinity is used to select the suitable phenotypes.

Indirect linking

With phage display methods the target protein is fused to a phagemid coat protein for selection purposes (Smith, 1985). Cells transformed by electroporation with phage DNA produce phage particles that are harvested and used in affinity binding selection methods known as 'panning' (Forrer *et al.*, 1999). Non-binding particles are washed away, while retained particles go through iterative cycles of infection, harvesting and selection under more stringent conditions. The phage particles displaying the selected fusion proteins harbour the DNA-tags, making genotype identification straightforward.

Target proteins can also be displayed on cells by fusing it with a membrane protein. Thus far *E. coli* (Francisco *et al.*, 1992; Boder *et al.*, 2000) and yeast (Boder and Wittrup, 1997) cells have been used for cell display selection. Screening is done with fluorescence activated cell sorting (FACS). The design of suitable screens can be complex, since fluorophores with quenching partners or fluorescent competitive substrates are required.

Compartmentalisation methods

Cell-based or genetic screens compartmentalise the genotype and phenotype in the same cell. This system is very powerful, providing suitable selection or screening methods can be devised. Coloured products or fluorogenic substrates are very useful as simple visual screens for novel enzyme catalytic reactions (MacBeath *et al.*, 1998; Joo *et al.*, 1999). Genetic selection can also be combined with gene knock-out *E. coli* strains to select for novel activities on suitable minimal growth media (Yano *et al.*, 1998). A similar method is devised where the DNA and protein produced by *in vitro* expression/translation systems are compartmentalised in an oil emulsion medium so that each droplet contains on average 1 DNA molecule (Tawfik and Griffiths, 1998; Cohen *et al.*, 2004). If the gene

product acts on its own gene, i.e. protection against digestion by methylation, it can be used as a tool selecting for enzyme function.

Spatial addressing

Gene products are often prepared as cell culture, crude cell extract or purified protein in a microwell or as a spot on a solid support such as a microarray protein chip. Each well or spot is assigned a unique address and can be traced back to the DNA source on the master plate. Conventional biochemical assays are performed routinely in wells at micro-scale. Rapid high-throughput protein purification can extend assays to characterisation of physico-chemical properties such as stability or protein structure determination (Heinemann *et al.*, 2001). Screens are typically performed using absorbance or fluorescence based spectroscopic assays (Sundberg, 2000; Pope *et al.*, 1999). Enzyme-linked immunosorbent assays (ELISA) (Engvall, 1977) are well established and also used frequently for screening of binding affinities using protein or antigens immobilised in microwells.

1.2.2.3 Population fitness considerations

A theoretical fitness landscape analogy is used to describe the multidimensional surface that depicts the relation between local or global fitness traits and sequence space. The structure of the fitness landscape is dependent on two experimentally deduced properties: additivity of mutations and coupling of amino acid residues. Mutations are considered additive if the combined fitness change is equal to the sum of individual contributions (Zhang *et al.*, 1995). Smooth fitness landscapes display an additive nature and can be searched easily with adaptive evolutionary walks. Coupling refers to the interactions between amino acid residues which can be complementary or conflicting in nature (Spiller *et al.*, 1999; LiCata and Ackers, 1995). An increase in coupling leads to an increase in the ruggedness of a fitness landscape.

Finding the global fitness peak on a landscape requires movements through sequence space. Jumps in sequence space produced by point mutagenesis are small and often get

trapped in local optima. Larger jumps as produced by recombination or high mutation rates on smaller regions can overcome the constraints of local optima (Crameri *et al.*, 1998).

Evolutionary optimisation using random point mutagenesis can be considered to be an adaptive walk on the fitness landscape. Optimal mutation rate, library size and landscape ruggedness are all important issues to consider when developing an optimisation strategy (Voigt, Kauffman and Wang, 2000). On a rugged landscape, mutants with fitness values lower than the parent are allowed through iterative steps to avoid optimisation towards a local optimum. An optimal mutation rate is required to speed up the adaptive walk and reduce the number of evolution cycles. The optimal mutation rate increases as the library size and landscape ruggedness increases (Voigt *et al.*, 2001). If the mutation rate is too high, the sequence space disperses rapidly and optimisation becomes stochastic rather than deterministic. Based on a theoretical model, the optimal mutation rate on a smooth landscape is 2 mutations per gene when screening a library of a 1000 mutants (Voigt, Kauffman and Wang, 2000).

Strategies using point mutations 'explore' sequence space by creating new diversity. Recombination strategies on the other hand 'exploit' the existing diversity by combining good mutations. Due to the additive nature of some mutations, recombination of point mutations can potentially save experimental effort (Moore *et al.*, 1997).

An optimal cross-over strategy is required to improve the fitness of the progeny when using recombination techniques. Uniform cross-over, where offspring receives the amino acid of either parent with equal probability is favoured on smooth fitness landscapes. Single-point cross-overs are favoured on rough landscapes (Voigt, Kauffman and Wang, 2000). Recombination is only successful as long as the disadvantage of separating beneficial amino acid interactions is less than combining good mutations (Feldman *et al.*, 1996). A theoretical model by Eiben and Back (1997) suggests that the optimal number of parents on a landscape of intermediate ruggedness are six to eight. As the ruggedness increases the optimal number of parents decrease.

Experimental and theoretical studies suggest that short adaptive walks with small libraries may provide the best method for optimising protein properties (Chen and Arnold, 1993; Matsuura *et al.*, 1998). Strategies combining point mutagenesis and recombination are also followed to achieve a balance between exploring sequence space and exploiting existing diversity (Tobin *et al.*, 2000).

1.3 Applications of protein engineering

Protein engineering facilitated, to a large extent, our increased understanding of protein structure-function relations. Perturbation of the protein tertiary structure followed by evaluation of the outcome is used to study first principles of protein folding, stability and activity. Application of protein engineering principles have enabled the protein chemist to design tailored proteins and enzymes required in the fields of biotechnology and biocatalysis. Five protein properties, important in industrial applications, will be discussed in this section.

1.3.1 Enzyme catalytic mechanism

Enzymatic catalysis relies on the action of the amino acid side-chains arrayed in the enzyme active sites. Usually, only two or three essential residues are directly involved in the bond making and breaking steps leading to product formation. Over the years enzymologists have been addressing the roles of such residues by changing them into chemically inert side-chains (Knowles, 1987; Plapp, 1995). In the times pre-dating protein engineering, catalytic residues were identified by modifications of residues with chemical reagents. Loss of protease activity for serine-dependent proteases subtilisin (Neet and Koshland, 1966) or chymotrypsin (Strumeyer *et al.*, 1963) have been reported when changing the catalytic serine into a cysteine or dehydroalanine residue respectively. Protein engineering simplified the task of site-specific chemical modification, used for elucidating catalytic mechanisms. Classic studies that provided proof of acid-base (Steyaert, 1997) or catalytic nucleophile (Carter and Wells, 1990) reaction mechanisms exemplify the value of protein

engineering in this area.

1.3.2 Substrate specific activity

Substrate specificity is the preference that an enzyme manifests for one substrate over competing substrates. Specificity is dependent both on substrate binding and on the utilisation of the binding energy for catalytic turnover. The binding of a substrate is necessary but not sufficient, for catalysis. The stereo- and regioselectivity of enzymes make them attractive tools for synthesising chiral synthon precursors and to this extent a large variety of new and non-natural substrates have been used in biocatalysis (Straathof *et al.*, 2002; Alexeeva *et al.*, 2003). Despite the widespread use of enzymes in asymmetric synthesis, factors determining their structural specificity and stereospecificity remain poorly understood. Protein engineering has been used successfully to change substrate specificity (Onuffer and Kirsch, 1995), cofactor requirements (Mittl *et al.*, 1994) and invert reaction stereochemistry (van Den Heuvel *et al.*, 2000). Strategies used include reshaping substrate or cofactor binding sites, positioning of charged residues to reorient substrates in the active site, or repositioning of catalytic residues inside the active site to redirect their mode of action.

1.3.3 New enzyme activities

It was suggested that chemistry, rather than binding specificity, is the dominant factor in the evolution of new enzymatic activities (Babbitt and Gerlt, 1997). As a consequence, proteins with similar folds can support very different chemical reactions after the incorporation of new catalytic groups. This is the main route for protein engineers to orient an enzyme toward an alternative mechanism. Many enzymes display promiscuous catalytic mechanisms, catalysing reactions distinctly different from their normal reaction (O'Brien and Herschlag, 1999). Tailoring the balance between the two reactions may improve one enzyme function compared with another one (Cedrone *et al.*, 2000).

1.3.4 Protein folding

Theoretical models such as lattice model, dynamic molecular simulations and Go-like models provide the means to rational design of protein folding (Shakhnovich, 1997). Site-directed mutagenesis studies perturbing protein structure complement theoretical studies, leading to greater understanding of folding pathways and rates (Dinner *et al.*, 2000). Increased understanding of protein folding has led to attempts of redesigning protein folding pathways or optimising core packing of novel protein scaffolds (Dahiyat and Mayo, 1997). Increased folding rates can be achieved with a rational approach whereby regions of a protein key in the rate-limiting nucleation-condensation process are stabilised (Guerois and Serrano, 2001). The folding rate of green fluorescent protein has been improved using an approach of random protein optimisation (Merkel and Regan, 2000).

1.3.5 Protein stability

Stability of a protein refers to its ability to remain in the folded state under conditions of increased denaturant, organic solvent temperature, salinity, acidity or alkalinity. There are two distinct definitions of *in vitro* protein stability. Thermodynamic (or conformational) stability is the resistance of protein to conformational denaturation (Tanford, 1968). Long term stability (also kinetic or thermostability) is the resistance to irreversible inactivation and aggregation (Ahern and Klibanov, 1988).

Improvement of biocatalyst stability is of almost universal interest, potentially broadening the applications of existing biocatalysts (Declerck *et al.*, 2000; Shaw *et al.*, 1999; van den Burg *et al.*, 1999). The amount of interactions contributing to stability makes protein engineering of this property a complex task. Comparison of homologous proteins from mesophilic and thermophilic sources has lead to the observation of recurring structural adaptations (Jaenicke, 1991; Vieille and Zeikus, 2001; Fields, 2001). The remainder of the introduction will be aimed at providing a more detailed view on protein stability as well as protein engineering strategies currently used for improvement of stability.

1.4 Engineering of protein stability

1.4.1 Estimation of protein stability parameters

Conformational stability of a protein can be estimated using denaturant or thermal induced unfolding curves, providing an indication of how much more stable the folded structure is compared to the unfolded conformation. These techniques rely on making measurements under equilibrium conditions. Proteins with multiple domains often display equilibrium unfolding curves with multiple transitions (Carra and Privalov, 1996).

A small protein following two-state unfolding is predominantly folded in absence of denaturant and unfolded in high concentrations denaturant. At the midpoint of unfolding the ratio of folded and unfolded species is 1:1 with no other species present. Denaturant is added to shift the equilibrium of the reaction towards unfolded. Measuring equilibrium constants at different concentrations allow stability estimation by means of the linear relation between the denaturant concentration and Gibbs energy (Pace, 1986). The equilibrium solvent denaturation method has two weaknesses. Thermodynamic calculations are made based on the methodological approach that unfolding follows a two-state conformational change. K values can only be determined accurately in the transition region of the two-state sigmoidal curve where significant amounts of both species are measurable. Subsequently, only these data points are used in estimating the linear dependence of ΔG on denaturant concentration. Extrapolation from the unfolding midpoint to 0 M denaturant assumes linear behaviour, even at low concentrations of denaturant. Several theoretical and experimental studies suggest that the linear assumption holds at low denaturant concentrations (Schellman, 1987; Pace and Vanderburg, 1979; Myers *et al.*, 1995). Despite this there are reports of discrepancies in Gibbs energies determined with solvent induced methods and the more accurate thermal induced denaturation studies using differential scanning calorimetry (DSC) (Santoro and Bolen, 1992; Johnson and Fersht, 1995).

DSC is the only technique that determines directly the energetics of the unfolding

process by measuring the excess heat capacity of a buffered protein solution relative to the buffer solution as a function of temperature. The obtained excess heat capacity is used to estimate entropic and enthalpic thermodynamic contributions to the Gibbs energy of unfolding (Privalov, 1989). Thermally induced denaturation is complicated by instrument response times, slow equilibrium unfolding and irreversible transitions. For thermodynamic principles to hold, the unfolding reaction needs to be reversible. DSC usually works well for smaller proteins, but covalent and conformational processes contributing to irreversible aggregation/inactivation are more pronounced for larger proteins (Freire *et al.*, 1990).

The mechanisms of chemical inactivation have been studied extensively and found to occur mainly under conditions of high temperature and alkalinity. Inactivation, is to an extent, also preceded by protein unfolding. Deamidation of asparagine (Asn) and glutamine (Gln) residues occur in the pH range of 4-8 (Zale and Klibanov, 1986). At acidic pH, aspartic acid catalyses the hydrolysis of the adjacent C-terminal peptide bond (Ahern and Klibanov, 1985). Disruption of disulfide bonds, oxidation of cystine and β-elimination occur at elevated temperatures in the pH range of 4-8. β-Elimination of disulfides produces free thiols that cause heat-induced disulfide interchange. Proteins with several disulfide bonds often are scrambled upon refolding and aggregate after becoming chemically trapped (Volkin and Klibanov, 1987).

Cis-configured peptide bonds are often found adjacent to Pro (X-Pro). Both the *cis* and *trans* orientations of proline have comparable intrinsic stabilities. Upon unfolding, the native *cis*-configuration can isomerise to the *trans*-orientation. The half-life of the reaction is 20 min at 0 °C and increases at a rate of 3.3 for every 10 °C (Creighton, 1993). When the rate of refolding is faster than *cis-trans* isomerisation, different refolding rates can be observed (fast and slow folding populations). A non-native *trans* peptide configuration can potentially lead to protein misfolding or formation of inactive enzyme.

Kinetic stability can be defined as the rate at which a protein degrades from the active native state (N) to a chemically inactivated state (I), usually determined by measuring for

biological activity . The rate of chemical inactivation is greatest for enzymes in partially or fully unfolded states (U), hence the reaction are written as $N \rightleftharpoons U \rightarrow I$ (Clarke and Fersht, 1993; Eijsink *et al.*, 2004). Since unfolding is an essential part of the process, kinetic stability depends on the energy barrier of unfolding (activation energy of unfolding) (Vieille and Zeikus, 2001). It can be said that kinetic stability of an enzyme is a measurement the unfolding rate leading to irreversible inactivation. Stability of proteins are typically compared by expressing kinetic stability as as the half-life of enzyme activity at a given temperature. Enzyme samples are incubated at an elevated temperature, removed at time-intervals, cooled down and activity tested. Protein samples lose activity due to chemical inactivation at elevated temperatures and aggregation during the cooling process. Residual activity data plotted against time produces a curve that fits a single exponential decay. Chemical inactivation displays a first-order rate constant and can be used to estimate the half-life of the protein under experimental conditions. Proteins with several inter-linked disulfide bonds adhere more closely to a protein concentration dependent second-order rate of inactivation (Perry and Wetzel, 1987). Use of the Arrhenius equation permits estimation of half-life values at different temperatures, useful for accelerated stability testing of protein formulations (Franks, 1994).

1.4.2 'Rules' for protein stabilisation

When increasing the stability of a protein there are two options: either increase the free energy of the unfolded form, or decrease the free energy of folded form. More than 20 years of protein engineering studies on the stability of small enzymes, displaying reversibly unfolding at high temperatures, have permitted thermodynamic assessment of several types of protein stabilisation interactions. This has led to the identification of several mechanisms that are involved in providing conformational stability.

1.4.2.1 Backbone restrictions

Glycine (Gly) lacks the β-carbon allowing the peptide backbone greater conformational flexibility than in the case of a bulky side-chain of any other amino acid. Gly can be replaced by alanine (Ala) reducing the free energy required to restrict the conformation in the folded form. The bulkier the amino acid the less the rotational freedom and entropy. Of all the amino acids, proline (Pro) places the largest restriction on conformational freedom. Substitution of a select few Gly→Ala or Xaa→Pro residues, causing minimal perturbation to the three-dimensional structure, can provide entropic stabilisation by rigidifications of the peptide backbone (Matthews *et al.*, 1987).

1.4.2.2 Disulfide bridges

Introduction of disulfide bridges can greatly increase the stability of a protein. The major effect of adding a cross-link on the relative stability of the folded conformation results from a decrease in the flexibility and conformational entropy of the unfolded conformation (Poland and Scheraga, 1965; Pace *et al.*, 1988; Matsumura *et al.*, 1989). As a result the relative stability of the native conformation is increased compared to that of the denatured one. Entropy loss is estimated by calculating the probability that two cross-linking ends will simultaneously occur in the same volume element, a sphere with diameter of approximately 4.8 $\mathring{A}$ in the case of S-S linking (Pace *et al.*, 1988). The loss of entropy increases with the number of residues between the linking ends (Poland and Scheraga, 1965; Lin *et al.*, 1984). Literature contains examples where introduction of disulfide bridges improved conformational stability by 1.2 to 4.1 kcal·mol^{-1} (Clarke and Fersht, 1993) and thermostability by 16.7 °C (Mansfeld *et al.*, 1997). Other studies, however, gave disappointing results, mainly attributed to side effects of the individual Xaa→Cys mutations (Wells and Powers, 1986; Mitchinson and Wells, 1989), or the introduction of strain resulting from suboptimal geometry of the disulfide bridge (van den Burg *et al.*, 1993; Matsumura *et al.*, 1989). Strain upon formation of disulfide bonds in the native form can reduce the overall conformational stability gain by reducing relative free energy

difference between native and unfolded forms. Although disulfide bonds can be used to increase stability they are also susceptible to oxidation inactivation at high temperatures or in alkaline conditions.

1.4.2.3 Helix-capping

Aspartic and glutamic acid are observed to occur, with higher than average frequency, toward the amino-termini of α-helices in known protein structures. Lysine and arginine occur with above average frequency toward the carboxy-termini of α-helices, although the preference is not as pronounced as for the acidic residues (Richardson and Richardson, 1988). These residues neutralise the partial positive and negative charges created by the α-helix dipole at the N and C termini respectively. Helix capping is the introduction of residues that interact favourable with the α-helix dipole (Nicholson *et al.*, 1988, 1991).

When selecting sites for helix-capping it is best to avoid positions that already have a negatively charged group located close to the helix N-terminus or positions where the proposed substitution would introduce unfavourable steric interference with the rest of the protein (Nicholson *et al.*, 1988). Several studies have shown consistently that Asp at the N-terminus of a α-helix stabilise the overall protein stability (Sali *et al.*, 1988; Serrano and Fersht, 1989; Marshall *et al.*, 2002).

It is also known that the helix-forming propensities of Gly and Ala depends on the position in the helix. Ala residues, when positioned internal, stabilises the helix markedly more than a Gly residue in the same position. Gly are favoured at the termini of the helix (Serrano *et al.*, 1992). Proline residues in the middle of an α-helix creates considerable distortion of the three-dimensional structure, also rationalising why in nature the frequency of Pro in middle of helix is the lowest of all amino acid residues (Blaber *et al.*, 1993).

1.4.2.4 Electrostatic forces

Thermodynamic evidence suggests that surface ion pairs make little contribution to protein stability (Hendsch and Tidor, 1994; Matthews, 1993; Honig and Nicholls, 1995). It is argued that the entropic cost of localising a pair of solvent-exposed charged groups on the surface of a protein largely offsets the interaction energy expected from the formation of a defined salt bridge (enthalpy). However, several studies have shown that electrostatic interactions engineered by either charge reversal or double mutant cycles can improve stability (Serrano *et al.*, 1990; Schwehm *et al.*, 2003). In these cases the entropic cost of making a surface salt bridge involving the protein backbone, buried salt bridges or salt bridges at dimer interfaces is reduced, since either one or both residues forming the interaction have already been immobilised upon protein folding (Anderson *et al.*, 1990; Horovitz *et al.*, 1990; Vetriani *et al.*, 1998). Residues selected for addition or alteration of electrostatic forces should comply with certain criteria. The sites of modification should be either exposed or partially exposed to solvent; mutations should not interfere with existing hydrogen-bonding or salt bridge interactions and finally substituted side-chains should permit a salt bridge to be constructed with good stereochemistry (Dao-pin *et al.*, 1991; Makhatadze *et al.*, 2003). Theoretical calculations invariably overestimate the stability contributions of coulombic interactions. Several studies empirically estimated electrostatic contributions to be rather small (typically -0.5 to -1.0 kcal·mol^{-1}) and that hydrophobic interactions contribute more to stability than even a buried salt bridge (Waldburger *et al.*, 1995).

It was shown by Grimsley et al that charge reversal of a hyper-exposed residue can lead to the improvement of protein stability of RNAse T1. RNAse T1 is an acidic protein with twice as many carboxylic acids than basic groups. By changing the hyper-exposed negative charge to neutral or positive lead to stability increases of -0.6 and -1.6 kcal·mol^{-1} respectively. Since hyper-exposed residues form no H-bonds or salt bridges, the stability improvements are contributed to more favourable long-range coulombic interactions. Using this approach, stability increases are considered smaller than theoretical predicted

values and occasionally leads to destabilisation of the native state instead. Theoretical calculations do not take into account coulombic interactions existing in the denatured state. Favourable charge-charge interactions may play an important role in the denatured state ensemble so that the free energy of the denatured state may be decreased more than that of the native state (Pace *et al.*, 2000).

1.4.2.5 Hydrophobic and aromatic interactions

The side-chains of phenylalanine and tyrosine residues in proteins are frequently found to be involved in pairwise interactions. These occur both within repeating elements of secondary structure and in tertiary and quaternary interactions. Hydrophobic interactions form the driving force for protein folding and are important in stabilising the hydrophobic core.

There is a distinction between hydrophobic interactions and aromatic-aromatic interactions. Aromatic-aromatic interactions (aromatic pairs) are defined by a distance of less than 7.0 Å (5.5 Å on average) between phenyl ring centroids. A typical aromatic-aromatic interaction contributes -0.6 to -1.3 kcal·mol^{-1} to protein stability (Serrano *et al.*, 1991). The redesign and repacking of the hydrophobic core can play a large role in stabilising proteins as shown by several successful examples of protein engineering (Burley and Petsko, 1985; Anderson *et al.*, 1993; Puchkaev *et al.*, 2003).

In a study conducted by Pace (1992), aliphatic side-chains, buried in folding, were sequentially shortened and the effect on protein stability determined. After normalisation of solvent accessibility to values for 100% burial, it was estimated that a fully buried methylene group contributes on average -1.27±0.51 kcal·mol^{-1} to protein stability. This indicates that a correctly placed mutation, filling an existing cavity in the hydrophobic core can increase protein stability significantly. However, mutations intended to fill cavities often lead to destabilisation of the protein since the increased size of the replacement side chain increases stearic effects, unfavourable van der Waals interactions or repacking/rearrangement of the core amino acids (Ishikawa *et al.*, 1993; Akasako *et al.*, 1997).

1.4.2.6 Observations from thermophilic enzyme structure

Apart from all the mechanisms mentioned in this section several other trends are observed when comparing mesophilic and thermophilic enzymes. In thermophilic proteins, loops are either shortened or better anchored to the rest of the protein. The shortening of loops can be the consequence of the extension of secondary structures (α-helices or β-sheets) or the deletion of loop residues (Russell *et al.*, 1997; Thompson and Eisenberg, 1999). It is argued that exposed loop sites have little interaction between residues or with the rest of the protein. Deletion of a residue from an exposed loop will have little effect on the enthalpy due to minimal contributions made by H-bonds. The solvent interaction will also be the same between native and denatured states so that a deletion will have minimal impact. A deletion will decrease the conformational entropy of both the native and denatured states. While a decrease in entropy of the native state will be localised to the loop region, the conformational entropy of the denatured state will be more extensive due to the multiplicative effect of the deletion on the conformations of the entire chain in the unfolded state. The overall effect will be a relative decrease in the entropy of the denatured state and increased protein stability (Thompson and Eisenberg, 1999).

Loop anchoring and the stabilisation of N and C termini is achieved with ion pairs, H-bonds, or hydrophobic interactions (Hennig *et al.*, 1995; Ermler *et al.*, 1997). Intersubunit interactions and oligomerisation are also considered potential stabilisation mechanisms for extremophile enzymes. A large number of hyperthermophilic proteins are known that have higher oligomerisation states than their mesophilic homologous (Vieille and Zeikus, 2001) and experimental evidence suggest the multimeric forms to be more stable (Thoma *et al.*, 2000) than their monomer subunits.

1.4.3 Engineering approaches for stability improvement

1.4.3.1 Rational

A rational approach to stability improvement requires a detailed inspection of the protein three-dimensional structure. Prediction of structure-based amino acid modifications are

made based on current understanding of first principles of protein stability. Though it is widely accepted that this approach often leads to failure there are reports where protein stabilities have been improved by altering the interactions listed in Section 1.4.2 (Eijsink *et al.*, 2004; van den Burg and Eijsink, 2002). Recently (Schymkowitz *et al.*, 2005) developed the FoldX software application that is capable of fast and accurate estimation of mutational free energy changes ($\Delta\Delta G$) on the stability of a protein. With a high resolution 3-D protein structure as input, FoldX is aimed at providing an *in silico* tool for high-throughput screening by estimating the stabilising effects of single amino acid mutations.

1.4.3.2 Random

Little or no information regarding protein structure is required when using this approach. Amino acids modifications contributing to the improved phenotype are often scattered throughout the the protein sequence. Strategies of single point mutations combined with recombination of the fittest phenotypes are typically used in this approach. The combinatorial nature of this approach is ideal for optimisation of properties, such as stability, where mutations often display an additive nature. This approach has been used successfully to improve protein thermostability (Arnold, 1998; Giver *et al.*, 1998), solvent tolerance (Moore *et al.*, 1997) and resistance to chemical modification (Matsumura *et al.*, 1999).

1.4.3.3 Rational-random

Three-dimensional protein structures often provide rational information regarding key residues involved in substrate binding or catalysis. Random saturation mutagenesis of these key residues has lead to suitably improved or altered enzyme properties (Geddie and Matsumura, 2004; Sio *et al.*, 2002). This approach circumvents the limited successes and numerous pitfalls when redesigning active sites using algorithms for ligand-docking and theoretical structure prediction. Hydrophobic core packing has been optimised with

this approach, improving the overall protein stability (Lee *et al.*, 1996).

The consensus approach is based on the assumption that conserved residues, as inferred from an alignment of homologous amino acid sequences, contribute more to stability than residues that are not conserved. Key residues involved in protein stabilisation can be predicted rationally and optimised in a random manner using site-directed mutagenesis. Several studies have used this approach successfully to improve the thermostability of industrially important enzymes (Lehmann *et al.*, 2000; Lehmann and Wyss, 2001).

1.4.3.4 Random-rational

Eijsink *et al.* (2004) adopted a combined random and rational approach to optimise protein thermostability. Under biocatalytic process conditions enzymes only unfold partially followed by thermal chemical inactivation. A strategy of random mutagenesis was used to identify the so-called 'weak spot' for partial unfolding. A rational approach was used to redesign the 'weak spot'. Backbone rigidification was used successfully to decrease conformational flexibility that contributed to partial unfolding.

1.4.3.5 Non-engineering

Approaches for protein stabilisation may also involve chemical modifications using polyethylene glycol (García-Arellano *et al.*, 2002) or acetylation reagents (Vinogradov *et al.*, 2001). Additives such as glycerol, sorbitol and manitol, generally referred to as polyols, are known to stabilise proteins in solution markedly. It is thought that stabilisation effect of polyols can be attributed to the increases in surface tension of water (Kaushik and Bhat, 1998) and the combination of non-specific excipient-protein interactions (Gekko and Timasheff, 1981). Formulation strategies involving low molecular weight poly-ion additives (MacLean *et al.*, 2002), liposome formulations (Balasubramanian *et al.*, 2000) and emulsification (Kang and Singh, 2003) have also been used successfully to improve the overall stability of proteins in solution. Enzymes used in organic solvents can be treated with crown ethers salts to protect catalytic activity and lipid coating is used to stabilise

proteins in two-phase and super-critical fluid systems. Highly stable cross-linked enzyme crystals (CLEC's) have been used for biotransformation reactions in aqueous or organic media (Bull *et al.*, 1999). Immobilisation on resin is also frequently used to stabilise proteins (Brocklebank *et al.*, 1999).

1.5 Screening of protein stability

The majority of useful biocatalysts are mesophilic in nature and not suitable for applications in industrial biocatalysis. Industrial process conditions for glucose to fructose conversion, using glucose isomerase, are pH 6-8 and 50-75 °C. The production of aspartame with thermolysin as biocatalyst is typically conducted in the presence of an organic solvent such as ethyl acetate at pH 6-8 and 40 °C. There are numerous more examples of reactions conducted in the presence of organic solvents or at high temperature (Sime, 1999; Takayama *et al.*, 1996). The large catalytic potential of enzymes in organic synthesis can only be harnessed if enzymes can be modified on demand to withstand process conditions. One of the main bottlenecks in the route for rapid engineering of stable biocatalysts, i.e. screening power, is manifested in the need for inexpensive and accessible direct high-throughput screens of protein stability. Discussed in this section are the stability screens currently available.

1.5.1 High-throughput stability assays

Optimisation of biocatalyst stability with directed evolution is done by monitoring residual enzyme activity after incubation at elevated temperatures (Giver *et al.*, 1998). This method has been used successfully in several instances producing variant proteins with improved thermal resistance (Kim *et al.*, 2000; Wintrode *et al.*, 2001; Zhao *et al.*, 1998). This screen relies on irreversible inactivation of the protein upon unfolding, and also the assumption that positive variants result from resistance to denaturation under the test conditions. This screen however, may not easily distinguish differences in stability for proteins that spontaneously refold when returned to the activity-assay conditions (Gray

et al., 2001).

Another indirect screen monitors protein aggregation (Won *et al.*, 1998) at elevated temperatures and is typically applied in formulation studies when attempting to optimise the shelf-life of therapeutic proteins in solution. Indirect screens like these have the advantage of simplicity. Light-scattering and residual activity are parameters that can be determined rapidly resulting in fast plate turn-over and large throughput values. Indirect screens essentially only collect data for two time events (i.e. before and after denaturation) making data processing and analysis a relatively simple and effective exercise. For directed evolution applications it is well known that 'you get what you screen for' (Schmidt-Dannert and Arnold, 1999). Due to the complex nature of biological systems, unexpected evolutionarily solutions may occur under the selection pressure of indirect screens. A high-throughput screen that directly measures protein stability is therefore preferred.

Recent advances in instrument technology currently allow the direct high-throughput measurement of protein stability using several techniques. Matrix-assisted laser desorption/ionisation mass spectrometry (MALDI-MS) has been used in conjunction with hydrogen/deuterium (H/D)-exchange to determine changes in protein molecular weight in the presence of different concentrations denaturant (Ghaemmaghami *et al.*, 2000). Provided the H/D-exchange rate is taken into account, sigmoidal unfolding curves are observed when plotting molecular weight against denaturant concentration. This data can be used to estimate the midpoints of unfolding for each variant in a mutant protein library and rank them according to stability. The MicroCal VP-Capillary differential scanning calorimetry (DSC) platform can perform very accurate stability measurements on protein samples in an automated manner (Weber and Salemme, 2003). Calorimetric measurements are made in a capillary chamber, increasing instrument response times. This technology enables the use of temperature gradients of up to 200 $°C \cdot hr^{-1}$, significantly increasing the throughput for this technique.

Direct screening methods evaluate the unfolding process itself where multiple data points are collected in order to calculate thermodynamic values of protein stability. This

implies more sample preparations, measurements, complex data analysis and the subsequent reduction in throughput. The existing MALDI-MS method provides a throughput of approximately 1000 samples per day. The instrumentation is however expensive and not all researchers have access to such dedicated equipment. The MicroCal DSC platform is capable of screening 50 samples per day. Though it is less expensive, it does not produce adequate throughput to screen a medium-sized mutant library in a reasonable time. Based on these figures a good initial sample throughput target for a novel direct screen will be 500-1000 per day. This will allow a sizeable directed evolution library to be screened in one working week.

1.6 Aims of study

This study has several aims that are addressed in three experimental chapters and laid out as follows.

High-throughput stability screen: The first aim was to develop a novel, automatable and inexpensive high-throughput screen for the direct measurement of protein stability. Chapter 3 describes the evaluation of high-throughput equipment, method optimisation and validation using well characterised, pure protein standards.

Unfolding characterisation of transketolase: The second aim, described in Chapter 4, was to conduct a preliminary study of the stability and unfolding characteristics of *E. coli* transketolase (*tkt*) under typical process conditions. A His_6-tag added onto the N-terminus of the wild-type *tkt* gene enabled rapid protein purification, required for both preparative and high-throughput experimentation. Mutant variants of the *tkt* gene, with reduced dimer stability, were designed and provide more insight into the unfolding characteristics of transketolase. The mutants are also useful for assessing the feasibility of optimising transketolase stability, using directed evolution strategies and the developed high-throughput screen.

High-throughput route for stability measurement: The final aim of this study was to optimise a generic, high-throughput route for screening protein stability using transketo-

lase as a model system. Chapter 5 describes optimisation of individual steps in the route which include cell-culturing, protein purification, buffer exchange and stability screening, all conducted in a high-throughput format. A novel dialysis apparatus was designed, tested and used routinely to facilitate high-throughput buffer exchange.

Chapter 2

Materials and Methods

2.1 Cytochrome *c* and BSA sample preparation

Cyt *c* was oxidised by adding 20 µl of 0.1 M potassium ferricyanide per 1 ml protein solution. The oxidised cyt *c* preparations were dialysed for 24 hr at 4 °C with two buffer exchanges using 10 kDa MWCO dialysis tubing (Perbio Science, Cheshire, UK). The absence of an absorbance peak at 550 nm confirms the complete oxidation of the heme moiety (Yonetani, 1965). Prepared cyt *c* samples were used immediately.

BSA samples containing various molar ratios of palmitic acid were prepared with the procedure of Brodersen *et al.* (1988). Briefly, 300 µM BSA solution in deionised H_2O was prepared and adjusted to pH 9 with NaOH. A 90 mM ethanolic solution of palmitic acid was prepared and sufficient volume added to the BSA solution to give BSA:palmitate molar ratios from 1:1 to 1:6. The preparations were dried using a centrifugal vacuum evaporator (Speed-Vac SVC100H, Savant Instruments, NY) and the dried material stored at -20 °C until used.

The concentrations of protein stock solutions were determined from absorbance measurements, using the molar absorption coefficients values of 106000 M^{-1} at 410 nm for cyt *c* (Margoliash and Frohwirth, 1959) and 43890 M^{-1} at 280 nm for BSA (Pace *et al.*, 1995). The denaturant stock buffers used were 100 mM Tris, pH 7.0, 6.5 M GdnHCl for cyt *c* and 60 mM sodium phosphate, pH 7.0, 9.0 M urea for BSA.

2.2 High-throughput fluorescence measurement

All baseline and protein unfolding transitions were measured in F-type, polystyrene, 96 well microtiter plates (Greiner Bio-One Ltd, Gloucestershire, UK) using a fluorescence microplate reader, Fluostar Optima (BMG Labtechnologies Ltd, Bucks, UK), fitted with a syringe pump for automation of reagent injection. Protein unfolding was monitored by intrinsic tryptophan fluorescence using a 340 ±10 nm emission window filter with excitation at 280 nm. A total fluorescence reading is obtained since the entire well content is sampled upon measurement.

2.3 Selection of microwell plate type

The baseline fluorescence and liquid volume were identified early on as a critical factors in experimental design. 50 µl of 0.1 M Tris-buffer pH 7.0 containing 0 M or 6.5 M GdnHCl were loaded into black well or clear well plates and fluorescence recorded at 340 nm after excitation at 280 nm. In the case of black well plates 4 µl additions were made followed by fluorescence measurement. For clear well plates the buffer additions increased in volume. The fluorescence signals were plotted against volumes for both solutions and plate types.

2.4 Auto-injector measurement

Several 3 µl and 5 µl injections were made into a single well of a pre-weighed 96 well microtiter plate (Section 3.2.2). The microtiter plate was weighed after each injection. Measurements were plotted as total volume injected against the total weight increase. In another test, larger volumes (10, 25 and 50 µl) were injected using the same procedure (Section 3.2.2). Weight measurements obtained were deemed reliable based on agreement with dispensed volumes and the density of water at 25 °C.

2.5 Microwell stability measurement

Two different protocols were used to generate protein unfolding curves in microtiter plates, one in which the final volume at each denaturant concentration was fixed and the other in which serial addition of denaturant to a protein solution gradually increased the total volume. Both protocols were automated, using the syringe pump to inject all the solutions. The two approaches allowed us to compare the effect of decreasing protein concentration on stability determination with that of keeping it constant.

Fixed volume denaturation: 20 μg cyt *c* in a volume of 50 μl, or 40 μg BSA in 20 μl was loaded into each well of a 96 well microtiter plate. For cyt *c*, wells were filled with different amounts of buffer containing 0 M, and then 6.5 M GdnHCl to a final volume of 215 μl, to generate a final denaturant concentration range of 0 to 3.8 M. For BSA, wells were filled using buffer containing 0 M and then 9 M urea to a final volume of 280 μl, for a urea concentration range of 0 M to 8.4 M. Each sample was repeated in triplicate (Section 3.2.3).

Serial addition of denaturant: With this protocol, protein was loaded into each well followed by the addition of small volumes of buffered denaturant stock. Each addition was followed by 30 seconds of mixing by orbital shaking at 350 rpm, and 15 minutes of equilibration prior to taking measurements. Volumes for serial addition were adjusted to provide nearly equidistant data points on the unfolding curve. The minimum volume of the initial protein solution that can cover the well bottom is 50 μl. Using this protocol, 20 μg cyt *c* or 40 μg BSA was loaded per well (Sections 3.2.3 and 3.2.4).

2.6 Estimation of stability parameters

All data fitting and analysis was performed using the software package SigmaPlot 8.0 (SPSS UK Ltd., Surrey, UK). For a two-state model the native and unfolded protein species are the only significant populations present. Pre- and post-unfolding baselines are determined using Equation 2.1 and Equation 2.2 where F_N° is the fluorescence signal for

the native state at 0 M denaturant, and m_N is the slope ($\delta F_N^\circ/\delta[D]$) for the native state baseline. F_U° and m_U are the corresponding parameters for the unfolded state.

$$F_{obs,N} = F_N^\circ + m_N[D] \tag{2.1}$$

$$F_{obs,U} = F_U^\circ + m_U[D] \tag{2.2}$$

The intercept and slope values were substituted into Equation 2.3 (Santoro and Bolen, 1988) where R is the gas constant (1.987 cal·mol^{-1}) and T the absolute temperature (298 K). $C_{1/2}$ (M) is the denaturant concentration where half the protein is in the unfolded state and m_G (kcal·mol^{-1}·M^{-1}) is the change in free energy of unfolding with respect to the denaturant concentration. The raw fluorescence data was fitted to Equation 2.3 and the two parameters, m_G and $C_{1/2}$, estimated with errors.

$$F_{obs} = \frac{(F_N^\circ + m_N[D]) + e^{\left(\frac{m_G[D-C_{1/2}]}{RT}\right)}(F_U^\circ + m_U[D])}{e^{\left(\frac{m_G[D-C_{1/2}]}{RT}\right)} + 1} \tag{2.3}$$

The free energy of protein unfolding was calculated by the linear extrapolation method (Pace, 1986), in which the free energy of unfolding is a function of denaturant concentration (Equation 2.4). The ΔG_{obs} values were calculated from data points in the transition region using Equation 2.5 and the relationship $\Delta G_{obs} = -RT \cdot ln \cdot K_{obs}$. The ΔG_{obs} values were plotted against denaturant concentrations and extrapolation towards 0 M determines the stability of the native protein (ΔG_{H_2O}).

$$\Delta G_{obs} = \Delta G_{H_2O} + m_G[D] \tag{2.4}$$

$$K_{obs} = \frac{F_{obs} - F_{obs,N}}{F_{obs,U} - F_{obs}} \tag{2.5}$$

2.7 Numerical methods for high-throughput indexing of relative protein stability

High-throughput data processing forms an essential part of the overall screening process. Three simple numerical methods were tested for their suitability to index the relative stabilities of normalised BSA unfolding curves (Section 3.2.5). BSA samples with different palmitate content and therefor different ligand induced stability values were used. The three numerical methods are described in this section.

2.7.1 Indexing with an apparent two-state model.

The unfolding curves of BSA do not follow the two-state model. Despite this, the data was fitted to Equation 2.3 in Section 2.6 to estimate an apparent $C_{1/2}$-values. Since normalised data is used the upper and lower baselines are fixed with $F_N^\circ = 0\%$ and $m_N = 0$ for Equation 2.1 and $F_U^\circ = 100\%$ and $m_U = 0$ for Equation 2.2. For clarity the method is depicted in Figure 2.1 using a hypothetical data set and results are shown in Section 3.2.5.

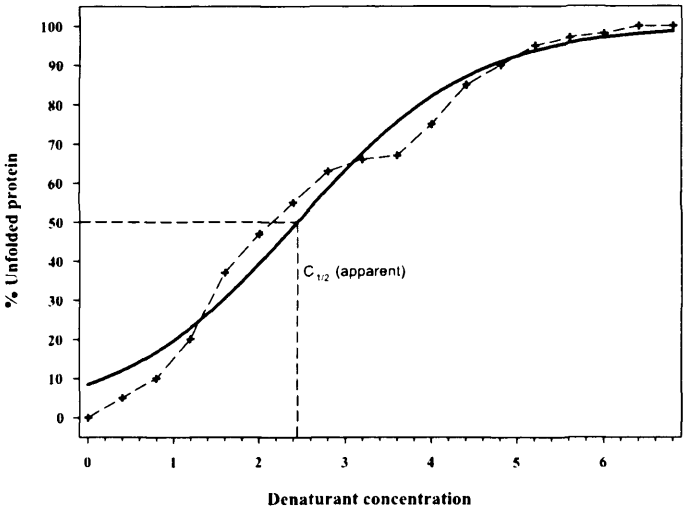


Figure 2.1: Theoretical example for estimation of apparent $C_{1/2}$-values for a non-two-state unfolding curve.

2.7.2 Indexing with area under unfolding curve

The area under an unfolding curve for a fixed interval on the x-axis was determined by numeric integration using the trapezoidal rule. A Sigma Plot macro was used to automate the estimations of the area under curve (AUC) values that served as the stability indices. The method of estimation using the trapezoidal rule is depicted in Figure 2.2. There curve is divided into segments as determined by the number and location of data points. The area for each segment is calculated by combining the area of a rectangle and triangle drawn inside the segment. Summing of the areas of all the segments provide the total area under the unfolding curve. The summed x-value interval of all unfolding curves has to be identical to allow comparison of indices. The area of the trapezoid for a given segment (i) is given by

$$Trapezoidal\ Area = y_i \cdot (x_{i+1} - x_i) + \frac{1}{2} \cdot (y_{i+1} - y_i) \cdot (x_{i+1} - x_i) \tag{2.6}$$

and the total AUC is calculated by summing all the trapezoidal areas using

$$AUC = \sum_{i=1}^{n} \left(y_i \cdot (x_{i+1} - x_i) + \frac{1}{2} \cdot (y_{i+1} - y_i) \cdot (x_{i+1} - x_i) \right) \tag{2.7}$$

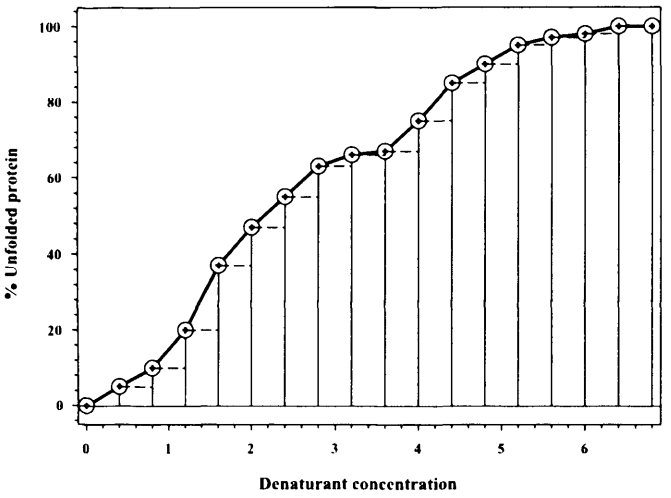


Figure 2.2: Theoretical example for estimation of the area under curve (AUC) using the trapezoidal rule.

2.7.3 Indexing with arbitrary unfolding point reference

An arbitrary reference point for the percentage unfolding was selected, for example 60 or 70% and the corresponding denaturant concentration taken as the index of stability. The method is explained at the hand of Figure 2.3 where the chosen reference point of 60% is used. The two data points (A & B), on either side of 60% were used to fit a straight line AB with equation

$$y = m_{60} * x + c_{60} \tag{2.8}$$

where the slope, m_{60}, is estimated using

$$m_{60} = \frac{x_B - x_A}{y_B - y_A} \tag{2.9}$$

and the intercept, c_{60}, equals y at x = 0 in Equation 2.8.

The stability index value (I_{60}) at reference point y = 60% is calculated with

$$I_{60} = \frac{60 - c_{60}}{m_{60}} \tag{2.10}$$

The same procedure can be followed to determine I_{70} at y = 70% using the points C & D that lie adjacent on either side of the 70%-mark, forming straight line CD. The arbitrary point used for indexing will differ from protein to protein and is positioned where the unfolding curve is most sensitive to changes in stability. Results for palmitate stabilised BSA unfolding curves are shown in Section 3.2.5.

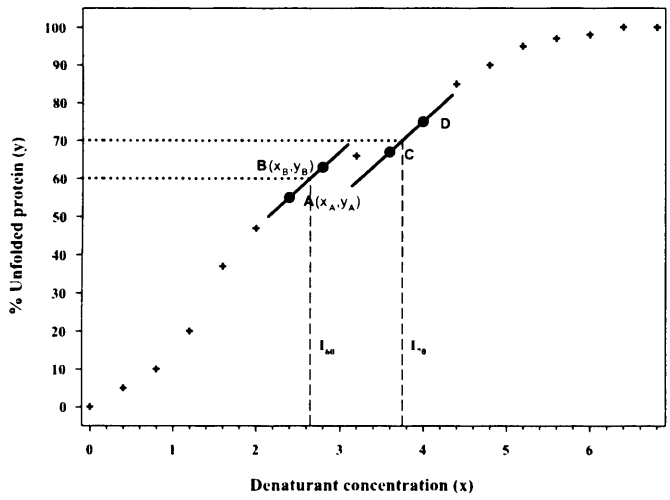


Figure 2.3: Theoretical example for estimation of stability index based on an arbitrary point of reference.

2.8 QuickChange^{TM} PCR reactions

The TK-15MutF and TK-15MutR DNA-oligomers (Table 2.1) were obtained from Amersham Pharmacia Biotech (Uppsala, Sweden) and used to insert an unique *Bgl*II restriction site into pQR711, creating pQR790 (Section 4.2.1). The TKHisF and TKHisR DNA-oligomers (Table 2.1) were obtained from Cruachem Limited (Glasgow, UK) and used for His-tag modification of pQR790, creating pQR791 (Section 4.2.1). The D381AFwd, D381ARev, Y440AFwd and Y440ARev DNA-oligomers (Table 2.1) were obtained from

Table 2.1: DNA-oligomers used for PCR amplifications and DNA sequencing.

Name	Sequence	T_m (°C)
TK-15MutF	$^{5'}$ccttcatcatcagatctggagtca$^{3'}$	58
TK-15MutR	$^{5'}$tgactccagatctgatgatgaaggg$^{3'}$	58
TKHisF	$^{5'}$atcagatctggagtcaaaatgcatcaccatcaccatcactcctcacgtaaagagcttgcc$^{3'}$	73
TKHisR	$^{5'}$ggcaagctctttacgtgaggagtgatggtgatggtgatgcattttgactccagatctgat$^{3'}$	73
D381AFwd	$^{5'}$ttcctcggcggttctgctgctctggcgccgtctaacctga$^{3'}$	78
D381ARev	$^{5'}$tcaggttagacggcgccagagcagcagaaccgccgaggaa$^{3'}$	78
Y440AFwd	$^{5'}$ccttcctgatgttcgtggaagctgcacgtaacgccgta$^{3'}$	74
Y440ARev	$^{5'}$tacggcgttacgtgcagcttccacgaacatcaggaagg$^{3'}$	74
TKN	$^{5'}$gatccagagatttctga$^{3'}$	50
SeqTK1180Fwd	$^{5'}$ctaaagcaatcaacgaagatgctg$^{3'}$	61
LibCeA	$^{5'}$cccatagggggcccccgg$^{3'}$	60

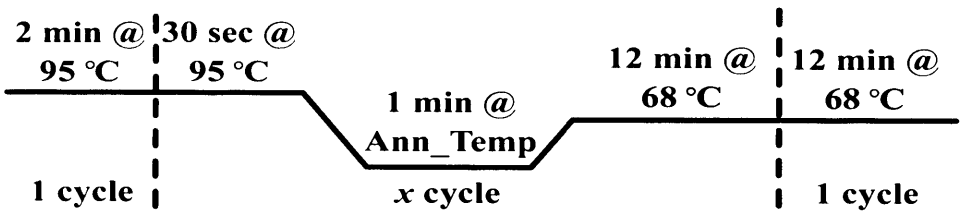


Figure 2.4: Generic thermocycler profile used for PCR reactions.

Table 2.2: Components of the PCR cocktails.

Volume (µl)	Reagent
(A) Two primer cocktail for PCR reactions	
1.0	DNA template
0.33	TK-15MutF (50 µM)
0.33	TK-15MutR (50 µM)
5.0	10X *Pfu* Reaction buffer
1.0	dNTP mix (100 mM)
42.34	H_2O (sterile deionised)
1.0	*PfuTurbo* (2.5 U·µl^{-1})
(B) Single primer cocktail for PCR reactions	
0.5	DNA template
0.2	Forward or Reverse primer (50 µM)
5.0	10X *Pfu* Reaction buffer
1.0	dNTP mix (8 mM)
43.3	H_2O (sterile deionised)
0.4	*PfuTurbo* (2.5 U·µl^{-1})

Table 2.3: Conditions for modified QuickChange™ PCR temperature cycles.

Plasmid Modification	Annealing Temperature (°C)	Cycles (1^{st} stage)	Cycles (2^{nd} stage)	ng DNA Template
His-insertion	70	10	18	18
D381A mutation	55	12	18	15
Y440A mutation	55	12	18	15

OPERON Biotechnologies GmbH (Cologne, Germany). These DNA-oligomers were used to create the D381A and Y440A mutants using the pQR791 plasmid as template (Section 4.2.1). All the DNA-oligomers sequences and melting temperatures (T_m) are listed in Table 2.1.

The *PfuTurbo* DNA-polymerase, *Pfu* reaction buffer and dNTP's were obtained from Stratagene, (La Jolla, CA, USA). The -15 *Bgl*II modification was produced using the standard Stratagene QuickChange^{TM} procedure. A PCR cocktail was prepared containing both the forward and reverse primers (Table 2.2 A). The polymerase chain reaction (PCR) was performed with the temperature cycling profile depicted in Figure 2.4 for 12 cycles with an annealing temperature of 55 °C.

A modified two-stage QuickChange^{TM} procedure (Wang and Malcolm, 1999) was used to produce all the other plasmid modifications. Forward and reverse primers were used separately in the first PCR amplification, producing modified single strand plasmid (Table 2.2 B). 25 µl from each of the forward and reverse primer single strand PCR products were pooled and 0.4 µl *PfuTurbo* added for a second round of PCR amplification. During the second amplification the single strand plasmid serves as template producing complete double stranded plasmid with the appropriate modification. The PCR amplifications were done using a generic temperature cycling profile (Figure 2.4). For each plasmid modification the annealing temperature and number of cycles that produced a successful reaction are listed in Tabel 2.3. The PCR reactions were carried out using a Techgene thermal cycler (Techne Limited, Cambridge, UK).

2.9 Restriction digestion reactions

Preparative restriction digestion was done by adding 1 µl of the *Dpn*I restriction enzyme (10 U·ml^{-1}) directly to each amplification to remove methylated parental plasmid DNA. The reaction mixture was incubated in a water bath at 37 °C for 1 hr. The remaining plasmid product was used for cell transformation.

A diagnostic restriction digestion was done with 1 µg plasmid DNA in 20 µl 1X universal

reaction buffer containing 1 U·ml^{-1} of both *Bgl*II and *Xba*I. The reaction mixture was incubated in a water bath for 90 min at 37 °C. Appropriate negative control reactions were included.

2.10 DNA gel electrophoresis and visualisation

All PCR products and plasmid preparations were analysed using 1.2% agarose-TBE gel electrophoresis. Plasmid DNA was mixed with 6X DNA loading buffer (Novagen, Madison, WI, USA) at a ratio of 3.0 µl:0.6 µl before loading. 1 µl of Perfect DNA Markers 0.5-12kb (Novagen) was used to determine the plasmid and PCR product size. Sufficient separation of samples and marker DNA was achieved when applying 60 V for 40-60 min using a EPS400/500 electrophoresis power supply (Pharmacia Fine Chemicals, Uppsala, Sweden). Agarose gels containing 0.005% (v/v) ethidium bromide were visualised under ultraviolet light and images captured using GelDoc-It Bio-Imaging System (UVP Ltd, Cambridge, UK).

2.11 Media and buffers for molecular biology

2.11.1 Luria Bertani (LB) media

LB media was prepared by dissolving, as (w/v), 1% NaCl, 1 % tryptone and 0.5 % yeast extract in deionised H_2O. Media pH was adjusted to 7.0 with 1.0 M NaOH and then autoclave sterilised.

2.11.2 SOC media

SOC media was prepared by dissolving, as (w/v), 0.05 % NaCl, 2 % tryptone and 0.5 % yeast extract in deionised H_2O. Media was autoclave sterilised. Prior to use 10 ml of both 1 M $MgCl_2$ and 1 M $MgSO_4$ as well as 1 ml of 2 M glucose, all filter sterilised, were added per liter media.

2.11.3 LB-Ampicillin Agar

LB agar was prepared by adding 1.5 % (w/v) agar from BDH Chemicals (Poole, Dorset, UK) to LB media followed by autoclave sterilisation. A stock solution of 50 $g \cdot l^{-1}$ ampicillin was prepared and filter sterilised. Autoclaved LB agar was allowed to cool to approximately 55 °C and 2 µl of the ampicillin stock was added for every 1 ml LB agar. Between 10 ml and 15 ml LB agar were poured into Petri dishes and allowed to set. The remainder of the ampicillin stock was stored at -20 °C.

2.11.4 TBE buffer

10X TBE buffer was prepared by dissolving 107.78 g Tris-base (0.89 M), 7.44 g Na_2-EDTA (0.02 M) and 55.0 g boric acid (0.89 M) in deionised H_2O to a final volume of 1 liter. The pH of 8.3 was confirmed. The stock was diluted 10-fold for preparation of agarose gels and to serve as reservoir buffer during gel electrophoresis.

2.12 Transformation of *E. coli* strains

2.12.1 XL1-Blue and BL21

Competent XL1-Blue or BL21 cells (Stratagene) were thawed and 100 µl transferred to a pre-chilled Falcon tube. β-Mercaptoethanol (β-ME) was prepared at a concentration of 1.42 M with deionised H_2O from a 14.2 M solution (Sigma Chemical Co.). 1.7 µl of the prepared β-ME was added and cells incubated on ice for 10 minutes. 1 µl of plasmid DNA was added to cells and incubated on ice for a further 30 minutes. The transformation reaction was heat-pulsed for 45 seconds at 42 °C and then incubated on ice for another 2 minutes. A total of 900 µl SOC media (heated to 42 °C) was added and cells incubated at 37 °C for 1 hr with shaking at 200 rpm. Between 5 µl and 100 µl transformed cells were plated out on LB-ampicillin agar. Colonies were picked from plates and incubated overnight in LB media containing 100 $mg \cdot l^{-1}$ ampicillin to produce master stock cultures. Master cultures were preserved in 40% (v/v) sterile glycerol and stored at -70 °C.

2.12.2 JM107

50 µl competent JM107 cells, obtained from J. Ward (UCL), was transferred into a pre-chilled Falcon tube. 1 µl plasmid DNA was added to the cells and incubated for 20 minutes on ice. The JM107 cells were heat shocked at 42 °C for 45 seconds and transferred back onto ice for 2 minutes. A total of 900 µl SOC media (heated to 42 °C) was added and cells incubated at 37 °C for 1 hr with shaking at 200 rpm. Between 50 µl and 200 µl transformed cells were plated out on LB-ampicillin agar. Colonies were picked from plates and incubated overnight in LB media containing 100 $mg{\cdot}l^{-1}$ ampicillin to produce master stock cultures. Master cultures were preserved in 40% (v/v) sterile glycerol and stored at -70 °C.

2.12.3 XL10-Gold

XL10-Gold ultracompetent cells (Stratagene) were thawed on ice and 50 µl aliquots transferred into a pre-chilled Falcon tube. 4 µl of the β-ME, provided with this kit, was added to each aliquot of cells and gently mixed while incubated on ice for 10 minutes. 2 µl of the PCR product was added to cell aliquots and incubated on ice for 30 minutes. The XL10-Gold cells were heat shocked in a 42 °C water bath for 30 seconds and transferred back onto ice for 2 minutes. 0.6 ml of preheated (42 °C) NZYM broth (Sigma Chemical Co.) was added and cells incubate at 37 °C for 1 hr with shaking at 200 rpm. Between 5 µl and 50 µl transformed cells were plated out on LB-ampicillin agar. Colonies were picked from plates and incubated overnight in LB media containing 100 $mg{\cdot}l^{-1}$ ampicillin to produce master stock cultures. Master cultures were preserved in 40% (v/v) sterile glycerol and stored at -70 °C.

2.13 DNA plasmid preparation and sequencing

Plasmid DNA was extracted from transformed cells using a commercial plasmid mini-preparation kit (Qiagen Ltd, West Sussex, UK). The final product was eluted into 75 µl

EB buffer. Plasmid preparations were diluted 1:10 for spectrophotometric quantification at 260 nm (UV/Vis2, Unicam Ltd Cambridge, UK) using a quartz microcuvette (Hellma UK Ltd, Essex, UK). The 260 nm:280 nm ratio was determined to verify DNA purity.

DNA sequencing was done by the Sequencing Service at the Wolfson Institute for Biomedical Research (UCL). The construction of pQR791 was verified using the TKN sequencing DNA-oligomer and both the D381A and Y440A mutations were confirmed using the LibCeA sequencing DNA-oligomer.

2.14 Ni-NTA spin columns purification of transketolase

BL21 and JM107 overexpressing His_6-transketolase were cultured overnight in two flasks each containing 100 ml LB media and 100 $mg \cdot l^{-1}$ ampicillin (Section 4.2.2). The final A_{600nm} was adjusted to 5.0 AU for both shake flasks using sterile LB media. 50 ml cells were harvested at 4000 rpm with Sorvall T-21 (Kendro Laboratory Products Plc, Hertfordshire, UK) and resuspended in 1 ml of 50 mM Tris, 18 mM $MgCl_2$, 0.5 mM TPP, pH 8.5. The cells were lysed on ice using a Soniprep 150 ultrasonicator (Sanyo Gallenkamp, Leicestershire, UK) with a protocol of 5 seconds on/off for 10 cycles at 8 μm amplitude. Cell debris was removed by centrifugation at 16000 rpm for 20 min. The transketolase was concentrated 50-fold in the cell-free extract since the volume decreased from 50 ml down to 1 ml.

His_6-transketolase from both BL21 and JM107 were purified using Qiagen Ni-NTA spin columns. For each purification segment the appropriate solution was loaded onto the spin column and passed through the column under centrifugal force created by centrifugation at 2000 g for 2 min. Initially a Ni-NTA spin column was equilibrated with 600 μl lysis buffer containing 50 mM Tris-base, 300 mM NaCl and 10 mM imidazole, pH 8.0. The cell-free extract was diluted 10-fold in the lysis buffer and 600 μl was loaded onto a spin column. The spin column was washed twice with 600 μl wash buffer containing 50 mM Tris-base, 300 mM NaCl and 20 mM imidazole, pH 8.0. His_6-transketolase was eluted from the spin column in two fractions of 200 μl with elution buffer containing 50 mM Tris-base,

300 mM NaCl and 250 mM imidazole, pH 8.0. The cell-free extract, flow-through, wash and elution fractions were collected for SDS-PAGE analysis (Sections 2.20 and 4.2.2).

2.15 Preparative purification of transketolase

Chelating Sepharose Fast Flow resin from Pharmacia Biotech Ltd. (Uppsala, Sweden) was suspended in deionised H_2O and packed in a XK16/20 column (Pharmacia Fine Chemicals) to a bed height of 10 cm (20 ml resin). The resin was packed at a flow rate of 1.5 ml·min^{-1} until the bed height stabilised. All segments of the purification were performed at a flow rate of 1 ml·min^{-1}. Column was equilibrated with 3 column volumes (CV) of deionised H_2O. Nickel was immobilised onto the resin by pumping 1 CV of 0.2 M $NiSO_4$ through the column. Excess Ni^{2+} was removed by washing the column with 5 CV deionised H_2O.

Column was equilibrated with 3 CV start buffer (0.25 M Tris-base, pH 8.5) before loading a 100 ml cell-free extract. The cell-free extract was prepared from 600 ml JM107 pQR791. Cells were harvested by centrifugation at 4000 rpm for 10 min. The cell paste was resuspended in 60 ml start buffer. The cells were lysed with an ultrasonication protocol of 10 seconds on/off at 8 μm amplitude for 10 cycles. Cell-free extract was cleared by centrifugation at 16000 rpm for 20 minutes and made up to 120 ml with buffer.

Column was washed with another 2 CV start buffer before the protein was eluted with a gradient of imidazole from 0 M to 1.0 M over 100 ml. The elution gradient was created with start buffer containing 1 M imidazole. Protein elution was monitored with absorbance at a wavelength of 280 nm. 5 ml fractions were collected and those containing transketolase were pooled. Eluted His_6-transketolase was dialysed using Snakeskin^{TM} dialysis tubing with a MWCO of 10 kDa (Perbio Science UK Ltd., Northumberland, UK). Imidazole was removed by dialysing against two exchanges of 1 liter of 0.25 M Tris-base, pH 7.5 buffer. The dialysed protein was divided into 1.5 ml aliquots and dried using a centrifugal vacuum evaporator (Speed-Vac SVC100H, Savant Instruments, NY). The dried material stored at -20 °C until used. After purification the column was regenerated with 5 CV of

0.05 M EDTA, 0.5 M NaCl and preserved in 0.5 M NaCl.

2.16 Size-exclusion chromatography

Size-exclusion chromatography (SEC) was used to resolve the TPP cofactor from the purified transketolase under alkaline conditions. Superose^{TM} 12 resin from Pharmacia Biotech Ltd. (Uppsala, Sweden) was suspended in deionised H_2O and packed in a HR 16/50 column (Pharmacia Fine Chemicals) to a bed height of 30 cm (50 ml resin). The resin was packed at a flow rate of 2 ml·min^{-1} for 60 min and then at 4 ml·min^{-1} until the bed height stabilised. All segments of the purification were performed at a flow rate of 1 ml·min^{-1}. Column was equilibrated with 3 CV of start buffer (0.25 M Tris-base, pH 8.5). Samples of 0.1 ml pure transketolase with concentration between 1 mg·ml^{-1} and 3 mg·ml^{-1} was injected onto the column and eluted as a single peak in a volume of approximately 2.5 ml. Protein elution was monitored with absorbance at a wavelength of 280 nm. Between runs the column was regenerated with 20 ml 0.1 M NaOH, 2 CV deionised H_2O and 3 CV start buffer. Column was stored in 20% ethanol (v/v).

2.17 Absorbance enzyme-linked assay

The enzyme-linked assay was done as reported by French and Ward (1995). Phosphoriboisomerase (PRI), α-glycerophosphate dehydrogenase-triosephosphate isomerase (GDTPI), D-ribulose-5-phosphate-3-epimerase (DR5P3E), nicotinamide adenine dinucleotide in the reduced form (NADH) and D-ribose-5-phosphate (DR5P) were obtained from Sigma Chemical Co. All the components for the enzyme-linked assay were prepared as stock solutions in 50 mM Tris-base, pH 8.5 buffer and stored at -20 °C. Stock concentrations, final assay concentrations and volumes per 1 ml assay mix are listed in Table 2.4. PRI, DR5P3E and DR5P were preincubated at room temperature for 2 hr, providing an equilibrated substrate mix. Transketolase samples were diluted appropriately in 50 mM Tris-base, 9 mM $MgCl_2$ 0.25 mM TPP, pH 8.5 and incubated for 30 min. Typically trans-

Table 2.4: Components for absorbance enzyme-linked assay.

Component	Concentration of Stock	Assay Volume (µl)	Concentration (Final)
PRI	10 $U \cdot ml^{-1}$	20	0.2 $U \cdot ml^{-1}$
DR5P3E	10 $U \cdot ml^{-1}$	20	0.2 $U \cdot ml^{-1}$
DR5P	30 mM	100	3 mM
GDTPI	2 $U \cdot ml^{-1}$	100	0.2 $U \cdot ml^{-1}$
NADH	1.54 mM	100	0.154 mM
Buffer	-	560	-
Transketolase	-	100	-

Table 2.5: Reaction mixtures for the transketolase inhibition studies using enzyme-linked and HPLC activity assays.

(A) Enzyme-linked assay		(B) HPLC Assay	
Component	Assay Volume (µl)	Component	Assay Volume (µl)
PRI	20		
DR5P3E	20	Glycolaldehyde (1 M)	100
DR5P	100	β-HPA (125 mM) &	
GDTPI	100	Inhibitor (625 mM)	400
NADH	100	Cell-free extract	500
Buffer with 500 mM inhibitor	500		
Buffer	60		
Cell-free extract	100		

ketolase samples were diluted to produce a reaction rate of no more than 0.1 δA·min^{-1}. GDTPI, NADH, transketolase and buffer was dispensed into a cuvette. The mixture was kept out of light and allowed to reach ambient temperature before rate determination. Reaction was started by adding the substrate mix and NADH consumption was recorded at a wavelength of 340 nm for 5 min. The NADH concentration was determined using a molar extinction coefficient of 6.22 mM^{-1}·cm^{-1} at 340 nm and enzyme activity per 1 ml reaction was expressed in U (µmol·min^{-1}) using Equation 2.11. Dilution factors and total volumes were taken into account before reporting the final activity as U·ml^{-1}. The assay mix was modified for the platereader by using twice as much NADH (3.08 mM) to increase the signal of A_{340nm} absorbance.

$$v = \frac{\delta A}{6.22\ \mu mol^{-1} \cdot min} \tag{2.11}$$

2.18 HPLC activity assay

This protocol was developed by C. Ingram (UCL). The HPLC system consisted of a GP50 gradient pump, an LC30 chromatography oven, a PC10 pneumatic controller and an AD20 absorbance detector (Dionex Corp.). PeakNet 5.1 software (Dionex Corp.) was used for instrumentation control and data collection. Samples were injected with an Endurance autosampler (Spark Holland BV).

The mobile phase was 0.1% (v/v) trifluoroacetic acid (TFA) and the flow rate was 0.8 ml·min^{-1}. Separation of the components was achieved using a 300mm Aminex HPX-87H ion-exclusion column (Bio-Rad Laboratories). The temperature of the column was maintained at 60 °C by the chromatography oven. The detector monitored the absorbance of the output stream from the column at 210 nm wavelength using the 0.1 AU sensitivity setting. The injection volume was 10 µl for all samples.

The reaction sample was diluted 1:4 by the addition of 0.2% (v/v) TFA. The addition of acid quenched the reaction by dropping the pH below the lower limit for transketolase activity (Mitra *et al.*, 1998). The concentration of L-erythrulose was determined by using a suitable standard curve.

2.19 Transketolase inhibition study

The levels of inhibition of transketolase by 250 mM imidazole or NaCl have been determined. Transketolase activity rates, in cell-free extract, were determined in absence and presence of 250 mM inhibitor using the enzyme-linked (Section 2.17) and HPLC activity assays (Section 2.18). The reaction mixture for the enzyme-linked assay inhibition study is listed in Table 2.5 A. The assay reaction was initiated by adding transketolase to avoid pre-incubation of enzyme and inhibitor. Table 2.5 B lists the reaction mixture of the HPLC inhibition study. The enzyme reaction was allowed to progress for 1 hr and stopped by addition of 0.2% (v/v) TFA. The percentage of inhibition was determined using control enzyme reactions containing no inhibitor.

2.20 SDS and native PAGE analyses

All PAGE analyses were performed using the Mini-PROTEAN II electrophoresis kit (Bio-Rad Laboratories Inc., Hercules, CA, USA). SDS-PAGE analysis followed the method of Laemmli (1970) using a 12% separating and 4% stacking gel. Gels and buffers were prepared according to the Mini-PROTEAN II electrophoresis cell instruction manual (Bio-Rad, Manual). If necessary, protein samples were clarified before analysis by centrifugation at 13000 g for 2 min with the Biofuge Fresco, Heraeus (Kendro Laboratory Products Plc, Hertfordshire, UK). Samples were diluted 1:1 with SDS reducing sample buffer (Fluka Chemie AG, Buchs, Switzerland) and heated in a near boiling water bath for 5 min before loading 20 µl sample per well. The protein separation was achieved with 150 V for 60 min.

Native-PAGE analysis was performed using a 8% separating and 4% stacking gel.

Buffers and gels were prepared as for SDS-PAGE analysis, but omitting the SDS denaturant. Pure transketolase samples were diluted 1:1 in native loading buffer, containing 0.2 M Tris-HCl, pH 6.8, 40% (v/v) glycerol and 0.004% (w/v) thymol blue (Sigma Diagnostics, St. Louis, MO, USA), before loading 20 µl sample per well. Electrophoresis was done with 150 V for 4-5 hr at 4°C.

Protein bands were visualised using either Coomassie Blue or silver staining. Gels were soaked for 30 min in an aqueous fixative containing 0.1% (w/v) Coomassie Blue R-250, 40% (v/v) methanol and 10% (v/v) acetic acid. After Coomassie staining, gels were destained with 2-3 changes of an aqueous solution containing 40% (v/v) methanol and 10% (v/v) acetic acid. Silver staining was performed according to the procedure of Blum *et al.* (1987).

2.21 Transketolase unfolding studies

Protein unfolding studies have been conducted on the wild-type transketolase as well as the D381A and Y440A mutants by measuring tryptophan fluorescence as a function of urea concentration (Section 4.2.7). Denaturant stock solutions with urea concentrations of 0 M to 8 M urea and 25 mM Tris buffer, pH 7.5 were prepared and stored at -20 °C for unfolding studies. 200 µl of 1 mg·ml^{-1} pure transketolase in 25 mM Tris buffer, pH 7.5 with 5 mM $MgCl_2$ and 0.5 mM TPP was added to 1.8 ml of each urea stock solution. The final 2 ml solution had a protein concentration of 0.1 mg·ml^{-1}. Samples were incubated for 30 min at ambient temperature allowing the unfolding equilibrium to be reached. The concentrations of the different transketolase stock solutions were determined using the monomer molecular weights (MW) and molar extinction coefficients (ϵ) as calculated by the method of Pace *et al.* (1995). The following molecular weight and molar extinction coefficient values were used: Wild-type MW = 72260.82 g·mol^{-1} and ϵ = 93905 M^{-1}·cm^{-1}; D381A MW = 72216.81 g·mol^{-1} and ϵ = 93905 M^{-1}·cm^{-1}; Y440A MW = 72168.72 g·mol^{-1} and ϵ = 92415 M^{-1}·cm^{-1}.

Tryptophan fluorescence was measured using a Perkin Elmer LS30 luminescence spec-

trophotometer (Perkin Elmer UK, Buckinghamshire, UK). Excitation wavelength was fixed at 280 nm and the emission spectra was recorded from 290 nm to 480 nm using LabView 7.1 Virtual Instrumentation routines (National Instruments, Austin, TX, USA). The protein solution was pumped into the fluorescence flow cell for 15 sec to ensure a stable fluorescence signal was obtained before data capture was initiated. The λ_{max} and fluorescence intensity at 340 nm were used for the purpose of data analysis.

2.22 Transketolase refolding study

Denaturant stock solutions with urea concentrations of 0 M to 3.8 M urea and 25 mM Tris buffer, 3.8M urea, pH 7.5 were used for the refolding study. 200 µl of 1 mg·ml^{-1} pure transketolase in 25 mM Tris buffer, 3.8 M urea, pH 7.5 with 5 mM $MgCl_2$ and 0.5 mM TPP was added to 1.8 ml of each urea stock solution. Samples were incubated for 60 min at ambient temperature allowing the refolding equilibrium to be reached before emission wavelength spectra were recorded. Data collection and the determination of the concentrations of the different transketolase stock solutions were performed as described in Section 2.21.

2.23 Spectral data analysis and stability calculations

In order to obtain accurate λ_{max} values from the recorded emission spectra, wavelength (λ) data in nm, were converted to wavenumbers (v), in cm^{-1}.

$$v = (\lambda \cdot 10^{-7})^{-1} \tag{2.12}$$

Raw fluorescence data from emission spectra were plotted against corresponding wavenumbers (Section 4.2.6). Data sets were fitted to a two-parameter lognormal distribution function representing the fluorescence emission spectra of tryptophan.

$$I(v) = I_m \cdot exp\left[-\frac{ln2}{ln^2 p} \cdot ln^2\left(\frac{a-v}{a-v_m}\right)\right] \quad \text{at} \quad v < a \tag{2.13}$$

A lognormal distribution is depicted in Figure 2.5 where I_m is the maximal intensity; v the wavenumber; p a band symmetry parameter with function $p = (v_m - v_-)/(v_+ - v_m)$ and a a function limiting point position $a = v_m + (p \cdot (v_+ - v_-))/(p^2 - 1)$. The parameter v_m is the wavenumber correlating to the position with maximal fluorescence intensity (I_m) while v_- and v_+ are the wavenumber positions for the half-maximal fluorescence intensity lower and higher than v_m respectively. The functions for a and p are substituted into Equation 2.13 in order to obtain unique solutions of v_m, v_- and v_+ for each of the recorded emission spectra (Burstein *et al.*, 2001). Estimated v_m values were converted back to λ_{max} using Equation 2.14 before plotting against urea concentration in order to obtain unfolding curves (Section 4.2.7).

$$\lambda_{max} = (v_m \cdot 10^{-7})^{-1} \tag{2.14}$$

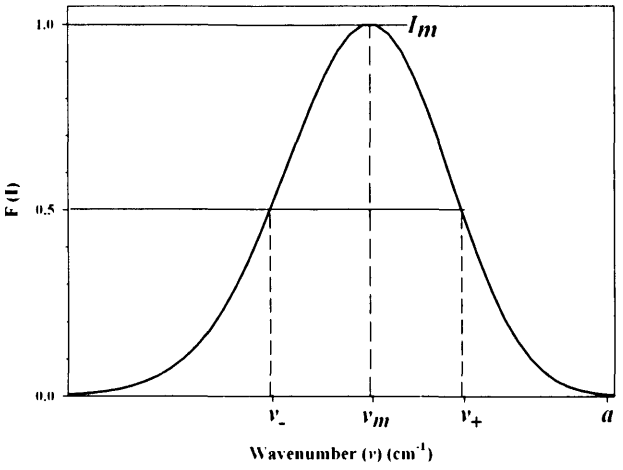


Figure 2.5: A typical lognormal distribution of the fluorescence emission (F) against wavenumber (v) of a protein. v_m is the wavenumber correlating to the position with maximal fluorescence intensity (I_m) while v_- and v_+ are the wavenumber positions for the half-maximal fluorescence intensity lower and higher than v_m respectively. A location parameter, a, is used to move the function away from the origin and also serves as function limiting parameter since only logarithms of positive numbers exist.

2.24 Estimation of dissociation midpoint

All data fitting and analysis was performed using the software package SigmaPlot 8.0 (SPSS UK Ltd., Surrey, UK). The dissociation midpoints were estimated to provide an indication of dimer stability for wild-type and mutant transketolases. The dissociation reaction can be written as

$$N_2 \rightarrow 2I \tag{2.15}$$

in which case the equilibrium dissociation constant (K_D) is

$$K_D = [I]^2/[N_2] \tag{2.16}$$

Since K_D is protein concentration dependent it can be rewritten as a function of the total protein concentration (P_t) and fraction dissociation (f_D). The relation between P_t, f_D and concentrations of N_2 and I can be written as reported by Bowie and Sauer (1989).

$$P_t = 2N_2 + I = (1 - f_D)P_t + f_D \cdot P_t \tag{2.17}$$

This relation can be used to express K_D as a function of P_t and f_D and gives

$$K_D = 2P_t(f_D)^2/(1 - f_D) \tag{2.18}$$

K_D as expressed in Equation 2.18 can then be substituted into $\Delta G_D = -RT \cdot ln \cdot K_D$. When the equation is rearranged as a function of f_D Equation 2.19 follows is derived (Bowie and Sauer, 1989).

$$f_D = \frac{\left(\left(e^{\left(\frac{m_G[D-C_{1/2}]}{RT}\right)}\right)^2 + 8P_t \cdot e^{\left(\frac{m_G[D-C_{1/2}]}{RT}\right)}\right)^{0.5} - e^{\left(\frac{m_G[D-C_{1/2}]}{RT}\right)}}{4P_t} \tag{2.19}$$

The unfolding data is normalised to represent the fraction dissociated dimer using Equations 2.1 and 2.2 in Section 2.6. The normalised curves were then fitted to Equation 2.19 to obtain $C_{1/2}$ values for the dissociation transition.

2.25 Fluorescence enzyme-linked assay

A more sensitive assay was required to determine the activity rate of the transketolase immobilised in microwells. Fluorescence monitoring of the NADH concentration is very sensitive and can be used to quantify low consumption rates, as would be expected with small amounts of transketolase. The linear range for fluorescence of NADH was determined for optimal accuracy in rate determination. A 100 $\mu g \cdot ml^{-1}$ NADH solution was prepared by diluting as stock solution of NADH to an A_{340nm} reading of 0.746 AU. The 100 $\mu g \cdot ml^{-1}$ NADH solution was used to prepare a dilutions range required to generate a calibration curve.

The concentrations and volumes per 1 ml assay mix for all the components are listed in Table 2.6. The PRI, DR5P3E and DR5P (substrate mix) were preincubated for 2 hr at room temperature, ensuring that transketolase substrates, D-ribose-5-phosphate and D-xylulose-5-phosphate, are at equilibrium concentrations. The remainder of the components (GDTPI, NADH and buffer or buffer/transketolase) were combined and 172 µl loaded into wells with immobilised transketolase. Reactions started by injecting 28 µl assay mix to provide a final well volume of 200 µl. Reaction progress was monitored with fluorescence emission at 450 nm after excitation at 340 nm. In addition a response curve was generated to determine the linear activity range for transketolase (Section 5.2.2). A concentration range of transketolase was created by mixing a standard concentration transketolase (x µl) with buffer containing cofactors (last column of Table 2.6).

Table 2.6: Components for fluorescence enzyme-linked assay.

Component	Concentration of Stock	Assay Volume (µl)	Concentration (Final)	Response Curve Volume (µl)
PRI	10 $U \cdot ml^{-1}$	20	0.2 $U \cdot ml^{-1}$	20
DR5P3E	10 $U \cdot ml^{-1}$	20	0.2 $U \cdot ml^{-1}$	20
DR5P	30 mM	100	3 mM	100
GDTPI	20 $U \cdot ml^{-1}$	100	2 $U \cdot ml^{-1}$	100
NADH	100 $\mu g \cdot ml^{-1}$	420	42 $\mu g \cdot ml^{-1}$	420
Buffer		340	-	340-x
Transketolase	13.72 $\mu g \cdot ml^{-1}$			x (1-100)

2.26 Immobilisation of transketolase in Ni-NTA microwells

His_6-transketolase was immobilised in Reacti-Bind Ni-NTA coated microplates (Perbio Science UK Ltd., Northumberland, UK), quantified and eluted for SDS-PAGE analysis. Wells were washed with 200 µl wash buffer (50 mM Tris-base, 300 mM NaCl, 20 mM imidazole. pH 8.0). 100 µl of approximately 0.1 $mg \cdot ml^{-1}$ purified transketolase or cell-free extract of 5 AU at A_{600nm} from *E. coli* JM107 containing transketolase was loaded into wells and shake-incubated for 1 hr at 600 rpm. Solutions were discarded and the wells washed twice with wash buffer. Bound protein was eluted from selected wells with elution buffer (50 mM Tris-base, 300 mM NaCl, 250 mM imidazole, pH 8.0). Eluted samples were analysed with 12% SDS-PAGE and visualised using silver staining (Section 2.20). The remainder of the wells were used for determination of immobilised protein content and enzyme activity. Protein content was determined with the Pierce Micro-BCA kit (Perbio Science UK Ltd.) (Sections 5.2.2). Transketolase activity measured with the fluorescence enzyme linked assay was used to calculate the variance in the immobilisation step.

2.27 Microwell cell-culturing

E. coli JM107 pQR791 and XL10-Gold containing mutant plasmids were cultivated in 96-well deep square well polypropylene plates with a well volume of 2 ml. Microplate wells were filled with 1 ml LB-media, covered with foil and autoclave sterilised. 5 µl of 50 $g \cdot l^{-1}$

ampicillin stock solution was added to each well before use. Each well was inoculated with 10-20 µl from over-night *E. coli* culture of the appropriate strain. Microplate was covered with an air-permeable plate sealer (Greiner Bio-One Ltd, Gloucestershire, UK), placed in a humidified incubator at 37 °C and agitated at 1400 rpm on a Variomag® (Variomag Daytona Beach, FL, USA) magnetic microplate shaker. Cultures were allowed to grow for 12 hr. 20 µl cell culture was transferred to a standard F-type clear microplate and diluted with 180 µl sterile LB-media. Variation in obtained cell densities were determined by measuring the absorbance at A_{600nm} on a platereader.

2.28 Estimation of oxygen flux and utilisation

To ascertain whether the target cell density can be achieved, the maximal theoretical ($Q_{O_2(max)}$) and required ($Q_{O_2(req)}$) oxygen fluxes were calculated for cultivation in microplates.

$$Q_{O_2(req)} = \frac{x\mu}{Y_{O_2}} \tag{2.20}$$

The oxygen flux required ($Q_{O_2(req)}$) to sustain the rate of metabolic oxygen utilisation at (x) as a given cell density, expressed as concentration dry cell weight ($g \cdot L^{-1}$) has been estimated with (μ) as the specific growth rate and (Y_{O_2}) as the Hill-coefficient (Section 5.2.4).

$$Q_{O_2(max)} = k_l a'(c_l^* - c_l) \tag{2.21}$$

The maximal theoretical oxygen flux ($Q_{O_2(max)}$) has been estimated where ($k_l a'$) is the oxygen mass-transfer coefficient; (c_l^*) the liquid-phase oxygen concentration in equilibrium with bulk gas phase and (c_l) the liquid-phase oxygen concentration (Section 5.2.4).

2.29 High-throughput protein purification

His_6-transketolase was captured using the Novagen RoboPop™ Ni-NTA His•Bind purification kit (Merck Biosciences Ltd., Nottingham, UK). The chemical lysis solution was pre-mixed and consisted of PopCulture™, rLysozyme™ (400 U·ml^{-1}) and Benzonase® Nuclease (250 U·ml^{-1}). 0.1 ml lysis solution was added to each well and gently shake-incubated for 15 min at room temperature and 200 rpm on the Variomag® magnetic microplate shaker.

Ni-NTA His•Bind Resin was suspended and 20 ml of the slurry transferred to a 50 ml Falcon tube. Resin was washed with 3 volumes of deionised H_2O for initial equilibration, 1 volume 0.2 M $NiSO_4$ for charging resin with metal ions and finally with 4 volumes deionised H_2O to remove excess Ni^{2+} ions. Between each wash step the resin was sedimented by centrifugation at 3000 rpm in a Centaur Heraeus centrifuge (Kendro Laboratory Products Plc, Hertfordshire, UK) and supernatant carefully decanted. Resin was resuspended in 1 volume of 25 mM Tris-base buffer, pH 8.5 ready to be use. After use the resin was salvaged, washed with 1 volume 0.05 M EDTA and 3 volumes deionised H_2O. For long term storage the resin was kept in an aqueous solution of 20% ethanol (v/v).

Either 100, 200 or 300 µl prepared Ni-NTA His•Bind Resin was added to the cell free extract in the deep square well plate. The resin was shake-incubated at room temperature and 1400 rpm for 2 hr.

The cell free extract and resin suspension was transferred to a 96-well filter plate. Cell free extract was removed by placing the filter plate on a vacuum manifold and applying a vacuum generated by a Vacuum/Pressure Pump (Millipore Corp., Bedford, USA). Resin in each well was washed three times with 1 ml of 25 mM Tris-base buffer, pH 8.5. The protein product was eluted into a regular 96-well plate using 200 µl 25 mM Tris-base buffer, 250 mM imidazole, pH 8.5 and the concentration determined spectrophotometrically at a wavelength of 280 nm. In addition protein elution profiles were generated for 100, 200 and 300 µl resin by sequential eluting in 100 µl fractions (Section 5.2.3). All the elution fractions were retained for protein quantification.

2.30 High-throughput dialysis

2.30.1 Dialysis apparatus design

A 12 mm aluminium baseplate was taken and a rectangular groove cut to the size of the 96-well microplate footprint. This allowed the 96-well plate to fit tightly and also support the well bottoms directly on the baseplate. 10 threaded holes were drilled in the baseplate allowing bolts to be screwed into it. A 12 mm Perspex^{TM} lid was cut to the same size as the baseplate and 96-holes made to match the wells of the microplate. Additional holes, three on the edge of either side of the lid, were used to secure the lid to the baseplate with the microplate sandwiched in between. A silicone gasket with matching 96 holes fitted between the 96-well plate and lid. The dialysis membrane is placed between the gasket and the microplate. Clear, polystyrene, shallow, flat bottomed, chimney well plates, (also known as F-shape plates) were obtained from Greiner Bio-One Ltd (Gloucestershire, UK). The bottoms of wells C4, C9, F4 and F9 were drilled out to accommodate the additional bolts that are required to achieve a tight seal of the gasket. These bolts are only temporary measure until the Perspex^{TM} of the lid is replaced with more rigid material, allowing for less curvature across the breadth of the plate (Section 5.2.5.1).

2.30.2 Dialysis procedure

Sodium chloride, imidazole, blue dextran and cytochrome *c* were obtained form Sigma Chemical Co. (St. Louis, MO, USA). All solutions were prepared using deionised H_2O from a Millipore Elix 5 (Millipore Corp., Bedford, USA) system.

Cellulose dialysis tubing (76 mm flat width) also obtained from Sigma Chemical Co. and had a molecular cut-off of 10 kDa. The tubing was cut into single membranes (75 mm x 120 mm) that cover all 96 wells of a microwell plate. Membranes were soaked in deionised H_2O, washed twice and allowed to dry off slightly before use. Two perpendicular slits the width of a well was cut into the membrane at positions C4, C9, F4 and F9 to allow bolts to pass through membrane without stretching or tearing.

A cylindrical glass container (200 mm diameter, 300 mm height) was filled with 4 liter appropriate dialysis buffer for all the experiments. The bulk dialysis buffer was agitated with a magnetic stirrer HI300N from Hanna Instruments Ltd (Bedfordshire, UK) at maximum speed using a 50 mm triangular stirrer bar. The agitation removes solute in close proximity of the membrane, thus maximising the concentration gradient across the membrane.

2.30.3 Determination of dialysis rate

Four separate experiments were conducted to determine the rate of dialysis of sodium chloride and imidazole in both directions across the membrane (Section 5.2.5.2). In the first experiment 200 µl of a 1 M sodium chloride solution was loaded into all the wells of a microwell plate and fitted into the dialysis apparatus. The wells were dialysed against deionised H_2O. Samples were taken in triplicate for analysis at appropriate intervals. Samples were diluted 30-fold with deionised H_2O and conductivity measurements made. The same experiment was conducted in reverse where 200 µl deionised H_2O was loaded per well and dialysed against 1 M NaCl.

Imidazole was prepared as a 1 M solution and 200 µl was loaded into each well of a microwell plate. The microwells were dialysed against deionised H_2O. Samples were taken in triplicate at appropriate intervals, diluted 5-fold with deionised H_2O and the imidazole concentration determined spectrophotometrically. The reverse experiment was conducted where deionised H_2O was loaded into the microwells and dialysed against 0.25 M imidazole in the bulk buffer solution.

Conductivity measurements of samples were made with a Pharmacia Conductivity Monitor 1. The conductivity meter was calibrated using the Pharmacia 12300 µS conductivity standard. All conductivity equipment and reagents were obtained from Amersham Biosciences UK Ltd (Buckinghamshire, UK). Conductivity measurements were made by injecting appropriately diluted samples into the flow cell with a 1 ml syringe. The flow cell was purged with deionised H_2O between sample readings. The imidazole concentration in

samples was determined with UV absorbance at a wavelength of 280 nm. All spectrophotometric measurements were made using an ATI UniCam UV2 spectrophotometer from Unicam Ltd (Cambridge, UK).

2.30.4 Dialysis rate calculations

The dialysis process is a uni-molecular reaction that can be written as Equation 2.22 where S_{well} is the solute in the well and S_{bulk} the solute in the bulk dialysis buffer.

$$S_{well} \rightarrow S_{bulk} \tag{2.22}$$

Since the volume of the bulk fluid is much greater than that of the wells, the solute (S) becomes infinitely dilute and the reaction irreversible. The data of the time course graphs follow exponential decay and can be fitted against a first order integrated rate law (Equations 2.23 and 2.24) where C_0 is the initial solute concentration, k the overall dialysis rate constant for the reaction and t is the time.

$$C = C_0 \times e^{(-kt)} \tag{2.23}$$

$$C = C_0 - C_0 \times e^{(-kt)} \tag{2.24}$$

2.30.5 Testing for cross-contamination and membrane integrity

Blue dextran and cytochrome c were prepared as 0.2 mg·ml^{-1} and 0.8 mg·ml^{-1} solutions respectively and 200 µl of either solution was loaded into the wells at different positions (Section 5.2.5.3). The microplate was dialysed overnight in deionised H_2O. Samples were taken from wells after dialysis and their absorbance spectra compared to the characteristic spectra of the original blue dextran and cytochrome c solutions to determine if cross-contamination occurred. The concentration of blue dextran was determined spectrophotometrically at a wavelength of 620 nm. Absorbance spectra for cytochrome c and blue dextran were recorded between 300 nm and 700 nm.

2.31 Transketolase concentration range for unfolding studies

Fixed volume denaturation was performed with different amounts of wild-type transketolase. 25, 50 and 100 µg transketolase in a volume of 120 µl 25 mM Tris buffer, pH 7.5 with 5 mM $MgCl_2$ and 0.5 mM TPP were loaded into 51 wells (3×17 data points) of a 96 well microplate. Wells were filled with different amounts of 25 mM Tris buffer, pH 7.5 containing 0 M or 8.5 M GdnHCl to a final volume of 300 µl, generating a final denaturant concentration range of 0 to 5.1 M. Each GdnHCl concentration was prepared in triplicate. Equivalent baseline readings were obtained and subtracted to obtain the fluorescence signal difference. The well contents were mixed by orbital shaking at 350 rpm for 30 seconds and allowed to equilibrate for 15 minutes at room temperature prior to taking measurements. Fluorescence emission was recorded at 340 ±10 nm after excitation at 280 nm (Section 5.2.1).

2.32 High-throughput unfolding of transketolase

Wild-type transketolase and mutant transketolases were prepared as 1 $mg \cdot ml^{-1}$ solutions in 25 mM Tris buffer, pH 7.5 with 5 mM $MgCl_2$ and 0.5 mM TPP. 50 µl of transketolase solution was loaded into wells. A denaturant solution of 25 mM Tris buffer, pH 7.5 with 8 M urea was added using the serial addition method. Each addition was followed by 30 seconds of mixing by orbital shaking at 350 rpm, and 15 minutes of equilibration prior to taking measurements. Volumes for serial addition were adjusted to provide nearly equidistant data points on the unfolding curve. Fluorescence data obtained at 340 nm was normalised over the $N_2 \rightarrow 2N$ transition to visualise the decrease in dimer stability of mutant transketolases (Section 5.2.6).

Chapter 3

High-throughput unfolding screen

3.1 Introduction

The aim of this Chapter was to develop of a novel, automatable and inexpensive high-throughput screen for the direct measurement of protein stability [1]. In principle protein unfolding can be conducted in microplates or microwells and unfolding curves recorded using fluorescence measurements. The ΔF measurements may then be used to estimate midpoint of unfolding and protein stability. Several key questions had to be addressed to ensure successful development of this high-throughput method. 1) What is a reasonable high-throughput value for a direct protein stability screen? 2) What is the best way to validate the accuracy of the screen? 3) What are the critical parameters that need evaluation for successful method development?

3.1.1 High-throughput range for stability assays

There are a number of high-throughput screening methods for evaluating protein stability, several of which are described in Chapter 1 (Section 1.5.1). Indirect screens interrogate thermostability of proteins at elevated temperatures resulting in a loss of activity (Giver *et al.*, 1998) or increase in aggregation/turbidity (Won *et al.*, 1998). Indirect screens

[1] Most of the results presented in this chapter are included in Jean P. Aucamp, Ana M. Cosme, Gary J. Lye, Paul A. Dalby, (2005) High-throughput Measurement of Protein Stability in Microtiter Plates. *Biotechnology and Bioengineering* **89** (5), 599-607.

like these have the advantage of simplicity. Light-scattering and residual activity are parameters that can be determined rapidly resulting in fast plate turn-over and large throughput values. Essentially, data for only two time events (i.e. before and after denaturation/inactivation) are collected, making data processing and analysis a relatively simple and effective exercise. One disadvantage of these screens is the potential for ambiguous data interpretation. For example, increased residual activity can be the result of either increased thermostability or increased ability for reversible unfolding.

Advances in instrument technology currently allow the direct high-throughput measurement of protein stability using matrix-assisted laser desorption/ionisation mass spectrometry (MALDI-MS) (Ghaemmaghami *et al.*, 2000) or differential scanning calorimetry (DSC) (Weber and Salemme, 2003). Direct screening methods evaluate the unfolding process itself leading to a reduction in throughput. The existing MALDI-MS method requires 8-12 sample preparations per protein and it is estimated that the stability parameters of up to a 1000 proteins can be determined per day. The MicroCal DSC platform is capable of screening 50 protein preparations per day. Based on these figures a respectable initial throughput target will be in the range of 500-1000 sample per day.

3.1.2 Validation of method using model proteins

Our aim was to develop and test the direct measurement of protein stability by monitoring tryptophan fluorescence in microplates, using proteins with well-defined unfolding properties and previously reported stability values. Use of previously characterised protein models has permitted us to assess the relevance of absolute stabilities obtained by this method, compared to the measurement of changes in stability upon, for example, mutation or addition of excipients. Several model proteins were initially screened and described in Appendix A. From these proteins cytochrome *c* and bovine serum albumin were selected for ease of preparation, availability of excitation and emission filters on instrumentation and their relevance to the study.

3.1.2.1 Cytochrome *c*

The first protein chosen, cytochrome *c* (cyt *c*) has applications in industrial biocatalysis (Vazquez-Duhalt, 1999) catalysing mainly oxidation and hydroxylation reactions. In the presence of hydrogen or organic peroxides, cyt *c* catalyses N- and O-demethylation, epoxidation and S-oxidation reactions that are cytochrome P450-like reactions (Akasaka *et al.*, 1993). Unlike most heme proteins, cyt *c* has a protoheme prosthetic group that is covalently attached by two thioether bridges (from cysteine residues of the protein) and the vinyl side chains of the heme. Since the heme is covalently linked to the protein it maintains activity in organic solvents and elevated temperatures where cytochrome P450, another very useful biocatalyst, is inactivated. Equine heart cyt *c* is active in 90% tetrahydrofuran and displays optimal activity at 80 °C.

The three-dimensional structure of oxidised cyt *c* is depicted in Figure 3.1. The prosthetic heme moiety can be in the oxidised or reduced form. The protein displays different unfolding characteristics depending on the state of the heme. It is known to unfold reversibly with the heme in the oxidised form (Xu *et al.*, 1998) along a apparent two-state mechanism. In the reduced form the protein unfolds via 2 intermediates (Myer *et al.*, 1980). The ability to screen variant cyt *c* libraries by fluorescence during unfolding is anticipated to allow the identification of mutations that extend the lifetime for this enzyme under biocatalytic process conditions. The need to identify differences in protein stability upon mutation is paramount in such directed evolution experiments, whereas the absolute determination of stability is not necessary.

3.1.2.2 Bovine serum albumin

Ligands, cofactors, metal ions, substrates and other excipients often confer extra stability upon binding to proteins. The thermodynamic stabilisation of bovine serum albumin (BSA) with fatty acid molecules is one such example (Ahmad and Qasim, 1995). Initially the topology of BSA was based on ligand binding properties, primary sequence and available proteolitic sites (Brown and Shockley, 1982), suggesting three folded domains

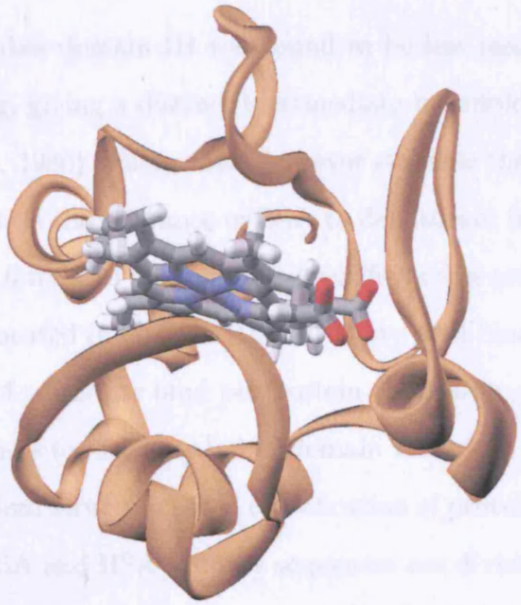


Figure 3.1: Schematic representation of the structure of oxidised equine cytochrome *c*. Heme moiety catalysis useful oxidation and hydroxylation reactions.

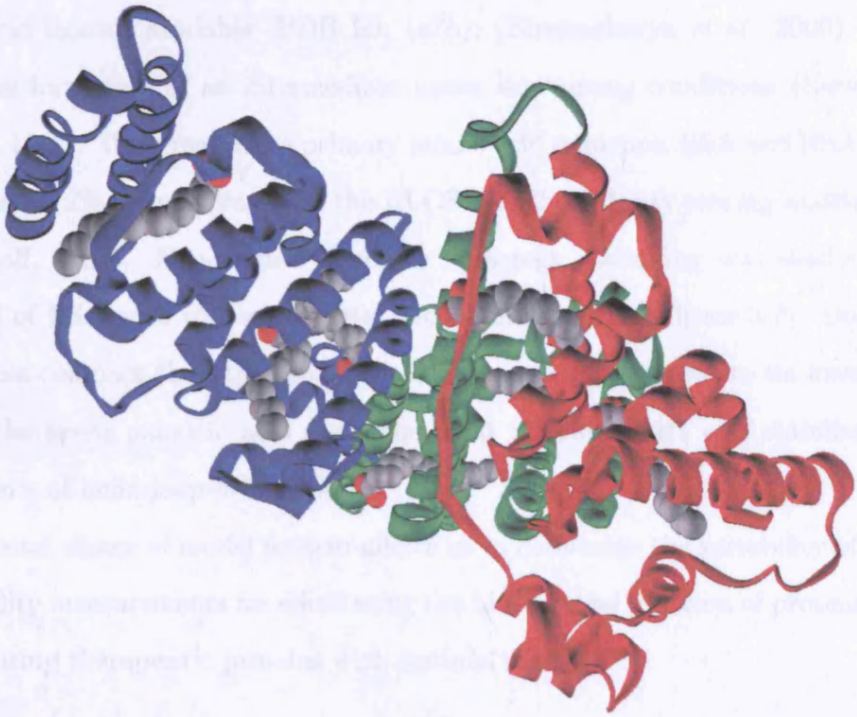


Figure 3.2: Schematic presentation of the structure of the BSA:palmitate complex obtained with homology modelling using an X-ray structure of human serum albumin saturated with seven palmitate molecules. The loosely structured Domain III (depicted in blue), known to be stabilised by fatty acids, contains three palmitate (depicted in grey) occupied fatty acid sites. The more compact Domains I (red) and II (green) contain two fatty acid binding sites each also occupied by palmitate molecules (grey).

(I-III). From several studies domain III was found to be less resistant to acid and denaturant induced unfolding, giving a distinct intermediate in unfolding analysis (Khan and Salahuddin, 1984; Khan, 1986). Fatty acids however stabilise the C-terminal domain III and have a marked effect on the tolerance of BSA to denaturant induced unfolding. Early attempts at locating the fatty acid binding sites identified three primary fatty acid binding sites, two of which are located in domain III. BSA-fatty acid binding studies determined that up to 7 molecules of palmitate bind per protein (Hamilton *et al.*, 1991; Reed, 1986) and the strongest binding sites are situated in domain III.

According to the modern structure based classification of protein domains, using CATH (Orengo *et al.*, 1997), BSA and HSA primary sequences are divided into six domains and contain three conserved fatty acid binding motifs based on PROSITE patterns in the Protein Data Bank (http://www.pdb.org/) (Berman *et al.*, 2000). No crystal structures are available of BSA. There is however a crystal structures for HSA with seven bound palmitic acid ligands available (PDB ID: 1e7h); (Bhattacharya *et al.*, 2000). HSA also displays the formation of an intermediate under denaturing conditions (Herskovits and Laskowski, 1962). With regards to primary amino acid sequence, BSA and HSA share 84% similarity and 72% identity based on the BLOSUM 62 similarity scoring matrix (Henikoff and Henikoff, 1992). Non-refined structure homology modelling was used to visualise domain III of BSA with respect to fatty acid binding motifs (Figure 3.2). Domain III is spatially less compact than the other two domains giving evidence to its lower stability. Three of the seven palmitic acid molecules bind to domain III and stabilise the loose arrangements of helix-loop-helix motifs.

This second choice of model protein allows us to determine the suitability of microtiter plate stability measurements for elucidating the biochemical function of proteins, and also for formulating therapeutic proteins with optimal shelf-life.

3.1.3 Critical parameters in fluorescence based screens

Fluorescence based assays are preferred for high-throughput screens, partly for the superior sensitivity allowing detection in low volumes or concentrations. In some modes of fluorescence, multiple output parameters are obtained from signal readouts allowing for correction/correlation of signals and serves as quality assurance for collected data (Pope *et al.*, 1999). Fluorescence based detection can be divided into macroscopic and microscopic methods. Macroscopic methods produces a readout from the averaged over a large part of the well volume used typically with microplates of the 96 to 1534 well format. Microscopic methods require an epi-illuminated confocal set-up and laser beam. The volume of detection is approximately 10^{-15} l. Macroscopic methods are not robust towards miniaturisation. The signal to baseline noise ratio decreases and the %CV increases as the assay volume decreases (Haupts *et al.*, 2000).

Macroscopic fluorescence intensity (FLINT) assays are susceptible to inner-filter and autofluorescence phenomena that are hard to predict, detect and correct for. The inner-filter effect is caused by absorption of either emission or excitation light by non-fluorescing compounds in solution with target analyte. At high concentrations of target analyte, the excitation light does not travel far enough through the sample producing non-representative readouts, also referred to as an inner-filter effect (Lakowicz, 1999). Several cellular metabolites including proteins display dim autofluorescent properties. At high concentrations (typically more than 10 μM) the autofluorescence contributes considerably to the overall signal, obscuring detection of the 'bright' signal from the target analyte (Billinton and Knight, 2001). Macroscopic FLINT methods do not discriminate according to 'brightness' and reliability is compromised at high protein concentrations.

The injection volume is also an important factor in the method development. Using titrating syringe pump technology allows for dispensing in the μl range. Injection apparatus based on piezo-electric ink-jet technology is able to pipette accurately to the sub-μl range (Brandt, 1998). The method reported here uses 96-well plates and FLINT mode to measure tryptophan emission as well as a titrating syringe pump for dispensing liquid.

3.2 Results and Discussion

3.2.1 Selection of microwell plate type

Solutions of Tris buffer or Tris GdnHCl were titrated incrementally into black well and clear well plates (Section 2.3). The fluorescence baseline was monitored with change in volume. Black well plates displayed a positive baseline drift that increases significantly above 140 µl. Clear well plates displayed a negative drift that is almost negligible over the range of the titration experiment (Figure 3.3). A GdnHCl concentration range was generated at four different well volumes. At constant volume, a concentration dependent increase in fluorescence baseline was observed above 3 M GdnHCl (Figure 3.4). The concentration and volume dependent drifts followed a similar pattern.

Conformational transition curves were generated for cyt *c* in black, clear-bottom black and clear well plates using the fixed volume method and results compared to literature values. Results obtained using clear well plates were found to be comparable to values reported in the literature (discussed in detail in Section 3.2.3; Table 3.1). Black well plates consistently under-estimates the m_G value for cyt *c* (Table 3.2). Clear-bottom black well plates showed mixed results where the m_G value is under-estimated the for equine cyt *c* and over-estimated for bovine cyt *c*. The $C_{1/2}$ values however are comparable to literature values.

Black plates produced the superior signal quality. It displayed lower background noise and the fluorescence signal change between native and unfolded protein was 3-5 times greater than for clear well plates. The performance of clear-bottom black well plates rated between black and clear well plates. Based on these results black well plates can only be used for serial addition type experiments if the complete unfolding can be achieved in less than 140 µl. Similarly, fixed volume experiments could be conducted if complete unfolding occurred below 3 M GdnHCl concentration and is still suitable for cyt *c* with a $C_{1/2}$ of ±2.5 M GdnHCl.

Since black well plates, typically used in fluorescence based assays, poses restrictions

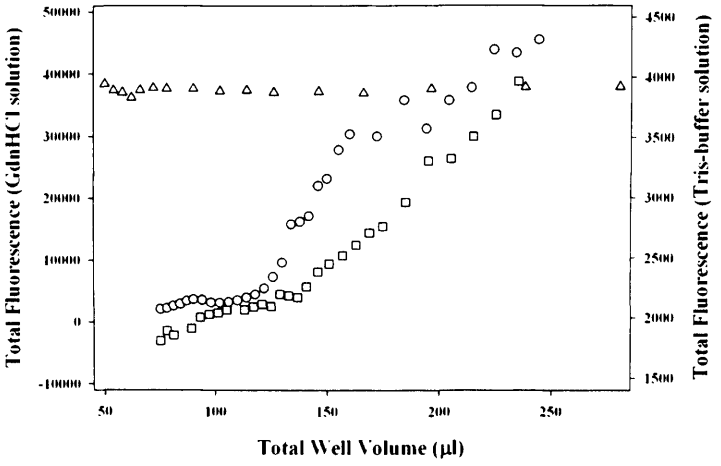


Figure 3.3: The well volumes were systematically increased and fluorescence signal recorded at 340 nm to determine the maximum working volume. Signal drifts are represented as follow: Black plates using (○) 0.1 M Tris-buffer, 6.5 M GdnHCl pH 7.0 or (□) 0.1 M Tris-buffer pH 7.0. Clear well plates using (△) 0.1 M Tris-buffer pH 7.0.

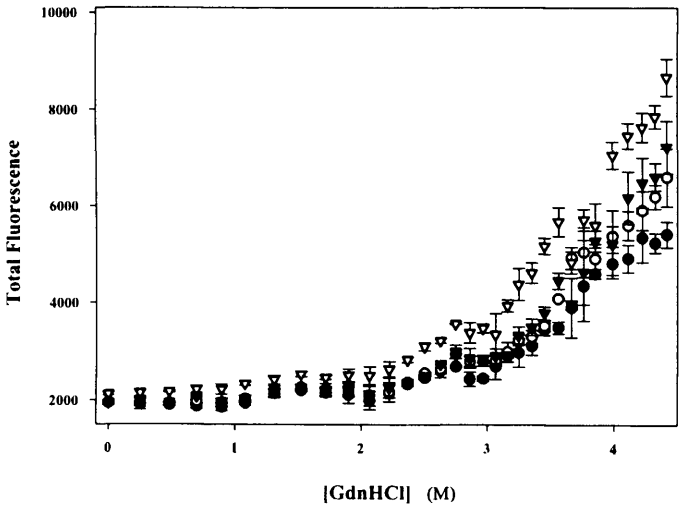


Figure 3.4: The GdnHCl concentration was increased while keeping the total well volume constant at (•) 75 µl, (○) 100 µl, (▼) 125 µl or (▽) 150 µl respectively.

on experimental design clear well plates were used for all the stability analyses instead.

3.2.2 Accuracy of microwell scale unfolding measurements

The determination of protein stability by equilibrium unfolding techniques requires that a large amount of measurements at different denaturant concentrations be made. The largest source of variation and experimental error originates from differences in the final protein concentration for each denaturant concentration solution. This can be eliminated by performing an unfolding experiment starting with a half-full cuvette, adding denaturant and correcting for the dilution effect on the signal if necessary (Schwehm and Stites, 1998). In doing so, accurate titration becomes the only major concern. This method was applied here in microwells and since the total fluorescence of each well was measured, dilution of the protein solution should have little effect on the signal (Section 2.2). Care has to be taken to ensure that the working protein concentration produces fluorescence in the linear response range of the instrument. High protein concentrations show a large fluorescence drift with dilution, most likely due to the inner filter effect (Lakowicz, 1999). Fluorescence signal drifts of only 2.15 % and 2.88 % were observed respectively over a four-fold dilution range for a 0.6 $mg{\cdot}ml^{-1}$ BSA (Figure 3.5) and a 0.25 $mg{\cdot}ml^{-1}$ bovine cyt c initial concentrations. Dilution effects on total fluorescence were therefore not corrected for in our calculations. Denaturant stock solutions were prepared at a concentration that permitted complete protein unfolding to be achieved without exceeding a four-fold dilution of the initial protein concentration. This requires some prior knowledge about the unfolding midpoints of proteins to be analysed, or a rough pre-screen of unfolding in microplates using various stock solutions.

The accuracy of both large and small volume titration using the syringe pump as determined by the manufacturer was verified, as this was expected to have the greatest effect on the accuracy of the protein unfolding data (Section 2.4). Water was auto-titrated into wells and the cumulative mass of water measured and plotted against the total volume dispensed (Figure 3.6). The average slope calculated was very similar to the density of

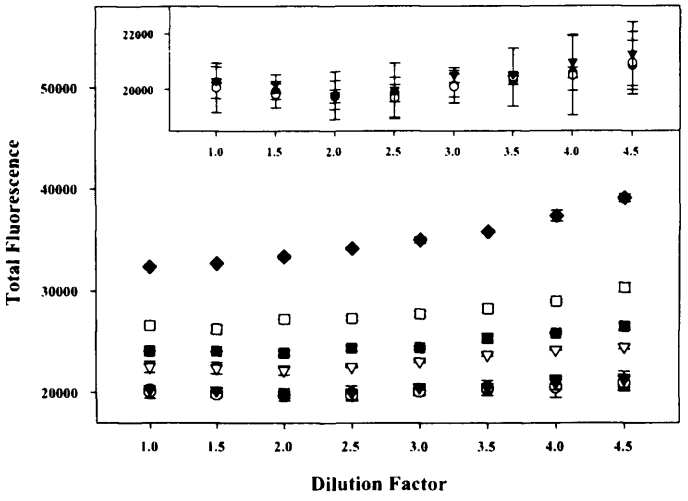


Figure 3.5: The effect of dilution on bovine serum albumin (BSA) fluorescence. Volumes of 50 µl of different concentrations of BSA solutions (•) 10 µg·ml^{-1}, (○) 30 µg·ml^{-1}, (▼) 60 µg·ml^{-1}, (▽) 100 µg·ml^{-1}, (■) 300 µg·ml^{-1}, (□) 600 µg·ml^{-1} and (♦) 1000 µg·ml^{-1} were pipetted into separate wells of a 96 well plate. Seven consecutive additions of 25 µl were implemented, up to a total volume of 225 µl, a total dilution of 4.5 fold. All experiments were performed in duplicate wells. The main graph shows all data and inlay shows data for only (•) 10 µg·ml^{-1}, (○) 30 µg·ml^{-1}, (▼) 60 µg·ml^{-1}, for greater clarity.

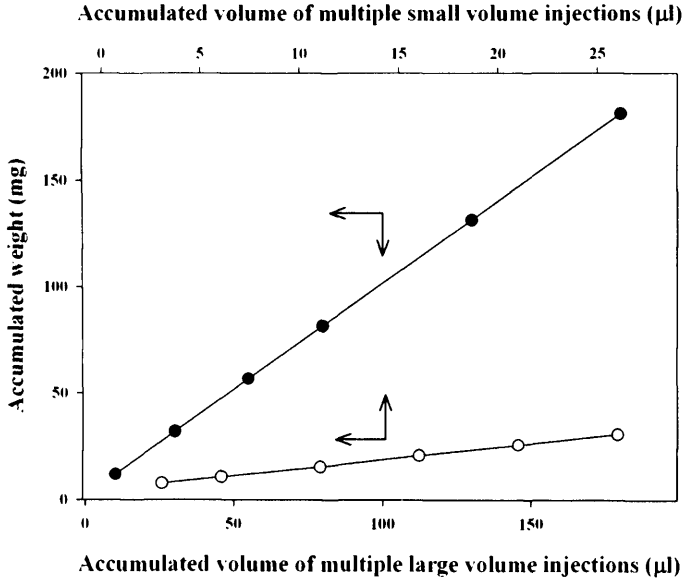


Figure 3.6: Accuracy of multiple liquid dispensing by the syringe pump into microtiter plates for various volume additions. Linear regressions (—) were fitted to the two sets of data. The slope, α, represents the accuracy and the standard error is a measure of the precision. For small injected volumes (○): $\alpha = 1.012 \pm 0.007$ g · cm^{-3}; and for large volumes (•): $\alpha = 0.995 \pm 0.001$ g · cm^{-3}.

water (0.997 mg·cm^{-3} at 25 °C), verifying the accuracy of titration (precision was not possible to determine for the low volumes dispensed). Linearity was observed down to 3 µl volume injections, making this the minimum accurately dispensable volume. This restricts the minimum denaturant concentration increase that could be made using serial addition. Incremental jumps in denaturant concentration were largest for the first injections, and the 3 µl volume injection minimum defined the number of useful data points that could be obtained for the initial baseline of the transition curve.

The equilibration time between injections of denaturant is critical to the the accurate determination of protein stability. The time required may vary considerably between proteins and can, therefore, adversely affect the throughput at which measurements can be made. Consequently it is expected that for rapidly equilibrating proteins, the serial addition mode would provide the highest throughput, whereas for slowly equilibrating proteins, the fixed volume mode would be best (Section 2.5). For unknown protein samples, a pre-screen of various equilibration times at a given denaturant concentration is, therefore, necessary to determine the the best experimental mode to use. The equilibration times were optimised for the serial addition mode, to ensure that cyt *c* and BSA unfolding both reached equilibrium. It was determined empirically that 15 minutes between injections were sufficient for both cyt *c* and BSA to fully equilibrate, resulting in serial addition experiments that took up to 10 hours for 25 denaturant concentrations. The equipment used produced a slight rise in temperature of 3 °C and approximately 5 % (w/w) evaporation over the duration of each experiment, highlighting potential improvements that could be made instrumentally.

The results reported here were all obtained using clear plates. All the wells had a uniform background signal, except for wells situated at the edge of the plate which had higher background signal. Only the inner wells were used for analysis with the fixed volume method to obtain reproducible replicates. The background noise was additive and constant for all readings taken from a single well, thus had no effect on results obtained by analysis with the serial addition method. This is exemplified by showing that an unfolding

curve obtained from an edge well (Figure 3.8, dataset ▲) has a similar profile to other unfolding curves obtained using inner wells. However, the $C_{1/2}$ obtained from the edge well (2.51 M) was distinctly lower than those obtained for all other wells, ranging from 2.63-2.75 M, and thus only inner wells should be used for improved accuracy and precision. The clear plates had a higher background noise, and the difference between native and unfolded baseline fluorescence signals were 3-5 times lower than those for black plates. However, black plates produced unexplained, inconsistent, volume dependant baseline drifts that distorted the m_G values obtained with stability calculations.

3.2.3 Conformational transition curves for cytochrome *c*

Conformational transition curves for cyt *c* from bovine and equine heart, obtained with the fixed volume method (Section 2.5), are plotted as the average fluorescence intensity from triplicate measurements (Figure 3.7). Curves obtained with the serial addition method are plotted as nine replicates performed in parallel (Figure 3.8). Stability values, $C_{1/2}$, m_G value (estimated with non-linear fitting) and ΔG (estimated with linear extrapolation) for both methods are summarised in Table 3.1. The reported stability values for the fixed volume method are obtained from one analysis and for the serial addition method they are averages of estimates for each of the nine parallel stability analyses. The inter-run variability was calculated from three separate experiments conducted on separate days, using different stock and protein solutions (Table 3.1).

$C_{1/2}$ estimations using the fixed volume method were accurate enough to rank protein variant stability in a population with similar $C_{1/2}$ values (typically ± 0.05 M). Estimations of $C_{1/2}$ and m_G values compare well with reported values in the literature (Table 3.1). Midpoints of unfolding for proteins can be estimated to an accuracy of ± 0.15 M using the serial addition method, but values of m_G in this case are significantly overestimated. Data obtained with the serial-addition, high-throughput method estimates the $C_{1/2}$ value of equine cyt *c* to be higher than that of the bovine counterpart. This is the opposite result as compared to the fixed volume and literature values where bovine cyt *c* has the

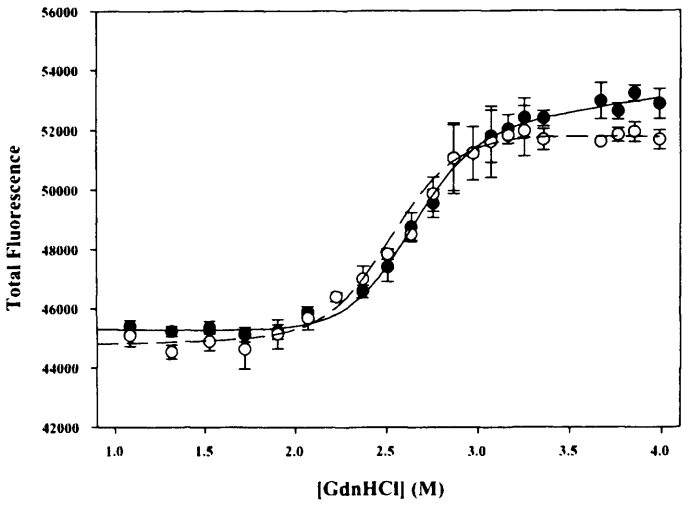


Figure 3.7: Conformational transition curves for oxidised cyt *c* obtained with the fixed volume method and GdnHCl. 20 μg (•) bovine heart and (○) horse heart cyt *c* were loaded per well. Wells were filled to a final volume of 215 μl with combinations of 100 mM TrisHCl buffer, pH 7.0 containing 0 M and 6.5 M GdnHCl to give the correct final denaturant concentration. Each concentration was measured in triplicate. Both datasets show curve fits as described in (Section 2.6).

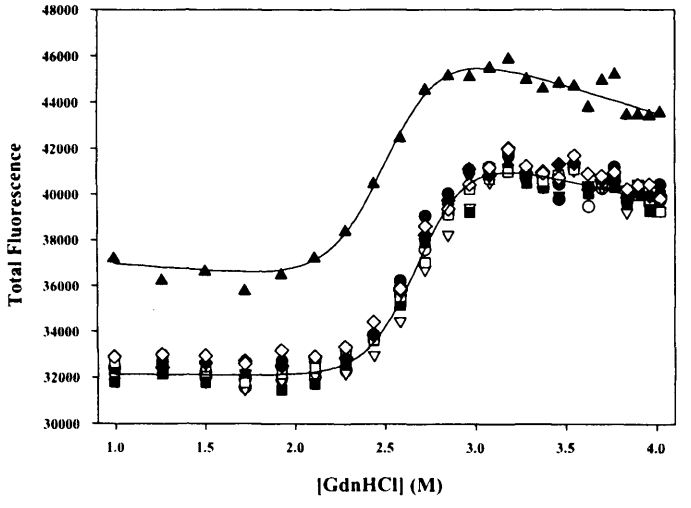


Figure 3.8: Conformational transition curves for oxidised cyt *c* obtained with the serial addition method and GdnHCl. Multiple replicates of conformational transition curves for horse heart cyt *c*. Each curve represents data obtained from a single well in which 20 μg cyt *c* was loaded. Wells were filled in small increments with the appropriate volume of 100 mM Tris, pH 7.0 containing 6.5 M GdnHCl to give the correct final denaturant concentration. Plates were agitated for 30 s followed by an equilibration period of 15 min before each measurement. One of the plots, (▲), represents data collected from an edge well. The curve fits are shown for the edge-well plot (▲), and for a typical example of the remaining non-edge well plots.

higher $C_{1/2}$ value (2.46 ±0.05 M and 2.51 ±0.19 M for bovine and equine cyt *c* respectively obtained from three separate runs of nine datasets each). However, the errors associated with $C_{1/2}$ values calculated by the serial addition method ($\pm 0.08 - 0.14$ M), are of similar magnitude to, and may partly account for, this discrepancy. Larger variation in pre- and post-transition baselines as well as evaporation and the small increase in temperature over the longer experiment time for serial addition are the most likely reasons for this occurrence. The ability to rank protein stabilities in terms of $C_{1/2}$ is sufficient for directed evolution screens, where the key requirement is to find enzymes that resist unfolding under the conditions required in a bioreactor (Burton *et al.*, 2002). It is also sufficient for therapeutic formulation in which a combination of excipients is desired that improve the resistance of the protein to unfolding (Remmele *et al.*, 1998).

The large error on m_G value calculation using the serial addition mode for bovine cyt *c* (6.2 ±1.8 kcal·mol^{-1}·M^{-1}) is a result of the inaccurate determination of the pre- and post-unfolding baselines. K_{obs} values, also determined by extension of the stable baselines into the transition region, and the subsequent calculation of ΔG_{obs} values are also adversely affected. Moreover, data used for linear extrapolation estimates show average regression coefficients of 0.93 ±0.11 (n = 36) suggesting that the large variations in ΔG_{H_2O} values originate from erroneous baseline determination. In this mode the initial additions cause large increments in the denaturant concentration, thus limiting the number of data points obtainable in this region. This limitation is a direct result of the minimum volume that can be accurately titrated by the syringe pump and suggests a critical area for further improvement. This leads us to conclude that with present instrumentation, ΔG values cannot be determined accurately with the serial addition method using linear extrapolation. ΔG_{H_2O} values obtained with the fixed volume method are reasonably accurate, but the ΔΔG for bovine and equine cyt *c* is underestimated using linear extrapolation (0.27 kcal·mol^{-1} calculated compared to 1.08 kcal·mol^{-1}) in literature.

Table 3.1: Summary of stability data for oxidised cyt *c* from bovine and equine heart obtained using clear microplates.

	$C_{1/2}$	m_G	ΔG	%CV		
	(M)	(kcal·mol^{-1}·M^{-1})	(kcal·mol^{-1})	$C_{1/2}$	m_G	ΔG
(A) Fixed volume mode						
Bovine	2.61 (0.03)[a]	3.47 (0.48)[a]	8.88 (1.27)[a]	1.15[b]	6.98[b]	13.38[b]
Equine	2.53 (0.03)[a]	3.34 (0.52)[a]	8.61 (1.51)[a]	4.03[b]	0.26[b]	9.00[b]
(B) Serial addition mode						
Bovine	2.50 (0.08)[c]	6.19 (1.83)[c]	14.26 (4.60)[c]	2.25[d]	22.66[d]	49.09[d]
Equine	2.67 (0.14)[c]	4.72 (1.19)[c]	11.83 (1.35)[c]	4.73[d]	4.01[d]	6.56[d]
(C) Reported literature values						
Bovine	2.63[e]	3.19[e]	8.38[e]	–	–	–
Equine	2.42[e]-2.60[f]	3.01[e]	7.27[e]-7.30[f]	–	–	–

[a] Standard deviations of the sample (σ) calculated from three datasets.
[b] Percentage coefficient of variation ($\sigma/\bar{x}$) calculated from three individual estimates.
[c] Standard deviations of the population calculated from nine datasets.
[d] Percentage coefficient of variation ($\sigma/\bar{x} \times 100$) calculated from twenty-seven datasets.
[e] (Knapp and Pace, 1974)
[f] (Brems *et al.*, 1982)

Table 3.2: Summary of stability data for oxidised cyt *c* from bovine and equine heart using fixed volume mode and different black microplates.

	$C_{1/2}$	m_G	ΔG	%CV		
	(M)	(kcal·mol^{-1}·M^{-1})	(kcal·mol^{-1})	$C_{1/2}$	m_G	ΔG
(A) Black Well						
Bovine	2.47 (0.18)[a]	1.96 (0.25)[a]	4.83 (0.26)[a]	7.13[b]	12.54[b]	5.43[b]
Equine	2.38 (0.23)[a]	2.07 (0.29)[a]	4.89 (0.20)[a]	9.90[b]	14.16[b]	4.18[b]
(B) Black and Clear Well						
Bovine	2.62	5.24	11.87	–	–	–
Equine	2.59	2.77	7.18	–	–	–

[a] Standard deviations of the sample (σ) calculated from three datasets.
[b] Percentage coefficient of variation ($\sigma/\bar{x}$) calculated from three individual estimates.

3.2.4 Conformational transition curves for BSA:palmitic acid mixtures

The conformational transition curves of BSA generated with the fixed volume method required a 9 M urea stock-solution to ensure complete unfolding of the most stable protein preparation (Section 2.5). GdnHCl, a stronger denaturant (Pace, 1986), was used instead of urea for the serial addition unfolding analysis for two practical reasons: i) the maximum urea concentrations attainable after serial dilution did not completely unfold BSA; ii) high concentrations of urea, approaching saturation can precipitate easily and block the needle of the syringe pump. Unfolding curves of various BSA:palmitate preparations obtained with fixed volume and serial addition methods are depicted in Figure A.5 and 3.10 respectively. BSA unfolding curves at high palmitate concentrations display at least one intermediate in the unfolding pathway, as seen from the deviation from a simple two-state unfolding curve. This intermediate is attributed to the early unfolding of the less stable C-terminal part of the protein. Fatty acid molecules partially stabilise the intermediate (Ahmad and Qasim, 1995), increasing the midpoint of urea induced unfolding. Interestingly, the fluorescence unfolding curves for BSA in GdnHCl in microtiter plates (Figure 3.10) can also detect the presence of intermediate states that occur as three palmitate molecules bind to each BSA molecule, i.e. at high palmitate concentrations. The unfolding intermediates in this and previously reported urea induced unfolding curves (Ahmad and Qasim, 1995) are less prominent than for the GdnHCl induced unfolding curves.

3.2.5 Numerical methods for high-throughput indexing of relative protein stability

The change in midpoints of unfolding ($\Delta C_{1/2}$) are accurate, reliable and straight-forward to calculate for proteins following a two-state unfolding. However, not all proteins follow the ideal two-state model of unfolding as is exemplified with BSA. Accurately comparing stabilities or $\Delta C_{1/2}$ values of variants with multiple or poorly defined transitions is much harder. High-throughput screening methods require simple, yet reliable, data processing

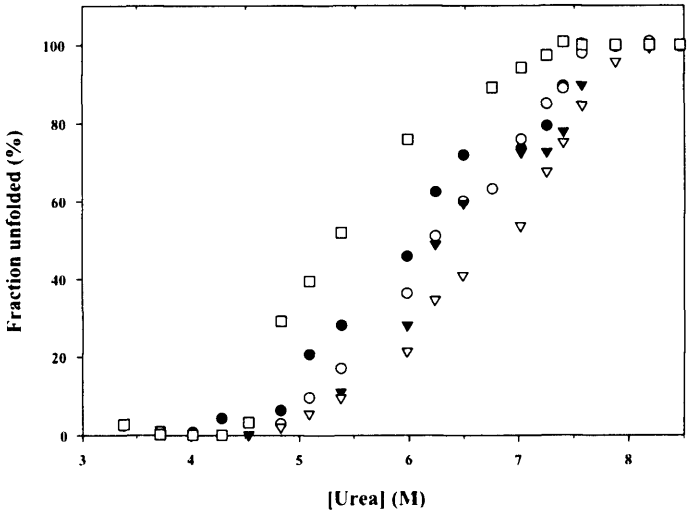


Figure 3.9: Conformational transition curves for bovine serum albumin (BSA) obtained with the fixed volume method. 40 μg of BSA stabilised with palmitic acid at molar ratios of (□) 1:0, (•) 1:2, (○) 1:3, (▼) 1:4, and (▽) 1:5 were loaded per well. Wells were filled to a final volume of 280 μl with combinations of 60 mM sodium phosphate buffer, pH 7.0 containing 0 M and 9.0 M urea to give the correct final denaturant concentration. Each sample was measured in triplicate.

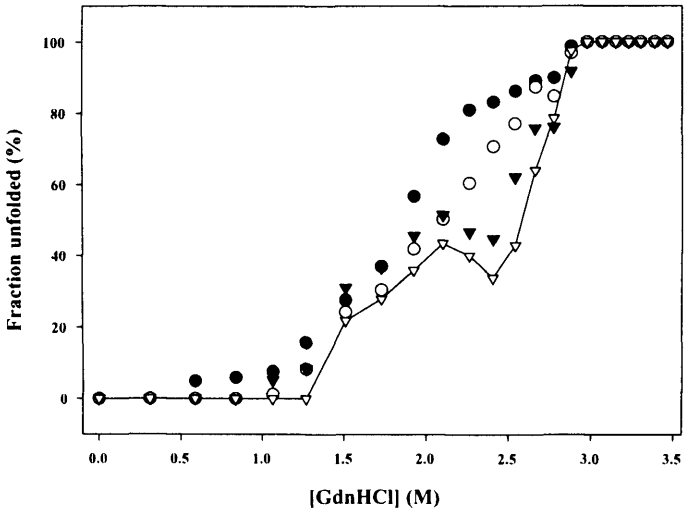


Figure 3.10: Conformational transition curves for bovine serum albumin (BSA) obtained with the serial addition method. 50 μl of a 0.8 mg·ml^{-1} BSA solution stabilised with palmitic acid at molar ratios of (•) 1:0, (○) 1:3, (▼) 1:5, (▽) 1:6 and buffered with 60 mM sodium phosphate, pH 7.0, were loaded per well. Unfolding was induced by titrating small volumes of a 5.5 M GdnHCl stock solution buffered with 60 mM sodium phosphate, pH 7.0 into each well. Plates were agitated for 30 s followed by an equilibration period of 15 min before each measurement. Each plot represents data for an unfolding curve obtained from a single well. Data points are interpolated for the highest BSA:palmitate molar ratio (▽), to demonstrate the deviation from a two-state transition.

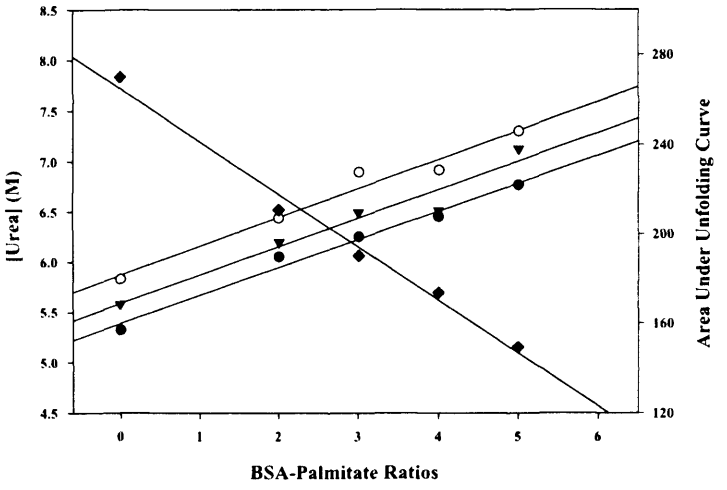


Figure 3.11: Stability indexing for palmitate stabilised bovine serum albumin (BSA) using different numerical methods. Each dataset in Figures A.5 was fit to the two-state unfolding equations 1-4, described in Section 2.6, even though it deviates from two-state transition. Linear increases in stability are represented by values calculated from: (•) denaturant concentration at apparent midpoint of unfolding ($[Urea] = 0.279 \cdot palmitate + 5.387$; $r^2 = 0.983$); (▼) urea concentration at 60 % unfolding ($[Urea] = 0.283 \cdot palmitate + 5.586$; $r^2 = 0.949$); (○) urea concentration at 70 % unfolding ($[Urea] = 0.287 \cdot palmitate + 5.869$; $r^2 = 0.969$); (♦) total area under unfolding curve ($AUC = -23.65 \cdot palmitate + 265$; $r^2 = 0.988$). Data obtained using fixed volume method and urea.

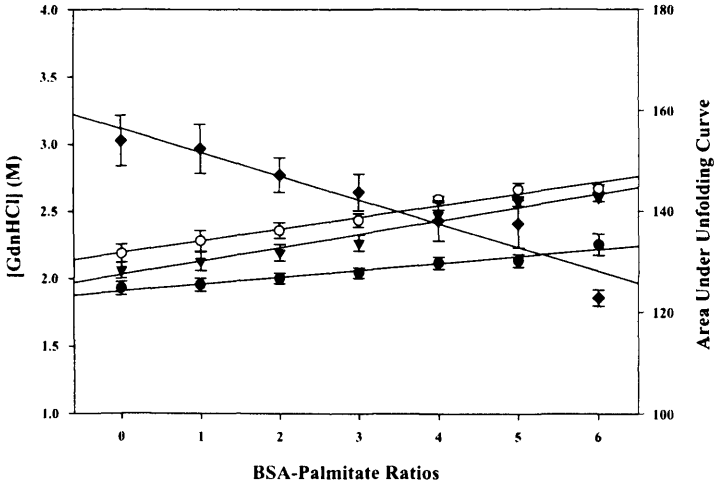


Figure 3.12: Stability indexing for palmitate stabilised bovine serum albumin (BSA) using different numerical methods. Each dataset in Figures 3.10 was fit to the two-state unfolding equations 1-4, described in Section 2.6, even for those cases that clearly deviate from two-state transition. Linear increases in stability are represented by values calculated from: (•) denaturant concentration at apparent midpoint of unfolding ($[Urea] = 0.051 \cdot palmitate + 1.905$; $r^2 = 0.958$); (▼) denaturant concentration at 60 % unfolding ($[Urea] = 0.099 \cdot palmitate + 2.029$; $r^2 = 0.962$); (○) denaturant concentration at 70 % unfolding ($[Urea] = 0.087 \cdot palmitate + 2.191$; $r^2 = 0.971$); (♦) total area under unfolding curve ($AUC = -4.73 \cdot palmitate + 156.45$; $r^2 = 0.915$). Data obtained using serial addition method and GdnHCl.

protocols to identify positive 'hits'. The method reported here will required numerical tools that will be able to index the relative stability of related proteins, regardless of the complexity of the unfolding profile.

Three, relatively simple, numerical methods were tested for their suitability to correctly index the stability of the different BSA preparations. These methods are i) the unfolding datasets were fitted to a two-state transition model, calculating an approximate estimate for a global $C_{1/2}$ value; ii) the areas below the unfolding curves were calculated using the trapezoidal rule and iii) 60% or 70% unfolding was selected as reference point and the denaturant concentration calculated using a linear equation defined by the two data points either side of the arbitrary set reference value. The numerical methods are described in detail in Section 2.7. The method estimating an apparent midpoint of unfolding (Section 2.7.1) is probably the least robust and insensitive to small changes in the equilibrium transitions of an unfolding curve displaying multiple states. The method estimating the area under the curve (Section 2.7.2) will detect stabilisation effects so long as the midpoint shifts to higher denaturant concentrations. It will be insensitive to effects that only affect the m_G-value of the transition state. The method making use of arbitrary set reference points is the most flexible method since the reference can be set in the region where the unfolding curve is most sensitive to changes in stability (Section 2.7.3). In the case of BSA, unfolding curves obtained by the serial addition method, shifted most significantly in the region of 60-80 % unfolding (Figure 3.12). This method thus requires some prior knowledge about the unfolding behaviour of the protein under investigation that can be obtained with a preliminary study.

All three numerical methods were tested on normalised unfolding curves of BSA obtained with either the fixed volume or serial addition methods. Figure 3.11 and 3.12 depicts the stability estimates using these techniques with urea and GdnHCl induced unfolding of BSA respectively. Although these are crude estimates of stability, they are still functional and effective at identifying fatty-acid induced stabilisation. All three numerical methods can confidently determine the order of BSA stabilities present at four

palmitate concentrations. This demonstrates microtiter plate protein unfolding as a useful high-throughput proteomics method for elucidating the ligand-binding aspects of protein function that alter protein stability.

3.3 Conclusion

We have developed and characterised an affordable high-throughput screening method for the direct measurement of the stability of proteins in a commercially available microplate reader. The method generates unfolding curves in microwells by autotitration of denaturant and measuring the resulting changes in tryptophan fluorescence. The $C_{1/2}$ estimate is the most reliable stability parameter obtained with both methods and is directly related to thermodynamic stability when substituted with [D] in Equation 2.4 (Section 2.6). Analysis of mutant stability requires estimations of ΔG values, calculated by substituting [D] with $C_{1/2}$ values in Equation 2.4 (Kellis *et al.*, 1989). This approach assumes that m_G values remain the same for all variants of a protein. The serial addition method is useful when screening for changes in $C_{1/2}$ values of greater than 0.15 M. Combined with a suitable numerical data analysis technique it can be a powerful tool for screening mutant biocatalyst libraries for improved stability. The fixed volume method produces more reliable datasets and parameter determination, but requires more protein, plates and solutions for the unfolding analysis. This method is useful for a more accurate characterisation of either small numbers of rationally designed mutants or a secondary screen, confirming hits from the higher throughput serial-addition screen for protein stability.

Though not all proteins are amenable to fluorescence monitored unfolding, this application can, at least in theory, be extended to absolute extinction coefficient and circular-dichroism based detection. These detection methods also allow for the screening of protein targets for which high-throughput activity screens are difficult to develop. The high-throughput determination of stability in microplates could be applied to the fields of directed evolution, proteomics and therapeutic protein formulation, to produce enzymes stable under process conditions, to characterise functions related to stability changes

upon ligand binding, or to ascertain optimal combinations of excipients that prolong the shelf-life of biopharmaceuticals.

The equipment used in this study is functional, and may be readily integrated with conventional automation platforms (Di *et al.*, 2004), but is not optimised for this type of work. It can be considered a prototype for dedicated instrumentation that would be modified to particularly accommodate the needs of the high-throughput unfolding analysis. Improvements to consider are accurate titration down to 1 µl volumes, solution dispensing close to the surface of the solution in each well, minimised evaporation, active temperature control, and multiple syringe pumps for dispensing various stocks in parallel. While the data in our studies was transferred to a more amenable software package for analysis, flexible software for the integrated data collection, unfolding analysis and stability indexing would also be preferable. All of these modifications are within the reach of current instrumentation technology.

As the high-throughput expression and purification of proteins in microwells is now fully automatable (Edgell *et al.*, 2003), they could be combined easily with the methods described here to produce a fully automated system for the high-throughput determination of protein stability in microwells, suitable for directed evolution. Such an approach would be highly complementary to related research in our laboratory on the automated microscale process evaluation of evolved biocatalyst libraries (Lye *et al.*, 2003) in linked sequences of fermentation (Elmahdi *et al.*, 2003), bioconversion (Doig *et al.*, 2002) and product recovery operations.

Chapter 4

Transketolase: Purification and Unfolding Study

4.1 Introduction

There is a considerable amount of sequence and structural information available for the transketolase enzyme. The potentially lucrative biomedical and biocatalytic applications provided the driving force for substantial characterisation efforts. Biocatalytic applications for transketolase have been identified early on, with the possibility of catalysing asymmetric and C-C bond formation. The transketolase reaction provides a means of producing complex sugars (Takayama *et al.*, 1997), carbohydrate analogues (Gijsen *et al.*, 1996), precursors for food flavours (Hecquet *et al.*, 1996) and compounds for pharmaceutical (Hecquet *et al.*, 1994b) and agrochemical industries (Myles *et al.*, 1991).

Hydroxypyruvate (HPA) is favoured as the ketol donor which forms carbon dioxide as a by-product, making the reaction effectively irreversible. *E. coli* transketolase accepts HPA much better than the yeast or spinach counterparts, resulting in enzyme reactions that are 5 to 10 times higher (Sprenger and Pohl, 1999). Although numerous potentially useful reactions for transketolase have been reported, many of the substrates and products were shown to be sensitive to alkaline environments (Mitra *et al.*, 1998). Due to the labile substrates used, transketolase reactions thus far have not be conducted under harsh

conditions (high temperature or extreme pH). However endeavours aimed at broadening transketolase substrate specificity are underway and may potentially require more stable biocatalysts. Stability refers to the ability to maintain the unique chemical and structural properties that are required for activity. Improving stability may also prolong the half-life of transketolase under mild reaction conditions to increase the reusability or allow one-pot, multi-step applications in the presence of organic solvents.

4.1.1 Physico-chemical properties of transketolase

Transketolase purified from both mammalian and microbial sources seems to be remarkably stable under modest conditions (4-50 °C and pH 4-10). Pure preparations of human erythrocyte transketolase retain 50% activity after incubation at 55 °C for 5 min . By adding BSA prior to heat incubation, the loss is reduced to 3% (Takeuchi *et al.*, 1986). Porcine liver transketolase shows no activity loss at 50 °C for 1 hr in the presence of thiamine pyrophosphate (TPP) (Philippov *et al.*, 1980). Stability of transketolase is pH dependent with mammalian transketolase being more stable under alkaline conditions. Human erythrocyte transketolase displays the greatest stability between pH 7.5-10 (Heinrich and Wiss, 1971). Likewise, porcine liver (Philippov *et al.*, 1980) and rat liver (Paoletti, 1983) transketolase are also most stable in the alkaline pH range of 7.8 and 8.2. Transketolase from bakers' yeast and *E. coli* appear to be less stable when stored under alkaline conditions. When stored at pH 7.9, yeast transketolase loses more than 30% of the activity after two weeks even at -20 °C formulated with glycerol and dithiothreitol (DTT) (Cavaliere *et al.*, 1975). However, when stored at -20 °C as a crystalline suspension in saturated ammonium sulphate the enzyme remained stable for months (Kochetov, 1982). The purified enzyme from *E. coli* in glycylglycine buffer at pH 8.5 could be stored at -20 °C in the presence of DTT and 20% glycerol for three months with less than 20% loss of activity (Sprenger *et al.*, 1995). At 4 °C in glycylglycine buffer at pH 8.5 the loss of activity was approximately 10% per month. *E. coli* transketolase is irreversibly denatured at pH values below 6.5 and also subject to oxidative deactivation at high concentrations

of glycolaldehyde (Mitra *et al.*, 1998).

The removal of TPP from mammalian transketolases requires acidic conditions, whereas TPP is easiest resolved from yeast (Srere *et al.*, 1958) and *E. coli* (Sprenger *et al.*, 1995) enzymes under alkaline conditions. After complete removal, the addition of TPP and a divalent metal ion are absolutely necessary to restore catalytic activity (Heinrich *et al.*, 1972). The addition of cofactors prevent the loss of activity during purification of transketolase. Cofactor binding studies indicated that both TPP and Mg^{2+} dissociate easily from *E. coli* holo-transketolase at all pH conditions suitable for biotransformation reactions (pH 6.5-9.5) (Mitra *et al.*, 1998).

The optimum pH for activity is similar in all characterised transketolases. For the enzyme from human erythrocytes (Heinrich and Wiss, 1971) the optimal pH is 7.7. A value between pH 7.5 and 7.6 has been reported for transketolases from bakers' yeast (Datta and Racker, 1961) and from spinach leaves (Villafranca and Axelrod, 1971). An optimal pH range between 7.4 and 8.2 has been observed for liver transketolase from various species (Paoletti, 1983; Philippov *et al.*, 1980; Simpson, 1960). The *E. coli* enzyme displays maximum activity in glycylglycine buffer at pH 8.0 to 8.5 (Sprenger *et al.*, 1995) and between pH 7.0 and 7.5 in phosphate buffer (Mitra *et al.*, 1998).

Besides TPP, the necessity of divalent metal ions for activity has been shown. For bakers' yeast transketolase (Schenk *et al.*, 1998) the reconstitution rate increases in the order $Ni^{2+} < Mg^{2+} < Co^{2+} < Mn^{2+} < Ca^{2+}$. The final catalytic activity is independent of the nature of the divalent metal ion (Heinrich and Wiss, 1971). A similar observation has been reported for *E. coli* transketolase (Sprenger *et al.*, 1995). Mammalian transketolases also require the presence of divalent cations for activity. Activity may be restored by Mg^{2+}, Ca^{2+}, Mn^{2+} and Co^{2+}, but not Zn^{2+} or Cu^{2+} (Jung *et al.*, 1988).

4.1.2 Development of the pQR711 *E. coli* transketolase expression vector

The *E. coli tkt* gene was first cloned by (Draths and Frost, 1990) as a 5 kb *Bam*HI genomic fragment into the low copy number pBR325 vector producing pKD112A. The

5 kb *Bam*HI fragment was excised from pKD112A and subcloned into a high copy number pUC18 *E. coli* expression vector (French and Ward, 1995). The *tkt* gene was oriented in both directions with respect to the *lac*Z promoter, giving rise to the pQR182 and pQR183 constructs respectively. The same author then amplified the *tkt* gene and *tkt* promoter from pQR182 as a 2.2 kb fragment and subcolned it into pCRScript SK(+). The *tkt* gene was inserted in both orientations with respect to the *lac* promoter, producing the pQR711 and pQR706 constructs. Of all the constructs, pQR711 produced the highest yields of transketolase in shake flask experiments and was subsequently the vector of choice. The pQR711 vector was used routinely in this study and all modified plasmids were derived from pQR711.

4.1.3 Purification protocols for transketolases

The purification protocols of transketolase from microbial, yeast and mammalian sources are very similar. The primary capture step involves a precipitation fractionation step and subsequent purification is done using anion-exchange chromatography. An additional polishing chromatographic step is used to produce homogenous transketolase. Ammonium sulfate fractionation was used for over-expressed *E. coli* transketolase (Sprenger *et al.*, 1995; Littlechild, 1995) as well as for rat liver (Horecker *et al.*, 1953) and porcine liver (Simpson, 1960) transketolases. Transketolase from bakers' yeast was enriched with acetone and ethanol fractionation steps (Srere *et al.*, 1958). Chromatographic purification involved consecutive steps of Q-Sepharose FF anion-exchange and Fractogel-EMD-DEAE anion-exchange (Sprenger *et al.*, 1995), Q-Sepharose FF anion-exchange and Superose 12 gel-filtration (Littlechild, 1995), DEAE-cellulose (Srere *et al.*, 1958) or DEAE-Sephadex and hydroxyapatite (Horecker *et al.*, 1953; Simpson, 1960). Erythrocyte transketolase has been partially purified with an acid precipitation step (Takeuchi *et al.*, 1986). A similar approach was suggested by (Mitra *et al.*, 1998) where the rate of denaturation for *E. coli* transketolase is much lower than the relative precipitation rate of the contaminating proteins at pH 5.5.

4.1.4 Dimeric proteins in biotransformation

Obligatory homodimeric proteins are often considered unsuitable for biotransformation reactions since an early loss in the protein quaternary structure results in a loss in enzyme activity. Many dimeric proteins denature with a two-state equilibrium transition, while others display stable intermediates in the process (Neet and Timm, 1994). For dimeric proteins displaying a single transition the conformational stability is significantly greater than that of the monomeric proteins. In many of these cases the increase in conformational stability is a result of the intersubunit interactions formed upon oligomerisation.

There are several examples of homodimeric biocatalysts used successfully in the biocatalytic industry. These include aspartate aminotransferase (Yano *et al.*, 1998), D-amino acid oxidase (Pilone and Pollegioni, 2002) and xylose isomerase (Hess *et al.*, 1998). Several methods are used to increase the stability of the quaternary structure including immobilisation (Brocklebank *et al.*, 1999), cross-linking (Hublin *et al.*, 2002; Davis, 2003) and lipid/micellar formulation (Pire *et al.*, 2004). Dimeric proteins have also been re-engineered to be functional monomers (MacBeath *et al.*, 1998). Obligatory homodimeric proteins, however exist in thermophile organisms and are known to be stable at temperatures of up to a 100 °C (Arnone *et al.*, 1997). The engineering of thermostable homodimeric proteins is an attainable goal, providing the principles of protein-protein interactions are understood (Chothia and Janin, 1975). Though interface residues are the most important contributors to dimer stability, general protein compactness (Bell *et al.*, 2002) and cofactor interaction (Risse *et al.*, 1992) are also factors to be considered. A comprehensive characterisation of the accessible surface area (ASA), hydrophobic patches, salt bridges and hydrogen bonds is required before a strategies for improving dimer stability can be proposed.

4.1.5 Unfolding characterisation of transketolase

This section aims to provide an initially characterisation of the unfolding properties of *E. coli* transketolase (*tkt*). A His_6-tag was added to the N-terminus of the transketolase gene

harboured on the pQR711 plasmid to allow rapid purification by affinity chromatography, since pure protein is required for both 'classical' and high-throughput unfolding experimentation. Unfolding parameters of transketolase are estimated using chemical induced denaturation. Mutant variants of wild-type transketolase, with reduced dimer stability, were also created. These mutants provide first insight into the unfolding characteristics of transketolase and are useful for direct comparison with microwell unfolding results obtained in Chapter 5. Transketolase used in this study, either wild-type or mutant, contained the N-terminal His_6-tag.

4.2 Results and Discussion

4.2.1 Modification of pQR711

The pQR711 plasmid, extracted from the *E. coli* JM107 pQR711 strain was modified in sequential steps (Section 2.8). The *Bgl*II site was inserted first, producing pQR790. The whole plasmid was amplified successful (Figure 4.1) with the QuickChange^{TM} protocol and transformed into *E. coli* XL1-Blue subcloning strain (Section 2.12.1). After selection on ampicillin LB agar plates, 5 random colony forming units were picked for plasmid isolation. Insertion of the restriction site into pQR790 was determined by means of a diagnostic double restriction enzyme digestion using *Bgl*II and *Xba*I (Section 2.9). The *Xba*I restriction site is in the pCRScript SK(+) backbone at the C-terminal end of the *tkt* gene, allowing for excision of the complete gene and subcloning if necessary. All five plasmid preparations harboured the restriction site (Figure 4.2). The His_6 tag was inserted into pQR790 creating pQR791. Complementary DNA-oligomers containing the insertion sequences were used to successfully amplify the whole plasmid (Figure 4.3). Competent XL1-Blue was transformed with the PCR product and grown on ampicillin-LB agar selection media (Section 2.12.1). Six single colonies were picked for plasmid isolation and sequencing. The DNA sequencing confirmed that four of the six transformants harboured the correct plasmid modification (Table 4.1). One of the successful constructs, HisTK6,

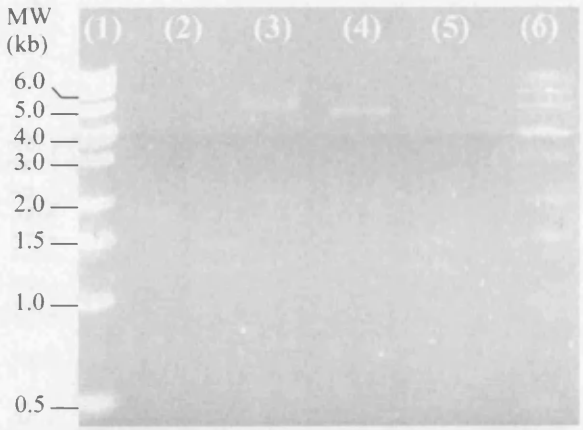


Figure 4.1: Analysis of QuickChange™ amplification products with a 0.9% TBE-agarose gel. Plasmid products should by modified to contain the -15 C→A mutation producing a *Bgl*II restriction site. The lanes are as follow: (1 and 6) Standard 0.5-12 kb markers; (2-5) are *Dpn*I digest resistant whole plasmid product from amplification reactions containing 5, 10, 15 and 20 ng of pQR711 template respectively. The product is expected to have a molecular weight of 5.4 kb, similar to that of pQR711. Only reactions containing 10 or 15 ng DNA template were successful.

was transformed into competent *E. coli* BL-21 and JM107 (Sections 2.12.1 and 2.12.2) for functional expression.

The D381A and Y440A mutants were created with the QuickChange™ protocol using pQR791 as template and appropriate DNA-oligomers (Section 2.8). The whole plasmid was amplified successfully (Figure 4.4) and *E. coli* XL-10 Gold transformed with the PCR product (Section 2.12.3). Transformants were grown under ampicillin selection pressure and three colonies of each mutant were picked for plasmid isolation. Mutations were confirmed with DNA sequencing (Table 4.2).

4.2.2 Overexpression of His_6-transketolase in *E. coli*

E. coli BL-21 and JM107, both harbouring the pQR791 plasmid, were cultured in identical shake-flask experiments and the cell density standardised. Cell cultures were harvested from equivalent volumes and used as starting material for Ni-NTA spin column purification (Section 2.14). Fractions from each purification step were retained for SDS-PAGE analysis (Section 2.20) and activity determination using the enzyme-linked activity assay (Section 2.17). Transketolase migrates as a double band in SDS-PAGE analysis (Fig-

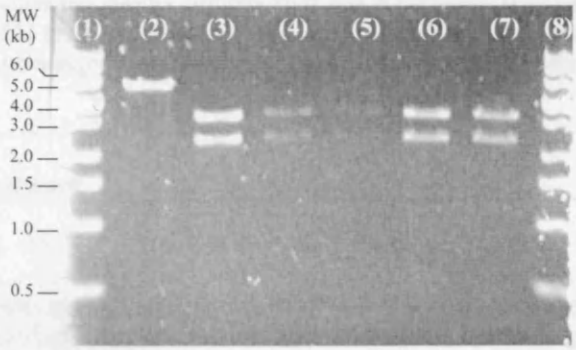


Figure 4.2: Introduction of a -15 *Bgl*II restriction site is confirmed by analysis of restriction digests with a 0.9% TBE-agarose gel. The lanes are as follow: (1 and 8) Standard 0.5-12 kb markers; (2) Linear plasmid product of pQR711 after a *Xba*I and *Bgl*II double restriction reaction, indicating absence of a *Bgl*II site; (3-7) Restriction fragments of purified plasmid from 5 transformants digested with *Xba*I and *Bgl*II. The fragments of approximately 2.3 kb and 3.1 kb are the *tkt* gene and pCRScript SK(+) vector respectively, indicating successful introduction of the *Bgl*II restriction site.

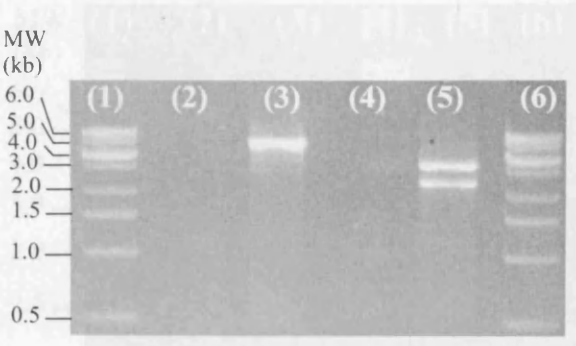


Figure 4.3: Analysis of QuickChange^{TM} amplification and restriction digestion products of N-terminal His_6-tag insertion with a 0.9% TBE-agarose gel. The lanes are as follow: (1 and 6) Standard 0.5-12 kb markers; (2) *Dpn*I digested QuickChange^{TM} reaction without *Pfu*-Turbo DNA-polymerase; (3) *Dpn*I digest resistant whole plasmid product amplified with *Pfu*-Turbo DNA-polymerase and expected product with molecular weight of 5.4 kb; (4 and 5) Products of lanes 2 and 3 respectively after the *Xba*I and *Bgl*II double restriction digestion. The fragments of approximately 2.3 kb and 3.1 kb are the *tkt* gene and pCRScript SK(+) vector respectively.

Table 4.1: The DNA sequencing results confirm that the N-terminal His_6-tag was inserted successfully producing pQR791. pQR711 is the original template and TKHisF the sequence of the sense strand DNA-oligomer used in the QuickChange™ reaction. HisTK1 to HisTK6 are the six transformants selected for DNA sequencing.

Template	Sequence
	-15 -5 5 15 25 35 45
	|....|.... |....|.... ....|....| ....|....| ....|....| ....|....| ....|
pQR 711	tcatccgatc tggagtcaaa atg------- ---------- -tcctcacgt aaagagcttg ccaat
HisTK1	tcatcagatc tggagtcaaa atgcatcacc atcaccatca ctcctcacgt aaagagcttg ccaat
HisTK2	tcatcagatc tggagtcaaa at-catcacc atcaccatca ctcctcacgt aaagagcttg ccaat
HisTK3	tcatcagatc tggagtcaaa atg------- ---------- -tcctcacgt aaagagcttg ccaat
HisTK4	tcatcagatc tggagtcaaa atgcatcacc atcaccatca ctcctcacgt aaagagcttg ccaat
HisTK5	tcatcagatc tggagtcaaa atgcatcacc atcaccatca ctcctcacgt aaagagcttg ccaat
HisTK6	tcatcagatc tggagtcaaa atgcatcacc atcaccatca ctcctcacgt aaagagcttg ccaat
TKHisF	--atcagatc tggagtcaaa atgcatcacc atcaccatca ctcctcacgt aaagagcttg cc---

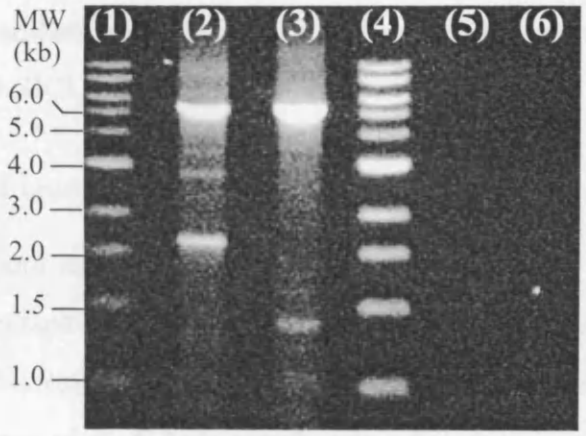


Figure 4.4: Analysis of QuickChange™ amplification products for producing the D381A and Y440A mutations with a 0.9% TBE-agarose gel. The lanes are as follow: (1 and 4) Standard 0.5-12 kb markers; (2 and 3) *Dpn*I digested QuickChange™ D381A and Y440A products amplified with *Pfu*-Turbo DNA-polymerase products with expected molecular weight of 5.4 kb; (5 and 6) *Dpn*I digested QuickChange™ reaction without *Pfu*-Turbo DNA-polymerase.

ure 4.5). This is most likely due to the decomposition of transketolase during the heat denaturation step prior to loading the samples onto the SDS gel (Sprenger *et al.*, 1995). The molecular weight of His_6-transketolase is estimated to be between 75-80 kD. This is in reasonable agreement with the theoretical value of 73.08 kDa, bearing in mind the moderate accuracy of molecular weight determination with SDS-PAGE analysis.

The total enzyme activities present in cell free extracts, as measured with the enzyme-linked assay (Section 2.17), were estimated to be 3.05×10^{-2} U·ml^{-1} for BL21 and 1.38×10^{-1} U·ml^{-1} for JM107. The activity results indicate that the JM107 strain produces approximately 4.5 fold more transketolase than the BL21 strain. This result is also corroborated with SDS-PAGE analysis. Lanes loaded with samples from the wash and elution steps of the JM107 strain contained more transketolase than the BL21 counterparts (Figure 4.6). Equal sample volumes were loaded in each lane for comparison of SDS gels. JM107 was retained as host strain for functional expression of His_6-tagged wild-type transketolase since it produced higher space-time yields under the conditions tested.

The D381A and Y440A mutants however were produced using *E. coli* XL-10 as expression host. This strain is reported to outperform JM107, attaining higher functional expression levels of transketolase with the pQR711 expression vector (personal communication O. Miller, 2004, UCL).

4.2.3 Inhibition of transketolase by NaCl and imidazole

The effects of NaCl and imidazole on the transketolase activity were evaluated (Section 2.19). In the standard Ni-NTA protocol NaCl and imidazole are included as purification and elution buffer components respectively. Transketolase activity was determined in absence and presence of these two compounds using the enzyme-linked and HPLC activity assays (Sections 2.17 and 2.18). Both buffer components were tested at a concentration of 250 mM, the typical working range required for purification. Inhibition effects were observed for both NaCl and imidazole using both assay methods. Different inhibition results were obtained with the two assay methods. The enzyme-linked assay re-

Table 4.2: The DNA sequencing results confirm mutations made to pQR791 plasmid, producing D381A and Y440A. ECTKT is the sequence of the original *tkt* gene. D381AFWD and Y440AFWD are the sense strand equivalent DNA-oligomers used to create the mutations with the QuickChange™ protocol.

Template	Sequence
D381A mutant	1125 1165
	|....|.... |....|.... |....|.... |....|.... |.
ECTKT	attcctcggc ggttctgctg acctggcgcc gtctaacctg ac
D381A	attcctcggc ggttctgctg ctctggcgcc gtctaacctg ac
D381AFWD	-ttcctcggc ggttctgctg ctctggcgcc gtctaacctg a-
ECTKT (aa seq)	F L G G S A D L A P S N L
D381A (aa seq)	F L G G S A A L A P S N L
Y440A mutant	1300 1340
	|....|.... |....|.... |....|.... |....|....|
ECTKT	accttcctga tgttcgtgga atacgcacgt aacgccgtac
Y440A	accttcctga tgttcgtgga agctgcacgt aacgccgtac
Y440AFWD	-ccttcctga tgttcgtgga agctgcacgt aacgccgta-
ECTKT (aa seq)	T F L M F V E Y A R N A V
Y440A (aa seq)	T F L M F V E A A R N A V

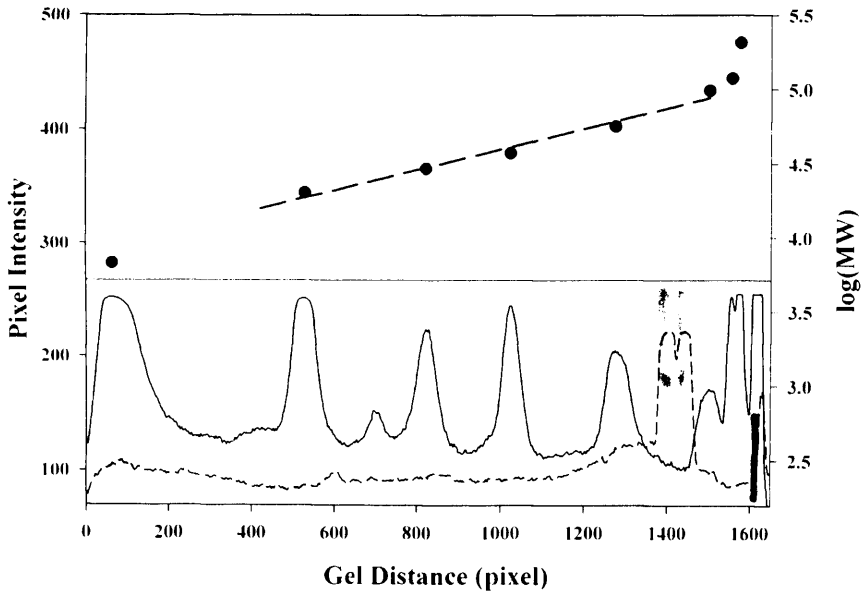


Figure 4.5: A typical SDS-PAGE analysis of purified His_6-transketolase using a 12% polyacrylamide gel. An electrophoresis peak profile was generated using ImageJ 1.32j image analysing software (http://www.rsb.info.nih.gov/ij/) with (—) representing the molecular weight marker lane and (---) the transketolase lane. A molecular weight calibration curve was created by plotting the log-values of molecular weight marker standards against migration distance. The molecular weight of His_6-transketolase was estimated to be between 75 and 80 kDa.

quires transketolase to utilise xylulose-5-phosphate and ribose-3-phosphate as substrates while for the HPLC activity assay transketolase utilises β-HPA and glycolaldehyde as substrates. Differences in substrate-inhibitor interactions may provide one explanation for the inhibition results. In the presence of 250 mM NaCl, transketolase activity was inhibited by 6.6% according to the HPLC assay and 25.8% using the enzyme linked assay. Other enzyme components of the enzyme-linked assay may also be inhibited, reducing the overall consumption rate of NADH further, thus increasing the apparent inhibition using this method. A concentration of 250 mM imidazole inhibited transketolase activity by 12.38% according to the HPLC assay and 8.3% as measured by the enzyme-linked assay. Based on these results NaCl was omitted from the purification protocol. Imidazole however is an essential elution component that cannot be omitted. The inhibitory mode of imidazole on transketolase has not been determined, however it suggests that imidazole acts as a ligand for the transketolase enzyme. Any ligand binding interaction has the potential of changing the measured protein stability. From this perspective it is essential to remove all imidazole before stability analysis to ensure a true measurement is made. Not only does imidazole strongly absorb UV light used for excitation of tryptophan residues, but it also quenches the emission reducing fluorescence intensity significantly (Vos and Engelborghs, 1994). Removal of imidazole is important in order to obtain suitably strong and background free emission signals for stability measurements. Alternatively to imidazole, a pH elution can be performed but, transketolase denatures irreversibly at the

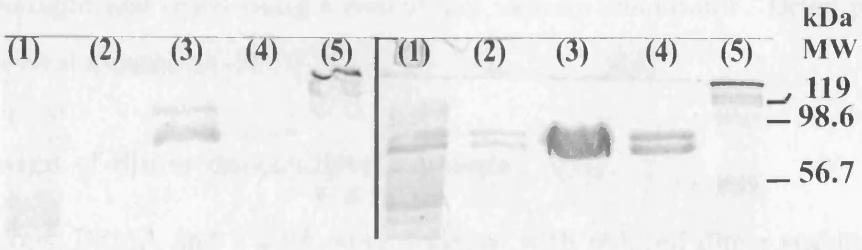


Figure 4.6: SDS-PAGE analysis, using a 12% acrylamide gel, of Ni-NTA spin column purified His_6-tagged wild-type transketolase overexpressed in (**A**) *E. coli* BL21 and (**B**) JM107. Equal amounts of starting material were used to compare expression levels in the two strains. The lanes are as follow: (1 and 5) Molecular weight markers; (2) Pooled fractions from wash segment; (3) First elution segment; (4) Second elution segment.

required pH of 4.0 (Mitra *et al.*, 1998) and the overall protein yield is usually poorer.

4.2.4 Purification of His_6-transketolase

The histidine-tag allows for simple one-step purification of proteins (Hochuli *et al.*, 1987). Several batches of pure wild-type and mutant transketolase were produced for activity and stability studies. A typical elution profile is depicted in Figure 4.7 where the transketolase elutes as a double peak around 430-450 mM imidazole concentration. Initial experiments showed that transketolase is displaced from the resin at an imidazole concentration of 250 mM. The steep gradient of 10 $mM \cdot ml^{-1}$ and 20 mL column volume explains the occurrence of the elution peak in the 450 mM concentration range. Purification runs with lower gradients of 2.5 $mM \cdot ml^{-1}$ showed that all non-specific proteins elute at an imidazole concentration of 100 mM or less. At the steeper gradient the impurities elute at approximately 300 mM concentration or less, providing adequate resolution between impurities and transketolase. The yields for transketolase range from 10 to 40 mg and varied between purification batches. Using the same amount of starting material, lower yields were observed for D381A mutant than for either the Y440A mutant or wild-type transketolase. Lower dimer stability or mutations elsewhere in the plasmid may have changed the level of functional expression for the D381A mutant.

SDS-PAGE analysis shows the transketolase to be essentially pure (Figure 4.8). Gels were overloaded in order to visualise small quantities of impurities. Eluted protein was dialysed overnight and dried using a centrifugal vacuum evaporator. Dried protein was stable for several months at -20 °C.

4.2.5 Design of dimer destabilised mutants

Two mutations, D381A and Y440A were designed with reduced dimer stability. D318A was chosen as the equivalent mutation (D382A) has been studied in yeast transketolase (Meshalkina *et al.*, 1997) and is involved in interactions at the dimer interface and also potentially in TPP cofactor stabilisation. Y440A was chosen as it is also located at the

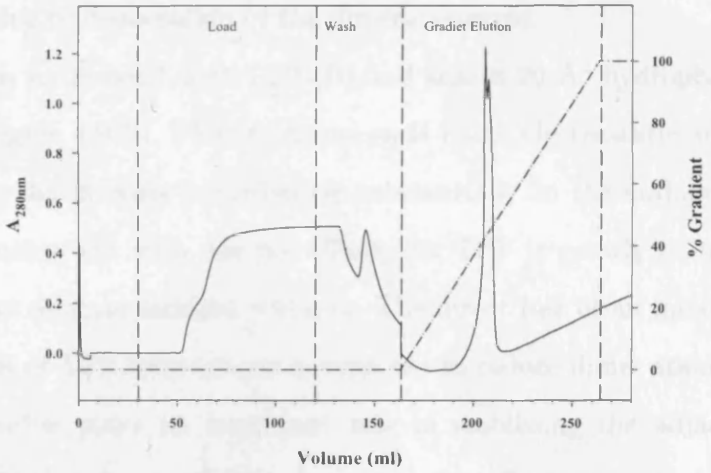


Figure 4.7: Wild-type transketolase was purified using a Pharmacia XK16/20 column packed with 20 ml Ni-NTA resin using an appropriate protocol. All segments of the purification protocol were performed at a flow rate of 1 ml·min^{-1}. Typically 100 ml cell free extract was loaded onto the column followed by a 100 ml wash segment, removing non-binding protein contaminants. During the elution segment transketolase was displaced with an imidazole gradient of 10 mM·min^{-1}. Peak profiles were obtained using UV light detection at 280 nm. The transketolase peaks were captured by fractionation. Cell free extract and wash segment flow-through were retained for SDS-PAGE analysis.

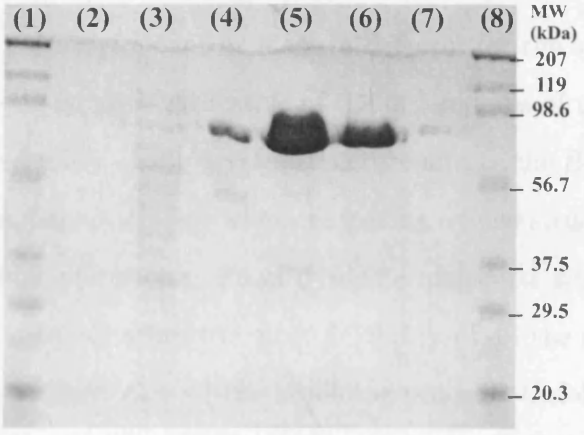


Figure 4.8: SDS-PAGE analysis, using a 12% polyacrylamide gel, of samples retained from each of segment of the purification process for wild-type transketolase. The lanes are as follow: (1 and 8) Molecular weight markers; (3) Flow-through from cell free extract load segment; (4) Flow-through from wash segment; (5) Fraction containing first cut of elution double peak; (6) Fraction containing second cut of the elution double peak.

dimer interface of transketolase, and was expected to primarily alter the dissociation transition and, therefore, enable us to determine which transition in the denaturation of transketolase is due to dissociation of the dimeric enzyme.

Y440(A) forms an H-bond with L116(B) and shares 20 $Å^2$ hydrophobic surface area with TPP(A) (Figure 4.9 B). TPP(A) forms an H-bond, electrostatic and van der Waals interactions with the B chain contributing substantially to the stability of the dimeric form. In substituting Tyr with Ala the affinity for TPP is greatly reduced and leads to a reduction of the cofactor induced stability. The direct loss of an interface H-bond and indirect reduction of TPP interactions is expected to reduce dimer stability of Y440A.

The D381 residue plays an important role in stabilising the adjacent loop region (residues 382-392) that forms a lid to the active site, effectively covering the TPP. The carboxyl group of D381 of chain A (D381(A)) forms two H-bonds with the B chain peptide bond N-atoms of D190(B) and L189(B), reducing the flexibility of the loop region (Figure 4.9 A). Substituting the Asp for an Ala removes these bonds. The equivalent mutation (D382 according to sequence alignment in Appendix A) has been studied in yeast transketolase (Meshalkina *et al.*, 1997). Reduced substrate and cofactor affinities have been determined, leading to an overall reduction in activity. The yeast D382A mutant crystal structure revealed a significant increase in B-factor for the loop region that forms the active site lid. The B-factor is indicative of the orderedness of residues in the crystal structure. Though the quality of the crystal structure affects the B-factor, it can still be used as an indication of entropy change when comparing refined structures that are closely related, such as for point mutations. From dynamic molecular simulations it has been shown that loops and coils, showing the most flexibility of all the secondary structures, are susceptible targets for initiation of the unfolding process. Unfolding of loop and coil regions lead to the fragmentation of the secondary structure units and the subsequent loss of tertiary interactions and packing (Daggett and Levitt, 1993). It has also been speculated that the negative charges of carboxyl groups play a role in neutralising electrostatic forces from TPP and substrate (Nikkola *et al.*, 1994).

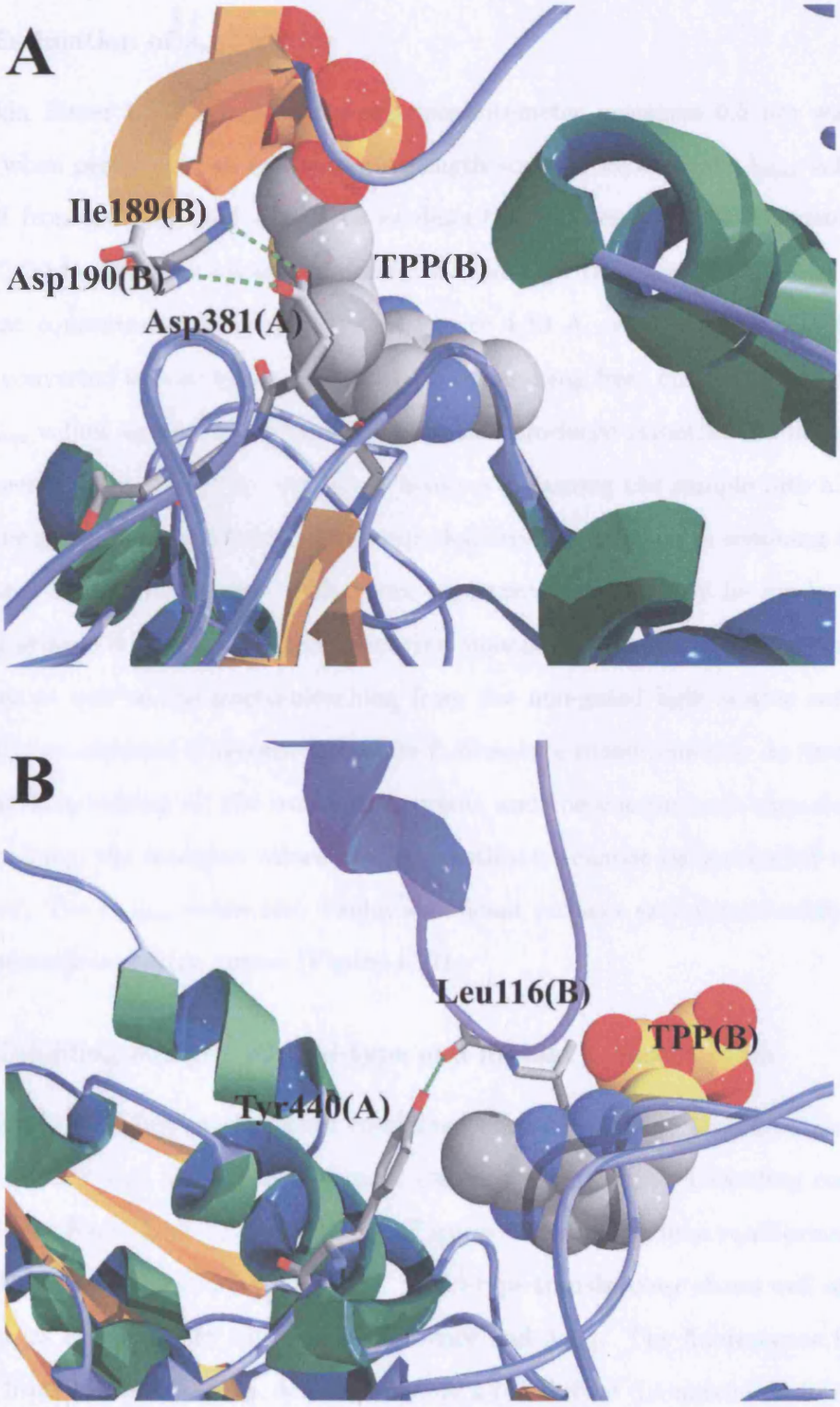


Figure 4.9: Schematic presentation of interactions at the dimer interface removed in mutants. (A) Carboxylic acid group of Asp381(A) forms H-bonds with peptide amide N atoms of Ile189(B) and Asp190(B); (B) Tyr440(A) forms H-bond with peptide carbonyl O atom of Ile116(B). Tyr440(A) also shares van der Waals interactions with B chain TPP cofactor.

4.2.6 Estimation of λ_{max} values

The Perkin Elmer LS30 luminescence spectrophotometer measures 0.5 nm wavelength intervals when performing an emission wavelength scan. More accurate λ_{max} values were estimated from the acquired data sets as described in (Section 2.23) by using Equations 2.12-2.14. The raw emission spectra of wild-type transketolase at four different denaturant concentrations are depicted in Figure 4.10 A, while Figure 4.10 B depicts the data converted to wavenumbers and the lines resulting from curve fits. Plotting estimated λ_{max} values against denaturant concentration produced smoother unfolding curves. Measurements made using the equipment involves siphoning the sample into a flow cell fixed in the path of the excitation light beam, followed by wavelength scanning and data acquisition. Due to instrument limitations, measurements can only be made once the siphoning process is completed. Delays between siphoning, wavelength scanning and data acquisition as well as the photo-bleaching from the non-gated light source contributes significantly to variation of recorded absolute fluorescence measurements. As result of the time delay in acquiring all the wavelength points and the concomitant time-dependant photo-bleaching, the recorded values and λ_{max} estimates cannot be considered to be entirely 'true'. The F_{340nm} values also display significant variance and plotted results do not produce smooth transition curves (Figure 4.12).

4.2.7 Unfolding analysis of wild-type and mutant transketolases

Unfolding and refolding studies were conducted on the wild-type transketolase, in the presence of TPP and Mg^{2+}, using urea as chemical denaturant. Unfolding curves are presented as a function of λ_{max} and F_{340nm} (Figures 4.11 A). The urea equilibrium denaturation using a concentration of 100 $\mu g \cdot ml^{-1}$ wild-type transketolase shows well separated transitions as monitored by intrinsic fluorescence and λ_{max}. The fluorescence intensity increases from 0 M to 4 M urea, accompanied by a red shift of the maximum from 332 to 347 nm.

With both spectroscopic methods (F_{340nm} and λ_{max}) the first transition superimposes

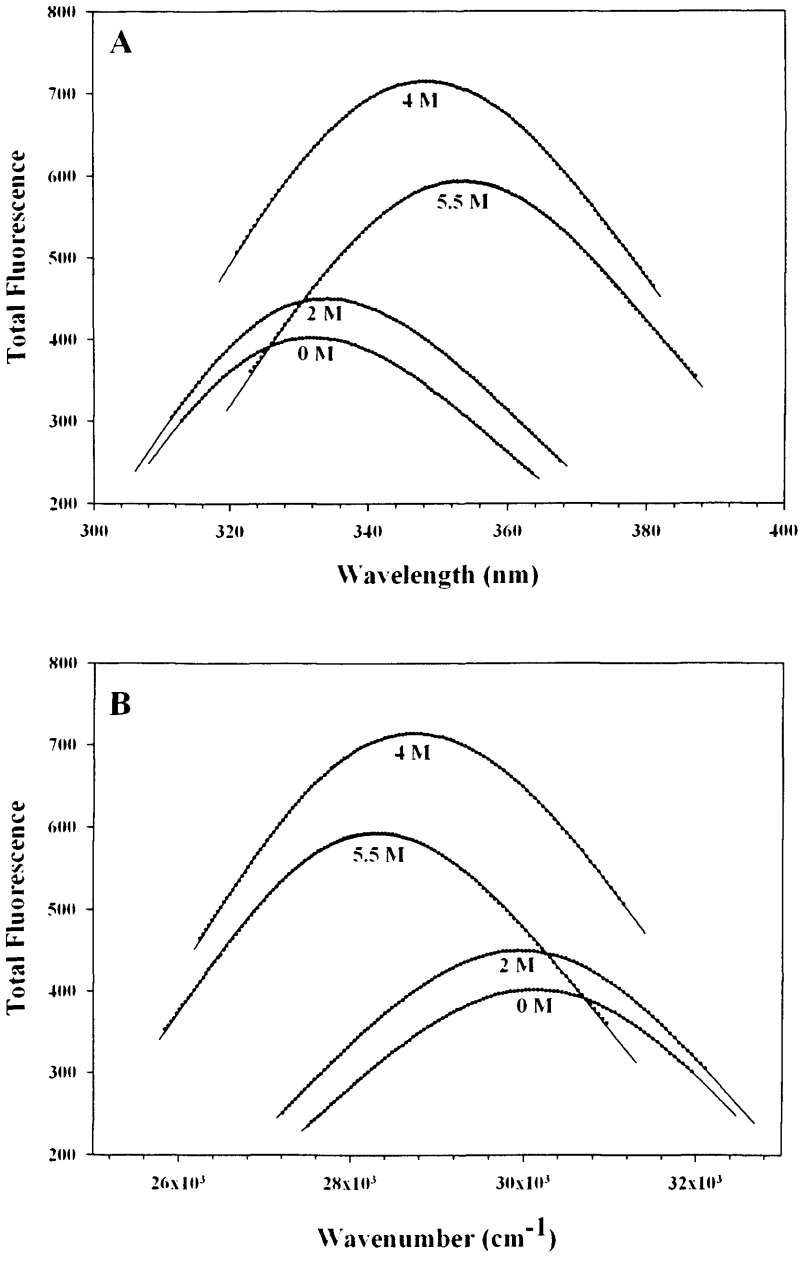


Figure 4.10: (**A**) Raw fluorescence data are plotted as total fluorescence against wavelength (nm). (**B**) Emission spectra are replotted by converting wavelength (nm) to the corresponding wavenumber (cm^{-1}) using Equation 2.12. Replots were fitted to a two-parameter lognormal distribution function (Equation 2.13) to estimate the λ_{max} values. All fitted curves show r^2 values greater than 0.999. The estimated λ_{max} values were obtained by converting υ_m back to the corresponding wavelength using Equation 2.14.

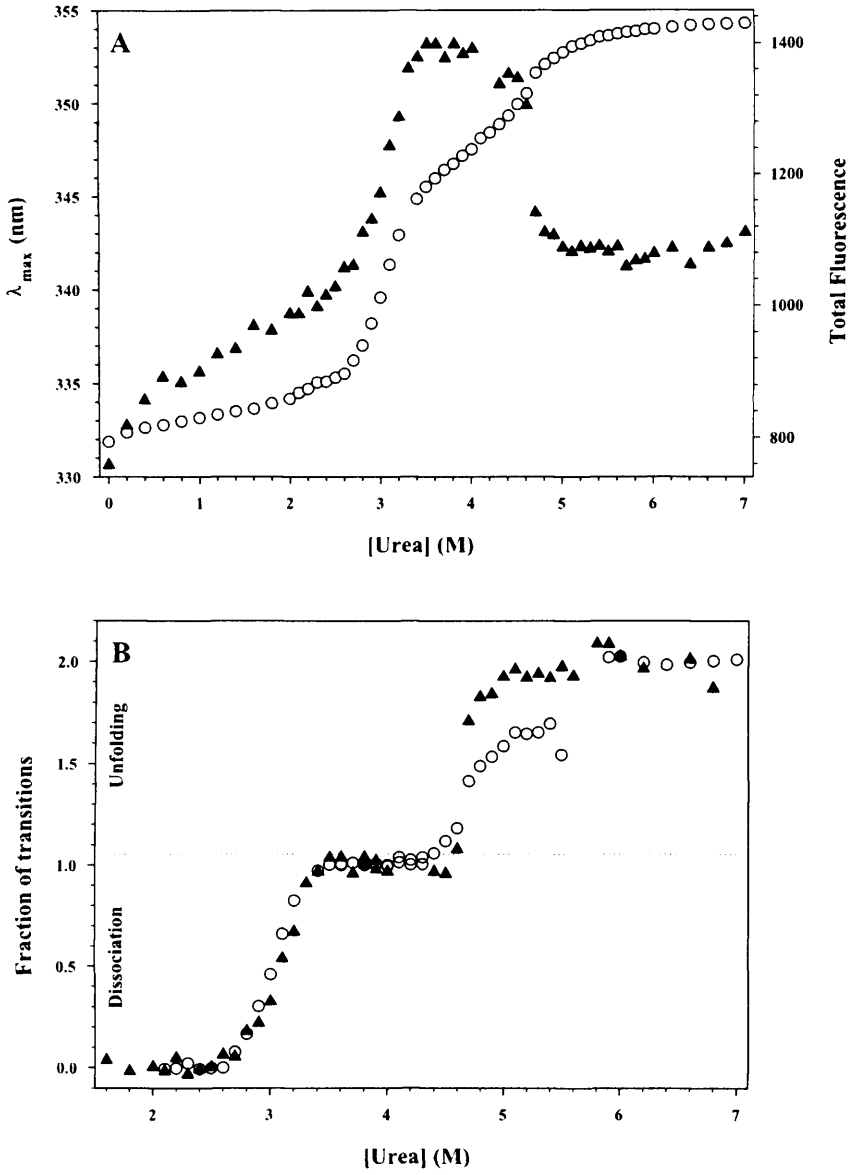


Figure 4.11: (**A**) Urea induced conformational transition curves of wild-type transketolase (▲) F_{340nm} and (○) λ_{max} both display two transition states. (**B**) Unfolding data for both (▲) F_{340nm} and (○) λ_{max} methods are normalised and superimposed. The first transition, normalised between 0 and 1 coincide well. The second transition, normalised between 1 and 2 vary greatly suggesting that multiple overlapping transitions may occur in this region. **Experimental conditions:** Denaturant stock solutions with urea concentrations of 0 M to 7 M urea and 25 mM Tris buffer, pH 7.5 were used for unfolding studies. 200 µl of 1 mg·ml^{-1} pure transketolase in 25 mM Tris buffer, pH 7.5 with 5 mM $MgCl_2$ and 0.5 mM TPP was added to 1.8 ml of each urea stock solution. Samples were incubated for 30 min at reached before emission wavelength spectra were recorded. The λ_{max} values were estimated by fitting tryptophan emission spectral data to a log-normal distribution function (Equation 2.13) and plotted against concentration urea to generate unfolding curve.

well (Figure 4.11 B), suggesting only one transition to be present in this region. The unfolding data of the second transition does not coincide well and a much sharper transition is observed for the F_{340nm} data than for the λ_{max} data. It is likely that two or more overlapping transitions occur within the single observable transition.

The F_{340nm} (Figure 4.12) and λ_{max} (Figure 4.13 A and B) unfolding curves of the dimer destabilised mutants Y440A and D381A were overlaid with the results obtained from wild-type transketolase. For both mutants a significant perturbation is observed from the corresponding first transition of wild-type transketolase. The unfolding curve of the Y440A mutant has two major transitions, similar to the wild-type and display a large shift in $C_{1/2}$ (Δ $C_{1/2}$ = 1.7 M) for the transition in the region between 0 and 4 M urea. The D381A mutant shows two transitions in the same region both with lower $C_{1/2}$ values and less cooperativity than the wild-type transition. This suggests the first transition to be the dissociation transition. In the case of transketolase, as with all obligatory dimeric enzymes, the stability of the dissociation transition has a great effect on the enzyme activity. Destabilisation of the dimer transition will lead to a decrease in specific enzyme activity. Preliminary activity measurements, using the enzyme-linked assay on samples with standardised protein concentrations are summarised in Table 4.3. From the results it is estimated that the D381A mutant is 56-fold less active than the wild-type transketolase. The Y440A mutant is even less active, displaying approximately 700-fold less activity than the wild-type transketolase. The kinetic parameters (K_m, V_{max}) for substrate and cofactor binding (K_d) were not determined.

The observed shifts in the mid-points of the first transition for both dimer destabilised mutants may assist in assigning transitions in wild-type transketolase. For wild-type transketolase, the transition between 3 M and 4 M is likely to be the dissociation of dimer form to monomer form and the second transition, between 4 M and 5 M, the unfolding of the monomer form. The dissociation transition is expected to be protein concentration dependent and unfolding experiments were conducted at different monomer transketolase concentrations, 0.1, 0.02 and 0.005 mg·ml^{-1} in an attempt to confirm the protein concen-

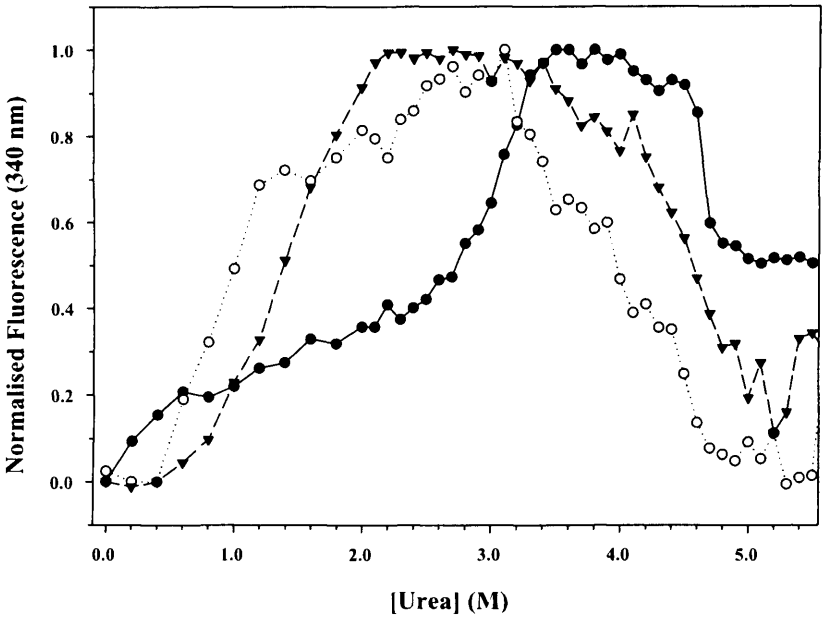


Figure 4.12: Urea induced F_{340nm} conformational transition curves of (•) wild-type, (○) D381A and (▼) Y440A mutant transketolases. Two conformational transitions can be observed for wild-type and Y440A unfolding curves, while at least three transitions are visible for D381A unfolding curves. Non-linear analysis for $C_{1/2}$ estimation of transitions are depicted in Figures 4.17 and 4.18. **Experimental conditions:** Denaturant stock solutions with urea concentrations of 0 M to 7 M urea and 25 mM Tris buffer, pH 7.5 were used for unfolding studies. 200 µl of 1 mg·ml^{-1} pure transketolase in 25 mM Tris buffer, pH 7.5 with 5 mM $MgCl_2$ and 0.5 mM TPP was added to 1.8 ml of each urea stock solution. Samples were incubated for 30 min at ambient temperature allowing the unfolding equilibrium to be reached before emission wavelength spectra were recorded. The F_{340nm} measurements were extracted from the results table and plotted.

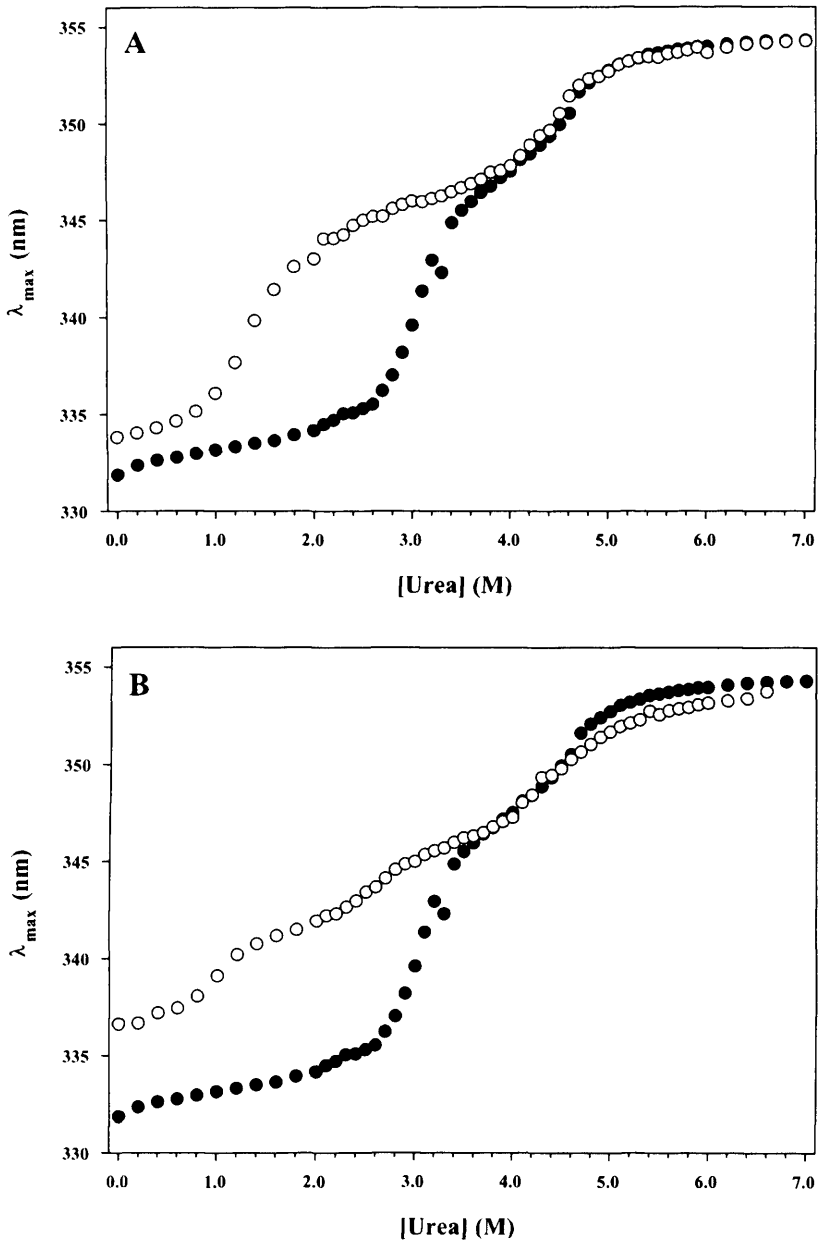


Figure 4.13: (**A**) Superimposed urea induced conformational transition curves of (•) wild-type and (○) Y440A mutant transketolases. (**B**) Superimposed urea induced conformational transition curves of (•) wild-type and (○) D381A mutant transketolases. When compared to the 1^{st} transition of wild-type transketolase, both mutants display transitions perturbed to lower stability. **Experimental conditions:** Denaturant stock solutions with urea concentrations of 0 M to 7 M urea and 25 mM Tris buffer, pH 7.5 were used for unfolding studies. 200 µl of 1 mg·ml^{-1} pure transketolase in 25 mM Tris buffer, pH 7.5 with 5 mM $MgCl_2$ and 0.5 mM TPP was added to 1.8 ml of each urea stock solution. Samples were incubated for 30 min at ambient temperature allowing the unfolding equilibrium to be reached before emission wavelength spectra were recorded. The λ_{max} values were estimated by fitting tryptophan emission spectral data to a log-normal distribution function (Equation 2.13) and plotted against concentration urea to generate unfolding curve.

Table 4.3: Summary of the enzyme and specific activities estimated for wild-type and mutant transketolases.

	Enzyme activity			Enzyme concentration				
	δA_{340}	Time[a] (min)	Activity[b] (U)	A_{280}	Volume[c] (µl)	Enzyme[d] (µg)	Specific activity ($U \cdot \mu g^{-1}$)	Fraction of Wild-type activity
Wild-type	0.106	2	8.5×10^{-3}	0.031	40	0.95[e]	8.92×10^{-3}	1
D381A	0.277	60	7.4×10^{-4}	0.060	100	4.61[f]	1.61×10^{-4}	0.018
Y440A	0.247	300	1.3×10^{-4}	0.336	40	10.50[g]	1.26×10^{-5}	0.0014

[a] Assay time of enzyme reactions increase with mutants that display low activity.
[b] Enzyme units (U) are calculated using Equation 2.11. 1 U activity = NADH consumption of 1 $\mu mol \cdot min^{-1}$.
[c] Volume of enzyme (µl) added to enzyme-linked assay mixture (Table 2.4).
[d] Total enzyme determined using $A = \epsilon cl$. Enzyme molar extinction coefficients (ϵ) calculated according to the method of Pace *et al.* (1995).
[e] Total Wild-type estimated with $\epsilon = 93905\ M^{-1} \cdot cm^{-1}$ and MW = 72260.82 $g \cdot mol^{-1}$
[f] Total D381A estimated with $\epsilon = 93905\ M^{-1} \cdot cm^{-1}$ and MW = 72216.81 $g \cdot mol^{-1}$
[g] Total Y440A estimated with $\epsilon = 92415\ M^{-1} \cdot cm^{-1}$ and MW = 72168.72 $g \cdot mol^{-1}$

tration dependence of the dissociation transition. However, no transition shift was observed (Figure 4.14 A and B). It is well documented that enzyme activity is still observed at the reported concentration and also for much lower protein concentrations, confirming the enzyme to be in the dimeric form. From sedimentation studies, using Baker's yeast transketolase, Cavaliere *et al.* (1975) found that the apo-transketolase dissociates at a concentration of approximately 0.1 $mg{\cdot}ml^{-1}$. However no dissociation was observed in the presence of TPP, even at lower protein concentrations (within experimental capabilities). It is possible that in the presence of TPP, the equilibrium of the association reaction may have shifted much further to the right and the protein concentrations used were not sufficiently low to observe a significant change in the reaction equilibrium. The lowest protein concentration used in this study was at the analytical limit of the equipment. A similar experiment with wild-type apo-transketolase or the destabilised Y440A mutant may reveal a shift in the equilibrium at lower protein concentrations.

The buffer conditions selected for unfolding are similar to those reported for a typical biotransformation when using transketolase (Mitra *et al.*, 1998). Under these conditions both the dissociation and reconstitution steps of forming dimeric holo-transketolase from monomeric apo-transketolase have been found to be reversible (Egan and Sable, 1981; Cavaliere *et al.*, 1975; Sprenger *et al.*, 1995). A refolding study was conducted to determine whether the first transition is reversible under the experimental conditions (Section 2.22). Transketolase was prepared in 3.8 M urea and diluted to various final urea concentrations using buffer containing 0 and 3.8 M urea. Under these experimental conditions the unfolding of the first transition was not reversible, results depicted in Figure 4.15. This suggests that the first transition may not solely be a dimer dissociation. Increased flexibility of the three domains of the monomer, relative to each other, may occur upon dimer dissociation in the presence of denaturant and result in a stable, partially unfolded intermediate. In such a case denaturant removal is more likely to lead to misfolding or aggregation. Partially unfolded proteins are also more susceptible to chemical inactivation. Based on the available evidence a three-state transition is initially proposed for

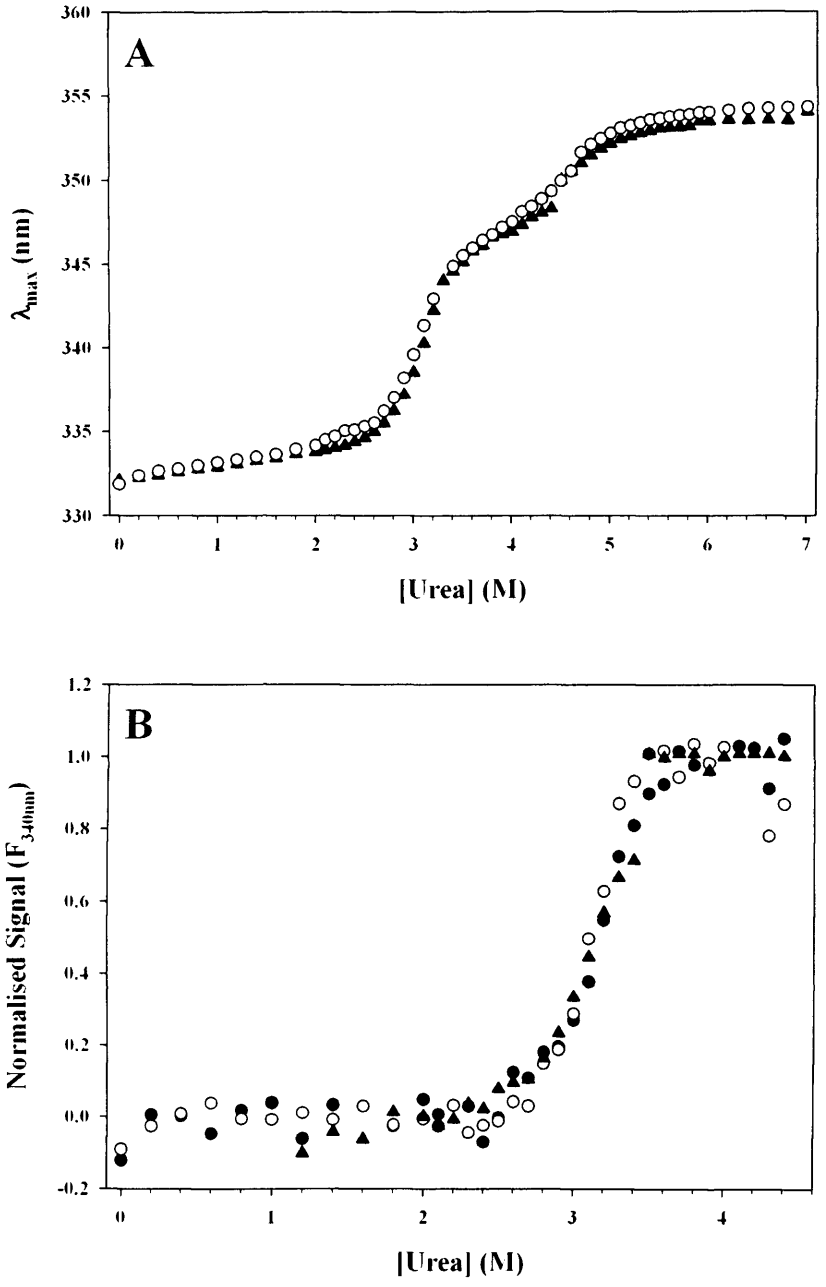


Figure 4.14: The dissociation transition is not dependent on the monomer transketolase concentration in the range of 0.1-0.005 mg·ml^{-1} (**A**) The superimposed λ_{max} conformational transition curves of *E. coli* wild-type transketolase at (○) 0.1 and (▲) 0.02 mg·ml^{-1} monomer protein respectively. (**B**) The superimposed F_{340nm} conformational transition curves of *E. coli* wild-type transketolase at (●) 0.1, (▲) 0.02 and (○) 0.005 mg·ml^{-1} monomer protein respectively.

wild-type transketolase where the dimer (N_2) dissociates to a stable monomer intermediate (I) and then proceed to the unfolded state (U) as shown in Equation 4.1. Since the first transition is not reversible the complete unfolding reaction is irreversible under biotransformation reaction conditions. In the case of irreversible unfolding, exact stability data cannot be obtained for wild-type and mutant transketolases using equilibrium unfolding experimentation. The respective $C_{1/2}$ values, however, when measured at a fixed pre-incubation period, may still serve as an indication of relative protein stability or tolerance to thermoinactivation, providing no large changes in m_G values occur. The changes in $C_{1/2}$ ($\Delta C_{1/2}$) will be used as the basis for evaluating changes in stability of transketolase mutants in high-throughput unfolding studies.

$$N_2 \rightarrow 2I \rightarrow 2U \tag{4.1}$$

Transition midpoints ($C_{1/2}$) of the λ_{max} and F_{340nm} unfolding curves were determined with non-linear fitting. For the first transition, ($N_2 \rightarrow$2I) a protein concentration dependent $C_{1/2}$ value was estimated using Equation 2.19; Section 2.24. The second transition (I→U) was estimated using Equation 2.3; Section 2.6. All the obtained results are summarised in Table 4.4. Wild-type transketolase displays two major transitions for both the λ_{max} and F_{340nm} unfolding curves. $C_{1/2}$ values estimated when using the λ_{max} data set, were 3.2 ±0.1 M and 4.9 ±0.1 M urea, while midpoints obtained from the F_{340nm} data set were 3.1 ±0.1 M and 4.7 ±0.1 M urea. Both λ_{max} and F_{340nm} unfolding curves of the D381A mutant display two transitions in the region between 0 and 4 M urea. Though reasonable curve fits can be made to determine the individual $C_{1/2}$ values, uncertainty exists over the correct equation to use since it is not clear whether only one or both transitions are involve in the dissociation step. From Figure 4.13B it seems that the 2^{nd} transition of the D381A mutant and the 1^{st} transition of the wild-type have the same endpoint (I). Therefor the combined 1^{st} and 2^{nd} transitions of the D381A mutant were fitted to Equation 2.19 to obtain an apparent $C_{1/2}$ value of 1.5 ±0.1 M and 1.2 ±0.1 M for λ_{max} and F_{340nm} data respectively. The third transition of the D381A mutant was

Table 4.4: Summary of the unfolding midpoint values estimated for wild-type and mutant transketolases. Values were determined using non-linear parameter estimation. Values in brackets represent the standard errors obtained from curve fittings.

Transition	$C_{1/2}$ (M)
λ_{max} method	
wt-TK ($N_2 \rightarrow 2I$)	3.2 (0.1)
wt-TK ($I \rightarrow U$)	4.9 (0.1)
Y440A ($N_2 \rightarrow 2I$)	1.5 (0.1)
Y440A ($I \rightarrow U$)	4.6 (0.1)
D381A (apparent $N_2 \rightarrow 2I$)	1.5 (0.1)
D381A ($I \rightarrow U$)	4.7 (0.1)
F_{340nm} method	
wt-TK ($N_2 \rightarrow 2I$)	3.1 (0.1)
wt-TK ($I \rightarrow U$)	4.7 (0.1)
Y440A ($N_2 \rightarrow 2I$)	1.4 (0.1)
Y440A (apparent $I \rightarrow U$)	4.4 (0.1)
D381A (apparent $N_2 \rightarrow 2I$)	1.2 (0.1)
D381A (apparent $I \rightarrow U$)	4.0 (0.1)

fitted as the unfolding transition (I→U) using the λ_{max} data set and Equation 2.3 to give a $C_{1/2}$ estimation of 4.7 ±0.1 M. For the Y440A mutant $C_{1/2}$ values were estimated to be 1.3 ±0.1 M and 4.6 ±0.1 M using the λ_{max} data set. For the F_{340nm} data set the I→U transition of both mutant transketolases display two or more partially populated intermediates. The transitions were fitted to an apparent two-state unfolding profile to obtain approximate values of $C_{1/2}$. The estimated apparent $C_{1/2}$ values obtained by more 'classical' means in this chapter provide a measure for comparing the F_{340nm} conformational transition curves (Figures 4.12-4.18) with high-throughput results, for method validation purposes. Unfolding results for transketolase, obtained in a high-throughput format by recording fluorescence emission at 340 nm, are described in Chapter 5 (Section 5.2.6).

4.2.8 Native PAGE analysis

Native PAGE analysis was explored as a tool for characterising the quaternary structure of purified transketolase. Theoretically, dimeric and monomeric forms can be separated based on size in gel electrophoresis. Samples were analysed under increasing urea concen-

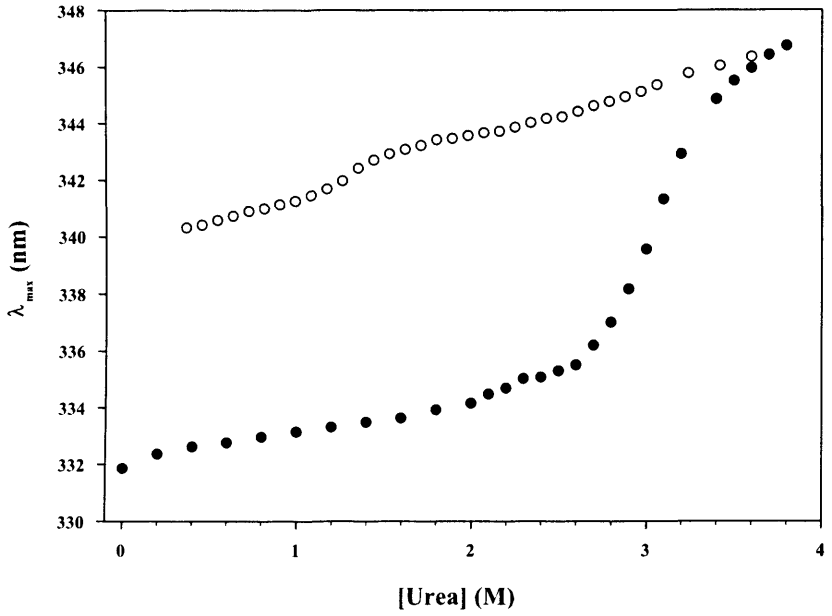


Figure 4.15: A refolding study was conducted to determine whether the first transition is reversible under selected process conditions. By comparing the (•) unfolding curve with the (○) refolding curve it is evident that the unfolding process is not reversible. **Experimental conditions:** Denaturant stock solutions with urea concentrations of 0 M to 3.8 M urea and 25 mM Tris buffer, 3.8M urea, pH 7.5 were used for the refolding study. 200 µl of 1 mg·ml^{-1} pure transketolase in 25 mM Tris buffer, 3.8 M urea, pH 7.5 with 5 mM $MgCl_2$ and 0.5 mM TPP was added to 1.8 ml of each urea stock solution. Samples were incubated for 60 min at ambient temperature allowing the refolding equilibrium to be reached before emission wavelength spectra were recorded. The λ_{max} values were estimated by fitting tryptophan emission spectral data to a log-normal distribution function (Equation 2.13) and plotted against concentration urea to generate unfolding curve.

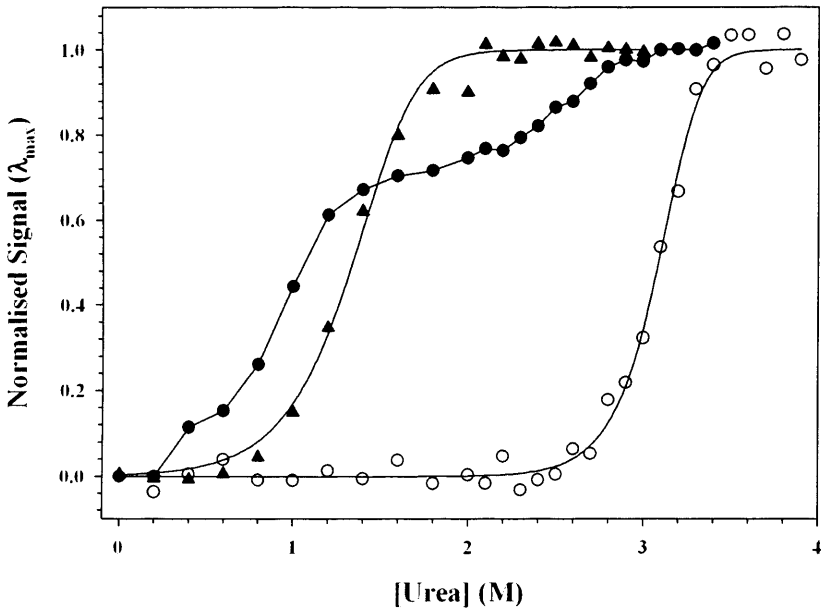


Figure 4.16: The $N_2 \rightarrow 2I$ conformational transitions of the λ_{max} urea induced unfolding data were normalised in order to perform non-linear estimation of the $C_{1/2}$ values using Equation 2.19. The $C_{1/2}$ values were only estimated for the (○) wild-type and (▲) Y440A mutant transketolases which display single dissociation transitions as well as an apparent $C_{1/2}$ value for the two transitions of the (●) D381A mutant. **Experimental conditions:** Denaturant stock solutions with urea concentrations of 0 M to 7 M urea and 25 mM Tris buffer, pH 7.5 were used for unfolding studies. 200 µl of 1 mg·ml^{-1} pure transketolase in 25 mM Tris buffer, pH 7.5 with 5 mM $MgCl_2$ and 0.5 mM TPP was added to 1.8 ml of each urea stock solution. Samples were incubated for 30 min at ambient temperature allowing the unfolding equilibrium to be reached before emission wavelength spectra were recorded. The F_{340nm} measurements were extracted from the results table and plotted.

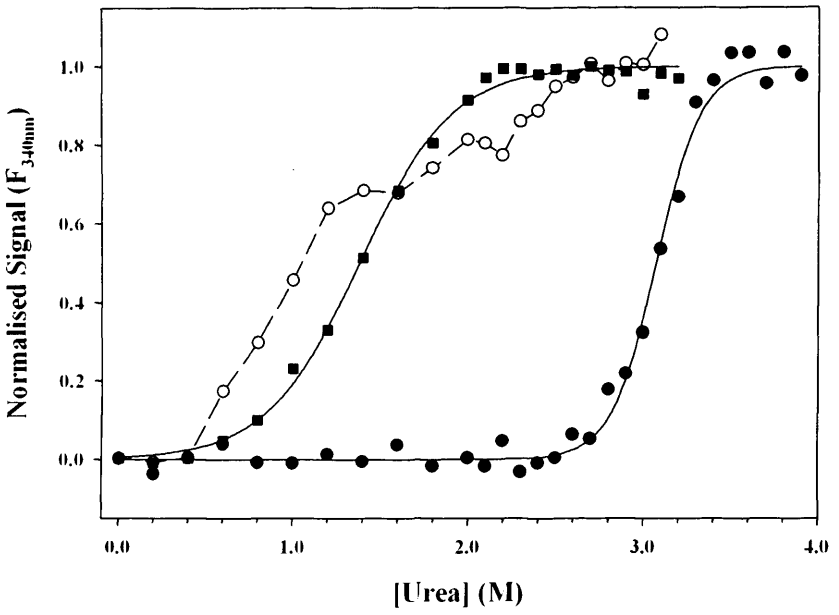


Figure 4.17: The $N_2 \rightarrow 2I$ conformational transitions of the F_{340nm} urea induced unfolding data were normalised for (•) wild-type, (○) D381A and (■) Y440A mutant transketolases. A single conformational transition is observed for wild-type and Y440A unfolding curves and non-linear analysis was used to estimate the $C_{1/2}$ values as 3.1 ±0.1 M and 1.4 ±0.1 M urea respectively. At least two conformational transitions are observed for D381A. For lack of clear transitions, non-linear analysis was used to estimate an apparent two-state $C_{1/2}$ = 1.2 ±0.1 M urea. The resulting curve-fit shows poor correlation to the actual data and is not shown. **Experimental conditions:** Denaturant stock solutions with urea concentrations of 0 M to 7 M urea and 25 mM Tris buffer, pH 7.5 were used for unfolding studies. 200 µl of 1 mg·ml^{-1} pure transketolase in 25 mM Tris buffer, pH 7.5 with 5 mM $MgCl_2$ and 0.5 mM TPP was added to 1.8 ml of each urea stock solution. Samples were incubated for 30 min at ambient temperature allowing the unfolding equilibrium to be reached before emission wavelength spectra were recorded. The F_{340nm} measurements were extracted from the results table and treated as explained in Section 2.6.

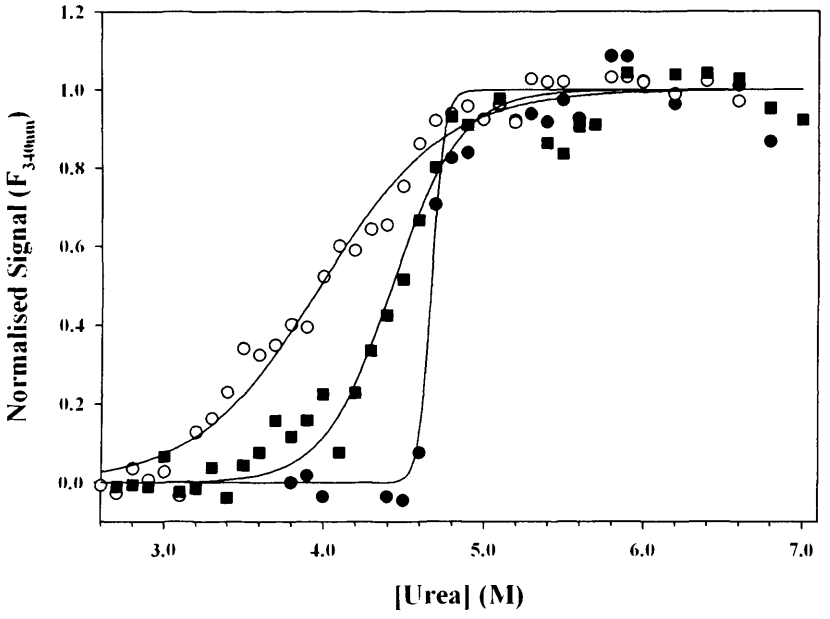


Figure 4.18: Urea induced I → U transition curves of (•) wild-type, (○) D381A and (■) Y440A mutant transketolases. According to the F_{340nm} unfolding curves, both mutants have lower monomer stability than the wild-type transketolase. Non-linear analysis was used to estimate the $C_{1/2}$ values of wild-type, Y440A and D381A as 4.7 ±0.1 M, 4.4 ±0.1 M and 4.0 ±0.1 M urea respectively. **Experimental conditions:** Denaturant stock solutions with urea concentrations of 0 M to 7 M urea and 25 mM Tris buffer, pH 7.5 were used for unfolding studies. 200 µl of 1 mg·ml^{-1} pure transketolase in 25 mM Tris buffer, pH 7.5 with 5 mM $MgCl_2$ and 0.5 mM TPP was added to 1.8 ml of each urea stock solution. Samples were incubated for 30 min at ambient temperature allowing the unfolding equilibrium to be reached before emission wavelength spectra were recorded. The F_{340nm} measurements were extracted from the results table and treated as explained in Section 2.6.

trations in an attempt to correlate band shifts on the gel with unfolding curve transitions, thus pinpointing the $N_2 \rightarrow 2I$ transition. All the native transketolase samples (wild-type and mutants) display three distinct bands (Figure 4.19). Upon addition of TPP and Mg^{2+}, bands 1 and 2 become less visible and band 3 more intense (Figure 4.19 A and B). This suggests that band 3 might be the holo-transketolase form, migrating faster because of the additional negative charges obtained from the pyrophosphate. The holo-transketolase band becomes less intense and bands 1 and 2 more visible when 4 M or 8 M urea is present. Bands 1 and 2 are most likely populations of dimeric and monomeric transketolase without any cofactor. Another band (not labelled) appears under denaturing conditions in the bottom of the lane well, possibly due to irreversible aggregation of transketolase caused by dilution of urea upon sample loading.

The relative activities, compared to the reconstituted transketolase prior to chromatography, are shown in each lane (Figure 4.19 C). The presence of band 3 however does not correlate well with activity measurements in experiments where TPP was resolved using size-exclusion chromatography (SEC). The species in bands 1 and 2 become slightly more abundant after SEC. The D381A mutant produces an additional band (4) which is possibly monomeric transketolase-TPP complex (Figure 4.19 D). At 4 M urea the wild-type transketolase seems to be more resistant against monomerisation and irreversible precipitation than either of the mutants (Figure 4.19 A,D and E). It was observed that the yeast apo-transketolase form still exists in the dimeric state after TPP was resolved based of SEC data (Saitou *et al.*, 1974). Addition of TPP and Mg^{2+} do not seem to have a marked effect on the band distribution in both mutants (Figure 4.19 D and E).

4.2.9 The transketolase dimer interface

All the residues involved in the formation of the dimer interface were determined based on changes in the accessible surface areas (ΔASA). Atomic and residue level ASA contributions were determined with an implementation of the algorithm of (Lee and Richards, 1971) using AreaIMol in the CCP4 program suite (Collaborative Computational Project,

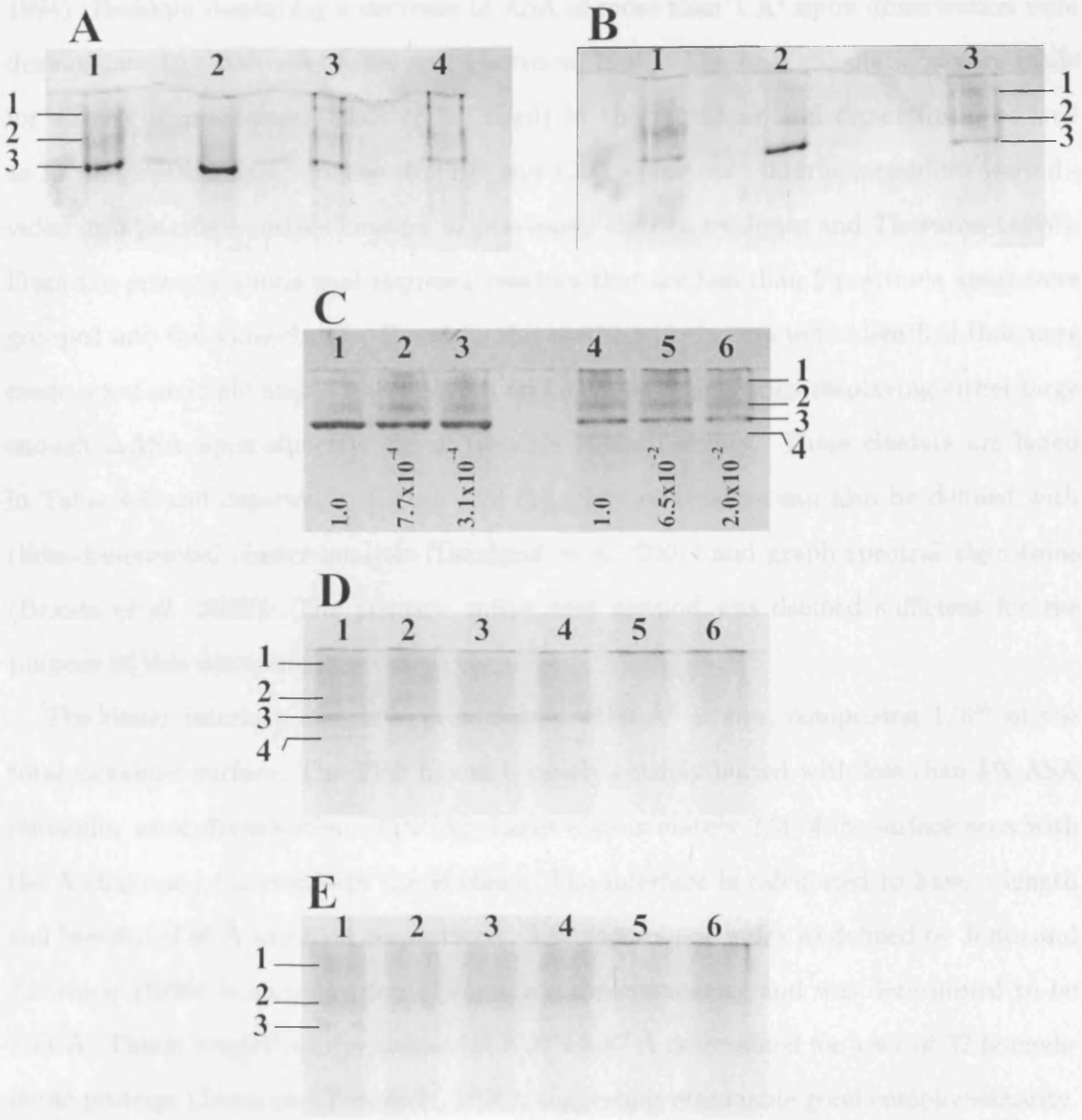


Figure 4.19: Native-PAGE analysis, using a 12% acrylamide gel of wild-type and mutant transketolase in absence and presence of urea. (**A**) (1) Native wild-type transketolase; (2) wild-type transketolase & TPP; (3) wild-type transketolase & 8 M urea; (4) wild-type transketolase, TPP & 8 M urea. (**B**) (1) Native wild-type transketolase; (2) wild-type transketolase & TPP; (3) wild-type transketolase and 4 M urea. (**C**) (1) Native wild-type transketolase, TPP & Mg; (2) wild-type transketolase after Ni-NTA purification; (3) wild-type transketolase after SEC; (4) D381A transketolase, TPP & Mg; (5) D381A transketolase after Ni-NTA purification; (6) D381A transketolase after SEC. Values in each lane states the relative activity after each chromatographic step compared to the holo-transketolase form. (**D**) (1) D381A transketolase & no cofactors; (2) D381A transketolase, TPP & Mg; (3) D381A transketolase and 1 M urea; (4) D381A transketolase and 2 M urea; (5) D381A transketolase and 4 M urea; (6) D381A transketolase and 6 M urea. (**E**) (1) Y440A transketolase & no cofactors; (2) Y440A transketolase, TPP & Mg; (3) Y440A transketolase & 1 M urea; (4) Y440A transketolase & 2 M urea; (5) Y440A transketolase & 4 M urea; (6) Y440A transketolase & 6 M urea.

1994). Residues displaying a decrease in ASA of more than 1 Å^2 upon dimerisation were deemed interface residues (Jones and Thornton, 1995). The ASA calculations were made for *E. coli* transketolase (PDB code: 1qgd) in the monomer and dimer form, as well as in the presence and absence of TPP and Ca^{2+} cofactors. Interface residues were divided into interface surface clusters as previously defined by Jones and Thornton (1995). From the primary amino acid sequence residues that are less than 5 positions apart were grouped into the same cluster. Based on this method 12 clusters were identified that were made up of multiple amino acid residues and 3 single amino acids displaying either large enough ΔASA upon dimerisation or possible H-bond ability. These clusters are listed in Table 4.5 and depicted in Figure 4.20 (1). Surface clusters can also be defined with three-dimensional cluster analysis (Landgraf *et al.*, 2001) and graph spectral algorithms (Brinda *et al.*, 2002). The primary amino acid method was deemed sufficient for the purpose of this discussion.

The dimer interface area is approximately 4100 Å^2 is size, comprising $1/6^{th}$ of the total monomer surface. The TPP ligand is nearly entirely buried with less than 5% ASA remaining after dimerisation. TPP(A) shares approximately 2/3 of its surface area with the A chain and the rest with the B chain. The interface is calculated to have a length and breadth of 65 Å and 55 Å respectively. The gap volume index as defined by Jones and Thornton (1995) is an indication of surface complementarity and was determined to be 1.71 Å. This is lower than the average of 2.20 $\pm$0.87 Å determined for a set of 32 homodimeric proteins (Jones and Thornton, 1996), suggesting reasonable good complementarity. There still is potential for improvement with examples of homodimeric proteins having a gap volume index of only 0.60 Å. The ratio of polar to non-polar atoms at the interface is 39% to 61% as calculated by the protein-protein interaction server (Jones and Thornton, 1996). The interface hydrophobic residues are also depicted in Figure 4.20 (2). The planarity of the interface was calculated as per method and software of Laskowski (1995) and found to have a 6.80 rms deviation from a plane. A high rms value indicates the surface area deviates substantially from a plane fitted through the subunit interface. The

transketolase monomer interface forms two planes crossing perpendicular and curving in opposite directions (Figure 4.21).

4.2.10 Analysis of interface clusters

A multiple sequence alignment (MSA) with 50 transketolase protein sequences was created using ClustalW (Thompson *et al.*, 1994). The generated MSA and phylogenetic tree served as input into the ConSurf algorithm (Armon *et al.*, 2001), used to calculate the estimated substitution rate (ESR) for each residue based on the mutation frequency and evolutionary distances covered. The ESR values of interface residues are depicted in Figure 4.20 (3) ranked as five groups ranging from highly (red) to least conserved (blue).

Transketolase is only active in the dimeric form with both subunits contributing to the formation of TPP cofactor and substrate binding sites. Care must be taken, since there is a potential overlap with surface residues that contribute to both ligand binding and interface formation. Functionally important residues (catalytic and ligand binding) and structural important residues (folding and dimerisation) are conserved (Valdar and Thornton, 2001). The ESR value of an interface cluster is calculated as the average of all the residue ESR values comprising the cluster and listed in Table 4.5. Clusters can be divided into two groups, namely highly conserved clusters (1,3,6-9) and moderate to poorly conserved clusters (2,4,5,10-15). A few of the highly conserved clusters are situated close to either the TPP binding sites (clusters 3,6,7 and 8) or substrate entrance (clusters 1 and 9). Only two conserved clusters form large diagonal bands across the dimer interface, suggesting that these two clusters are conserved for protein-protein interaction purposes. Negative ESR values signify conserved residue positions and positive values indicate residues positions with high substitution rates.

Comparative analysis suggests that proteins adapt through mutations to an extent where they can fulfill their function at the physiological conditions they occur in. While maintaining structure, flexibility, ligand binding and activity, proteins only have marginal stability for optimal half-life turnover allowing sensitivity for homeostasis (Jaenicke, 1991).

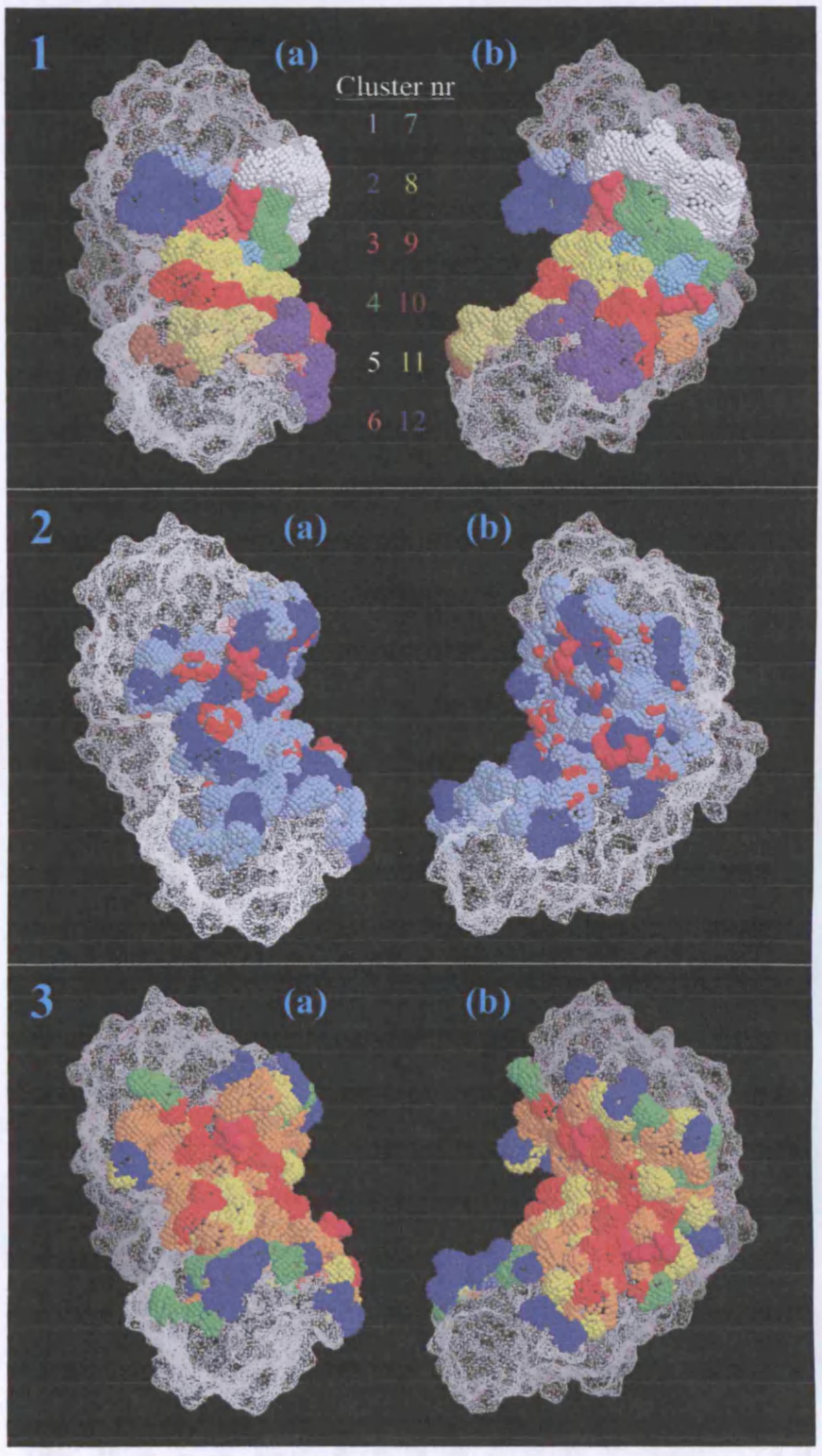


Figure 4.20: A schematic presentation of interface residues of *E. coli* transketolase. (**1**) Residues are divided into clusters based on primary amino acid and coloured according to cluster number. (**2**) Distribution of H-bonds (red) and hydrophobic residues (dark blue) across the interface area. Distinct hot spots can be observed for H-bond formation. (**3**) Surface residues are coloured based on conservation rate. Red represents a highly conserved, green moderately conserved and blue the least conserved residues positions.

Table 4.5: Summary of results for analysis of transketolase interface residue clusters.

Cluster	Amino acid sequence	ΔASA	Average ESR	Hydrogen bonds
1	A22-S24-H26	38.2	-0.71175	0
2	R91-Q92-L93-H94-P98-G99-H100-E102-K105	392.9	-0.378	4
3	T112-T113-G114-P115-L116-Q118	95.1	-1.04267	1
4	G156-M158-M159-E160-G161-I162-H164-E165-S168-L169-G171-T172-L173	436.1	-0.311	6
5	G194-G195-W196-F197-T198-D199-D200-M203-R204-E206-A207-Y208	714.9	0.01973	9
6[a]	S379-**D381**-L382-A383-P384-S385	164.7	-0.68067	3
7	G408-V409-R410-E411-F412	264.7	-0.84114	4
8[a]	M436-F437-E439-**Y440**-R442-N443-R446-M447-A449-L450	573.5	-0.6974	6
9	E468-D469-P471-T472-H473-Q474-V476-E477-Q478-A480-S481-R483-V484-T485-P486-N487	637.9	-0.83756	5
10	K603-T606-R608	41.2	0.03033	0
11	I615-D617-Y618-Y620-K621-Y622-V623-G624-L625-N626	411.6	0.8701	2
12	T633-T634-F635-E637-S638-A639-P640-L643-L644-E647	370.9	0.2075	2
13	H258	8.4	0.741	0
14	N397	8	0.822	0
15	D590	0	-0.279	(1)[b]

[a] Positions of mutations D381A and Y440A are highlited in clusters 6 and 8 respectively.
[b] Value in brackets lists a H-bond interaction that may possibly form in solution with more conformational flexibility.

Since structural and sequence homology between between psychrophilic, mesophilic and thermophilic only highlight the conserved interface residues, the added or reduced stability may be contributed by appropriate adaption of less conserved residues. From several studies it is shown that these thermophilic improvements is not the result of one single influence but rather a combination of stabilising effects (Shaw and Bott, 1996). In this regard studying homodimeric proteins in psychrophilic and thermophilic enzymes may prove useful. The calculated average ESR values for surface clusters may be an indication of the overall structural and functional importance. It also highlights regions that can be targeted for improving dimer stability without disrupting function.

Under its natural selection pressure *E. coli* transketolase is evolutionarily optimised to be marginally stable and suitably active at moderate temperatures. In the case of transketolase the few clusters with dual functionality (ligand binding and interface interaction) such as residue Y440 may be sufficient to provide adequate stability.

4.2.11 Improving homodimer stability with directed evolution

Improving stability of the dimer interface in itself is a difficult task. Increases in ASA and hydrophobic interaction as well as improvements of geometric and electrostatic complementarity between surfaces must be considered. It has been shown, particularly in systems with large surface interfaces, that only a small number of residues contribute to the majority of the binding energy (Clackson and Wells, 1995). These hot spots tend to be conserved both in residue type and position and display a high degree of burial at the interface (Bogan and Thorn, 1998). From Figure 4.20 (2) and Table 4.5 it is evident where these hot spots of H-bond and hydrophobic interactions are located. Further more it has been shown that interface residues are more conserved than random surface residues in most protein families (Valdar and Thornton, 2001) with a few exceptions (Caffrey *et al.*, 2004). From multiple sequence alignments it was observed that some interface contacts range from highly to very poorly conserved. Highly conserved residues may indicate an important interface stabilising residue. It may also highlight residues involved in ligand

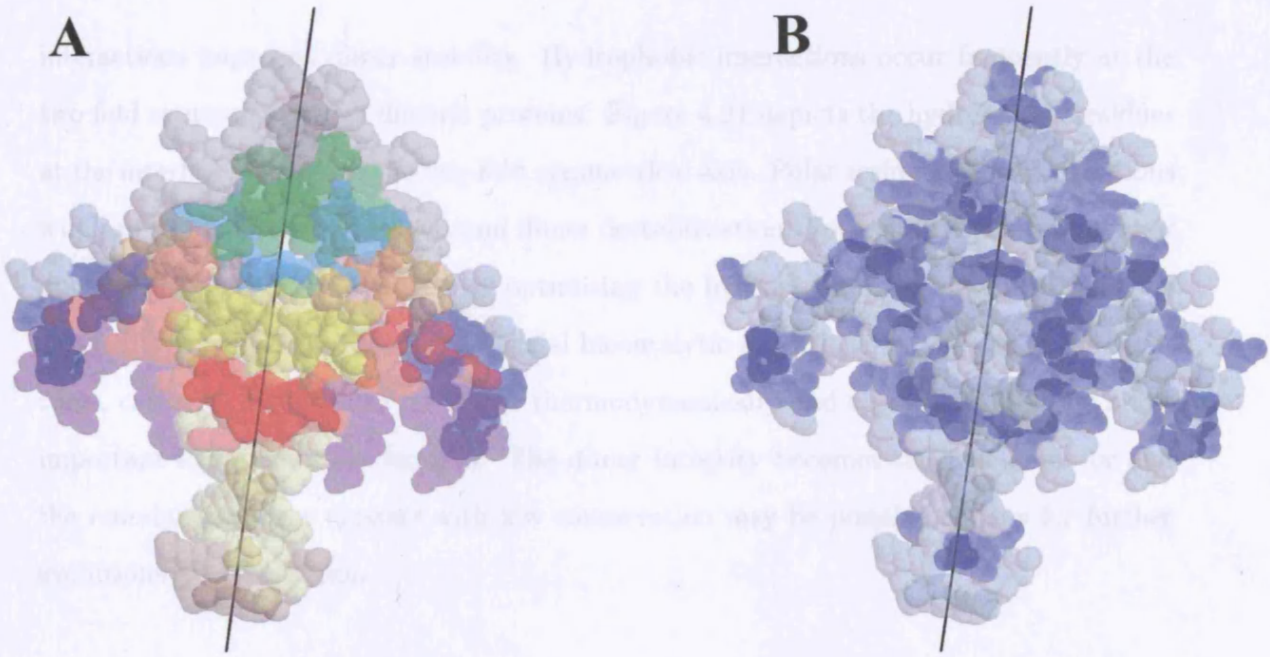


Figure 4.21: Subunit interface surfaces are superimposed to reveal residues located on the two-fold symmetry axis. (A) Colour coded surface clusters are superimposed and overlapping clusters with same colour that visualise darker make up the twofold symmetry, represented by the black line. These residues can be targeted for protein engineering, particularly to improve (B) hydrophobic interactions and create more stable homodimers.

binding or formation of the active site in the interface area in the case of obligatory dimeric proteins. Poor conservation scoring of residues may be the result of large variation or gaps in the multiple sequence alignment at that position (Valdar and Thornton, 2001).

The key to designing directed evolution strategies for improving homodimer interaction may lie in targeting less conserved interface residues. By targeting a smaller population of residues using a mutational strategy such as saturation site-directed mutagenesis (SSDM), smarter libraries can be created with a more comprehensive sequence space coverage of the targeted areas. Hydrophobic interactions are speculated to be the major driving force behind protein folding, albeit via the hydrophobic collapse or nucleation condensation pathways (Daggett and Fersht, 2003). Hydrophobic interactions also play a role in driving protein complex formation and contributes to stabilisation thereof. Reports of the effects of hydrophobic interactions at the dimer interface are mixed. Examples exist where both an increase (Ohkuri and Yamagishi, 2003) and decrease (Bell *et al.*, 2002) of hydrophobic

interactions improved dimer stability. Hydrophobic interactions occur frequently at the two fold symmetry axis of dimeric proteins. Figure 4.21 depicts the hydrophobic residues at the interface relative to the two-fold symmetrical axis. Polar residues at these positions will lead to repelling interactions and dimer destabilisation. From a rational point of view dimer stability can be improved by optimising the hydrophobic interactions around the two-fold symmetry axis. Under industrial biocatalytic conditions with elevated temperatures, catalytic activity can be driven thermodynamically and substrate affinity is a less important factor for consideration. The dimer integrity becomes the crucial factor and the remaining surface clusters with low conservation may be possible targets for further evolutionary optimisation.

4.3 Conclusion

E. coli transketolase previously subcloned into the pQR711 vector has been modified by inserting an N-terminal His_6 sequence. Wild-type and mutant transketolase were purified using a single step Ni-NTA protocol for activity and stability studies. Transketolase displays an unfolding profile with a three-state transitional nature. The two transitions are initially defined as the $N_2 \rightarrow 2I$ and $I \rightarrow U$ unfolding events with midpoints at 3.2 M and 4.9 M respectively when using urea as chemical denaturant. Two mutants, D381A and Y440A, have reduced dimer stability and catalytic activity when compared to the wild-type transketolase. Reduced activity results from a direct loss of inter subunit H-bonds and indirect loss of cofactor stabilising interactions. Analysis of properties of interface residues (conservation, H-bond and hydropathy) will lead to a possible a strategy of directed evolution that may improve the stability of homodimeric proteins used in industrial biotransformation processes.

Chapter 5

High-throughput route for stability analysis of transketolase

5.1 Introduction

The development of protocols for high-throughput protein production is the subject of extensive research focus in the post-genomic era. In several fields, effective mining of available genomic data is reliant on the analyses of pure recombinant protein samples in a high-throughput or parallel manner. These research areas include protein engineering (Jermutus *et al.*, 2001), enzyme catalysis (Koga *et al.*, 2003), studies pertaining to the understanding gene function (Osterman and Overbeek, 2003), protein expression patterns (Page *et al.*, 1999), physical characterisation of gene products with unknown function (Martzen *et al.*, 1999) as well as the study of protein-protein interaction and networking (Blackstock and Weir, 1999). Our main interest lies with producing pure protein samples in sufficient yields for directed evolutionaliry improvement of biocatalyst stability and formulation of therapeutic protein samples[2]. Production of pure recombinant protein samples involve three main steps: 1) PCR and subcloning, 2) cell culturing and expression, 3) purification and buffer-exchange. This section provides a brief overview of each steps.

[2]Part of the results presented in this chapter were included in Jean P. Aucamp, Gary J. Lye, Paul A. Dalby, (2004) 'Engineering Novel High-throughput Screening Routes for Biocatalyst Development Programmes' 5th European Symposium of Biochemical Engineering Science, Stuttgart, Germany, September 8-11.

5.1.1 PCR and subcloning

Traditionally ligation was an unavoidable step when subcloning DNA fragments into plasmid vectors for propagation and expression. Ligation steps are time consuming with low success rates, often reducing the overall throughput and efficiency of the subcloning process. This problem has been overcome with the developed of 'ligation-free' cloning systems allowing high efficiency cloning of gene targets into suitable vectors (Doyle *et al.*, 2002; Hartley *et al.*, 2000). Expression hosts have a great affect on the functional expression and solubility of recombinant proteins. The expression of eukaryotic genes in microbial systems are often problematic resulting in low functional expression, misfolding or product aggregation (Hartley and Kane, 1988; Lilie *et al.*, 1998). Recombination cloning provides a 'ligation-free' strategy for moving gene sequences between host-specific plasmid vectors (Lesley, 2001) and multi-system vectors with hybrid promoters allow for gene expression in different hosts without additional cloning steps (Chambers *et al.*, 2004). These techniques simplify the process of expression host optimisation.

In this study the transketolase gene is already expressed in a suitable expression system. Libraries of variant transketolase genes can be produced in 'ligation-free' strategies with whole plasmid epPCR (Fromant *et al.*, 1995), saturation site-directed mutagenesis (SSDM) (O'Donohue and Kneale, 1994; Zheng *et al.*, 2004), megaprimer PCR of whole plasmid (MEGAWHOP) (Miyazaki and Takenouchi, 2002) or "focused" error prone PCR (fepPCR) (Miller, 2004). Alternatively variants that have undergone gene sequence truncation (Ostermeier *et al.*, 1999a,b) or homology recombination (Stemmer, 1994; Crameri *et al.*, 1998) can be obtained using strategies involving one ligation step. Though ligation will reduce efficiency in these strategies, the real bottleneck for biocatalyst improvements with directed evolution lies with the time consuming screening processes (Dove, 1999; Vaschetto *et al.*, 2003). The first mutation library, if well-designed, may serve as starting point for improvement or alteration of several biocatalyst properties.

5.1.2 Functional expression

Optimisation of the expression of recombinant proteins requires consideration of several factors: 1) use of induced or constitutive expression 2) use of fusion tags 3) choice of expression host and 4) required protein yield.

Tight control of expression under an inducible promoter allows for production of gene products that are toxic to the host, but adds an additional induction step to the protein production route. Non-toxic biocatalysts may be expressed constitutively to make the process route less complex. Fusion tags can have dual functionality by facilitating purification with affinity chromatography and promoting the stable and functional fold of proteins. Several studies showed that fusion with tags such as glutathione-S-transferase (GST) or maltose binding protein (MBP) can improve the solubility and expression level of eukaryotic proteins (Murby *et al.*, 1996; Smith, 2000; Kapust and Waugh, 1999). Other non-affinity fusion proteins such as green fluorescent protein (GFP) (Rücker *et al.*, 2001) or ubiquitin (Catanzariti *et al.*, 2004) can serve the same purpose. The choice and construction of the fusion protein depends on the functionalities that are desired from the fusion tag.

Microbial cells such as *E. coli* are still the most widely used expression host, although insect cells (Amegadzie *et al.*, 1993) or *Pichia* (Gellissen, 2000) are also being used. Protein production in expression hosts provide yields (typically low milligram amounts) sufficient for functional analyis, proteomic studies (Doyle *et al.*, 2002) and crystallisation (Kim *et al.*, 2004; Schmid, 2002). Cell-free expression systems produce low protein yields (typically microgram amounts), enough for screens such as expression optimisation (Knaust and Nordlund, 2001; Nguyen *et al.*, 2004). Subsequently these expression systems are popular for screens that identify gene constructs with improved levels of functional expression or soluble protein. Pre-screening approaches can be timesaving by circumventing unnecessary cloning, cultivation and purification steps (Busso *et al.*, 2003). The required yields of soluble, functional protein will determine which expression system will be effective for an application.

5.1.3 Protein purification

Recombinant protein fused with peptide tags can be obtained in high purity and yield using affinity chromatography. This technique harnesses the high selectivity of fusion tags for their natural ligands to provide a rapid single-step purification protocol. A variety of suitable small and large fusion tags have been identified and are used routinely. Examples of small tags are His_6-tag, S-tag, ZZ-tag and calmodulin binding peptide (CBP) while large tags are glutathione-S-transferase (GST), thioredoxin, biotinylation domain, maltose binding domain (MBP) and chitin binding protein (Braun and LaBaer, 2003; Lesley, 2001; Terpe, 2003). GST and thioredoxin are the most popular large affinity tags also used to enhance functional expression. Several studies suggest that MBP is more successful in facilitating correct folding of eukaryotic gene products when compared to the His, CBP and GST tags (Braun *et al.*, 2002).

The His_6-tag is generally regarded as being unobtrusive with no significant effect on the structure or function of recombinant proteins in solution. There are however reports where small tags such as His_6 and CBP were inaccessible under non-denaturing conditions due to misfolding, leading to ineffective enrichment (Braun *et al.*, 2002; Hammarström *et al.*, 2002). Additionally, unstructured terminal residues can reduce the crystallisation efficiency of proteins and may need to be removed with suitable protease cleavage sites (Heinemann *et al.*, 2001). Despite these disadvantages, the His_6-tag is by far the most commonly used tag for affinity purification and was also used in this study. The affinity tag of His_6-transketolase is accessible for affinity binding since it was used successfully in Chapter 4 to purify wild-type and mutant transketolases. The His_6-tag is also unlikely to influence ΔF stability measurements. Larger tags such as MBP would contribute significantly to the overall fluorescence signal.

Purification steps using affinity chromatography and fusion tags require elution with excess amounts of a competing ligands such as imidazole, glutathione (GSH), β-mercapto-ethanol (β-ME), biotin, maltose or thiol to displace the protein from the affinity column (Braun and LaBaer, 2003). Long term contact with these ligands may have adverse effects

on protein structure and function or interfere with the screening process. High-throughput buffer exchange, using dialysis or ultra-filtration (UF), is required to remove these elution compounds. In this study buffer exchange was done using dialysis.

5.1.4 Dialysis buffer exchange

Dialysis is the process where a solute at a high concentration in one compartment migrates across a semipermeable membrane into another compartment with a lower solute concentration. The dialysis rate is dependent on the concentration gradient of the chemical species across the membrane as well as viscosity, temperature, mixing rate, molecular size and charge (Atkins, 2001).

Several systems are available where parallel sample dialysis can be achieved in a multiwell format, 96 wells being the maximum well density. All the systems belong to either one of two designs. The first design is a multiwell plate (24 or 96 well format) where each well has a dialysis membrane as bottom (The Nest Group Inc, BD Biosciences; Harvard Apparatus). Samples are loaded onto the membrane of each well and allowed to float in the bulk buffer. Dialysis occurs across the membrane interface at well bottom. This is an open system and evaporation is limited with a 96-well lid. Easy access to samples make these plates suitable for time course studies as needed for equilibrium dialysis. These plates are specially produced and costly disposables in routine high-throughput screens.

A second apparatus, Spectra/Por® 96-Well MicroDialyzer, design by Spectrum (Breda, Netherlands) is much more complex. A custom made 96-hole frame is sealed at one side with a membrane and gasket, forming wells into which the samples are loaded. Underneath the membrane is a reservoir chamber with 96 holes, aligned with the 96-holes in the frame. The reservoir is filled with buffer through an inlet port. An outlet port, pump and stirrer bar create circulation inside the reservoir chamber. The 96-hole frame is compatible with standard multi-channel pipettes for liquid handling purposes. The apparatus is intricate and running parallel high-throughput experiments will be very costly.

A 'home-made' dedicated equilibrium dialysis apparatus has been designed by Kariv

et al. (2001). This device is used for testing serum protein binding properties of pharmaceutical compounds in high-throughput manner. The apparatus was designed for studies that require analytical precision and was validated by testing drugs with known protein binding properties.

These systems are open to the atmosphere so that liquid crosses the membrane into the wells until osmotic pressure and hydrostatic pressure are balanced (Thain, 1967). These increases in sample volumes result in unnecessary dilution effects. For all mentioned systems, samples need to be transferred to and from the dialysis apparatus increasing the liquid handling steps in an automated process. With the design described in this study, there is no need for liquid transfer since the sample is dialysed directly in a 96-well microplate. This reduces liquid-handling steps and the loss of material associated with sample transfer. In addition, this design is substantially different from the other designs, reduces cost of disposables and is cost-effective to run in parallel.

5.1.5 Protein unfolding

Protein unfolding procedures used in this section have been characterised and discussed in detail in Chapter 3. The fixed volume method has been used to determine the working range for transketolase. The serial addition method has been used to analyse the dimer stabilities of the wild-type transketolase and both the D381A and Y440A mutants. All the transketolase proteins used in this study were His_6-tagged on the N-terminus. The proposed high-throughput route, depicted in Figure 5.1, starts with cloning and ends with the screen. Optimisation of the route occurs in the opposite direction by first determining the amount of protein required for screening. With proteins, the intrinsic fluorescence intensity is dependent on the amount and micro-environment of the tryptophan residues. In Chapter 3 the amounts of protein used were 20 μg and 40 μg per well for cyt *c* and BSA respectively. This provides a good starting point for determining the quantity of transketolase required for stability measurements.

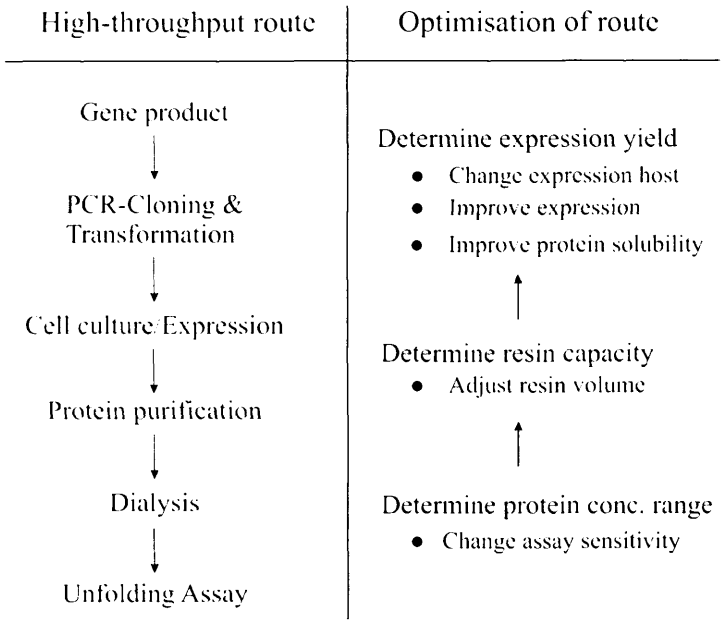


Figure 5.1: Schematic presentation of the proposed high-throughput route screening for protein stability.

5.2 Results and Discussion

5.2.1 Protein concentration range

Adequate fluorescence intensity is required to detect changes in tryptophan emission spectral properties induced by unfolding. Since fluorescence intensity is dependent on the quantity of protein per well it is important to determine the minimum amount of protein required to conduct the stability screens. Unfolding curves were generated using different amounts of pure wild-type transketolase and the fixed volume method (Section 2.5). Results show that between 50 and 100 µg transketolase per well will generate unfolding curves suitable for analysis (Figure 5.2). The protein is loaded into wells as a 50 µl initial volume, thus the transketolase concentration has to be at least 1 mg·ml^{-1} after the dialysis step.

The two major pitfalls, best avoided when developing macroscopic fluorescence intensity (FLINT) assays, are protein autofluorescence and inner-filter effects. The transketolase solution was diluted and the change in fluorescence intensity recorded. Small drifts

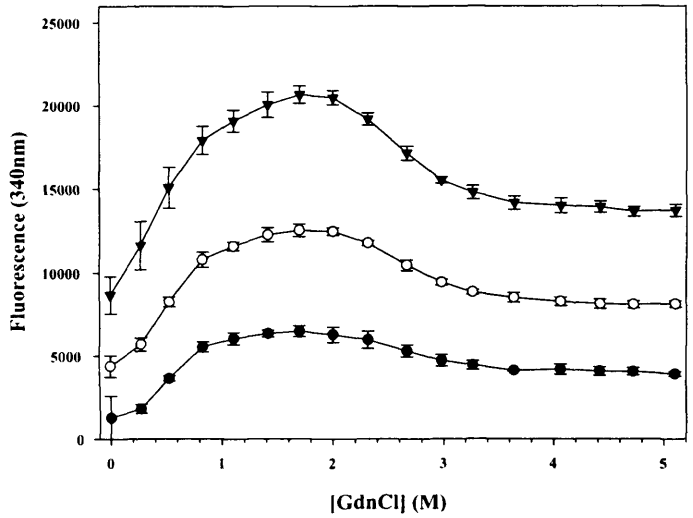


Figure 5.2: High-throughput GdnHCl induced unfolding of various concentrations of wild-type transketolase. (•) 25 µg, (○) 50 µg and (▼) 100 µg were loaded per well in an initial volume of 50 µl. A denaturant concentration range was created by adding 25 mM Tris, pH 7.5 containing either 0 M or 8 M GdnHCl. Error bars represent standard deviation (σ) from the mean obtained from triplicate measurements.

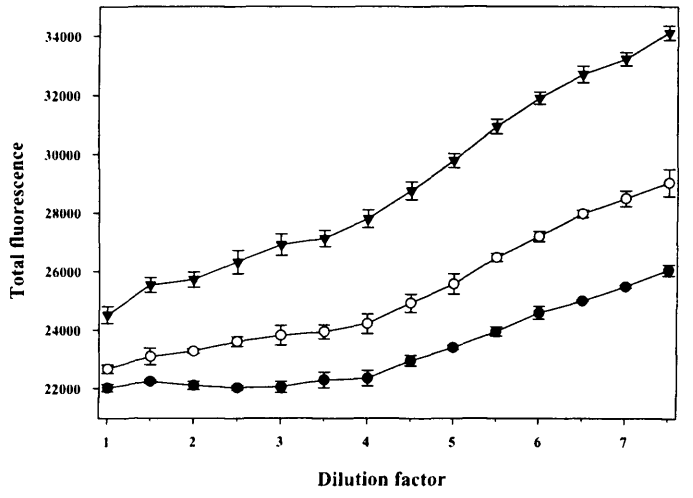


Figure 5.3: Dilution causes drifts in the protein fluorescence signal as a result of autofluorescence and inner-filter effects. Fluorescence drifts were obtained for various quantities of wild-type transketolase. A total of (•) 25, (○) 50 and (▼) 75 µg protein were loaded per well and fluorescence emission measured at 340 nm after excitation at 280 nm. Error bars represent standard deviation (σ) from the mean obtained from triplicate measurements.

indicate that the concentration of the target analyte is not affected by either of the above mentioned phenomena. The fluorescence signal from 25 μg per well had a 7% decrease over the initial 4-fold dilution (Figure 5.3). Fluorescence signals of wells containing 50 and 75 μg displayed drifts amounting to 11% and 33% increases over a 4-fold dilution respectively. By combining the results of Figures 5.2 and 5.3 it is evident that an initial transketolase concentration of 1 $mg \cdot ml^{-1}$ is best suited for the stability screen. A 4-fold dilution is sufficient to unfold transketolase completely when using 8 M urea stock solution.

5.2.2 Immobilisation of transketolase in Ni-NTA microwells

The initial aim was to capture His_6-transketolase in microplates with Ni-NTA coated wells. Elution, dialysis and unfolding steps would all be conducted in the same microplate. To this extent, the protein binding capacity and reproducibility of protein immobilisation were evaluated for Ni-NTA wells. The total immobilised protein content, determined using the bicinchoninic acid (BCA) protein quantification procedure, was less than 0.25 μg transketolase per well (Figure 5.4). The BCA assay is suitable for quantifying protein that are free in solution or immobilised on surfaces such as in wells (Sorensen and Brodbeck, 1986) or on resins (Stich, 1990). An excess amount of pure His_6-transketolase was loaded into Ni-NTA wells and the immobilised protein eluted for SDS-PAGE analysis (Figure 5.5). BCA protein quantification provided similar results (0.3-0.6 μg per well) for these experiments, suggesting the maximum loading capacity of the wells to be less than 1 μg. This value is at the lower end of the sensitivity for the BCA assay (Redinbaugh and Turley, 1986), making it an unreliable procedure for assessing the variability of protein binding. Paborsky *et al.* (1996) also determined the binding capacity of Pierce Ni-NTA plates (Reacti-Bind) to be low, estimating it in the order of 5-50 nM protein per well.

The fluorescence enzyme-linked assay was used as an indirect method for estimating relative amounts of immobilised transketolase in wells. Variation in reaction rates should be the result of varied amounts of transketolase immobilised in wells. With enzyme-linked

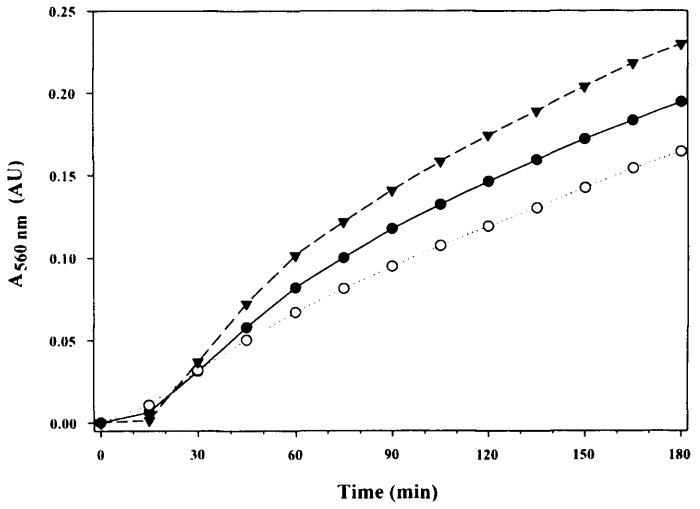


Figure 5.4: Time course of BCA protein quantification assay over 3 hr measures the formation of the Cu^{+1}:bicinchoninic acid colour complex at A_{560nm}. The rate and intensity of absorbance is an indication of the total protein content per well. As protein standards, 100 µl BSA solutions at concentrations of (○) 0 µg·ml^{-1} and (▼) 2.5 µg·ml^{-1} were loaded per well. (●) 100 µl cell-free extract was used to immobilise transketolase in Ni-NTA coated wells. The cell free extract contained approximately 500 µg·ml^{-1} transketolase, enough to saturate the binding capacity of the Ni-NTA coated wells. Results indicate that less than 0.25 µg transketolase is immobilised per well.

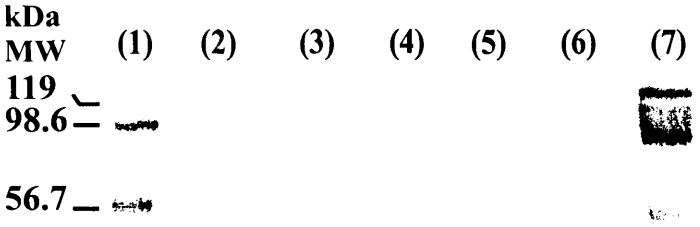
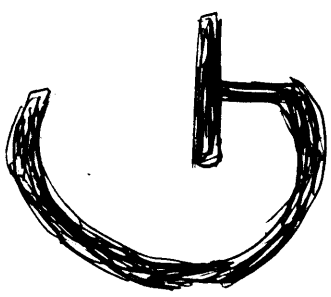


Figure 5.5: SDS-PAGE analysis, using a 12% polyacrylamide gel, of samples loaded and eluted from Ni-NTA wells. The lanes are as follow: (1 and 7) Molecular weight markers; (2) Purified transketolase at a concentration of 70 µg·ml^{-1} in 50 mM Tris, 20 mM imidazole, pH 8.5; (3) Same preparation as in lane 2 diluted 1:1 with 50 mM Tris, 20 mM imidazole, pH 8.5; (4 and 5) Transketolase eluted from Ni-NTA wells after immobilisation using pure transketolase preparations of lanes 2 and 3; (6) Transketolase eluted from a Ni-NTA well after immobilisation using cell-free extract obtained from an overnight culture of *E. coli* JM107 pQR791 with an original cell density of 5 AU at A_{600nm}.



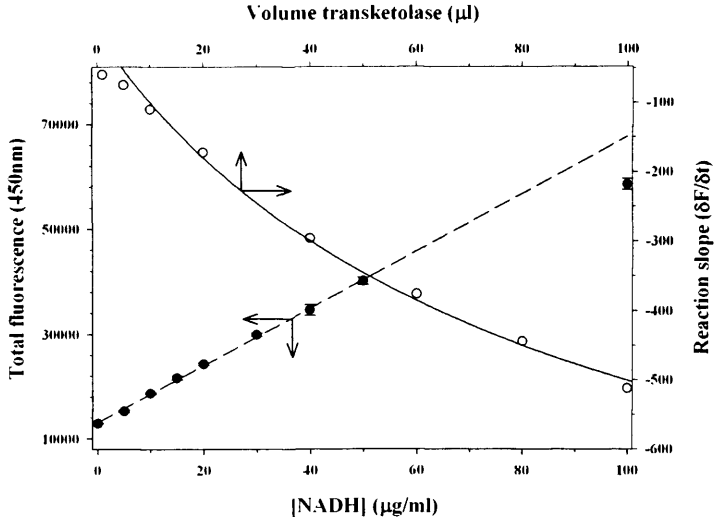


Figure 5.6: (• - -) Calibration curve for NADH obtained from fluorescence readings at 450 nm after excitation at 340 nm. Averages from triplicate readings are plotted in the concentration range of 0-100 $\mu g \cdot ml^{-1}$. Error bars represent the standard deviation from the mean (σ); r^2=0.998. The fluorescence increase is linear from 0-50 $\mu g \cdot ml^{-1}$. (o —) Response curve for the fluorescence enzyme-linked assay using various volumes of pure transketolase. Reaction rate is reported as the consumption of NADH with the decrease of fluorescence at 450 nm having a negative slope ($\delta F/\delta t$). The reaction rate was plotted against the volume of transketolase in the reaction. The response curve was fitted to the hyperbolic equation: $y = (a \times x)/(b + x)$ with a=-909.02 b=81.63 and r^2=0.985. The response curve displays slightly poorer fit to the linear equation ($y = -4.71x - 69.18$) with r^2=0.983.

assays care has to be taken that the observed reaction rate is indicative of the enzyme of interest, in this case transketolase, and not of that of the reporter enzyme. Care was taken to work in the linear fluorescence concentration range of NADH which is between 0 and 50 $\mu g \cdot ml^{-1}$ (Figure 5.6) as measured with the microplate reader. A response curve was generated for the fluorescence enzyme-linked assay using a concentration range of transketolase. The slopes of the reactions, representing the reaction rates, were plotted against the volume of transketolase present in the reaction (Figure 5.6) producing a hyperbolic trend. Transketolase was immobilised in wells by loading 100 µl of a cell-free extract obtained from an over-night culture of *E. coli* JM107 pQR791. The variation of δF was determined for two sets of conditions each calculated from eight wells. For the first set 200 µl assay mix (Table 2.6) was loaded per well for activity measurement. In the second set 100 µl assay mix and 100 µl buffer were combined. The percentages variance were calculated to be 15.7% and 17.3% for the first and second set respectively. The intrinsic variance of the assay was determined by measuring the reaction rates for a fixed amount of purified transketolase. Results taken from 21 wells produced a variance of approximately 4%. The magnitude of the overall variance was not deemed a problem, as in the final stability assay the allowed range is 50-100 µg protein per well. SDS-PAGE analysis shows the eluted transketolase to be of high purity. However the amount of immobilised protein was too little (approx. 1 µg per well) and an alternative protein purification protocol was required.

5.2.3 High-throughput protein purification using Ni-NTA resin

Ni-NTA resin suspensions are used routinely for high-throughput protein purification (Gottstein and Forde, 2002; Doyle *et al.*, 2002) and can potentially provide larger yields of pure protein. His_6-transketolase was purified in a high-throughput manner using the RoboPop kit (Novagen) as well as a vacuum manifold and pump. The resin provided with the kit is the standard Ni-NTA FastFlow^{TM} resin supplied by Qiagen. The transketolase binding capacity of the resin was determined in order calculate the optimal volume of

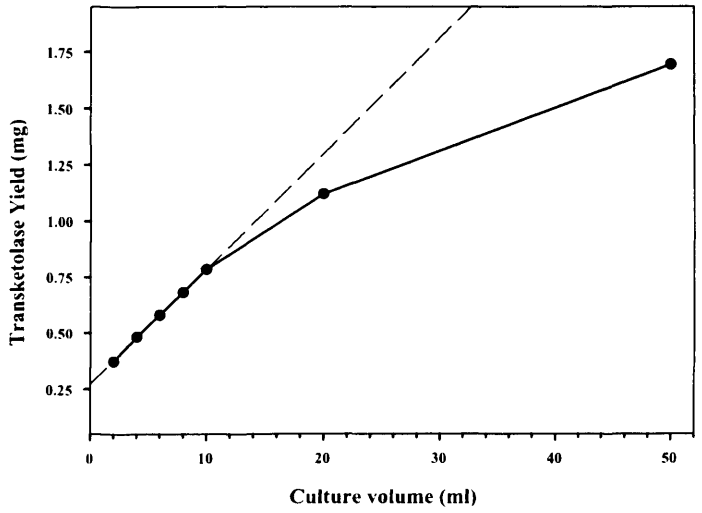


Figure 5.7: (• -) Wild-type transketolase binding capacity of Ni-NTA resin was determined by incubating 2 ml resin with various volumes of cell-free extract of 1 AU at 600 nm. Less than 1 mg transketolase binds per ml resin. The linear slope (- -) obtained for data points at lower volumes, where no saturation occurs, indicates that the transketolase yield is 0.05 $mg \cdot ml^{-1} \cdot OD^{-1}_{600nm}$ using the E. coli JM107 pQR791 expression system. Linear fit displayed a correlation coefficient of r^2=0.999.

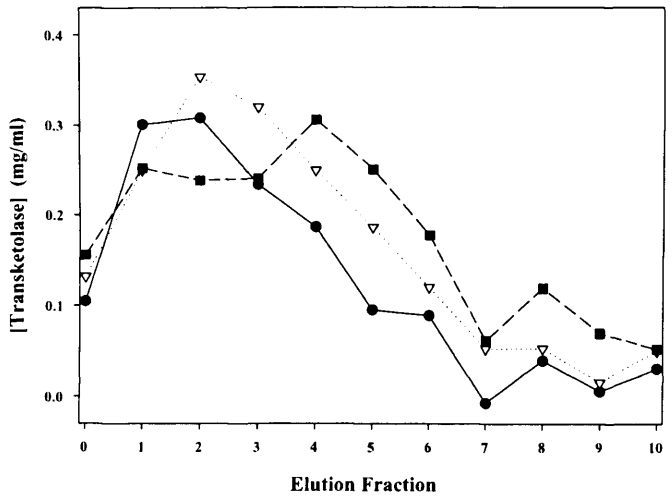


Figure 5.8: Elution profiles were obtained for (•) 100, (▽) 200 and (■) 300 µl resin. The elution buffer was added in 100 µl fractions and recovered using the vacuum manifold. Absorbance measurements were made at 280 nm using the Hellma quartz microcuvette. Fraction 0 is taken as the A_{280nm} reading from the final wash step prior to elution.

resin required to attain the target protein concentration of 1 $mg{\cdot}ml^{-1}$. An overnight cell culture of *E. coli* JM107 pQR791 was diluted to have an A_{600nm} measurement of 1 AU. The culture was lysed using the RoboPop lysis reagents, (PopCulture™, rLysoyme™ and Benzonase® Nuclease). 2 ml Ni-NTA resin was shake-incubated for 2 hr with various volumes of the prepared cell-free extract. The resin was washed; the transketolase eluted and dialysed. The transketolase concentrations were quantified with A_{280nm} measurements. Resin saturation data are shown in Figure 5.7 and indicate that approximately 1 mg transketolase bind per ml resin. Binding capacities reported in literature vary greatly, ranging from 0.1-2 $mg{\cdot}ml^{-1}$ (Doyle *et al.*, 2002) to 80-100 $mg{\cdot}ml^{-1}$ resin (Draveling *et al.*, 2001). It can also be determined that the *E. coli* JM107 pQR791 expression system yields a transketolase titer of 0.05 $mg{\cdot}ml^{-1}$ cell culture with absorbance of 1.0 AU at A_{600nm}. This value is useful in calculating the required cell culture density for the miniaturised cell culturing step of the route.

Initially 200 µl elution fractions were taken according to the standard protocol. However, protein concentrations were found to be lower than expected. To investigate this, elution profiles were obtained for 100, 200 and 300 µl volumes resin. The elution buffer was added in 100 µl fractions resin with immobilised transketolase and recovered using the vacuum manifold. Absorbance measurements were made at 280 nm. Control elutions were done using resin without immobilised transketolase to obtain an baseline absorbance readings for the imidazole. Figure 5.8 displays the estimated protein concentration obtained from each elution fraction. From the results it is evident that the protein concentration only reaches a concentration of 0.5 $mg{\cdot}ml^{-1}$. The low concentration is probably the result of several factors: 1) low binding capacity; 2) resin has a void volume that further dilutes protein; 3) not all the protein is displaced instantaneously.

The elution step requires more optimisation. The target protein concentration may possibly be reached by using more resin and eluting with a higher concentration of imidazole. Alternatively an ultrafiltration (UF) step must me introduced. UF can be used for both buffer exchange and concentration of the protein.

5.2.4 High-throughput cell culturing

The required amount of transketolase is estimated to be 200 µg per well and can be captured with 200 µl resin, providing that the elution step is optimised sufficiently. The titer yield is 50 µg protein per 1 ml culture of 1 AU at A_{600nm} and the working culture volume is 1 ml. From this the target cell density is calculated to be 4 AU at A_{600nm}. The strategy for cell culturing is based on work done by S. Doig (UCL). To ascertain whether the target cell density can be achieved, the maximal theoretical ($Q_{O_2(max)}$) and required ($Q_{O_2(req)}$) oxygen fluxes were calculated using Equations 2.20 and 2.21. The cell density (x) was calculated with the conversion factor where cell culture with 1 $AU_{600nm} = 0.4$ g·L^{-1} dry cell weight. The required oxygen flux was estimated with μ as 0.5 generation·hr^{-1} and Y_{O_2} as 1.5 g cells formed per 1 g O_2 consumed. The maximal theoretical oxygen flux was determined with the microplate specific ($k_l a'$) set to 150 hr^{-1} as determined by S. Doig (UCL). The oxygen concentration gradient was set at its maximum by using the O_2 saturation concentration at 37°C in water (6×10^{-3} g·L^{-1}) (Washburn, 1926) for (c_l^*) and 0 g·L^{-1} for (c_l).

The required oxygen flux ($Q_{O_2(req)}$) was estimated as 0.53 g·L^{-1}·hr^{-1} and the the maximum oxygen flux ($Q_{O_2(max)}$) as 0.9 g·L^{-1}·hr^{-1}. Thus $Q_{O_2(max)} > Q_{O_2(req)}$ and the target cell density is attainable without oxygen limitations. *E. coli* JM107 pQR791 cultured overnight reached a final A_{600nm} of 4.3 with a percentage coefficient of variation of 2.01% as determined from 8 wells.

5.2.5 High-throughput dialysis

5.2.5.1 Design

The dialysis apparatus, presented schematically in Figure 5.9, consist of a base plate, silicone gasket and a lid. The base plate (A) is made of aluminium and has a hollowed out groove the exact size of the footprint of a 96-well format microwell plate. A microwell plate fits snugly into this groove and a dialysis membrane (B) is placed on top of the microwell plate. The dialysis membrane is covered with a silicone gasket with 96 holes

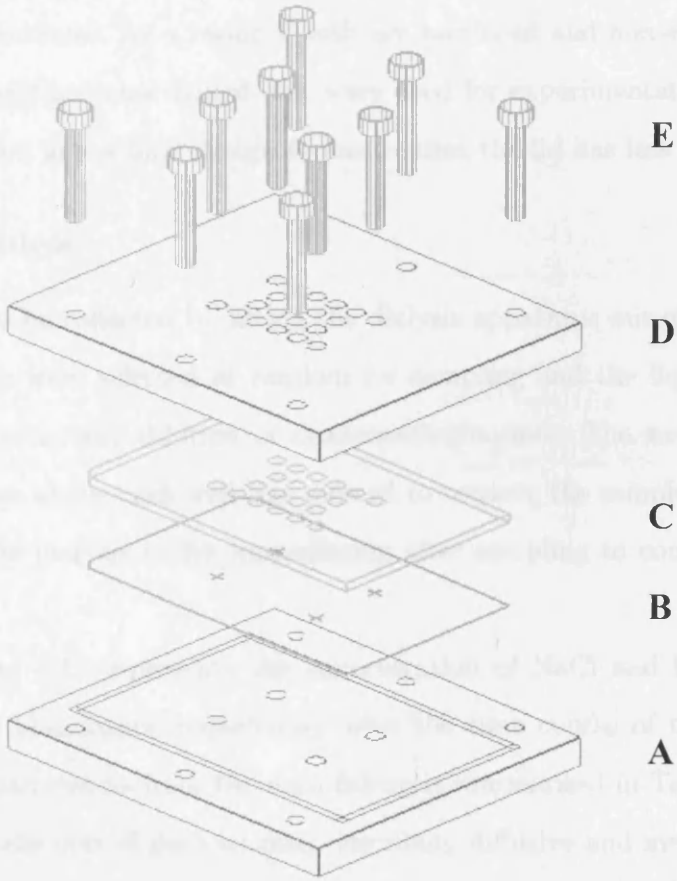


Figure 5.9: Schematic presentation of the high-throughput dialysis apparatus. The components are: (A) Aluminium base plate with threaded holes and rectangular groove for microplate footprint; (B) Dialysis membrane; (C) 96-hole silicone gasket; (D) Perspex^{TM} lid with 96-holes aligned with microwells; (E) Ten bolts to connect the base plate and lid, compressing the membrane and gasket onto a 96-well format microwell plate. The microplate is placed between component (A) and (B). The drawing was created in IntelliCAD2002 (CADopia LLC, San Diego, CA, USA; http://www.cadopia.com)

aligned to the microwells (C). A Perspex^{TM} (acrylic polymer) lid with 96 holes (D) covers the gasket. The microwell plate, membrane and gasket are then sandwiched together between the base plate and lid with 10 bolts (E). This ensures that every well seals of individually. Initially the design needs four additional bolts in the middle of the plate to ensure that the flexible Perspex^{TM} lid does not arch too much, pressing evenly on the gasket and membrane. As a result 4 wells are sacrificed and non-standard microwell plates, with four well bottoms drilled out, were used for experimentation. A more rigid material will be used in the final design to ensure that the lid has less flexibility.

5.2.5.2 Rate of Dialysis

Samples could only be collected by lifting the dialysis apparatus out of the bulk dialysis buffer. Three wells were selected at random for sampling and the liquid on top of the wells removed to avoid any dilution or cross-contamination. The wells were sacrificed when the membrane above each well was pierced to retrieve the samples. The apparatus was returned to the dialysis buffer immediately after sampling to continue time course studies.

Figures 5.10 and 5.11 represents the concentration of NaCl and imidazole, plotted as conductivity or absorbance respectively, over the time course of the of the dialysis experiment. The parameters from the data fitting is summarised in Table 5.1. This rate constant provides the overall dialysis rate, describing diffusive and membrane transport rates of the system as a single constant and therefore not very useful for diagnostic purposes. For a symmetrical membrane, in a closed system where no osmosis occurs, the rate constant should be identical in both directions. This is the case with NaCl but not with imidazole where different initial solute concentrations were used.

5.2.5.3 Cross-contamination and membrane integrity

The molecular weights of cytochrome *c* (cyt *c*) and blue dextran are 12.4 kDa and 2000 kDa respectively and the molecular weight cut-off (MWCO) of the membrane is

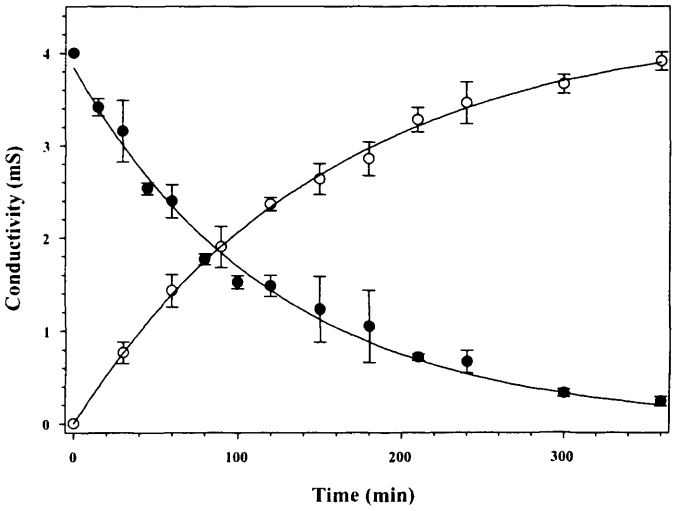


Figure 5.10: Dialysis time course of 1 M sodium chloride both into and out of microwells. (•) Removal of 1 M sodium chloride from the microwells, when dialysed against deionised H_2O, as monitored with conductivity. (○) Reverse process where 1 M sodium chloride diffused back into wells, containing only deionised H_2O.

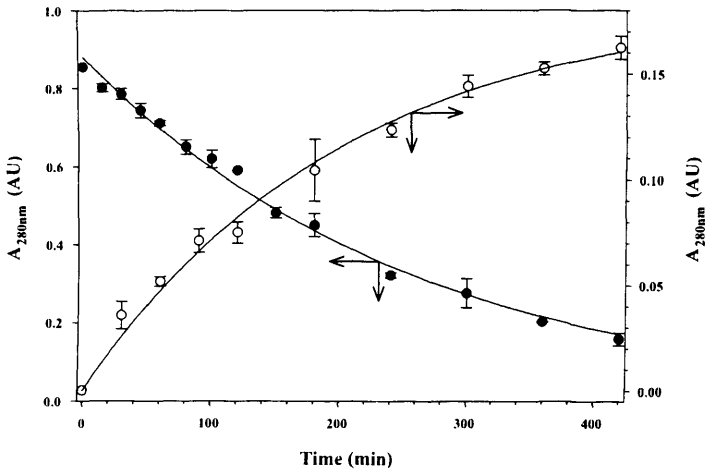


Figure 5.11: Dialysis time course of imidazole into and out of microwells. (•) Removal of 1 M imidazole from the microwells, when dialysed against deionised H_2O, as monitored with absorbance at 280 nm. (○) Reverse process where 0.25 M imidazole diffused back into wells, containing only deionised H_2O.

Table 5.1: Summary of the dialysis rate parameters obtained by fitting dialysis time course data from Figures 5.10 and 5.11 to the first order integrated rates (Equations 2.22 and 2.23; Section 2.30.4). Correlation coefficients indicate that data adheres to the first order rate law.

Experiment	k $(min)^{-1}$	r^2
NaCl (out)	0.0082	0.976
NaCl (in)	0.0081	0.982
Imidazole (out)	0.0039	0.990
Imidazole (in)	0.0063	0.974

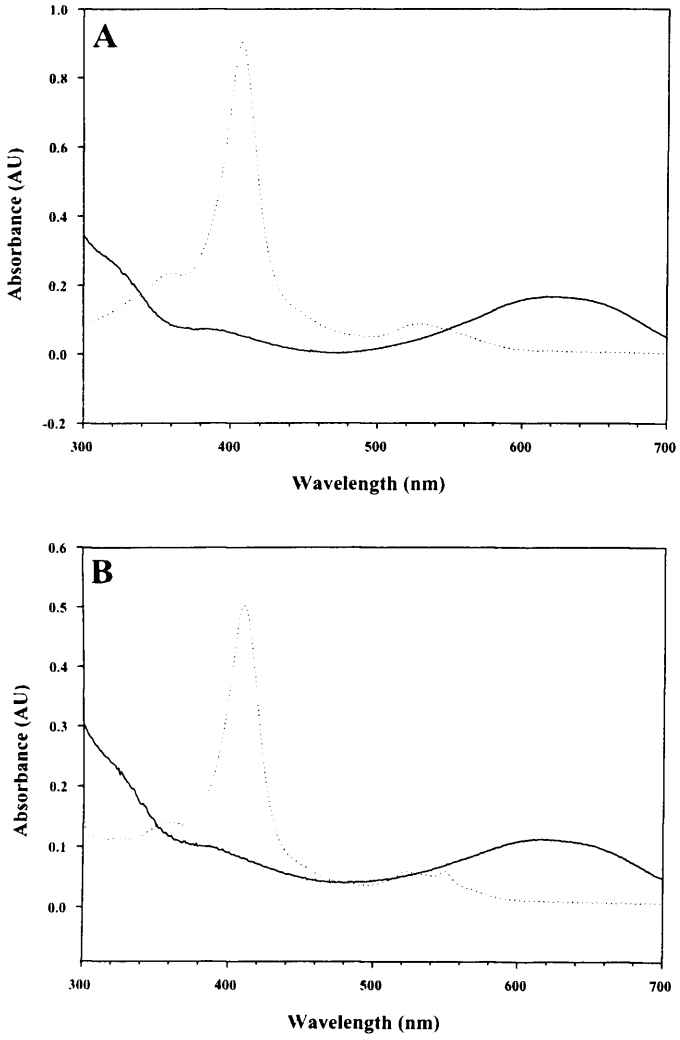


Figure 5.12: Typical absorbance spectra of blue dextran (—) and cyt *c* (- - -) respectively. Spectra are determined (A) before and (B) after dialysis and indicate that no cross-contamination occurred during the dialysis process.

10 kDa. The MWCO is defined as the molecular weight that is retained 90% by the membrane (Mulder, 1991), suggesting that cyt *c* will not be retained completely. Blue dextran on the other hand would not pass through the membrane unless the integrity of the membrane is compromised. Spectral analysis in the range of 300 nm to 700 nm for cyt *c* and blue dextran show characteristic absorbance peaks at 408 nm and 620 nm respectively. In the case of cross-contamination both these unique peaks would be present in the absorbance spectra of post-dialysis samples. The absorbance spectra obtained before and after analysis were very similar (Figures 5.12 A and B). The cyt *c* concentration after dialysis was less since the membrane did not retain it completely. Absorbance at 620 nm, used as an indication of the concentration, was measured for all wells that contained blue dextran. Statistical analysis of the 12 wells containing dextran blue estimated a coefficient of variation of 6.77%. This value is low and probably due to liquid handling error. A much larger error value would be expected if the membrane leakage occurred, suggesting that the membrane remained intact despite the pressure used to secure it between the well edge and a silicone gasket.

5.2.6 High-throughput unfolding of wild-type and mutant transketolases

The midpoints of the dissociation transition, as an indication of dimer stability, have been determined for wild-type transketolase and mutants using the serial addition method (Section 2.5). Fluorescence data measurements of protein samples obtained with the high-throughput protein purification procedure are depicted in Figure 5.13. The high-throughput purification yielded protein concentrations that were too low (as discussed Section 5.2.3) to provide accurate unfolding data. Instead, pure protein samples obtained with large scale column purification were used for the 'proof of principle' stability screening. $C_{1/2}$ values estimated from unfolding curves (Figure 5.14) have been compared with results obtained in Chapter 4 (Figure 4.17).

Wild-type transketolase has a $C_{1/2}$ of 2.8 $\pm$0.1 M for the $N_2 \rightarrow 2I$ transition, marginally lower than the 3.1 $\pm$0.1 M obtained with the more 'classical' ΔF_{340nm} method in Chapter

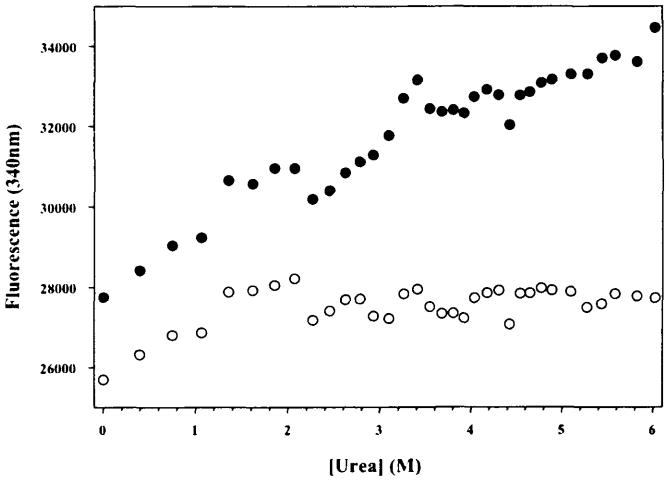


Figure 5.13: Urea induced F_{340nm} unfolding curves of (•) wild-type and (○) D381A transketolase samples obtained from the high-throughput protein purification step. Generated unfolding curves were not suitable for data analysis as a result of too low protein concentrations in the final purified samples.

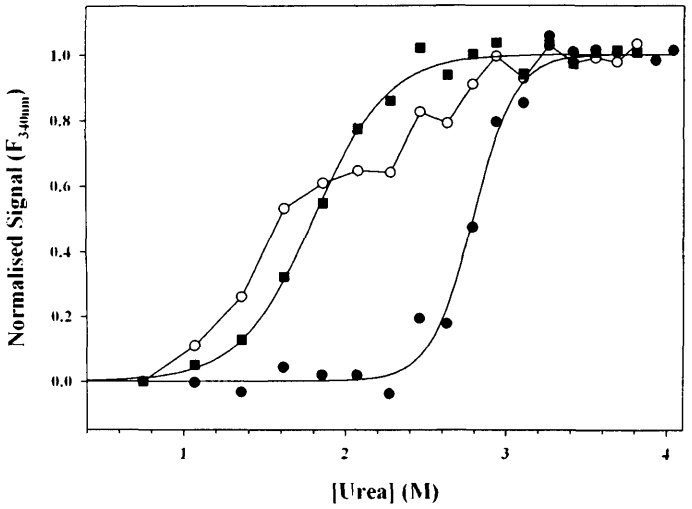


Figure 5.14: Urea induced F_{340nm} unfolding curves of the $N_2 \rightarrow 2I$ transitions of (•) wild-type, (○) D381A and (■) Y440A transketolases. Curves were generated with serial addition high-throughput stability screen and are similar to the ΔF_{340nm} unfolding analysis obtained in Chapter 4 (Section 2.5 and Figure 4.17). **Experimental conditions:** Wild-type and mutant transketolases were prepared as 1 $mg{\cdot}ml^{-1}$ solutions in 25 mM Tris buffer pH 7.5 with 5 mM $MgCl_2$ and 0.5 mM TPP. 50 µl of transketolase solution was loaded into wells. A denaturant solution of 25 mM Tris buffer pH 7.5 with 8 M urea was added using the serial addition method. Samples were mixed for 30 seconds by orbital shaking at 350 rpm, and reaction allowed to equilibrate for 15 minutes. Fluorescence measurements were recorded at 340 ±10 nm.

4 (Section 4.2.7). From the graph it is evident that both the D381A and Y440A mutants display reduced stabilities. Both the 1^{st} and 2^{nd} transitions, comprising the $N_2 \rightarrow 2I$ transition, are visible in the unfolding curve of the D381A mutant and were used to estimate an apparent $C_{1/2}$ of of 1.8 $\pm$0.1 M. The estimated $C_{1/2}$ value for the 1^{st} transition of the Y440A mutant was also 1.8 $\pm$0.1 M. The $N_2 \rightarrow 2I$ transitions for the mutants show a similar pattern as found in Chapter 4 with the 1^{st} and 2^{nd} transitions of the D381A mutant occurring on either side of the Y440A $N_2 \rightarrow 2I$ transition (Figure 4.17). The results obtained in Chapter 4 (see Section 4.2.7; Table 4.4) estimate D381A to be slightly less stable than Y440A with $C_{1/2}$ values of 1.2 $\pm$0.1 M and 1.4 $\pm$0.1 M respectively. The results obtained with the high-throughput method estimate D381A and Y440A to have, within experimental error, very similar stability values. As reported in Chapter 3 (Section 3.2.3), the high-throughput unfolding screen is not sensitive enough to accurately rank proteins with very similar unfolding midpoints ($\pm$0.15 M) such as the case for D381A and Y440A. From the unfolding curves it is observed that the D381A mutant first transition occurs at a lower denaturant concentration than that of the Y440A mutant, suggesting the former to be less stable.

5.3 Conclusion

The use of robotic systems are relatively commonplace for purification of nucleic acids and are increasingly being adapted for application in protein purification. Several robotic systems with functions relevant to protein purification are commercially available and have aspirate, liquid-dispense, pipet, vacuum filter, and plate-shake facilities on a relatively compact platform (Lesley, 2001). These functions can be adapted to perform cell lysis and chromatography steps from 1-2 ml of bacterial culture. Specialised 96-well filter plates clear cell debris via vacuum filtration and are also used to retain resin for chromatographic separations.

The aim was to optimise and combine high-throughput protein production and purification steps in an attempt to produce a route for automated high-throughput protein

stability screening. This will be beneficial for screening large amounts of variant biocatalysts in directed evolutionary improvement programmes. It was determined that 50 μg transketolase per well will be enough for fluorescence detection of unfolding. Ni-NTA coated microwells immobilised too little protein, estimated to be less than 1 μg. A fluorescence based enzyme link assay was implemented for transketolase. This assay was sensitive and useful for evaluating the variance in protein immobilisation in wells.

A high-throughput protein purification procedure was implemented using Ni-NTA resin and 96-well filter plate. The capacity of Ni-NTA resin for transketolase was estimated to be approximately 1 $mg \cdot ml^{-1}$. The concentration of transketolase in the eluate was not high enough to do accurate stability screening. The low concentration can be attributed to the low binding capacity and void volume of the resin and requires further optimisation.

The unfolding curves obtained with the serial addition method display a similar trend to the curves obtained by more accurate measurement. Both the D381A and Y440A mutants display reduced dimer stability.

The dialysis apparatus has been designed and tested under typical operating conditions. Dialysis of high concentrations imidazole or NaCl can be achieved in 8 hr. This design uses existing 96-well format microwell plates which reduces screening cost. Liquid handling steps are reduced in automated multistep sample processing sequences typically associated with protein purification, protein-ligand and protein-protein interaction studies (Fernandes, 1998). The apparatus can be stacked vertically, allowing multiple parallel studies to be conducted simultaneously. As main disadvantage, special care has to be taken to ensure that samples drop onto membrane upon inversion of the apparatus and form an interface through which dialysis can take place.

Chapter 6

Conclusion and Future Recommendations

Screening of large protein libraries forms an integral part of many proteomic and protein engineering research programmes. Despite technological advances in automation and miniaturisation, high-throughput screening is still considered a bottleneck. To increase the screening throughput, many assays rely on indirect modes of measurement. Indirect screening, particularly in the area of directed evolution may lead to failure. Screening for improved biocatalyst stability, using directed evolution techniques, rely on elevated temperature with the concomitant chemical inactivation, irreversible aggregation and measurement of residual enzyme activity as a method for selecting protein variants with increased thermostability. This does not consider the possibility that improved reversible denaturation may also increase the level of residual activity observed. One of the aims of this research was to develop a high-throughput screening method that directly measures the stability of proteins.

To this extent, unfolding curves were generated in microwells by addition of chemical denaturants using an autotitrating syringe pump. Conformational changes were monitored by measurement of the resulting changes in tryptophan fluorescence. The $C_{1/2}$ estimate was the most robust stability parameter obtained in high-throughput format

and can be related directly to thermodynamic stability. The serial addition method was the least accurate method validated, useful when screening for changes in $C_{1/2}$ values of greater than 0.15 M. Combined with suitable numerical data analysis techniques this method can be a powerful tool for screening mutant biocatalyst libraries for improved stability.

The fixed volume method generated more reliable datasets, and parameter estimation gave results comparable to reported literature values. This method can be used as a more accurate medium-throughput method suitable for characterising either small numbers of rationally designed mutants or as a secondary screen, confirming 'hits' from the higher throughput serial-addition screen for protein stability.

High-throughput determination of stability in microplates can be applied in the fields of directed evolution, proteomics and therapeutic protein formulation. Within the scope of research activities in our laboratory, this method will be used in biocatalyst stability improvement programmes using directed evolution. *E. coli* transketolase is one biocatalyst studied extensively by our research group. Though a considerable amount of structural data is available for transketolase, no unfolding data has yet been reported in the literature.

The pQR711 plasmid vector expressing *E. coli* transketolase has been modified by inserting an N-terminal His_6 sequence. This facilitated rapid purification using a single metal affinity chromatography step, providing product with purity suitable for activity and unfolding studies. Urea induced unfolding studies on wild-type transketolase elucidated a three-state transitional profile. The two transitions were defined as the $N_2 \rightarrow 2I$, being the monomerisation of the holo-dimer (N_2) to a stable, partially unfolded intermediate (I) and $I \rightarrow U$, being the unfolding transition. The two transitions have midpoints at 3.2 M and 4.9 M urea respectively.

Two transketolase mutants, D381A and Y440A, were designed to probe the nature of the $N_2 \rightarrow 2I$ transition. Compared to the wild-type transketolase, both mutations displayed reduced catalytic activity and dimer stability based on the observed lowering in

the $C_{1/2}$ values. Stability is reduced as a result of the direct loss of intersubunit hydrogen bonds. In addition, the TPP cofactor stabilises the dimer by forming interactions with both subunits. Reduced affinity for TPP has an indirect destabilising effect in both mutants. More evidence should be gathered to determine if any secondary structure is lost in the $N_2 \rightarrow 2I$ transition, whether the TPP cofactor remains bound to the I form and to determine the main cause of the irreversible unfolding (aggregation, chemical inactivation or misfolding). To this extent further studies should be conducted with aims to characterise the transition states better.

In working towards the goal of screening a mutant transketolase library, a high-throughput protein expression and purification route had to be implemented. The route involved micro-scale cell-culturing, Ni-NTA affinity purification and dialysis steps, all requiring optimisation. The overall process had to yield a final enzyme concentration of 1 mg·ml^{-1} since 50μg transketolase loaded per well in 50μl initial volume was deemed optimal for fluorescence detection of unfolding. It was estimated that sufficient quantities of transketolase (more than 200μg) will be produced by the expression system, *E. coli* JM107 pQR791. Ni-NTA resin had a low binding capacity (1 mg·ml^{-1}) for transketolase, limiting the final concentration of enzyme to no more than 0.5 mg·ml^{-1}. Insufficient protein concentration prevented stability screening using samples prepared at high-throughput. Literature reports suggested the proposed route should yield reasonable protein concentrations. However, with transketolase the yield was unusually low, possible due to the molecular size or dimeric nature of the protein. It is however essential that the protein capture and elution step be optimised to complete the high-throughput route. Two possible solutions for the problem would be a combination of more resin per well and higher a concentration of imidazole elution buffer, or a replacement of the dialysis with an ultrafiltration step. An ultrafiltration step will allow imidazole removal and product concentration at the same time.

Instead, pure protein samples were used to generate transketolase unfolding curves using the serial addition method. Unfolding curves displayed similar profiles and stability

trends as observed when using larger volumes and more accurate measurement techniques. These results indicate that the method is suitable for estimation of dimer stability when screening transketolase variants for an improvement of this property.

The high-throughput route was designed to be integrated readily with an automation platform. Several robotic systems with functions relevant to protein purification are commercially available and have facilities to aspirate, dispense, pipet, vacuum filter, and plate-shake on a relatively compact platform. Improvements worth considering for current equipment are accurate titration down to 1 μl volumes, solution dispensing close to the surface of the solution in each well, minimised evaporation, active temperature control, and multiple syringe pumps for dispensing various stocks in parallel. The dialysis apparatus has also been designed with the automated process in mind. Liquid handling steps and expensive consumables can be reduced since dialysis is conducted in a regular 96-well microplate. In a suitable dialysis chamber, multiple parallel studies can be conducted simultaneously by stacking the apparatus vertically.

This route, though optimised for protein stability may also be applied to other problems. Occasionally biocatalytic reactions from wild-type or recombinant source are obscured by background activity in cell-free extracts. Partial protein purification using conventional ion-exchange chromatography may be readily implemented with this route and provide adequate removal of unwanted background reactions or enrichment of the studied enzyme.

Bibliography

Ahern, T. J. and Klibanov, A. M. (1985) The mechanisms of irreversible enzyme inactivation at 100°C. *Science* **228** (4705), 1280–1284.

Ahern, T. J. and Klibanov, A. M. (1988) Analysis of processes causing thermal inactivation of enzymes. *Methods of Biochemical Analysis* **33**, 91–127.

Ahmad, F. and Bigelow, C. C. (1982) Estimation of the free-energy of stabilization of ribonuclease A, lysozyme, α-lactalbumin, and myoglobin. *Journal of Biological Chemistry* **257** (21), 2935–2938.

Ahmad, N. and Qasim, M. A. (1995) Fatty-acid binding to bovine serum albumin prevents formation of intermediate during denaturation. *European Journal of Biochemistry* **227** (1-2), 563–565.

Akasaka, R., Mashino, T. and Hirobe, M. (1993) Cytochrome P450-like substrate oxidation catalyzed by cytochrome *c* and immobilized cytochrome *c*. *Archives of Biochemistry and Biophysics* **301** (2), 355–360.

Akasako, A., Haruki, M., Oobatake, M. and Kanaya, S. (1997) Conformational stabilities of *Escherichia coli* RNase HI variants with a series of amino acid substitutions at a cavity within the hydrophobic core. *Journal of Biological Chemistry* **272** (30), 18686–93.

Alexeeva, M., Carr, R. and Turner, N. J. (2003) Directed evolution of enzymes: new biocatalysts for asymmetric synthesis. *Organic & Biomolecular Chemistry* **1** (23), 4133–4137.

Amegadzie, B. Y., Jiampetti, D., Craig, R. J., Appelbaum, E., Shatzman, A. R., Mayer, R. J. and DiLella, A. G. (1993) High-level production of biologically active human cytosolic phospholipase A2 in baculovirus-infected insect cells. *Gene* **128** (2), 307–308.

Amiza, M. A. and Apenten, R. K. O. (1996) Urea and heat unfolding of cold-adapted Atlantic cod (*Gadus morhua*) trypsin and bovine trypsin. *Journal of the Science of Food and Agriculture* **70** (1), 1–10.

Anderson, D. E., Becktel, W. J. and Dahlquist, F. W. (1990) pH-induced denaturation of proteins: a single salt bridge contributes 3-5 kcal/mol to the free energy of folding of T4 lysozyme. *Biochemistry* **29** (9), 2403–2408.

Anderson, D. E., Hurley, J. H., Nicholson, H., Baase, W. A. and Matthews, B. W. (1993) Hydrophobic core repacking and aromatic-aromatic interaction in the thermostable mutant of T4 lysozyme Ser 117→Phe. *Protein Science* **2** (8), 1285–1290.

André, C., Bolte, J. and Demuynck, C. (1998a) Syntheses of L-threose and D-erythrose analogues modified at position 2. *Tetrahedron-Asymmetry* **9** (8), 1359–1367.

André, C., Guérard, C., Hecquet, L., Demuynck, C. and Bolte, J. (1998b) Modified L-threose and D-erythrose as substrates of transketolase and fructose-1,6-bisphosphate aldolase. Application to the synthesis of new heptulose analogues. *Journal of Molecular Catalysis B-Enzymatic* **5** (5-6), 459–466.

Armon, A., Graur, D. and Ben-Tal, N. (2001) ConSurf: an algorithmic tool for the identification of functional regions in proteins by surface mapping of phylogenetic information. *Journal of Molecular Biology* **307** (1), 447–463.

Arnold, F. H. (1998) Enzyme engineering reaches the boiling point. *Proceedings of the National Academy of Sciences of the United States of America* **95** (5), 2035–2036.

Arnone, M. I., Birolo, L., Pascarella, S., Cubellis, M. V., Bossa, F., Sannia, G. and Marino, G. (1997) Stability of aspartate aminotransferase from *Sulfolobus solfataricus*. *Protein Engineering* **10** (3), 237–248.

Atkins, P. W. (2001) *The elements of physical chemistry: with applications in biology.* 3^{rd} edn W.H. Freeman New York.

Babbitt, P. C. and Gerlt, J. A. (1997) Understanding enzyme superfamilies. Chemistry as the fundamental determinant in the evolution of new catalytic activities. *Journal of Biological Chemistry* **272** (49), 30591–30594.

Balasubramanian, S. V., Bruenn, J. and Straubinger, R. M. (2000) Liposomes as formulation excipients for protein pharmaceuticals: A model protein study. *Pharmaceutical Research* **17** (3), 344–350.

Bashford, D. and Case, D. A. (2000) Generalized Born models of macromolecular solvation effects. *Annual Review of Physical Chemistry* **51**, 129–152.

Bell, G. S., Russell, R. J., Connaris, H., Hough, D. W., Danson, M. J. and Taylor, G. L. (2002) Stepwise adaptations of citrate synthase to survival at life's extremes. From psychrophile to hyperthermophile. *European Journal of Biochemistry* **269** (24), 6250–6260.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. (2000) The Protein Data Bank. *Nucleic Acids Research* **28** (1), 235–242.

Bessler, C., Schmitt, J., Maurer, K. H. and Schmid, R. D. (2003) Directed evolution of a bacterial α-amylase: toward enhanced pH-performance and higher specific activity. *Protein Science* **12** (10), 2141–2149.

Bhattacharya, A. A., Grüne, T. and Curry, S. (2000) Crystallographic analysis reveals common modes of binding of medium and long-chain fatty acids to human serum albumin. *Journal of Molecular Biology* **303** (5), 721–732.

Billinton, N. and Knight, A. W. (2001) Seeing the wood through the trees: A review of techniques for distinguishing green fluorescent protein from endogenous autofluorescence. *Analytical Biochemistry* **291** (2), 175–197.

Bio-Rad (Manual) Mini-PROTEAN® II electrophoresis cell instruction manual.

Blaber, M., Zhang, X. J. and Matthews, B. W. (1993) Structural basis of amino acid α-helix propensity. *Science* **260** (5114), 1637–1640.

Blackstock, W. P. and Weir, M. P. (1999) Proteomics: quantitative and physical mapping of cellular proteins. *Trends in Biotechnology* **17** (3), 121–127.

Blum, H., Beier, H. and Gross, H. J. (1987) Improved silver staining of plant-proteins, RNA and DNA in polyacrylamide gels. *Electrophoresis* **8** (2), 93–99.

Boder, E. T., Midelfort, K. S. and Wittrup, K. D. (2000) Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. *Proceedings of the National Academy of Sciences of the United States of America* **97** (20), 10701–10705.

Boder, E. T. and Wittrup, K. D. (1997) Yeast surface display for screening combinatorial polypeptide libraries. *Nature Biotechnology* **15** (6), 553–557.

Bogan, A. A. and Thorn, K. S. (1998) Anatomy of hot spots in protein interfaces. *Journal of Molecular Biology* **280** (1), 1–9.

Bongs, J., Hahn, D., Schorken, U., Sprenger, G. A., Kragl, U. and Wandrey, C. (1997) Continuous production of erythrulose using transketolase in a membrane reactor. *Biotechnology Letters* **19** (3), 213–215.

Bornscheuer, U. T., Altenbuchner, J. and Meyer, H. H. (1999) Directed evolution of an esterase: Screening of enzyme libraries based on pH-Indicators and a growth assay. *Bioorganic & Medicinal Chemistry* **7** (10), 2169–2173.

Bowie, J. U. and Sauer, R. T. (1989) Equilibrium dissociation and unfolding of the Arc repressor dimer. *Biochemistry* **28** (18), 7139–43.

Brandt, D. (1998) Microfluidics and high-density microplates - The drug discovery process: Why fix what isn't broken? *Genetic Engineering News* **18** (7), 30–31.

Braun. P., Hu, Y. H., Shen, B. H., Halleck, A., Koundinya, M., Harlow, E. and LaBaer, J. (2002) Proteome-scale purification of human proteins from bacteria. *Proceedings of the National Academy of Sciences of the United States of America* **99** (5), 2654–2659.

Braun, P. and LaBaer, J. (2003) High throughput protein production for functional proteomics. *Trends in Biotechnology* **21** (9), 383–388.

Brems, D. N., Cass, R. and Stellwagen, E. (1982) Conformational transitions of frog-heart ferricytochrome *c*. *Biochemistry* **21** (7), 1488–1493.

Brinda, K. V., Kannan, N. and Vishveshwara, S. (2002) Analysis of homodimeric protein interfaces by graph-spectral methods. *Protein Engineering* **15** (4), 265–277.

Brocklebank, S., Woodley, J. M. and Lilly, M. D. (1999) Immobilised transketolase for carbon-carbon bond synthesis: biocatalyst stability. *Journal of Molecular Catalysis B-Enzymatic* **7** (1-4), 223–231.

Brodersen, R., Honoré, B. and Andersen, S. (1988) Palmitate binding to serum albumin, measured by rate of dialysis. *European Journal of Biochemistry* **174** (1), 45–50.

Brown, J. R. and Shockley, P. (1982) *Lipid-protein interactions* Vol. 1 John Wiley, New York ; Chichester.

Bull, A. T., Bunch, A. W. and Robinson, G. K. (1999) Biocatalysts for clean industrial products and processes. *Current Opinion in Microbiology* **2** (3), 246–251.

Burley, S. K. and Petsko, G. A. (1985) Aromatic-aromatic interaction: A mechanism of protein structure stabilization. *Science* **229** (4708), 23–28.

Burstein, E. A., Abornev, S. M. and Reshetnyak, Y. K. (2001) Decomposition of protein tryptophan fluorescence spectra into log-normal components. I. Decomposition algorithms. *Biophysical Journal* **81** (3), 1699–1709.

Burton, S. G., Cowan, D. A. and Woodley, J. M. (2002) The search for the ideal biocatalyst. *Nature Biotechnology* **20** (1), 37–45.

Busso, D., Kim, R. and Kim, S. H. (2003) Expression of soluble recombinant proteins in a cell-free system using a 96-well format. *Journal of Biochemical and Biophysical Methods* **55** (3), 233–240.

Caffrey, D. R., Somaroo, S., Hughes, J. D., Mintseris, J. and Huang, E. S. (2004) Are protein-protein interfaces more conserved in sequence than the rest of the protein surface? *Protein Science* **13** (1), 190–202.

Carra, J. H. and Privalov, P. L. (1996) Thermodynamics of denaturation of staphylococcal nuclease mutants: an intermediate state in protein folding. *FASEB Journal* **10** (1), 67–74.

Carter, P. and Wells, J. A. (1990) Functional interaction among catalytic residues in subtilisin BPN'. *Proteins* **7** (4), 335–342.

Catanzariti, A. M., Soboleva, T. A., Jans, D. A., Board, P. G. and Baker, R. T. (2004) An efficient system for high-level expression and easy purification of authentic recombinant proteins. *Protein Science* **13** (5), 1331–1339.

Cavaliere, S. W., Neet, K. E. and Sable, H. Z. (1975) Enzymes of pentose biosynthesis. The quaternary structure and reacting form of transketolase from baker's yeast. *Archives of Biochemistry and Biophysics* **171** (2), 527–532.

Cedrone, F., Ménez, A. and Quéméneur, E. (2000) Tailoring new enzyme functions by rational redesign. *Current Opinion in Structural Biology* **10** (4), 405–410.

Chambers, S. P., Austen, D. A., Fulghum, J. R. and Kim, W. M. (2004) High-throughput screening for soluble recombinant expressed kinases in *Escherichia coli* and insect cells. *Protein Expression and Purification* **36** (1), 40–47.

Chauhan, R. P., Woodley, J. M. and Powell, L. W. (1996) *In situ* product removal from *E.coli* transketolase-catalyzed biotransformations. *in* Enzyme Engineering Xiii Vol. 799 of *Annals of the New York Academy of Sciences* pp. 545–554.

Chen, K. and Arnold, F. H. (1993) Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Proceedings of the National Academy of Sciences of the United States of America* **90** (12), 5618–5622.

Chothia, C. and Janin, J. (1975) Principles of protein-protein recognition. *Nature* **256** (5520), 705–708.

Clackson, T. and Wells, J. A. (1995) A hot spot of binding energy in a hormone-receptor interface. *Science* **267** (5196), 383–386.

Clarke, J. and Fersht, A. R. (1993) Engineered disulfide bonds as probes of the folding pathway of barnase: Increasing the stability of proteins against the rate of denaturation. *Biochemistry* **32** (16), 4322–4329.

Coco, W. M., Levinson, W. E., Crist, M. J., Hektor, H. J., Darzins, A., Pienkos, P. T., Squires, C. H. and Monticello, D. J. (2001) DNA shuffling method for generating highly recombined genes and evolved enzymes. *Nature Biotechnology* **19** (4), 354–359.

Cohen, H. M., Tawfik, D. S. and Griffiths, A. D. (2004) Altering the sequence specificity of *Hae*III methyltransferase by directed evolution using *in vitro* compartmentalization. *Protein Engineering Design & Selection* **17** (1), 3–11.

Collaborative Computational Project, Number 4. (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallographica Section D-Biological Crystallography* **50** (Pt 5), 760–763.

Crameri, A., Raillard, S. A., Bermudez, E. and Stemmer, W. P. C. (1998) DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* **391** (6664), 288–291.

Creighton, T. E. (1993) Proteins: structures and molecular properties. 2nd edn W.H. Freeman New York pp. 181, 311–312.

Cull, M. G., Miller, J. F. and Schatz, P. J. (1992) Screening for receptor ligands using large libraries of peptides linked to the C terminus of the *lac* repressor. *Proceedings of the National Academy of Sciences of the United States of America* **89** (5), 1865–1869.

Daggett, V. and Fersht, A. R. (2003) Is there a unifying mechanism for protein folding? *Trends in Biochemical Sciences* **28** (1), 18–25.

Daggett, V. and Levitt, M. (1993) Protein unfolding pathways explored through molecular dynamics simulations. *Journal of Molecular Biology* **232** (2), 600–619.

Dahiyat, B. I. and Mayo, S. L. (1996) Protein design automation. *Protein Science* **5** (5), 895–903.

Dahiyat, B. I. and Mayo, S. L. (1997) Probing the role of packing specificity in protein design. *Proceedings of the National Academy of Sciences of the United States of America* **94** (19), 10172–10177.

Dalby, P. A. (2003) Optimising enzyme function by directed evolution. *Current Opinion in Structural Biology* **13** (4), 500–505.

Dao-pin, S., Sauer, U., Nicholson, H. and Matthews, B. W. (1991) Contributions of engineered surface salt bridges to the stability of T4 lysozyme determined by directed mutagenesis. *Biochemistry* **30** (29), 7142–7153.

Datta, A. G. and Racker, E. (1961) Mechanism of action of transketolase. I. Properties of the crystalline yeast enzyme. *Journal of Biological Chemistry* **236**, 617–623.

Davis, B. G. (2003) Chemical modification of biocatalysts. *Current Opinion in Biotechnology* **14** (4), 379–386.

Declerck, N., Machius, M., Wiegand, G., Huber, R. and Gaillardin, C. (2000) Probing structural determinants specifying high thermostability in *Bacillus licheniformis* α-amylase. *Journal of Molecular Biology* **301** (4), 1041–1057.

Desmet, J., Demaeyer, M., Hazes, B. and Lasters, I. (1992) The dead-end elimination theorem and its use in protein side-chain positioning. *Nature* **356** (6369), 539–542.

Di, L., McConnell, O. J., Kerns, E. H. and Sutherland, A. G. (2004) Rapid, automated screening method for enzymatic transformations using a robotic system and supercritical fluid chromatography. *Journal of Chromatography B* **809** (2), 231–235.

Dinner, A. R., Šali, A., Smith, L. J., Dobson, C. M. and Karplus, M. (2000) Understanding protein folding via free-energy surfaces from theory and experiment. *Trends in Biochemical Sciences* **25** (7), 331–339.

Doig, S. D., Pickering, S. C. R., Lye, G. J. and Woodley, J. M. (2002) The use of microscale processing technologies for quantification of biocatalytic Baeyer-Villiger oxidation kinetics. *Biotechnology and Bioengineering* **80** (1), 42–49.

Dove, A. (1999) Drug screening–beyond the bottleneck. *Nature Biotechnology* **17** (9), 859–863.

Doyle, S. A., Murphy, M. B., Massi, J. M. and Richardson, P. M. (2002) High-throughput proteomics: A flexible and efficient pipeline for protein production. *Journal of Proteome Research* **1** (6), 531–536.

Draths, K. M. and Frost, J. W. (1990) Synthesis using plasmid-based biocatalysis - Plasmid assembly and 3-deoxy-D-arabino-heptulosonate production. *Journal of the American Chemical Society* **112** (4), 1657–1659.

Draveling, C., Ren, L., Haney, P., Zeisse, D. and Qoronfleh, M. W. (2001) SwellGel: An affinity chromatography technology for high- capacity and high-throughput purification of recombinant-tagged proteins. *Protein Expression and Purification* **22** (2), 359–366.

Edgell, M. H., Sims, D. A., Pielak, G. J. and Yi, F. (2003) High-precision, high-throughput stability determinations facilitated by robotics and a semiautomated titrating fluorometer. *Biochemistry* **42** (24), 7587–7593.

Edinger, S., Cortis, C., Shenkin, P. and Friesner, R. (1997) Solvation free energies of peptides: Comparison of approximate continuum solvation models with accurate solution of the Poison-Boltzmann equation. *Journal of Physical Chemistry B* **101**, 1190–1197.

Egan, R. M. and Sable, H. Z. (1981) Transketolase kinetics. The slow reconstitution of the holoenzyme is due to rate-limiting dimerization of the subunits. *Journal of Biological Chemistry* **256** (10), 4877–83.

Eiben, A. E. and Back, T. (1997) Empirical investigation of multiparent recombination operators in evolution strategies. *Evolutionary Computation* **5** (3), 347–365.

Eijsink, V. G., Bjørk, A., Gåseidnes, S., Sirevåg, R., Synstad, B., van den Burg, B. and Vriend, G. (2004) Rational engineering of enzyme stability. *Journal of Biotechnology* **113** (1-3), 105–120.

Eisenberg, D. and McLachlan, A. D. (1986) Solvation energy in protein folding and binding. *Nature* **319** (6050), 199–203.

Elmahdi, I., Baganz, F., Dixon, K., Harrop, T., Sugden, D. and Lye, G. J. (2003) pH control in microwell fermentations of *S-erythraea* CA340: Influence on biomass growth kinetics and erythromycin biosynthesis. *Biochemical Engineering Journal* **16** (3), 299–310.

Engvall, E. (1977) Quantitative enzyme immunoassay (ELISA) in microbiology. *Medical Biology* **55** (4), 193–200.

Ermler, U., Merckel, M., Thauer, R. and Shima, S. (1997) Formylmethanofuran: tetrahydromethanopterin formyltransferase from *Methanopyrus kandleri* - new insights into salt-dependence and thermostability. *Structure* **5** (5), 635–646.

Feldman, M. W., Otto, S. P. and Christiansen, F. B. (1996) Population genetic perspectives on the evolution of recombination. *Annual Review of Genetics* **30**, 261–295.

Fernandes, P. B. (1998) Technological advances in high-throughput screening. *Current Opinion in Chemical Biology* **2** (5), 597–603.

Fields, P. A. (2001) Review: Protein function at thermal extremes: balancing stability and flexibility. *Comparative Biochemistry and Physiology A-Molecular & Integrative Physiology* **129** (2-3), 417–431.

Forrer, P., Jung, S. and Plückthun, A. (1999) Beyond binding: using phage display to select for structure, folding and enzymatic activity in proteins. *Current Opinion in Structural Biology* **9** (4), 514–520.

Francisco, J. A., Earhart, C. F. and Georgiou, G. (1992) Transport and anchoring of β-lactamase to the external surface of *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* **89** (7), 2713–2717.

Franks, F. (1994) Accelerated stability testing of bioproducts: attractions and pitfalls. *Trends in Biotechnology* **12** (4), 114–117.

Freire, E., van Osdol, W. W., Mayorga, O. L. and Sanchez-Ruiz, J. M. (1990) Calorimetrically determined dynamics of complex unfolding transitions in proteins. *Annual Review of Biophysics and Biophysical Chemistry* **19**, 159–188.

French, C. and Ward, J. M. (1995) Improved production and stability of *Escherichia coli* recombinants expressing transketolase for large-scale biotransformation. *Biotechnology Letters* **17** (3), 247–252.

Fromant, M., Blanquet, S. and Plateau, P. (1995) Direct random mutagenesis of gene-sized DNA fragments using polymerase chain-reaction. *Analytical Biochemistry* **224** (1), 347–353.

Ganem, B. (1978) Shikimate-derived metabolites. 4. From glucose to aromatics - recent developments in natural-products of shikimic acid pathway. *Tetrahedron* **34** (23), 3353–3383.

García-Arellano, H., Valderrama, B., Saab-Rincón, G. and Vazquez-Duhalt, R. (2002) High temperature biocatalysis by chemically modified cytochrome c. *Bioconjugate Chemistry* **13** (6), 1336–1344.

Geddie, M. L. and Matsumura, I. (2004) Rapid evolution of β-glucuronidase specificity by saturation mutagenesis of an active site loop. *Journal of Biological Chemistry* **279** (25), 26462–26468.

Gekko, K. and Timasheff, S. N. (1981) Mechanism of protein stabilization by glycerol: Preferential hydration in glycerol-water mixtures. *Biochemistry* **20** (16), 4667–4676.

Gellissen, G. (2000) Heterologous protein production in methylotrophic yeasts. *Applied Microbiology and Biotechnology* **54** (6), 741–750.

Ghaemmaghami, S., Fitzgerald, M. C. and Oas, T. G. (2000) A quantitative, high-throughput screen for protein stability. *Proceedings of the National Academy of Sciences of the United States of America* **97** (15), 8296–8301.

Gijsen, H. J., Qiao, L., Fitz, W. and Wong, C. H. (1996) Recent advances in the chemoenzymatic synthesis of carbohydrates and carbohydrate mimetics. *Chemical Reviews* **96** (1), 443–474.

Giver, L., Gershenson, A., Freskgard, P. O. and Arnold, F. H. (1998) Directed evolution of a thermostable esterase. *Proceedings of the National Academy of Sciences of the United States of America* **95** (22), 12809–12813.

Godzik, A. (1995) In search of the ideal protein sequence. *Protein Engineering* **8** (5), 409–416.

Goldstein, R. F. (1994) Efficient rotamer elimination applied to protein side-chains and related spin glasses. *Biophysical Journal* **66** (5), 1335–1340.

Gottstein, C. and Forde, R. (2002) Affinity chromatography system for parallel purification of recombinant protein samples. *Protein Engineering* **15** (10), 775–777.

Gray, K. A., Richardson, T. H., Kretz, K., Short, J. M., Bartnek, F., Knowles, R., Kan, L., Swanson, P. E. and Robertson, D. E. (2001) Rapid evolution of reversible denaturation and elevated melting temperature in a microbial haloalkane dehalogenase. *Advanced Synthesis & Catalysis* **343** (6-7), 607–617.

Greener, A., Callahan, M. and Jerpseth, B. (1997) An efficient random mutagenesis technique using an *E.coli* mutator strain. *Molecular Biotechnology* **7** (2), 189–195.

Guérard, C., Alphand, V., Archelas, A., Demuynck, C., Hecquet, L., Furstoss, R. and Bolte, J. (1999) Transketolase-mediated synthesis of 4-deoxy-D-fructose 6- phosphate by epoxide hydrolase-catalysed resolution of 1,1- diethoxy-3,4-epoxybutane. *European Journal of Organic Chemistry* (12), 3399–3402.

Guerois, R. and Serrano, L. (2001) Protein design based on folding models. *Current Opinion in Structural Biology* **11** (1), 101–106.

Hamilton, J. A., Era, S., Bhamidipati, S. P. and Reed, R. G. (1991) Locations of the three primary binding sites for long-chain fatty acids on bovine serum albumin. *Proceedings of the National Academy of Sciences of the United States of America* **88** (6), 2051–2054.

Hammarström, M., Hellgren, N., van den Berg, S., Berglund, H. and Härd, T. (2002) Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli. *Protein Science* **11** (2), 313–321.

Harbury, P. B., Plecs, J. J., Tidor, B., Alber, T. and Kim, P. S. (1998) High-resolution protein design with backbone freedom. *Science* **282** (5393), 1462–1467.

Hartley, D. L. and Kane, J. F. (1988) Properties of inclusion bodies from recombinant *Escherichia coli*. *Biochemical Society Transactions* **16** (2), 101–102.

Hartley, J. L., Temple, G. F. and Brasch, M. A. (2000) DNA cloning using *in vitro* site-specific recombination. *Genome Research* **10** (11), 1788–1795.

Haupts, U., Rüdiger, M. and Pope, A. J. (2000) Macroscopic versus microscopic fluorescence techniques in (ultra)-high-throughput screening. *Drug Discovery Today* pp. 3–9.

Hawkins, C. F., Borges, A. and Perham, R. N. (1989) A common structural motif in thiamin pyrophosphate-binding enzymes. *FEBS Letters* **255** (1), 77–82.

Hecquet, L., Bolte, J. and Demuynck, C. (1994a) Chemoenzymatic synthesis of 6-deoxy-D-fructose and 6-deoxy-L-sorbose using transketolase. *Tetrahedron* **50** (29), 8677–8684.

Hecquet, L., Bolte, J. and Demuynck, C. (1996) Enzymatic synthesis of "natural-labeled" 6-deoxy-L-sorbose precursor of an important food flavor. *Tetrahedron* **52** (24), 8223–8232.

Hecquet, L., Lemaire, M., Bolte, J. and Demuynck, C. (1994b) Chemoenzymatic synthesis of precursors of fagomine and 1,4- dideoxy-1,4-imino-D-arabinitol. *Tetrahedron Letters* **35** (47), 8791–8794.

Heinemann, U., Illing, G. and Oschkinat, H. (2001) High-throughput three-dimensional protein structure determination. *Current Opinion in Biotechnology* **12** (4), 348–354.

Heinrich, P. C., Steffen, H., Janser, P. and Wiss, O. (1972) Studies on the reconstitution of apotransketolase with thiamine pyrophosphate and analogs of the coenzyme. *European Journal of Biochemistry* **30** (3), 533–541.

Heinrich, P. C. and Wiss, O. (1971) Transketolase from human erythrocytes. Purification and properties. *Helvetica Chimica Acta* **54** (8), 2658–2668.

Hendsch, Z. S. and Tidor, B. (1994) Do salt bridges stabilize proteins? A continuum electrostatic analysis. *Protein Science* **3** (2), 211–226.

Henikoff, S. and Henikoff, J. G. (1992) Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences of the United States of America* **89** (22), 10915–10919.

Hennig, M., Darimont, B., Sterner, R., Kirschner, K. and Jansonius, J. N. (1995) 2.0 Å structure of indole-3-glycerol phosphate synthase from the hyperthermophile *Sulfolobus solfataricus*: possible determinants of protein stability. *Structure* **3** (12), 1295–1306.

Herskovits, T. T. and Laskowski, M., Jr. (1962) Location of chromophoric residues in proteins by solvent perturbation. I. Tyrosyls in serum albumins. *Journal of Biological Chemistry* **237**, 2481–2492.

Hess, J. M., Tchernajenko, V., Vieille, C., Zeikus, J. G. and Kelly, R. M. (1998) *Thermotoga neapolitana* homotetrameric xylose isomerase is expressed as a catalytically active

and thermostable dimer in *Escherichia coli*. *Applied and Environmental Microbiology* **64** (7), 2357–2360.

Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K. and Pease, L. R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77** (1), 51–59.

Hochuli, E., Dobeli, H. and Schacher, A. (1987) New metal chelate adsorbent selective for proteins and peptides containing neighboring histidine-residues. *Journal of Chromatography* **411**, 177–184.

Honig, B. and Nicholls, A. (1995) Classical electrostatics in biology and chemistry. *Science* **268** (5214), 1144–1149.

Horecker, B. L., Smyrniotis, P. Z. and Klenow, H. (1953) The formation of sedoheptulose phosphate from pentose phosphate. *Journal of Biological Chemistry* **205** (2), 661–682.

Horovitz, A., Serrano, L., Avron, B., Bycroft, M. and Fersht, A. R. (1990) Strength and co-operativity of contributions of surface salt bridges to protein stability. *Journal of Molecular Biology* **216** (4), 1031–1044.

Horsfall, M. J., Gordon, A. J., Burns, P. A., Zielenska, M., van der Vliet, G. M. and Glickman, B. W. (1990) Mutational specificity of alkylating agents and the influence of DNA repair. *Environmental and Molecular Mutagenesis* **15** (2), 107–122.

Hublin, A., Gradisar, H., Friedrich, J. and Vasic-Racki, D. (2002) Stability and stabilisation of *Doratomyces microsporus* keratinase. *Biocatalysis and Biotransformation* **20** (5), 329–336.

IbarraMolero, B. and SanchezRuiz, J. M. (1997) Are there equilibrium intermediate states in the urea-induced unfolding of hen egg-white lysozyme? *Biochemistry* **36** (31), 9616–9624.

Ishikawa, K., Nakamura, H., Morikawa, K. and Kanaya, S. (1993) Stabilization of

Escherichia coli ribonuclease HI by cavity-filling mutations within a hydrophobic core. *Biochemistry* **32** (24), 6171–8.

Jaenicke, R. (1991) Protein stability and molecular adaptation to extreme conditions. *European Journal of Biochemistry* **202** (3), 715–728.

Jermutus, L., Honegger, A., Schwesinger, F., Hanes, J. and Plückthun, A. (2001) Tailoring *in vitro* evolution for protein affinity or stability. *Proceedings of the National Academy of Sciences of the United States of America* **98** (1), 75–80.

Jespers, L., Jenné, S., Lasters, I. and Collen, D. (1997) Epitope mapping by negative selection of randomized antigen libraries displayed on filamentous phage. *Journal of Molecular Biology* **269** (5), 704–718.

Johnson, C. M. and Fersht, A. R. (1995) Protein stability as a function of denaturant concentration: ther thermal stability of barnase in the presence of urea. *Biochemistry* **34** (20), 6795–6804.

Jones, D. T. (1994) *De novo* protein design using pairwise potentials and a genetic algorithm. *Protein Science* **3** (4), 567–574.

Jones, S. and Thornton, J. M. (1995) Protein-protein interactions: a review of protein dimer structures. *Progress In Biophysics and Molecular Biology* **63** (1), 31–65.

Jones, S. and Thornton, J. M. (1996) Principles of protein-protein interactions. *Proceedings of the National Academy of Sciences of the United States of America* **93** (1), 13–20.

Joo, H., Lin, Z. and Arnold, F. H. (1999) Laboratory evolution of peroxide-mediated cytochrome P450 hydroxylation. *Nature* **399** (6737), 670–673.

Jordan, F. (2003) Current mechanistic understanding of thiamin diphosphate dependent enzymatic reactions. *Natural Product Reports* **20** (2), 184–201.

Juillerat, A., Gronemeyer, T., Keppler, A., Gendreizig, S., Pick, H., Vogel, H. and Johnsson, K. (2003) Directed evolution of O^6-alkylguanine-DNA alkyltransferase for effi-

cient labeling of fusion proteins with small molecules *in vivo*. *Chemistry & Biology* **10** (4), 313–317.

Juneja, J. and Udgaonkar, J. B. (2002) Characterization of the unfolding of ribonuclease A by a pulsed hydrogen exchange study: evidence for competing pathways for unfolding. *Biochemistry* **41** (8), 2641–54.

Jung, E. H., Takeuchi, T., Nishino, K. and Itokawa, Y. (1988) Studies on the nature of thiamine pyrophosphate binding and dependency on divalent cations of transketolase from human erythrocytes. *International Journal of Biochemistry* **20** (11), 1255–1259.

Kang, F. R. and Singh, J. (2003) Conformational stability of a model protein (bovine serum albumin) during primary emulsification process of PLGA microspheres synthesis. *International Journal of Pharmaceutics* **260** (1), 149–156.

Kapust, R. B. and Waugh, D. S. (1999) *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Science* **8** (8), 1668–1674.

Kardos, J., Bodi, A., Zavodszky, P., Venekei, I. and Graf, L. (1999) Disulfide-linked propeptides stabilize the structure of zymogen and mature pancreatic serine proteases. *Biochemistry* **38** (38), 12248–12257.

Kariv, I., Cao, H. and Oldenburg, K. R. (2001) Development of a high throughput equilibrium dialysis method. *Journal of Pharmaceutical Sciences* **90** (5), 580–587.

Kaushik, J. K. and Bhat, R. (1998) Thermal stability of proteins in aqueous polyol solutions: Role of the surface tension of water in the stabilizing effect of polyols. *Journal of Physical Chemistry B* **102** (36), 7058–7066.

Kellis, J. T., Nyberg, K. and Fersht, A. R. (1989) Energetics of complementary side-chain packing in a protein hydrophobic core. *Biochemistry* **28** (11), 4914–4922.

Khan, M. Y. (1986) Direct evidence for the involvement of domain III in the N-F transition of bovine serum albumin. *Biochemical Journal* **236** (1), 307–310.

Khan, M. Y., Agarwal, S. K. and Hangloo, S. (1987) Urea-induced structural transformations in bovine serum albumin. *Journal of Biochemistry (Tokyo)* **102** (2), 313–7.

Khan, M. Y. and Salahuddin, A. (1984) Lack of N–F transition in the N-terminal fragment (domain I + II) of bovine serum albumin. *European Journal of Biochemistry* **141** (3), 473–475.

Kim, G. J., Cheon, Y. H. and Kim, H. S. (2000) Directed evolution of a novel N-carbamylase/D-hydantoinase fusion enzyme for functional expression with enhanced stability. *Biotechnology and Bioengineering* **68** (2), 211–217.

Kim, Y., Dementieva, I., Zhou, M., Wu, R., Lezondra, L., Quartey, P., Joachimiak, G., Korolev, O., Li, H. and Joachimiak, A. (2004) Automation of protein purification for structural genomics. *Journal of Structural and Functional Genomics* **5** (1-2), 111–118.

Knapp, J. A. and Pace, C. N. (1974) Guanidine hydrochloride and acid denaturation of horse, cow, and *Candida krusei* cytochromes *c*. *Biochemistry* **13** (6), 1289–1294.

Knaust, R. K. C. and Nordlund, P. (2001) Screening for soluble expression of recombinant proteins in a 96-well format. *Analytical Biochemistry* **297** (1), 79–85.

Knowles, J. R. (1987) Tinkering with enzymes: what are we learning? *Science* **236** (4806), 1252–1258.

Kochetov, G. A. (1982) Transketolase from yeast, rat liver, and pig liver. *Methods in Enzymology* **90** (Pt E), 209–223.

Koehl, P. and Delarue, M. (1994) Application of a self-consistent mean field theory to predict protein side-chains conformation and estimate their conformational entropy. *Journal of Molecular Biology* **239** (2), 249–275.

Koga, Y., Kato, K., Nakano, H. and Yamane, T. (2003) Inverting enantioselectivity of *Burkholderia cepacia* KWI-56 lipase by combinatorial mutation and high-throughput screening using single-molecule PCR and *in vitro* expression. *Journal of Molecular Biology* **331** (3), 585–592.

Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227** (259), 680–685.

Lakowicz, J. R. (1999) *Principles of fluorescence spectroscopy.* 2^{nd} edn Kluwer Academic/Plenum New York.

Landgraf, R., Xenarios, I. and Eisenberg, D. (2001) Three-dimensional cluster analysis identifies interfaces and functional residue clusters in proteins. *Journal of Molecular Biology* **307** (5), 1487–1502.

Laskowski, R. A. (1995) SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. *Journal of Molecular Graphics* **13** (5), 323–330, 307–308.

Lazar, G. A., Desjarlais, J. R. and Handel, T. M. (1997) *De novo* design of the hydrophobic core of ubiquitin. *Protein Science* **6** (6), 1167–1178.

Lazaridis, T. and Karplus, M. (1999) Effective energy function for proteins in solution. *Proteins* **35** (2), 133–152.

Lee, B. and Richards, F. M. (1971) The interpretation of protein structures: estimation of static accessibility. *Journal of Molecular Biology* **55** (3), 379–400.

Lee, K. N., Park, S. D. and Yu, M. H. (1996) Probing the native strain in α_1-antitrypsin. *Nature Structural Biology* **3** (6), 497–500.

Lehmann, M., Pasamontes, L., Lassen, S. F. and Wyss, M. (2000) The consensus concept for thermostability engineering of proteins. *Biochimica et Biophysica Acta-Protein Structure and Molecular Enzymology* **1543** (2), 408–415.

Lehmann, M. and Wyss, M. (2001) Engineering proteins for thermostability: the use of sequence alignments versus rational design and directed evolution. *Current Opinion in Biotechnology* **12** (4), 371–375.

Lesley, S. A. (2001) High-throughput proteomics: Protein expression and purification in the postgenomic world. *Protein Expression and Purification* **22** (2), 159–164.

LiCata, V. J. and Ackers, G. K. (1995) Long-range, small magnitude nonadditivity of mutational effects in proteins. *Biochemistry* **34** (10), 3133–3139.

Lilie, H., Schwarz, E. and Rudolph, R. (1998) Advances in refolding of proteins produced in *E.coli*. *Current Opinion in Biotechnology* **9** (5), 497–501.

Lin, S.H., Konishi, Y., Denton, M.E. and Scheraga, H.A. (1984) Influence of an extrinsic cross-link on the folding pathway of Ribonuclease A. Conformational and thermodynamic analysis of cross-linked (Lysine^{7}-Lysine^{41})-Ribonuclease A.

Lindqvist, Y., Schneider, G., Ermler, U. and Sundström, M. (1992) Three-dimensional structure of transketolase, a thiamine diphosphate dependent enzyme, at 2.5 Å resolution. *The EMBO Journal* **11** (7), 2373–2379.

Lingen, B., Grötzinger, J., Kolter, D., Kula, M. R. and Pohl, M. (2002) Improving the carboligase activity of benzoylformate decarboxylase from *Pseudomonas putida* by a combination of directed evolution and site-directed mutagenesis. *Protein Engineering* **15** (7), 585–593.

Littlechild, J. (1995) Crystallization and preliminary X-ray crystallographic data with *Escherichia coli* transketolase. *Acta Crystallographica Section D-Biological Crystallography* **51** (Pt 6), 1074–1076.

Lutz, S., Ostermeier, M., Moore, G. L., Maranas, C. D. and Benkovic, S. J. (2001) Creating multiple-crossover DNA libraries independent of sequence identity. *Proceedings of the National Academy of Sciences of the United States of America* **98** (20), 11248–11253.

Lye, G. J., Ayazi-Shamlou, P., Baganz, F., Dalby, P. A. and Woodley, J. M. (2003) Accelerated design of bioconversion processes using automated microscale processing techniques. *Trends in Biotechnology* **21** (1), 29–37.

MacBeath, G., Kast, P. and Hilvert, D. (1998) Redesigning enzyme topology by directed evolution. *Science* **279** (5358), 1958–1961.

MacLean, D. S., Qian, Q. S. and Middaugh, C. R. (2002) Stabilization of proteins by low molecular weight multi-ions. *Journal of Pharmaceutical Sciences* **91** (10), 2220–2229.

Makhatadze, G. I., Loladze, V. V., Ermolenko, D. N., Chen, X. and Thomas, S. T. (2003) Contribution of surface salt bridges to protein stability: guidelines for protein engineering. *Journal of Molecular Biology* **327** (5), 1135–1148.

Mansfeld, J., Vriend, G., Dijkstra, B. W., Veltman, O. R., Van den Burg, B., Venema, G., Ulbrich-Hofmann, R. and Eijsink, V. G. (1997) Extreme stabilization of a thermolysin-like protease by an engineered disulfide bond. *Journal of Biological Chemistry* **272** (17), 11152–11156.

Margoliash, E. and Frohwirth, N. (1959) Spectrum of horse-heart cytochrome *c*. *Biochemical Journal* **71** (3), 570–572.

Marshall, S. A., Morgan, C. S. and Mayo, S. L. (2002) Electrostatics significantly affect the stability of designed homeodomain variants. *Journal of Molecular Biology* **316** (1), 189–199.

Martzen, M. R., McCraith, S. M., Spinelli, S. L., Torres, F. M., Fields, S., Grayhack, E. J. and Phizicky, E. M. (1999) A biochemical genomics approach for identifying genes by the activity of their products. *Science* **286** (5442), 1153–1155.

Matsumura, I., Wallingford, J. B., Surana, N. K., Vize, P. D. and Ellington, A. D. (1999) Directed evolution of the surface chemistry of the reporter enzyme β-glucuronidase. *Nature Biotechnology* **17** (7), 696–701.

Matsumura, M., Signor, G. and Matthews, B. W. (1989) Substantial increase of protein stability by multiple disulphide bonds. *Nature* **342** (6247), 291–293.

Matsuura, T., Yomo, T., Trakulnaleamsai, S., Ohashi, Y., Yamamoto, K. and Urabe, I. (1998) Nonadditivity of mutational effects on the properties of catalase I and its application to efficient directed evolution. *Protein Engineering* **11** (9), 789–795.

Matthews, B. W. (1993) Structural and genetic analysis of protein stability. *Annual Review of Biochemistry* **62**, 139–160.

Matthews, B. W., Nicholson, H. and Becktel, W. J. (1987) Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. *Proceedings of the National Academy of Sciences of the United States of America* **84** (19), 6663–6667.

Mayo, S., Olafson, B. and Goddard, W. (1990) Dreiding - A generic force-filed for molecular simulation. *Journal of Physical Chemistry* **94**, 8897–8909.

Merkel, J. S. and Regan, L. (2000) Modulating protein folding rates *in vivo* and *in vitro* by side-chain interactions between the parallel β strands of green fluorescent protein. *Journal of Biological Chemistry* **275** (38), 29200–29206.

Meshalkina, L., Nilsson, U., Wikner, C., Kostikowa, T. and Schneider, G. (1997) Examination of the thiamin diphosphate binding site in yeast transketolase by site-directed mutagenesis. *European Journal of Biochemistry* **244** (2), 646–652.

Miller, O. J. (2004) Directed evolution of transketolase a carbon-carbon bond-forming enzyme. PhD thesis. University College London.

Mitchinson, C. and Wells, J. A. (1989) Protein engineering of disulfide bonds in subtilisin BPN'. *Biochemistry* **28** (11), 4807–4815.

Mitra, R. K., Woodley, J. M. and Lilly, M. D. (1998) *Escherichia coli* transketolase-catalyzed carbon-carbon bond formation: biotransformation characterization for reactor evaluation and selection. *Enzyme and Microbial Technology* **22** (1), 64–70.

Mittl, P. R., Berry, A., Scrutton. N. S., Perham, R. N. and Schulz, G. E. (1994) Anatomy of an engineered NAD-binding site. *Protein Science* **3** (9), 1504–1514.

Miyazaki, K. and Takenouchi, M. (2002) Creating random mutagenesis libraries using megaprimer PCR of whole plasmid. *Biotechniques* **33** (5), 1033–1034, 1036–1038.

Moore, J. C., Jin, H. M., Kuchner, O. and Arnold, F. H. (1997) Strategies for the *in vitro* evolution of protein function: Enzyme evolution by random recombination of improved sequences. *Journal of Molecular Biology* **272** (3), 336–347.

Mulder, M. (1991) *Basic principles of membrane technology.* 1^{st} edn Kluwer Academic Dordrecht, Netherlands; Boston.

Murby, M., Uhlén, M. and Ståhl, S. (1996) Upstream strategies to minimize proteolytic degradation upon recombinant production in *Escherichia coli*. *Protein Expression and Purification* **7** (2), 129–136.

Myer, Y. P., MacDonald, L. H., Verma, B. C. and Pande, A. (1980) Urea denaturation of horse heart ferricytochrome *c*. Equilibrium studies and characterization of intermediate forms. *Biochemistry* **19** (1), 199–207.

Myers, J. K., Pace, C. N. and Scholtz, J. M. (1995) Denaturant *m* values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding. *Protein Science* **4** (10), 2138–2148.

Myles, D. C., Andrulis, P. J. and Whitesides, G. M. (1991) A transketolase-based synthesis of (+)-*exo*-brevicomin. *Tetrahedron Letters* **32** (37), 4835–4838.

Neet, K. E. and Koshland, D. E. Jr. (1966) The conversion of serine at the active site of subtilisin to cysteine: a "chemical mutation". *Proceedings of the National Academy of Sciences of the United States of America* **56** (5), 1606–1611.

Neet, K. E. and Timm, D. E. (1994) Conformational stability of dimeric proteins: quantitative studies by equilibrium denaturation. *Protein Science* **3** (12), 2167–74.

Nguyen, H., Martinez, B., Oganesyan, N. and Kim, R. (2004) An automated small-scale protein expression and purification screening provides beneficial information for protein production. *Journal of Structural and Functional Genomics* **5** (1-2), 23–27.

Nicholson, H., Anderson, D. E., Dao-pin, S. and Matthews, B. W. (1991) Analysis

of the interaction between charged side chains and the α-helix dipole using designed thermostable mutants of phage T4 lysozyme. *Biochemistry* **30** (41), 9816–9828.

Nicholson, H., Becktel, W. J. and Matthews, B. W. (1988) Enhanced protein thermostability from designed mutations that interact with α-helix dipoles. *Nature* **336** (6200), 651–656.

Nikkola, M., Lindqvist, Y. and Schneider, G. (1994) Refined structure of transketolase from *Saccharomyces cerevisiae* at 2.0 Å resolution. *Journal of Molecular Biology* **238** (3), 387–404.

Nilsson, U., Hecquet, L., Gefflaut, T., Guérard, C. and Schneider, G. (1998) Asp^{477} is a determinant of the enantioselectivity in yeast transketolase. *FEBS Letters* **424** (1-2), 49–52.

O'Brien, P. J. and Herschlag, D. (1999) Catalytic promiscuity and the evolution of new enzymatic activities. *Chemistry & Biology* **6** (4), R91–R105.

O'Donohue, M. J. and Kneale, G. G. (1994) Site-directed and site-saturation mutagenesis using oligonucleotide primers. *Methods in Molecular Biology* **30**, 211–225.

Ohkuri, T. and Yamagishi, A. (2003) Increased thermal stability against irreversible inactivation of 3-isopropylmalate dehydrogenase induced by decreased van der Waals volume at the subunit interface. *Protein Engineering* **16** (8), 615–621.

Ono, T., Negishi, K. and Hayatsu, H. (1995) Spectra of superoxide-induced mutations in the *lacI* gene of a wild-type and a *mutM* strain of *Escherichia coli* K-12. *Mutation Research* **326** (2), 175–183.

Onuffer, J. J. and Kirsch, J. F. (1995) Redesign of the substrate specificity of *Escherichia coli* aspartate aminotransferase to that of *Escherichia coli* tyrosine aminotransferase by homology modeling and site-directed mutagenesis. *Protein Science* **4** (9), 1750–1757.

Orengo, C. A., Michie, A. D., Jones, S., Jones, D. T., Swindells, M. B. and Thornton,

J. M. (1997) CATH - a hierarchic classification of protein domain structures. *Structure* **5** (8), 1093–1108.

Osterman, A. and Overbeek, R. (2003) Missing genes in metabolic pathways: a comparative genomics approach. *Current Opinion in Chemical Biology* **7** (2), 238–251.

Ostermeier, M., Nixon, A. E. and Benkovic, S. J. (1999a) Incremental truncation as a strategy in the engineering of novel biocatalysts. *Bioorganic & Medicinal Chemistry* **7** (10), 2139–2144.

Ostermeier, M., Nixon, A. E., Shim, J. H. and Benkovic, S. J. (1999b) Combinatorial protein engineering by incremental truncation. *Proceedings of the National Academy of Sciences of the United States of America* **96** (7), 3562–3567.

Paborsky, L. R., Dunn, K. E., Gibbs, C. S. and Dougherty, J. P. (1996) A nickel chelate microtiter plate assay for six histidine-containing proteins. *Analytical Biochemistry* **234** (1), 60–65.

Pace, C. N. (1986) Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods in Enzymology* **131**, 266–280.

Pace, C. N. (1992) Contribution of the hydrophobic effect to globular protein stability. *J Mol Biol* **226** (1), 29–35.

Pace, C. N., Alston, R. W. and Shaw, K. L. (2000) Charge-charge interactions influence the denatured state ensemble and contribute to protein stability. *Protein Science* **9** (7), 1395–1398.

Pace, C. N., Grimsley, G. R., Thomson, J. A. and Barnett, B. J. (1988) Conformational stability and activity of ribonuclease T_1 with zero, one, and two intact disulfide bonds. *Journal of Biological Chemistry* **263** (24), 11820–5.

Pace, C. N., Vajdos, F., Fee, L., Grimsley, G. and Gray, T. (1995) How to measure and predict the molar absorption coefficient of a protein. *Protein Science* **4** (11), 2411–2423.

Pace, C. N. and Vanderburg, K. E. (1979) Determining globular protein stability: guanidine hydrochloride denaturation of myoglobin. *Biochemistry* **18** (2), 288–292.

Page, M. J., Amess, B., Rohlff, C., Stubberfield, C. and Parekh, R. (1999) Proteomics: a major new technology for the drug discovery process. *Drug Discovery Today* **4** (2), 55–62.

Paoletti, F. (1983) Purification and properties of transketolase from fresh rat liver. *Archives of Biochemistry and Biophysics* **222** (2), 489–496.

Perry, L. J. and Wetzel, R. (1987) The role of cysteine oxidation in the thermal inactivation of T4 lysozyme. *Protein Engineering* **1** (2), 101–105.

Philippov, P. P., Shestakova, I. K., Tikhomirova, N. K. and Kochetov, G. A. (1980) Characterization and properties of pig liver transketolase. *Biochimica et Biophysica Acta* **613** (2), 359–369.

Pilone, M. S. and Pollegioni, L. (2002) D-amino acid oxidase as an industrial biocatalyst. *Biocatalysis and Biotransformation* **20** (3), 145–159.

Pire, C., Marhuenda-Egea, F. C., Esclapez, J., Alcaraz, L., Ferrer, J. and Bonete, M. J. (2004) Stability and enzymatic studies of glucose dehydrogenase from the Archaeon *Haloferax mediterranei* in reverse micelles. *Biocatalysis and Biotransformation* **22** (1), 17–23.

Plapp, B. V. (1995) Site-directed mutagenesis: a tool for studying enzyme catalysis. *Methods in Enzymology* **249**, 91–119.

Pokala, N. and Handel, T. M. (2001) Review: protein design–where we were, where we are, where we're going. *Journal of Structural Biology* **134** (2-3), 269–281.

Poland, D.C. and Scheraga, H.A. (1965) Statistical mechanics of noncovalent bonds in polyamino acids. VIII. Covalent loops in proteins. *Biopolymers* **3**, 379–399.

Ponder, J. W. and Richards, F. M. (1987) Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. *Journal of Molecular Biology* **193** (4), 775–791.

Pope, A. J., Haupts, U. M. and Moore, K. J. (1999) Homogeneous fluorescence readouts for miniaturized high- throughput screening: theory and practice. *Drug Discovery Today* **4** (8), 350–362.

Privalov, P. L. (1989) Thermodynamic problems of protein structure. *Annual Review of Biophysics and Biophysical Chemistry* **18**, 47–69.

Puchkaev, A. V., Koo, L. S. and Ortiz de Montellano, P. R. (2003) Aromatic stacking as a determinant of the thermal stability of CYP119 from *Sulfolobus solfataricus*. *Archives of Biochemistry and Biophysics* **409** (1), 52–58.

Redinbaugh, M. G. and Turley, R. B. (1986) Adaptation of the bicinchoninic acid protein assay for use with microtiter plates and sucrose gradient fractions. *Analytical Biochemistry* **153** (2), 267–271.

Reed, R. G. (1986) Location of long chain fatty acid-binding sites of bovine serum albumin by affinity labeling. *Journal of Biological Chemistry* **261** (33), 5619–5624.

Reidhaar-Olson, J. F., Bowie, J. U., Breyer, R. M., Hu, J. C., Knight, K. L., Lim, W. A., Mossing, M. C., Parsell, D. A., Shoemaker, K. R. and Sauer, R. T. (1991) Random mutagenesis of protein sequences using oligonucleotide cassettes. *Methods in Enzymology* **208**, 564–586.

Remmele, R. L. Jr., Nightlinger, N. S., Srinivasan, S. and Gombotz, W. R. (1998) Interleukin-1 receptor (IL-1R) liquid formulation development using differential scanning calorimetry. *Pharmaceutical Research* **15** (2), 200–208.

Richardson, J. S. and Richardson, D. C. (1988) Amino acid preferences for specific locations at the ends of α helices. *Science* **240** (4859), 1648–1652.

Risse, B., Stempfer, G., Rudolph, R., Schumacher, G. and Jaenicke, R. (1992) Characterization of the stabilizing effect of point mutations of pyruvate oxidase from *Lactobacillus plantarum*: protection of the native state by modulating coenzyme binding and subunit interaction. *Protein Science* **1** (12), 1710–1718.

Roberts, R. W. and Szostak, J. W. (1997) RNA-peptide fusions for the *in vitro* selection of peptides and proteins. *Proceedings of the National Academy of Sciences of the United States of America* **94** (23), 12297–12302.

Rücker, E., Schneider, G., Steinhäuser, K., Löwer, R., Hauber, J. and Stauber, R. H. (2001) Rapid evaluation and optimization of recombinant protein production using GFP tagging. *Protein Expression and Purification* **21** (1), 220–223.

Russell, R. J., Ferguson, J. M., Hough, D. W., Danson, M. J. and Taylor, G. L. (1997) The crystal structure of citrate synthase from the hyperthermophilic archaeon, *Pyrococcus furiosus* at 1.9 Å resolution. *Biochemistry* **36** (33), 9983–9994.

Saito, Y. and Wada, A. (1983) Comparative study of GuHCl denaturation of globular proteins. I. Spectroscopic and chromatographic analysis of the denaturation curves of ribonuclease A, cytochrome c, and pepsinogen. *Biopolymers* **22** (9), 2105–22.

Saitou, S., Ozawa, T. and Tomita, I. (1974) The purification and some properties of brewer's yeast apotransketolase. *FEBS Letters* **40** (1), 114–118.

Sali, D., Bycroft, M. and Fersht, A. R. (1988) Stabilization of protein structure by interaction of α-helix dipole with a charged side chain. *Nature* **335** (6192), 740–743.

Santoro, M. M. and Bolen, D. W. (1988) Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl α-chymotrypsin using different denaturants. *Biochemistry* **27** (21), 8063–8068.

Santoro, M. M. and Bolen, D. W. (1992) A test of the linear extrapolation of unfolding free energy changes over an extended denaturant concentration range. *Biochemistry* **31** (20), 4901–4907.

Schellman, J. A. (1987) The thermodynamic stability of proteins. *Annual Review of Biophysics and Biophysical Chemistry* **16**, 115–137.

Schenk, G., Duggleby, R. G. and Nixon, P. F. (1998) Properties and functions of the thiamin diphosphate dependent enzyme transketolase. *International Journal of Biochemistry and Cell Biology* **30** (12), 1297–1318.

Schmid, M. B. (2002) Structural proteomics: the potential of high-throughput structure determination. *Trends in Microbiology* **10** (10), S27–S31.

Schmidt-Dannert, C. and Arnold, F. H. (1999) Directed evolution of industrial enzymes. *Trends in Biotechnology* **17** (4), 135–136.

Schwehm, J. M., Fitch, C. A., Dang, B. N., García-Moreno, E. B. and Stites, W. E. (2003) Changes in stability upon charge reversal and neutralization substitution in staphylococcal nuclease are dominated by favorable electrostatic effects. *Biochemistry* **42** (4), 1118–1128.

Schwehm, J. M. and Stites, W. E. (1998) Application of automated methods for determination of protein conformational stability. *Methods in Enzymology* **295**, 150–170.

Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F. and Serrano, L. (2005) The FoldX web server: an online force field. *Nucleic Acids Res* **33** (Web Server issue), W382–8.

Selivanov, V. A., Kovina, M. V., Kochevova, N. V., Meshalkina, L. E. and Kochetov, G. A. (2003) Studies of thiamin diphosphate binding to the yeast apotransketolase. *Journal of Molecular Catalysis B-Enzymatic* **26** (1-2), 33–40.

Selivanov, V. A., Kovina, M. V., Kochevova, N. V., Meshalkina, L. E. and Kochetov, G. A. (2004) Kinetic study of the H103A mutant yeast transketolase. *FEBS Letters* **567** (2-3), 270–274.

Serrano, L., Bycroft, M. and Fersht, A. R. (1991) Aromatic-aromatic interactions and

protein stability. Investigation by double-mutant cycles. *Journal of Molecular Biology* **218** (2), 465–475.

Serrano, L. and Fersht, A. R. (1989) Capping and α-helix stability. *Nature* **342** (6247), 296–299.

Serrano, L., Horovitz, A., Avron, B., Bycroft, M. and Fersht, A. R. (1990) Estimating the contribution of engineered surface electrostatic interactions to protein stability by using double-mutant cycles. *Biochemistry* **29** (40), 9343–9352.

Serrano, L., Neira, J. L., Sancho, J. and Fersht, A. R. (1992) Effect of alanine versus glycine in α-helices on protein stability. *Nature* **356** (6368), 453–455.

Shakhnovich, E. I. (1997) Theoretical studies of protein-folding thermodynamics and kinetics. *Current Opinion in Structural Biology* **7** (1), 29–40.

Shaw, A. and Bott, R. (1996) Engineering enzymes for stability. *Current Opinion in Structural Biology* **6** (4), 546–550.

Shaw, A., Bott, R. and Day, A. G. (1999) Protein engineering of α-amylase for low pH performance. *Current Opinion in Biotechnology* **10** (4), 349–352.

Sime, J. T. (1999) Applications of biocatalysis to industrial processes. *Journal of Chemical Education* **76** (12), 1658–1661.

Simpson, F. J. (1960) Preparation and properties of transketolase from pork liver. *Canadian Journal of Medical Science* **38**, 115–124.

Sio, C. F., Riemens, A. M., van der Laan, J. M., Verhaert, R. M. and Quax, W. J. (2002) Directed evolution of a glutaryl acylase into an adipyl acylase. *European Journal of Biochemistry* **269** (18), 4495–4504.

Smith, D. B. (2000) Generating fusions to glutathione S-transferase for protein studies. *Methods in Enzymology* **326**, 254–270.

Smith, G. P. (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228** (4705), 1315–1317.

Soh, Y., Song, B. J., Jeng, J. and Kallarakal, A. T. (1998) Critical role of Arg^{433} in rat transketolase activity as probed by site-directed mutagenesis. *Biochemical Journal* **333** (Pt 2), 367–372.

Sorensen, K. and Brodbeck, U. (1986) Assessment of coating-efficiency in ELISA plates by direct protein determination. *Journal of Immunological Methods* **95** (2), 291–293.

Spiller, B., Gershenson, A., Arnold, F. H. and Stevens, R. C. (1999) A structural view of evolutionary divergence. *Proceedings of the National Academy of Sciences of the United States of America* **96** (22), 12305–12310.

Sprenger, G. A. and Pohl, M. (1999) Synthetic potential of thiamin diphosphate-dependent enzymes. *Journal of Molecular Catalysis B-Enzymatic* **6** (3), 145–159.

Sprenger, G. A., Schörken, U., Sprenger, G. and Sahm, H. (1995) Transketolase A of *Escherichia coli* K12. Purification and properties of the enzyme from recombinant strains. *European Journal of Biochemistry* **230** (2), 525–532.

Srere, P., Cooper, J. R., Tabachnick, M. and Racker, E. (1958) The oxidative pentose phosphate cycle. I. Preparation of substrates and enzymes. *Archives of Biochemistry and Biophysics* **74** (2), 295–305.

Stemmer, W. P. C. (1994) DNA shuffling by random fragmentation and reassembly: *In vitro* recombination for molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America* **91** (22), 10747–10751.

Steyaert, J. (1997) A decade of protein engineering on ribonuclease T1–atomic dissection of the enzyme-substrate interactions. *European Journal of Biochemistry* **247** (1), 1–11.

Stich, T. M. (1990) Determination of protein covalently bound to agarose supports using bicinchoninic acid. *Analytical Biochemistry* **191** (2), 343–346.

Straathof, A. J., Panke, S. and Schmid, A. (2002) The production of fine chemicals by biotransformations. *Current Opinion in Biotechnology* **13** (6), 548–556.

Strop, P., Marinescu, A. M. and Mayo, S. L. (2000) Structure of a protein G helix variant suggests the importance of helix propensity and helix dipole interactions in protein design. *Protein Science* **9** (7), 1391–1394.

Strumeyer, D. H., White, W. N. and Koshland, D. E. Jr. (1963) Role of serine in chymotrypsin action. Conversion of the active serine to dehydroalanine. *Proceedings of the National Academy of Sciences of the United States of America* **50**, 931–935.

Sundberg, S. A. (2000) High-throughput and ultra-high-throughput screening: solution- and cell-based approaches. *Current Opinion in Biotechnology* **11** (1), 47–53.

Taguchi, S., Ozaki, A. and Momose, H. (1998) Engineering of a cold-adapted protease by sequential random mutagenesis and a screening system. *Applied and Environmental Microbiology* **64** (2), 492–495.

Takayama, S., Livingston, P. O. and Wong, C. H. (1996) Synthesis of the melanoma-associated ganglioside 9-*O*-acetyl GD3 through regioselective enzymatic acetylation of GD3 using subtilisin. *Tetrahedron Letters* **37** (52), 9271–9274.

Takayama, S., McGarvey, G. J. and Wong, C. H. (1997) Microbial aldolases and transketolases: New biocatalytic approaches to simple and complex sugars. *Annual Review of Microbiology* **51**, 285–310.

Takeuchi, T., Nishino, K. and Itokawa, Y. (1986) Purification and characterization of, and preparation of an antibody to, transketolase from human red blood cells. *Biochimica et Biophysica Acta* **872** (1-2), 24–32.

Tanford, C. (1968) Protein denaturation. *Advances in Protein Chemistry* **23**, 121–282.

Tawfik, D. S. and Griffiths, A. D. (1998) Man-made cell-like compartments for molecular evolution. *Nature Biotechnology* **16** (7), 652–656.

Terpe, K. (2003) Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology* **60** (5), 523–533.

Thain, J. F (1967) *Principles of osmotic phenomena* Vol. 13 of *Royal Institute of Chemistry monographs for teachers* Royal Institute of Chemistry.

Thoma, R., Hennig, M., Sterner, R. and Kirschner, K. (2000) Structure and function of mutationally generated monomers of dimeric phosphoribosylanthranilate isomerase from *Thermotoga maritima*. *Structure* **8** (3), 265–276.

Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22** (22), 4673–4680.

Thompson, M. J. and Eisenberg, D. (1999) Transproteomic evidence of a loop-deletion mechanism for enhancing protein thermostability. *Journal of Molecular Biology* **290** (2), 595–604.

Tobin, M. B., Gustafsson, C. and Huisman, G. W. (2000) Directed evolution: the 'rational' basis for 'irrational' design. *Current Opinion in Structural Biology* **10** (4), 421–427.

Tsong, T. Y. (1976) Ferricytochrome *c* chain folding measured by the energy transfer of tryptophan 59 to the heme group. *Biochemistry* **15** (25), 5467–73.

Turner, N. J. (2000) Applications of transketolases in organic synthesis. *Current Opinion in Biotechnology* **11** (6), 527–531.

Ulmer, K. M. (1983) Protein engineering. *Science* **219** (4585), 666–671.

Valdar, W. S. and Thornton, J. M. (2001) Protein-protein interfaces: Analysis of amino acid conservation in homodimers. *Proteins* **42** (1), 108–124.

van den Burg, B., de Kreij, A., Van der Veek, P., Mansfeld, J. and Venema, G. (1999) Characterization of a novel stable biocatalyst obtained by protein engineering. *Biotechnology and Applied Biochemistry* **30**, 35–40.

van den Burg, B., Dijkstra, B. W., van der Vinne, B., Stulp, B. K., Eijsink, V. G. and Venema, G. (1993) Introduction of disulfide bonds into *Bacillus subtilis* neutral protease. *Protein Engineering* **6** (5), 521–527.

van den Burg, B. and Eijsink, V. G. H. (2002) Selection of mutations for increased protein stability. *Current Opinion in Biotechnology* **13** (4), 333–337.

van Den Heuvel, R. H., Fraaije, M. W., Ferrer, M., Mattevi, A. and van Berkel, W. J. (2000) Inversion of stereospecificity of vanillyl-alcohol oxidase. *Proceedings of the National Academy of Sciences of the United States of America* **97** (17), 9455–9460.

Vaschetto, M., Weissbrod, T., Bodle, D. and Guner, O. (2003) Enabling high-throughput discovery. *Current Opinion in Drug Discovery & Development* **6** (3), 377–383.

Vazquez-Duhalt, R. (1999) Cytochrome *c* as a biocatalyst. *Journal of Molecular Catalysis B-Enzymatic* **7** (1-4), 241–249.

Vetriani, C., Maeder, D. L., Tolliday, N., Yip, K. S., Stillman, T. J., Britton, K. L., Rice, D. W., Klump, H. H. and Robb, F. T. (1998) Protein thermostability above 100°C: A key role for ionic interactions. *Proceedings of the National Academy of Sciences of the United States of America* **95** (21), 12300–12305.

Vieille, C. and Zeikus, G. J. (2001) Hyperthermophilic enzymes: Sources, uses, and molecular mechanisms for thermostability. *Microbiology and Molecular Biology Reviews* **65** (1), 1–43.

Villafranca, J. J. and Axelrod, B. (1971) Heptulose synthesis from nonphosphorylated aldoses and ketoses by spinach transketolase. *Journal of Biological Chemistry* **246** (10), 3126–3131.

Vinogradov, A. A., Kudryashova, E. V., Grinberg, V. Y., Grinberg, N. V., Burova, T. V. and Levashov, A. V. (2001) The chemical modification of α-chymotrypsin with both hydrophobic and hydrophilic compounds stabilizes the enzyme against denaturation in water-organic media. *Protein Engineering* **14** (9), 683–689.

Voigt. C. A., Gordon, D. B. and Mayo, S. L. (2000) Trading accuracy for speed: A quantitative comparison of search algorithms in protein sequence design. *Journal of Molecular Biology* **299** (3), 789–803.

Voigt, C. A., Kauffman, S. and Wang, Z. G. (2000) Rational evolutionary design: the theory of *in vitro* protein evolution. *Advances in Protein Chemistry* **55**, 79–160.

Voigt, C. A., Mayo, S. L., Arnold, F. H. and Wang, Z. G. (2001) Computationally focusing the directed evolution of proteins. *Journal of Cellular Biochemistry* pp. 58–63.

Volkin, D. B. and Klibanov, A. M. (1987) Thermal destruction processes in proteins involving cystine residues. *Journal of Biological Chemistry* **262** (7), 2945–2950.

Vos, R. and Engelborghs, Y. (1994) A fluorescence study of tryptophan-histidine interactions in the peptide anantin and in solution. *Photochemistry and Photobiology* **60** (1). 24–32.

Waldburger, C. D., Schildbach, J. F. and Sauer, R. T. (1995) Are buried salt bridges important for protein stability and conformational specificity? *Nature Structural Biolology* **2** (2), 122–128.

Wang, W. and Malcolm, B. A. (1999) Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange^{TM} site-directed mutagenesis. *Biotechniques* **26** (4), 680–682.

Washburn, E. W. (1926) *International critical tables of numerical data, physics, chemistry and technology* Published for the National Research Council by the McGraw- Hill Book Company, New York.

Weber, P. C. and Salemme, F. R. (2003) Applications of calorimetric methods to drug discovery and the study of protein interactions. *Current Opinion in Structural Biology* **13** (1), 115–121.

Wells, J. A. and Powers, D. B. (1986) *In vivo* formation and stability of engineered disulfide bonds in subtilisin. *Journal of Biological Chemistry* **261** (14), 6564–6570.

Wikner, C., Meshalkina, L., Nilsson, U., Nikkola, M., Lindqvist, Y., Sundström, M. and Schneider, G. (1994) Analysis of an invariant cofactor-protein interaction in thiamin diphosphate-dependent enzymes by site-directed mutagenesis. Glutamic acid 418 in transketolase is essential for catalysis. *Journal of Biological Chemistry* **269** (51), 32144–32150.

Wintrode, P. L., Miyazaki, K. and Arnold, F. H. (2001) Patterns of adaptation in a laboratory evolved thermophilic enzyme. *Biochimica et Biophysica Acta-Protein Structure and Molecular Enzymology* **1549** (1), 1–8.

Won, C. M., Molnar, T. E., McKean, R. E. and Spenlehauer, G. A. (1998) Stabilizers against heat-induced aggregation of RPR 114849, an acidic fibroblast growth factor (aFGF). *International Journal of Pharmaceutics* **167** (1-2), 25–36.

Xu, Y., Mayne, L. and Englander, S. W. (1998) Evidence for an unfolding and refolding pathway in cytochrome *c*. *Nature Structural Biology* **5** (9), 774–778.

Yano, T., Oue, S. and Kagamiyama, H. (1998) Directed evolution of an aspartate aminotransferase with new substrate specificities. *Proceedings of the National Academy of Sciences of the United States of America* **95** (10), 5511–5515.

Yonetani, T. (1965) Studies on cytochrome *c* peroxidase. II. Stoichiometry between enzyme, H_2O_2, and ferrocytochrome *c* and enzymic determination of extinction coefficients of cytochrome *c*. *Journal of Biological Chemistry* **240** (11), 4509–4514.

Zaccolo, M., Williams, D. M., Brown, D. M. and Gherardi, E. (1996) An approach to

random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. *Journal of Molecular Biology* **255** (4), 589–603.

Zale, S. E. and Klibanov, A. M. (1986) Why does ribonuclease irreversibly inactivate at high temperatures? *Biochemistry* **25** (19), 5432–5444.

Zhang, J. H., Dawes, G. and Stemmer, W. P. (1997) Directed evolution of a fucosidase from a galactosidase by DNA shuffling and screening. *Proceedings of the National Academy of Sciences of the United States of America* **94** (9), 4504–4509.

Zhang, X. J., Baase, W. A., Shoichet, B. K., Wilson, K. P. and Matthews, B. W. (1995) Enhancement of protein stability by the combination of point mutations in T4 lysozyme is additive. *Protein Engineering* **8** (10), 1017–1022.

Zhao, H., Giver, L., Shao, Z., Affholter, J. A. and Arnold, F. H. (1998) Molecular evolution by staggered extension process (StEP) *in vitro* recombination. *Nature Biotechnology* **16** (3), 258–261.

Zheng, L., Baumann, U. and Reymond, J. L. (2004) An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Research* **32** (14), e115.

Appendix A

Preliminary microplate unfolding study

A.1 Introduction

The results reported in this appendix provides the motivation for selecting cytochrome *c* and BSA for analysis in the final study as is presented in Chapter 3. A preliminary study was conducted to determine the feasibility of using combined auto-titrating and microplate reader equipment to produce protein unfolding curves at micro-scale. This work was conducted mainly by Ana Cosme with assistance from Jean Aucamp and Dr. Paul Dalby. Apart from the results shown here, Ana Cosme also looked into automated unfolding experiments and optimisation of parameters such as protein concentration and sample volumes. These results are not deemed necessary for the purpose of this appendix and therefore excluded.

From literature, we identified several proteins with well studied unfolding properties as well as published thermodynamic data and unfolding curves. A list was compiled of only proteins that are readily available commercially as reasonably pure preparations (Table A.1). A purity of 90 % was considered as a reasonable cut-off criteria in the light that high-throughput protein purification protocols may yield similar purities. The purities of α-chymotrypsin and α-chymotrypsinogen, though not stated by the supplier, were con-

sidered good enough based on the protein purification process used (four crystallisation steps). The literature pertaining to unfolding data and conditions for these proteins are summarised in Table A.2.

A.2 Materials and Methods

A.2.1 Protein concentration determination

The proteins studied in the present report, the abbreviation codes of their names, purities, molar absorption coefficients and molecular weights are listed in Table A.1. Protein concentrations were determined using the molar absorption coefficients and molecular weights for each protein and absorbance measurements made using a UV/Vis Spectrometer from ATi UNICAM UV2. The molar absorption coefficients used were obtained either from the reference articles or by calculation from equation A.1 (Pace *et al.*, 1995) where $\epsilon_{(280)}$ represents the molar extinction coefficient at 280 nm in $(M^{-1} \cdot cm^{-1})$ and # the number of the respective residues in the protein.

$$\epsilon_{(280)} = (\#Trp)(5,500) + (\#Tyr)(1,490) + (\#cystine)(125) \tag{A.1}$$

A.2.2 Reduction and dialysis of cytochrome *c*

The heme iron in equine and bovine heart cytochrome *c* was oxidised to Fe^{3+} by addition of potassium ferricyanide after which it was removed from the solutions by overnight dialysis against buffer. Potassium ferricyanide was purchased from Sigma Chemical Co. Ltd, with a purity of 100 %. The fully oxidised state of the cytochrome *c* was confirmed before each unfolding experiment by observing the absence of a peak at 550 nm in the absorbance spectrum.

Table A.1: Summary of literature survey for commercial proteins used in unfolding studies.

Protein	Source	Abbrev.	Sigma Product nr	Purity	ϵ	MW (g/mol)
Serum Albumin	Bovine	BSA	A0281	99 %	43 961 (ϵ_{280})	≈66 350
Serum Albumin (fat free)	Bovine	BSA-ff	A3912	96-99 %	43 961 (ϵ_{280})	66 350
α-Chymotrypsin	Bovine pancreas	ChySin	C4129	-[a]	49 614 (ϵ_{280})	25 185
α-Chymotrypsinogen	Bovine pancreas	ChyGen	C4879	-[a]	28 460 (ϵ_{280})	25 666
Cytochrome *c*	Equine heart	eCyt*c*	C7752	97 %	106 000 (ϵ_{280})	12 384
Cytochrome *c*	Bovine heart	bCyt*c*	C2037	100 %	106 000 (ϵ_{410})	12 327
Lysozyme	Hen egg white	LZM	L6876	95 %	37 970 (ϵ_{280})	14 314
Ribonuclease A	Bovine pancreas	RBN	R5500	90 %	9 880 (ϵ_{280})	13 690
Trypsin	Bovine pancreas	TRS	T6763	90 %	37 650 (ϵ_{280})	23 800

[a]Protein purity not reported by Sigma Chemical Co. Ltd., but deemed to be of reasonable good purity from purification that includes four crystallisation steps.

A.2.3 Inhibition of proteolytic enzymes

PMSF with a purity of 99 %, from Sigma Chemical Co., was used to prepare phenylmethanesulphonyl chymotrypsin (PMSF-ChySin) by addition of 100 µl of 0.2 M PMSF, dissolved in acetonitrile, per 50 mg of protein in buffer. Reactions for PMSF-ChySin were allowed to proceed for at least 30 min (Santoro and Bolen, 1988). For inhibition of TRS and ChyGen, 100 µl of 0.2 M PMSF, dissolved in acetonitrile was added to 10 mg of protein dissolved in buffer. The reactions for producing PMSF-ChyGen and PMSF-TRS were allowed to proceed for at least 3 hr.

A.2.4 Analysis of protein homogeneity

SDS-PAGE was used to verify protein homogeneity and the efficacy of inhibiting proteolytic activity of ChySin and TRS. Cast gels contained 15% and 6% acrylamide in the resolving gel and stacking gels respectively.

Table A.2: Summary of unfolding conditions for selected proteins obtained from literature.

Protein	Buffer	pH	Denaturant	[Protein]	Unfold	Excitation	Emission	Reference
BSAff	60 mM phosphate	7.0	Urea	42	-	282 nm	340 nm	(Ahmad and Qasim, 1995)
BSA	60 mM phosphate	7.0	Urea	200	-	293 nm	340 nm	(Khan *et al.*, 1987)
ChySin	50 mM acetate, 144 mM NaCl	4.0	GdmCl, urea	< 30	1h	295 nm	327 nm	(Santoro and Bolen, 1988)
ChyGen	50 mM acetate, 144 mM NaCl	4.0	GdmCl	< 30	overnight	-	-	copy conditions of ChySin assay; (Kardos *et al.*, 1999)
eCyt *c*	100 mM Tris	7.0	GdmCl	25	-	280 nm	340 nm	(Tsong, 1976) (Brems *et al.*, 1982)
bCyt *c*	100 mM Tris	7.0	GdmCl	-	-	280 nm	340 nm	copied method of (Tsong, 1976)
RBN	50 mM Tris	8.0	GdmCl	110 $\mu g \cdot ml_{-1}$	5h (10°C)	280 nm	320 nm	(Juneja and Udgaonkar, 2002)
TRS	50 mM Tris	8.0	GdmCl	200	3h	295 nm	327 nm	(Amiza and Apenten, 1996)
LZM	100 mM Citrate 100 mM NaCl	2.9	Urea urea	46	-	280 nm	360 nm	(IbarraMolero and SanchezRuiz, 1997)

A.2.5 Conditions of unfolding

The conditions used in the unfolding experiments were based on methods published in papers for each protein. The conditions, denaturants, buffers, excitation and emission wavelengths are listed in Table A.2. The pH of the buffers was measured using a Mettler Toledo pH meter, with a combination electrode.

The excitation and emission wavelengths used in this project were subject to the availability of filters provided with equipment. A 280 nm filter was available for excitation, while only two emission filters, 320 nm and 340 nm were available. The 340 nm emission filter was used for all proteins, except for ChyGen and RBN, where a 320 nm emission filter was used. The measurements were performed on a microplate reader: FLUOstar OPTIMA, from BMG Labtechnologies Ltd. In this study Sarstedt No 82.1581 clear polystyrene flat-bottom 96-well microwell plates from Sarstedt (NC, USA) were used.

A.2.6 Fixed volume determination of transition curves

Denaturant solutions of decreasing concentration (concentration range) were prepared in microplates through dilution of a stock denaturant solution. Equal volumes of protein solution were added to the wells of the microwell plate, using the pipetting syringe pump. The protein was left to equilibrate in the microwell plate, covered with parafilm to minimise evaporation, until fluorescence was measured.

A.2.7 Automated determination of transition curve

The software of the microplate reader was insufficient to perform the automated procedure of unfolding of the proteins over time. The software of the auxiliary stacker equipment was installed on the computer controlling the microplate reader. This allowed programming of automated unfolding procedures, by creating scripts (example on page 200) that could be adjusted for each particular unfolding experiment. The protein solution was injected in a microwell plate using the syringe of the FLUOstar OPTIMA and a script then executed to control sequential addition of denaturant, automating the unfolding process.

A.2.8 Unfolding curves analysis

The analysis of the unfolding curves was done as described by Santoro and Bolen (1988) and explained in Chapter 3. The normalised data were fitted to a two-state model according to Equation 2.3 using SigmaPlot 8.0.

A.3 Results

A.3.1 Unfolding curves of proteins

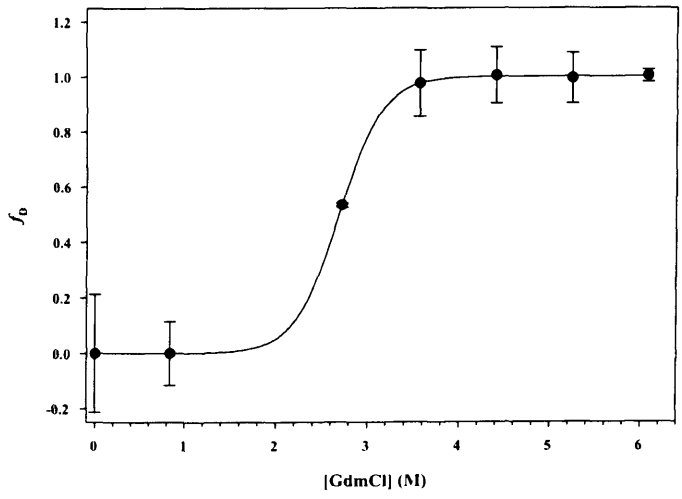


Figure A.1: Conformational transition curve for RNAse A (RBN) obtained with the fixed volume method. 140 µg of RNAse A was loaded per well. Wells were filled to a final volume of 250 µl with combinations of 50 mM Tris buffer, pH 8.0 containing 0 M and 8.0 M GdmCl to give the correct final denaturant concentration. Each data point was measured in triplicate.

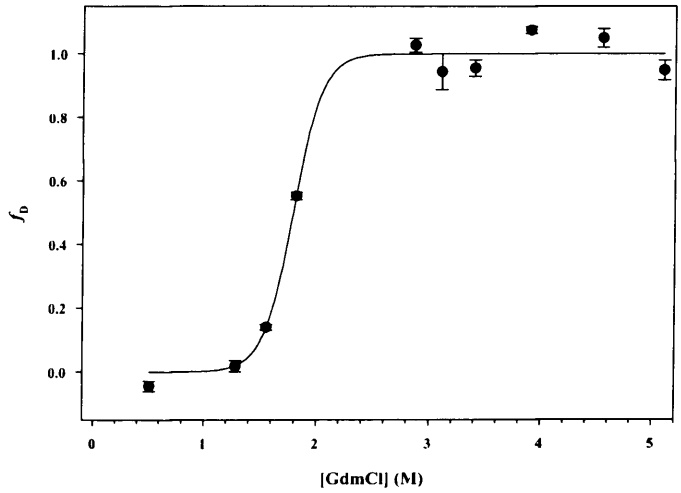


Figure A.2: Conformational transition curve for chymotrypsin (ChySin) obtained with the fixed volume method. 130 µg of ChySin was loaded per well. Wells were filled to a final volume of 250 µl with combinations of 50 mM acetate buffer, pH 4.0 containing 0 M and 8.0 M GdmCl to give the correct final denaturant concentration. Each data point was measured in triplicate.

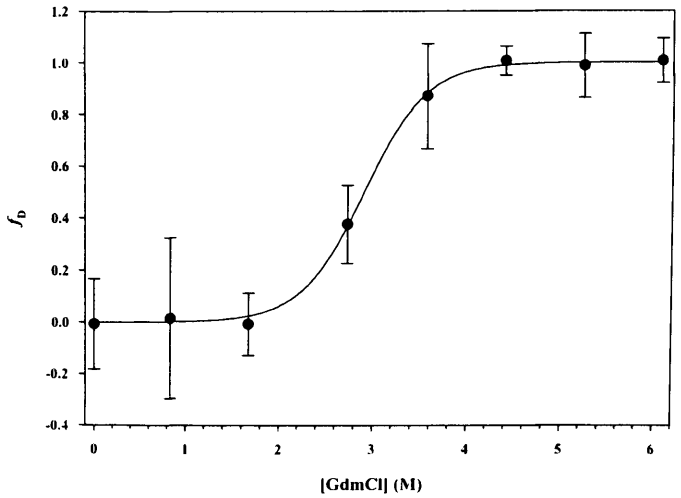


Figure A.3: Conformational transition curve for chymotrypsinogen (ChyGen) obtained with the fixed volume method. 20 µg of ChyGen was loaded per well. Wells were filled to a final volume of 250 µl with combinations of 50 mM acetate buffer, pH 4.0 containing 0 M and 8.0 M GdmCl to give the correct final denaturant concentration. Each data point was measured in triplicate.

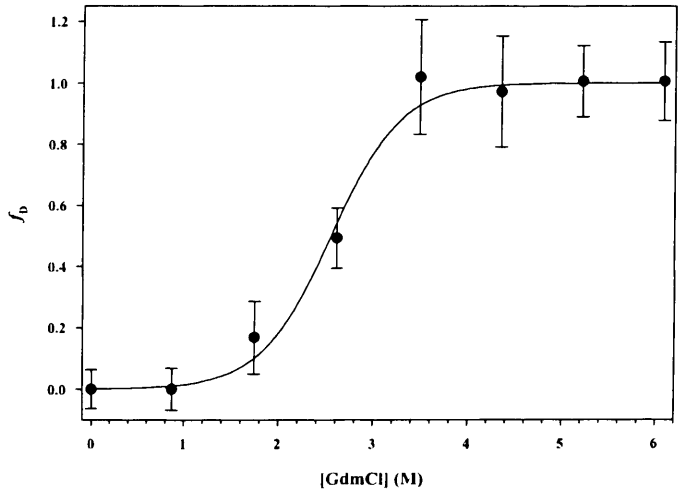


Figure A.4: Conformational transition curve for bovine cytochrome *c* (bCyt*c*) obtained with the fixed volume method. 140 μg of bCyt*c* was loaded per well. Wells were filled to a final volume of 250 μl with combinations of 100 mM Tris buffer, pH 7.0 containing 0 M and 8.0 M GdmCl to give the correct final denaturant concentration. Each data point was measured in triplicate.

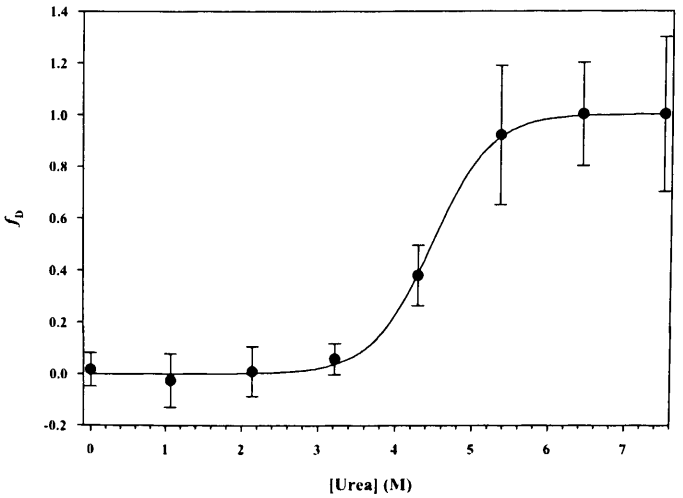
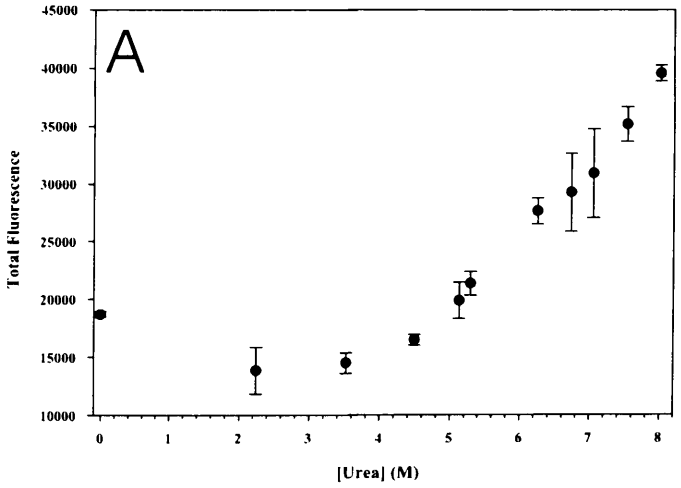


Figure A.5: Conformational transition curve for fat free bovine serum albumin (BSA-ff) obtained with the fixed volume method. 194 μg of BSA-ff was loaded per well. Wells were filled to a final volume of 300 μl with combinations of 60 mM sodium phosphate buffer, pH 7.0 containing 0 M and 9.0 M urea to give the correct final denaturant concentration. Each data point was measured in triplicate.



B

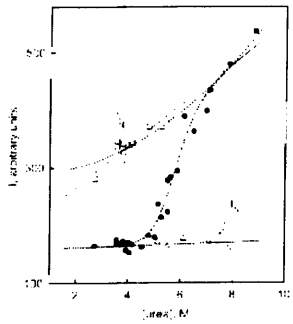


FIGURE 5: The urea-induced equilibrium unfolding of HEW lysozyme in citrate buffer, pH 2.9, as monitored by fluorescence emission. Three data sets are included in this figure: (●) fluorescence intensity values measured at equilibrium, (□) intensity values for the unfolded state calculated as the zero-time extrapolations of the fluorescence-detected refolding kinetic profiles recorded after denaturant concentration jumps from strongly unfolding conditions (8 M urea, citrate buffer, pH 2.9), (△) intensity values for the native state calculated as the zero-time extrapolations of the fluorescence-detected unfolding profiles recorded after denaturant concentration jumps from native conditions (citrate buffer, pH 2.9, in the absence of urea). In all cases (●, □, and △) the protein concentration is 0.945 mg mL. The continuous lines represent the best simultaneous fit of eq 7 and linear equations for the unfolded and native fluorescence baselines, to the experimental I, I_U, and I_N data, respectively. The results of the same global fit, but assuming a quadratic unfolded baseline, is shown with dotted lines.

Figure A.6: (**A**) Conformational transition curve for lysozyme (LZM) obtained with the fixed volume method. 100 µg of LZM was loaded per well. Wells were filled to a final volume of 300 µl with combinations of 100 mM acetate buffer, pH 2.9 containing 0 M and 9.0 M urea to give the correct final denaturant concentration. Each data point was measured in triplicate. (**B**) Conformational transition curve for lysozyme (LZM) obtained with identical conditions and reprinted from (IbarraMolero and SanchezRuiz, 1997). The unfolding curve obtained in our experiment is very similar to published unfolding curve with a poorly defined post-transitional baseline.

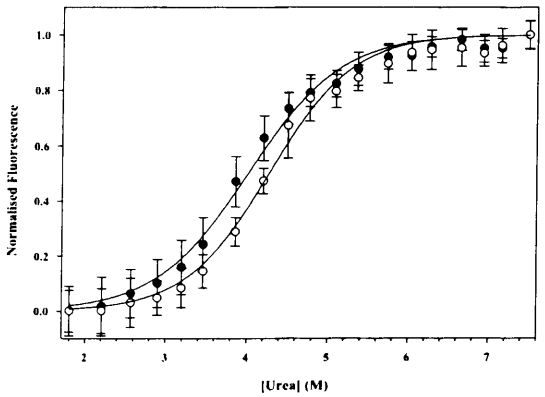


Figure A.7: Conformational transition curves for fat free bovine serum albumin (BSA-ff) and commercial BSA with unknown amount of fat (BSA) obtained with the serial addition method. 40 µl of 1.5 mg·ml^{-1} of BSA protein was loaded per well. 3 - 10 µl additions of 60 mM sodium phosphate buffer, pH 7.0 containing 9.0 M urea were made to produce the serial denaturant concentration range. Signal average was measured from three samples.

Table A.3: Comparison of literature stability data with experimental data obtained using the fixed volume method.

	Reported values		Experimental values[a]	
	$C_{1/2}$ [M]	m_G (kcal·mol^{-1}·M^{-1})	$C_{1/2}$ [M]	m_G (kcal·mol^{-1}·M^{-1})
[b]eCyt *c*	2.68	4.10	1.85 ± 0.02	3.02 ± 0.23
[c]bCyt *c*	2.6	-	2.35 ± 0.04	2.04 ± 0.23
[d]ChySin	2.0 ± 0.02	4.19 ± 0.53	1.75 ± 0.03	1.70 ± 0.14
[e]ChyGen	-	-	2.91 ± 0.03	7.50 ± 0.68
[f]RBN	3.56	2.3		
[b]RBN	2.97	3.45	2.69 ± 0.08	10.4 ± 7.5
[g]RBN	3.60	1.27		
[b]LZM	4.24	2.33	-	-

[a] For the determination of m_G-values, a temperature of 20 °C is used.
[b] Data calculated from spectroscopic and chromatographic data (Saito and Wada, 1983).
[c] Data calculated from fluorescence emission spectra (Brems *et al.*, 1982).
[d] Unfolding transition followed by fluorescent measurements at pH 4.0 (Santoro and Bolen, 1988).
[e] Unfolding conditions same as for ChySin and compared to ChySin stability data.
[f] Unfolding transition followed by intrinsic tyrosine fluorescence, exciting 280 nm, emission 320 nm (Juneja and Udgaonkar, 2002).
[g] Unfolding transition followed by observing changes in the UV spectra, at pH 7.0 (Ahmad and Bigelow, 1982).

A.3.2 SDS-PAGE analysis of protein samples

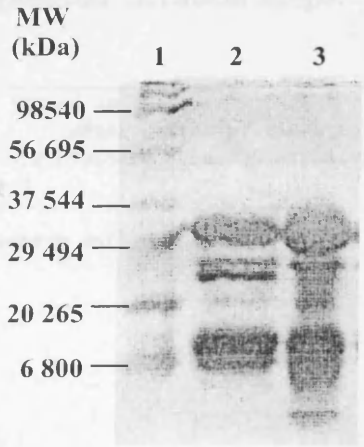


Figure A.8: SDS-Polyacrylamide gel electrophoresis of (2) bovine trypsin, 40 mg protein; and (3) PMSF-inhibited bovine trypsin after dialysis. A 15% resolving gel and 6% stacking gel. (1) is the molecular weight reference. Trypsin (1) was dissolved directly in denaturing and boiling electrophoresis buffer in order to prevent further auto-proteolysis.

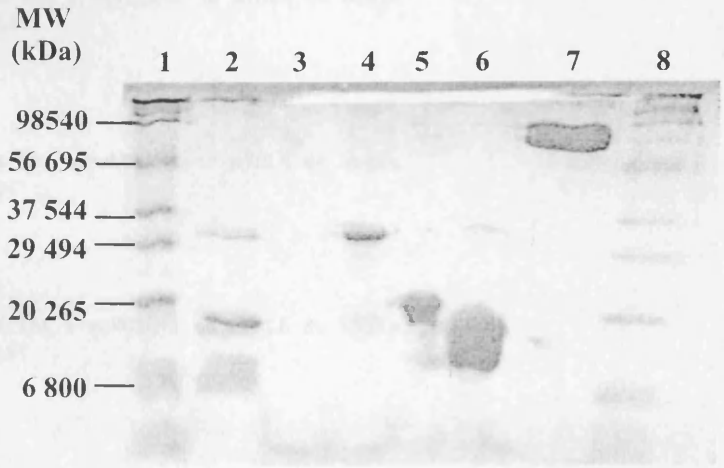


Figure A.9: SDS-Polyacrylamide gel electrophoresis, using a 15% resolving gel and 6% stacking gel, of (2) chymotrypsin, (4) chymotrypsinogen, (5) ribonuclease A, (6) bovine cytochrome *c* and (7) BSA. (1) and (9) are the molecular weight references. Chymotrypsin (2) was dissolved directly in denaturing and boiling electrophoresis buffer in order to prevent further auto-proteolysis.

A typical example of script code used for running automated unfolding experiments employing sequential titration steps.

```
 VOL1:=4          ;enter volume to inject in the
first injection group nVOL1:=4         ;enter number of times to
inject VOL1 VOL2:=6 nVOL2:=2 VOL3:=12 nVOL3:=4 VOL4:=20 nVOL4:=2
VOL5:=30 nVOL5:=1 VOL6:=42 nVOL6:= 2

ID1:= "0.200 mg/ml" ID2:="<1+>, lysozym unf"

for i:=1 to nVOL1 do begin
  F_run "INJECT VOL1"
  F_run "SHAKE"
  F_run "Measure"
  ID3:= "<time>"
end

nVOL2:=nVOL1+ nVOL2 for i:=nVOL1+1 to nVOL2 do begin
  F_run "INJECT VOL2"
  F_run "SHAKE"
  F_run "Measure"
end

nVOL3:=nVOL2+ nVOL3 for i:=nVOL2+1 to nVOL3 do begin
  F_run "INJECT VOL3"
  F_run "SHAKE"
  F_run "Measure"
end

nVOL4:=nVOL3+ nVOL4 for i:=nVOL3+1 to nVOL4 do begin
  F_run "INJECT VOL4"
  F_run "SHAKE"
  F_run "Measure"
end

nVOL5:=nVOL4+ nVOL5 for i:=nVOL4+1 to nVOL5 do begin
  F_run "INJECT VOL5"
  F_run "SHAKE"
  F_run "Measure"
end

nVOL6:=nVOL5+ nVOL6 for i:=nVOL5+1 to nVOL6 do begin
  F_run "INJECT VOL6"
  F_run "SHAKE"
  F_run "Measure"
end
```

A.4 Discussion and Conclusion

Proteolytic enzymes are important in the biocatlysis industry and α-chymotrypsin would be an excellent candidate for the proof of principle study, bearing in mind the intended purpose of the high-throughput application in biocatalysis and formulation. In addition, $\Delta C_{1/2}$ values could be obtained when comparing $C_{1/2}$ values with the inactive precursor, chymotrypsinogen or after activation and stabilisation with the addition of Ca^{2+} ions. This could be used to show the method to be suitable for formulation studies. However, SDS-PAGE indicated that both ChySin and PMSF-ChySin protein preparations contained truncated forms of the protein due to auto-proteolysis. The PMSF-chymotrypsin fraction would have to be purified first, using size-exclusion chromatography to produce homogenous preparations that could be used to obtain accurate $C_{1/2}$ values. (Santoro and Bolen, 1988) followed the unfolding using an emission wavelength of 327 nm. Though 320 nm is closer to 327 nm, the fluorescence emission of chymotrypsin was followed at 340 nm instead. This is because no change in fluorescence signal was observable at 320 nm for the unfolding process.

Bovine trypsin did not produce a significant ΔF signal at either 320 or 340 nm. Optimal fluorescence change is observed at 327 nm for which we do not have a wavelength filter. Significant amounts of both trypsin and PMSF-trypsin were truncated due to auto-proteolysis. This means that size-exclusion chromatography of PMSF-trypsin has to be done to obtain homogenous protein preparations, therefor not making a good candidate.

RNB is a good candidate since it follow a simple two-state unfolding mechanism, is pure and unfolding can be monitored with the available 320 nm filter. It would however be better if it could be compared to a homologue such as RNAse T1 for instance so that $\Delta C_{1/2}$ can be measured. RNAse T1 could not be obtained commercially and is not conducive to fluorescence unfolding since it contains no tryptophan residues.

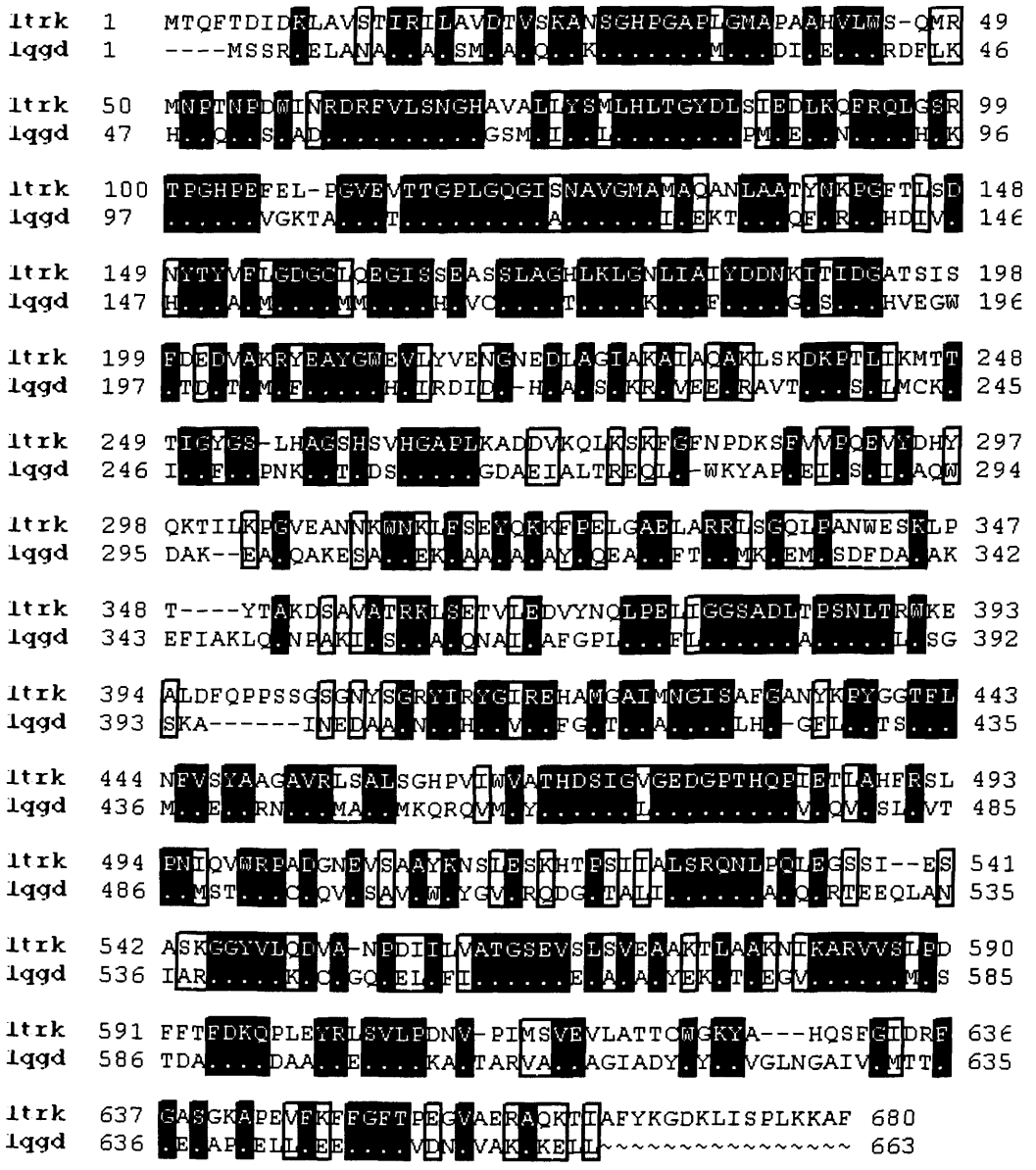
Our literature search only identified one study where fluorescence was used to monitor unfolding of lysozyme. Due to the extreme pH of 2.9 and atypical unfolding curve obtained under these conditions it was considered best not to use lysozyme for 'proof of principle'

studies. It should however be noted that our method produced a similar unfolding curve under the same conditions (Figure A.6).

The heme part of bovine and equine cytochrome *c* were oxidised and protein then dialysed to remove the maximum of the oxidising agent, potassium ferricyanide. Oxidised cytochrome *c* follows an apparent two-state unfolding and is obtained as relative pure preparations. The bovine and equine variants have different $C_{1/2}$ values and ideal for testing the ability of the screen to distinguish variants with changes in stability. In addition cytochrome *c* is used as biocatalyst in the industry making it a relevant and highly suitable candidate for the study.

BSA has the ability to bind to fatty acids and its stability increases with increase in fatty acid content. BSA with an unknown amount of fatty acid and fat-free BSA were compared and the screen was able to detect a change in unfolding midpoints. BSA is occasionally used to stabilise protein/enzyme formulations making it relevant for this study. In addition formulation studies often make use of ligands to improve the stability of therapeutic proteins. Successful detection of ligand induced stabilisation suggests that this screen may also be used successfully in similar studies on therapeutic proteins.

Appendix B



Sequence alignment of yeast (1trk) and *E. coli* (1qgd) transketolases.