Characterisation of the structural stability of transketolase under biocatalytically relevant conditions

A thesis submitted to University College London
For the degree of

DOCTOR OF PHILOSOPHY

By

Rubén Julio Martínez Torres

The Advanced Centre for Biochemical Engineering
Department of Biochemical Engineering
University College London
Torrington Place
London
WC1E 7JE

2008

Abstract

The enzyme transketolase (TK; E.C. 2.2.1.1) from *Escherichia coli* occupies a pivotal place in metabolic regulation. TK catalyses the interconversion of sugars by transferring a two-carbon ketol unit from a ketose donor substrate to an aldolase acceptor substrate. It is also an important biocatalyst in stereo-specific carbon-carbon bond synthesis with potential industrial application for the synthesis of pharmaceuticals, agrochemicals and fine chemicals. Although many useful reactions have been reported for TK, many of the substrates and products are unstable or insoluble at the pH or temperature for which the enzyme has optimum activity. Understanding the structural stability of transketolase under bioprocess conditions will improve our capacity to comprehend and ultimately to engineer it to make it work in a broader range of pH or temperature to potentially help in the reduction of process time and to increase the quality and solubility of products. In this research I characterised the early events on the urea denaturation pathway of *E. coli* transketolase, providing new insights into the mechanisms of enzyme deactivation that occur under biocatalytic conditions.

Equilibrium denaturation measurements by fluorescence intensity and circular dichroism (CD), combined with size-exclusion and dynamic light scattering studies, have revealed three transitions in the denaturation pathway for holo-TK. The first step, at low urea concentration corresponds to the local restructuring of the thiamine diphosphate (TPP) binding-sites. Next, the dissociation of the TPP cofactors and partial loss of secondary structure produces a form, which is most consistent with a partially denatured dimeric enzyme. While the enzyme is deactivated initially by changes in structure associated with the cofactors, this event does not release the cofactor from the enzyme, consistent with the intermediate formed during the reconstitution of holo-TK from apo-TK. Improvement of biocatalytic processes using TK over prolonged reaction times would, therefore, need to address the formation of this

cofactor-associated intermediate state. Equivalent results were also observed with a high throughput microplate-based fluorescence method that uses less enzyme and time.

The equilibrium denaturation of holo and apo-TK at different temperatures and pH was also investigated for further insights into the enzyme stability and to provide a benchmark for assessing any future enzyme variants with altered pH or temperature optima. In an effort to enhance the stability of the enzyme I subsequently used bioinformatical, statistical and multivariate analyses of protein sequences and associated properties to determine the most likely residues to affect temperature and pH optima in biocatalysis. The outcome of this first parametric statistical analysis (Pearson's *r* test) rendered 20 different points. Promising mutation points were selected based on the correlation coefficient (*r*) results taken into account a level of significance $\alpha = 0.05$. TK *E. coli* selected points were then mutated and screened by site-saturated mutagenesis (SSM) and automated techniques respectively. Finally alternative statistical correlation methods were examined, including a non-parametric statistical analysis (Kendall's τ or tau test), principal component analysis (PCA) and partial least square (PLS), for their potential to generate TK mutant variants with either enhanced pH or temperature stability.

Acknowledgements

I would like to express my gratitude to:

My supervisor, Dr. Paul A. Dalby, whose expertise, understanding, and patience, guided me through to my PhD making it a very enjoyable trip. I appreciate his vast knowledge in the protein field and his assistance in writing reports, papers and scholarship applications.

To all the beautiful Foster Court people and all my colleagues at UCL I met during my studies. Particularly to Janahan Paramesvaran, who helped me to understand so many things, specially the English language!

In particular, I recognise that this research would not have been possible without the financial assistance of CONACyT (The Mexican National Council for Science and Technology) Scholarships fund, and express my gratitude to this agency.

I want to thank to all my friends from Mexico for all their support.

Finally, I would also like to thank my Parents, my brothers and sisters for the support they provided me through my PhD studies. This would not have been possible without their love and understanding.

I, Julio Martínez-Torres, declare that the dissertation, submitted in partial fulfilment of the requirements for the degree of Doctorate of Philosophy represents my own work and has not been previously submitted to this or any other institution for any degree, diploma or other qualification.

Signature_____

Content

List of Figures

List of Tables

Abbreviations

Ala	Alanine
Amp^+	ampicilin resistance
Apo-TK	Apo-transketolase
Arg	arginine
Asn	asparagine
Asp	aspartic acid
BLAST	basic local alignment search tools
BRENDA	BRaunschweig ENzyme DAtabase
CAPS	N-cyclohexyl-3-aminopropanesulfonic acid
CD	Circular dichroism
Cys	cysteine
ddH_2O	d destilate H_2O
DLS	dynamic light scattering
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide A/C/G/T triphosphate
DTT	dithiothreitol
EC	Enzyme commision
EDTA	ethylenediaminetetraacetic acid
FASTA	text-based format
GA	glycoaldehyde
Gln	glutamine
Glu	glutamic acid
Gly	Glycine
HCL	hydrochloric acid
His	histidine
Ho	null hypothesis
Holo-TK	Holo-transketolase
HPA	hydroxypyruvate
HPLC	high-performance liquid chromatography
Ile	isoleucine
J	Pearson's skewness coefficient
Leu	Leucine
Li	Lithium
Lys	lysine
Met	methionine
Mg	magnesium
$MgCl_2$	magnesium chloride
MRE	mean residue ellipticity
MW	molecular weights
NaCl	Sodium chloride
NaOH	Sodium hydroxide
NCBI	National Center for Biotechnology Information
Ni	Nickel
NTA	nitrilotriacetic acid

PCA	principal component analysis
PCR	polymerase chain reaction
PDB	proteine data bank
Phe	phenylalanine
PLS	partial least square
PLSR	partial least squares regression
PP	pyrophosphate-binding domain
PP	*PPi-binding domain*
Pro	Proline
Pyr	pyridinium-binding domain
QCM	QuickChange™ Site-Directed Mutagenesis System
r	Pearson's correlation coefficient
SCA	statistical coupling analysis
SDS	Sodium dodecyl sulfate
SDS PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SDSM	site-directed saturation mutagenesis
SEC	size exclusion chromatography
Ser	serine
τ	Kendall's correlation coefficient
Taq	*Thermos Aquaticus*
TBE	Tris/Borate/EDTA
TFA	trifluoroacetic acid
Thr	threonine
TK	transketolase
TPP	thiamine pyrophosphate
Tris	hydroxymethylaminomethane
Trp	tryptophan
Tyr	tyrosine
Val	valine
wt	wild-type
wt-RT	wild-type at room temperature

Chapter 1

1 Introduction

1.1 Biocatalysis

Biocatalysis has received increasing attention in recent years as a powerful tool available to the synthetic chemist (Rozzell, 1999; Buckland et al., 2000; Thayer, 2001). While the use of biological catalysts has become widespread in the laboratory, large scale implementation requires the development of process techniques and a framework within which they can be applied (Woodley et al., 1996; Brocklebank et al., 1999). The range of potential application is broadening on all frontiers with increasing complexity of target structures, while the inherent limitations of enzymes are being taken into account and approaches are being devised to overcome these restrictions (Fessner, 1998)

In this field progress is expected to be fuelled by whole genome sequencing and by an almost explosive growth in structural knowledge for all important classes of enzymes in complex with substrates and analogs, to allow a more detailed understanding of substrate recognition and catalytic function. C-C bond forming enzymes are now candidates for rational engineering and directed molecular evolution (Morris et al., 1996; Woodley et al., 1996), which could lead to improved catalytic and physical properties and the development of novel catalytic functions, and in particular will probably soon lift the current restriction from narrow donor specificities (Fessner, 1998).

Biocatalytic C-C bond formation has become invaluable for the asymmetric synthesis of complex multifunctional molecules as for example the production of synthetic sugar molecules (Brocklebank et al., 1999). However, for process design and development, an intimate knowledge base is required,

on subjects such as substrate tolerance for donor and acceptor components or on the influence of substrate structure on enantioselectivity and diastereoselectivity of C-C bonding (Littlechild et al., 1995). It is evident; however, that the technology is well accepted in the chemical community and is beginning to fulfil its potential, not least because early processes are now close to industrial commercialization (Fessner, 1998).

1.2 Enzymes as Biocatalysts

Over 3,000 enzymes have so far been identified, and this number may be greatly augmented in the wake of genomic and proteomic research (Borman, 2000; MacBeath, 2001). So far one of the most important challenges in bioindustry has been to find a biocatalyst that catalyzes efficiently a biochemical reaction. In other words, endow them with new features that are not found in natural sequences because they confer no evolutionary advantages.

Biocatalysts carry out the chemistry of life. Classically, the subset of proteins with catalytic activity - the enzymes- has been the focus of biocatalysis research (Walsh, 2001). Work during the past decade has shown that there are surprisingly few barriers to the use of enzymes and whole cells as biocatalysts in organic synthesis. Isolated enzymes are typically used for hydrolytic or isomerization reactions (Schmid et al., 2001)

Enzymes are remarkable catalysts: capable of accepting a wide range of complex molecules as substrates, and very selective, catalyzing reaction with unparalleled chiral (enantio-) and positional (regio-) selectivities. As a result, biocatalysts can be used in both simple and complex transformations without the need for the tedious blocking and deblocking steps that are common in enantio- and regioselective organic synthesis (Schmid et al., 2001).

As a significant point every catalyst, and thus also every biocatalyst, can be characterized by the three basic dimensions of merit, namely activity, selectivity, and stability (Bomma & Riebel, 2004) and mainly stability plays an important roll so that it will receive special attention in this work.

1.2.1 Ideal Biocatalysts

Enzymes are class of macromolecules with the ability both to bind small molecules and to effect reactions. Stabilizing forces such as hydrophobic effects only slightly dominate destabilizing forces as Columbic forces of equal polarity; thus the Gibbs free enthalpy of formation of proteins, ΔG formation, is only weakly negative (Pace, 1990; Bomma & Riebel, 2004)

Several parameters have been identified that affect the practicality of enzymatic reactions. Of particular importance are the specific activity (quantified by *kcat*), specificity (determined by *kcat/Km*) and stability of the enzyme. In addition, the degree of inhibition by substrate or product (determined by their affinity to the enzyme) may be particularly important in the outcome of a reaction. In an ideal scenario, the enzyme used would have high specific activity and stability, and would be subject to minimal substrate and product inhibition (Hibbert et al., 2005). Furthermore, the extent of substrate specificity can determine whether a given enzyme will have general synthetic utility, with stereo specificity perhaps the most important parameter under consideration.

Almost all the biocatalysts in use today came from the small fraction of organisms that can be grown under controlled conditions. Some of these organisms live in harsh environments and their catalysts exhibit remarkable and useful properties, including the ability to function under extreme conditions of temperature, salt and pH (Jaenicke, 1981; Szilagyi & Zavodszky, 2000). At present with the synergism of the new biomolecular tools and Bioinformatics,

these special features (pH, temperature, salt concentration, and so on) can be exploited to enhance the natural environmental characteristics of an enzyme such as transketolase which has been identified as an important biocatalyst.

1.3 Transketolase

The enzyme transketolase (TK; E.C. 2.2.1.1) from *Escherichia coli* (Littlechild et al., 1995) *Saccharomyces cerevisiae* (de la Haba et al., 1955) and from others organisms (Kochetov et al., 1975; Philippov et al., 1980; Takeuchi et al., 1986; Masri et al., 1988; Sprenger et al., 1995; Veitch et al., 2004) has been widely characterised. TK occupies a pivotal place in metabolic regulation, providing a link between the glycolytic (degradative) and pentose phosphate pathways. The enzyme has a controlling role in the supply of ribose units for nucleoside biosynthesis, and (in microorganisms) in the supply of erythrose-4-phosphate into the shikimate pathway for aromatic amino acid biosynthesis.

According with X-ray studies (Fig. 1-1), each of the subunits of the dimeric transketolse molecule consists of three domains: N-, or PP-domain (residues 3-322); the middle, or Pyr-domain (residues 323-538), and the C-domain (residues 539-680). The first two domains are involved in coenzyme binding, while the function of the C-domain remains unclear (Kochetov, 2001). TK is homodimeric and requires two TPP and two bivalent metal ion cofactors bound to the two active sites for catalytic activity. The cofactor binding, subunit interactions, and catalytic mechanism of transketolase have also been studied extensively and are well understood, in particular for yeast TK (Datta & Racker, 1961; de la Haba et al., 1955; Nilsson et al., 1997; Meshalkina et al., 1997; Fiedler et al., 2001). The TPP cofactor is covalently bound to the enzyme in the less energetically favourable but activated 'V' conformation (Muller et al., 1993), which promotes the deprotonation of the C2 atom in the thiazolium ring, by the amino group of the pyrimidine in the same molecule. Structures of yeast apo-TK and holo-TK differ in the conformation of two loops (residues 187–198 and

381–394), which are more flexible in the apo-TK form (Sundstrom et al., 1992). Both of these loops are in contact with the TPP cofactor in holo-TK, and become more ordered upon TPP-binding in holo-TK (Nikkola et al., 1994). The dimeric structure of the enzyme is maintained mainly by the PP and Pyr domains.

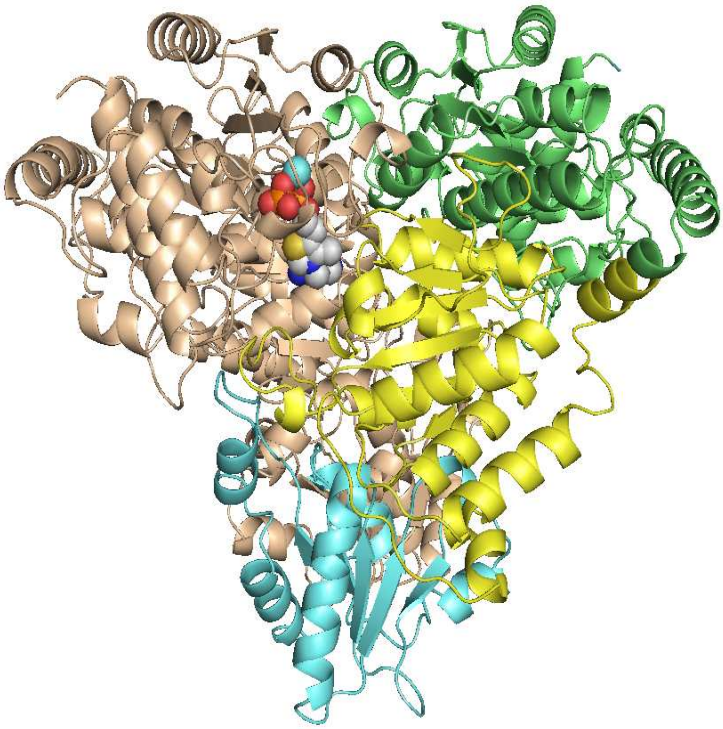


Figure 1-1 Structure of *E. coli* transketolase.
PDB file was obtained from 1qgd.pdb (Littlechild et al., 1995). One monomer is shown as light brown ribbons. The second monomer is shown as ribbons with the PP-domain (green), Pyr domain (yellow), and the C-terminal domain (light blue). For one active-site only, the TPP cofactor is highlighted in CPK colours and as spheres, and the metal ion cofactor is highlighted as a cyan sphere. Figure generated with PyMol (DeLano, W.L. (2002), The PyMOL Molecular Graphics System on World Wide Web http://www.pymol.org).

TK catalyses the interconvertion of sugars by transferring a two-carbon ketol unit from a ketose donor substrate to an aldolase acceptor substrate. Xylulose 5-phosphate, fructose 6-phosphate, sedoheptulose 7-phosphate, erytrulose, etc. serve as the donor substrate. Ribose 5-phosphate, erytrose 4-phosphate, glyceraldehydes 3 phosphate, glyceraldehydes, etc. are used as the acceptor substrates. Common features for the donor substrates are the following: a keto group splitting to the C-C bond, a hydroxil group and the C1 atom and *trans* orientation of the hydroxyl groups at C3 and C4 asymmetrical carbon atoms. Hydroxypyruvate and dioxyacetone, wich contain no asymmetrical carbon atoms, are the exceptions. The Transketolase reaction is reversible; however, the process becomes irreversible in the case of hydroxypyruvate (Figure 1-2) (HPA), wich is subject to decarboxylation due to the production of CO_2 (Datta & Racker, 1961; Schenk et al., 1998; Kochetov, 2001).

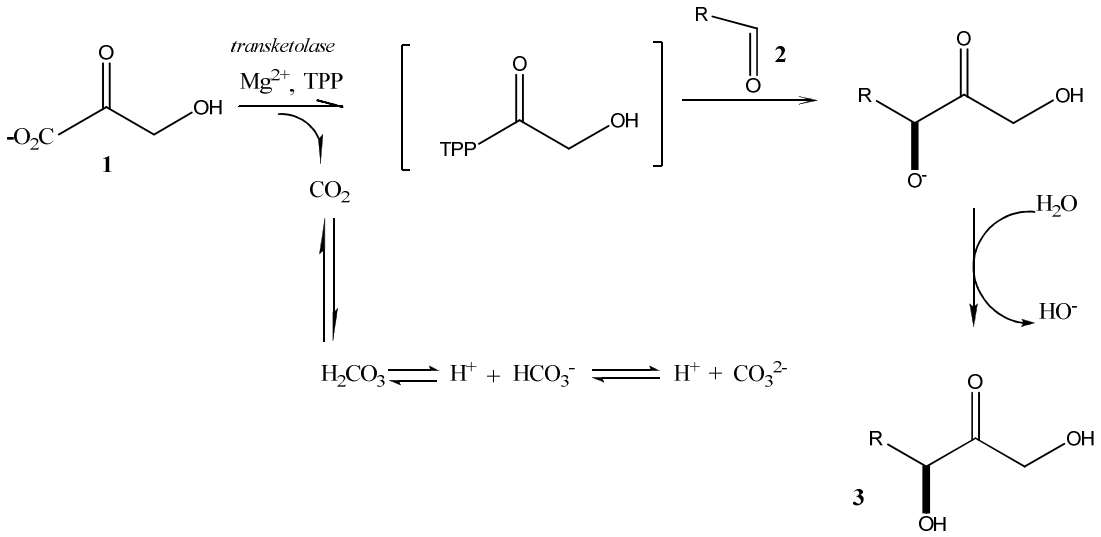


Figure 1-2. Transketolase catalysed reaction.
TK catalyses the reaction of hydroxypyruvate **1** with an aldehyde **2** to yield a keto-diol **3**. The reaction is rendered irreversible by the release of CO_2 by hidroxypiruvic acid **2** and it is accompanied by an increase of pH due to the consumption of one proton per molecule of product formed. (R) denotes variable residues. Figure generated with ChemDraw v.11 (CambridgeSoft, Cambridge, UK).

TK is a member of the group of transferases. It is a homodimer with molecular weight of 148.4 kDa and it has two active sites of equal catalytic activity. Its cofactors are thiamine diphosphate (ThDP) and bivalents metal ions. Native holoenzyme isolated from baker's yeast contain only calcium but bivalent cations such as magnesium, manganese, cobalt, and some others can also serve as cofactors. Changing one cation for another does not alter significantly catalytic activity of the enzyme, but it does affect the character of interaction between the coenzyme and the opoenzyme (Schenk et al., 1998; Kochetov, 2001).

The chemical structure of ThDP (Figure 1-3) is that of an aromatic methylaminopyrimidine ring, linked via a methylene bridge to a methylthiazolium ring with a pyrophosphate group attached to a hydroxyethyl side chain. In non-enzymatic model studies it has been demonstrated that the thiazolium ring can catalyse reactions which are similar to those of ThDP-dependent enzymes but several orders of magnitude slower (Schenk et al., 1998). It has been shown that the dissociation of the proton from C2 of the thiazolium ring is necessary for catalysis; the abstraction of the proton leads to the formation of a carbanion (yield) with the potential for nucleophilic attack on the carbonyl group of the substrate (Schenk et al., 1998).

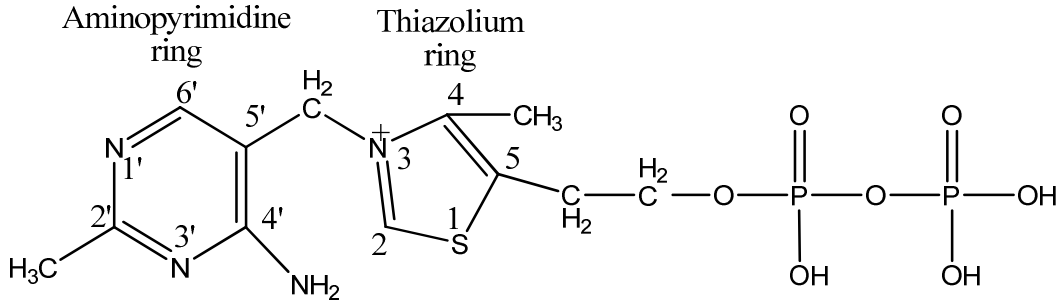


Figure 1-3. Thiamin diphospahate structure (ThDP).
Figure generated with ChemDraw v.11 (CambridgeSoft, Cambridge, UK).

The low specificity for the aldehyde coupled with the high stereoselectivity in the C-C bond-forming step combine to make transketolase a potencially important biocatalist for asymmetric C-C bond synthesis. Carbon-Carbon bond with complete stereo chemical control is of utmost importance in organic synthesis due to the great contribution to this regard (Littlechild et al., 1995; Morris et al., 1996; Koeller & Wong, 2001). The transfer is accomplished by the formation of a covalent intermediate between the ketol moiety and the thiazole ring of a thiamine diphosphate cofactor (Figure 1-4). The structure, cofactor binding sites, subunit interactions, and catalytic mechanism of transketolase have been studied extensively and are well understood (Lindqvist et al., 1992). The thiamine pyrophosphate (ThDP) cofactor molecule is covalently bound to the enzyme in its activated form. Upon reaction, the thiazolium ring adds to the carbonil to perform a C2 addition. The mechanism of the coenzyme activation, which is common for all the thiamine diphosphate-dependent enzymes, includes two closely connected processes: a deprotonation of C2 of the ThDP thiazole ring and functioning of the amino acid pyramiding part of the coenzyme supporting the deprotonation of C2 of the thiazole ring (Kochetov, 2001).

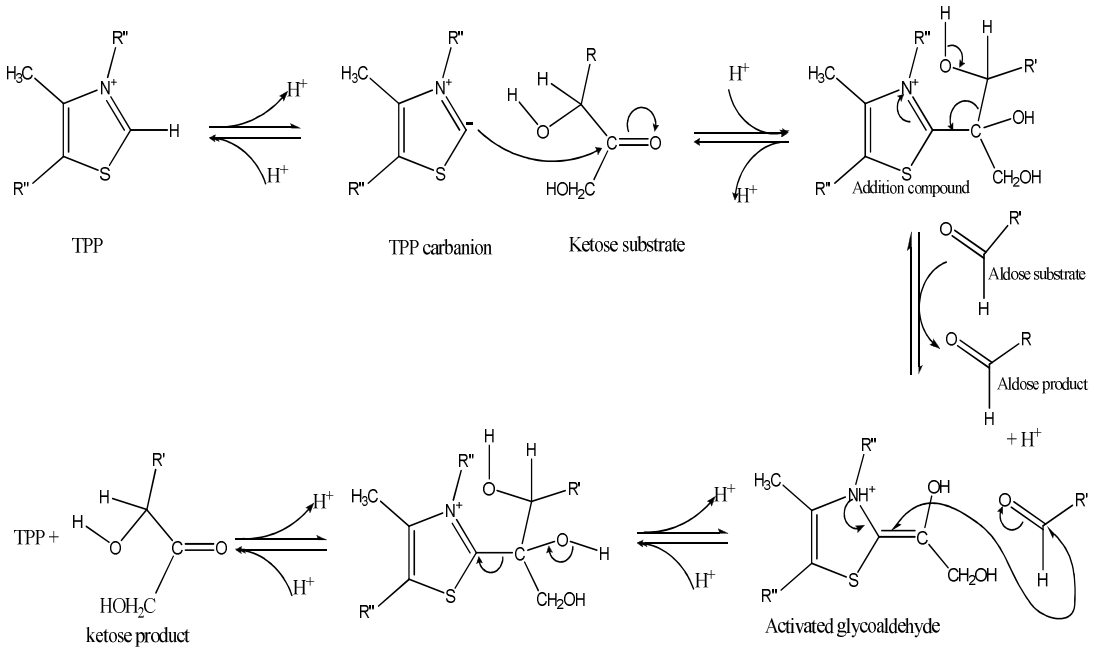


Figure 1-4. Transketolase mechanism reaction.
The carbanion of thiamine pyrophosphate (TPP) attacks the ketos substrate. Cleavage of a carbon-carbon bond frees the alodose product and leaves a two-carbon fragment joined to TPP. This activated glycoaldehyde intermediate attacks the aldose substrate to form a new carbon-carbon bond. The ketose product is released, freeing the TPP for the next reaction cycle. R denotes variable residues. Figure generated with ChemDraw v.11 (CambridgeSoft, Cambridge,UK).

1.3.1 Relevance of TK in industry

TK finds an increasing number of applications for industrial purposes, in particular in the synthesis of chemicals for the food industry (Hecquet et al., 1996); pharmaceuticals (Hecquet et al., 1994) and agrochemical industries (Myles et al., 1991). One example is the biosynthesis of the aromatic aminoacids L-phenylalanine, L-tryptophan and L-tyrosine from D-glucose. These aminoacids are used as precursors for the organic synthesis of various products; for instance: 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). TK has the potential for use in industrial applications, for effective application a large-scale production (Schenk et al., 1998).

TK is also important biocatalyst in stereo-specific carbon-carbon bond synthesis (Woodley et al., 1996). TK accepts a large number of aldehydes as substrates, in particular those containing an alpha-hydroxy group in 2R configuration; the products are ketoses with a 3S, 4R configuration (threo configuration). Moreover, there is currently great interest in the use of biocatalysis for preparing flavour and fragrance components because of the desire to produce 'natural' molecules that can command a premium price as food additives. TK from spinach (Villafranca & Axelrod, 1971) has been employed in a chemo enzymatic synthesis of 6-deoxy-L-sorbose, which is a known precursor of furaneol, a compound with caramel-like flavour, this compound has been found as an important industrial aromatic product used in the food industry (Schenk et al., 1998; Turner, 2000).

Those are only a few of the many examples of the use of transketolase and for that reason recent analyses have outlined the requirements for a bioreactor carrying out TK-catalysed reactions (Woodley et al., 1996;

Brocklebank et al., 1999). Further optimization for large-scale operation at high concentrations of reactants would be facilitated by engineering TK to enhance the affinity for substrates such as hydroxypyruvate and glycolaldehyde, while increasing the turnover rate and decreasing the potential for substrate inhibition by hydroxypyruvate (Morris et al., 1996; Schenk et al., 1998).

Finally, TK activity is crucial for the flux of carbohydrates through the non-oxidative limb of the PPP and understanding its control is critical to the industrial application of TK in the production of biosynthetic molecules. Recombinant mutants possessing desired features such as broader subtrates range, higher thermostability or increased turnover rate may be designed for use in industrial processes (Dalby et al., 2007). Generation of recombinant microbial systems expressing TK has commenced and some encouraging results have been reported (Zimmermann et al., 1999).

1.3.2 Inactivation of TK

Even though some many important reactions have been reported, many of the substrates and products have shown to be sensitive to alkaline environments owing to the labile substrate used. However, as *E. coli* TK is gradually deactivated during prolonged biocatalytic processes. Previous studies have highlighted oxidation (Brocklebank et al., 1999), deactivation by aldehyde substrates (Bongs et al., 1997), irreversible denaturation at pH values below 6.5 (Mitra et al., 1998), and loss of TPP cofactor (Mitra et al., 1998) as the most prominent stability issues in the biocatalytic process for *E. coli* TK. Enzyme oxidation and deactivation by substrate can potentially be minimised respectively, by the addition of a reducing agent such as dithiothreitol (DTT), and by the feeding of substrate in biocatalytic processes. As for loss of cofactors, enzyme activity decreased when excess cofactor was removed by gel filtration, indicating that both TPP and Mg^{2+} dissociate easily from *E. coli* holo-transketolase under all pH conditions suitable for biotransformation reactions

(pH 6.5–9.5) (Mitra et al., 1998). Although TK requires the cofactor TPP, it is regenerated in each cycle of the reaction and so its use is not prohibitive for biocatalysis as is typically observed for other cofactors such as NADH.

1.4 Enzyme stability

The increasing interest in applying enzymes in industrial processes has encouraged the search for biocatalysts with new or improved properties (Kirk et al., 2002; Turner, 2003). Due to the unique capacity of enzymes to catalyse reactions with high velocity and unmet specificity under a variety of conditions, the development of new biocatalytic processes is principally feasible and potentially profitable. Unfortunately, naturally available enzymes are usually not optimally suited for industrial applications. This incompatibility often relates to the stability of the enzymes under process conditions. Although it sometimes is beneficial to adapt industrial processes to mild and environmentally benign conditions favoured by the enzyme, the use of more extreme conditions is often desirable. For example, the use of high process temperatures may be beneficial with respect to factors such as substrate and product solubility, viscosity, process speed and microbial contamination. Regardless of process conditions, the stability of the biocatalyst often is an important economic factor.

The stability of an enzyme is affected by many factors, such as temperature, pH, oxidative stress, the solvent, binding of metal ions or co-factors, and the presence of surfactants (Matthew et al., 1985; Jaenicke, 1988; Sandberg & Terwilliger, 1989; Pace et al., 2004; Dalby et al., 2007). The effect of surfactants is extremely important from an industrial point of view since the detergent area is the largest application area of industrial enzymes. The effect of organic solvents is important since the presence of such solvents is often essential when applying enzymes for the production of fine chemicals.

Among all the potential deactivation factors, temperature is the best studied. At elevated temperatures many enzymes become (partly) unfolded and/or inactivated, meaning that they are no longer able to perform the desired tasks. This can be caused by incompatibility of the optimum temperature for activity or relate to the intrinsic stability of the enzymes (Danson et al., 1996; Daniel et al., 2001; Peterson et al., 2004). It is often assumed that enzymes with improved thermal stability also become more resistant to other denaturing factors (Wang et al., 2000). However, this correlation is not absolute, especially not when it comes to denaturation processes which do not or to a minor extent depend on folding stability (e.g. oxidation of surface residues, temperature-induced deamidation of Asn and Gln).

1.4.1 Enzyme engineering "rules"

Since the beginning of large-scale (recombinant) enzyme production for industrial applications, protein-engineering methods have been applied to improve enzyme properties. Studies on the stability of small enzymes which unfold reversibly at high temperatures have permitted thermodynamic assessment of certain types of interactions for protein stability (Matthews, 1993a). This has led to the identification of several general strategies for protein stabilization, e.g. "entropic stabilization" (rigidification) by Gly → Ala, Xxx → Pro mutations or the introduction of disulfide bridges (Clarke & Fersht, 1993), "helix capping" by introducing residues that interact with the alpha-helix dipole (Marshall et al., 2002), other types of helix optimisation (Serrano et al., 1992), the introduction of salt bridges (Serrano et al., 1990; Schwehm et al., 2003), and the introduction of clusters of aromatic–aromatic interactions (Puchkaev et al., 2003).

In the past decade, large numbers of enzymes isolated from extremophiles have been studied and compared to their counterparts from mesophilic sources. The conclusion from a large collection of such comparative

studies, only few of which are substantiated by mutagenesis work, is that nature has employed many different structural strategies for obtaining high stability.

There are now several examples of proteins which have been stabilized by the introduction of numerous mutations with cumulative small stabilizing effects (D'Amico et al., 2003), which may be taken to suggest that it is useful to use sequence statistics or statistical structural comparisons to discover "rules" for protein stability. It has been shown that stability differences between homologous enzymes may be due to very few (out of many) of the naturally occurring sequence variations (Serrano et al., 1993; Sandgren et al., 2003). Obviously, these latter observations indicate that statistical comparisons of the sequences and structures of proteins with varying stabilities may not be that useful.

The stabilities of proteins that unfold completely and reversibly under denaturing conditions may be assessed by equilibrium measurements, permitting quantification of mutational effects in terms of changes in the ΔG of folding (Pace, 1990). Many proteins do not unfold reversibly and if they do so in the laboratory, they do not necessarily behave likewise in the more complex environment of an enzyme reactor. Therefore, for most industrial enzymes, the only stability parameters that are relevant and assayable relate to kinetic (and not thermodynamic) stability.

1.5 Stabilising routes

There are three major and principally different routes to obtain enzyme variants with improved stability: i) isolating enzyme variants from organisms living in appropriate extreme environments (Vieille & Zeikus, 2001; Schiraldi & De Rosa, 2002); ii) rational-based mutagenesis (enzyme engineering) (Matthews, 1995; van den Burg & Eijsink, 2002), and iii) directed evolution (Arnold, 1998; Dalby, 2003). While directed evolution is based on the generation of diversity followed

by selection/screening, rational protein engineering utilises information on enzyme structure and the molecular basis of stability to predict stabilising mutations.

1.5.1 Natural evolution of enzymes

In biology, evolution is the process of change in the inherited traits of a population of organisms from one generation to the next. The genes that are passed on to an organism's offspring produce the inherited traits that are the basis of evolution. Life exists almost everywhere on the earth, the organisms that inhabit and have adapted to these extreme and diverse environments are often classified by their altered habitat, such as temperature adaptations (psychrophiles to hyperthermophiles), high salinity adaptations (halophiles), pH adaptations (acidophiles and alkaliphiles), and pressure adaptation (barophiles), to name a few groups. In case of adaptations to extremes of pH, salinity, and pressure, membrane components and protective small molecules often play an important role (Jaenicke & Bohm, 1998) and these have been studied quite extensively (Holthauzen & Bolen, 2007).

The microbes and other life forms that are capable of surviving such as harsh environments have adapted their physiology by evolving enzymes, proteins and other characteristics that enable them to remain active and functional. Enzymes from such organisms can be candidate-enzymes to be applied in industrial processes that occur under conditions that are similar to the conditions in their natural habitats (Kirk et al., 2002; Straathof et al., 2002; Robertson & Steer, 2004).

While enzymes from extremophilic organisms may provide useful biocatalysts, their genes may be even more valuable (Schiraldi & De Rosa, 2002). These genes provide important input to genetic pools that are needed for directed evolution experiments. For example, one might attempt to create a

"super-enzyme" by recombining genes from different extremophilic environments, thus creating enzyme variants (Ness et al., 1999) with several favourable characteristics

1.5.2 Directed evolution

Directed evolution (Figure 1-2) has emerged in just a few years as one of the most effective approaches to adapting biocatalysts to the performance requirements of industrial and medical applications (Schmidt-Dannert & Arnold, 1999; Schmid et al., 2001; Zhao et al., 2002; Dalby, 2003). Recently, *in vivo* directed evolution of enzymes, using random genetic mutation and recombination, followed by screening or selection for a desired trait, has been explored as a more generally applicable approach to the modification of enzymes properties (Arnold, 1998; Arnold et al., 2001; Sylvestre et al., 2006).

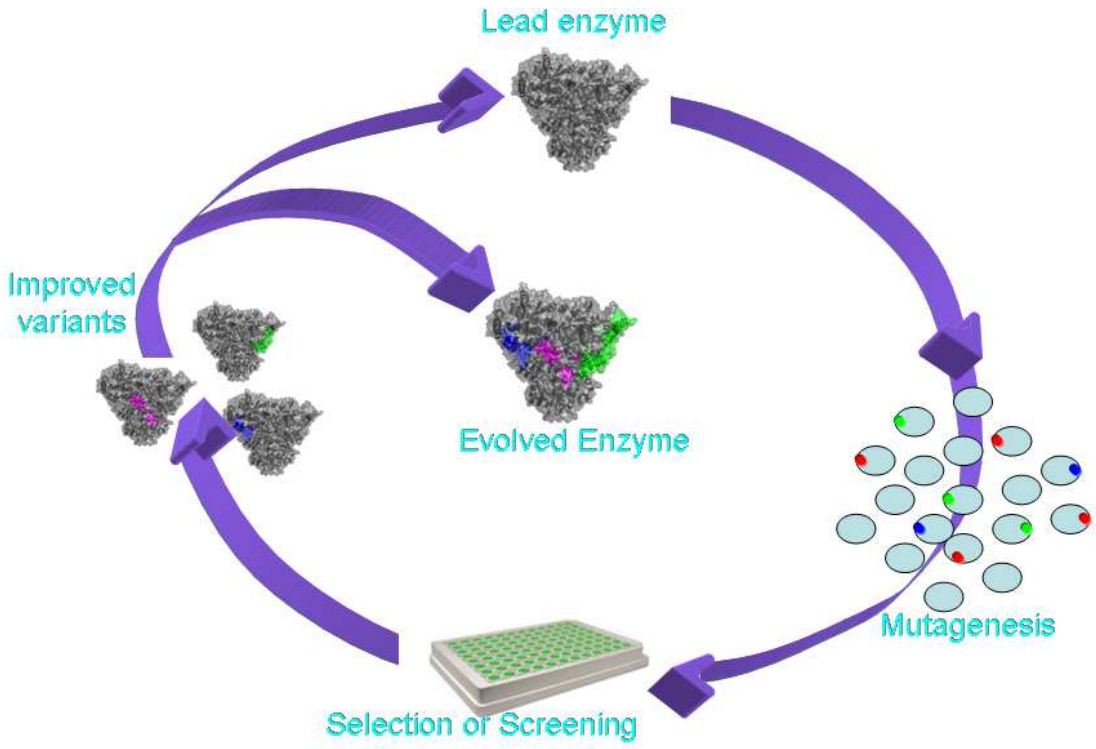


Figure 1-5. The two-essentail steps of directed evolution of enzymes.
(1) mutagenesis of the gene encoding enzyme, and (2) selection of a particular variant based on desired properties.

The precise three-dimensional architecture of enzymes permits almost unerring selectivity in physical and chemical steps to impose remarkable rate accelerations and specificity in product-determining reactions. Many enzymes are members of families that carry out related chemical transformation and offer opportunities for directed *in vitro* evolution, to tailor catalytic properties to particular functions (Walsh, 2001).

Unlike natural evolution, laboratory evolution is directed. A "generation" of new molecules can be developed in a few days, with large number of progeny. Because the molecules are produced in recombinant cell and are decoupled from their biological functions, they can be developed for non-natural but useful properties, including the ability to carry out reactions on substrates not encountered in nature, or to function under highly unusual conditions. This approach is particularly attractive for engineering industrial biocatalysis (Arnold, 2001; Kirk et al., 2002).

To select an enzyme for a given reaction, one can start with one enzyme capable of catalyzing that specific type of reaction, optimize the reaction conditions, and further improve the catalyst trough directed evolution cycle (Dalby, 2003). Once an enzyme has been evolved to have a detectable and new desirable activity, additional rounds of *in vitro* evolution can improve its stability and robustness. (Schmidt-Dannert, 2001; Walsh, 2001). In turn this will aid in enzyme evolution to select and detect new activities and then to incorporate improved catalytic efficiency, attributes of specificity, and structural features optimized to a given operating microenvironment. Small-molecule chemical transformations catalysed by enzymes from microorganisms that live in unusual environments or conduct chemical warfare against their neighbours have been and are likely to remain good hunting grounds for new enzyme transformations. Taking into account these approaches, site-directed mutagenesis and directed evolution have also been used for the improvement of thermal and/or solvent stability of enzymes.

1.5.3 Recombinant DNA

A revolution in biotechnological design possibilities was unleashed by the advent of recombinant DNA technology, with which one can manipulate DNA sequences in a highly specific fashion and express their protein products in a variety of organisms, from animals to bacteria. This provides a means to redesign nature's catalysts at the molecular level according to detailed specifications, and to produce them in large quantities in fast-growing micro-organisms. Broadly speaking, one can identify two philosophies: either existing biocatalysts can be fine-tuned by rational redesign, or combinatorial techniques can be used to search for useful functionality in libraries generated at random and improved by suitable selection methods (Arnold, 2001; Zhao et al., 2002; Hibbert & Dalby, 2005).

1.5.4 Site Specific *in vitro* Mutagenesis

In vitro mutagenesis comprises a group of molecular biological techniques used for structural and functional analysis of nucleic acids and/or proteins by introducing mutation *in vitro* on a segment of DNA in a controlled manner. The mutagenesis studies provide key information required for designing DNA, RNA, or protein with a novel function, tailored specificity, and altered stability to heat or organic solvents (Miyazaki & Arnold, 1999).

Mutagenesis is called "random" if one or more bases are modified throughout the sequence. Mutagenesis is called "site specific" when a base or bases are altered within a restricted target region (Reetz & Carballeira, 2007). Whereas random mutagenesis is useful to identify the location and boundaries of a particular function and is most profitably applied to poorly characterized DNA and proteins, site-specific mutagenesis provides a means to explore the role of specific nucleotides or amino acids in relatively well-characterized DNA and proteins. Random mutagenesis generally relies on chemical mutagens to

modify bases, whereas site-specific mutagenesis usually resorts to sequence-specific oligonucleotide.

Site-specific mutagenesis continues to offer great scope for protein stability improvement where the structure of the target protein (or closely-related protein) is known. This is specially so in situations where screening of a large population of mutants is not possible or feasible (O'Fagain, 2003).

1.5.5 Rational design

Rational design has been mainly used in biotechnology to improve the properties (especially thermostability) of natural enzymes. The analysis of proteins for rational design involves sequence comparison of the protein family, which gives information about the structurally important amino acids and variability at each amino acid site. In engineering, e.g., thermostability, earlier mutational studies, comparison of amino acid sequences, total amino acid content, and crystal structures between mesophilic and thermophilic enzymes may give information on the key factors behind the elevated stability (Hakulinen et al., 2003; Eijsink et al., 2004). When the protein structure is known, computer simulations can be used to study the active site properties, substrate binding, thermostability, and unfolding of the enzymes. Simulations provide information that is useful in planning mutations. For example, molecular dynamics simulations have been used to identify flexible regions in proteins (Daggett & Levitt, 1993), and subsequently, the protein stability has been increased considerably by introducing a disulfide bridge into such a region. One of the simplest stabilization methods is actually the construction of a disulfide bridge into a protein structure. It has to be done by rational design because random mutations are not likely to form the required simultaneous double mutation (Fenel et al., 2004).

A new semi-rational method to improve enzyme stability is to make a consensus sequence to the protein family. This approach is based on the assumption that conserved amino acid properties have been selected in nature because of their impact on protein stability. The consensus method efficiently explores the local sequence space (Lehmann et al., 2000; Lehmann & Wyss, 2001). In limited areas, rational design methods can be very effective in improving the properties of industrial enzymes.

1.6 Screening enzyme libraries

The method used for screening novel or improved enzyme traits is extremely important, because the screening process is costly and because the screening process has many pitfalls (Robertson & Steer, 2004). It is important that the evolutionary pressure that is put on the library of enzyme variants precisely and exclusively reflects the property or properties that one wants to evolve. One of the simplest potential pitfalls consists of expression problems that may interfere with sub optimally designed screening methods that are based on measuring enzyme activity. With respect to thermal stability, the distinction between thermal tolerance and real thermal stability is important. The first term refers to the ability to withstand incubation at elevated temperatures, without necessarily being active at those temperatures. Real-thermal stability refers to enzymes that not only withstand elevated temperatures, but that also retain activity at these temperatures. Clearly, these two types of properties need different screening regimes.

It is important to ensure that the screening procedure accounts for all the properties that one wishes to improve, while ensuring that other qualities that are important for the process in question (for instance activity) are preserved. If the experimental strategy involves some kind of rational considerations the chances of getting unwanted effects are smaller, for instance because diversity is generated at sites known to be far from the active site (Morley & Kazlauskas,

2005). Prior to the generation of genetic diversity, one should carefully consider which screening procedure would be most suited with respect to handling capacity, sensitivity and achievement of the envisioned goals.

The screening challenge may be approached in numerous different ways. In some cases, screening may be done by selection in a biological system, but in most cases microtiter plate-based (Wahler & Reymond, 2001), filter-based or other in vitro formats need to be used (Olsen et al., 2000). New methods are being developed at high speed and most of these yield faster and less costly screening and/or permit the use of larger library sizes.

1.6.1 Screening for stability

Most strategies for screening for enzyme stability rely on measuring residual enzyme activity after exposure to a denaturing challenge (e.g. high temperature, organic solvents, extreme pH) on a filter or in microtiter plate wells (Cirino & Georgescu, 2003). An alternative and potentially simpler strategy would be to screen directly for activity under denaturing conditions, but this may pose some practical problems. For example, in the case of thermal stability, one faces the limitation that most high throughput microplate readers can only reach temperatures of about 50 °C, as well as problems caused by the fact that high temperatures cause evaporation, pH changes and/or substrate instability. Also, potential problems caused by variation in expression levels need to be addressed, e.g. by screening at two temperatures and looking at ratios. The general picture emerging from studies on engineered thermal stability is that enzymes that withstand higher temperatures usually also have higher temperature optima for activity (Zhao & Arnold, 1999). There are, however, exceptions (Arnott et al., 2000) and it has been pointed out that enzymes may have "true" temperature optima for activity, which are not necessarily related to the conformational stability assessed in stability assays (Daniel et al., 2001).

During protein evolution, functional properties are under different selection pressure. As a result, during selection for either activity or stability, the best mutation should improve either one of these without negatively affecting the other. Theoretical considerations concerning the relationship between activity and stability, suggest that the proteins evolve function more efficiently when the stability requirements are gradually increased than when there is constant selection for high stability (Bloom et al., 2004). In practice, this means that screening schemes may need a separated activity and stability component.

1.7 Protein aggregation

Protein aggregation can be merely a trouble factor in many *in vitro* studies of proteins or it can cause major economic and technical problems in the biotechnology and pharmaceutical industries. Protein aggregation is intimately tied to protein folding and stability, and also, in the cell, to molecular chaperones. The prevalence of protein aggregation is probably much higher than much higher than generally realised – it is often ignored or worked around, and in protein folding experiments its presence may not even be realised (Linding et al., 2004).

1.7.1 Protein molten globular state

A variety of proteins have been observed under certain conditions to exist in stable conformations that are neither fully folded nor fully unfolded. These conformations have sufficient similarities to suggest that they are different manifestations of a third stable conformational state (Kunihiro, 1989). The most common properties are: (1) The overall dimensions of the polypeptide chain are much less than those of a random coil and only marginally greater than those of the fully folded state. (2) The average content of secondary structure is similar to that of the folded state. (3) The interior side-chains are in homogeneous surroundings, in contrast to the asymmetric environments they have in the fully folded state. (4) Many interior amide groups exchange hydrogen atoms with the solvent more rapidly than in the folded state, but more slowly than in the fully unfolded state. (5) Its enthalpy is very nearly the same as that of the fully unfolded state, substantially different from that of the native state. (6) Interconversions with the fully unfolded state are rapid and nonco-operative, but slow and co-operative with the fully folded state.

If these observations are applicable to a homogeneous structure, they suggest a collapsed molecule with native-like secondary structure and a liquid-like interior, i.e. a 'molten globule'. Detailed studies of one such protein, a-lactalbumin, however, have demonstrated that portions of the hydrophobic interior are in relatively stable, well-ordered three-dimensional conformations, with amide groups highly protected from hydrogen exchange, whereas other parts are much less structured (Baum et al., 1989). The molten globule state has also been reported in proteins subjected to proteolysis, mutagenesis or chemical modifications (Goto & Nishikiori, 1991; Chaffotte et al., 1991).

It therefore seems reasonable to conclude that most of the interactions within the folded protein are more favourable energetically, in both enthalpy and free energy, than the corresponding interactions of the unfolded state. They should therefore contribute to the net stability of the folded state, and it is not surprising that hydrogen bonds and salt-bridges have been found to do so (Alber et al., 1987; Serrano & Fersht, 1989; Anderson et al., 1990).

1.8 Bioinformatics

An unprecedented wealth of biological data has been generated by the human genome project and sequencing projects in other organism. The huge demand for analysis and interpretation of these data is being managed by the evolving science of bioinformatics. Bioinformatics is defined as the application of tools of computation and analysis to the capture and interpretation of biological data. It is an interdisciplinary field, which harnesses computer science, mathematics, physics and biology. Bioinformatics is essential for management of data in modern biology.

1.8.1 Bioinformatics tools

The main tools of bioinformatician are computer software programs and the internet. A fundamental activity is sequence analysis of DNA and proteins using various programs and databases available on the worldwide web such as BRENDA, BLAST, and CLUSTALW and so on. Bioinformatician now use software for retrieving, sorting out, analyzing, predicting, and storing DNA and protein sequence data.

The growth of bioinformatics has been a global venture, creating computer networks that have allowed easy access to biological data and enabled the development of software programs for effortless analysis. Bioinformatics in this work will guide and help us to capitalize on the advantages brought by computational biology and database.

1.8.1.1 BRENDA

The rapid sequencing of a large number of genomes has made it imperative to organize the available data in ways that facilitate easy analysis. BRENDA (BRaunschweig ENzyme DAtabase at http://www.brenda.uni-koeln.de/) was created in 1987 at the German National Research Centre for Biotechnology at Raunschweig. As of February 2003, it provides information on 40,000 different enzymes represented by 4,087 EC numbers, and present in more than 9000 different organisms as of February, 2005.

The enzyme content of BRENDA can be accessed using the enzyme EC number, the enzyme name and the organism name (Schomburg et al., 2002). Once a particular enzyme has been searched, information can be obtained for a large number of fields. These include not only the reaction catalyzed by the enzyme in terms of its substrates and products, but also the organisms and the

tissues in which the enzyme is present. In addition to this primary information, functional parameters such as optimum pH range, optimum temperature range, and molecular properties of the enzyme can also be obtained (Pharkya et al., 2003). Crucial information for enzyme handling, preparation and application can also be acquired from the database. For instance, the field 'application' cites information on the established and potential uses of an enzyme. The field 'protein engineering' lists enzyme variants and helps to compare their properties with that of the wild-type enzyme. BRENDA is quite comprehensive in terms of both the number of enzymes included in the database and the broad range of properties on which information about an enzyme can be extracted (Schomburg et al., 2004). It also enables a user to get all the data compiled from different sources at the same place which is far more useful than retrieving data from a single reference at a time (http://emp.mcs.anl.gov/). Furthermore, with a single query, it is possible to retrieve the enzyme properties in all organisms where the enzyme is present. BRENDA is a great tool for obtaining a wide range of information on enzymes and understanding their biological functions.

1.8.1.2 BLAST

One of the simplest and better known search tools is called BLAST (Basic Local alignment search tools at http://www.ncbi.nlm.nih.gov/BLAST/). This algorithm software is capable of searching database for genes with similar nucleotide structure and allows comparison of an unknown DNA or aminoacid sequence with hundred of thousands of sequences from humans and other organisms until a match is found. Database of know sequences are thus used to identify similar sequences, which may be homologues of the query sequences. Homology implies that sequences may be related by divergent from a common ancestor or to share common functional aspects. When a database is searched with a newly determined sequence (the query sequence), local alignment occurs between the query sequence and any similar sequence in the database. The result of the search is sorted in order of priority on the basis of maximum

similarity. The sequences with the highest score in the database of known genes are the homologue.

1.8.1.3 ClustalW

In order to identify some similarities between sequences a multiple alignment should be done using the sequences collected. In this work Clustal W was used to produce the multiple alignment (Thompson et al., 1994). ClustalW is a general multiple sequence alignment program for DNA or proteins. It produces biologically meaningful multiple sequence alignments of divergent sequences. It calculates the best match for the selected sequences, and lines them up so that the identities, similarities and differences can be seen. The alignment is progressive and considers the sequence redundancy and on the other hand, depending of the aim of the research project, a phylogenetic analysis can be done by using the phylogenetic tree, which is generated from multiple alignments. To produce an alignment a comparison matrix is commonly used such as PAM, Blossum62, or Gonnet. In this work Blossum62 was used as a default matrix to produce the multiple alignment

The procedure for calculating a BLOSUM matrix is based on a likelihood method estimating the occurrence of each possible pairwise substitution. Any further information could be found at http://www.ebi.ac.uk/clustalw/.

1.8.1.4 NCBI

The NCBI (http://www.ncbi.nlm.nih.gov/) has been charged with creating automated systems for storing and analyzing knowledge about molecular biology, biochemistry, and genetics; facilitating the use of such databases and software by the research and medical community; coordinating efforts to gather biotechnology information both nationally and internationally; and performing

research into advanced methods of computer-based information processing for analyzing the structure and function of biologically important molecules.

The NCBI offers into one of its subprograms the software BLAST. BLAST (Basic Local Alignment Search Tool), provides a method for rapid searching of nucleotide and protein databases. Since the BLAST algorithm detects local as well as global alignments, regions of similarity embedded in otherwise unrelated proteins can be detected. Both types of similarity may provide important clues to the function of uncharacterized proteins.

Using this program the results are given under the format FASTA which is required to done the alignment in ClustalW. This sort of format allows to ClustalW identify the sequences and therefore the alignment can be then done. Otherwise the sequences could not be recognized by ClustalW and as a consequence the alignment will not be generated.

1.8.1.5 PDB

The Protein Data Bank (PDB; http://www.rcsb.org/pdb/) is the single worldwide repository for the processing and distribution of 3-D structure data of large molecules of proteins and nucleic acids. Using this databank the 3-D structure of Transketolase, reported by Jennifer Littlechild, 1995, was acquired.

1.8.1.6 PyMol

PyMOL (http://pymol.sourceforge.net/) is a Python-enhanced molecular graphics program (DeLano, 2002). It excels at 3D visualization of proteins, small molecules, density, surfaces, and trajectories. It also includes molecular editing, ray tracing, and movies.

1.8.1.7 BioEdit

BioEdit is a biological sequence alignment. An intuitive multiple document interface with convenient features makes alignment and manipulation of sequences relatively easy. Several sequence manipulation and analysis options and links to external analysis programs facilitate a working environment which allows you to view and manipulate sequences with simple point-and-click operations.

1.9 Alternative analysis applied to protein engineering

There are two main reasons to try to predict an enzyme's function from its sequence. The first is to identify the components and thus the functional capabilities of an organism; the second is to create enzymes with specific properties.

Many computational methods are now being tested for their ability to identify functionally important residues. These residues can then be used to help functional assignment of proteins, or to predict regions or residues involved in protein–protein interactions or catalysis. Functional residue predicting methods typically use one or more of the following strategies: i) Use of phylogenetic analysis to define evolutionary distances between homologs (del Sol Mesa et al., 2003; Costelloe et al., 2008) and cluster them based on evolutionarily related sub-families. Evolutionary trace, sequence motif and conserved functional group analysis of homologous proteins to identify putative active sites (Di, V et al., 1996); ii) Degree of amino acid conservation at positions in related proteins (Casari et al., 1995); iii) Use of principal component analysis of sequences (Casari et al., 1995); iv) Use of structural information such as proximity of observed changes to active site (Hibbert et al., 2007); v) Use of structural information such as conservation of specific residues in specific three-

dimensional positions (Innis et al., 2004); vi) Use of substitution matrices that describe the relative likelihood of amino acid conversions (Chelliah et al., 2004); vii) Protein structure prediction using multivariate statistical methods such as PLS (Clementi et al., 1997)

1.9.1 Prediction function from sequence

Mathematical and data-mining tools developed in other fields of engineering have now been applied to analyze sequence –activity relationship of peptides and proteins and to assist in the design of proteins and peptides with specified properties. Decreasing costs of DNA sequencing in conjunction with methods to quickly synthesize statistically representative sets of proteins allow modern heuristic statistics to be applied to proteins engineering. Comparisons of natural protein and DNA sequences, particularly those using the powerful technique of PCA and PLS, can be used to identify residues that are important for specific functionality within a protein (Gustafsson et al., 2003).

One of the most exciting advances in elucidating the relationship between sequence and function in the use of dimension-compressing techniques such as principal component analysis and partial least square to identify key sequence elements that correlate with specific activities. These techniques were pioneered to assist in functional assignments for naturally occurring proteins (Casari et al., 1995).

Statistical analysis of mutations distributed throughout several enzymes has been used to identify the contribution of those changes to function of the protein and predict the sequence with the best function(Jonsson et al., 1993; Gustafsson et al., 2003).

Since sequence-activity relationship require the testing of orders of magnitude fewer variants than traditional approaches (Zhao et al., 2002; Eijsink et al., 2004), the use of such predictive algorithm will permit us to design lower through screens that more accurately reflect the properties that we are aiming for.

1.10 Project aims

Transketolase catalyses the stereoselective transfer of a two-carbon ketol group to an aldose sugar, which results in asymmetric carbon–carbon bond synthesis making the enzyme of particular interest for the biocatalytic synthesis of complex carbohydrates, and other high value compounds. In this research a biophysical analysis of the enzyme biocatalyst transketolase was carried out to provide insight into the mechanism of enzyme deactivation that occurs during biocatalytic conditions. Further research then aimed to enhance the thermal and pH stability of the enzyme by combining bioinformatic and statistical tools with protein engineering. A general overview of each chapter in the thesis follows:

Chapter 3 describes stability studies of the complex enzyme transketolase. Wild-type transketolase was: (i) over-expressed by using the plasmid pQR791; (ii) purified using an AKTA purification system equipped with unicorn software; (iii) finally, the stability and denaturation pathway of the enzyme was assessed by using biophysical techniques such as circular dichroism, intrinsic fluorescence, dynamic light scattering, and size exclusion chromatography.

Chapter 4 analysed the effect of different pH's and temperatures on activity and the secondary structure of wild-type transketolase. Wild-type transketolase was exposed to a range of temperatures and pH and the residual activity was measured by using HPLC as a quantitative high-throughput method. Additionally, changes in the secondary structure of the protein were monitored using circular dichroism and dynamic light scattering.

Chapter 5 describes an attempt to design rationally transketolase stability using bioinformatics to guide the mutagenesis. Different approaches such as statistical, multivariate and bioinformatical analyses of protein sequences were combined in an effort to associate enzyme properties with protein sequence to determine the most likely amino-acid sequence sites that affect key protein properties such as temperature and pH optima in biocatalysis. These analyses identified 20 sites which were randomly mutated using site-saturation mutagenesis methods and the resulting libraries of enzymes screened using automated techniques such as automated colony-picker, HPLC and a TECAN handling liquid robot.

Chapter 2

2 Materials and Methods

2.1 General notes

All chemicals were obtained from Sigma–Aldrich UK unless otherwise stated. DTT was obtained by Fisher-Scientific UK. Water was purified to 15 MΩ cm resistivity using an Elix 5 water purification system (Millipore Corp).

2.2 Preparation of media, buffer and reagents.

2.2.1 Luria-Bertani (LB) medium

The following method was used to prepare 1 L of LB medium. 10 g of Tryptone; 5 g of yeast extract and 10 g of NaCl were measured out. Solids were then dissolved in RO water to make up 1litre. Prior to autoclaving, pH was checked and adjusted to pH 7 with sodium hydroxide when necessary. The medium was autoclaved at 121 ^{o}C for 15 min. After cooling, the media was ready to use.

2.2.2 Luria Bertani (LB) agar and agar plates

LB agar was prepared by adding 20 g L^{-1} select agar to LB medium. The agar preparation was then autoclaved. Once cooled to between 40-50 °C, 20 mL of media was poured into each standard-sized petri dish.

2.2.3 Ampicillin

Ampicillin was dissolved in pure water to a concentration of 150 g L^{-1}. Stocks were sterilised by filtration and stored at -20 ˚C. Ampicillin was used at a concentration of 150 μg mL^{-1} and added to either LB medium or LB agar before

it was set to select bacteria carrying the plasmid pQR791. Preparations containing this concentration of ampicillin are labelled Amp^+.

2.2.4 250 mM and 25 mM Tris-HCL buffer (pH 7.5)

250 mM or 25 mM Tris (hydroxymethylaminomethane) buffer with a pH of 7.5 were prepared by dissolving 31.75 or 3.175 mg mL^{-1} Tris hydrochloride and 5.9 or 0.59 mg mL^{-1} Tris base in pure water.

2.2.5 Standard transketolase cofactor solution

Standard cofactor solution were usually prepared on the day by dissolving 2.5 mg mL^{-1} (27mM) of $MgCl^{+2}$ and 3.3 mg mL^{-1} (7.2mM) TPP in 250mM Tris-buffer. pH solution was adjusted by adding a concentrated solution of either sodium chloride or hydrochloric acid.

2.2.6 Standard transketolase substrate solution

Standard substrate solution was usually prepared on the day by dissolving 16.5 mg mL^{-1} (150 mM) of Li-HPA and 9.0 mg mL^{-1} (150mM) of GA in 250 mM Tris-buffer. pH Solution was adjusted by adding a concentrated solution of either sodium chloride or hydrochloric acid.

2.2.7 Standard *E. coli* Transketolase reaction

Usually 50 µL of either pure enzyme or clarified lysate were mixed with 50 µL of standard cofactor solution (section 2.2.5) by pippeting it up and down in a 96 shallow well plate, and incubated for 30 minutes, prior to the addition of substrate solution to allow full reconstitution of HoloTK from ApoTK. Reaction was started by adding 50 µL of substrate solution (section 2.2.6). After 60 minutes a 0.2% (v/v) solution of TFA was added in a one to one ratio (1:1) to

quench the reaction. The concentration of erythrulose, reaction product, in this quenched sample was determined by HPLC (section 2.2.8). Reaction was performed at room temperature and pH 7.5 unless otherwise stated.

2.2.8 HPLC method to estimate L-Erythrulose concentration

This method was developed by Christine Ingram (Department of Biochemical Engineering, UCL). 10 µL of samples reaction (section 2.2.7) were analysed by HPLC (Dionex, CA, USA) and UV detection at 210 nm. Sample was injected onto either a 300 mm Aminex HPX-87H ion-exclusion column (Bio-Rad Laboratories) and maintained at 60 °C using a LC30 chromatography oven (Dionex Corp.) for a 16 min high accuracy screening or using a 50 mm PL Hi-Plex H guard column (Polymer Laboratories Ltd., UK) and maintained at 25 ^{o}C for a 5 min fast primary screening. Methods were used to separate the components Li-HPA (substrate) and L-erythrulose (product), with 0.1% (v/v) TFA mobile phase at a flow rate of 0.6 mL min^{-1}.

2.2.9 Retention time and calibration curve

The retention time of Li-HPA and L-erythrulose products are depicted in Table 2-1 see below. The progress of the reaction was followed by the appearance of L-erythrulose product, the peak area of which was used in subsequent analysis. Standard curve of 0–40 mM of L-erythrulose in the same conditions reaction was used to obtain the L-erythrulose concentrations. Figure 2-1 below illustrates the calibration curve for L-erythrulose HPLC using high accuracy method (section 2.2.8).

Table 2-1 Retention time of Li-HPA and L-Erythrulose using high accuracy and primary screening method.

	Retention time (min)	
Method	Li-HPA	L-Erythrulose
High accuracy method	8.5	11.9
primary screening	1.5	2.3

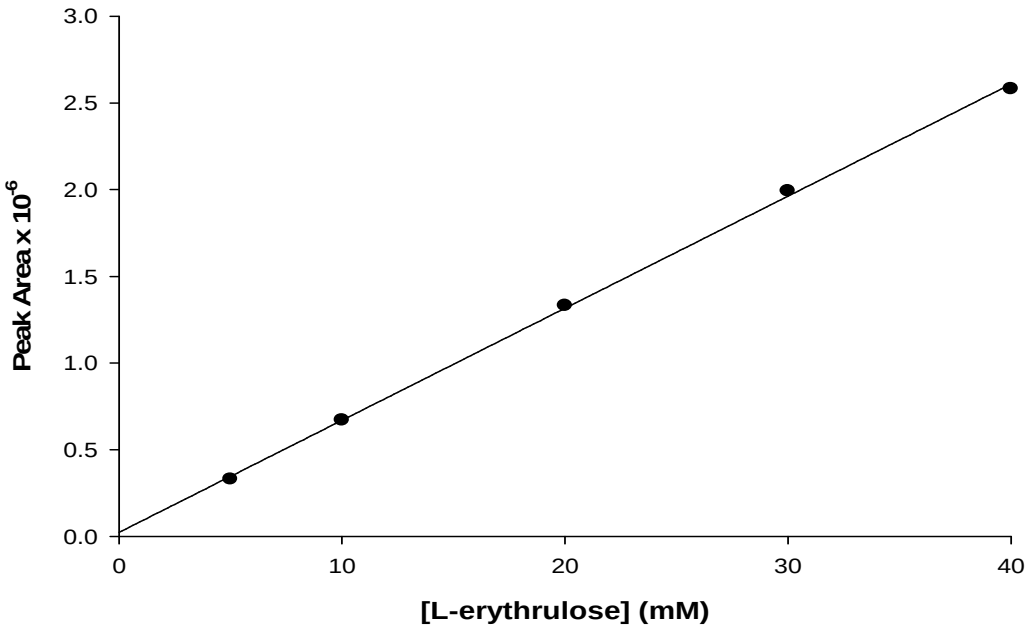


Figure 2-1 Calibration curve of L-Erythrulose determined on the 300 mm Aminex HPX-87H ion-exclusion column.
L-erythrulose was prepared in 250 mM Tris buffer at pH 7.5. Figure shows a linear relationship between peak area and L-erythrulose concentration with a R2 of 0.9997.

2.3 Standard Molecular Biology Procedures

2.3.1 Streaked plates

A culture was streaked out on a Petri dish of LB agar (Amp^+) using a wire loop. The plate was incubated overnight at 37 ˚C and stored at 4 ˚C.

2.3.2 Over night cultures

A single *E. coli* colony was picked from a gar plate and transferred into 5 mL of LB medium contained in a 50 mL falcon tube. The tube was incubated for 16 h at 37 ^oC with 220 rpm agitation.

2.3.3 Shake flask culture

6 mL of an overnight culture was added to 594 mL of LB medium (Amp^+) in a sterile 2 L shake flask. The shake flask was incubated for 16 h at 37 ^oC with 220 rpm agitation.

2.3.4 Glycerol stocks

A 20% (v/v) glycerol stock was prepared by adding filter-sterilised or autoclaved 40% (v/v) glycerol to an overnight culture in a one to one (1:1) volume ratio. 100 µL aliquots were prepared in a 0.75 mL eppendorf tube and stored at -80 °C.

2.3.5 Sonication

The cells were lysed with an ultrasonication protocol of 10 seconds on/off at 8 µm amplitude for 10 cycles (Soniprep150, Sanyo, UK). Cell-free extract was cleared by centrifugation at 13000 rpm for 20 minutes in a bench

microcentrifuge (Kendro Laboratory,UK). Cells must be placed on ice to stop overheating.

2.3.6 Preparation of plasmid pQR791

Plasmid DNA was extracted from 2 mL of an overnight culture using a commercial plasmid mini-preparation kit (Qiagen Ltd, UK). Plasmid DNA was extracted according to the manufacturer's protocol. The final product was eluted into 50 µL EB buffer. Plasmid DNA was stored at -20 ^{o}C.

2.3.7 Plasmid pQR791

In this thesis, plasmid pQR791 was used as a template for production of wild-type transketolase and mutants. The pQR711 plasmid (French & Ward, 1995) was modified previously (Aucamp, 2005) by site-directed mutagenesis using the QuickChange™ kit (Stratagene, La Jolla CA) to introduce a BglII restriction site in the tkt gene promoter region (producing pQR790), and then an N-terminal His6 tag on transketolase to create pQR791. This new engineered plasmid allowed to purify large enzyme concentration of either wild-type or transketolase mutants by using Ni-affinity columns.

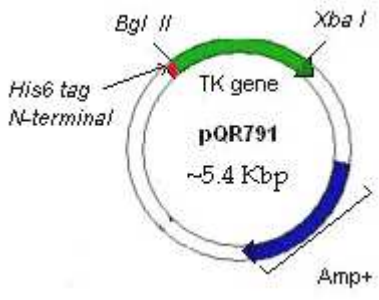


Figure 2-2 Schematic representation of plasmid pQR791.
The ampicilin resistance (Amp^{+}) is represented in blue; the *tkt* gene is represented in green, and the poli His tag tail is coloured in red. The total plasmid size is 5.4 Kbp.

2.3.8 Transformation by heat-shock

XL10-Gold ultra-competent cells were routinely used to perform heat-shock transformations (Stratagene, La Joya, USA). Cells were thawed on ice and 40 µL aliquots transferred into a 15 mL pre-chilled Falcon tube. 2 µL of the β-ME was added to each aliquot of cells and gently mixed every 2 minutes while incubated on ice for 10 minutes. 1.5 µL of the either plasmid pQR791 or quick change product was added to cell aliquots and incubated on ice for 30 minutes. The XL10-Gold cells were heat shocked in a 42 ^{o}C water bath for 30 seconds and immediately placed on ice for 5 minutes. 0.5 mL of preheated (42 ^{o}C) NZYM broth (Sigma-Aldrich, UK) was added and cells incubate at 37 ^{o}C for 1 h with shaking at 220 rpm. 5 µL, 10 µL, 30 µL and 50 µL of transformed cells were plated out on LB-agar plates (Amp^{+}). Colonies were picked from plates and incubated overnight in LB media containing 150 µg mL^{-1} ampicillin to produce master stock cultures. Master cultures were preserved in 20% sterile glycerol and stored at -80 ^{o}C.

2.3.9 Agarose gel electrophoresis

Double-stranded DNA fragments were separated by gel electrophoresis in agarose gels (0.7-1.5% (w/v) depending on the size of fragments to be analysed. Agarose was suspended in 0.5x TBE buffer and heated in a microwave oven. When agarose was dissolved and solution cooled to less than 45 ^{o}C, ethidium bromide (0.5 µg/mL final concentration) was added and the agarose poured into an electrophoresis tray. When the gel had hardened, it was placed horizontally in a gel electrophoresis tank containing enough 0.5x TBE buffer to cover the surface of the gel to a depth of 1 mm. DNA samples were mixed with 0.2 volumes of the 6X gel loading buffer (Sigma-Aldrich, UK) before loading into the wells in the gel. Electrophoresis was carried out for 1 and a half hour at a constant voltage of 75 V. Fragment sizes were estimated by comparison with tracks of a standard super-coiled DNA ladder (Invitrogen, UK) (Figure 2-3).

DNA gels were UV-visualised and photographed using a Gel DOC 2000 system (Bio-Rad Laboratories).

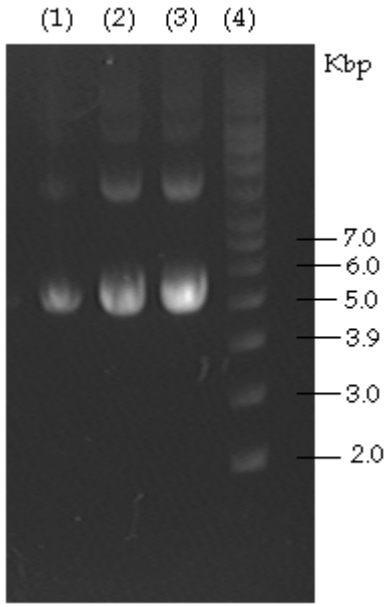


Figure 2-3 Digital image of plasmid pQR791 in a 0.6% (w/v) agarose gel. Lanes are as follow: (1); (2); (3) lanes correspond to plasmid pQR791 extracted from 3 different *E. coli* colonies. Lane (4) corresponds to the DNA marker. The darker band corresponds to ~5.4 Kbp super-coiled plasmid pQR791.

2.3.10 SDS-page

SDS-Page analysis was performed following method of Laemmli using a 7.5% separating and 4% stacking gel (Table 2-2). Gels were prepared according to the ProtoGel's protocol (National Diagnostics, UK). Gel cassette was assembled according to the manufacturer manual (Bio-Rad,UK). Clarified lysate or pure enzyme were mixed (1:1) with 2X protein sample buffer (Sigma-Aldrich, UK) and heated in boiling water for 5 min before loading 10 µL sample per well. The protein separation was then achieved by applying first 8 V/cm of stacking gel for 20 min, and 15 V/cm of resolving gel for a total of 60 min.

Table 2-2 Volumes of ProtoGel to achieve a 7.5% casting gel

Stacking gel		Resolving gel	
Chemical	[Volume] (mL)	Chemical	[Volume] (mL)
Protogel	1.3	Protogel	13.35
4x stacking buffer	2.5	4x resolving buffer	12.5
Deionised water	6.1	Deionised water	23.6
10% (w/v) ammonium persulfate*	0.05	10% (w/v) ammonium persulfate*	0.5
TEMED	0.01	TEMED*	0.05

*Chemicals typically added just before casting the gel.

Protein bands were visualised using Coomassie Blue. The gel was placed in a beaker, with 50 mL of coomassie blue staining solution 0.1% (w/v) Coomassie Blue R-250, 40% (v/v) methanol and 10% (v/v) acetic acid- and microwaved on full power for 1 minute. The stain was poured away and the gel was destained by boiling it for 10 minutes for every 100 mL of water. The gel was photographed using a Gel Doc 2000 system (Bioimaging systems, Cambridge).

2.3.11 Determination of DNA concentration

Two microlitres of the DNA sample solution were loaded in a nanodrop ND-1000 spectrophotometre (Fisher scientific,UK). The nanodrop was adjusted to zero using the same buffer used to dilute the DNA. Protein contamination was estimated from the A_{260}/A_{280} ratio. DNA quantity was determined by measuring absorbance at 260 nm. Concentration was calculated internally by the spectrophotometer by considering that 1.0 A_{260} unit of double-stranded DNA is equivalent to 50 μg/mL. DNA concentration was usually given in ng/μL.

2.3.12 DNA sequencing

DNA sequencing was performed by the sequencing service of the Wolfson Institute for Biomedical Research. 12 μL of DNA sample was given to the service at a concentration of 16 fmoles in a 1.5 mL eppendorf tube. Three main customs

primers were used to sequence the E. coli transketolase: TKN 5′-GATCCAGA GATTTCTGA-3′; TKmedSeq 5′-GTATGAAAGGCGAAAT GCCG TCTGACT-3′ and TKC 3′-CAAAAGAACTGCTGTAA-5′. Primers were prepared at a concentration of 5 pmoles/µL Results were usually obtained in *.ab1 format and subsequently analysed using BioEdit available on world wide web: (www. mbio.ncsu.edu/BioEdit/bioedit.html).

2.3.13 Medium scale purification of *E. coli* transketolase

His-tagged wild-type transketolase was over-expressed from E. coli XL10-Gold (section 2.3.3) containing the engineered plasmid pQR791 (section 2.3.7). Cell cultures were centrifuged at 2000 × g for 20 min. Cell paste was then resuspended in 250 mM Tris–HCl, pH 7.5, and the cells lysed by sonication (Section 2.3.5). Transketolase was purified from the cell-free extract using an XK 16/20 column, attached to an AKTA basic purification station (Amersham Bioscience, UK), packed with 25 mL of charged Ni-NTA resin (Amersham Biosciences, UK). Wild-type transketolase was eluted in an imidazole range of 420–450 mM (Figure 2-4), and then dialysed for 24 h against 250 mM Tris–HCl, pH 7.5, at 4 °C (section 2.3.15) . No protein impurities were visible by SDS-PAGE analysis (Figure 2-5).

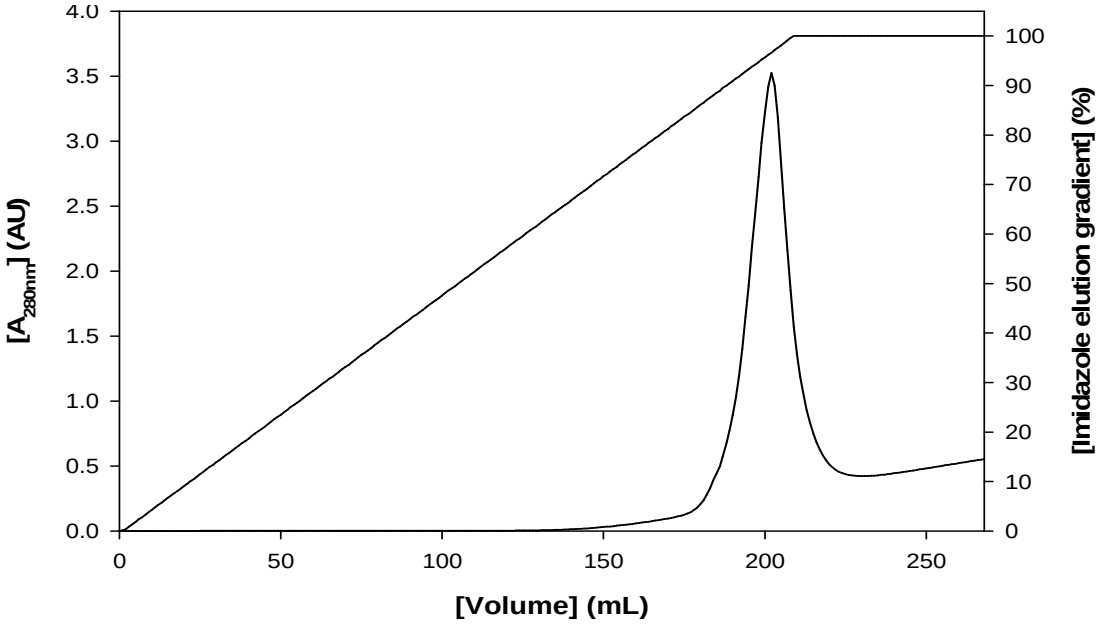


Figure 2-4 Wild-type transketolase purification elution profile.
Purification was performed using a XK 16/20 column properly packed with 25 mL of charged Ni-NTA Chelating Sepharose Fast Flow® resin (Amersham, Biosciences, Sweden) at a flow rate of 2.0 mL min-1. Equilibration, washing and elution steps were performed at a flow rate of 1.5 mL min-1 and monitored at 280nm. Protein was eluted by using an imidazole gradient at a flow rate of 15 mM min-1. Wild-type transketolase was eluted in an imidazole range of 430-450 mM.

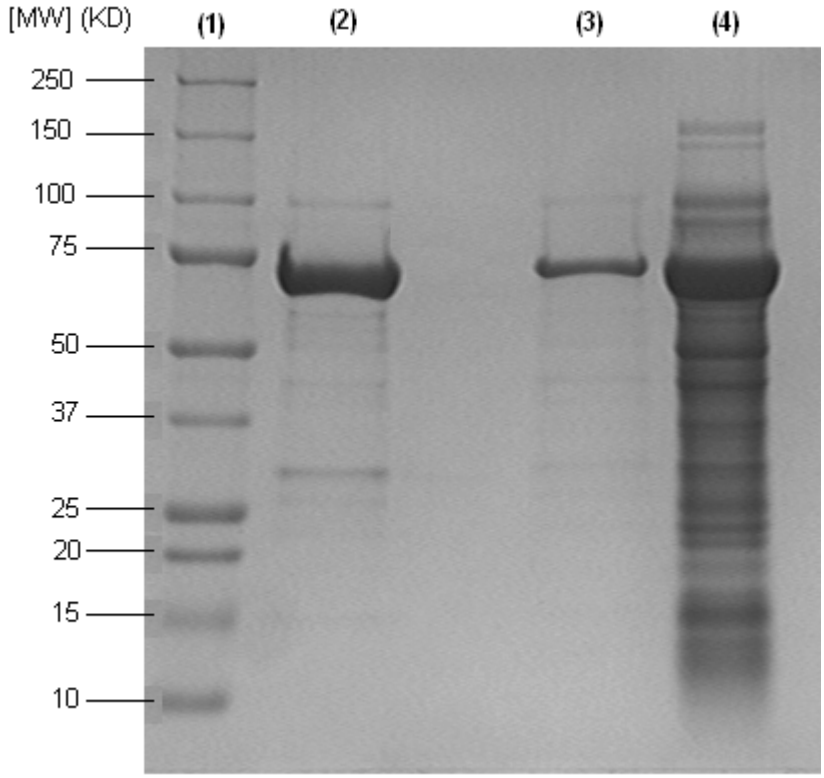


Figure 2-5 Top view of a 7.5% SDS-page gel of purified wild-type transketolase. Enzyme was purified by using a XK 16/20 chromatographic column. Sample lines are as follow: (1) Protein standard marker; (2) Purified transketolase; (3) A 1:30 purified transketolase; (4) Clarified lysate before purification. The darkest band at ~75KD correspond to the transketolase monomer.

2.3.14 Bench scale purification of *E. coli* transketolase

His-tagged mutant transketolase purification was performed using a His-Bind® Quick 900 cartridge (Novagen, Germany), packed with a pre-charged large diameter cellulose matrix with tethered Ni^{+2} complex immobilised on NTA. An over-night culture, usually 10 mL, was centrifugated at 2000 x g for 20 min. Cell paste was then resuspended in 250 mM Tris–HCl, pH 7.5, and the cells lysed by sonication (section 2.3.5). Particulates were removed from the cell extract by filtration through a 0.45µm syringe filter (Whatman scientific, UK). Purification was carried out after equilibrating the cartridge with binding buffer according to the manufacture's protocol. After unbound proteins were washed away, transketolase was recovered by elution with a solution of 40 mM of EDTA at pH 7.0, and then dialysed for 24 h against 250 mM Tris–HCl, pH 7.5, at 4 °C

(section 2.3.15). No protein impurities were visible by SDS-PAGE analysis (Figure 2-6). Total enzyme concentration was estimated by loading 2 μL of sample onto nanodrop equipment (section 0) followed by band densitometry using a Gel-Doc-it bioimaing systems with labworks 4.5 (Bioimaging systems, Cambridge) for image acquisition and quantification where appropriate.

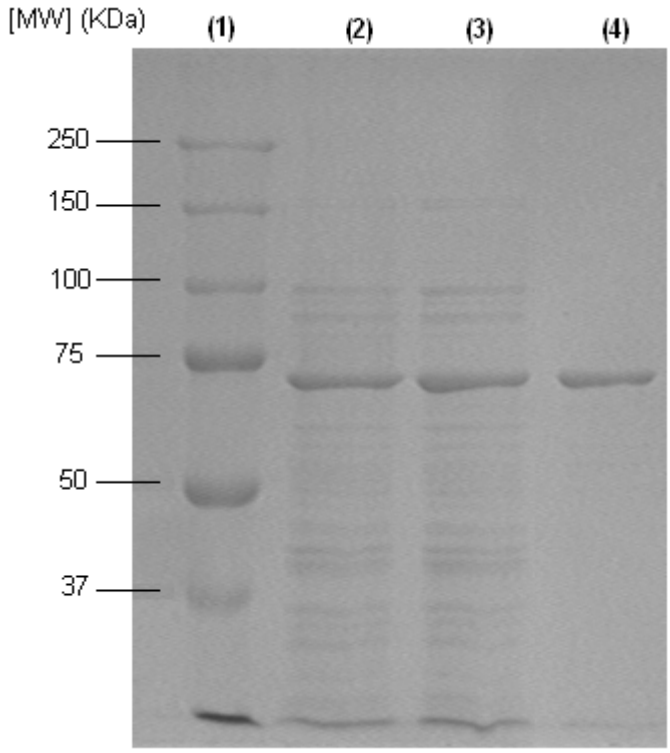


Figure 2-6 Top view of a 7.5% SDS-page gel of purified wild-type transketolase using a bench scale method.
Samples were purified by using a 2mL pre-packed purification column. The gel has 4 different lines. Sample lines are as follow: (1) Protein standard marker; (2) & (3) Clarified lysate; (4) purified transketolase. The darkest band at ~75KD correspond to the transketolase monomer.

2.3.15 Protein Dialysis

Purified protein was usually transferred to SnakeSkin dialysis tubing with a 10,000-Da molecular-mass cutoff (Pierce Biotechnology Inc., USA) and dialysed for 24 h against 250 mM Tris–HCl, pH 7.5, at 4 °C. Buffer was replaced at least two times during dialysis.

2.3.16 Protein Concentration

Protein concentration was determined by absorbance at 280 nm by loading 2 μL of sample in a nanodrop ND-1000 spectrophotometre (Fisher scientific,UK), assuming a monomeric molecular weight (MW) of 72260.82 g mol^{-1} and an extinction coefficient (ε) of 93905 L mol^{-1} cm^{-1} (Pace et al., 1995). Pure transketolase was stored in 250 mM Tris–HCl, pH 7.5, at 4 °C for a maximum of 2 weeks without loss of activity, and with no precipitation visible.

Chapter 3

3 Structural stability of *E. coli* transketolase to urea denaturation

3.1 Introduction

Many studies have demonstrated the industrial potential of *E. coli* TK for biocatalysis (Hobbs et al., 1996; Bongs et al., 1997; Mitra et al., 1998; Alexandre et al., 2002). TK catalyses the stereo selective transfer of a two-carbon ketol group to an aldose sugar, which results in asymmetric carbon–carbon bond synthesis (Datta & Racker, 1961) making the enzyme of particular interest for the biocatalytic synthesis of complex carbohydrates, and other high value compounds (Villafranca & Axelrod, 1971; Bolte et al., 1987; Myles et al., 1991; Kobori et al., 1992; Hecquet et al., 1994; Hecquet et al., 1996; Andre et al., 1998; Humphrey et al., 2000). Such stereo chemical control of carbon–carbon bond formation is of utmost importance in organic synthesis due to the increasing demand for enantiopure pharmaceuticals (Morris et al., 1996; Koeller & Wong, 2001).

Even though *E. coli* TK has a potential for industrial application, it is gradually deactivated during prolonged biocatalytic processes. Previous studies have highlighted oxidation (Brocklebank et al., 1999), deactivation by aldehyde substrates (Bongs et al., 1997), irreversible denaturation at pH values below 6.5 and loss of TPP cofactor (Mitra et al., 1998) as the most prominent stability issues in the biocatalytic process for E. coli TK. Enzyme oxidation and deactivation by substrate can potentially be minimised respectively, by the addition of a reducing agent such as dithiothreitol (DTT), and by the feeding of substrate in biocatalytic processes. As for loss of cofactors, enzyme activity decreased when excess cofactor was removed by gel filtration, indicating that both TPP and Mg^{2+} dissociate easily from *E. coli* holo-transketolase under all

pH conditions suitable for biotransformation reactions (pH 6.5–9.5) (Mitra et al., 1998). However, the results did not confirm whether the cofactor is directly dissociated from the enzyme, or whether an intermediate conformation is first formed with cofactor still bound in an inactive state, as observed during the reconstitution of yeast holo-TK from the apo-enzyme (Selivanov et al., 2003; Esakova et al., 2005). The loss of cofactor would require future modification of the enzyme to maintain stability, in a manner that would depend on the mechanism of cofactor dissociation or protein denaturation. For example, if the first denaturation step involves dimer dissociation, then chemical cross-linking or enzyme immobilisation would potentially be useful for stabilising transketolase. However, if the first step were the dissociation of the cofactors then protein engineering for improved stability would be more useful.

For the first time, in this work the early events that occur on the denaturation of *E. coli* TK have been characterised. Fluorescence intensity and CD spectroscopy was used to monitor the unfolding process of homodimeric transketolase. All the experiments using intrinsic fluorescence were supported by using size exclusion chromatography (SEC) to determine the different species of the enzyme under denaturant conditions and Dynamic Light Scattering (DLS) to assess the molecular size of the globular enzyme under physiological and denaturant conditions to understand the deactivation mechanism of TK under biocatalytic conditions. During the studies, urea was used as the chemical denaturant to determine the structural relationships between species on an unfolding pathway. It is worth emphasizing that even though urea is not used in the bioprocess, the TK unfolding pathway described, might potentially be similar to those that occur with other denaturants such as temperature and pH.

This has a significant impact on the use of *E. coli* TK as a biocatalyst, for which it is valuable to understand the conformational changes that may lead to the gradual inactivation of the enzyme under biocatalytic process conditions

(Dalby et al., 2007). Such an understanding is required before enzyme variants or alternative process conditions can be rigorously explored which improve the stability during biocatalysis.

Currently advances in protein engineering, for instance directed evolution (Dalby, 2003), will allow the creation of a large number of enzyme variants in a search for hits with improved stability. Certainly, in a more industrial aspect of protein design for proteins in industrial processes, several biophysical properties often have to be optimised simultaneously (Lilie, 2003). For example there may be a need for stability and activity at both high temperature and either alkaline or acid pH. For the reason described above, a high throughput determination of protein stability would be advantageous to reduce mainly time-consuming and ultimately volume sample. Finally, the fixed volume method (Aucamp et al., 2005) was evaluated using a commercial plate reader.

3.2 Materials and Methods

3.2.1 Preparation of protein samples

Purified transketolase was diluted to 1 mg mL^{-1} (13.8 µM) in a total of 200 µL of 25 mM Tris–HCl, pH 7.5, and 5 mM DTT, either with or without 5 mM $MgCl_2$, 0.5 mM TPP cofactors. Samples were then diluted to 2000 µL in denaturation experiments by the addition of urea solutions to obtain the required final urea concentration, or cofactor solutions to initiate binding reactions. Urea or cofactor solutions were prepared on the day of the experiments in 25 mM Tris–HCl, pH 7.5, and 5 mM DTT. The final sample concentrations were 0.1 mg mL^{-1} (1.38 µM) transketolase, 25 mM Tris–HCl, pH 7.5, and 5 mM DTT, either with or without 0.5 mM $MgCl_2$ and 0.05 mM TPP, in a total volume of 2000 µL. Saturating concentrations of cofactors were used at 6-fold, and 50-fold greater than the native dissociation constants for TPP and Mg^{2+}, respectively (Sprenger et al., 1995).

3.2.2 Activity of holo-TK at 0, 2 and 3.8 M urea

Holo-TK was pre-equilibrated with cofactors at pH 7.5 for 1 h at 25 °C then incubated with urea and DTT at pH 7.5 for a further hour, giving final concentrations of 0.1 mg mL^{-1} (1.38 µM) transketolase, 25 mM Tris–HCl, pH 7.5, 0.5 mM $MgCl_2$, 0.05 mM TPP, 5 mM DTT and either 0, 2 or 3.8 M urea. Reactions were initiated by addition to a final concentration of 50 mM Li-hydroxypyruvate (HPA), and 50 mM glycolaldehyde (GA) pre-incubated in 25 mM Tris–HCl, pH 7.5, 5 mM DTT (section 2.2.7). L-erythrulose production was then monitored by using the short 5 min HPLC method (section 2.2.8).

3.2.3 Cofactor binding-time measured by intrinsic fluorescence

Apo-TK samples were prepared as described in section 3.2.1, before addition to final concentrations of 0.5 mM $MgCl_2$ and 0.05 mM TPP at 25 °C, to initiate the binding reaction. Samples were then pumped through the fluorescent flow cell of an LS30 spectrophotometer (Perkin Elmer, Beaconsfield, UK) for 10 s to prior to measurement. Fluorescence intensity measurements at 340 nm were taken every 30 s for 15 min, then every 60 s for a further 35 min, using excitation at 280 nm. All kinetic data were fitted by linear regression using Kaleidagraph (Synergy Software, Reading, USA).

3.2.4 Time-course to attain denaturation equilibrium at 3.8 and 7.2 M urea, using intrinsic fluorescence

The time taken for denatured TK samples to reach equilibrium was determined by dilution of enzyme to 3.8 M and 7.2 M urea. Samples and urea solutions were prepared as described in section 3.2.1. Final concentrations in 2 mL, were 0.1 mg mL^{-1} (1.38 µM) transketolase, 25 mM Tris–HCl, pH 7.5, 0.5 mM $MgCl_2$, 0.05 mM TPP, 5 mM DTT, and either 3.8 M or 7.2 M urea. Fluorescence intensity at 340 nm was measured every 5 min for 140 min at 3.8 M urea, and every 20 min for 10 h at 7.2 M urea, using an LS30 spectrophotometer with excitation at 280 nm, all at 25 ˚C. All kinetic data were fitted by linear regression using the Kaleidagraph software package to double- or single-exponential curves.

3.2.5 Equilibrium denaturation monitored by fluorescence intensity

Preparation of protein samples was carried out as described above using fresh urea solutions ranging from 0 to 7.2 M. Samples were incubated for 4 h at 25 ˚C, as suggested by the time taken to reach equilibrium at 7.2 M urea, and fluorescence intensity at 340 nm was measured using an LS30 spectrophotometer with excitation at 280 nm.

3.2.6 Equilibrium refolding of transketolase from 3.8 M urea

Apo-TK or holo-TK was prepared at 3.8 M urea (close to mid-point of third transition), 1 mg mL^{-1} enzyme in 200 µl of 5 mM DTT, 25 mM Tris–HCl, pH 7.5, with 5 mM $MgCl_2$ and 0.5 mM TPP for holo-TK, and without cofactors for apo-TK. Protein was incubated for 1 h at 25 ˚C then refolded to urea concentrations ranging down to 0.38 M urea, or further denatured up to 4.2 M urea, by the addition of 1.8 mL of various urea solutions in 5 mM DTT, 25 mM Tris–HCl, pH 7.5. Fluorescence intensity was measured at 340 nm with excitation at 280 nm using a LS30 luminescence spectrophotometer after incubation for 1, 4 and 8 h.

3.2.7 Equilibrium denaturation monitored by circular dichroism (CD)

CD spectra (190–300 nm) were recorded on an AVIV 202 SF spectrometer (AVIV Associates, Lakewood, NJ) at 25 ˚C using a quartz precision cell cuvette with a path length of 0.1 mm. One volume of wild-type transketolase at 0.5 mg mL^{-1} (with 2.5 mM $MgCl_2$, and 0.25 mM TPP for holo-TK), in 25 mM Tris–HCl, pH 7.5 and 5 mM DTT was added to four volumes of urea solution in 25 mM Tris–HCl, pH 7.5 and 5 mM DTT, to give various final concentrations (0–7.2 M) and incubated for 4 h. CD spectra were recorded at 1 nm intervals and averaged for 4 s at each wavelength. A spectrum for 25 mM Tris–HCl, pH 7.5 and 5 mM DTT buffer was subtracted from each recording. Mean residue ellipticity (θ_{MRE}, deg cm^2 $dmol^{-1}$) at 222 nm were used to monitor urea-induced denaturation transitions for both apo-TK and holo-TK.

3.2.8 Size-exclusion chromatography (SEC)

Size exclusion chromatography under denaturant conditions was performed with a protein pak column 300SW, 7.5 mm × 300 mm, MW range 10,000–400,000 kDa (Waters Corporation, MA, USA) on a Dionex AD20 (Dionex Corporation, CA, USA) HPLC system with UV detection. Apo-TK or holo-TK was prepared at 0.1–2 mg mL^{-1} (27.6 µM), in 100 µL 25 mM Tris–HCl, pH 7.5, 5 mM DTT and, for holo-TK only, 10 mM $MgCl_2$ and 1.0 mM TPP. Each sample was incubated in 0, 2, 3.8 or 7.2 M urea at 25 ˚C for: 30 min at 0 M urea; 1 h at 2 M and 3.8 M urea; and 4 h at 7.2 M urea, injected as a 10 µL sample onto the column pre-equilibrated at the appropriate denaturant concentration, and eluted at 0.5 mL min^{-1}, 25 ˚C.

3.2.9 Dynamic light scattering (DLS)

The particle size distribution of transketolase under denaturant conditions was measured at 25 ˚C with a Zetasizer Nano S (Malvern Instruments Ltd., Worcestershire, UK). Samples of 1 mg mL^{-1} protein were prepared in 25 mM Tris–HCl, pH 7.5, 5 mM DTT without cofactors, and also with cofactors (0.5 mM TPP and 5 mM $MgCl_2$) at 0, 3.8 and 7.2 M urea. Samples were incubated for 1 h before data was acquired. The effect of urea viscosity was subtracted from each recording to eliminate the noise produced by it. Data were acquired in triplicate, with a low volume disposable sizing cuvette with a path length of 1 cm, and the hydrodynamic diameters and molecular weights of each sample were calculated from the averaged-measurements using the Zetasizer Nano Series software V.4.20 (Malvern Instruments Ltd., Worcestershire, UK).

3.2.10 Equilibrium denaturation of *E.coli* TK monitored by intrinsic fluorescence using a commercial plate reader

High throughput denaturation experiments were performed using the fixed-volume method described somewhere (Aucamp et al., 2005). Protein samples were prepared in 250mM Tris-HCl at pH 7.5 with or without cofactors (section 3.2.1) and 5mM DTT. Pure transketolase samples were manually loaded into a 96-well microplate (Greiner Bio-one Ltd, UK) at a concentration of 1 mg mL^{-1}. Wells were filled with the same buffer with different urea concentrations ranged from 0-7.2 M to give a final protein concentration of 0.1 mg mL^{-1} in a final volume of 300 µL. Fluorescence emission was recorded at 340 nm (± 10 nm bandpass) after excitation at 280 nm using a FLOUstar optima (BMG Labtechnologies Ltd., UK) microplate reader fitted with a syringe pump for automation of denaturant injection. Samples were mixed at 350 rpm for 60 sec and incubated for 4 h as suggested in section 3.3.2.

3.3 Results and discussion

3.3.1 Reconstitution of holo-TK from apo-TK

The reconstitution of *E. coli* TK might be expected, though not assumed, to proceed via the same two-step mechanism that is observed for yeast TK (Scheme 1). Assuming the principle of microscopic reversibility, the denaturation of holo-TK may also then be expected to proceed via a similar inactive and cofactor-associated $(TK\lozenge TPP)_2$ intermediate. Both of these hypotheses are directly investigated in this chapter. The reconstitution of *E. coli* holo-TK from 0.1 mg mL^{-1} apo-TK, 0.5 mM $MgCl_2$ and 0.05 mM TPP, at pH 7.5 monitored using fluorescence intensity (Figure 3-1), fits best to a double-exponential curve with the rate constants k_1 = 0.64 (±0.02) min^{-1} ($t_{1/2}$ = 1.6 min), and k_2 = 0.041 (±0.005) min^{-1} ($t_{1/2}$ = 24 min), and negative amplitudes of 67 (74%) and 24 (26%) fluorescence units, for the fast and slow phase respectively, in agreement with the two-step process previously observed for *S. cerevisiae* TK. This shows that 60 min is the minimum time required to fully reconstitute holo-TK.

$$(\text{apo-TK})_2 + \text{TPP} \underset{K_{-1}}{\overset{K_1}{\rightleftharpoons}} (\text{TK}\diamond\text{TPP})_2 \xrightarrow{K_2} (\text{holo-TK})_2$$

Scheme 3-1. The reconstitution of yeast holo-TK was initially shown to proceed by two-step mechanism and either via inactive TPP-bound dimeric intermediate $(TK\diamond TPP)_2$ at higher consentrations.

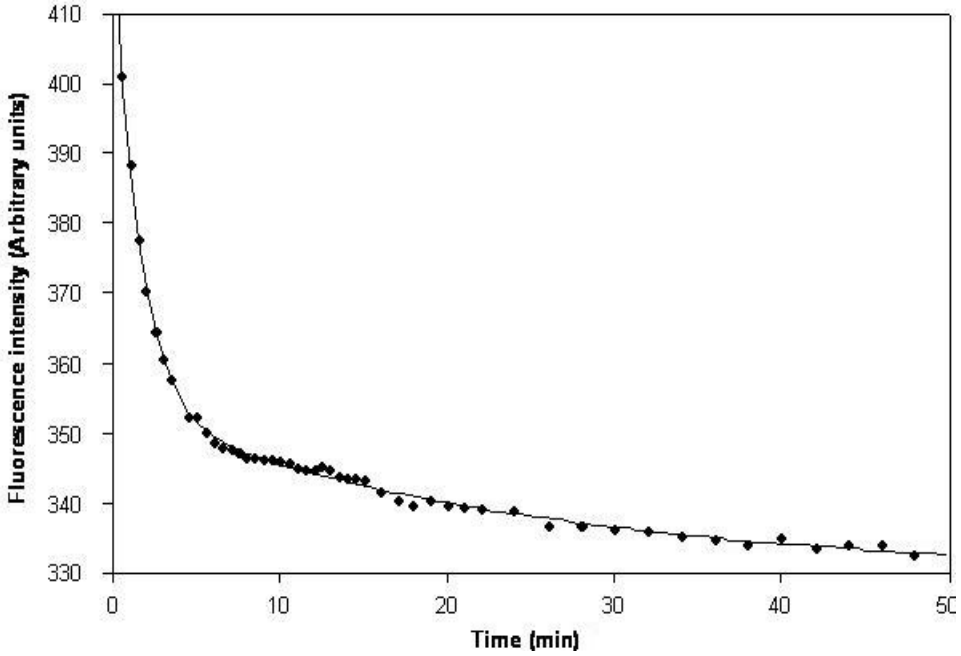


Figure 3-1 Reconstitution of Holo-TK from Apo-TK.
Reconstitution of E. coli holo-TK from 0.1 mg mL−1 apo-TK, 0.5 mM MgCl2 and 0.05 mM TPP, at pH 7.5 monitored using fluorescence intensity. Data fit best to a double-exponential curve.

As the cofactor binds to apo-TK the fluorescence intensity decreases, indicating an increase in the polarity of the environment around one or more tyrosine or tryptophan residues, as has been observed previously for the yeast TK (Kochetov, 2001). Several potentially fluorescence quenching interactions are formed upon cofactor binding, including the close proximity between: Tyr182 and Mg^{2+} (6.4 Å); Tyr182 and the TPP phosphate (6.8 Å); Tyr440 and the methyl group of the TPP pyridine ring (3.4 Å); and between Trp196 and the TPP phosphate (7.6 Å), TPP thiazolium ring (8.5 Å), and the Mg^{2+} ion (8.4 Å). Two loops (residues 187-198, and 383-394) were identified previously for yeast TK as becoming more ordered upon cofactor binding (Nikkola et al., 1994) and (Sundstrom et al., 1992). The Trp196 residue is in an equivalent loop in *E. coli* TK, whereas the other equivalent loop contains residue Trp390, which is in close proximity to the polar side-chains of Asn403 and Glu366 in the holo-TK structure (Figure 3-2). The formation of these interactions is thus likely to cause the observed decrease in fluorescence intensity upon cofactor binding.

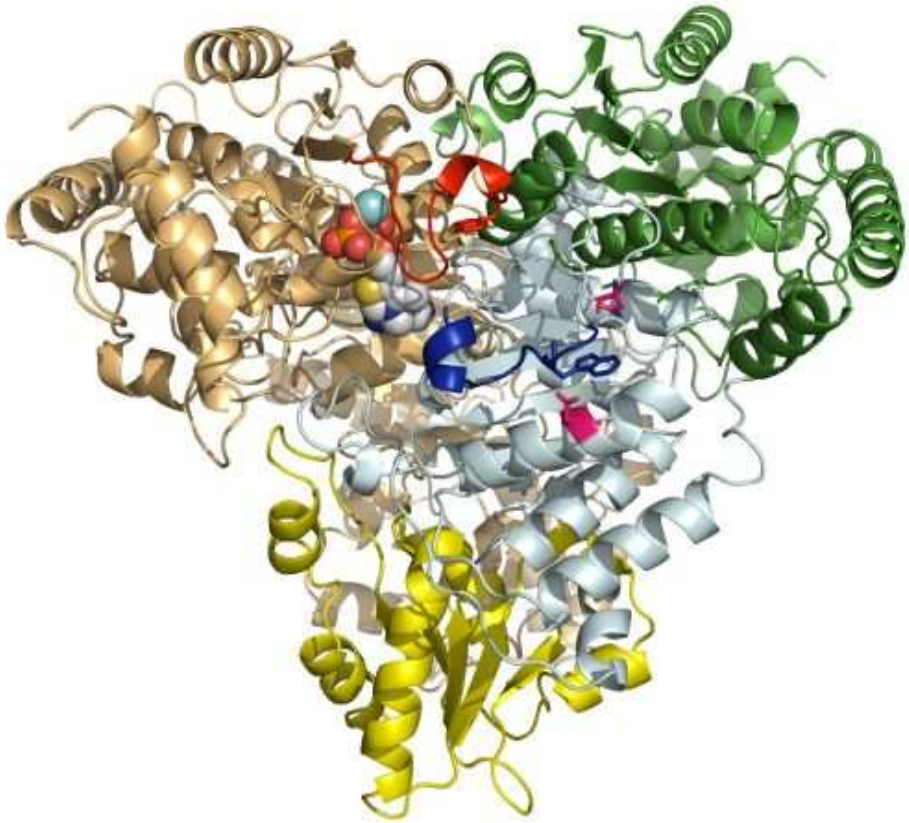


Figure 3-2 Structure of *E. coli* transketolase.
PDB file was obtained from 1qgd.pdb (Littlechild et al., 1995). One monomer is shown as ribbons in light brown. The second monomer is shown in ribbons with the PP-domain (green), Pyr domain (light blue), and the C-terminal domain (yellow). The loops equivalent to those not structured in yeast apo-TK are highlighted in red (residues 187–198) and dark blue (residues 383–393), and one tryptophan residue is highlighted in each loop (Trp196 and Trp390 respectively). Two polar residues (Asn403 and Glu366) which interact with Trp390 are highlighted in magenta. For one active-site only, the TPP cofactor is highlighted in CPK colours and as spheres, and the metal ion cofactor is highlighted as a cyan sphere. Figure generated with PyMol (DeLano, W.L. (2002), The PyMOL Molecular Graphics System on World Wide Web http://www.pymol.org).

3.3.2 Time-course to attain denaturation equilibrium at 3.8 and 7.2 M urea measured by intrinsic fluorescence

Time courses for the denaturation of holo-TK in the presence of cofactors, at both 3.8 M and 7.2 M urea, were monitored by the fluorescence intensity at 340 nm to determine the time at which the unfolding equilibrium was reached. The denaturation of holo-TK at 3.8 M urea (Figure 3-3) fits a double-exponential curve with rate constants $k_3 = 0.37 \pm 0.05$ min^{-1} ($t_{1/2} = 2.7$ min), and $k_4 = 0.08 \pm 0.01$ min^{-1} ($t_{1/2} = 12.5$ min), and positive amplitudes of 156 (58%) and 111 (42%) fluorescence units, respectively. By contrast, the denaturation of holo-TK at 7.2 M urea (Figure 3-4) fits a single-exponential curve with the rate constant $k_5 = 0.015 \pm 0.001$ min^{-1} ($t_{1/2} = 67$ min), and a negative amplitude of 36 fluorescence units. The equilibration within 1 h at 3.8 M urea, and 4 h at 7.2 M, contrasts with only 10 min previously seen for yeast TK (Esakova et al., 2005). The two denaturation transitions at 3.8 M urea are expected to occur more rapidly at higher denaturant concentrations, and so the observation of a single transition at 7.2 M urea suggests that they are now within the dead-time of the experiment, and that the third and slowest transition only becomes apparent at above 3.8 M urea. The opposing fluorescence amplitudes also suggest that the two fast steps are distinctly different from the slower step, where the intensity increased during the dead-time at 7.2 M urea for the two fast steps and then decreased more slowly for the observable third step. It is possible that denaturation at lower temperatures would make the first two rapid steps slower and therefore observable at 7.2 M urea.

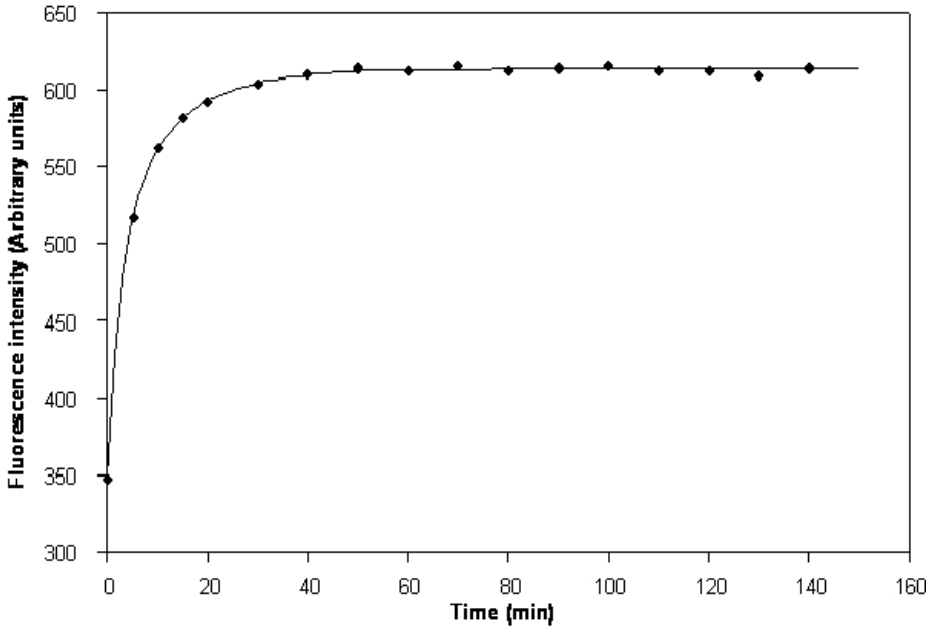


Figure 3-3 Time courses for the denaturation of holo-TK in the presence of cofactors at 3.8 M urea.
Denaturation of holo-TK (0.1 mg mL−1; 0.5 mM MgCl2 and 0.05 mM TPP, at pH 7.5) at 3.8 M urea was monitored by fluorescence intensity at 340 nm to determine the time at which the unfolding equilibrium was reached. Data fit best to a double-exponential curve.

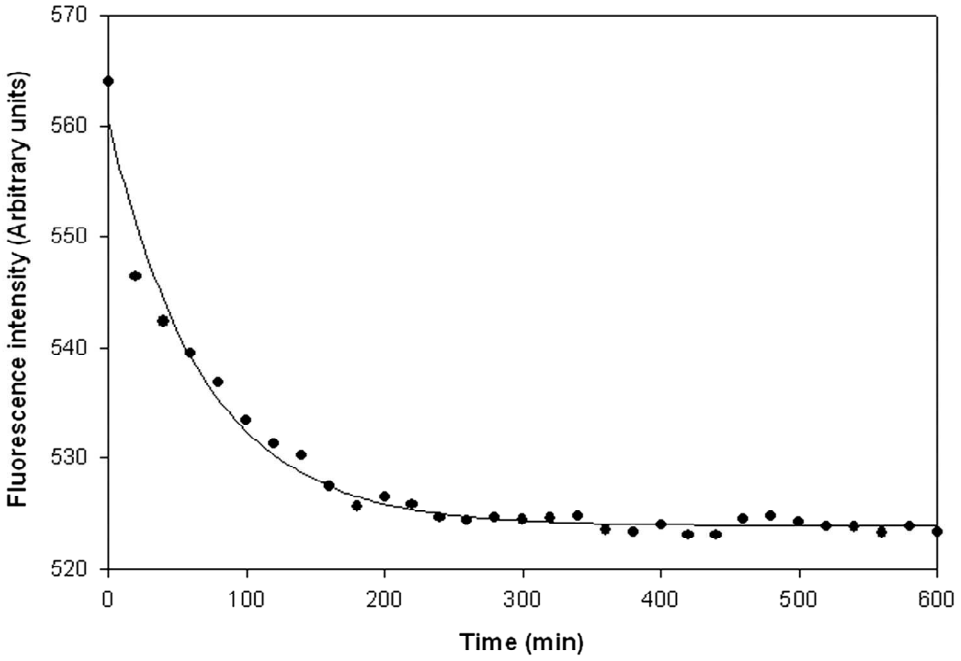


Figure 3-4 Time courses for the denaturation of holo-TK in the presence of cofactors at 7.2 M urea.
Denaturation of holo-TK (0.1 mg mL−1; 0.5 mM MgCl2 and 0.05 mM TPP, at pH 7.5) at 7.2 M urea was monitored by fluorescence intensity at 340 nm to determine the time at which the unfolding equilibrium was reached. Data fit best to a single-exponential curve.

3.3.3 Equilibrium denaturation monitored by fluorescence

Denaturation equilibrium transitions of the homodimeric *E. coli* TK in the absence of cofactors (apo-TK), the presence of each individual cofactor (TPP-TK, and Mg-TK), and the presence of both cofactors (holo-TK), were monitored by fluorescence intensity at 340 nm (Figure 3-5). Apo-TK and holo-TK both show two major denaturation transitions occurring at 2.0–3.2 M urea, and 3.2–5.5 M urea which result in an overall increase and the decrease in fluorescence, respectively. In addition, holo-TK shows a significant deviation from the signal of apo-TK at 0–2 M urea, which indicates the presence of at least one further transition at the start of the denaturation process for holo-TK. The amplitudes of these three transitions are in agreement with the three steps observed by the

denaturation kinetics at 3.8 and 7.2 M urea for holo-TK. The multiphase denaturation of *E. coli* TK differs substantially from the single transition of increasing fluorescence seen previously for yeast holo-TK (Esakova et al., 2005). However, these differences in the fluorescence signal alone are not sufficient to conclude that yeast-TK and E. coli-TK follow different denaturation pathways without evidence from other probes of protein unfolding on the yeast TK, such as circular dichroism.

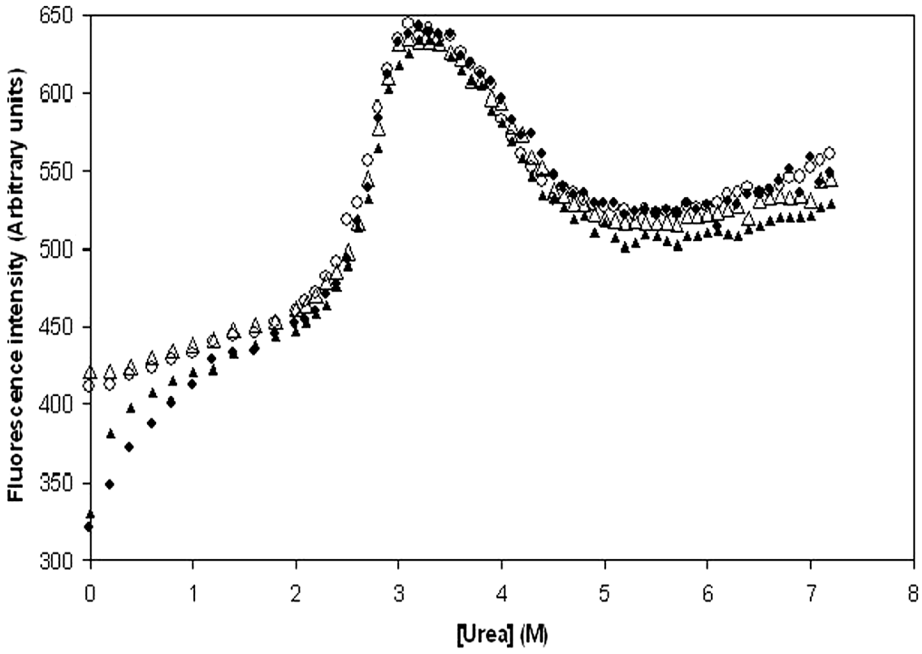


Figure 3-5 Equilibrium urea-denaturation of transketolase monitored by intrinsic fluorescence.
Samples of 0.1 mg mL^{-1} (○) apo-TK, (●) holo-TK, (▲) TPP-TK, and (△) Mg-TK were prepared at different concentrations of urea (0–7.2 M), in 5 mM DTT, 25 mM Tris–HCl buffer, pH 7.5, and where appropriate 0.05 mM TPP, and 0.5 mM $MgCl_2$. Fluorescence intensity was measured at 340 nm with excitation at 280 nm after incubation at 25 ˚C for 4 h.

3.3.4 Equilibrium refolding

The reversibility of the early transition for TK denaturation was determined by refolding both apo-TK and holo-TK from 3.8 M urea and monitoring their intrinsic fluorescence (Figure 3-6) after 1–8 h. In addition, some samples were

further denatured at up to 4.2 M urea. The fluorescence signals from unfolding and refolding diverge for both apo-TK and holo-TK below 3 M urea and do not change beyond 1 h, suggesting that refolding does not go to completion, and that some of the protein forms a misfolded or soluble aggregate state. Protein precipitation was not observed at any time. The presence of cofactors suppresses refolding more than in their absence, with recovery of only 60% of the native fluorescence signal at 0.38 M for holo-TK, and 72% of the native apo-TK fluorescence signal in the absence of cofactors. The timescale for complete reconstitution of holo-TK from folded apo-TK, is much shorter than the equilibrium times used here, which implies that the inactive holo-TK intermediate is not accumulating in this case. Previous work has indicated that exposure to pH below 6.5 results in the irreversible denaturation of TK (Mitra et al., 1998). Our results suggest that the enzyme may also become partially inactivated by misfolding or aggregation during biocatalysis at pH 7.5, but only if a significant degree of denaturation occurs in the bioreactor, potentially through shear effects or interaction with high concentrations of chemical reagents.

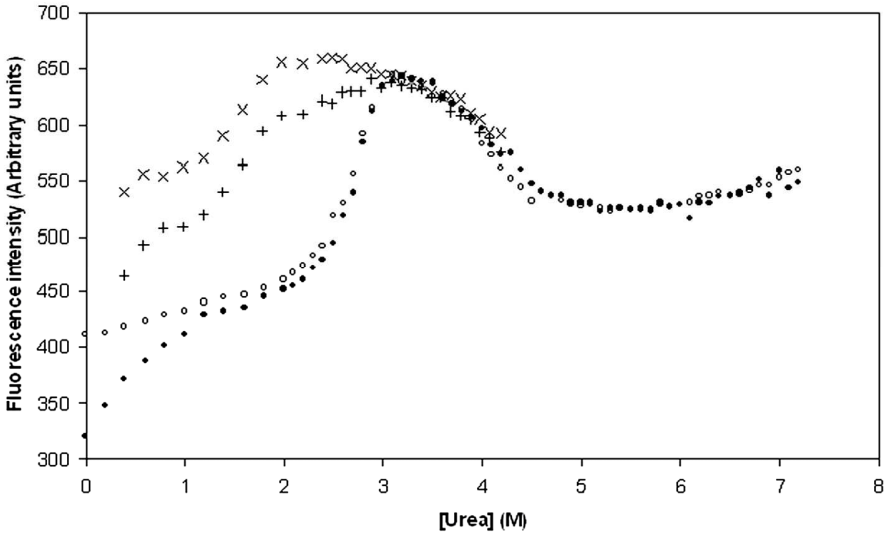


Figure 3-6 Equilibrium refolding from 4.2 M urea of holo-TK and apo-TK.
Samples of holo-TK (×) in 0.5 mM $MgCl_2$, 0.05 mM TPP, and apo-TK (+) without cofactors, at 0.1 mg mL^{-1} (1.38µM) in 5 mM DTT, 25 mM Tris–HCl, pH 7.5. Equilibrium denaturation of (●) holo-TK and (○) apo-TK is shown for comparison. Protein was incubated for 1 h at 25 °C then refolded by 10-fold dilution into appropriately buffered urea. Fluorescence intensity was measured at 340 nm with excitation at 280 nm.

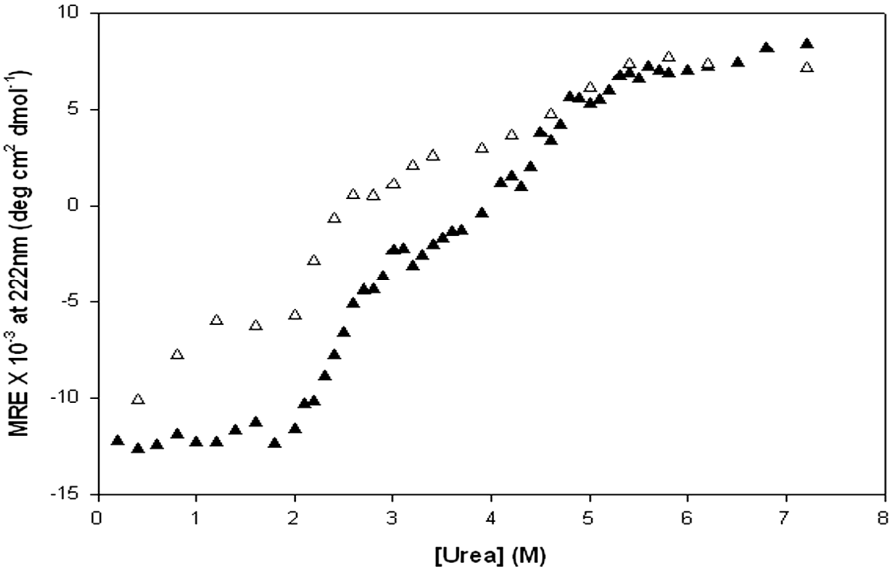


Figure 3-7 Equilibrium urea-denaturation of holo- and apo-transketolase, monitored by circular dichroism.
Denaturation of (▲) holo-TK and (△) apo-TK at 222 nm. Holo-TK (0.5 mg mL^{-1}) was prepared in 25 mM Tris–HCl, pH 7.5, 2.5 mM $MgCl_2$, 0.25 mM TPP and 5 mM DTT and the appropriate concentration of urea. Apo-TK (0.1 mg mL^{-1}) was prepared in 25 mM Tris–HCl, pH 7.5, and 5 mM DTT and the appropriate concentration of urea. Samples were incubated for 4 h at 25 °C before full spectra (195–300 nm) were acquired.

3.3.5 Equilibrium denaturation monitored by circular dichroism

The CD spectra of holo-TK and apo-TK at 0 M urea both display a peak of negative ellipticity with a minimum at 222 nm (Figure 3-8), consistent with the mostly α-helical protein structure (Littlechild et al., 1995) and (Manavalan & Johnson, 1983). Apo-TK contains marginally less secondary structure content which most likely reflects the increased flexibility of the two cofactor-binding loops as observed for yeast apo-TK (Nikkola et al., 1994) and (Sundstrom et al., 1992). Denaturation of holo-TK and apo-TK with urea concentrations ranging from 0 to 7.2 M resulted in increasing loss of secondary structure as shown for holo-TK in (Figure 3-9). The secondary structure content in both cases changed from mostly α-helical at 0 M urea to essentially random-coil at 7.2 M. The equilibrium denaturation of both holo-TK and apo-TK is shown in (Figure 3-7) as a function of mean residue ellipticity at 222 nm for 0–7.2 M urea. Two cooperative transitions were apparent for holo-TK at 2.0–3.5 M urea and 3.5–5.5 M urea, with only a baseline drift in ellipticity over the 0–2 M urea range. The same transitions were observed for apo-TK, with a baseline drift at 0–2 M urea, a cooperative transition at 2–3.5 M urea, but unfolding with relatively poor cooperativity above 3.5 M urea. The CD signals for apo-TK and holo-TK at 222 nm only converge at above 4 M urea.

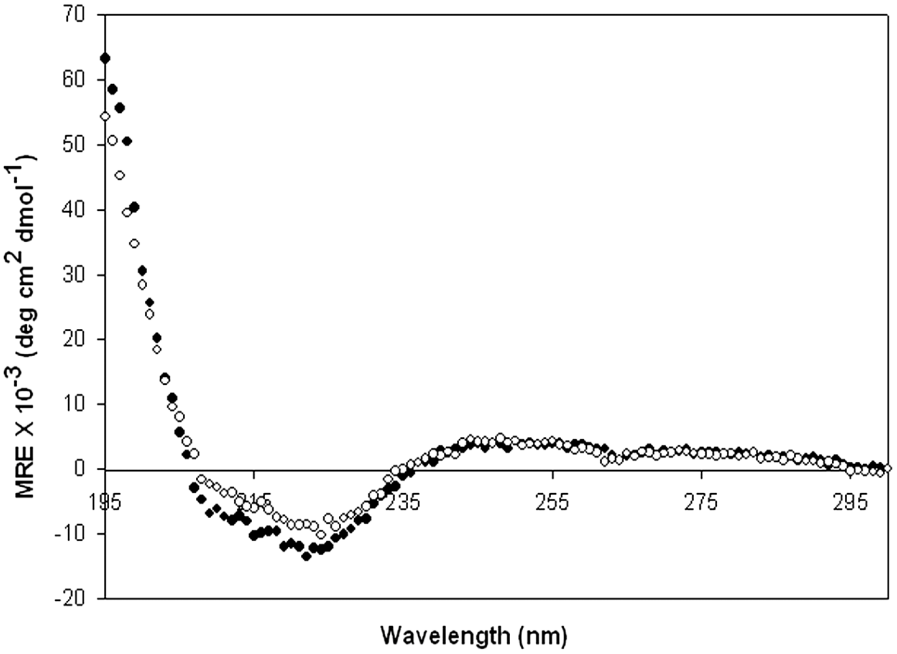


Figure 3-8 Comparison of circular dichroism spectra of holo- and apo-TK. Samples of (●) holo-TK (2.5 mM $MgCl_2$, 0.25 mM TPP) and (○) apo-TK at 0.5 mg mL^{-1} (6.9µM) in 25 mM Tris–HCl, pH 7.5, and 5 mM DTT, 0 M urea. Samples were incubated for 4 h at 25 °C before full spectra (195–300 nm) were acquired.

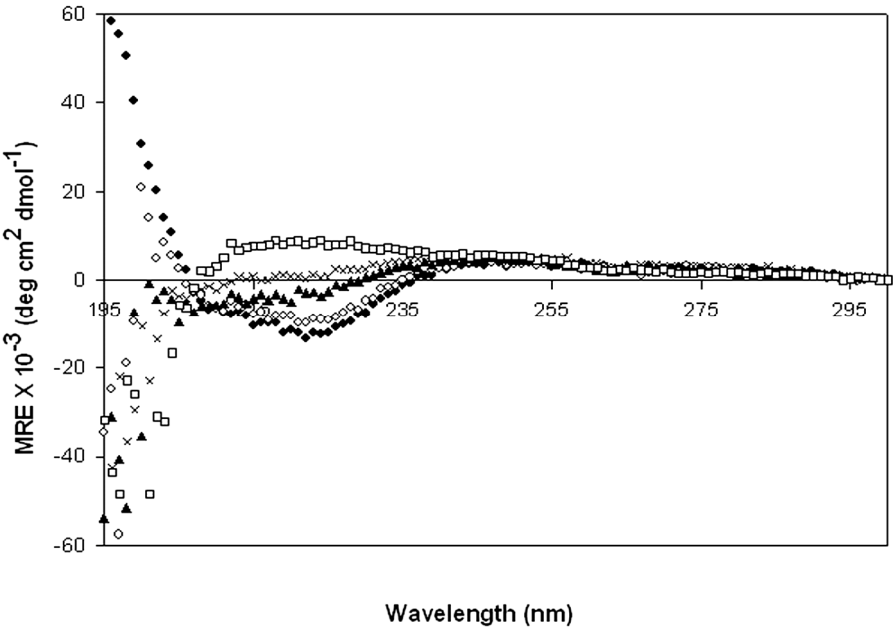


Figure 3-9 Circular dichroism spectra of holo-transketolase.
Holo-TK, 0.5 mg mL^{-1} (6.9µM), in 25 mM Tris–HCl, pH 7.5, 2.5 mM $MgCl_2$, 0.25 mM TPP and 5 mM DTT, at: (●) 0 M urea; (○) 2.0 M urea; (▲) 3.2 M urea; (×) 3.8 M urea; and (□) 7.2 M urea. Samples were incubated for 4 h at 25 °C before full spectra (195–300 nm) were acquired.

3.3.6 Denaturation transition at 0–2 M urea

Comparison of the fluorescence intensity of the apo-TK at 0 M urea, to those obtained by the addition to apo-TK of Mg^{2+} only (Mg-TK), TPP only (TPP-TK), or both cofactors (holo-TK), indicates that TK becomes more structured in the presence of TPP alone, but not with Mg^{2+} alone (Figure 3-5). In addition, the binding of TPP produces a degree of structure similar to that of the fully reconstituted holo-TK dimer at 0 M urea. This is in agreement with previous studies on yeast TK in which TPP bound to apo-TK in the absence of the metal ion, though in a catalytically inactive form (Datta & Racker, 1961), and where Mg^{2+} was necessary for the catalytic activation of TPP-associated apo-TK (Selivanov et al., 2003) and (Egan & Sable, 1981).

Convergence of the holo-TK and TPP-TK fluorescence intensities with those of apo-TK and Mg-TK as urea is increased to 2 M (Figure 3-5), indicates that TPP has either dissociated from holo-TK, or that the local structure around the TPP binding-site has undergone a structural rearrangement. By contrast, the CD signal at 222 nm indicated only a slight loss of secondary structure for holo-TK at 2 M urea (Figure 3-9) and (Figure 3-7), and the secondary structure content of apo-TK remains consistently lower than that of holo-TK at 0–2 M urea, which suggests that at least one of the TPP and Mg^{2+} cofactors is still bound at 2 M urea. The fluorescence intensity change in holo-TK as urea concentration is increased from 0 to 2 M urea, may be caused by a local structural change which alters the polarising environment of the bound TPP cofactor, but without a significant change in secondary-structure content. It is not clear whether $MgCl_2$ has dissociated at 2 M urea as no difference in fluorescence can be seen between the apo-TK and the Mg-TK. However, the TPP-TK form of the enzyme is marginally less stable than the fully reconstituted holo-TK as indicated by the increase in fluorescence signal of TPP-TK at slightly lower urea, and confirms that Mg^{2+} partially stabilises the holo-TK as observed previously for yeast TK (Esakova et al., 2005).

The size-exclusion chromatography (SEC) for holo- and apo-TK at 0, 2, 3.8 and 7 M urea (Table 3-1), provided further evidence that holo-TK does not denature to apo-TK at 2 M urea. Holo-TK (Figure 3-10) was in a fully compact and active dimeric form at 0 M urea, with a single peak observed at 14 min. At 2 M urea, holo-TK also gave a single peak at 13.5 min suggesting that the protein retained the native-like compact dimer form, though in a slightly more expanded form than at 0 M urea. At 3.8 M urea holo-TK gave two peaks with 24% of the protein in the native-like compact dimer form at 13.2 min, and 76% in an expanded form at 9.5 min. Apo-TK (Figure 3-11) at 0 M urea gave a single peak at 11.1 min corresponding to a native structure that is more expanded and distinct from the native holo-TK. At 2 M urea, apo-TK gave two peaks with 97% of the protein in the native compact dimer form at 11.2 min, and 3% eluting as the same expanded form at 9.5 min as observed for holo-TK. Holo-TK does not co-elute with the native apo-TK form, and the intermediate state observed by fluorescence and CD for holo-TK at 2 M urea is still as compact as the native holo-TK form.

Table 3-1 Retention times and fraction of total protein of holo- and apo-TK estimated by size exclusion chromatography (SEC).

[Urea] (M)	Retention time, minutes (Fraction of total protein, %)					
	Holo-TK			Apo-TK		
0	14.0	(100%)		11.1	(100%)	
2	13.5	(100%)		11.2	(97%)	9.5 (3%)
3.8	9.5	(76%)	13.2 (24%)	9.5	(97%)	11.2 (3%)
7.2	10.8	(100%)		10.5	(100%)	

Samples contained 2 mg mL^{-1} apo-TK or holo-TK in 0, 2, 3.8 or 7.2 M urea, 5 mM DTT, 25 mM Tris-HCl, pH 7.5, and for holo-TK, 1 mM TPP, and 10 mM $MgCl_2$. Samples were eluted through a protein pak column 300SW (7.5 mm i.d x 300 mm length, MW range 10,000–400,000 kDa).

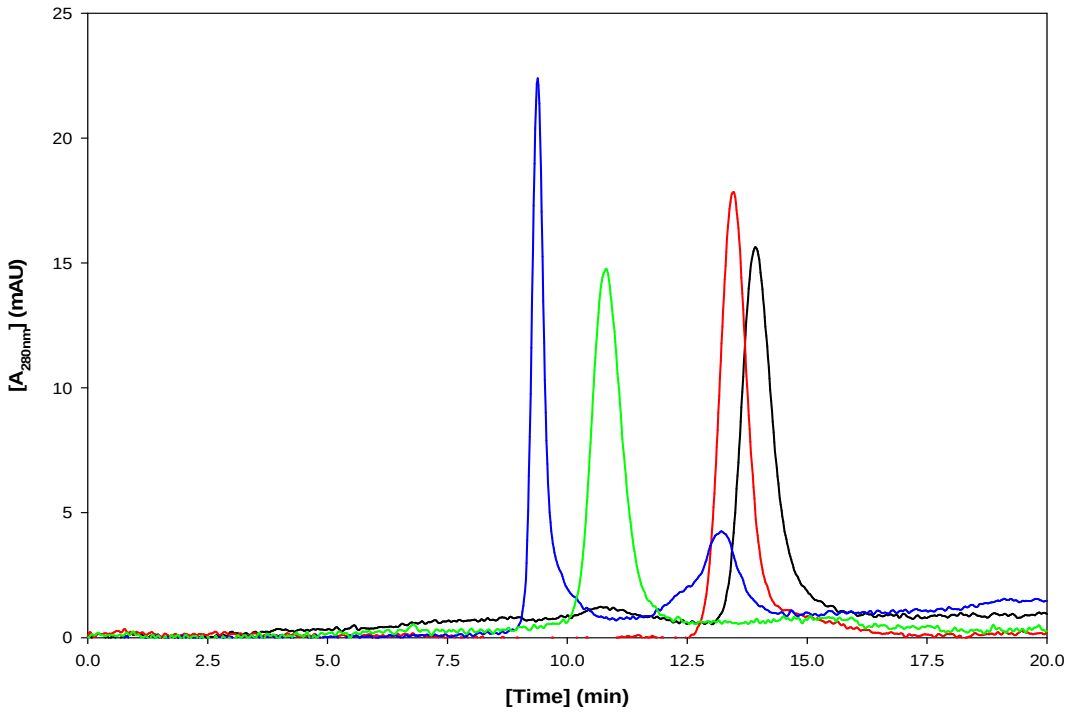


Figure 3-10 Size-exclusion chromatography of holo-transketolase.
Holo-TK prepared at 2 mg mL^{-1} in 25 mM Tris.HCl, pH 7.5, 10 mM $MgCl_2$, 1.0 mM TPP and 5 mM DTT, eluted by size exclusion at: Solid lines are: (Black) 0 M urea; (Red) 2 M urea; (Blue) 3.8 M urea; and (Green) 7.2 M urea respectively. Samples were incubated for (0 M urea) 30 min; (2 M urea) and (3.2 M urea) 1 h and (7.2 M urea) 4 h at 25 °C prior to elution through a protein pak column 300SW (7.5 mm i.d x 300 mm length, MW range 10,000–400,000 kDa).

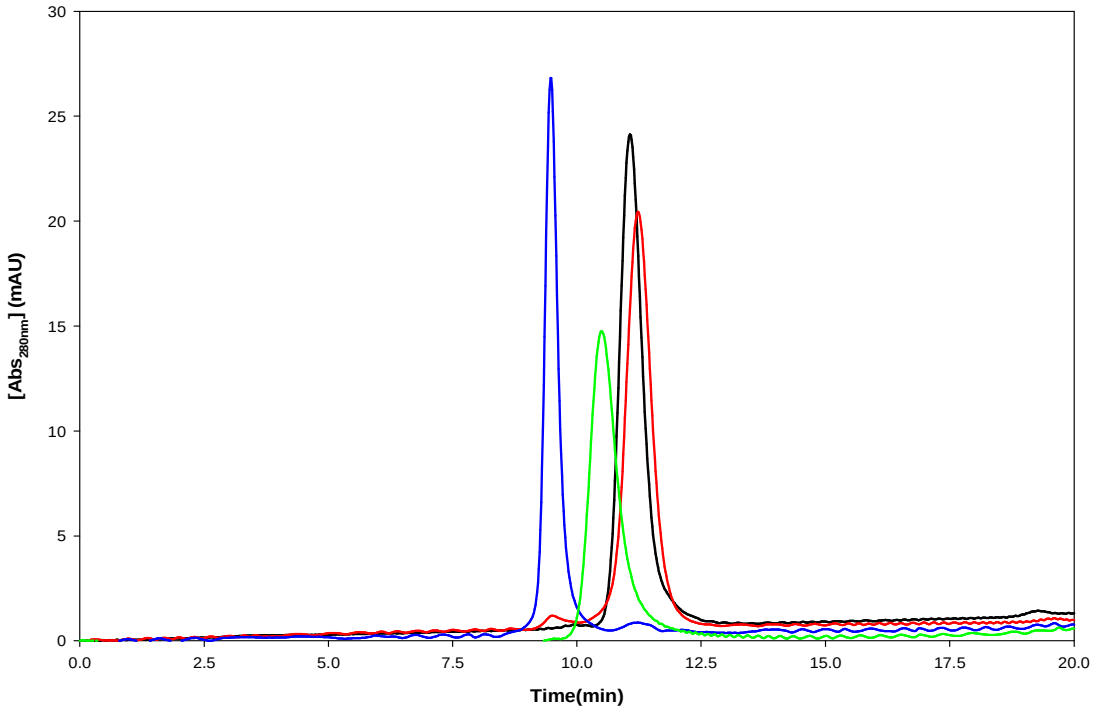


Figure 3-11 Size-exclusion chromatography of apo-transketolase.
Apo-TK prepared at 2 mg mL^{-1} in 25 mM Tris.HCl, pH 7.5 and 5 mM DTT, eluted by size exclusion at: (Black) 0 M urea; (Red) 2 M urea; (Blue) 3.8 M urea; and (Green) 7.2 M urea. Samples were incubated for (0 M urea) 30 min; (2 M urea) 1 h and (3.2 M urea) 4 h at 25 ^{o}C prior to elution through a protein pak column 300SW (7.5 mm i.d x 300 mm length, MW range 10,000–400,000 kDa).

Partial denaturation of the two cofactor-binding loops (Asn185-Trp196, and Leu382-Gly392) in *E. coli* holo-TK at 2 M urea would expose the intrinsically fluorescent Trp196 and Trp390 residues to solvent but also remove them from their native polar environments, resulting in the observed increase in fluorescence intensity (Figure 3-12). In contrast, the marginal loss of secondary structure would not be easily observed by CD or SEC. Furthermore, the fluorescence intensity of yeast TK shows only one denaturation transition, and at higher concentrations of urea, which could be explained by the presence of a tryptophan (Trp196) in one of the cofactor-binding loops of *E. coli* TK that is not present in yeast TK, whereas both enzymes have an equivalent tryptophan residue in the second cofactor-binding loop (Figure 3-12). This suggests that the early denaturation transitions may simply be unobservable by fluorescence intensity for yeast TK, and also that the fluorescence change observed for *E. coli*

TK at 0–2 M urea is due to at least partial denaturation of the first cofactor-binding loop (Asn185-Trp196).

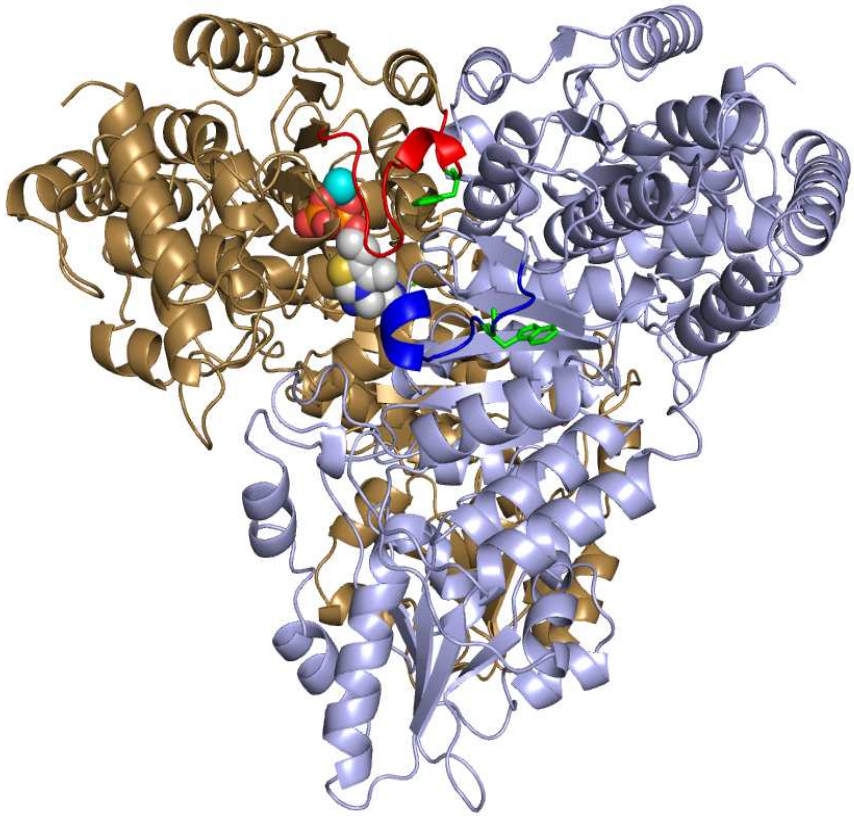


Figure 3-12 Loops not structures in *E.coli* transketolase.
PDB file was obtained from 1qgd.pdb (Littlechild et al.,1995). Monomer are shown as ribbons in light brown and light blue. The loops equivalent to those not structured in yeast apo-TK are highlighted in red (residues 187-198) and dark blue (residues 383-393), and one tryptophane residue is highlighted (green) in each loop (Trp196 and Trp390) respectively. For one active-site only, the TPP cofactor is highlighted in CPK colours and as spheres, and the metal ion cofactor is highlighted as a cyan sphere. Figure generated with PyMol (DeLano,W.L., 2002), The PyMol Molecular Graphics System on the World Wide Web htt://www.pymol.org).

An intermediate state was formed during the reconstitution of yeast and now also *E. coli* holo-TK from apo-TK in which, for yeast TK at least, the cofactors were bound in an inactive form which slowly converted to the fully active native-state. It is possible that this intermediate is similar to that seen during the

equilibrium denaturation of *E. coli* holo-TK at 2 M urea. Indeed holo-TK incubated for 1 h at various urea concentrations gave 90% conversion to l-erythrulose after 1 h reaction with β-HPA and GA at 0 M urea, whereas no activity was observed at 2 or 3.8 M urea (data not shown), consistent with the inactive intermediate formed during reconstitution. The intermediate observed at 2 M urea may also have been present in previous experiments under biocatalytic process conditions, for which cofactors were reported to dissociate easily from *E. coli* TK at all pHs, as measured by activity after removal of excess cofactors by gel filtration (Mitra et al., 1998). The cofactors may dissociate via the inactive intermediate form upon removal of the excess cofactor in solution, which has significant implications for the stability of the enzyme during biocatalysis.

3.3.7 Denaturation transition at 2–3.5 M urea

The fluorescence intensity for *E. coli* TK increases to a maximum at 3.2 M urea (Figure 3-5), indicating further spatial distancing of aromatic amino acids from the quenching capability of internal highly polar groups. The transitions coincide with those observed by CD as urea is increased from 2 to at least 3.5 M urea (Figure 3-7) and are sigmoidal, cooperative transitions. The very close proximity of the 2–3.5 M urea transition observed by CD, to the next one at above 3.5 M urea, indicates that the fluorescence intensity peak at 3.2 M urea may not be at the exact point for which an intermediate is most populated. The precise position of the fluorescence peak is dependent on the destructive interference and relative signal intensity changes that occur in opposing directions for the two transitions. The CD data indicate that an intermediate state occurs at closer to 3.5 M urea and has approximately 50% of the native holo-TK secondary structure.

The CD signals at 222 nm for apo-TK and holo-TK did not converge until above 4.0 M urea. This may at first suggest that the transition for holo-TK at 2–3.5 M urea ends with a different equilibrium denaturation state to that present for apo-TK. Also, on first inspection this transition might reasonably be expected to be due to dimer dissociation, by analogy with previous observations for pyruvate decarboxylase unfolding (Killenberg-Jabs et al., 2002). However, further analysis by size-exclusion chromatography and dynamic light-scattering suggest otherwise.

The size-exclusion chromatography of apo-TK and holo-TK (Table 3-1) at 3.8 M urea gave similar results with respectively 97% and 76% of the protein eluting in the expanded form at 9.5 min. The remaining protein elutes as their respective native states at 11.2 min for apo-TK and at 13.2 min for holo-TK, supporting the cooperative nature of the transition from different compact native states in the two TK samples at 2 M urea to a common intermediate at 3.8 M urea. This also indicates the complete dissociation of cofactors from holo-TK at 3.8 M urea without formation of the compact form of apo-TK. The observation that both apo-TK and holo-TK undergo similar fluorescence and CD transitions from 2.0 to 3.5 M urea, which are only both convergent at above 3.8 M urea suggests that the dissociation of cofactors from holo-TK is accompanied by significant denaturation of structure which also unfolds in apo-TK. The observation of a residual population of the native-like proteins by SEC at 3.8 M urea may contribute to the slightly different CD signals at 222 nm.

The decreased elution time of 9.5 min for the 3.8 M urea intermediate indicates a structure that is less compact than either of the native proteins, as is often caused by an increased volume of the protein molecules upon the binding of and denaturation by urea (Endo et al., 1983), or potentially by low-order aggregates formed from partially denatured intermediates (Pohl et al., 1994). Similarly, dynamic light-scattering (DLS) of holo-TK and apo-TK at 0 M urea estimated their molecular weights (MW) to be 145–157 kDa (data not shown)

with hydrodynamic diameters of 10.3 nm (Table 3-2, in agreement with the 10.2 nm dimeric protein (Littlechild et al., 1995). At 3.8 M urea, both apo- and holo-TK gave an apparent particle size of 32–33 nm. The increased size of the protein at 3.8 M urea most favours an intermediate state that is a partially unfolded and expanded form of the dimeric protein, rather than an even more greatly expanded monomer. Such a dimeric intermediate state has also been observed recently for the homodimeric enzyme transaminase (Deu & Kirsch, 2007). DLS measures the diffusion coefficient of a particle in solution, and the hydrodynamic radius is estimated based on a spherical approximation, which can be assumed for a typical globular protein (Murphy, 1997). However, the monomeric TK would be an elongated, non-globular form with a length of at least 9.6 nm as measured across the longest axis in the compact structure of holo-TK (1qgd.pdb), and while denaturation increases the particle diameter, any deviation from the hard-sphere approximation of a globular structure would lead to a decreased apparent particle diameter as estimated by DLS. Characterisation of the denaturation to 7.2 M urea further supports the existence of an expanded dimer form at 3.8 M urea as discussed below.

Table 3-2 Hydrodynamic diameter (nm) of holo- and apo-TK estimated by dynamic light scattering (DLS).

	Hydrodynamic diameter (d), nm	
[Urea] (M)	Holo TK	Apo TK
0	10.3	10.4
3.8	33.0	31.8
7.2	22.7	22.3

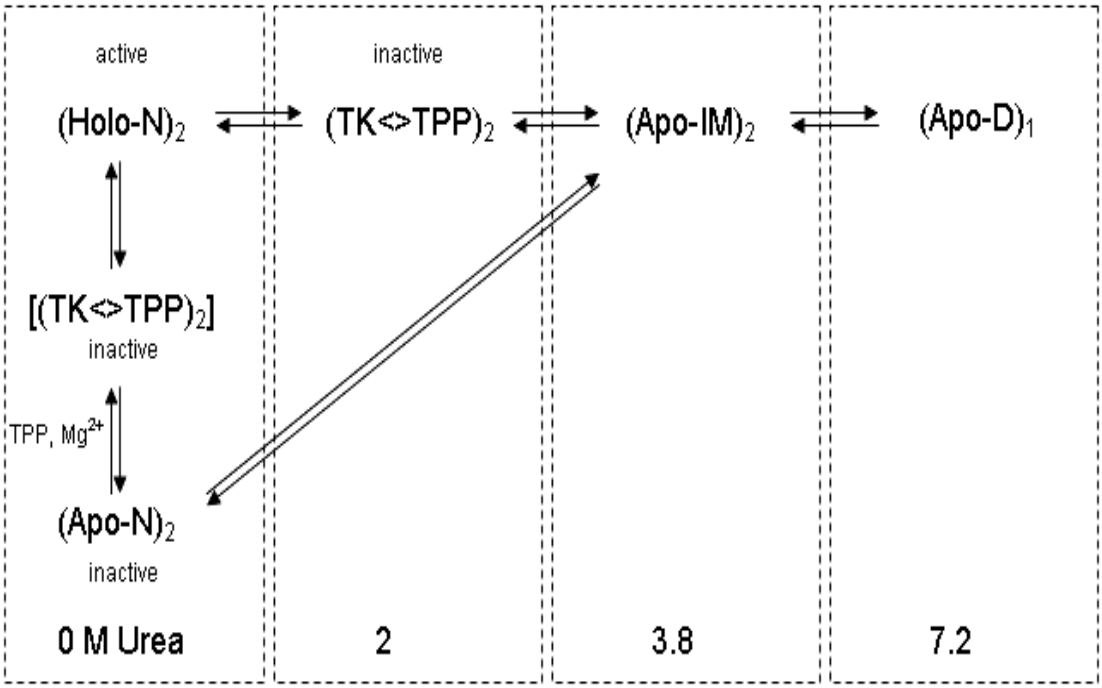
Samples contained 1 mg mL^{-1} protein in 25 mM Tris-HCl, pH 7.5, 5 mM DTT with or without cofactors (0.5 mM TPP and 5 mM $MgCl_2$) and 0, 3.8 or 7.2 M urea.

3.3.8 Denaturation transition at 3.8–7.2 M urea

The apparently cooperative decreases in fluorescence intensity for both apo-TK and holo-TK as urea increased from 3.2 to 7.2 M, were accompanied by less cooperative CD signal transitions, particularly for apo-TK. These deviations hint at the independent denaturation of more than one secondary structure element, only one of which has an effect on the intrinsically fluorescent amino acids. It is possible that dimer dissociation and also the monomer unfolding could both occur in the 3.8–7.2 M urea region, and the loss of cooperativity may reflect these events occurring independently, though there is no direct evidence for this. The CD spectra of both apo-TK and holo-TK at 7.2 M urea contain essentially random-coil, suggesting that the protein is fully denatured. However, the order of events for dimer dissociation and monomer denaturation at 3.8–7.2 M urea is not conclusively determined here. No protein concentration dependence was observed down to the detection limits for all transitions measured by fluorescence and SEC over the range 0.02–2.0 mg mL^{-1} transketolase (data not shown), indicating that much lower concentrations would be required to affect the dimer dissociation transition.

Holo-TK and apo-TK at 7.2 M urea both eluted by SEC at 10.5–10.8 min, indicating a common denatured form that is more compact, and/or of smaller size than the intermediate form observed in 3.8 M urea at 9.4 min (Table 3-1). This is also confirmed by the smaller radius shown in DLS at 7.2 M (Table 3-2). The observation of a smaller form at 7.2 M urea additionally supports the existence of a dimer at 3.8 M urea, as it is unlikely to have arisen from the formation of a more compact state at the higher urea concentration.

Based on a collective examination of the data, the denaturation pathways of apo-TK and holo-TK are proposed to follow those shown in Scheme 3-2, where N, IM and D denote native, intermediate and denatured states, respectively.



Scheme 3-2 Proposed denaturation pathways of apo-TK and holo-TK ; where N, IM and D denote native, intermediate and denatured states, respectively.

3.3.9 High Throughput equilibrium denaturation of tranketolase using a fixed-volume method.

Equivalent denaturation transition curves were observed using the fixed-volume method (Aucamp et al., 2005). Those curves obtained by the high throughput method and with a classical fluorimeter are depicted in Figure 3-13. The microplate method reproduces the curves obtained by classical fluorimeter including the deviation in fluorescence between apo-TK and Holo-TK in the range from 0M to 2M urea.

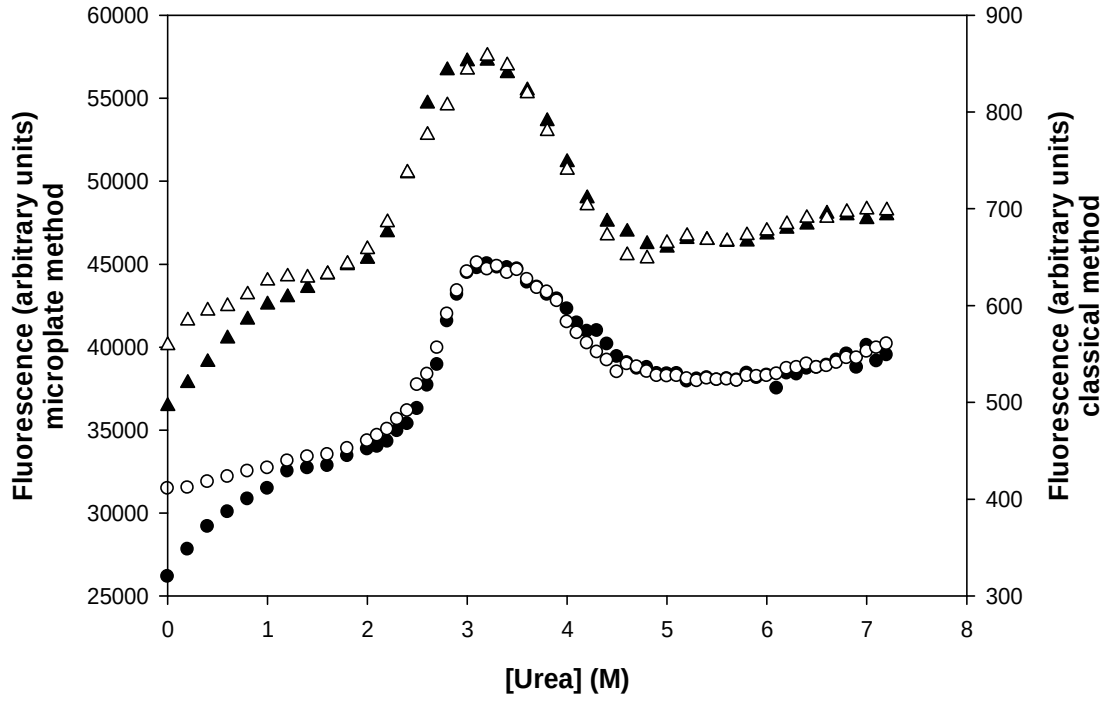


Figure 3-13 Comparison of microplate-based and classical fluorescence intensity.
Data were obtained with the fixed volume mode in microplates for wild-type (▲) holo-TK and (△) apo-TK, and in a standard fluorimeter for (●) holo-TK and (○) apo-TK. Samples of 0.1mgmL^{-1} TK were prepared at 0-7.2M urea in 5mM DTT, 25mM Tris-HCl, pH 5, and for Holo-TK with 0.05mM TPP, and 0.5mM $MgCl_2$, then equilibrated for 4 h prior to measurement. Measurements were performed at 340 nm as a function of urea concentration.

3.4 Conclusions

The denaturation of holo-TK by urea displays at least three transitions, where only the final equilibrium denaturation transition is the same for both apo-TK and holo-TK. This result is different to previous studies on yeast TK in which only a single transition was observed as measured by fluorescence intensity. The additional transitions observed for *E. coli* TK most likely result from a tryptophan residue (Trp196) that is present in the first cofactor-binding loop, but not present in the equivalent loop for yeast TK. This suggests that yeast TK may also denature by the same pathway as *E. coli* TK but that not all of the steps are observable by fluorescence alone.

Overall, data suggest a TK denaturation pathway shown in Scheme 3-2 that begins with the reverse of the reconstitution pathway shown in Scheme 1. The first denaturation event for holo-TK appears to be a local change in conformation at the cofactor-binding site that does not significantly alter the compactness or secondary structure of the homodimer. This holo-TK intermediate is more compact and contains more secondary structure than the native apo-TK structure, indicating that the cofactors are still bound. The kinetic and equilibrium denaturation data, coupled with size-exclusion chromatography and dynamic light-scattering suggest that an expanded dimer forms at 3.2–3.8 M urea from the denaturation of both apo-TK and holo-TK, indicating that the cofactors are no longer bound. However, further experiments at various TK concentrations are required to resolve the exact structural nature of the expanded intermediate form at 3.8 M urea.

The loss of TK activity upon removal of excess cofactor by gel filtration was previously ascribed to the complete dissociation of cofactor (Mitra et al., 1998). Our results explain this observation further as the dissociation of cofactor via the formation of a cofactor-associated but inactive intermediate. A similar intermediate is also observed during the kinetic reconstitution from apo-TK.

These observations provide potentially valuable information for improving biocatalytic processes that use TK, and imply that enzyme deactivation during prolonged reactions could be minimised by improving the interactions of the cofactors and their binding loops.

Finally, the fixed-volume method for denaturation of protein in microplate (Aucamp et al., 2005) gave a denaturation profile that were equivalent to that classical ones. This method reduces the sample volume and time without compromising the precision of the data obtained as all the unfolding transition curves were accurately detected.

Chapter 4

4 Characterisation of the stability of *E. coli* transketolase to different pH and Temperature

4.1 Introduction

Stability refers to the maintenance of a defined functional state (chemical and structural properties that are required for activity) under extreme conditions (Pace, 1990; Jaenicke & Bohm, 1998). On one hand, the stability of enzymes is often interpreted simplistically as thermal stability, i.e., a temperature beyond which the enzyme loses stability. Although this quantity is important, every statement of stability at a certain temperature depends on the exposure time and is thus often ambiguous. Also, for biocatalytic process applications, a more important quantity is the process or operational stability, which is the long-term stability under specified conditions (Bomma & Riebel, 2004). The stability of proteins is also known to be affected significantly dependant on salt concentration and pH (Matthew et al., 1985; Yang & Honig, 1993; Brocklehurst, 1994; Antosiewicz et al., 1994). Moreover, there are many structural factors that affect the stability of proteins such as the number of salt bridges, hydrogen bonds, hydrophobic core packing, among others that contribute to the protein free energy of unfolding ($\Delta G_{D,T}$) (Stigter et al., 1991; Jaenicke, 1991). For that reason when studying the stability of an enzyme in terms of either temperature, pH or any other destabilising agent, all these features must be taken into account.

Transketolase (TK) from *E. coli* has the potential for use in industrial applications, for the efficient large-scale production of sugar analogs (Woodley et al., 1996; Schenk et al., 1998). Examples include the biosynthesis of the aromatic amino acids (Draths et al., 1992), food industry (Hecquet et al., 1996); pharmaceuticals (Hecquet et al., 1994) and agrochemical industries TK is also an

important biocatalyst in stereo-specific carbon-carbon bond synthesis as it accepts a large number of aldehydes as substrates (Schorken & Sprenger, 1998; Turner, 2000). Even though many important reactions have been reported, many of the substrates and products were shown to be labile to alkaline environments owing to the labile substrate used (Mitra et al., 1998). 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), and is subject to oxidative deactivation at high concentrations of glycolaldehyde (Mitra et al., 1998), Many studies described in chapter 3 showed also that TK may also become partially inactivated by misfolding or aggregation during biocatalysis at pH 7.5 where a significant degree of denaturation can occur in the bioreactor, due for example to a high concentration of chemical reagents (Martinez-Torres et al., 2007).

A broader practical implementation of Holo-TK is, impeded by the fact that, it retains a high catalytic activity in a relatively narrow pH (6-9) or temperature (4-50^{o}C). The range shape of the pH and temperature optimum is determined by both the protein structure and the nature of the catalytic mechanism of the enzyme e.g. through a radical change in the pH or temperature, entropically driven denaturation or unfolding can quickly occur. When denaturation occurs it may happen that inter-polymer hydrophobic interactions can quickly lead to a non-specific and typically irreversible aggregation of the denatured polypeptide chains.

Similarly optimum pH have been reported for tranketolases ranging from 7.0 to 7.6 for *E. coli* TK using free-buffer (Mitra et al., 1998) or phosphate buffer (Schenk et al., 1998); pH 7.5- 7.6 for TK from baker's yeast (Datta & Racker, 1961); for human erythrocytes a pH of 7.7 (Himmo et al., 1988); for rabbit liver 7.8-8.0 (Masri et al., 1988); and a pH from 8.0 to 8.5 when using a glycylglycine buffer (Sprenger et al., 1995). It seems that the buffer effect is also important as it influences enzyme activity so considerations must be taken into

account and optimised in order for an enzymatic reaction to be accurate and reproducible when using different buffers in bioprocess conditions.

The optimum temperature for activity varies between transketolases depending on the conditions. For example, for rabbit liver TK it is around 40°C (Masri et al., 1988); for human erythrocyte transketolase after incubation at 55°C for 5 min it retained 50% activity (Takeuchi et al., 1986); whereas Porcine liver transketolase shows no activity loss at 50°C for 1h in the presence of thiamine pyrophosphate (TPP) (Philippov et al., 1980).

In this work, retaining the structural stability of the TK with minimal enzymatic inactivation in harsh conditions, such as high temperature and pH, is the principal target. The structural stability of TK under extremes of pH and temperature are studied using both functional (enzymatic activity) and an optical method to monitor structural changes. In both cases the pH and thermal inactivation of holo-TK is measured by circular dichroism and the decrease in enzymatic activity. Furthermore, the thermodynamics (stability) of the holo-TK are measured.

The study of conformational changes and the stability of native Holo-TK is a must when industrial applications have to be considered. Thus, understanding the inactivation mechanism linking it to denaturation phenomena is a major objective from both technical and scientific points of view. For example, they can be exploited for the engineering of new enzymes for important bioengineering applications and for the improvements of industrial processes. Indeed, improving stability may also prolong the half-life of transketolase under mild reaction conditions to increase its reusability in a process.

4.2 Materials and Methods

4.2.1 Thermal deactivation or residual activity of *E. coli* transketolase

Samples of 50 µL, usually ~0.5 mgmL^{-1} (6.9 µM), of pure or clarified lysate of wild-type tranketolase were subjected to heat-treatment to temperatures between 40 ^{o}C and 65 ^{o}C. Deactivation was initiated by placing samples into a water bath pre-heated to the desired temperature. Temperatures inside the samples were monitored using a digital wired-thermometer (Topac, USA). Samples were taken out from the water bath after 30 and 60min and immediately placed on ice, then they were incubated and assayed for relative activity at room temperature (25^{o}C) (section 2.2.7) and the concentration of L-erythrulose reaction product was estimated by the high accuracy HPLC method (section 2.2.8). Thermal deactivation or residual activity was determined by the amount of L-erythrulose produced by the enzyme after having been heated-up.

4.2.2 Thermal denaturation of *E. coli* Holo transketolase monitored by circular dicrohism

CD spectra (185–300 nm) and 222nm single wavelength were recorded on an AVIV 400 SF spectrometer (AVIV Associates, Lakewood, NJ), using a quartz precision cuvette with path lengths of 1mm and 0.1mm respectively. One volume of wild-type transketolase at 0.5 mg mL^{-1} (6.9 µM) (with 2.5 mM $MgCl_2$, and 0.25 mM TPP for holo-TK) was prepared in 25 mM Tris–HCl at pH 7.5. CD spectra were recorded at 0.5nm intervals and averaged for 4s at each wavelength. A spectrum for 25 mM Tris–HCl, pH 7.5 buffer was subtracted from each recording. The temperature was raised from 5° to 95°C at 1^{o}C intervals and then lowered to the initial temperature to check the reversibility by measuring the ellipticity. Mean residue ellipticity (θ_{MRE}, deg cm^2 dmol^{-1}) at 222 nm were used to monitor temperature-induced denaturation transitions for

holo-TK. Samples were incubated at each temperature for 2 min before taking either the whole wavelength or single measurement.

Protein stability curves as a function of protein stability were evaluated by fitting the signal at 222nm versus temperature to an apparent two-state model even though more than two states exist.

The equilibrium constant $K_{D,T}$ and the free energy of unfolding, $\Delta G_{D,T}$, were calculated defining two linear base-lines for the nature and denature states and using the Eq. (4-1) and (4-2) described below:

$$y_{N,T} = m_N T + b_N \tag{4-1}$$

$$y_{D,T} = m_D T + b_D \tag{4-2}$$

Where m_D and b_N are the slope and y-intercept of the native base-line, m_D and b_D are the values for the denatured base-lines, and T is the absolute temperature. By comparing the signal, A_T, at the different temperatures to the native and denatured base lines, the fraction of protein, $\alpha_{D,T}$, were determined by using Eq. (4-3):

$$\alpha_{D,T} = \left(\frac{A_T - y_{N,T}}{y_{D,T} - y_{N,T}} \right) \tag{4-3}$$

To estimate the apparent equilibrium constant for denaturation, $K_{D,T}$, equation (4-4) was used:

$$K_{D,T} = \frac{[D]}{[N]} = \frac{\alpha_{D,T}}{1 - \alpha_{D,T}} \tag{4-4}$$

Alternatively, T_m apparent, was also calculated by fitting the data to a two-state transition between a folded dimmer and unfolded monomer at a given protein concentration (Greenfield, 2006), assuming ΔCp=0, using software package SigmaPlot10.0 (SPSS UK Ltd., Woking,Surrey, England).

4.2.3 Thermal denaturation of *E. coli* transketolase monitored by dynamic light scattering

The thermal denaturation (Tm) of purified wild-type transketolase was measured with a Zetasizer Nano S (Malvern Instruments Ltd., UK). Samples of 1 mg mL^{-1} (13.8 µM) protein were prepared in 25 mM Tris–HCl, pH 7.5 with cofactors (0.5 mM TPP and 5 mM $MgCl_2$). Samples were incubated for 1 h before data was acquired. The protein solutions were spun at 4000 rpm for 10 min and filtered through a 100 nm pore size filter membrane (Whatman).

The temperature was raised from 20° to 60°C at 0.5°C intervals. The effect of Tris-HCl was subtracted from each recording to eliminate the noise produced by it. Data were acquired with a low volume disposable sizing cuvette with a path length of 1 cm, and the hydrodynamic diameters and molecular weights of each sample were calculated from the averaged-measurements using the Zetasizer Nano Series software V.4.20 (Malvern Instruments Ltd., Worcestershire, UK).

4.2.4 Structural Analysis of *E.* coli transketolase at different pH's monitored by circular dichroism

pH dependant stability of pure wild-type holo-transketolase, 0.5 mg mL^{-1} (6.9 µM) was monitored using circular dichroism in an Aviv CD spectrometer model 400 (AVIV Associates, Lakewood, NJ). The far-UV signal was measured at 222 nm, and the spectrum was recorded between 300 nm and 180 nm. The structural analysis of 0.5 mgmL wild-type transketolase were performed after incubating the sample, usually for 45 min, in 25 mM Tris–HCl with 5 mM $MgCl_2$, 0.5 mM TPP cofactors at different pH from pH 4 to pH 10.

4.2.5 Deactivation of *E. coli* holo-transketolase at different pHs

Pure and clarified lysate holo-transketolase solutions, 0.5 $mgmL^{-1}$ (6.9 µM), in 100 mM Tris–HCl with 2.5 mM $MgCl_2$, 0.25 mM TPP cofactors at pH 6.0, 7.0, 8.0 and 9.0 were incubated for 45 and 120min. Similarly, samples at pH 4.0 and 5.0 or pH 10.0 and 11.0 were prepared in 100 mM Bis-Tris or 100 mM CAPS buffer respectively with 2.5 mM $MgCl_2$, 0.25 mM TPP cofactors and incubates for 45 and 120 min. Substrate solutions, previously adjusted to the corresponding pH, were added to each sample to start the reaction. Usually a 1M HCl or a 1M NaOH solutions were used to adjust the solutions pH. After 45, 90 and 120 min a 0.2% (v/v) solution of TFA was added in a one to one ratio (1:1) to quench the reaction (section 2.2.7). Product concentration was estimated by HPLC (section 2.2.8)

4.2.6 Calculation of pKa values of all *E. coli* transketolase ionisable groups using PROPKA program

The PROPKA web interface 2.0 on World Wide Web (http: //propka.ki.ku.dk/ ~drogers/) program was used to calculate the pKa values of ionisable groups in *E. coli* transketolase amino acids. The structure of *E. coli* transketolase was obtained from 1qgd.pdb (Littlechild et al., 1995). The PDB file was first prepared and uploaded according to the website tutorial instructions. Results were acquired through the analysis of a PROPKA-file generated internally by the program (Delphine et al., 2008; Li et al., 2005).

4.3 Results and discussion

4.3.1 Thermal inactivation of wild-type *E. coli* transketolase

The effects of temperature on the activity of both the purified and clarified-lysate enzyme were analysed at a range from 40 to 65^{o}C. The inactivation consisted of incubating the samples for 30 and 60 min at the desired temperature (40-60^{o}C) (section 4.2.1), and then assaying for the relative activity at room temperature (25^{o}C). It should be emphasized that this approach measures either the ability of TK to refold upon cooling, or to retain the native structure at elevated temperatures. As the temperature was increased from 40 to 55 ^{o}C, and in each case as the time of incubation was increased from 0.5 to 1 hour, the measured residual activity at 25 ^{o}C for the pure enzyme also increased (Figure 4-1). The activity of the enzyme after incubation at 55 ^{o}C for 1 hour was around three times higher than that after incubation at 40 ^{o}C for 0.5 hours. Above 55 ^{o}C, the activity of the enzyme decreased precipitously to values 3-fold lower after incubation at 60 ^{o}C for 1 hour, and to no activity after incubation at 65 ^{o}C for 0.5 hours. An almost identical effect was observed for holo-transketolase in a clarified lysate.

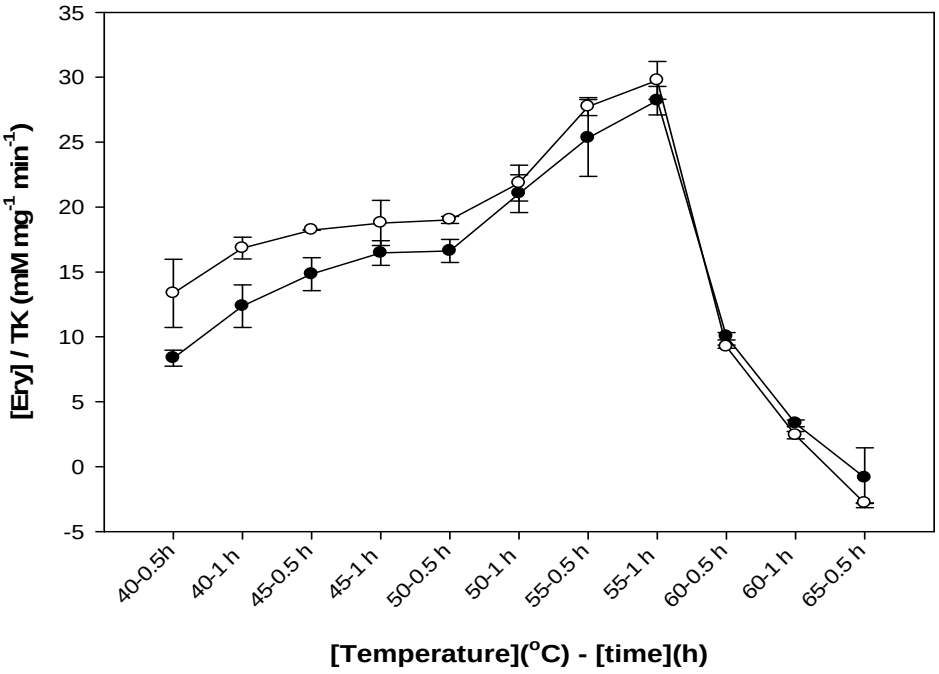


Figure 4-1 Temperature dependence of wild-type holo-transketolase.
Residual activity of both (●)pure wild-type and (○) clarified lysate were measured in triplicate over a range of temperatures from 40°C to 65°C in Tris-HCl pH 7.5. Samples were heated up to the desired temperature and time and immediately cooled down in an ice-bath. Residual activity was measured using GA and HPA as a substrates at 25°C and pH 7.5.

Under the condition tested (section 4.2.1), it is possible that when Holo-TK is exposed to temperatures of 40-55 °C it undergoes a refolding, re-annealing, de-aggregation, or other structural event which leads to a higher recovery of active enzyme (Lencki et al., 1992). At these temperatures, it is clear that holo-transketolase is resistant to irreversible unfolding or aggregation that would otherwise lead to a loss of activity. The increase in activity suggests that heating the enzyme up initially may also be unfolding or partially mis-folded structures, and/or dissociating low-order aggregates. The elevated temperature thus increases the chances of a successful "collision" between residues to form the correct hydrogen bonds, ion pairs, hydrophobic and Van der Waals interactions of the native and active enzyme (Querol et al., 1996; Vetriani et al., 1998; Lee et al., 2007). However, this refolding propensity changes at temperatures greater than 60 °C. Above this point, the kinetic energy of the enzyme and water molecules is so great that the structure of the enzyme either denatures and

misfolds irreversibly or even begins to aggregate irreversibly, thus leading to a decrease in activity upon re-cooling to 25 ^{o}C. Such inactivation of enzymes through protein denaturation or aggregation with increased temperature is more commonly expected (Daniel et al., 2008).

The time-course for inactivation of Holo-TK was assessed at 60 and 65 ^{o}C. Samples were incubated at the desired temperature for various times, and then re-cooled and assayed at 25 ^{o}C. From Figure 4-2 it can be seen that at 60 ^{o}C the activity of Holo-TK decreases with time with slight activity variations for the first 15min, but then consistently decreases for times above 15 min and to up to 1h. The incubation at 60 ^{o}C gives an inactivation constant of k_1= 0.0315±0.03 min^{-1} ($t_{½}$= 22±0.3min) as calculated by plotting the *ln* (Initial activity/Residual activity) versus time (Figure 4-2 inset). The activity of the enzyme incubated at 65 ^{o}C decreased more abruptly over 0-15 min, with complete loss of activity after 15 min, and gave an inactivation constant of k_2=0.188±0.05 min^{-1} ($t_{½}$ =3.7±0.1min). The inactivation constant calculated for 60 and 65 ^{o}C was used to estimate the energy of inactivation as E_{ai}=2.207x10^{-3} kcal mol^{-1}. These values are useful for comparing any mutants with improved thermal stability (Chapter 5).

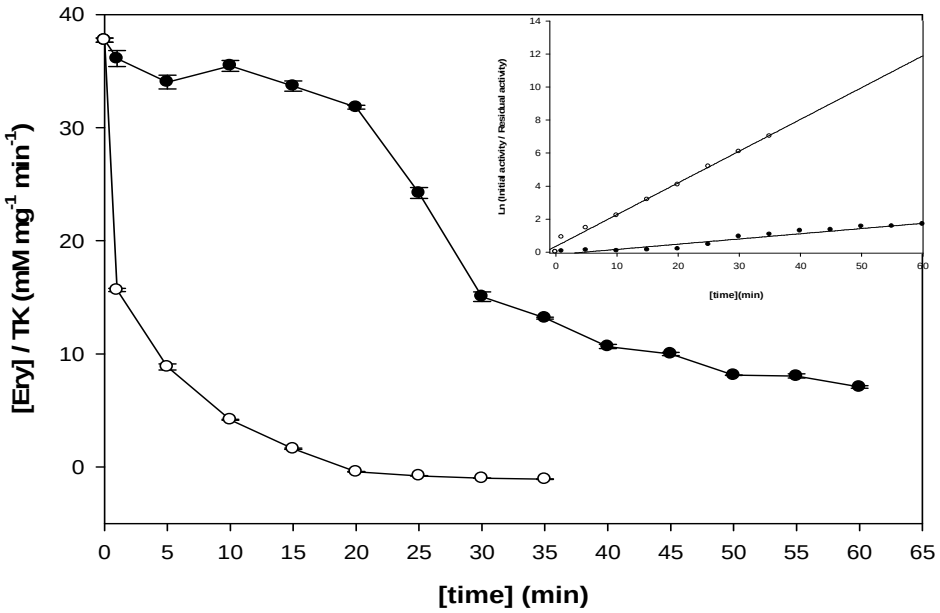


Figure 4-2 Thermal deactivation of wild-type holo-TK at 60 and 65°C.
Residual activity of pure wild-type transketolase was measured in triplicate at temperatures of (●) 60°C and (○) 65°C in Tris-HCl pH 7.5. Samples were heated up to the desired temperature and time and immediately cooled down in an ice-bath. Residual activity was measured using GA and HPA as a substrates at 25°C and pH 7.5. The inset shows the thermal inactivation of Holo-TK at 60 and 65°C. The experimental curves were fitted to a single exponential first-order reaction

In general, the contribution of hydrophobic and hydrophilic residues within the protein are of utmost importance in controlling the denaturation and aggregation of the enzyme at high temperature. Indeed, it is well accepted that protein aggregation at elevated temperatures is attributed to favourable interaction between hydrophobic regions of partly unfolded proteins molecules (Tomazic & Klibanov, 1988; De Bernardez Clark, 1998). In other words, hydrophobic residues that are normally buried in the native form of the enzyme become exposed to the solvent and interact with hydrophobic residues from other unfolding enzyme molecules to minimise their exposure to the solvent, thus leading to aggregation.

Previous analyses on the three dimensional structure of Holo-TK clearly show that the thiazole ring of TPP within holoTK is bound to both subunits of the enzyme mainly by hydrophobic interactions. Of those residues within an 8Å

sphere of the TPP, 30 are hydrophilic and 33 are hydrophobic. The distribution of hidrophilic residues were: Pro (3),Trp (1),Gly (11),Ala (2),Met (5),Phe (4),Leu (2),Val (2) and Ile (3) located around the substrate-binding pocket in a highly hydrophobic environment. Among those residues are Ile187, Ile189, Val193, Trp196 and Phe197 which belong to the loops that are involved in TPP-binding in Holo-TK (Nikkola et al., 1994), and are unstructured in the Apo-TK form. The early disruption of these loops could potentially explain the loss of activity at temperatures greater than 60 ^{o}C if their local stability is less than the overall stability of the enzyme.

4.3.2 Thermal denaturation of *E. coli* Holo transketolase monitored by circular dichroism

The CD spectra measured from 25^{o}C shows a minimun signal at 208 and 222nm. This profile is consistent with an α-helical protein which, according to the crystal structure of *E. coli* transketolase, contains 46% helical structure (Littlechild et al., 1995). The spectra for wild-type holo-TK recorded at several temperatures using a temperature ramp of 0.5^{o}C/min (Figure 4-3) indicated that the enzyme undergoes a large, 45-50%, spectral change by the time the temperature reaches 60±2^{o}C.

Analysis of the spectrum in Figure 4-3 showed that the ellipticity decreases along the entire spectrum, and there is no clear isosbestic point which is a signature for a two-state (Folded-Unfolded) model (Kelly & Price, 2000). Even though the ellipticity signal continued changing, the changes between 65-95^{o}C did not seem to be very substantial as it can be seen from Figure 4-3. Certainly, the shape of the spectrum measured at 95 ^{o}C, does not correspond to a random coil state, but rather a compact denatured state (Carra & Privalov, 1996; Killenberg-Jabs et al., 2002) containing apparent secondary structure. However , it should be noted that the CD signal of an aggregated protein can be misleading in terms of apparent secondary structure content (Figure 4-3). This

suggests that the denaturation may precede the formation of intermediate or aggregate products. Indeed, the material at 70 °C was visibly aggregated. These results are different from those previously reported using urea as a chemical denaturant (Chapter 3) where the CD-signal changed to a much higher values of ellipticity (i.e. ellipticity values >> 0), and where no aggregation was observed.

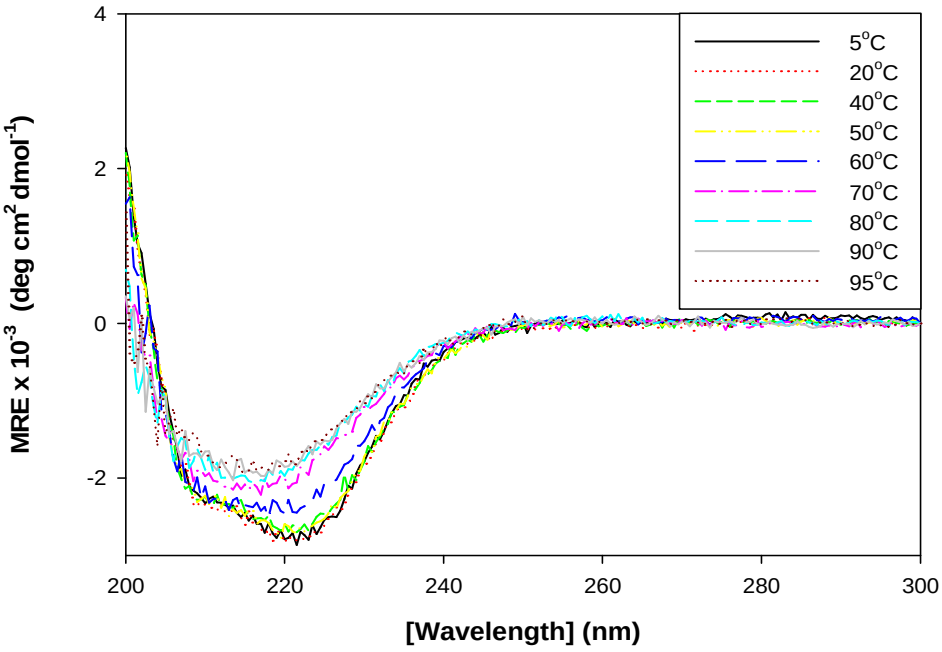


Figure 4-3 Temperature-dependant spectra of Holo-TK.
The value of ellipticity at 222 nm decreases with increasing temperature. Holo-TK, 0.1 mg mL^{-1} (1.38 µM), was prepared in 25 mM Tris–HCl, pH 7.5, 0.5 mM $MgCl_2$, 0.05 mM TPP. CD spectra was collected at a temperature range from 5 to 95°C using a 1.0mm cuvette. Samples were incubated for 2 min at the appropriate temperature before full spectra (190–300 nm) were acquired. The inset box shows the temperature and its corresponding spectra. Dynavolts showed to be between the permitted ranges during the whole scan.

The decrease in ellipticity was consistent with a loss of protein due to temperature induced protein precipitation. Inspection of the cuvette following heating at temperatures above 70 ^{o}C showed that the sample contained a large amount of precipitate. The observation that wild-type Holo-TK unfolded irreversibly suggests that the loss of activity of the enzyme (section 4.3.1 and Figure 4-2) at temperatures above 60 ^{o}C was mainly due to the formation of inactive aggregates.

The thermal denaturation/aggregation profile of wild-type holo-TK was evaluated by measuring the temperature-dependent CD response at 222nm from 5 to 95 ^{o}C (Figure 4-4). This profile fits to an apparent two-state transition from native to non-native form(s), and so it is useful to analyse the thermodynamics by two-state equations even though more than two states are likely to exist. The midpoint of unfolding (Figure 4-4), T_m, was determined by fitting the ellipticity signal and temperature to a two-state unfolding reaction mechanism (Rufus et al., 1966), and applying the necessary corrections to the data as described previously (Allen & Pielak, 1998; Greenfield, 2006). The T_m for holo-TK was estimated as 58.3 ^{o}C. This is consistent with results seen in Figure 4-3; where a more substantial change begins to occur above this temperature.

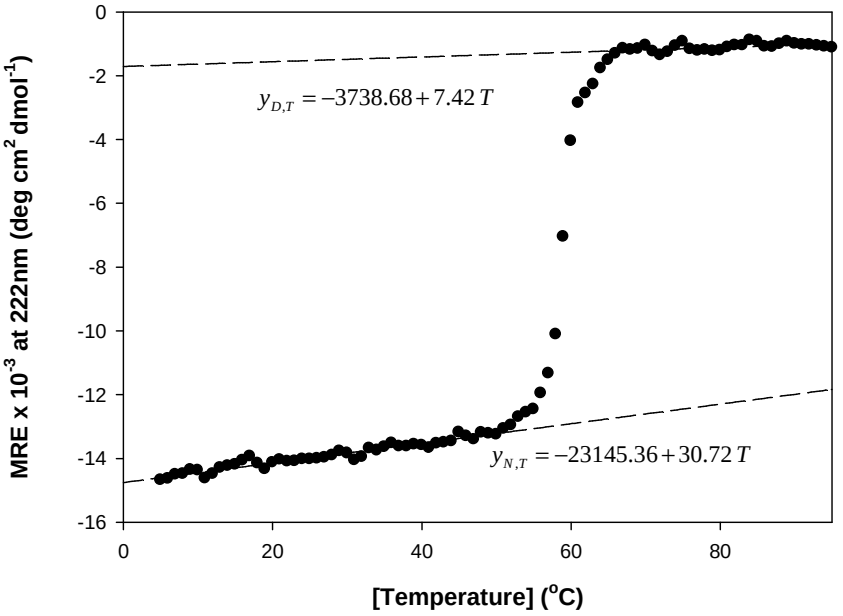


Figure 4-4 Temperature-dependent denaturation of Holo-TK observed by CD at 222nm.
Holo-TK, 0.1 mg mL^{-1} (1.38 μM), was prepared in 25 mM Tris–HCl, pH 7.5, 0.5 mM $MgCl_2$, 0.05 mM TPP. Samples were incubated for 2 min at the appropriate temperature before full spectra (190–300 nm) were acquired. Values of MRE at [θ_{222nm}] were taken for denaturation analysis. Broken lines and equations are based on least-squares analysis of the pre- and post-transition baselines.

4.3.3 Thermal denaturation of *E. coli* transketolase monitored by dynamic light scattering

A further analysis of the holo-TK thermal denaturation was performed using the more sensitive dynamic light scattering (DLS) technique. The sensitivity of DLS is sufficient to distinguish different oligomeric and quaternary protein states present at low concentrations, and it is ideally suited for monitoring the stability of the protein structure to denaturing conditions such as elevated temperature. The protein thermal denaturation mid-point (T_m) is defined as the temperature at which the protein denatures to 50% of the native structure population. The change in size that accompanies protein denaturation is easily identified by DLS (Murphy, 1997). Figure 4-5 shows the temperature dependent Z-average diameter (the mean diameter based on the intensity of scattered light) and scattering intensity (in kilo counts per second) of holo-TK in Tris-buffer at

pH 7.5. Indeed, at temperatures below T_m=46 °C, the size and scattering intensity are constant, suggesting a stable tertiary and quaternary structure. Above 46 °C and up to 49 °C both the size and scattering intensity increase exponentially with temperature, indicating the presence of denatured aggregates. Interestingly, the Z-average does not change at all at temperatures above 49 °C. However, the intensity shows a small shoulder at 49-51 °C suggesting the presence of a second stage in the denaturation process.

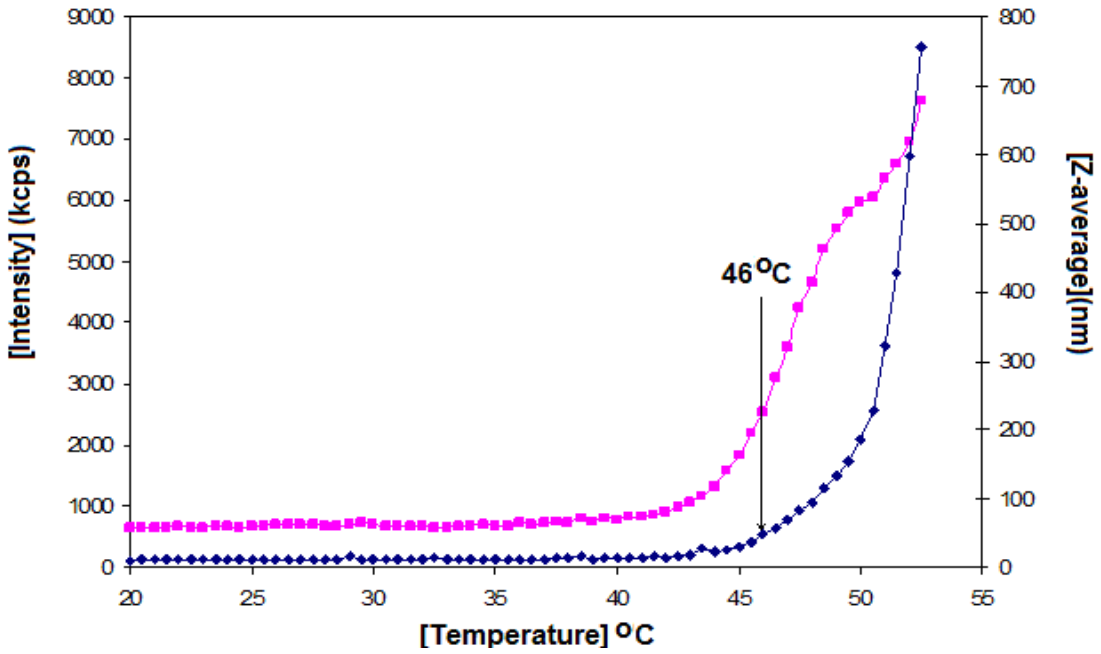


Figure 4-5 Melting point trace and hysteresis for holo-TK.
Sample of 1 mg mL−1 (13.8 µM) protein were prepared in 25 mM Tris–HCl, pH 7.5 with cofactors (0.5 mM TPP and 5 mM $MgCl_2$) and incubated for 1 h before data was acquired. The mean diameter based on the intensity of scattered light (Z-average) of holo-TK (■) intensity (in kilo counts per second) of holo-TK (♦).

It seems that 46 °C marks the beginning of the denaturaration-aggregation process. This change was not observed by CD measurement. However, DLS was unable to detect further changes above 55 °C where the signal became extremely noisy (data not shown). Conversely, CD results at 222 nm Figure 4-4 section 4.3.2 showed that the denaturation begins at around 50 °C. This temperature agrees with that at which a second change is visible using dynamic light

scattering. These final results fully support the hypothesis that there are more than a two states in the denaturation process of TK.

Since denaturation depends strongly on conditions such as pH, ionic strength, contamination, and the history of the sample, the role of these factors must be studied in more detail. Presently, the DLS parameters can be used to characterise any possible interesting TK variants.

4.3.4 Structural analysis of *E. coli* transketolase at different pH's monitored by circular dichroism

The secondary structure content of holo-TK was measured using CD spectroscopy at different pHs ranging from 4 to 11 (Figure 4-6). The structure of holo-TK displays a peak of negative ellipticity with a minimum peak at 222 nm at pH 7.0, a characteristic property usually found in α-helical protein structure (Manavalan & Johnson, 1983; Littlechild et al., 1995). The structure of holo-TK exhibited mostly α-helical structure and seems to be very stable in a pH range of 5 to 9.

Increasing or decreasing the pH by the order of 2 pH units on either side of neutral pH (~7.0) did not have a considerable impact on the secondary structure of the protein. However, a significant decrease of α-helical structure occurred at both pH 4.0 and pH 10.0. The structure changed from mostly α-helical to eventually "random coil". It is worth noticing that at pH 11.0 the secondary structure of the protein seemed to recover itself almost completely to its apparent original α-helical structure when compared to that at neutral pH. This suggests that at pH 11 the presence of large water-soluble aggregates is present rather than reformation of a native-like structure (see discussion below).

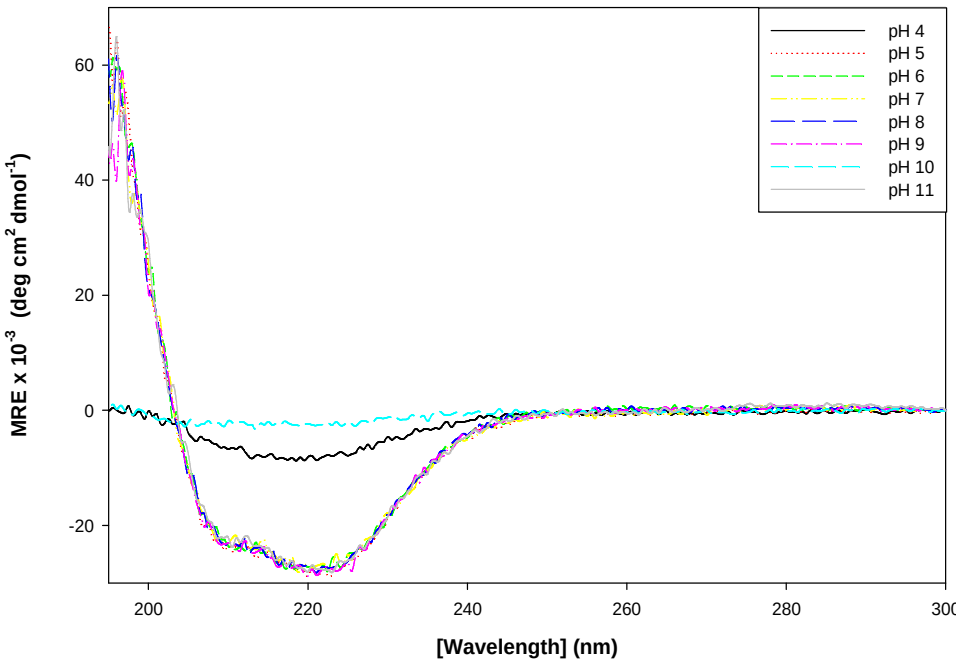


Figure 4-6 Comparison of circular dichroism spectra of Holo-TK at different pH's.
The inset box shows the different pHs and their correspondent spectra. Samples were incubated with 2.5 mM $MgCl_2$, 0.25 mM TPP at different pHs for 45 min at 25 °C before full spectra (190-300 nm) were acquired. Changes in the secondary structure were monitored using a circular dichroism instrument (AVIV Biomedical, USA).

At first glance, the variation of pH between 5 and 9 did not seem to have a significant impact on the structure of the enzyme. However, the residual activity of the enzyme at different pHs, suggests otherwise (section 4.3.5).

Changes in the secondary structure of holo-TK are shown as a function of ellipticity at 222 nm for pH 4-11 (Figure 4-7). The molar ellipticity obtained at 222 nm was used to determine the α-helicity of the protein using the Eq. 4-1 below for a quick estimation of the α-helical content of the globular protein (Chen et al., 1972).

$$[\theta]_{222nm} = -\theta_{\max} f_H - \theta_{\min} \qquad \text{Eq. (4-1)}$$

where $[\theta]_{222nm}$ is the molar ellipticity of the protein, f_H the fraction of α-helix, $\theta_{\min}$ (2340) is the minimum value of the molar ellipticity at 222 nm calculated for

the "unordered" fraction of five proteins, and θ_{max} (-30 300) is the maximum value for the ellipticity at 222 nm as measured for the helical fraction of five proteins (Chen et al., 1972). The estimated α-helicity of the enzyme at different pHs is described in Table 4-1.

Table 4-1 Estimated content of α-helical structure of Holo-TK after incubation for 45 min at different pHs.

pH	MRE X 10^{-3}, at 222nm*	f_{Helix}, %
4	-8.149112579	19
5	-26.92608934	81
6	-26.74624985	80
7	-27.89462253	84
8	-26.88925475	81
9	-27.07992795	82
10	-2.643423916	1
11	-27.38543842	83

*Changes in the secondary structure were monitored by circular dichroism (AVIV Biomedical, USA) at 222 nm. Samples were incubated with 2.5 mM $MgCl_2$, 0.25 mM TPP at different pHs for 45 min at 25 °C prior to measurement.

The fraction of α-helical structure content does not seem to deviate by more than 2-3% across the range pH 5-9. However, this does not prove that it is exactly the "same" secondary structure (see discussion). There is a decrease in α-helical structure content to 19% at pH 4 and 1% at pH 10, compared to 84% at pH 7 (Figure 4-7).

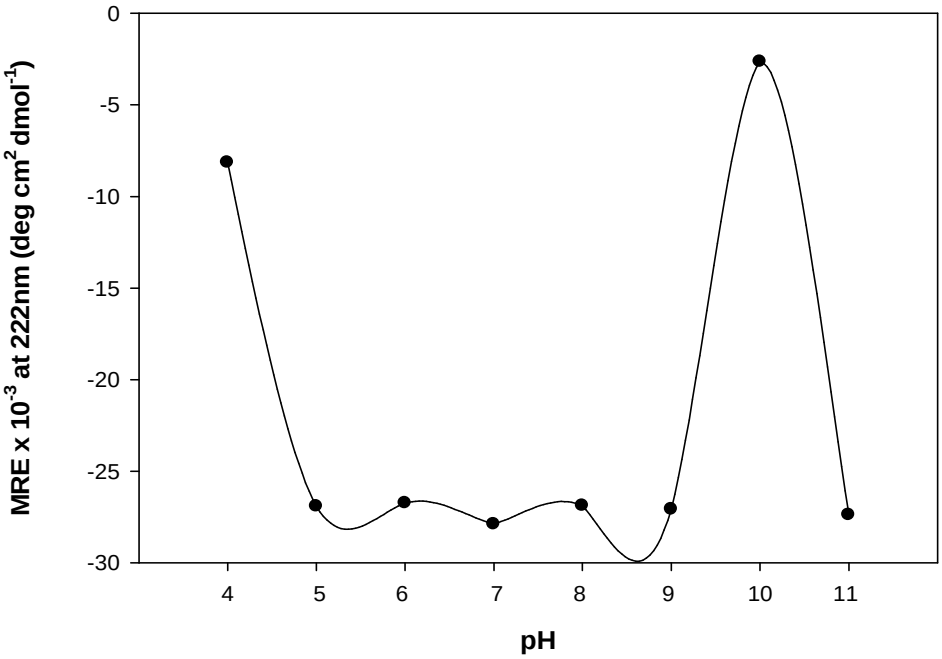


Figure 4-7 Effect of pH on *E.coli* holo-transketolase secondary structure. Pure transketolase (0.5 mg mL^{-1}) with cofactors (TPP and Mg^{+2}) were prepared at different pHs ranging from pH 4 to pH 11. Mean residue ellipticity was measured by circular dichroism (AVIV Biomedical, USA) at 222nm.

Analysis of Figure 4-6 and Table 4-1 showed that the content of secondary structure at pH 4 (~19%) seems to be more structured than that at pH 10 (~1%). A CD spectrum of holo-TK at pH 4 (Figure 4-8 A) was obtained as a function of time to gain an insight into the acid-denaturation mechanism of the enzyme at pH 4. Holo-TK secondary structure is shown as a function of ellipticity at 222 nm and 190 nm (Figure 4-8 B) at pH 4. Holo-TK was incubated at pH 4 for 45 min before the first circular dichroism spectrum was acquired. Data from Table 4-1 and Figure 4-7 at pH 7 provide a benchmark comparison point. As it can be seen from the figure that after 45 minutes of incubation the protein contained ~40% secondary structure (compared to 19% previously in Table 4-1).

This structural content continued to decrease such that after 65 min of incubation the secondary structure content was only ~2.6%. However, from 65 to 75 minutes of incubation the protein seemed to partially recover structure to 37±2% of the initial secondary structure content (Table 4-1), though as discussed previously the observed change in CD signal is more likely to have resulted from aggregation of the protein. No further difference in secondary structure was found for CD spectra taken after 85min, 105min and 24h of incubation. These results imply that the initial loss of structure at pH 4 is slightly slower than at pH 10, but that both reach almost complete "denaturation" at some point. However, a second transition occurs after a lag phase at pH 4 such that if the data for pH 4 in Figure 4-6 (also Figure 4-7 and Table 4-1) had been taken 30min later, then the apparent secondary structure content would have been higher than 19% (i.e. 37%). A similar effect may have occurred at pH 11 where the α-helical structure content appears to have recovered from the complete denaturation at pH 10, to the same level as that at pH 7 (Figure 4-7). Again, this does not prove that the secondary structures are identical at pH 7 and pH 11, only that there are equivalent amounts of secondary structure (see discussion below). Taken together it appears that both the acid and base denaturations of holo-TK are complex time-dependent processes that warrant further work beyond the scope of this thesis.

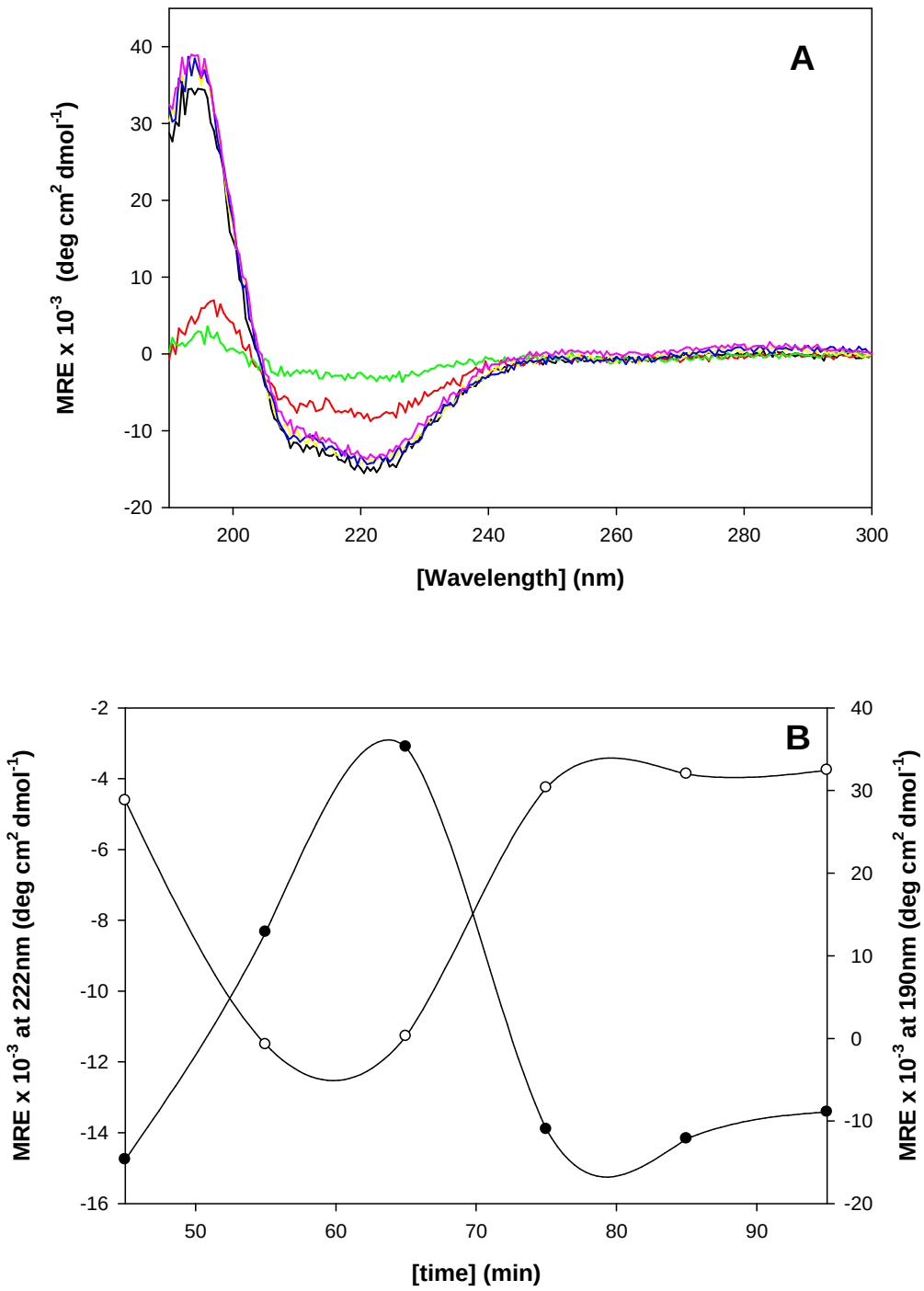


Figure 4-8 Effect of incubation time at pH 4 on *E. coli* transketolase secondary structure content.
Samples were monitored at (A) 190-300nm and (B) at 222 and 190nm. Solid lines in (A) correspond to time: (-)45;(-)55;(-)65;(-)75;(-)85;(-)105&24h respectively, in (B) at (●) 222 and (○) 190nm respectively. 0.5 mgmL^{-1} of *E. coli* transketolase was prepared with cofactors (TPP and Mg^{+2}) and incubated at pH 4 for 45 min before the first circular dichroism spectrum was acquired. Changes in secondary structure were monitored every 10 min for 60 min and then at 24 hours.

The effect of pH on protein structure is not general, varying from one protein to another. It seems that the acid or base denaturations of holo-TK secondary structure give rise to very different unfolded states from that obtained with the high concentration of urea used in Chapter 3 (Martinez-Torres et al., 2007). These different results between acid/basic and urea denaturations were perhaps expected as it has been previously documented that acid/basic denaturated states of proteins are often less unfolded than those obtained with high concentration of urea (Goto & Fink, 1989; Debnath et al., 2005; Pace et al., 2004).

Depending on the environmental conditions, the enzyme (holo-TK) may adopt various conformational states (Dubey & Jagannadham, 2003). At pH (6-9) and room temperature (25^{o}C), the enzyme maintains homo-dimeric native state with a probably rigid tertiary structure and predominantly α-helical secondary structure but with loss of activity (section 4.3.5) especially at pH 6 and pH 9.

It is suggested but not assumed that at acidic pH (below pH 5) and room temperature, the protein is converted into a partially unfolded state with disrupted tertiary structure and enhanced, non-native α-helical secondary structure (Figure 4-8). It has been reported for *Staphylococcal* nuclease (Fink et al., 1994), apomyoglobin, β-lactamase and cytochrome *c* (Goto et al., 1990) that when increasing the concentration of H^+ the enzyme can go through three different states: i) a native-folded structured state (N); ii) an acid-unfolded state (U_A) and iii) a molten-globule state (A), which has a compact state with substantial secondary structure but little or no native-like tertiary structure (Goto & Fink, 1989; Fink et al., 1994; Dubey & Jagannadham, 2003). The molten-globule state is produced essentially by the presence of H^+ which makes the enzyme become more positively charged, thus causing intra-molecular repulsion between the positively charged groups leading to unfolding (U_A) (Goto et al., 1990). This can be happening in TK according to Figure 4-8 after the first 65 min of incubation at pH 4, where the enzyme has not yet fully attained

equilibrium with H^+. Conversely, after 75 min the A state may be present, resulting in the protein folding up to adopt a compact structure with decreased hydrophobic surface area as electrostatic interactions with positively charged centres of the enzymes are occurring (Goto et al., 1990; Stigter et al., 1991; Fink et al., 1994). Alternatively, an initially accumulated A-state may be slowly forming a low order soluble aggregate that has more secondary structure content than the monomeric A-state.

To fully support these findings a more thorough analysis has to be performed, such as the use of size exclusion chromatography, dynamic light scattering for the determination of hydrodynamic radius, and intrinsic fluorescence for elucidation of the tertiary structure. Regrettably, those experiments are beyond scope of this work.

In the same way, at alkaline pH and room temperature (Figure 4-6 and Figure 4-7), it can be seen that holo-TK may undergo two base-induced transitions. The first transition (between pH 9 and 10) results in an apparent partial disruption of secondary structure. The second transition (pH >10) is maybe accompanied by a further disruption of tertiary structure and formation of extensive, though non-native α-helical secondary structure, analogous to the A-state often observed at low pH (Goto & Fink, 1989; Stigter et al., 1991). Similar molten-globule states have been observed in pepsinogen at pH 11 (McPhie, 1982). Thus, it is likely that at pH 11, as described previously for holo-TK at pH 4, the presence of large water-soluble aggregates or an A-state is present, resulting in the retention (or re-formation) of similar alpha-helical content to that at pH 7.

4.3.5 Effect of pH on the deactivation of *E. coli* holo transketolase

A concentration of 0.5 mg mL^{-1} of pure TK and that in clarified lysate was incubated in the presence of cofactors (2.5 mM $MgCl_2$, 0.25 mM TPP) and used to estimate the pH deactivation profile of wild-type tranketolase. To study the acid (H^+) and basic effect (OH^-) of pH on the enzyme activity a pH range from 4 to 11 were studied. Tris-HCl buffer was utilised for pH 6 to 9; Bis-Tris for pH 4 to 5, CAPS for pH 10 to 11, at a concentration of 100 mM (section 4.2.5). Even though clarified lysate seems to have slightly more activity at some pH (i.e. pH 6.0 and 7.0), the difference is not that significant; so the same analysis is applicable for pure holo-transketolase.

Figure 4-9 shows the results obtained after samples were incubated for 45 min. Comparable results were obtained after incubating samples for 90 and 120 min (data not shown). Purified and clarified lysate enzyme showed the same characteristic "bell-shape" curve as at least two different amino acids groups of the enzyme are being titrated (see discussion below) at different states of ionization for the different pH values (Stigter et al., 1991; Nelson & Cox, 2004). Pure and clarified lysate showed the highest activity at pH 7.0. At pH above or below 7.0, the enzyme activity decreased, with a more rapid decrease at pH below 7.0, than at above 7.0.

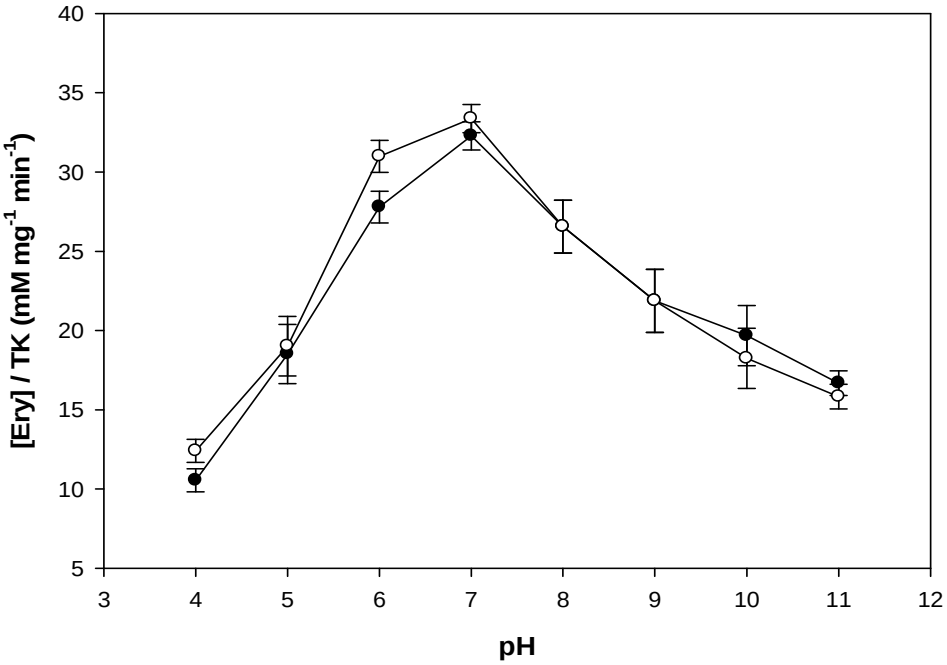


Figure 4-9 Effect of pH on Holo-transketolase activity.
Residual activity of both (●)pure wild-type and (○) clarified lysate were measured in triplicate over a range of pH from 4.0 to 11.0 in 100 mM Bis-Tris for pH 4-5; 100 mM Tris-HCl for pH 6-9, and 100 mM CAPS for pH 10-11. Residual activity was measured using GA and HPA as a substrates, in their respective buffers, at 25^{o}C.

The results also showed that at pH≤ 5.0 and pH ≥ 10 the enzyme activity is reduced to ~50%. An enzyme activity ≥ 80% was maintained between pH 6.0-8.0.

Previous work has shown that exposure to pH below 6.5 results in a decrease in activity of TK, that was assumed to be due to denaturation (Mitra et al., 1998). However, it is more likely to be due to the titration of ionisable catalytic residues in cases between pH 5 and 7 as the secondary structure content was unchanged in Figure 4-8. Below pH 5, denaturation of secondary structure was observed (Figure 4-8), and so may provide a secondary cause of enzyme inactivation. The pKa values of ionisable groups in the protein can potentially be modified significantly by the environment, to the extent that residues which are basic in solution may act as proton donors, and vice versa (Stigter et al., 1991; Harris & Turner, 2002). Therefore, it is likely that the activity

of TK can be affected by changing, among other factors, the pKa of catalytic residues rather than being influenced by dramatic changes in the secondary structure of the enzyme (section 4.3.4).

Residues 12 Å away from the active site (Figure 4-10), previously identified as essential in binding the cofactors -TPP and $MgCl_2$ - and that play an important role in the catalytic process of the enzyme (Kochetov, 2001; Costelloe et al., 2008) were analysed. Among them are residues that belong to the two cofactor-binding loops (Asn185-Asp199, and Leu382-Gly392) in *E. coli* (Nikkola et al., 1994).

Even though there are 20 natural amino acids only those that are normally used in protein pKa calculation (i.e. Asp, Glu, Arg, Lys, His, Cys, and Tyr) were taken into account (Davies et al., 2006; Nielsen, 2007; Delphine et al., 2008), as they contain ionisable groups and are usually involved in substrate or product binding and/or catalysis, and, are therefore most likely to determine the pH activity profile of an enzyme (Fersht, 1985).

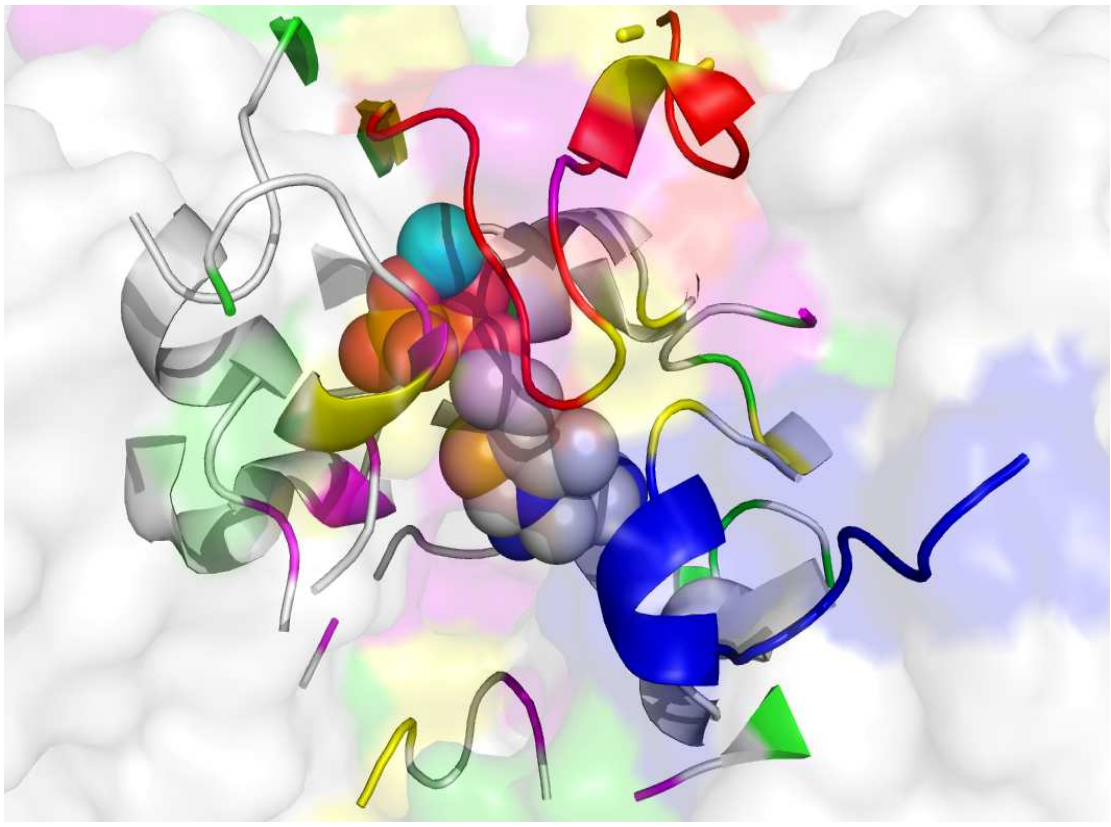


Figure 4-10. Relevant amino acids, 12Å away from active site, that are more likely to determine the pH activity profile of *E.coli* transketolase.
PDB file was obtained from 1qgd.pdb (Littlechild et al., 1995). For one active site only, the TPP cofactor is highlighted in CPK colours and as spheres, and the metal ion cofactor is highlighted as a cyan sphere. Amino acids such as Cys, Tyr, Arg and Lys that are affected in an alkaline environment are coloured in green; those such as Asp and Glu which are more likely to be affected by an acid environment are coloured in yellow. Amino acid His which can be altered by either acid or alkaline environments are coloured in magenta. Among those amino acids are residues that belong to the cofactor-binding loops Asn185-Asp199 (Red) and Leu382-Gly392 (blue) (see Table 4-2). In the back ground amino acids that are > 12Å away from active site but are also affected depending on the pH environments. Figure generated with PyMol (DeLano, W.L. (2002), The PyMOL Molecular Graphics System on World Wide Web http://www.pymol.org).

Table 4-2 Relevant amino acids that are 12 Å way from the *E. coli* transketolase active site.

AA in Holo-TK 12 Å away from active site	Residue located in chain	Position	Estimated pKa using Propka software		pKa values normally used in a protein pKa calculation*
			Holo -TK	Apo-TK	
Asp183	A	Buried	4.87	6.26	
Asp184	A	Buried	5.82	4.93	
Asp190**	A	Buried	2.46	2.19	3.8
Asp259	A	Surface	3.64	3.64	
Asp381	B	Buried	6.67	6.26	Range 3-6†
Asp469	B	Buried	4.19	3.9	
Glu160	A	Buried	16.48	16.14	
Glu194**	A	Surface	6.02	5.25	
Glu165	B	Buried	6.01	5.3	4.3
Glu411	B	Buried	5.14	6.65	Range 3-6†
Glu468	B	Buried	3.74	3.73	
Arg91	A	Buried	10.46	10.71	
Arg204	A	Buried	12.32	12.43	
Arg442	A	Buried	10.63	10.64	
Arg446	A	Buried	12.22	12.31	12
Arg358	B	Buried	9.66	9.79	
Arg410	B	Buried	8.55	8.81	Range 11-13†
Arg442	B	Buried	10.62	10.63	
Arg520	B	Buried	11.8	11.92	
Lys244	A	Surface	10.5	10.5	10.5
Lys254	A	Surface	10.22	10.22	Range 7-9&9-11†

*pKa values taken from (Nelson & Cox, 2004).

** Residues in the two cofactor-binding loops (Asn185-Asp199 and Leu382-Gly392) (Nikkola et al., 1994)

† pKa values range normally used in protein calculations

Table 4-2 continued

AA in Holo-TK 12 Å away from active site	Residue located in chain	Position	Estimated pKa using Propka software		pKa values normally used in a protein pKa calculation*
			Holo	Apo-TK	
His26	A	Buried	-3.41	-3.14	
His66	A	Buried	0.77	0.14	
His100	A	Buried	-1.72	-1.45	
His192**	A	Surface	6.43	6.43	
His219	A	Buried	4.83	4.91	6.08
His258	A	Surface	5.21	6.43	
His261	A	Buried	2.39	2.35	Range 5-8†
His406	B	Buried	6.62	6.8	
His461	B	Buried	-0.54	-0.36	
His473	B	Buried	0.61	1.22	
Cys157	A	Buried	6.68	6.41	8.28 (-SH)
Cys243	A	Buried	11.25	11.16	Range8-11†
Tyr182	A	Buried	17.37	17.71	
Tyr440	A	Buried	14.55	14.3	
Tyr404	B	Buried	13.52	13.37	10.01
Tyr407	B	Buried	13.85	13.83	
Tyr440	B	Buried	14.51	14.26	Range 9-12†
Tyr459	B	Buried	14.1	14	

*pKa values taken from (Nelson & Cox, 2004).

** Residues in the two cofactor-binding loops (Asn185-Asp199 and Leu382-Gly392) (Nikkola et al., 1994)

† pKa values range normally used in protein calculations

Table 4-2 shows, in general, that there are more buried ionisable amino acids than at the surface in the active site. This is not unusual as proteins often have buried residues with highly perturbed pKas which can contribute to the denaturation or loss of activity in an enzyme (Elcock, 1999). In this work those residues in the active site and so more likely to alter the activity of the enzyme (Perez-Jimenez et al., 2004) are considered. The surface-charged residues which are more critical for protein-protein association (Sheinerman et al., 2000) are mentioned briefly.

The pKa values estimated using PROPKA web-based service (section 4.2.6), and also those which are normally used in protein pKa calculations, are depicted in Table 4-2. Those values in combination helped us to acquire more knowledge of the pKa of ionisable residues in an effort to predict the protonation state of the side chain residue at a given pH, which is often essential in the prediction and understanding of protein properties (Warshel et al., 2006).

pKa values were estimated based on the interaction with neighbouring residue side chains or ligands such as TPP or Mg2+ (Delphine et al., 2008). However, this program is under constant revision it has been found that in some cases the pKa values differ by ≤1 (Li et al., 2005), so some of the values are not reliable and were either discarded or treated with discretion as a reference only. However, most of the pKa's estimated using PROPKA do not differ from those usually used in the pKa calculations (Table 4-2). Figure 4-9 shows that when holo-TK is in a slightly alkaline environment (pH > 8.5) it loses over 40% activity. This is likely due to the environmental effect on the residue side chain groups from Lys (Amino group); Tyr (Phenolic); Cys (Thiol) or Arg (Guanidil) that may control the activity of the enzyme at pKa around that pH environment.

When the pH is more alkaline, then the enzyme is less active due to the deprotonation of the side chain groups that may produce the breaking of ionic or hydrogen bonds between neighbouring residues. This is probably caused by the deprotonation of the amino, guanidine and the phenol groups of lysine, arginine and tyrosine residues respectively (Bushueva & Tonevitsky, 1987; Nath & Rao, 2001). Furthermore, In the case of Cys, it can be caused by a partial destruction of cysteine residues due to base catalysed β-elimination reactions or may be protonation of the thiol (Nelson & Cox, 2004; Witt et al., 2008). Deprotonation of these residues could result in changes that are unfavorable to the native state, including electrostatic repulsions, destruction of salt bridges, and the formation of buried isolated charges, eventually leading to functional loss of activity. In contrast, pH values slightly less alkaline (i.e. close to pH neutral) the enzyme seems to be more active as it has a more stabilising effect creating important ionic or hydrogen bonds.

Consideration is given to Histidine residues, with special attention to His192 that may play an important role in the loss of stability of the enzyme at pH > 8.5 as it is a crucial amino acid contained in one of the cofactor-binding loops (residues 185-199 in *E. coli* TK) (Nikkola et al., 1994). If the protonation state of His192 is modified then this cofactor-binding loop may become partially or completely unstructured, leading to weaker cofactor-binding and loss of activity. Alternatively, His192 might be involved in crucial acid/base reactions that are critically related to its pKa. As the pKa is close to neutral pH, His192 can potentially both accept and donate protons. If this pKa or surrounding pH is perturbed then it may act as only a donor or acceptor but not both. Although His192 does not contribute directly to the binding of substrate or either of the cofactors TPP and Mg^{2+}, it could have a structural effect on catalysis. It forms potential hydrogen bond interactions with Gly186, Gly256, and Glu194 which are important residues involved in stabilising one of the cofactor-binding loops. There may also be

a distant influence on the pKa of His258 which is 6.4 Å away and is directly involved in substrate binding. Perturbations in pKa can occur for example during exposure of the charged side-chain to a more alkaline environment (Pace et al., 1996). A similar effect can occur in an acidic environment. A negative effect on enzyme activity might be expected for other ionisable residues that are close to the enzyme active site, through local structural or catalytic changes. However, more long-range structural effects cannot be ruled out.

The Glu and Asp residues surrounding the active site are shown in Table 4-2. The total numbers of Glu and Asp residues in the two *E. coli* chains A and B are: 95 and 35 respectively. Those that are within 12Å of the active site represent 6.7% of the total number that are distributed through the whole enzyme. In acid milieu (pH<5.5) holo-TK showed the most rapid decline in activity. The residue side chains that may be affected are most likely to be the carboxyl groups of Aspartate (Asp) and Glutamate (Glu). Perturbing the pH at these residues may result in either an increase or decrease in enzyme activity. For example, at pHs below their pKa (~4-5), the enzyme may be irreversibly deactivated due to the hydrolysis of the labile peptide bonds often found next to aspartic or glutamic acid residues; while at pH values above it, the enzyme might be most active with the carboxyl side chain ionised with a net negative charge (Antosiewicz et al., 1994; Forsyth et al., 2002).

Of particular relevance are residues Asp190 and Glu194, which are contained in one of the cofactor-binding loops (residues 185-199) which is structured in the active state. Loss of structure and correct cofactor interactions would lead to a loss of activity (Mitra et al., 1998; Martinez-Torres et al., 2007). As the acidity of the solvent increases, the deionised state is more likely to be present which can affect hydrogen-bonding of for example, Asp190, to its neighbours His406 and Ala383, resulting in a

destabilised cofactor-binding loop. In the case of Glu194 it is located on the surface, but also potentially hydrogen-bonds to His192 as discussed above. Thus perturbation of either His192 or Glu194 could lead to destabilisation of the cofactor-binding loop, and subsequently to the loss of enzyme activity.

For the other ionisable residues, there are a total of 44 Tyr, 9 Cys, 71 Lys and 55 Arg residues in the *E. coli* transketolase homodimer. Those that are within 12Å of the active site, are also included in Table 4-2 and represent 10% of the total number distributed through the whole enzyme.

Finally, in either environment, acid or basic, increasing hydrogen ion concentration may, additionally, increase the successful competition of hydrogen ions for any metal cation binding sites of the enzyme, reducing the bound metal ion concentration. Decreasing hydrogen concentration leads, to increasing hydroxyl ion concentration which competes against the enzymes' ligands for divalent and trivalent cations causing their conversion to hydroxides and, at high hydroxyl concentration, their complete removal of the enzyme (Stigter et al., 1991; Kajander et al., 2000; Pace et al., 2004).

4.4 Conclusions

The optimum temperature of holo-TK, in the conditions tested, was observed upon re-cooling after incubating at 55 ^{o}C for 1 hour, as it shows around three times more activity than that after an incubation at 40^{o}C for 1 hour, suggesting that the enzyme has a high propensity to refold from up to 55 ^{o}C, and that some misfolded or aggregated enzyme is re-annealed to the active native state at these temperatures. Conversely, the enzyme partially loses activity after it is incubated at 60 ^{o}C, and completely loses activity after incubation at 65 ^{o}C for only 15 minutes. This loss of activity is most likely due to irreversible denaturation and/or aggregation.

Analysis of the secondary structure of holo-TK revealed a T_m=58.3^{o}C, in agreement with the temperature at which it started losing activity. At T_m the enzyme is expected to be half folded (active) and half denatured enzyme. In the denatured state, it appears to form aggregates without activity, yet containing a considerable level of secondary structure even at temperatures above 70 ^{o}C.

Additionally, the optimum pH of the enzyme activity was confirmed to be at pH 7. It was suggested that the decrease in TK activity at basic pHs is due to deprotonation of the residues His, Tyr, Lys, Cys and Arg; whereas the acid inactivation is most likely due to the protonation of His, Glu and Asp residues.

Structural analysis at different pHs suggested that holo-TK displays a high level of helical structure content at pH < 5 and pH> 10, perhaps due to the formation of an A state or molten-globule which is mostly inactive. On the other hand, almost no structural changes were seen at pH between 5 and 9. This supports the hypothesis that the effect of pH upon activity of the enzyme in this range is mostly due to titration of catalytic residues or to small local structural effects such as destabilisation of the cofactor binding loops. In either case when

extreme temperatures or pH are used the enzyme tends to aggregate as a protective mechanism against complete denaturation.

All information acquired in this chapter will be extremely useful for assessing any mutant with increasing either pH or temperature stability.

Chapter 5

5 Combining bioinformatics and statistics for the rational design of *E. coli* transketolase stability

5.1 Introduction

The stability of enzymes and proteins in vitro remains a critical issue in biotechnology; both storage and operational stabilities affect the usefulness of enzyme-based products. Storage stability, or shelf life, refers to an enzyme's maintaining its catalytic activity in the period between manufacture and eventual use. Operational stability describes the persistence of enzyme activity in process conditions such as extreme temperature or pH, oxidative stress, organic solvents, metal ions and surfactants (O'Fagain, 2003). Although it is sometimes beneficial to adapt industrial processes to mild and environmentally benign conditions favoured by the enzyme, the use of more extreme conditions is also often desirable (Chaplin & Bucke, 1990). For example, the use of high process temperatures may be beneficial with respect to factors such as substrate and product solubility, viscosity, process speed and microbial contamination. Regardless of process conditions, the stability of the biocatalyst is typically an important economic factor. The engineering of transketolase stability for its use in industrial process conditions, by for example, improving the resistance of the enzyme structure to temperature and pH, is an important goal.

In the previous chapters I have highlighted the importance of *E. coli* transketolase as an important biocatalyst in stereo-specific carbon-carbon bond synthesis (Datta & Racker, 1961). TK finds an increasing number of applications for industrial purposes, in particular in the synthesis of chemicals important for the food industry (Hecquet et al., 1996); pharmaceuticals (Hecquet et al., 1994) and agrochemical industries (Myles et al., 1991). However, its use is still limited to very mild bioprocess conditions (Mitra et al., 1998). Furthermore a substantial analysis of the structural stability of the enzyme has been performed

(see Chapter 3) in an effort to understand the mechanism of deactivation under bioprocess conditions such as high temperature or extreme pH. Such insights begin to make it possible to engineer the enzyme and obtain mutants with increased operational stability in the harsh conditions present in bioprocesses (Dalby et al., 2007).

There are three major and principally different routes to obtain enzyme variants with improved stability: i) isolating enzyme variants from organisms living in appropriate extreme environments (Vieille & Zeikus, 2001; Schiraldi & De Rosa, 2002); ii) rational-based mutagenesis (enzyme engineering) (Matthews, 1995; van den Burg & Eijsink, 2002), and iii) directed evolution (Arnold, 1998; Dalby, 2003). While directed evolution is based on the generation of diversity followed by selection/screening, rational protein engineering utilises information on enzyme structure and the molecular basis of stability to predict stabilising mutations.

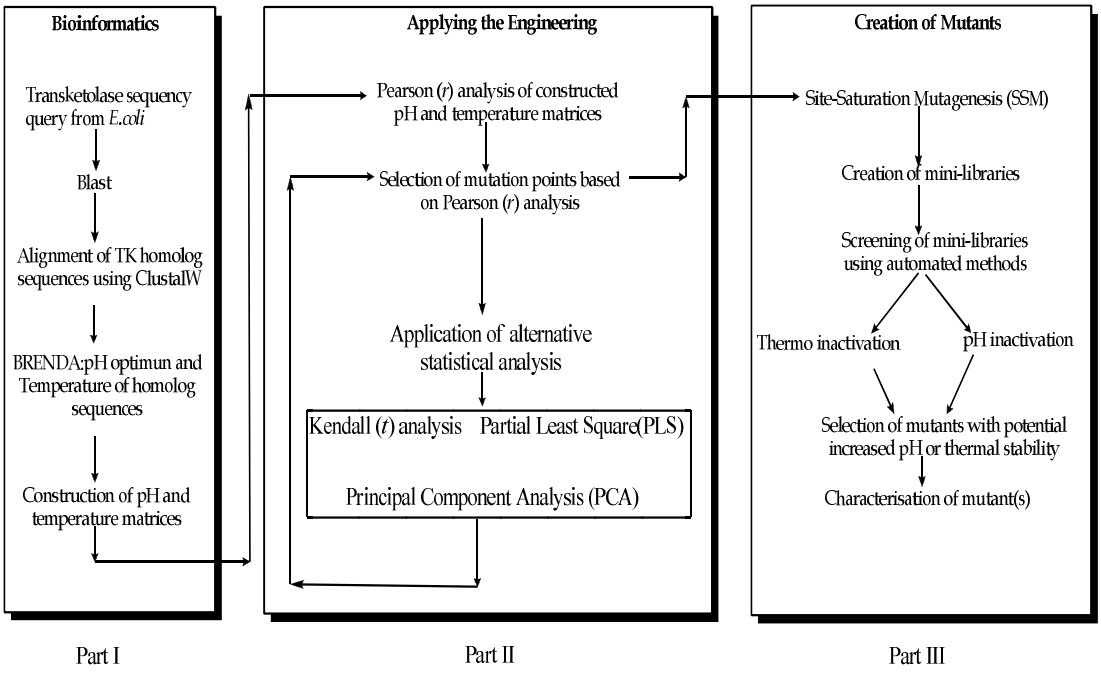
Several concepts for engineering enzyme stability have been established and are found to be applicable to a wide range of proteins. Stabilization by the insertion of prolines (Suzuki, 1999), by cavity filling (Karpusas et al., 1989), and by helix capping (Aurora & Rose, 1998) are all well know concepts. However, the combinatorial large number of possible sequences and the incomplete understanding of the factors that control protein structure are the primary obstacles in protein design (Dahiyat & Mayo, 1996).

Today, with the every day growing field of bioinformatics, there are a considerable amount of physicochemical, sequence and structural information available not only for *E. coli* transketolase but for many other enzymes (Pharkya et al., 2003). Using such information it should be possible to establish a new way to guide the rational mutagenesis of transketolase to enhance its stability. In this work a quantitative structure-temperature/pH relationship analysis was performed using, first a simple parametric (Pearson's coefficient) and then a non-parametric (Kendall's coefficient) statistical analyses to predict single

changes in the *E. coli* transketolase sequence. First, all possible transketolase homolog sequences were extracted from different databases. Next, the optimum temperature or pH of the enzyme was obtained for each sequence. For those transketolase sequences whose physico-chemical properties (i.e. pH or temperature) were not possible to acquire, then the following procedure was followed: i) Identify both the sequence whose either pH or temperature optimum was not found and its corresponding microorganism that produces it; ii) Retrieve all the possible enzymes produced for that microorganism, identified in step i, with their corresponding optimum pH and temperature; iii) For a single microorganism, group the information in two columns named temperature and pH; iv) Apply a frequency distribution analysis to single pH or temperature column to estimate potential optimum temperature or pH of the homolog transketolase sequence; v) Assign estimated optimum temperature or pH to its corresponding sequence.

To assess the utility of the statistical methods, their predicted single changes to the transketolase protein structure were introduced using site-saturation mutagenesis. Preparation and screening of reaction plates that contain transketolase variants was performed by automated high-throughput methods using robots. Interesting mutants were purified and their activity re-tested. Finally, those with "apparent" improved activity at extreme pH or temperature were sent to be sequenced. TK sequence variants were compared to the wild type to confirm the effect of mutations. Finally, two more statistical analyses were carried out, namely multivariate principal component analyses and a partial least-square analysis, to assess possible future routes for improved site prediction.

A general overview of this chapter is depicted in Scheme 5-1 below.



Scheme 5-1. Sequence of events in chapter 5.

Part I and II describe the bioinformatic and statistical tools used in this chapter in an effort to select potential important points within *E. coli* transketolase sequence; aimed to increase either thermal or pH stability in a rational maner, whereas Part III is aimed to apply recombinant DNA techniques and auntomated methods for the creation and screening of mutant mini libraries on *E.coli* transketolase.

5.2 Materials and Methods

5.2.1 Data Collection and sequence alignment

Homologous sequences were retrieved from Basic Local alignment search tools (BLAST**)** (Ye et al., 2006), and acquired in FASTA format using *E. coli* transketolase sequence as a query. This format includes not only the homolog sequence but also the microorganism that produces it. Subsequently, the sequences were multiply aligned using the program ClustalW (Thompson et al., 1994). Sequences sharing less than 65% residue identity to that from *E. coli* transketolase were removed and the remainders were realigned. Such identity thresholds guarantee a sufficiently accurate alignment and structural homology (Vogt et al., 1995). For the same reason, some sequences were also excluded e.g. those from plants, mammalian cells and putative sequences were not taken into account owing to the ambiguous definition of their function, or temperature and pH optima, such as in plants.

5.2.2 Estimation of optimum pH and temperature for aligned sequences

For some sequences it was not possible to find an exact pH or temperature optimum and so it was necessary to look at the microorganisms that produce them. In these cases, the optimum pH and temperatures for all enzymes in the microorganism were obtained using the BRENDA database (Schomburg et al., 2002; Pharkya et al., 2003). These data were then collected and statistically treated using the standard z-score method to estimate the lower and upper limit for both the optimum pH and temperature of each homolog of the *E. coli* transketolase sequence. The standard z-score (Eq. 5-1) for normally distributed frequencies, is a very useful statistic as it allows us to calculate the probability of a score occurring within our normal distribution.

The standard normal distribution converts the group of data in our frequency distribution such that the mean is 0 and the standard deviation ≡1.

$$z = \frac{x - \mu}{\sigma} \qquad \text{Eq. (5-1)}$$

where, x is a single value; μ and σ are the estimated mean and standard deviation of a population respectively.

Skewness has also provides a better understanding of the data (Donna, 1986) rather than assuming that the frequency distribution is symmetric about the mean. The normal distribution has a skewness of zero. But in reality, data points are not perfectly symmetric so an understanding of the skewness of the dataset indicates whether deviations from the mean are going to be positive or negative.

To estimate the symmetry of the data or the level of skewness the Pearson's skewness coefficient (*J*) was calculated using Eq. 5-2 below:

$$J = \frac{3\,(\mu - median)}{\sigma} \qquad \text{Eq. (5-2)}$$

i) If $J = 0$, then there is no skewness
ii) If J is positive, the skewness is also positive.
iii) If J is negative, the skewness is also negative.

In addition, a confidence interval was calculated for each of the samples by using Eq. 5-3 and Eq. 5-4 depicted below:

$$Z = \frac{X - \mu}{\sigma / \sqrt{n}} \qquad \text{Eq. (5-3)}$$

$$P(-z \leq Z \leq z) = 1 - \alpha = 0.95 \qquad \text{Eq. (5-4)}$$

Where μ is the estimated mean (*Ho*) obtained using a grouped frequency distribution, *X* and σ are the sample mean and standard deviation, and *n* is the number of samples respectively.

Finally, the *p*-value was calculated to reject or accept the so called *null* hypothesis (*Ho*), in other words, to see whether the calculated mean (μ) from each of the frequency distributions is true or false. All estimations were performed with a level of significance α=0.05. Generally, one rejects (*Ho*) if the *p*-value is smaller than or equal to the significance level ($p \leq 0.05$).

5.2.3 Generation of a numerical pH or temperature matrix

After the alignment of all transketolase sequences using ClustalW, and the known or estimated temperature and pH optima were assigned to each of them, a numerical matrix for each property was generated by representing each amino acid contained in the sequence alignment matrix with a number. Each number reflects the composition, polarity and molecular volume of each amino acid, and was taken from the literature (Grantham, 1974). For example, Ala, Val, Ser were represented by 27, 68, 74, respectively. The two resulting matrices contain an X (independent variable) column for each residue position in the *E. coli* transketolase sequence used to create the alignment, and then one Y (dependent variable) column for either the calculated optimum pH or Temperature.

5.2.4 Calculation of the Pearson (*r*) correlation coefficient.

The two numerical matrices (section 5.2.3) were loaded individually into MathLab v. R2007a (The MathWorks, Inc. USA) and de-gapped completely (i.e. columns with gaps were removed from the matrix). A subsequent parametric analysis to determine the Pearson's correlation coefficient (*r*) was performed. Residues were selected based on the correlation coefficient (*r*) results which

were taken into account at a level of significance of α = 0.05. Equation 5-5 below was used to estimate the *t* statistic (*p*-value) depending on the sample Pearson's correlation coefficient (*r*):

$$t = r\sqrt{\frac{n-2}{1-r^2}} \qquad \text{Eq. (5-5)}$$

where *r* is the sample Pearson's correlation coefficient and *n* is the number of pairs of data respectively.

5.2.5 Primer Design

Primers were designed to generate mutations at the residue positions identified by the Pearson (*r*) analysis. The *E. coli* transketolase sequence was used to devise primers for the PCR-based amplification of target sites. The software package AnnHyb 3.5 (Fromant et al., 1995) was used to ensure that the physical properties of these primers were conducive to PCR by meeting the following criteria (Innis, 1990) : (a) 17–28 bases in length; (b) 50–60% G/C content; (c) melting temperature (Tm) of 55–80°C (target was 60°C); and (d) no significant secondary structure. The sequences and physical properties of the primers are depicted in Table 5-1.

Table 5-1 Primers used for site-saturation mutagenesis.

A

Position		Sequence		Tm
NNS186	5'-	GATTGCATTCTACGATGACAAC**NNS**ATTTCTATCGATGGTCACGTTG	-3'	71.98
	3'-	CAACGTGACCATCGATAGAAAT**SNN**GTTGTCATCGTAGAATGCAATC	-5'	71.98
NNS418	5'-	CGGTATGACCGCGATT**NNS**AACGGTATCTCCCTGC	-3'	74.04
	3'-	GCAGGGAGATACCGTT**SNN**AATCGCGGTCATACCG	-5'	74.04
NNS640	5'-	CGGTGAATCTGCT**NNS**GCAGAGCTGCTGT	-3'	73.55
	3'-	ACAGCAGCTCTGC**SNN**AGCAGATTCACCG	-5'	73.55
NNS167	5'-	GCATCTCCCACGAAGTT**NNS**TCTCTGGCGGGTACG	-3'	76.87
	3'-	CGTACCCGCCAGAGA**SNN**AACTTCGTGGGAGATGC	-5'	76.87
NNS558	5'-	GGAACTGATTTTCATCGCTACC**NNS**TCAGAAGTTGAACTGGCTGTTG	-3'	73.45
	3'-	CAACAGCCAGTTCAACTTCTGA**SNN**GGTAGCGATGAAAATCAGTTCC	-5'	73.45
NNS429	5'-	ACGGTGGCTTCCTG**NNS**TACACCTCCACCTTCC	-3'	74.6
	3'-	GGAAGGTGGAGGTGTA**SNN**CAGGAAGCCACCGT	-5'	74.6
NNS513	5'-	GAGCGTCAGGACGGC**NNS**ACCGCACTGATCCTCTC	-3'	77.56
	3'-	GAGAGGATCAGTGCGGT**SNN**GCCGTCCTGACGCTC	-5'	77.56
NNS475	5'-	GGGCCGACTCACCAG**NNS**GTTGAGCAGGTCGCTTC	-3'	77.08
	3'-	GAAGCGACCTGCTCAAC**SNN**CTGGTGAGTCGGCCC	-5'	77.08

Forward and Reverse NNS* primers for Temperature (A) and pH (B) mutants were designed for the sites selected from Pearson statistical analysis.

*NNS primer: allows the incorporation of all 20 amino acids while keeping the total number of codon possibilities low, at 32 (where S is G or C).

Table 5-1 Continued

B

Position		Sequence		Tm
NNS146	5'-	GCCACGACATTGTC**NNS**CACTACACCTACGC	-3'	72.58
	3'-	GCGTAGGTGTAGTG**SNN**GACAATGTCGTGGC	-5'	72.58
NNS104	5'-	GGTCACCCGGAAGTG**NNS**TACACCGCTGGTGTG	-3'	75.84
	3'-	CACACCAGCGGTGTA**SNN**CACTTCCGGGTGACC	-5'	75.84
NNS130	5'-	GGTATGGCGATTGCA**NNS**AAAACGCTGGCGGCG	-3'	74.6
	3'-	CGCCGCCAGCGTTTT**SNN**TGCAATCGCCATACC	-5'	74.6
NNS630	5'-	CCTGAACGGTGCTATC**NNS**GGTATGACCACCTTCG	-3'	74.04
	3'-	CGAAGGTGGTCATACC**SNN**GATAGCACCGTTCAGG	-5'	74.04
NNS112	5'-	CTGGTGTGGAAACC**NNS**ACCGGTCCGCTG	-3'	74.52
	3'-	CAGCGGACCGGT**SNN**GGTTTCCACACCAG	-5'	74.52
NNS175	5'-	CGGGTACGCTGAAG**NNS**GGTAAACTGATTGCAT	-3'	71.26
	3'-	ATGCAATCAGTTTACC**SNN**CTTCAGCGTACCCG	-5'	71.26
NNS347	5'-	GCGAAAGAGTTCATCGCT**NNS**CTGCAGGCTAATCCGG	-3'	74.66
	3'-	CCGGATTAGCCTGCAG**SNN**AGCGATGAACTCTTTCGC	-5'	74.66
NNS134	5'-	GCAGAAAAAACGCTG**NNS**GCGCAGTTTAACC	-3'	69.93
	3'-	GGTTAAACTGCGC**SNN**CAGCGTTTTTTCTGC	-5'	69.93
NNS173	5'-	CTCTCTGGCGGGTACG**NNS**AAGCTGGGTAAACTGATTGC	-3'	75.21
	3'-	GCAATCAGTTTACCCAGCTT**SNN**CGTACCCGCCAGAGAG	-5'	75.21
NNS638	5'-	CCACCTTCGGTGAA**NNS**GCTCCGGCAGAG	-3'	74.52
	3'-	CTCTGCCGGAGC**SNN**TTCACCGAAGGTGG	-5'	74.52
NNS657	5'-	CTTCACTGTTGATAACGTTGTT**NNS**AAAGCAAAAGAACTGCTGTAAT	-3'	69.38
	3'-	ATTACAGCAGTTCTTTTGCTTT**SNN**AACAACGTTATCAACAGTGAAG	-5'	69.38
NNS311	5'-	CGCATGGAACGAGAAA**NNS**GCTGCTTACGCGAA	-3'	72.11
	3'-	TTCGCGTAAGCAGC**SNN**TTTCTCGTTCCATGCG	-5'	72.11

5.2.6 Mutant-library creation for sites predicted from Pearson's coefficient analysis

The statistically selected sites were mutated by site-directed saturation mutagenesis (SDSM). Mutations were carried out using the QuickChange™ Site-Directed Mutagenesis System (QCM) developed by Stratagene (La Jolla, CA, USA). Plasmid pQR791 (5.4 kb) (French & Ward, 1995) was used as the mutagenesis template. All primers were synthesized and purified by Eurofins MWG Operon (Ebersberg, Germany).

The PCR reactions were carried out using a thermocycler (TECHNE, Cambridge, UK). The 50 µL PCR reaction was carried out with 50–100 ng templates, 5µL 10X reaction buffer, 1 µL (~ 500ng) of forward and backward primer pair, 1µL dNTP mix, ddH_2O to a final volume of 50µL, and 1µL of *PfuTurbo*® (Stratagene , USA) DNA polymerase (2.5 U/µL). The PCR reaction was initiated by pre-heating the reaction mixture to 95°C for 0.5min; 95°C for 0.5min, 55°C for 1 min and 68°C. Usually between 16-18 cycles were performed, ~4 minutes/kb of plasmid length. The PCRs were purified by QIAquick™ PCR purification kit (Qiagen, Germany) and further treated with restriction enzyme 10 U/µL DpnI (Roche Diagnostics, Mannheim, Germany). The PCR products were evaluated by agarose gel electrophoresis.

5.2.7 Transformation of mutant plasmids into *E. coli* cells

Purified PCR product (4 µL) was directly transformed into 50 µL of either XL10-Gold ultracompetent cells or TOP10 competent cells (section 2.3.8). Transformants were incubated in 5 mL LB medium for 1h and then plated out in pre-prepared LB-agar plates containing 150 µg/mL ampicillin and incubated overnight. From the typically >200 colonies obtained for each mutation site, a total of 10 random colonies were selected and their plasmid libraries were isolated by mini-prep and used for DNA sequencing to assess the mutagenesis. All plasmids (mutants) were sequenced by Scientific Support Services - Wolfson

Institute for Biomedical Research (UCL). Samples were prepared according to the DNA service guidelines (section 2.3.12)

5.2.8 Picking and growing colonies in a 96 deep-well plate

Microwell fermentation was carried out in 2 mL 96 deep-well plates (DWP). Wells were filled with 90 µL LB Amp^+ (150 µg/mL) medium and inoculated with individual colonies of *E. coli* XL10-Gold pQR711. Transfer of colonies from LB-agar plates into 96 deep-well plates was performed using a programmed Qpix robot (Genetix Ltd, Hampshire UK). Usually 90 *E. coli* potential mutants were picked, 3 wells were left empty as a control and the rest inoculated with either TOP10 or XL10-Gold expressing wild-type *E. coli* transketolase from pQR711. Each plate was sealed by taking an identical empty plate, inverting it, and taping it over the top. The sealed plates were incubated for 16 hours at 900 rpm agitation speeds at 37°C.

5.2.9 Automated preparation of reaction and glycerol plates using TECAN handling robot work station.

Preparation of 50 µL reaction and 100 µL glycerol stock plates, and OD_{600nm} measurements of cell density were carried out using a programmed automated workstation Tecan Genesis platform (Tecan Trading AG, Switzerland). Using 96 shallow-well plates (SWP), four reaction and two 20%-glycerol stock plates were prepared from the 96-DWP overnight culture. Additionally, one SWP for each library was prepared with overnight culture and water or fresh LB-media at a 1:10 dilution for OD_{600nm} measurements using a Tecan plate reader included in the TECAN work station. Reactions and glycerol stock plates were then heat-sealed with foil, labelled and immediately placed in a -80°C freezer until required.

5.2.10 Lysis of reaction plates

Reaction plates were thawed from −80 °C resulting in freeze–thaw lysis of the 50 µL of cell culture. This method was performed at least twice to each plate before use.

5.2.11 Screening of mutants for increased thermal or pH resistance.

For thermal inactivation assays, reaction plates were first placed in a pre-heated water bath for 1h at 60 ^{o}C (see results section 4.3.1) and immediately place on ice (section 4.2.1). The edge trims of the plate were cut off to increase the heat transfer between the pre-heated water-bath and sample plates. The temperature inside the wells was monitored at all times using a digital wired-thermometer in at least 6 different wells. Screening of pH mutant libraries was carried out by adjusting the pH at pH 4 and pH 10 using Bis-Tris and CAPS buffer respectively. The same buffers were used to adjust the cofactor and substrate solutions to pH 4 and pH 10 respectively. Finally, the standard *E. coli* transketolase reaction was carried out (section 2.2.7) to estimate residual transketolase activities of the mutants. The concentration of L-erythrulose reaction product (residual activity) was estimated by the 5 min fast primary screening (section 2.2.8). Interesting mutants were re-evaluated using the 16min high accuracy HPLC method (section 2.2.8).

5.3 Evaluation of non-parametric statistical methods.

5.3.1 Calculation of Kendall (τ) correlation coefficient.

The same pH and temperature matrices described earlier (section 5.2.3) were used to estimates the Kendall's correlation coefficient (τ) using previously described procedures (Kendall & Gibbons, 1990). Matrices were loaded into MathLab v. R2007a (The MathWorks, Inc. USA). Results were evaluated with a level of significance $\alpha=0.05$. The Kendall tau coefficient (τ) has the following properties:

i) If the agreement between the two rankings is perfect (i.e., the two rankings are the same) the coefficient has value 1.

ii) If the disagreement between the two rankings is perfect (i.e. one ranking is the reverse of the other) the coefficient has value −1.

iii) For all other arrangements the value lies between −1 and 1, and increasing values imply increasing agreement between the rankings. If the rankings are completely independent, the coefficient has value 0 on average.

5.3.2 Principal component analysis and Partial least square

Two different prediction methods were selected in an effort to identify new possible mutation targets in the *E. coli* transketolase sequence. Principal components analysis (PCA) and partial least-squares (PLS) were performed with PLS_Toolbox v.3.5 (Eigenvector Research, Inc., USA) installed in MathLab v. R2007a (The MathWorks, Inc. USA). PLS was used to identify any possible relationship between the X (independent variable) and Y (dependent variable) elements that constitute the pH or temperature matrix (section 5.2.3). PCA provided a more exploratory data analysis. Calculations included in the present study were performed after mean centering the data for each attribute.

5.4 Results and Discussion

5.4.1 Data collection and sequence alignment

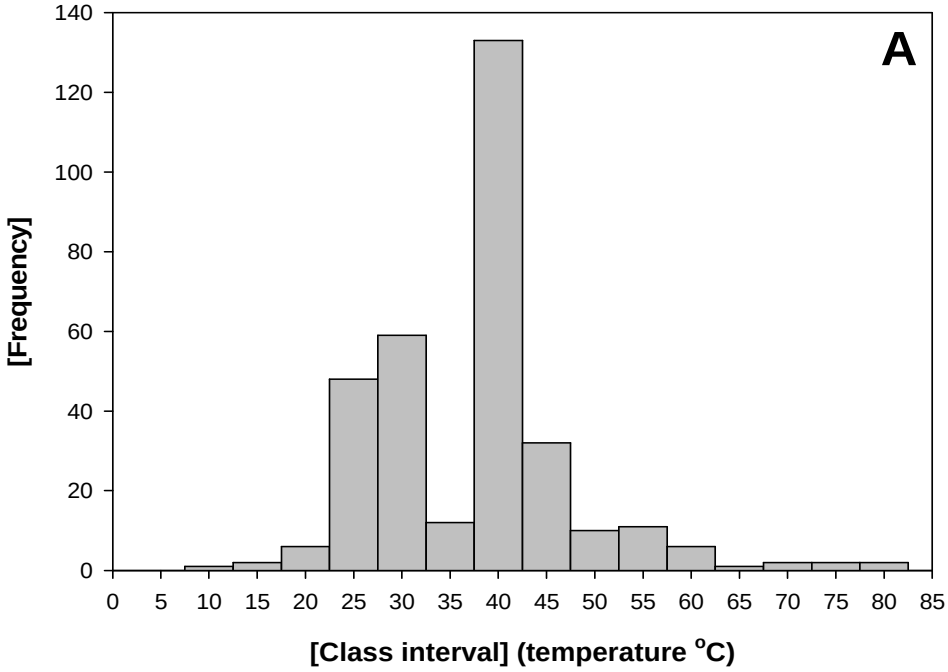
A total of 90 enzyme sequences were obtained in FASTA format, with the conditions described in section 5.2.1, by using the default program parameters (gap initiation penalty of 11, gap extension penalty of 1), using *E. coli* transketolase sequence as a query. Each sequence was used as a query in the BRENDA database to obtain the optimum pH and temperature from all enzymes produced in the corresponding microorganism where neither pH nor temperature was already known for the enzyme. Finally, only 50 and 43 out of 90 sequences were kept to build-up the pH and temperature matrices respectively.

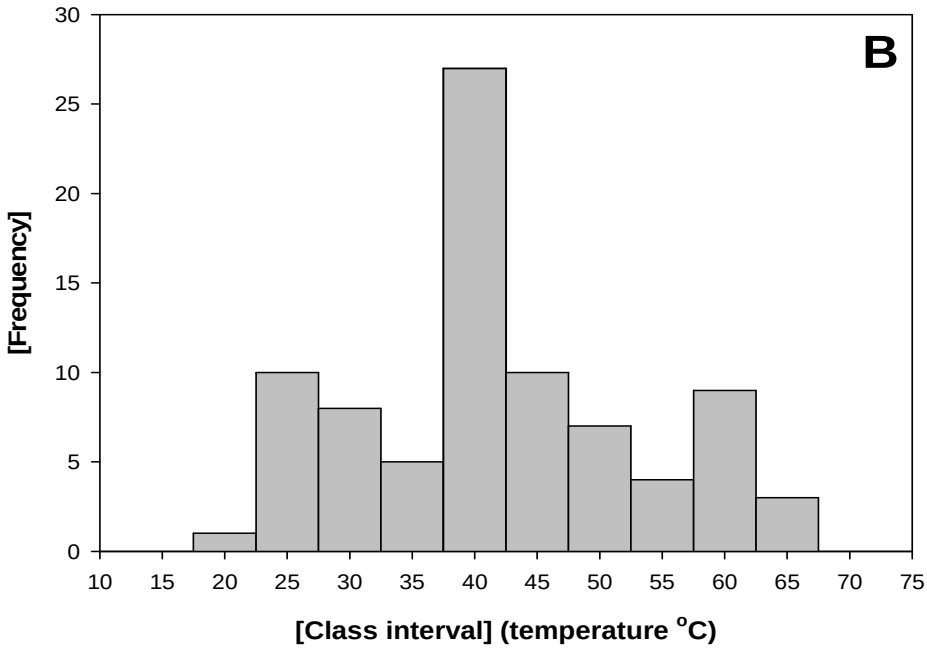
5.4.2 Estimation of optimum pH and temperature for aligned sequences

The final set of data from each matrix was used to estimate the apparent optimum pH or temperature of some of the homolog sequences by applying a frequency distribution analysis and validating them by comparing the *p*-value with the level of confidence alpha (α). Typical frequency distribution histograms of selected microorganism are presented in Figure 5-1 and Figure 5-2, for the optimum temperature or pH data obtained from the BRENDA database.

It can be seen in Figure 5-1 and Figure 5-2, that the distribution on the right and left-sides of the mean taper differently, indicating some level of skewness present. For the temperature histograms (Figure 5-1 A, B and C), the level of skewness, as calculated using Pearson's skewness coefficient (*J*), from each set of data is depicted in Table 5-2; where *E. coli* (A) shows a slightly negative skewed distribution as the values are biased towards the negative end of the distribution; whereas *Bacillus subtilis* (B) and *Saccharomyces cerevisiae* (C) are slightly positively skewed as the extreme values are biased towards the

positive end of the distribution. For pH, (Figure 5-2 A, B and C) it can be seen that there is a slightly negatively skewed distribution in each case.





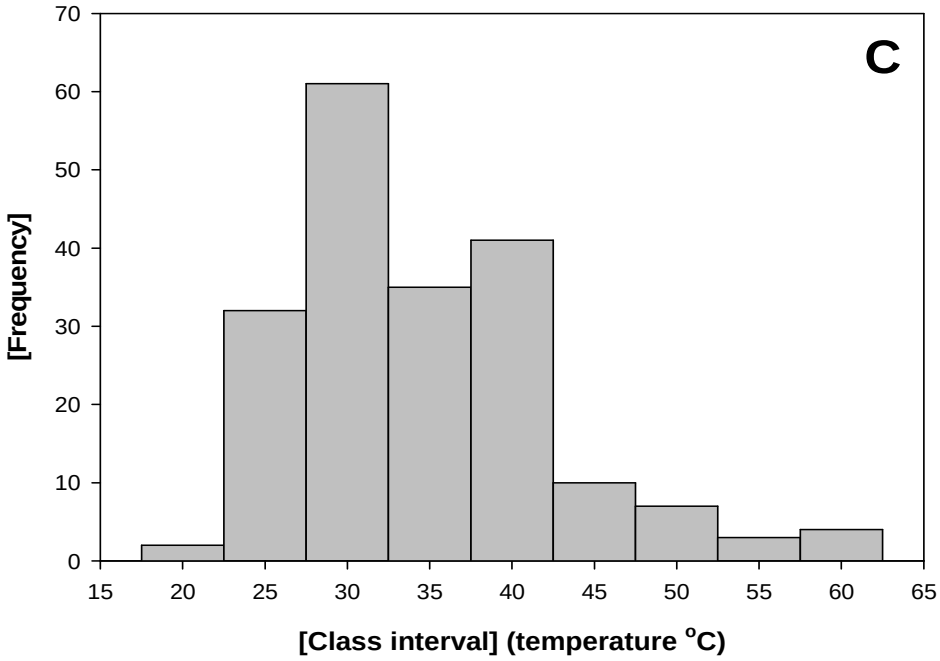
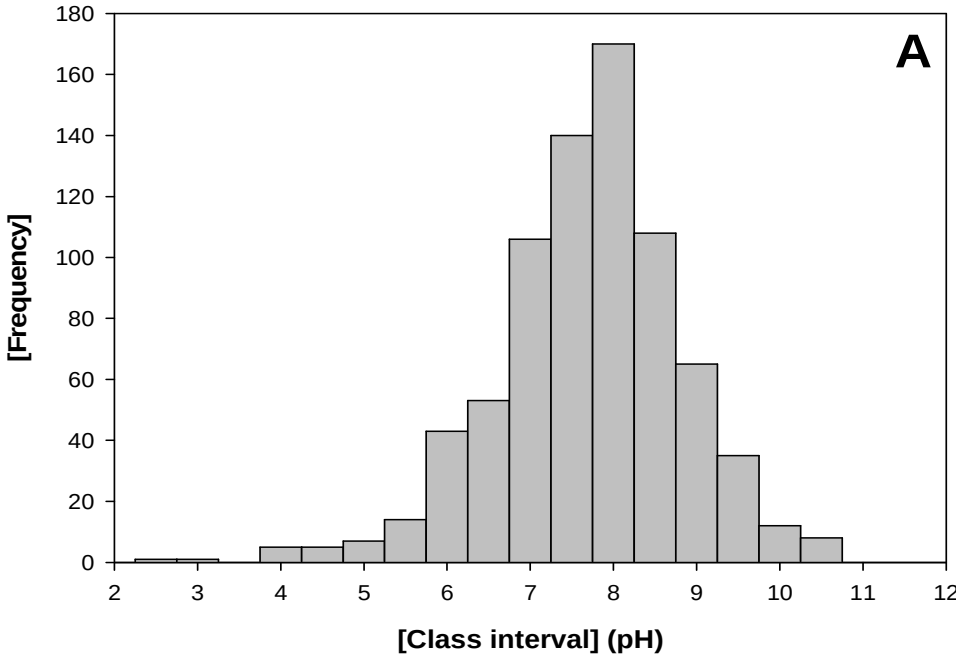
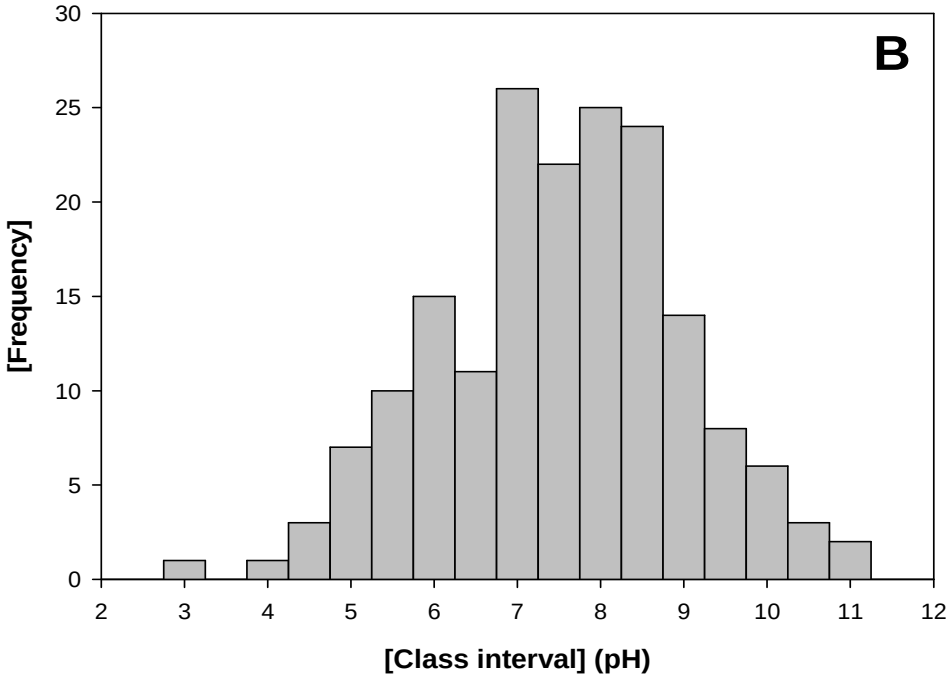


Figure 5-1 Frequency distribution histograms for temperature.
Figures are as follows: *E. coli* (A); *Bacillus subtilis* (B), and *Saccharomyces cerevisiae* (C), respectively. Data were acquired from BRENDA by using the microorganism name associated with each aligned sequence.


A
[Frequency]
[Class interval] (pH)
0
20
40
60
80
100
120
140
160
180
2
3
4
5
6
7
8
9
10
11
12


B
[Frequency]
[Class interval] (pH)
0
5
10
15
20
25
30
2
3
4
5
6
7
8
9
10
11
12

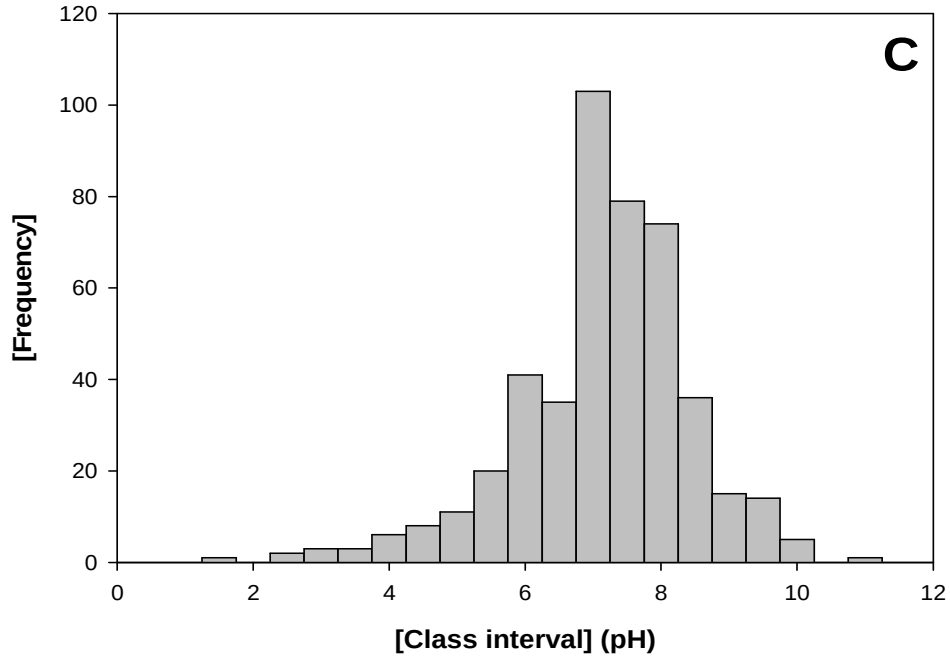


Figure 5-2 Frequency distribution histograms for pH.
Figures are as follows: *E. coli* (A); *Bacillus subtilis* (B), and *Saccharomyces cerevisiae* (C), respectively. Data were acquired from BRENDA by using the microorganism name associated with each aligned sequence.

Table 5-2 Central tendency parameter for temperature and pH data.

A. Temperature

Parameters		*E. coli*	*B. subtilis*	*S. cerevisiae*
Mean	(*Ho*)	35.58 ℃	39.9 ℃	32.56 ℃
average		36.19	40.7	33.57
median		36.8	38.83	30.8
lower limit	(-z)	34.96	37.51	32.32
upper limit	(+z)	37.2	43.4	34.82
Skewness	(*J*)	-0.365	0.285	0.66
p-value		>0.05	>0.05	>0.05
Known optimum		37-40	30-37	30-35

B. pH

Parameters		*E. coli*	*B. subtilis*	*S. cerevisiae*
Mean	(*Ho*)	7.5	7.19	6.92
Average		7.62	7.42	7.059
Median		7.58	7.39	7.02
lower limit	(-z)	7.5	7.21	6.9
upper limit	(+z)	7.69	7.64	7.17
Skewness	(*J*)	-0.2065	-0.291	-0.209
p-value		>0.05	>0.05	>0.05
Known optimum		7-8	7-8	7-8

The skewness parameter helps to estimate the distance between the mean and the median and consequently the symmetry of the dataset (see Eq. 5-2 section 5.2.2). Usually when the mean is greater than the median, then it is considered a positive skewed distribution; whereas when the mean is smaller than the median then it is considered a negative skewed distribution (Table 5-2), and when skewness is equal to zero then it is said that the data are symmetric or normally distributed, which is clearly not the case for any of the Figures above. If the probability distribution of a random variable is not too different from the normal distribution (e.g. its cumulative distribution function does not have any discontinuities and its skewness is moderate) (Donna, 1986; Cadogan & Sutton, 1994; Dytham, 1999), then even with only a few dozen observations in the sample, an approximate confidence interval for a population mean can be constructed, relying on the central limit theorem.

For transketolase, the sample size of each dataset from the different microorganisms was always equal to or greater than 30 (some contained up to 700 enzyme data) then it can be assumed that the samples are large enough such that the approximation is expected to be reliable. As it can be seen from Table 5-2 those conditions are present. Indeed, usually the estimated mean (*Ho*) (i.e. population mean not arithmetical average) always happened to be between the lower and the upper limit, usually determined as a function of *Z* (Eq. 5-3).

For those data that exhibit moderate skewness (i.e. $J \geq \pm 0.5$) such as *S. cerevisiae* (see Table 5-2) one approach was applied which consisted of some type of "transformation" to try to make the data normally distributed, or more nearly normal. The square root of the data set was taken and then the analysis was performed. After applying the "transformation" most of the data presented a *p*-value>0.05. In general, all estimated mean (*Ho*) laid within the confidence interval and more importantly, they presented a *p*-value>>0.05 (Table 5-2). Statistically speaking, the calculated means from the frequency distribution analyses of each data set can be "approved" with 95% confidence (Nist/Sematech, 2008).

Previous studies have found that the relationship between the optimum pH or temperature of the source microorganism is, in many cases, weakly correlated with the optimum pH or temperature for any given enzyme (Schiraldi & De Rosa, 2002; Daniel et al., 2008). Our new and simple approach to estimate these optima for a given enzyme uses the known enzyme optima data rather than the known growth optima for each micro-organism. This approach is expected to improve the correlations, particularly for transketolase for several reasons. Firstly, the transketolase optima have been directly measured for X cases used, and those for *E. coli* and *S cerevisiae* approximate closely to the mean values for enzymes in these micro-organisms (Table 5-2). Second, the statistical analysis of the distributions of enzymes within each micro-organism shown that most of the enzymes fall within ±5 °C and ±1 pH unit of the mean values (Figure 5-1 and Figure 5-2). Finally, as transketolase is an ancient housekeeping enzyme it is more likely to have optimum activity under the average conditions for each micro-organism.

5.4.3 Generation of a numerical pH or temperature matrix

Separate matrices were generated for temperature and pH optima. Each matrix contains m variables (amino acid position) and n individual sequences with their associated pH or temperature optimum values, giving a total of $n \times m$ data points. For each matrix value, Xrj represents the rth observation on the jth variable where r= 1 to n and j=1 to m.

After the alignment of the homolog sequences was performed, it was partially de-gapped (i.e. columns with no data or less than 30 were removed) and each amino acid was assigned a number as in Table 5-3, derived from the composition, polarity and molecular volume of each amino acid (Grantham, 1974). This interchange to numbers enabled statistical methods to be applied more easily.

Table 5-3 Values assigned to each amino acid.

Residue	Value	Residue	Value
Ala	27	Leu	98
Arg	103	Lys	102
Asn	91	Met	87
Asp	108	Phe	114
Cys	169	Pro	1*
Glu	93	Ser	74
Gln	76	Thr	38
Gly	42	Trp	147
His	77	Tyr	110
Ile	95	Val	68

Each number involves composition, polarity and molecular volume of each amino acid (Grantham, 1974).
*Residue taken as a base.

The final matrix obtained is called the *data matrix* and is denoted by X_{pH} or X_{temp}, for pH and temperature respectively. Thus the dimensions are:

$$X_{pH} = (n \; x \; m) = 50 \; x \; 573$$

$$X_{temp} = (n \; x \; m) = 43 \; x \; 573$$

Where, n is equal to the total number of rows and m is equal to the total numbers of columns. Even though the number of columns is the same, it does not necessarily mean that we are talking about the same residue at certain position as they contain some different sequences within each alignment (i.e. not all the sequences aligned in the pH matrix are present in that for temperature and vice versa. The rest of the colums were not taken into account due to lack of data (i.e. less than 10 data out of 50)

5.4.4 Calculation of the Pearson (r) correlation coefficient.

In an effort to look for a relationship between amino acid sequence position and either optimum temperature or pH, the estimated pH (Y_{pH}) and temperature (Y_{temp}) values (section 5.4.2) were coupled to its corresponding matrix (X_{pH} or X_{temp}), and the Pearson correlation coefficient (r) was calculated for each of them.

Usually Pearson coefficient (*r*) indicates the strength and direction of a linear relationship between two random variables, where it is possible to have either a negative or positive correlation between pair of normally distributed data. As it can be seen from Figure 5-3 and Figure 5-4, in both cases, the estimated Pearson correlation coefficients are predominantly not higher than 0.5 which essentially reflects a medium correlation between data (Donna, 1986; Ennos, 2000).

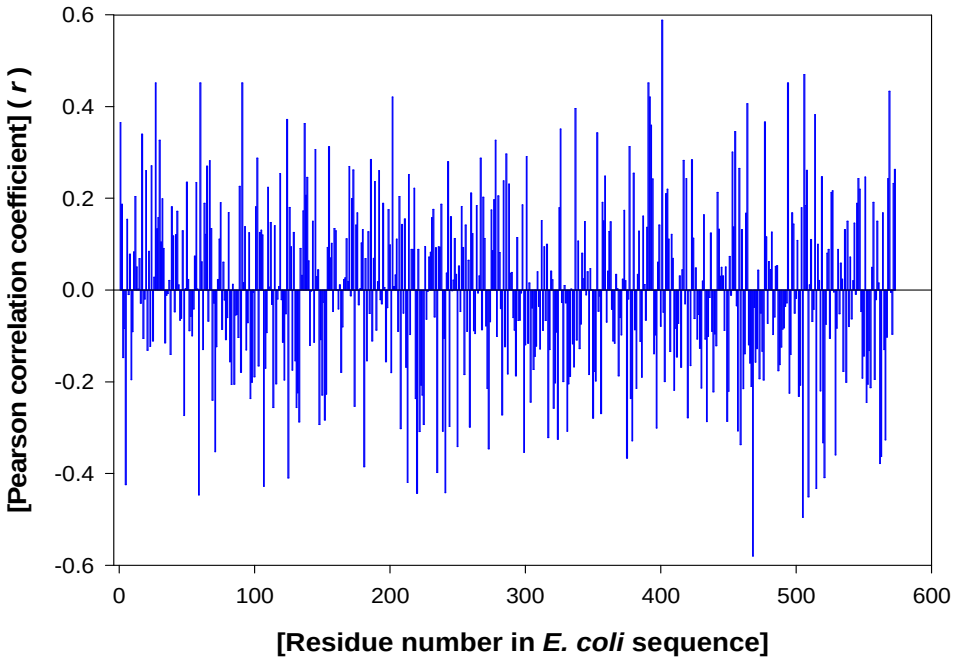


Figure 5-3 Estimated Pearson correlation coefficient (*r)* for temperature matrix.

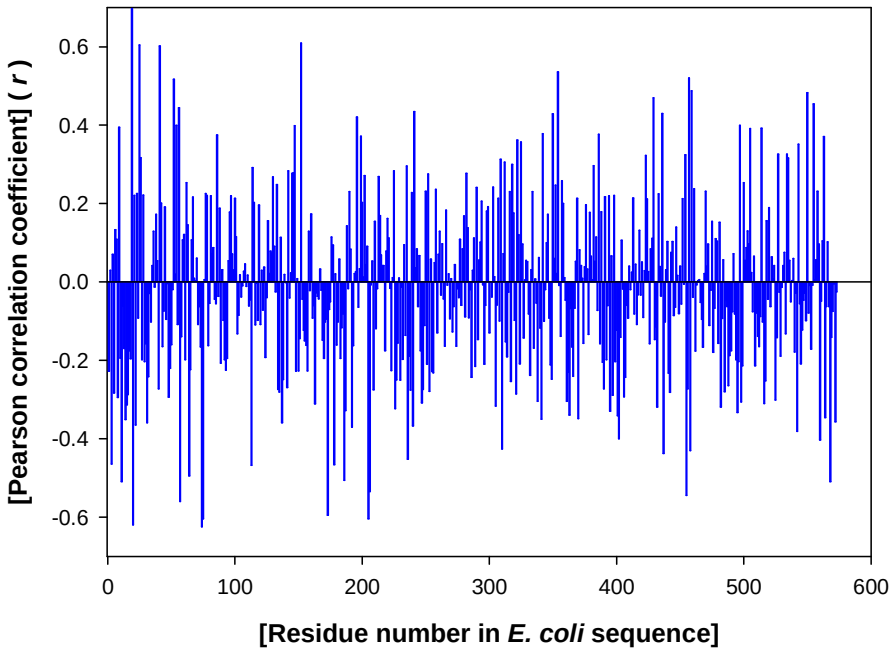


Figure 5-4 Estimated Pearson correlation coefficient (*r*) for pH matrix.

Residues with statistically significant correlations were selected by comparing the *t* statistic (dependent on the sample Pearson correlation coefficient , *r*) (Eq. 5-5 section 5.2.4) with the level of significance (α=0.05). In this case, the *p*-value is the observed significance level of the test (*t*). Initially a total of 160 single residues for temperature and 106 for pH matrices were selected, those that were observed to have a significance level (*t*) higher than the chosen significance level (i.e. *p*-values = $t > 0.05$). As it was assumed that the data followed a parametric distribution, then the *t* values higher than alpha (α) are considered to have some level of correlation. Nevertheless, Pearson correlation indicates the strength of a linear relationship between two variables, its value alone may not be sufficient to evaluate this relationship (Anscombe, 1973; Cohen, 2003), especially in the case where the assumption of normality is believed to be incorrect so a further individual examination of selected residues was performed and a final set of 8 (temperature) and 12 (pH) single residues were selected (Table 5-4) for site-saturation mutagenesis.

Table 5-4 Selected residues based on Pearson correlation coefficient values with a level of confidence α=0.05.

A

Residue	Pearson (*r*)	Position	Residue in *E. coli* TK	Most frequent amino acids in the column	Position	*t* values† (*p*-value)
186	0.2688	Surface	Gly	H,G	Loop	1.677
418	-0.346*	Buried	Ala	A,M	Helix	2.157
640	-0.2232*	Surface	Pro	D,P	Loop	1.392
167	0.2819	Buried	Cys	A,C	Helix	1.758
558	-0.2364*	Buried	Gly	G	Helix	1.475
429	0.2383	Buried	Pro	A,P	Sheet	1.487
513	0.2494	Buried	Pro	N,P	Loop	1.556
475	-0.2908*	Buried	Pro	A,P	Loop	1.814

B

Residue	Pearson (*r*)	position	Residue in *E. coli* TK	Most frequent amino acids in the colum	position	*t* values† (p-value)
146	0.5168	Surface	Asp	N,D	Loop	3.532
104	0.6998	Surface	Gly	F,G	Loop	4.772
130	0.6019	Surface	Glu	A,E	Helix	4.110
630	0.4693	Surface	Val	I,V	Sheet	3.033
112	0.2251	Buried	Thr	T	Loop	1.542
175	-0.6042*	Buried	Leu	L	Loop	4.126
347	-0.6042*	Surface	Lys	E,K	Helix	4.126
134	0.2908	Surface	Ala	G,A	Helix	1.991
173	-0.6244*	Surface	Leu	W,L	Loop	4.261
638	-0.4372*	Buried	Ser	S	Loop	2.991
657	-0.54456*	Surface	Ala	E,A	Helix	3.721
311	-0.2702*	Surface	Phe	L,F	Helix	1.850

Table A and B represent the selected residues aimed to increase the temperature and pH stability of the enzyme respectively. Residues were selected with a level of significance α=0.05. Calculated *t* values for each of them were shown to be greater than α for these cases.

† The absolute value of negative *t* was taken for analysis purposes.

* Negative Pearson values represent a negative relationship.

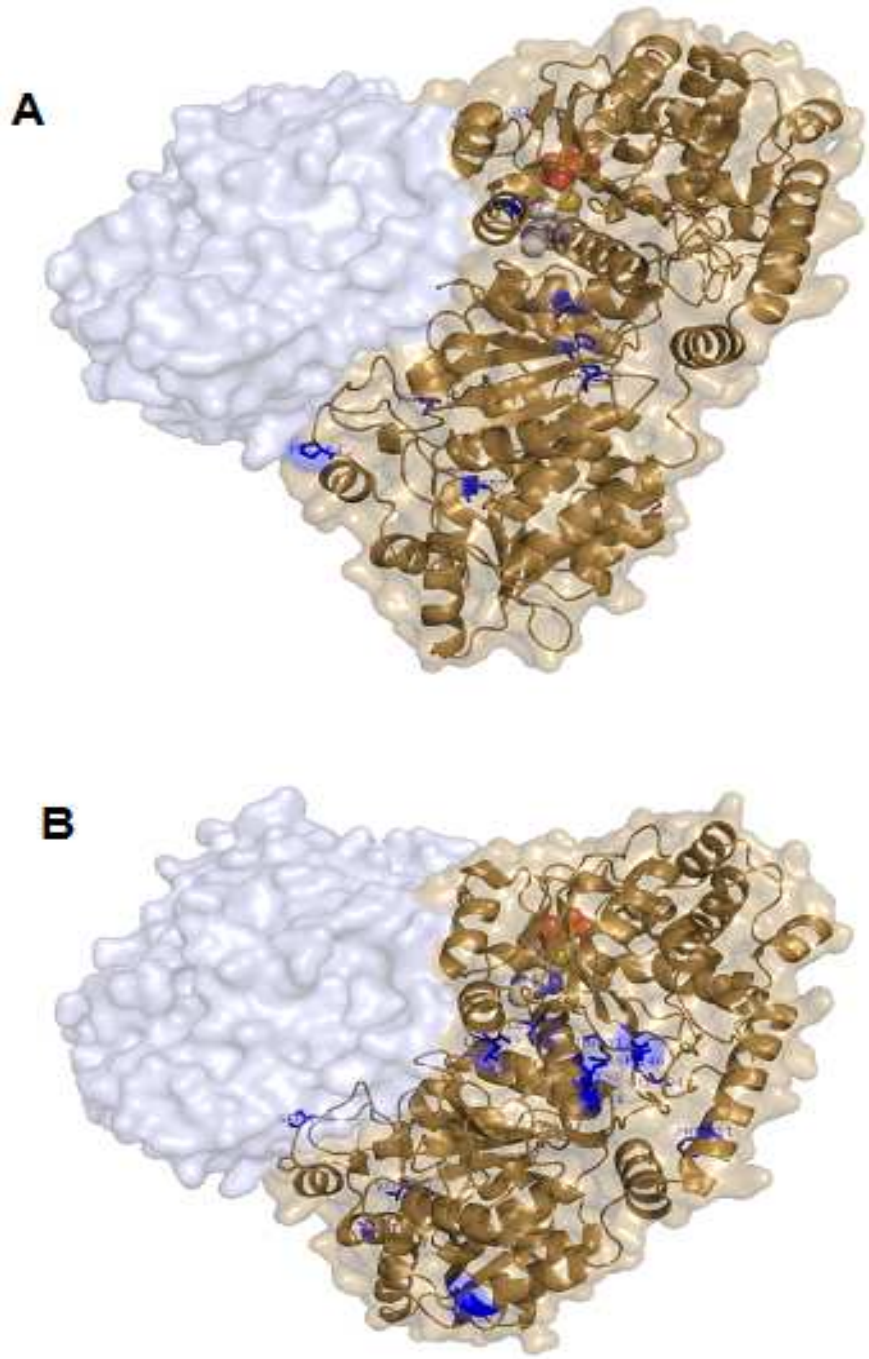


Figure 5-5 Selected amino acids within *E.coli* transketolase aimed to increase temperature and pH stability of enzyme.
Selected (**A**) temperature and (**B**) pH amino acids are coloured in blue (see Table 5-4). PDB file was obtained from 1qgd.pdb (Littlechild et al., 1995). One monomer is shown as cartoon/surface in light brown with selected amino acids. The second monomer is shown as surface in light blue. For one active site only, the TPP cofactor is highlighted in CPK colours and as spheres. Figure generated with PyMol (DeLano, W.L. (2002), The PyMOL Molecular Graphics System on World Wide Web http://www.pymol.org).

The rest of the non-selected residues with smaller *p*-values though still greater than alpha (α) exhibited in most cases a very good level of correlation (-0.6>*r* and *r*>0.6) (data not shown). Hence, a further visual analysis of each set of data revealed that those residues were either highly conserved (~90%) (i.e. only between 2-5 of the total residues in that position were different) or semi-conserved (~50%) (i.e. only two or three residues, of the total number in that single position, were present in the set). Those sets of data with "high" Pearson correlation coefficients and yet very low correlation when visually analysed are deemed to be false positives (Anscombe, 1973). However, these residues might be considered important if the aim were to identify regions with conserved or semi conserved residues within the enzyme alignment or if a different analysis is applied (i.e. non-parametric analysis). While the final chosen residues (Table 5-4) are shown to be statistically significant, the Pearson values for pH gave much stronger correlations (below -0.4 or above 0.6), than for temperature (below -0.2 or above 0.2).

This method is very useful when data are normally distributed. However, the level of correlation was no higher than expected and also the Pearson's (*r*) interpretation is a bit obscure. It might be necessary to apply a different statistical approach in an attempt to increase the level of correlation or to use a more simple analysis that is not so ambiguous. For example, Kendall's correlation coefficient (τ) can be used as it is a powerful analysis when it is found that pairs of data do not follow a normal distribution (Kendall, 1976).

5.4.5 Library creation for Pearson's (*r*) coefficient results.

Site-saturation mutagenesis was successfully performed by using the designed plasmids for pH and temperature (section 5.2.5) and conditions described in section 5.2.6. The plasmids obtained after site-saturation mutagenesis (Figure 5-7) gave the open circular form.

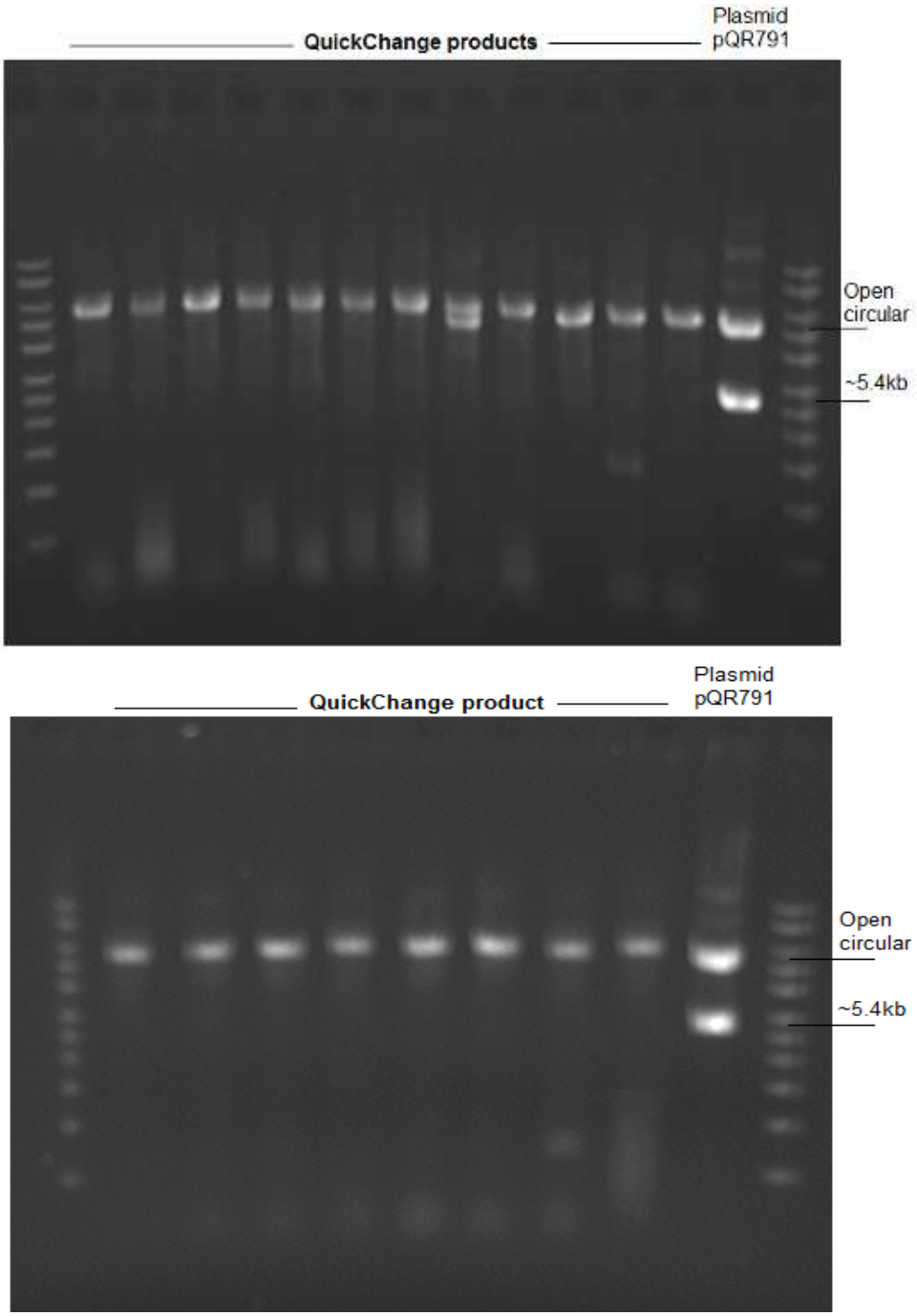


Figure 5-6 Electrophoresis gel of QuickChange products.
Agarose gels (0.6%) show the QuickChange™ products performed by using site-saturation primers for temperature (top) and pH (bottom) respectively. Plasmid pQR791 (far right site) was used to compare it with plasmid DNA products.

The QuickChange™ products for both pH and temperature were transformed into XL10-Gold *E. coli* ultracompetent cells (section 5.2.7) for library creation. For each library 90 colonies were picked into a 96 deep-well plate along with 3 wild-type controls as described in section 5.2.8. By sequencing the DNA of 10

random samples from each 96 deep-well plate it was found that at least 70% of the sequences contained a mutation in the targeted position, whereas the remaining 30% were either wild-type clones or contained a stop codon. That suggested that the PCR parameters and primers used were adequate for targeting single mutations.

After mutations were confirmed for each of the targeted points, then the preparation of two glycerol and four reaction plates were carried out using the automated TECAN handling robot (section 5.2.9). Essentially, 6 plates were prepared for each mutant library so 20 single mutants times 6 equals to 120 plates in total. By using automated methods for picking and preparing the reaction and glycerol plates, the average time for preparing one library was around two and a half days.

5.4.6 Screening of potential mutants for increased thermal or pH resistance based on Pearson (*r*)analysis.

In the case of temperature, the selection process consisted of incubating the mutant libraries for 1h at 60 ^{o}C as described in Chapter 4, and then assaying for the relative activity at room temperature (25^{o}C). It should be emphasized that this approach does not directly select for mutants that are active at elevated temperatures and consequently for a higher T_m, but rather it identifies variants with an increased propensity for refolding (as compared with wild-type enzyme) or/and those that are able to retain their native structure at elevated temperature. After screening 8 plates (~768 colonies) each with a single mutation as described above, the quick HPLC method was used (section 2.2.8) for primary screening.

From the total plates screened only five from eight showed variants that exhibited some activity. Variants were selected from plates A418 (1); G558 (3); P429 (4); P513 (1); P475 (3) respectively (Figure 5-7), where the number between brackets stand for the number of variants within that sample plate.

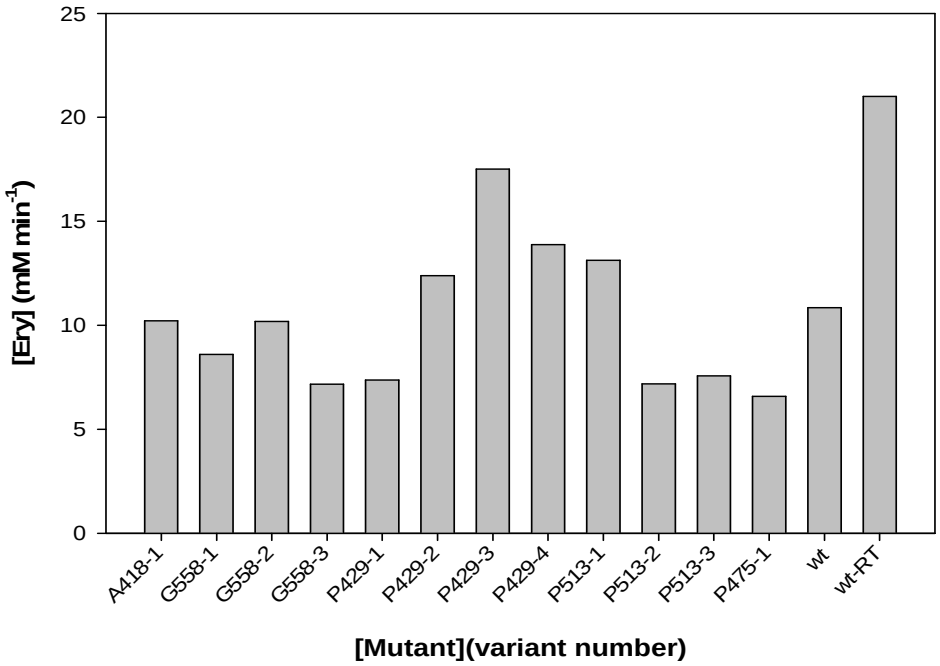


Figure 5-7 Comparison of activity of mutants selected from primary screening. TK variants were first heated-up and then cooled down at 60 ^{o}C for 1h. Activities of the variants were measured in cell lysate with 50mM HPA, 50 mM GA, and 50 mM Tris-HCl, 2.4 mM TPP, 9 mM $MgCl_2$, pH 7 at 25 ^{o}C in a final volume of 300 µL. Activity of wild-type TK at room temperature (wt-RT) was also included.

Some of the variants showed either lower or higher activity relative to that from wild-type (wt). This difference may be caused by variations in enzyme concentration as crude lysates were used in this first screen. Selected mutants were re-tested with the 16 min high accuracy HPLC method (section 2.2.8). This time the enzymes were purified using the bench scale method described in section 2.3.14 and the enzyme concentration determined at OD_{280nm} and by SDS-PAGE band densitometry (section 2.3.14) to obtain the specific activity.

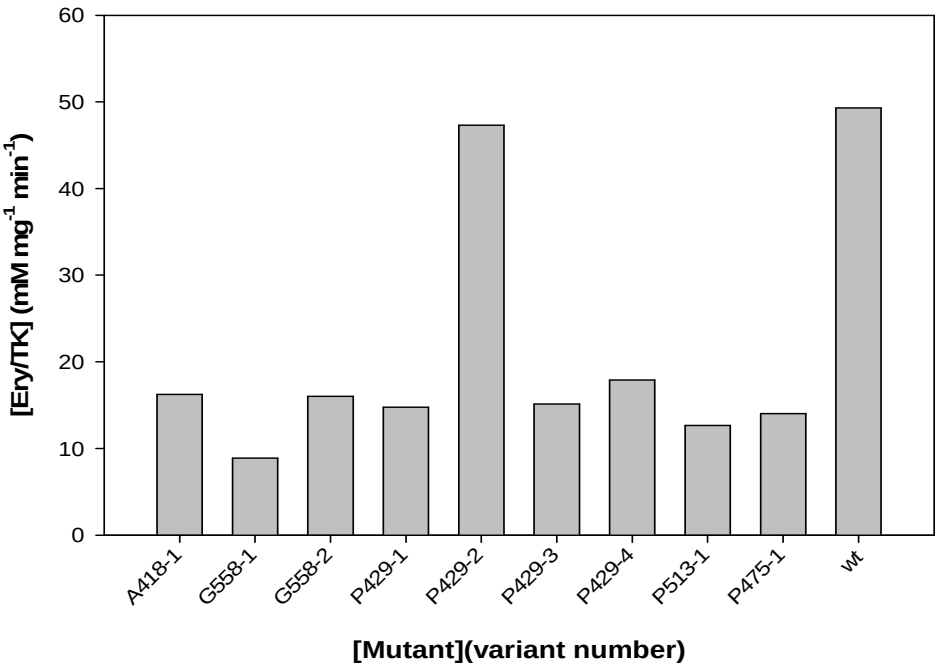


Figure 5-8 Comparison of the activity of mutants obtained by using the high accuracy screening.
Pure TK and variants were first heated-up and then cooled down at 60 ^{o}C for 1h. Activities of the wild-type and variants were measured in with 50 mM HPA, 50 mM GA, and 50 mM Tris-HCl, 2.4 mM TPP, 9 mM $MgCl_2$, pH 7 at 25 ^{o}C in a final volume of 300 µL.

Figure 5-8 shows the results of the high accuracy screening. Only one mutant (P429-2) showed greater activity than the rest of the variants, but not with respect to that from wild-type TK. That single mutant was prepared to be sequenced. Sequencing results showed that the "variant" was also a wild-type clone so it was not further analysed.

It was not very surprising to find that the majority of the variants showed a lower propensity to refold after being heated at 60 ^{o}C as most of the mutations were mutated from a proline residue. The removal of proline from the enzyme could cause an increment in its flexibility and consequently a reduction of TK compactness according to the "proline" rule (Macarthur & Thornton, 1991; Suzuki, 1999). Conversely, the only variant that showed almost the same activity or high propensity to refold as wild-type (P429) after being heated-up turned out to be a clone. It can be seen that these proline mutations may play an important role in the stability of the enzyme at high temperatures. Moreover, three of the sites, P475, P513 and P640 are found in a loop which potentially becomes more disordered upon proline removal. In the case of C167, its removal from the enzyme might also have a destabilising effect in terms of stability from the loss of a hydrogen bond, as it does not have a neighbouring cysteine to produce disulphide bond. For A418 and G558 a few variants were present in the final screening with some residual activity, showing that they are important for transketolase stability. By contrast, for G186 no variants were found.

Variation in protein is so great, and the intermolecular interaction so complex (Querol et al., 1996), that there has been a little success in developing simple and widely applicable amino acid substitution "rules" by which protein stability may be increased. Previous works (Matthews, 1993b) have led to great increases in the understanding of the way that amino acid substitutions influence structure and stability. On the one hand the majority of point mutation in proteins have little or no effect on stability, suggesting that only a minority of amino acids are involved in significant stabilizing interactions. On the other, quite extensive efforts to identify single point mutations that enhance

protein stability have been relatively unsuccessful in increasing half-lives by amounts corresponding to stabilizing energy differences as much as 5kJ.mol^{-1} (Perutz & Raidt, 1975). Nevertheless accumulated point mutations have led to stability increases of more than 15kj.mol^{-1} (Dasgupta & Bell, 1993). There may therefore be a possibility that most single mutations do not affect beneficially the stability of the enzyme and that combining some of them could attain this goal.

The conformational stability of the enzyme can be increased by rigidifying the enzyme structure by such means as cross-linking and immobilization (Germain et al., 1989), but such a loss of flexibility is also likely to reduce enzyme activity. Thus the most useful approach might be to rigidify the enzyme outside the active-site region, and thus leave the active site flexible enough for high activity. Regrettably, this may leave the active region of the enzyme as the most susceptible to degradation. Strategies are also available for reducing degradation, for example by mutating susceptible residues such as Asp to less labile, but structurally similar, amino acid Glu (Ahern et al., 1987; George-Nascimento et al., 1990). Like these and other approaches, as there are many potential types of mutation that can enhance the thermal stability of an enzyme, the key goal is to rationalise the methods of choosing enzyme residues so that fewer enzyme variants have to be screened. Our approach looked for correlations between residues and their temperature optima. However, it may be that even with a positive correlation, it may be that the wild-type sequence already contains the best residue, or that other residues must be changed simultaneously due to a functional link between them before any improvement can be observed.

For the pH variants the same screening criteria was applied to selected mutants with a potential increased stability at extreme pH (i.e. pH 5 or pH 10), where wild-type TK was found to lose activity by an average of ~50% Chapter 4. First a quick primary screening (section 2.2.8) was performed using cell lysate, followed by a secondary high accuracy 16 minute HPLC method (section 2.2.8) to re-test the selected variants obtained from the previous screening. Secondary and final screenings were performed with a clarified lysate and pure variants respectively. The concentration of both clarified and wild-type variants was calculated as above to obtain specific activities.

Results from the secondary and final screening of TK variants at pH~5 are shown below. Selection of the potential variants with increased pH stability was carried out by coupling their obtained activity at pH 4.5 versus that obtained at pH 4.75. This method allowed the identification of variants in different regions: A) low activity at pH 4.5 and high at pH 4.75; B) Low activity at both pHs 4.5 and 4.75, and C or D) between high and moderate activity at both pH's. Different wild-type enzymes (wt) were used as a reference for defining the graph limits. For this case, the variants chosen were those that fell into the A, C and D areas (Figure 5-9), and were subsequently purified and re-tested.

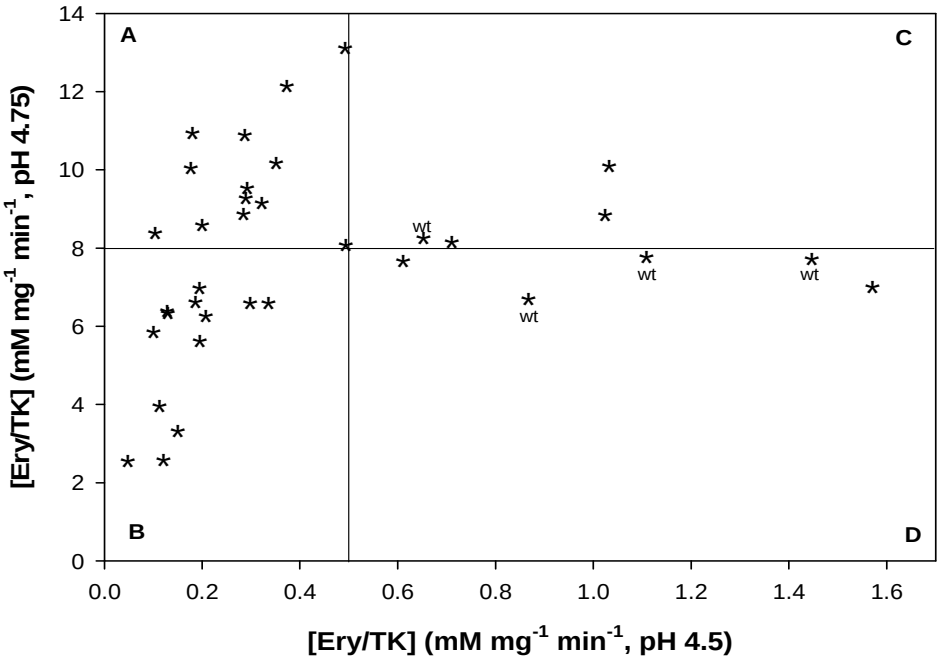


Figure 5-9 Comparison of specific activity of variants and wild-type TK at pH 4.5 and 4.75 (clarified lysate).
TK variants were incubated at pH 4.5 and 4.75 for 45 min before activity was assessed. Activities of the variants were measured in cell lysate with 50 mM HPA, 50 mM GA, and 50 mM Tris-HCl, 2.4 mM TPP, 9 mM $MgCl_2$, pH 4.5 and 4.75 at 25 ^oC in a final volume of 300 µL. Wild-type enzymes were also assessed and are labelled as (wt) in the figure. Concentration of enzymes was determined by OD at 280nm and SDS-PAGE band densitometry.

Nine final variants from Figure 5-9 were selected. Selected mutants were re-tested at pH 4.5 with the 16 min high accuracy HPLC method (section 2.2.8). This time the enzyme were purified using the bench scale method decribed in section 2.3.14 and the enzyme concentration determined at OD_{280} and also by SDS_PAGE band densitometry (section 2.3.14) to obtain the specific activity.

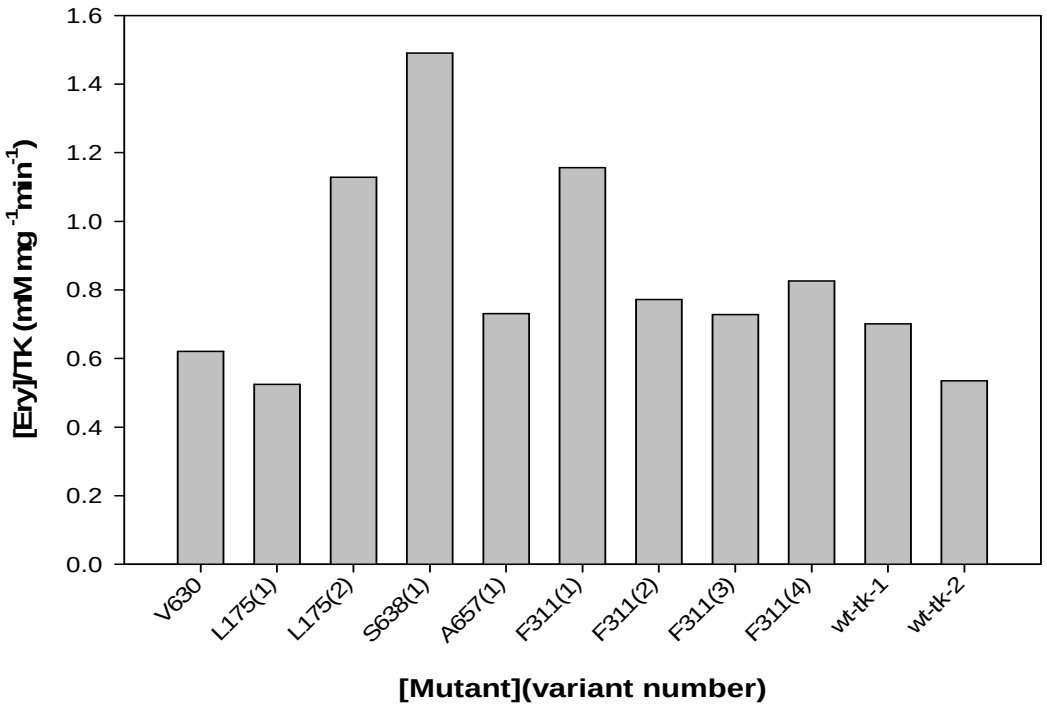
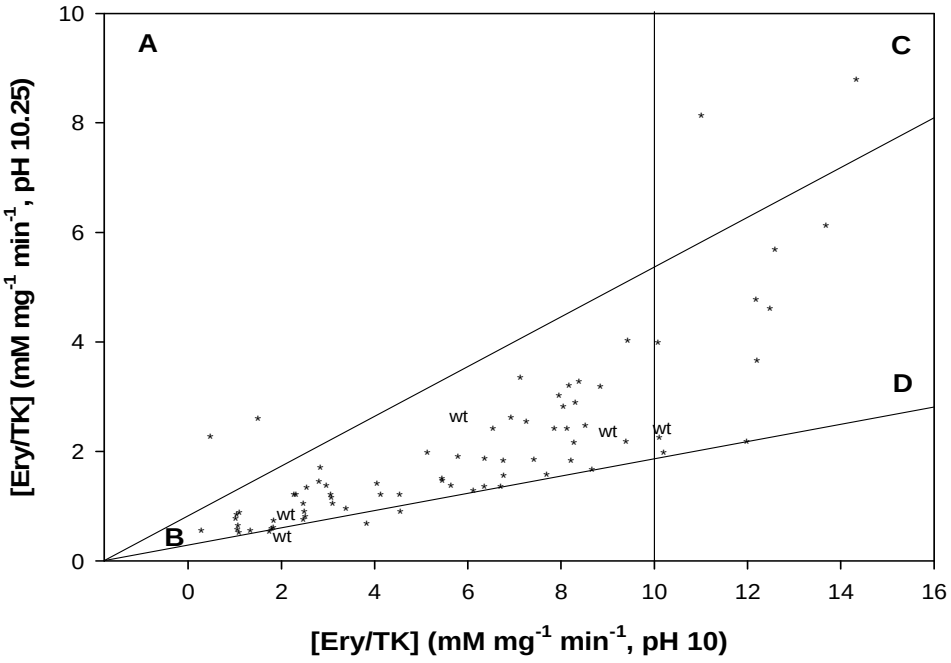


Figure 5-10 Comparison of specific activity of variants and wild-type TK at pH 4.5 (pure enzymes).
Pure TK variants were incubated at pH 4.5 for 45 min before activity was assessed. Activities of the variants were measured in with 50 mM HPA, 50 mM GA, and 50 mM Tris-HCl, 2.4 mM TPP, 9 mM $MgCl_2$, pH 4.5 and 4.75 at 25 °C in a final volume of 300 µL. For this experiment wild-type enzymes were also assessed and are labelled as (wt) in the figure. Concentration of enzymes was determined by OD at 280nm and SDS-PAGE band densitometry.

Figure 5-10 shows the final nine variants that were selected and sent to be sequenced. The variants are: V630 (1); L175 (2); S638 (1); A657 (1); F311 (4) where the number between brackets corresponds to the number of variants found and tested from that mutant. After sequencing the variants, it revealed the presence of stop codons in the first three listed above. The last two (A657, F311) turned to be wild-type TK clones.

For variants at pH ≥10, the same selection criteria were followed as described above. Variants were screened at pH 10, 10.25 and 10.5. From primary screening (data not shown) mutants that showed similar or higher activity than that wild-type were selected. Figures below shows the results obtained from the 16min high accuracy secondary screening. Clarified lysates were incubated at pH 10, 10.25 and 10.5. Residual activity at pH 10 versus 10.25 and pH 10.25 versus 10.5 were then compared (Figure 5-11). The areas within the graph (depending on the pH comparison) stand for: A) low activity at either pH 10 and pH 10.25 and high or moderate pH at pH 10.25 and 10.5; B) Low activity at both pH 10, 10.25 and 10.5, and C or D) between high and moderate activity at tested pH's. Different wild-type enzymes (wt) were used as a reference for defining the graph limits. For pH 10 screening, the variants chosen were those that fell into areas C and D, but also some that were in area A.



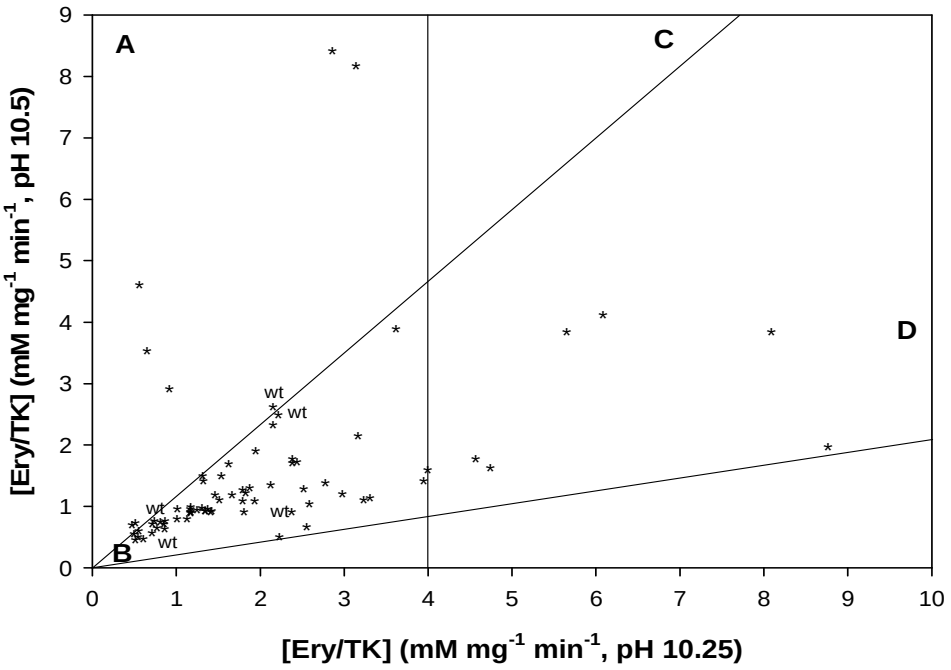


Figure 5-11 Comparison of specific activity of variants and wild-type TK at pH 10 vs. pH 10.25 and pH 10.25 vs. pH 10.5 (clarified lysate).
TK variants were incubated at pH 10, 10.25 and 10.5 for 45 min before activity was assessed. Activities of the variants were measured in cell lysate with 50 mM HPA, 50 mM GA, and 50 mM Tris-HCl, 2.4 mM TPP, 9 mM $MgCl_2$, pH 10, 10.25 and 10.5 at 25 ^{o}C in a final volume of 300 µL. Wild-type enzymes were also assessed and are labelled as (wt) in the figure. Concentration of enzymes was determined by OD at 280nm and SDS-PAGE band densitometry.

The final selected mutants for secondary screening were purified and re-tested in the same conditions as described above (i.e. pH 10 and 10.25). The concentration of enzymes was determined by OD at 280nm (section 2.3.16) and SDS-PAGE band densitometry (section 2.3.14).

Figure 5-12 shows the residual activity results from purified variants and wild-type (wt) as tested at pH 10.5. Those final variants showed higher activity than any of the wild-types used as a control.

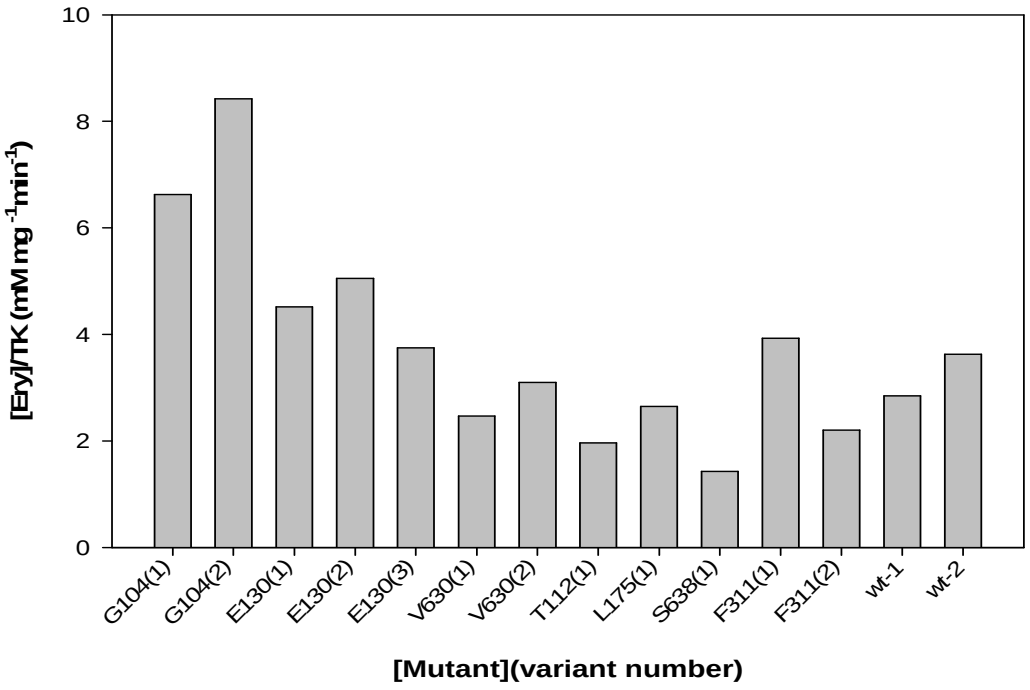


Figure 5-12 Comparison of specific activity of variants and wild-type TK at pH 10.5 (pure enzymes).
Pure TK variants were incubated at pH 10.5 for 45 min before activity was assessed. Activities of the variants were measured in with 50 mM HPA, 50 mM GA, and 50 mM Tris-HCl, 2.4 mM TPP, 9 mM $MgCl_2$, pH 10 and 10.25 at 25 °C in a final volume of 300 µL. For this experiment wild-type enzymes were also assessed and are labelled as (wt) in the figure. Concentration of enzymes was determined by OD at 280 nm and SDS-PAGE band densitometry.

The final variants that were selected and prepared to be sequenced were: G104 (2), E130 (2), and F311 (1). The results indicated that the first two variants on the list contained a stop codon and the last one was a wild-type TK clone. Therefore stop codons were present in most of the final candidates screened for increased stability.

As we can see from the results none of the variants tested turned out to be a potential mutant with increased pH or thermal stability. Even though the Pearson statistical method was applied correctly, it is commonly aimed to be used with normally distributed data (see section 5.4.4). The use of a new method that accepts non-normally distributed data is expected to have more interesting results as it was discussed before (section 5.4.4). Additionally, these two statistical methods do not consider biological aspects which are important in the engineering of the enzyme such as number of salt bridges, neighbouring amino acids, etc. so the use of some other parameters would increase the robustness of the program, and certainly the implementation of techniques such as statistical coupling analysis would in future helps us to understand in a better way the effect of a single mutation within the whole sequence, or the effect of combining the mutations that have been reported in this work .

Although there are many examples of enzymes that have been stabilised by introduction of only one or two mutations for increased thermal stability (Williams et al., 1999; Gerk et al., 2000; Bjork et al., 2004) or pH (Ness et al., 1999; Richardson et al., 2002; Boer & Koivula, 2003) and despite many successful efforts to understand the structural basis of protein stability, there is still no universal strategy to stabilise any protein by a limited number of rationally designed mutations.

5.5 Evaluation of non-parametric statistical method

5.5.1 Calculation of Kendall (τ) correlation coefficient.

For a non-parametric analysis of the data, the Kendall correlation coefficient (τ) was used, taking as a starting point the data sets from the Pearson analysis, to generate two new data matrices. This time a more rigorous treatment of data was performed (i.e. matrices were completely de-gapped so only those columns with a complete set of data were taken into account). The final matrix dimension was reduced for pH and temperature as follows:

$$X_{pH} = (n \; x \; m) = 50 \; x \; 506$$

$$X_{pH} = (n \; x \; m) = 43 \; x \; 505$$

The same level of confidence (α=0.05) was applied for the selection of the final Kendall correlation coefficient (τ), evaluated using the formula described in section 5.2.4 (Eq.5-5) for each data set. Usually the Kendall coefficient (τ) indicates the degree of correlation between two rankings and assesses its significance.

Estimated Kendall correlation coefficients for both temperature and pH are shown in Figure 5-13 and Figure 5-14 respectively. As it can be seen, the level of association between the residues and either temperature or pH, is only moderate as no values greater than 0.4 or less than -0.4 were found. However, some data turned out to be statistically significant as the tau (τ) value was higher than the selected level of confidence alpha (α=0.05).

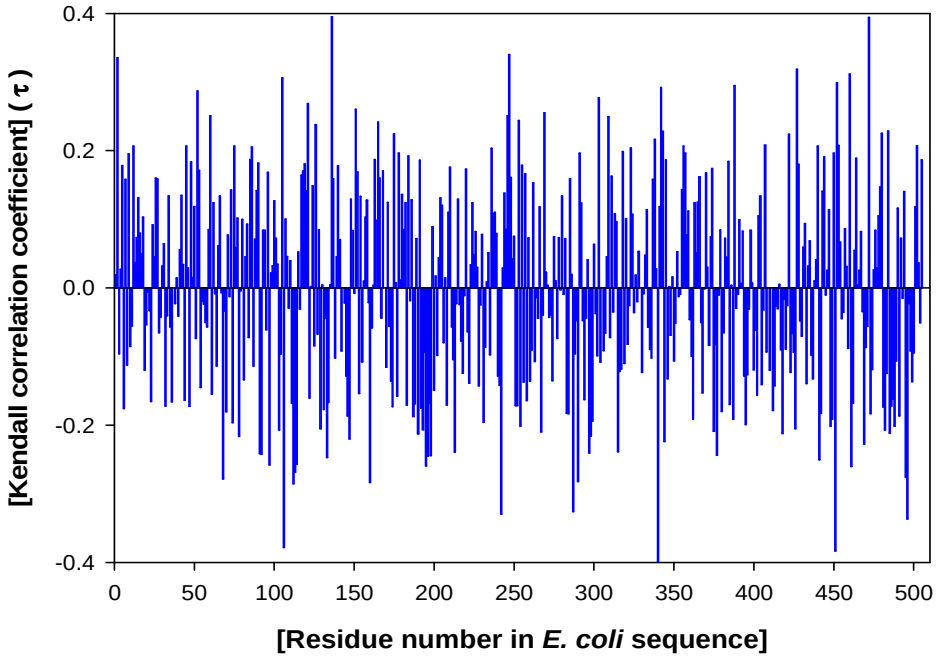


Figure 5-13 Estimated Kendall correlation coefficient (τ) for temperature matrix.

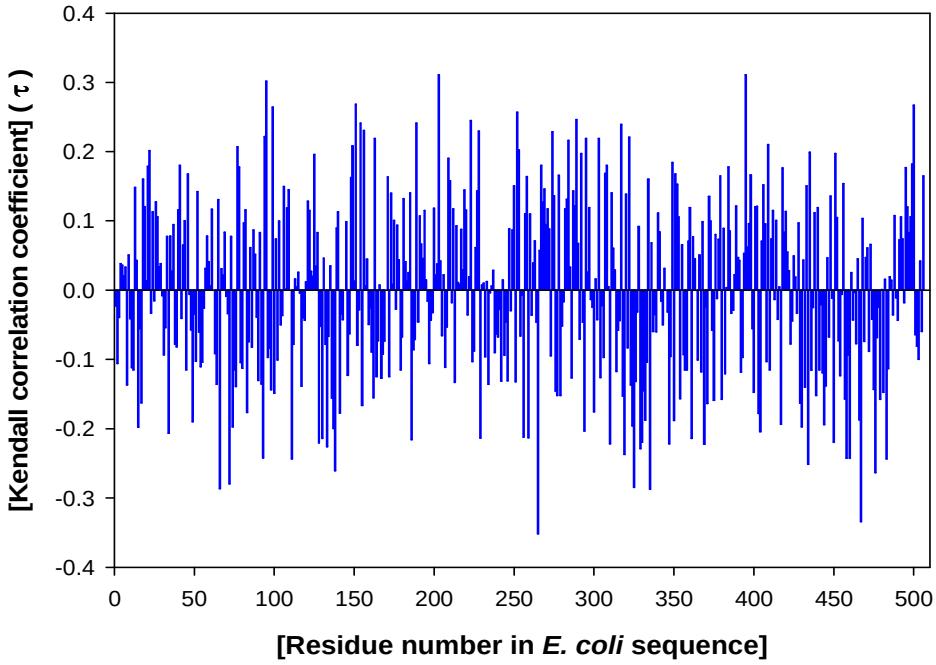


Figure 5-14 Estimated Kendall correlation coefficient (τ) for pH matrix.

A total of 17 and 8 single points for temperature and pH were selected (Table 5-5) respectively. They were selected by comparing the single Kendall's coefficient against the level of confidence. Residues selected had a *t* value greater than the level of significance (0.05) so they could be considered as statistically significant. As it was assumed that there was no association between pair of data set) (i.e. sequence residue and pH or temperature), then that hypothesis is rejected, so it can be said that they show some significant level of correlation (Kendall, 1976; Kendall & Gibbons, 1990). The final residues (candidates) for mutation are depicted in Table 5-5.

Table 5-5 Selected residues based on Kendall correlation coefficient values with a level of confidence α=0.05

A

			Residue in *E coli*			*t*-value[†]
Residue	Kendall (τ)	Position	TK	One letter	Position	(*p*-value)
23	0.239	Surface	Lys	K	Loop	0.317
132	-0.236*	Buried	Thr	T	Helix	0.533
137	-0.242*	Surface	Phe	F	Helix/Loop	0.134
153	-0.257*	Buried	Met	M	Sheet/Loop	0.679
192	-0.247*	Surface	His	H	Loop	0.050
246	0.261	Surface	Ile	I	Loop	0.249
268	-0.245*	Surface	Ala	A	Helix	0.249
270	-0.244*	Buried	Ile	I	Helix	0.095
286	-0.239*	Surface	Ile	I	Loop	0.329
372	-0.326*	Buried	Leu	L	Loop	0.207
395	0.277	Surface	Ala	A	Loop	0.606
432	-0.303*	Buried	Ser	S	Sheet	0.068
480	0.235	Buried	Ala	A	Helix	0.801
550	0.281	Surface	Pro	P	Loop	0.298
593	-0.383*	Surface	Asp	D	Loop	0.710
641	0.229	Surface	Ala	A	Helix/Loop	0.122
653	-0.337*	Surface	Asp	D	Helix	0.423

B

Residue	Kendall (τ)	Position	Residue in *E coli* TK	One letter	Position	*t*-value† (*p*-value)
192	-0.226*	Surface	His	H	Loop	0.121
328	0.308	Buried	Arg	R	Helix	0.078
368	0.265	Buried	Phe	F	Helix	0.805
384	-0.312*	Surface	Pro	P	Helix	0.952
428	-0.229	Buried	Leu	L	Sheet/Loop	0.228
521	-0.283*	Surface	Gln	Q	Loop	0.734
544	-0.241*	Surface	Lys	K	Sheet/Loop	0.200
644	0.252	Buried	Leu	L	Helix	0.450

Table A and B represent the selected residues aimed to increase the temperature and pH stability of the enzyme respectively.
Residues were selected with a level of significance where their *t* values were greater than α=0.05.
† The absolute value of negative *t* was taken for analysis purposes
* Negative Kendall values represent a negative association

A visual analysis of the residue column revealed three different types of "association". Association 1: represents two vertical clusters of data or highly conserved residue. Association 2: three vertical clusters or semi-conserved residues. Association 3: scattered set of data. For the three types of association the Kendall's coefficient was evaluated against the level of confidence alpha=0.05. Data that failed to pass the hypothesis test were no longer considered (i.e. *p*-value <0.05). In general, most of the residues (columns) that fell into the Association 3 category failed this test.

The most popular correlation is the Pearson correlation coefficient (r). However, Pearson's (r) was found not to be appropriate for the correlation of either pH or temperature (i.e. dependent variable) with their corresponding residue or set of data (i.e. independent variable). The pairs, X and Y in this work were found to frequently violate the statistical assumption in a combination of ways. For example, the presence of outliers or non-constant variance over the range of scores can destabilise these classical statistical methods and make findings difficult to replicate or interpret (Donna, 1986).

There are two reasons for reporting a correlation, statistical testing and assessing the degree of relationship. When testing for a relationship, (for example temperature or pH with mutagenic variation at each residue), issues of type I error, which is the error of rejecting a null hypothesis when it is actually true or correct, rates and statistical power predominate. In many cases the two correlations provide a similar protection for type I error. However, Pearson's (*r*) is unstable and behaves poorly when faced with pairs of data that significantly violate its statistical assumption (i.e. the normal distribution). Kendall's correlation is more consistent and offers better protection against type I error. Kendall's correlation also sometimes errors on the conservative side. A common belief is that a non-parametric test is less powerful than an appropriate parametric test such as the Pearson correlation. Unfortunately, none of the assessed pairs of data appeared normally distributed so the advantage of using the Pearson correlation may be undermined.

The choices are clearer when the purpose of calculating the correlation is to estimate the degree of relationship. Kendall's tau is better from a statistical standpoint as it is unbiased. On the other hand, Pearson's coefficient *r* is easily generalised to multivariate situations (e.g. multiple regression, partial correlation). While, the square of the Pearson's correlation indicates the amount of variance that overlaps between two variables, this only makes sense when the variability of the sample is constant (Dytham, 1999). We saw in this analysis that a non-normal distribution was usually present so the interpretation of Pearson's *r*-value is problematic when rating the degree of relationship.

Kendall's coefficient has a simple meaning. In general, it is the probability that any pair of observations will have the same ordering i.e. a high or low level of association depending on the Kendall's value (-1 to 1). When simple correlations are required Kendall's tau possesses several advantages. It appears to maintain adequate protection against type I errors even under severe conditions. It also provides a more relevant value for expressing the level of association.

In all cases, the residues identified in this analysis were made without using knowledge of the structural properties of the enzyme. It is well known that most homology-based engineering attempts frequently lead to the substitution of rigidly conserved amino acids that do not affect the desired properties and some times render the enzyme inactive because of changes in protein conformation. Such methods might have overlooked key amino acid residues (Chen, 2001), whereas our approach avoids the simple assumption that high conservation is linked to the desired properties. An interesting point is that most of the selected residues were far from the more highly conserved active site. Some authors have highlighted that when trying to increase the stability of an enzyme in terms of temperature, for example, then residues that lie far from active site are preferred (Morley & Kazlauskas, 2005).

5.5.2 Principal component analysis and Partial least square

Since the operation of PCA is often thought of as revealing the internal structure of the data in a way which best explains the variance within them (Härdle, 2007), then the assumption for this work is that large variances might contain important information and that low variances are the result of inconsistent data. In other words, PCA is used to look for outliers. Additionally, as PCA is a non-parametric analysis then the answer or outcome can be regarded as unique and independent of any hypothesis about the data probability distribution (Jolliffe, 2002). Figure 5-15 shows the results obtained from PCA of the pH matrix. It is a plot of the first two columns of scores that shows the ratings data projected onto the first two principal components. Each data point represents a single residue within the protein sequence alignment. Scores were usually mean centred to minimise the mean square error of approximating the data (Miranda et al., 2008)

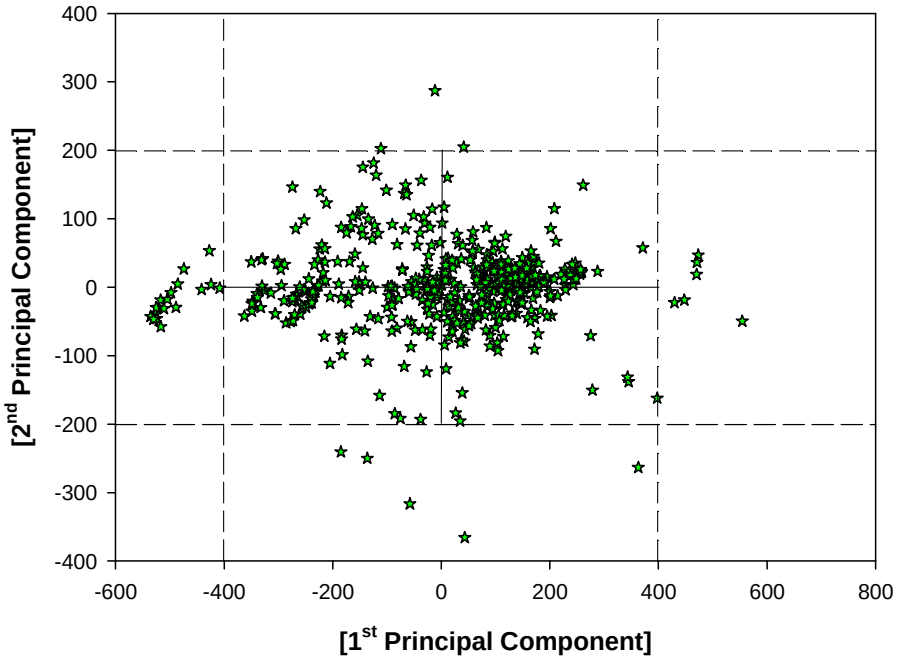


Figure 5-15 First and second principal component estimated using PCA for the pH matrix.

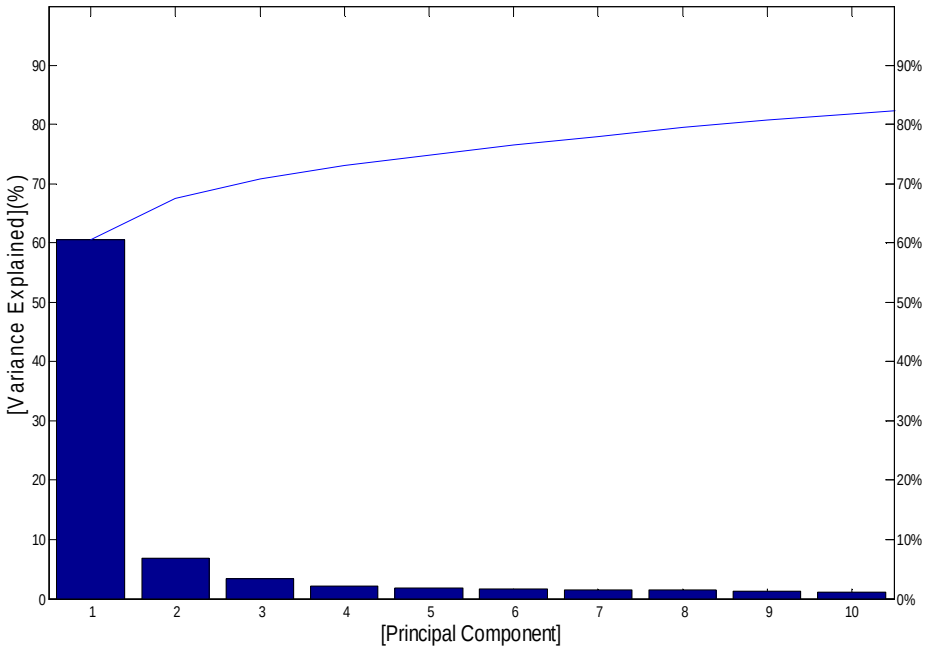


Figure 5-16 Screed plot of the percent variability explained by each principal component for pH. The line represents the cumulative variance explained

Figure 5-16 shows that the only clear difference in the amount of variance accounted for by each component is between the first and second components. The first component by itself explains about 60% of the variance.

Similarly, the two principal components obtained from the temperature matrix were plotted (Figure 5-17). Following the same criteria as above, the points that present the highest variation with respect to the mean were selected (i.e. outliers).

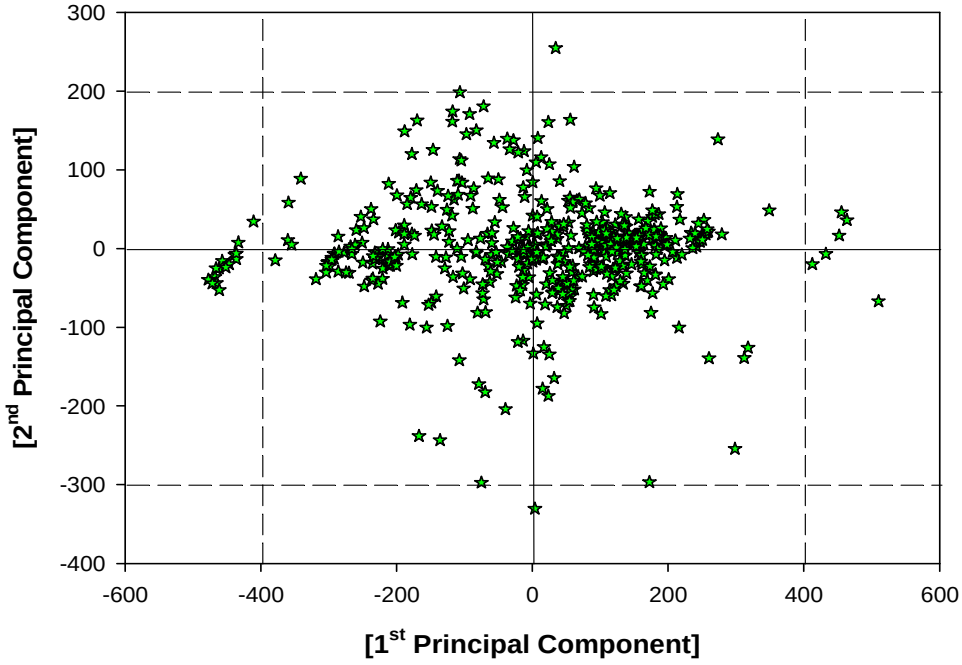


Figure 5-17 First and second principal component as estimated using PCA for temperature matrix.

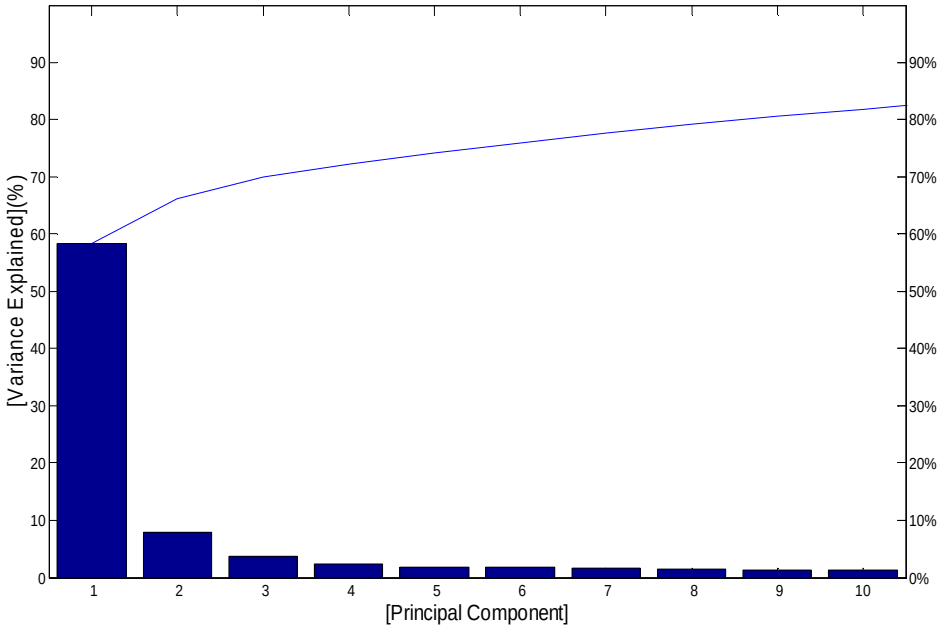


Figure 5-18 Screed plot of the percent variability explained by each principal component for temperature.

Figure 5-18 shows the amount of variance accounted for by each component is between the first and the second component.

It can be concluded that in both screed plots, for pH and temperature, the first two principal components explain more than two thirds of the total variability in the standardized ratings. Consequently, using only these two components is a reasonable way to reduce the dimensions in order to visualize the data from the principal component analysis (Lay, 2000; Miranda et al., 2008).

Typically, the data set was reduced after two rounds of applying PCA. In the first round some of the outliers were removed. Those outliers were found to have in common a very high level of dispersion, i.e. for a single position some of the residues showed more than 10 of the 20 natural amino acids, representing no distinct preferences. The second round allowed us to reduce the data set further such that outliers were now found to have a moderate level of dispersion. A final third round was performed, but after analysing them they were found to have a trend in terms of variation or distribution of data within a single residue (column) position so it was decided to stop removing data points.

The final selected points were those that were outside the dashed lines in each plot, and are shown in Table 5-6.

Table 5-6 Selected points from temperature and pH data set by using Principal Component Analysis (PCA).

PCA	
temperature	pH
Met33	Ala37
Ser68	Ser68
Gly104	Gly104
His142	Leu173
Leu173	Tyr182
Ala180	Lys244
Tyr182	Pro264
Ala207	Pro283
Asp220	Gln301
Lys244	Met329
Ala323	Ile345
Ala346	Ala346
Ala356	Glu398
Ser393	Ala444
Ala416	Ala448
Thr431	Gln526
Ala448	Ala556
Ala515	Ser584
Ala524	Pro602
Gln526	Ala610
Ala564	Ala639
Ala567	
Thr572	

Selected points for temperature and pH from Principal Component analysis (PCA). Those points represent the outliers obtained after the third round.

The intention, in using PCA, was to extract the underlying effects in the X data (i.e. single residue column). In this way, we could guarantee that only independent effects were used, and that low-variance noise effects were excluded. This improved the quality of the model significantly. However, PCA still has a problem, if the underlying effects relevant to temperature and pH optima are small in comparison with other irrelevant ones, then they would not appear among the first few principal components. So, we are still left with a component selection problem where we cannot just include the first two principal components, as these may serve to degrade the performance of the model. Instead, we have to extract all components, and determine whether adding each one of these improves the model. This is a complex problem that can be solved using Partial Least Squares Regression (PLSR).

The PLS algorithm used examines both X (i.e. residue position) and Y (i.e. temperature or pH column) data and extracts components (now called factors), which are directly relevant to both sets of variables. These are extracted in decreasing order of relevance. It is then necessary to extract the smallest number of factors that adequately model the relevant underlying effects.

Figure 5-19 and Figure 5-20 show the result after applying the PLS regression. The regression factor that belongs to the model constructed by PLS for both pH and temperature is plotted versus the residue position within the TK sequence. Those outliers that were above or below the dashed line were selected in both cases. However, caution has to be taken as those residues can not always be considered as a single/isolated residue as the obtained value is highly affected for some of the values (residues) that are either to the right or the left side of it.

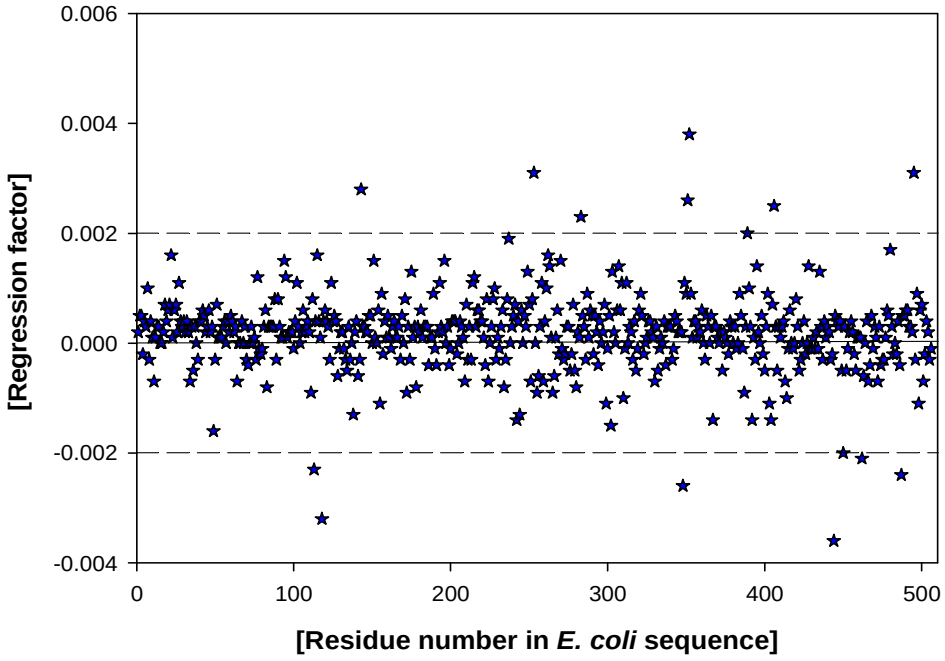


Figure 5-19 Regression function values calculated by using PLS for pH matrix.

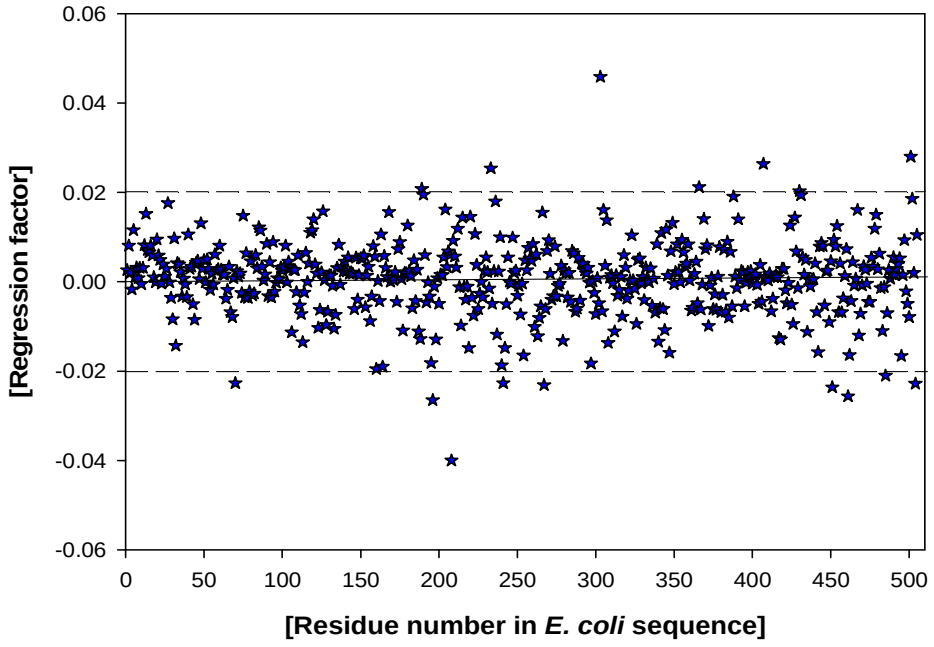


Figure 5-20 Regression function values calculated by using PLS for temperature matrix.

Additionally, as it can be seen from Figure 5-21 and Figure 5-22 the model in either case turned out to be very accurate as when the estimated pHs and temperatures (section 5.4.2) were compared with those estimated by PLS they were found to be very similar to each other.

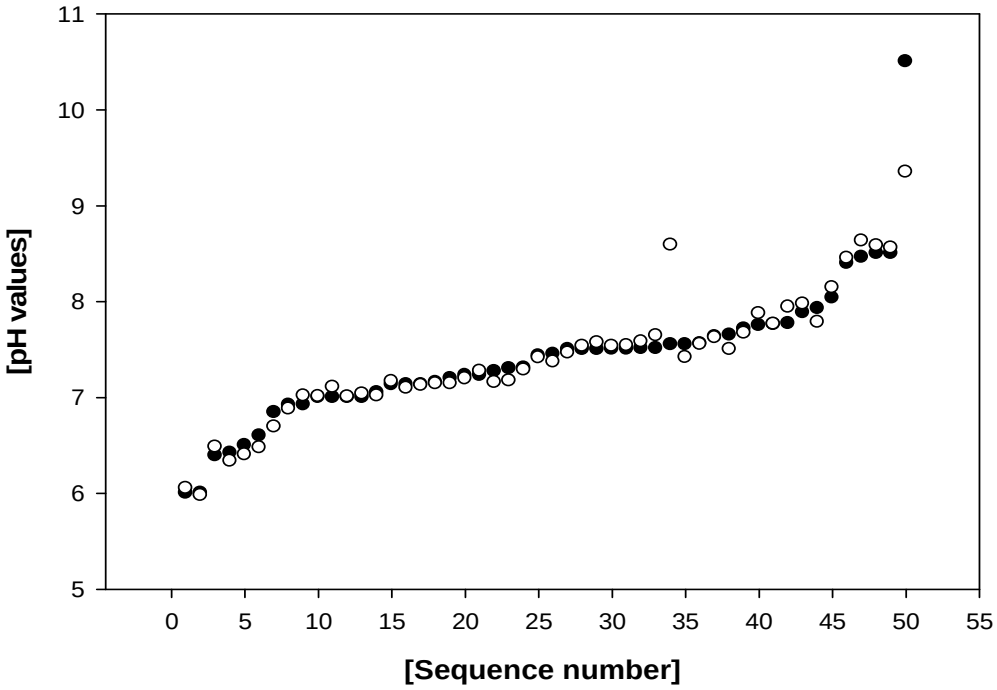


Figure 5-21 Predicted pH optimum values using the PLS model.
(●) Estimated pH values using a frequency distributed analysis, and (○) the pH predicted by PLS model. The two methods correlate with an r^2=0.9789.

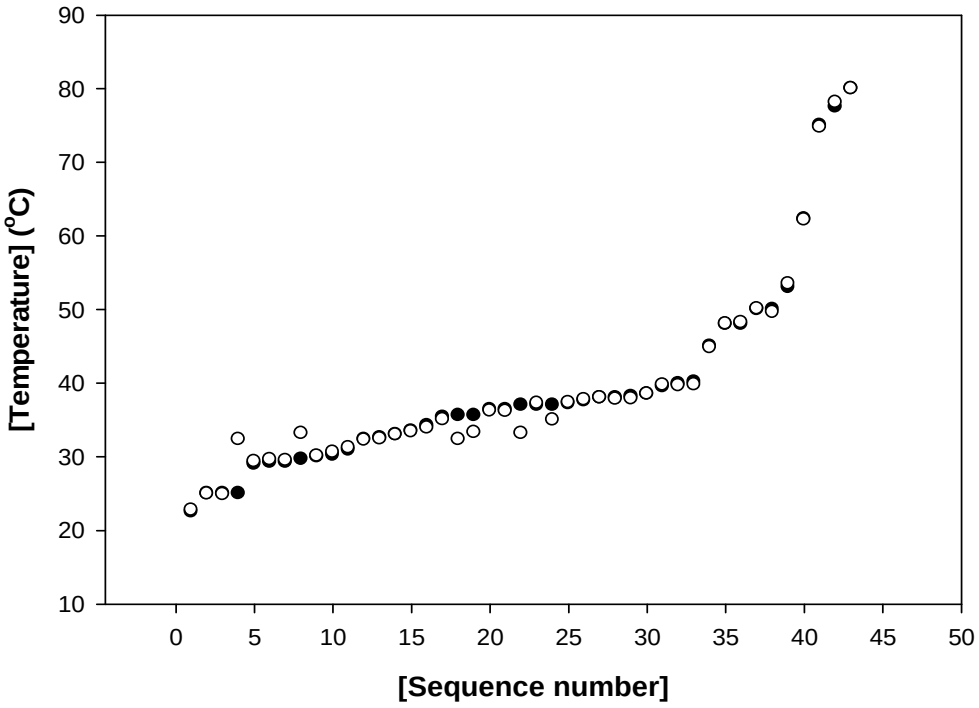


Figure 5-22 Predicted temperature values using PLS model.
(●) Estimated temperature values using a frequency distributed analysis and (○) the temperature predicted by PLS model. The two methods correlate with an r^2=0.9889.

Finally, the number of latent variables used for temperature and pH matrices was 5. Increasing the number of latent variables in the PLS model would lead us to have a more inaccurate model (Agnar, 1988). The results of the PLS analysis were evaluated in terms of goodness of fit (r^2=0.9899), analogous to Pearson correlation coefficient (Johan et al., 1998). The final selected residues for pH and temperature are listed in Table 5-7.

Table 5-7 Selected points from temperature and pH data set by using Partial Least Square (PLS).

PLS	
temperature	pH
Arg42	Asp81
Gln50	Leu169
Thr78	Leu173
Tyr182	Met203
Lys225	Ala317
Glu229	Arg328
Lys280	Met333
Glu309	Ala339
Ala312	Ser361
Lys316	Ala367
Ala317	Ser393
Lys330	Met451
Leu389	Gln455
Met451	Ser585
Gln453	Lys591
Ala480	Lys603
Val507	Leu643
Ala537	Thr651
Glu569	
Asp593	
Lys603	
Val609	
Glu642	
Lys658	
Ala659	
Glu661	

Selected points for temperature and pH from Partial Least Square (PLS) analysis. Those points represent the outliers obtained. The data set was first filtered by using the results of PCA.

The primary difference between PCA and PLS is that PCA is performed on one data matrix (e.g., X or Y) and PLS evaluates both (X and Y) simultaneously to both develop a predictive model (e.g., predict Y from X) and to evaluate relationships between specific X and Y variables. PCA is first used to identify characteristics of data in either the X- or Y-matrix. The principal outcome of this analysis is the identification of data with the highest covariance and thus are assumed to have a systematic relationship, whereas PLS was used to predict temperature or pH from the *E. coli* transketolase alignment.

The analysis of large data sets with many observations and many variables is difficult. The real difficulties with large data sets are still probably hidden, since the known problems still are based mainly on extrapolations from small data sets. Still, both methods seemed to be quite robust in terms of handling non-normally distributed data. In conclusion, PCA was used to filter out noisy data. The new filtered sets of data for either case pH or temperature were used in the PLS analysis.

Three of the four methods used in this work, namely Pearson, Kendall and PCA, account just for a single mutation, without looking at the effect that it might produce within the whole sequence. In other words, an amino acid change at one position may have a functional consequence of changing other amino acids in the protein. Therefore the only predictive sequence-function method used that accounts for this issue is PLS as the selections result from analysing the contribution of neighbouring and even distant residues within the same sequence.

The total number of selected residues from all the statistical analyses can only practically be tested as single mutations. However, it would be interesting to see how much the single mutations alter the stability of the enzyme compared to a combination of them. To date, there are a number of computational methods that could be used (Jones, 1994; Desjarlais & Handel, 1995; Dahiyat & Mayo, 1996). Statistical coupling analysis (SCA) (Lockless &

Ranganathan, 1999) may be one approach to use in future as it quantifies how much the amino acid distribution at position *i* changes upon a perturbation of the amino acid distribution at another position j. For this analysis the resulting statistical coupling energy indicates the degree of evolutionary dependence between the residues, with higher coupling energy corresponding to increased dependence (Lockless & Ranganathan, 1999). This analysis has been proved to be successful for the PDZ domain families and also for transketolase from *E. coli* (unpublished data- Stratford, J., 2009).

To test the theoretical results, the next stage was the creation of mutant libraries based on the results obtained from all the statistical analyses. Mutants were created in the same order as those statistical analyses were performed. First, mutants from Pearson analysis will be created and screened, then the ones from Kendall, PCA and finally for PLS.

5.6 Conclusions

The use of bioinformatics was an essential tool as most of the information used in this work was gathered from some of the data bases that are focused on proteins for instance BRENDA and NCBI. Without the use of such data bases the statistical analyses (i.e. estimation of optimum pH or temperature for some proteins) would not have been possible and as a consequence the number of sequences used would have been much lower. To take advantage of all that information statistical methods were used to try to find a relationship between protein sequence and optimum temperature. The first two methods used can give an idea of how well a set of data are correlated to each other. These were the Pearson coefficient(r) for use with parametric data and the Kendall's coefficient (τ), a less common but more powerful analysis for non-parametric data

Using the Pearson coefficient method to predict the sites to mutate for either increased temperature or pH, did not prove to be as robust as it had been first thought. Pearson's coefficient is the method of choice for data that are normally distributed. The Kendall's coefficient accounts for non-normal distributions and so may in future result in a better identification of relevant residues.

Two more statistical approaches were also preliminarily applied in this work, PCA and PLS. The first one showed to be very powerful in terms of reducing noisy data. However, it does not look for any relationship between for example a dependant (Y) and an independent (X) variable as it performs a single analysis for either X or Y. By contrast, PLS can include both variables X and Y and predict a model that better explains the data by using a multiple regression and then validating the model with an r^2 (similar to that from Pearson) correlation to the raw data. PCA can be used as a filter before the data used in the PLS in order to increase the prediction model more accurately.

In this work only the set of variants revealed from the Pearson's analysis were tested. None of them gave improvements in either pH or temperature stability. A reason of that result might be attributed to the number of data evaluated for example even though the pH and temperature matrices met the criteria of containing more than 40 data, it is very likely that a larger set of data could have improved the analysis .

The results obtained using Kendall's coefficient, PCA and PLS methods were not investigated further due to a lack of time. It will be interesting to see in future whether those correlations results proved to be more significant experimentally. However, it is encouraged that for the next set of variants obtained by the non-parametric methods such as Kendall, PLS and PCA, the use of protein modelling software such as Discovery Studio® is used to estimate the changes in stability upon mutation as a preliminary method to eliminate some of the selected mutants that might not have an impact on stability before moving on to the library creation.

Chapter 6

6 Future recommendations

Following the conclusions from each chapter in this thesis, there are some directions in which the project may be taken for a better understanding of the mechanism of deactivation that occurs for TK under harsh bioprocess conditions.

First, with respect to thermal stability, the distinction between thermal tolerance and real thermal stability is important. The first term refers to the ability to withstand incubation at elevated temperatures, without necessarily being active at those temperatures. Real thermal stability refers to enzymes that not only withstand elevated temperatures, but that also retain activity at these temperatures. The approach used for thermal stability analysis in Chapter 4 did not measure the activity of TK directly at elevated temperatures but rather the propensity of the enzyme to either refold or retain the native structure at elevated temperature. Clearly, these two types of properties need to be assessed with different methods. The next experiments may be addressed for a more formal evaluation of thermal stability and inactivation of the enzyme, for example, carrying out some time-dependent thermo-inactivation experiments through a temperature range from 30 to 75 ^{o}C both with and without cofactors (i.e. TPP and Mg^{2+}). In addition, to track any changes in the size, and potential aggregation of the enzyme, dynamic light scattering may be helpful.

In the case of pH, a titration curve of TK using NMR may give a better understanding of the effect of pH changes to pKa values of ionisable residue groups. These experiments could be performed on both apo- and holo-TK for assessing the effect of cofactors on the enzyme. These experiments can be supported by dynamic light scattering and size exclusion chromatography to

track changes in size and different enzyme conformations before and after aggregation.

During the stability studies carried out in Chapter 4 (i.e. thermal and pH stability), it was suggested that TK tends to form aggregates when exposed to extreme temperatures ≥ 65^{o}C or pH (i.e. 4 or 11). That aggregation is most likely to arise from association of specific partially unfolded intermediates. However, an important consequence of this is that aggregation seems to be favoured by temperatures or pHs that increase the population of these intermediates, and hence it is the properties of these intermediates that are important in determining how aggregation occurs. Furthermore, the characteristics and properties of the intermediates may be significantly different from those of the native (and unfolded) conformation. Some different approaches such as dynamic light scattering, size-exclusion chromatography, stopped-flow CD, small-angle X ray scattering and, for example, for a more formal and deep aggregation analysis using Fourier transform spectroscopy (FTIR) experiments could be performed. These methods are likely to provide an explanation in more detail about the aggregation process that TK suffers from when exposed to extreme pH and temperature conditions.

In Chapter 5 a different set of potential mutation points were identified within *E. coli* transketolase by using different statistical approaches such as partial least square, principal component analysis and Kendall's correlation coefficient. These targets are aimed to enhance the thermal or pH stability of transketolase. Among all these selected mutation points, it is likely that some are "real" potential candidates that will influence the stability of the enzyme, whereas other may not have an important effect. It is recommended that before applying any mutation, the selected points are evaluated computationally. For example, DiscoveryStudio® can be used to assess potential positive or negative changes in ΔG caused by each single mutation. This software would allow us to reduce the number of "non-effective" mutations being screened for stability. A second option would be to assess the mutations using statistical coupling

analysis (SCA) to quantify how much the selected residue at each position changes upon the perturbation of the residue distributions at other positions. This would determine the extent to which selected sites require simultaneous mutations at other coupled sites.

Finally if none of the selected points turned out to be a potential candidate for enhancing the stability of the enzyme after being assessed by the two software previously mentioned, then it may be recommendable to follow a different *in vitro* evolutionary approach such as combining rational design and directed evolution, along with a more detailed structural study of the enzyme.

References

Agnar,H. (1988). PLS regression methods. *SO: Journal of Chemometrics*, 2(3), 211-228.

Ahern,T.J., Casal,J.I., Petsko,G.A., & Klibanov,A.M. (1987). Control of oligomeric enzyme thermostability by protein engineering. *Proc Natl.Acad.Sci U.S.A*, 84(3), 675-679.

Alber,T., Dao-pin,S., Wilson,K., Wozniak,J.A., Cook,S.P., & Matthews,B.W. (1987). Contributions of hydrogen bonds of Thr 157 to the thermodynamic stability of phage T4 lysozyme. *Nature*, 330(6143), 41-46.

Alexandre,F.R., Pantaleone,D.P., Taylor,P.P., Fotheringham,I.G., Ager,D.J., & Turner,N.J. (2002). Amine-boranes: effective reducing agents for the deracemisation of DL-amino acids using L-amino acid oxidase from Proteus myxofaciens. *Tetrahedron Letters*, 43(4), 707-710.

Allen,D. & Pielak,J. (1998). Baseline length and automated fitting of denaturation data. *Protein Science*, 7(5), 1262-1263.

Anderson,D.E., Becktel,W.J., & 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.

Andre,C., Demuynck,C., Gefflaut,T., Guerard,C., Hecquet,L., Lemaire,M., & Bolte,J. (1998). Fructose-1,6-bisphosphate aldolase and transketolase: Complementary tools for the *de novo* syntheses of monosaccharides and analogues. *J.Mol.Catal.B: Enzym.*, 5(1-4), 113-118.

Anscombe,F.J. (1973). Graphs in statistical analysis. *American Statistician*, 27, 17-21.

Antosiewicz,J., McCammon,J.A., & Gilson,M.K. (1994). Prediction of pH-dependent Properties of Proteins. *Journal of Molecular Biology*, 238(3), 415-436.

Arnold,F.H. (1998). Design by directed evolution. *Acc.Chem.Res.*, 31, 125-131.

Arnold,F.H. (2001). Combinatorial and computational challenges for biocatalyst design. *Nature*, 409(6817), 253-257.

Arnold,F.H., Wintrode,P.L., Miyazaki,K., & Gershenson,A. (2001). How enzymes adapt: lessons from directed evolution. *Trends Biochem.Sci*, 26(2), 100-106.

Arnott,M.A., Michael,R.A., Thompson,C.R., Hough,D.W., & Danson,M.J. (2000). Thermostability and Thermoactivity of Citrate Synthases from the Thermophilic and Hyperthermophilic Archaea, Thermoplasma acidophilum and Pyrococcus furiosus. *Journal of Molecular Biology*, 304(4), 657-668.

Aucamp, J. P. Enginnering a new high-throughput screening for direct evolution of enzyme stability. University College London-Thesis. 2005.
Ref Type: Generic

Aucamp,J.P., Cosme,A.M., Lye,G.J., & Dalby,P.A. (2005). High-throughput measurement of protein stability in microtiter plates. *Biotechnology and Bioengineering*, 89(5), 599-607.

Aurora,R. & Rose,G.D. (1998). Helix capping. *Protein Science*, 7(1), 21-38.

Baum,J., Dobson,C.M., Evans,P.A., & Hanley,C. (1989). Characterization of a partly folded protein by NMR methods: studies on the molten globule state of guinea pig .alpha.-lactalbumin. *Biochemistry*, 28(1), 7-13.

Bjork,A., Dalhus,B., Mantzilas,D., Sirevag,R., & Eijsink,V.G. (2004). Large improvement in the thermal stability of a tetrameric malate dehydrogenase by single point mutations at the dimer-dimer interface. *J Mol Biol*, 341(5), 1215-1226.

Bloom,J.D., Wilke,C.O., Arnold,F.H., & Adami,C. (2004). Stability and the Evolvability of Function in a Model Protein. *Biophysical Journal*, 86(5), 2758-2764.

Boer,H. & Koivula,A. (2003). The relationship between thermal stability and pH optimum studied with wild-type and mutant Trichoderma reesei cellobiohydrolase Cel7A. *Eur.J Biochem.*, 270(5), 841-848.

Bolte,J., Demuynck,C., & Samaki,H. (1987). Utilization of enzymes in organic chemistry: Transketolase catalysed synthesis of ketoses. *Tetrahedron Lett.*, 28(45), 5525-5528.

Bomma,A.S. & Riebel,B.R. (2004). *Biocatalysis: Funadamentals and applications*. USA: Wiley-VCH.

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

Borman,S. (2000). Proteomics: taking over where genomics leaves off. *Chemical and Engineering News*, 31-37.

Brocklebank,S., Woodley,J.M., & Lilly,M.D. (1999). Immobilised transketolase for carbon-carbon bond synthesis: biocatalyst stability. *J.Molec.Catal.B: Enzymatic*, 7, 223-231.

Brocklehurst,K. (1994). A sound basis for pH-dependent kinetic studies on enzymes. *Protein Engineering Design and Selection*, 7(3), 291-299.

Buckland,B.C., Robinson,D.K., & Chartrain,M. (2000). Biocatalysis for pharmaceuticals--status and prospects for a key technology. *Metab Eng*, 2(1), 42-48.

Bushueva,T.L. & Tonevitsky,A.G. (1987). The effect of pH on the conformation and stability of the structure of plant toxin -- ricin. *FEBS Letters*, 215(1), 155-159.

Cadogan,A. & Sutton,R. (1994). *Maths for Advenced Biology*. London: Nelson co.

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

Casari,G., Sander,C., & Valencia,A. (1995). A method to predict functional residues in proteins. *Nat Struct Biol*, 2(2), 171-178.

Chaffotte,A., Guillou,Y., Delepierre,M., Hinz,H.J., & Goldberg,M.E. (1991). The isolated C-terminal (F2) fragment of the Escherichia coli tryptophan synthase .beta.2-subunit folds into a stable, organized nonnative conformation. *Biochemistry*, 30(32), 8067-8074.

Chaplin,M.F. & Bucke,C. (1990). *Fundamentals of enzyme kinetics*. London,UK.: Cambridge University Press.

Chelliah,V., Chen,L., Blundell,T.L., & Lovell,S.C. (2004). Distinguishing Structural and Functional Restraints in Evolution in Order to Identify Interaction Sites. *Journal of Molecular Biology*, 342(5), 1487-1504.

Chen,R. (2001). Enzyme engineering: rational redesign versus directed evolution. *Trends Biotechnol.*, 19(1), 13-14.

Chen,Y.H., Yang,J.T., & Martinez,H.M. (1972). Determination of the secondary structures of proteins by circular dichroism and optical rotatory dispersion. *Biochemistry*, 11(22), 4120-4131.

Cirino,P.C. & Georgescu,R. (2003). Screening for Thermostability. *Directed Enzyme Evolution* (pp. 117-125).

Clarke,J. & 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.

Clementi,M., Clementi,S., Cruciani,G., Pastor,M., Davis,A.M., & Flower,D.R. (1997). Robust multivariate statistics and the prediction of protein secondary structure content. *Protein Eng*, 10(7), 747-749.

Cohen,J.C.P.W.S.G.&.A.L.S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ.: Lawrence Erlbaum Associates.

Costelloe,S., Ward,J., & Dalby,P. (2008). Evolutionary Analysis of the TPP-Dependent Enzyme Family. *Journal of Molecular Evolution*, 66(1), 36-49.

D'Amico,S., Gerday,C., & Feller,G. (2003). Temperature Adaptation of Proteins: Engineering Mesophilic-like Activity and Stability in a Cold-adapted [alpha]-Amylase. *Journal of Molecular Biology*, 332(5), 981-988.

Daggett,V. & Levitt,M. (1993). Protein unfolding pathways explored through molecular dynamics simulations. *J Mol Biol*, 232(2), 600-619.

Dahiyat,B. & Mayo,S. (1996). Protein design automation. *Protein Science*, 5(5), 895-903.

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

Dalby,P.A., Aucamp,J.P., George,R., & Martinez-Torres,R.J. (2007). Structural stability of an enzyme biocatalyst. *Biochemical Society Transactions*, 035(6), 1606-1609.

Daniel,R.M., Danson,M.J., Eisenthal,R., Lee,C.K., & Peterson,M.E. (2008). The effect of temperature on enzyme activity: new insights and their implications. *Extremophiles.*, 12(1), 51-59.

Daniel,R.M., Danson,M.J., & Eisenthal,R. (2001). The temperature optima of enzymes: a new perspective on an old phenomenon. *Trends in Biochemical Sciences*, 26(4), 223-225.

Danson,M.J., Hough,D.W., Russell,R.J.M., Taylor,G.L., & Pearl,L. (1996). Enzyme thermostability and thermoactivity. *Protein Engineering Design and Selection*, 9(8), 629-630.

Dasgupta,S. & Bell,J.A. (1993). Design of helix ends. Amino acid preferences, hydrogen bonding and electrostatic interactions. *Int.J Pept.Protein Res*, 41(5), 499-511.

Datta,A.G. & Racker,E. (1961). Mechanism of Action of Transketolase .1. Properoperties of Crystalline Yeast Enzyme. *J.Biol.Chem.*, 236(3), 617-623.

Davies,M., Toseland,C., Moss,D., & Flower,D. (2006). Benchmarking pKa prediction. *BMC Biochemistry*, 7(1), 18.

De Bernardez Clark,E. (1998). Refolding of recombinant proteins. *Current Opinion in Biotechnology*, 9(2), 157-163.

de la Haba,G., Leder,I.G., & Racker,E. (1955). Crystalline transketolase from bakers' yeast: isolation and properties. *J.Biol.Chem.*, 214, 409-426.

Debnath,D.K., Mukhopadhyay,K., & Basak,S. (2005). Acid-induced denaturation and refolding of prothrombin. *Biophysical Chemistry*, 116(2), 159-165.

del Sol Mesa,A., Pazos,F., & Valencia,A. (2003). Automatic Methods for Predicting Functionally Important Residues. *Journal of Molecular Biology*, 326(4), 1289-1302.

DeLano, W. L. The PyMOL Molecular Graphics System. 2002. DeLano Scientific, San Carlos, CA, USA.
Ref Type: Computer Program

Delphine,C.B., David,M.R., & Jan,H.J. (2008). Very fast prediction and rationalization of pKa values for protein-ligand complexes. *SO: Proteins: Structure, Function, and Bioinformatics*, 9999(9999), NA.

Desjarlais,J.R. & Handel,T.M. (1995). De novo design of the hydrophobic cores of proteins. *Protein Science*, 4(10), 2006-2018.

Deu,E. & Kirsch,J.F. (2007). Cofactor-Directed Reversible Denaturation Pathways: The Cofactor-Stabilized *Escherichia coli* Aspartate Aminotransferase Homodimer Unfolds through a Pathway That Differs from That of the Apoenzyme. *Biochem.*, in press.

Di,F., V, Garnier,J., & Munson,P.J. (1996). Improving protein secondary structure prediction with aligned homologous sequences. *Protein Science*, 5(1), 106-113.

Donna, S. Elementary Statistics. Chapter 1. 1986. USA, Addison-Wesley. Ref Type: Generic

Draths,K.M., Pompliano,D.L., Conley,D.L., Frost,J.W., Berry,A., Disbrow,G.L., Staversky,R.J., & Lievense,J.C. (1992). Biocatalytic Synthesis of Aromatics from D-Glucose - the Role of Transketolase. *J.Amer.Chem.Soc.*, 114(10), 3956-3962.

Dubey,V.K. & Jagannadham,M.V. (2003). Differences in the Unfolding of Procerain Induced by pH, Guanidine Hydrochloride, Urea, and Temperature. *Biochemistry*, 42(42), 12287-12297.

Dytham,C. (1999). *Choosing and using statistics. A biologist's guide*. Oxford: Blackwell Science.

Egan,R.M. & Sable,H.Z. (1981). Transketolase kinetics - the slow reconstitution of the holoenzyme is due to rate-limiting dimerization of the subunits. *J.Biol.Chem.*, 256(10), 4877-4883.

Eijsink,V.G.H., Bj°rk,A., Gσseidnes,S., Sirevσg,R., Synstad,B.r., Burg,B.v.d., & Vriend,G. (2004). Rational engineering of enzyme stability. *Journal of Biotechnology*, 113(1-3), 105-120.

Elcock,A.H. (1999). Realistic modeling of the denatured states of proteins allows accurate calculations of the pH dependence of protein stability. *J Mol Biol*, 294(4), 1051-1062.

Endo,S., Saito,Y., & Wada,A. (1983). Denaturant-Gradient Chromatography for the Study of Protein Denaturation - Principle and Procedure. *Anal.Biochem.*, 131(1), 108-120.

Ennos,R. (2000). *Statistica and data handlings skills in biology.* New York: Prentice Hall.

Esakova,O.A., Meshalkina,L.E., & Kochetov,G.A. (2005). Effects of transketolase cofactors on its conformation and stability. *Life Sci.*, 78(1), 8-13.

Fenel,F., Leisola,M., Janis,J., & Turunen,O. (2004). A de novo designed N-terminal disulphide bridge stabilizes the Trichoderma reesei endo-1,4-beta-xylanase II. *J Biotechnol.*, 108(2), 137-143.

Fersht,A.R. (1985). *Enzyme structure and mechanism*. New York.

Fessner,W.D. (1998). Enzyme mediated C-C bond formation. *Curr.Opin.Chem.Biol.*, 2(1), 85-97.

Fiedler,E., Golbik,R., Schneider,G., Tittmann,K., Neef,H., Konig,S., & Hubner,G. (2001). Examination of donor substrate conversion in yeast transketolase. *Journal Of Biological Chemistry*, 276(19), 16051-16058.

Fink,A.L., Calciano,L.J., Goto,Y., Kurotsu,T., & Palleros,D.R. (1994). Classification of Acid Denaturation of Proteins: Intermediates and Unfolded States. *Biochemistry*, 33(41), 12504-12511.

Forsyth,W.R., Antosiewicz,J.M., & Robertson,A.D. (2002). Empirical relationships between protein structure and carboxyl pKa values in proteins. *Proteins*, 48, 388-403.

French,C. & Ward,J.M. (1995). Improved production and stability of *E.coli* recombinants expressing transketolase for large scale biotransformation. *Biotechnol.Lett*, 17(3), 247-252.

Fromant,M., Blanquet,S., & Plateau,P. (1995). Direct Random Mutagenesis of Gene-Sized DNA Fragments Using Polymerase Chain Reaction. *Analytical Biochemistry*, 224(1), 347-353.

George-Nascimento,C., Lowenson,J., Borissenko,M., Calderon,M., Medina-Selby,A., Kuo,J., Clarke,S., & Randolph,A. (1990). Replacement of a labile aspartyl residue increases the stability of human epidermal growth factor. *Biochemistry*, 29(41), 9584-9591.

Gerk,L.P., Leven,O., & Muller-Hill,B. (2000). Strengthening the dimerisation interface of Lac repressor increases its thermostability by 40 deg. C. *J Mol Biol*, 299(3), 805-812.

Germain,P., Slagmolen,T., & Crichton,R.R. (1989). Relation between stabilization and rigidification of the three-dimensional structure of an enzyme. *Biotechnol.Bioeng.*, 33(5), 563-569.

Goto, Calciano,L., & Fink,A. (1990). Acid-induced folding of proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 87(2), 573-577.

Goto,Y. & Fink,A.L. (1989). Conformational states in .beta.-lactamase: molten-globule states at acidic and alkaline pH with high salt. *Biochemistry*, 28(3), 945-952.

Goto,Y. & Nishikiori,S. (1991). Role of electrostatic repulsion in the acidic molten globule of cytochrome c. *Journal of Molecular Biology*, 222(3), 679-686.

Grantham,R. (1974). Amino Acid Difference Formula to Help Explain Protein Evolution. *Science*, 185(4154), 862-864.

Greenfield,N.J. (2006). Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nat.Protoc.*, 1(6), 2527-2535.

Gustafsson,C., Govindarajan,S., & Minshull,J. (2003). Putting engineering back into protein engineering: bioinformatic approaches to catalyst design. *Curr.Opin.Biotechnol.*, 14(4), 366-370.

Hakulinen,N., Turunen,O., Janis,J., Leisola,M., & Rouvinen,J. (2003). Three-dimensional structures of thermophilic beta-1,4-xylanases from Chaetomium thermophilum and Nonomuraea flexuosa. Comparison of twelve xylanases in relation to their thermal stability. *Eur.J Biochem.*, 270(7), 1399-1412.

Härdle,W.S.L. (2007). *Applied Multivariate Statistical Analysis*. Belgium: Springer.

Harris,T.K. & Turner,G.J. (2002). Structural basis of perturbed pKa values of catalytic groups in enzyme active sites. *IUBMB.Life*, 53(2), 85-98.

Hecquet,L., Bolte,J., & 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., & Demuynck,C. (1994). Chemoenzymatic Synthesis of Precursors of Fagomine and 1,4-Dideoxy-1,4-Imino-D-Arabinitol. *Tetrahedron Lett.*, 35(47), 8791-8794.

Hibbert,E.G., Baganz,F., Hailes,H.C., Ward,J.M., Lye,G.J., Woodley,J.M., & Dalby,P.A. (2005). Directed evolution of biocatalytic processes. *Biomolecular Engineering*, 22, 11-19.

Hibbert,E.G. & Dalby,P.A. (2005). Directed evolution strategies for improved enzymatic performance. *Microbial Cell Factories*, 4.

Hibbert,E.G., Senussi,T., Costelloe,S.J., Lei,W., Smith,M.E.B., Ward,J.M., Hailes,H.C., & Dalby,P.A. (2007). Directed evolution of transketolase activity on non-phosphorylated substrates. *Journal of Biotechnology*, 131(4), 425-432.

Himmo,S.D., Thomson,M., & Gubler,C.J. (1988). Isolation of transketolase from human erythrocytes. *Prep.Biochem.*, 18(3), 261-276.

Hobbs,G.R., Mitra,R.K., Chauhan,R.P., Woodley,J.M., & Lilly,M.D. (1996). Enzyme-catalysed carbon-carbon bond formation: Large-scale production of *Escherichia coli* transketolase. *J.Biotechnol.*, 45(2), 173-179.

Holthauzen,L.M. & Bolen,D.W. (2007). Mixed osmolytes: The degree to which one osmolyte affects the protein stabilizing ability of another. *Protein Science*, 16(2), 293-298.

Humphrey,A.J., Parsons,S.F., Smith,M.E.B., & Turner,N.J. (2000). Synthesis of a novel N-hydroxypyrrolidine using enzyme catalysed asymmetric carbon-carbon bond synthesis. *Tetrahedron Lett.*, 41(22), 4481-4485.

Innis,C.A., Anand,A.P., & Sowdhamini,R. (2004). Prediction of Functional Sites in Proteins Using Conserved Functional Group Analysis. *Journal of Molecular Biology*, 337(4), 1053-1068.

Innis,M.A.G.D.H.S.J.J.W.T.J. (1990). *PCR Protocols: Guide to Methods and Applications*. *New York*: *Academic Press*.

Jaenicke,R. (1981). Enzymes under extremes of physical conditions. *Annu.Rev.Biophys Bioeng.*, 10, 1-67.

Jaenicke,R. (1988). Stability and self-organization of proteins. *Naturwissenschaften*, 75(12), 604-610.

Jaenicke,R. (1991). Protein stability and protein folding. *Ciba Found.Symp.*, 161, 206-216.

Jaenicke,R. & Bohm,G. (1998). The stability of proteins in extreme environments. *Curr.Opin.Struct.Biol*, 8(6), 738-748.

Johan,A.W., Theodora,K., & John,F.M. (1998). Analysis of multiblock and hierarchical PCA and PLS models. *SO: Journal of Chemometrics*, 12(5), 301-321.

Jolliffe,I.T. (2002). *Principal Component Analysis*. NY: Springer.

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

Jonsson,J., Norberg,T., Carlsson,L., Gustafsson,C., & Wold,S. (1993). Quantitative sequence-activity models (QSAM)--tools for sequence design. *Nucleic Acids Res.*, 21(3), 733-739.

Kajander,T., Kahn,P.C., Passila,S.H., Cohen,D.C., Lehtio,L., Adolfsen,W., Warwicker,J., Schell,U., & Goldman,A. (2000). Buried charged surface in proteins. *Structure*, 8(11), 1203-1214.

Karpusas,M., Baase,W.A., Matsumura,M., & Matthews,B.W. (1989). Hydrophobic packing in T4 lysozyme probed by cavity-filling mutants. *Proc Natl.Acad.Sci U.S.A*, 86(21), 8237-8241.

Kelly,S.M. & Price,N.C. (2000). The use of circular dichroism in the investigation of protein structure and function. *Curr.Protein Pept.Sci*, 1(4), 349-384.

Kendall,M. & Gibbons,J. (1990). *Rank Correlation methods*. New York: Oxford University Press.

Kendall,M.G. (1976). *Rank Correlation Methods*. Griffin.

Killenberg-Jabs,M., Kern,G., Hubner,G., & Golbik,R. (2002). Folding and stability of different oligomeric states of thiamin diphosphate dependent homomeric pyruvate decarboxylase. *Biophys.Chem.*, 96(2-3), 259-271.

Kirk,O., Borchert,T.V., & Fuglsang,C.C. (2002). Industrial enzyme applications. *Current Opinion in Biotechnology*, 13(4), 345-351.

Kobori,Y., Myles,D.C., & Whitesides,G.M. (1992). Substrate specificity and carbohydrate synthesis using transketolase. *J.Org.Chem.*, 57, 5899-5907.

Kochetov,G.A. (2001). Functional flexibility of the transketolase molecule. *Biochem.(Mosc.).*, 66(10), 1077-1085.

Kochetov,G.A., Philippov,P.P., Razjivin,A.P., & Tikhomirova,N.K. (1975). Kinetics of reconstruction of holo-transketolase. *FEBS Lett.*, 53(2), 211-212.

Koeller,K.M. & Wong,C.H. (2001). Enzymes for chemical synthesis. *Nat.*, 409(6817), 232-240.

Kunihiro,K. (1989). The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. *Proteins: Structure, Function., and Genetics*, 6, 87-103.

Lay,D. (2000). *Linear Algebra and It's Applications.* New York.: Addison-Wesley.

Lee,C.K., Daniel,R.M., Shepherd,C., Saul,D., Cary,S.C., Danson,M.J., Eisenthal,R., & Peterson,M.E. (2007). Eurythermalism and the temperature dependence of enzyme activity. *FASEB J*, 21(8), 1934-1941.

Lehmann,M., Pasamontes,L., Lassen,S.F., & Wyss,M. (2000). The consensus concept for thermostability engineering of proteins. *Biochim.Biophys Acta*, 1543(2), 408-415.

Lehmann,M. & Wyss,M. (2001). Engineering proteins for thermostability: the use of sequence alignments versus rational design and directed evolution. *Curr.Opin.Biotechnol.*, 12(4), 371-375.

Lencki,R.W., Arul,J., & Neufeld,R.J. (1992). Effect of subunit dissociation, denaturation, aggregation, coagulation, and decomposition on enzyme inactivation kinetics: II. Biphasic and grace period behavior. *Biotechnol.Bioeng.*, 40(11), 1427-1434.

Li,H., Robertson,A.D., & Jensen,J.H. (2005). Very Fast Empirical Prediction and Rationalization of Protein pKa values. *Proteins*, 55, 689-704.

Lilie,H. (2003). Designer proteins in biotechnology. International Titisee Conference on protein design at the crossroads of biotechnology, chemistry and evolution. *EMBO Rep.*, 4(4), 346-351.

Linding,R., Schymkowitz,J., Rousseau,F., Diella,F., & Serrano,L. (2004). A comparative study of the relationship between protein structure and beta-aggregation in globular and intrinsically disordered proteins. *Journal of Molecular Biology*, 342(1), 345-353.

Lindqvist,Y., Schneider,G., Ermler,U., & Sundstrom,M. (1992). 3-Dimensional Structure of Transketolase, A Thiamine Diphosphate Dependent Enzyme, at 2.5 Angstrom Resolution. *Embo Journal*, 11(7), 2373-2379.

Littlechild,J.A., Turner,N.J., Hobbs,G.R., Lilly,M.D., Rawas,A., & Watson,H. (1995). Crystallization and preliminary-X-ray crystallographic data with *Escherichia-coli* transketolase. *Acta Crystallogr., Sect D: Biol.Crystallogr.*, 51, 1074-1076.

Lockless,S.W. & Ranganathan,R. (1999). Evolutionarily Conserved Pathways of Energetic Connectivity in Protein Families. *Science*, 286(5438), 295-299.

Macarthur,M.W. & Thornton,J.M. (1991). Influence of proline residues on protein conformation. *J Mol Biol*, 218(2), 397-412.

MacBeath,G. (2001). Proteomics comes to the surface. *Nat.Biotechnol*, 19(9), 828-829.

Manavalan,P. & Johnson,W.C. (1983). Sensitivity of Circular-Dichroism to Protein Tertiary Structure Class. *Nat.*, 305(5937), 831-832.

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

Martinez-Torres,R.J., Aucamp,J.P., George,R., & Dalby,P.A. (2007). Structural stability of E. coli transketolase to urea denaturation. *Enzyme and Microbial Technology*, 41(5), 653-662.

Masri,S.W., Ali,M., & Gubler,C.J. (1988). Isolation of transketolase from rabbit liver and comparison of some of its kinetic properties with transketolase from other sources. *Comp Biochem.Physiol B*, 90(1), 167-172.

Matthew,J.B., Gurd,F.R., Garcia-Moreno,B., Flanagan,M.A., March,K.L., & Shire,S.J. (1985). pH-dependent processes in proteins. *CRC Crit Rev.Biochem.*, 18(2), 91-197.

Matthews,B.W. (1993a). Structural and Genetic Analysis of Protein Stability. *Annual Review of Biochemistry*, 62(1), 139-160.

Matthews,B.W. (1993b). Structural and genetic analysis of protein stability. *Annu.Rev.Biochem.*, 62, 139-160.

Matthews,B.W. (1995). Studies on Protein Stability With T4 Lysozyme. InC.B.Anfinsen (Ed.), *Advances in Protein Chemistry* (pp. 249-278). Academic Press.

McPhie,P. (1982). Swine pepsinogen folding intermediates are highly structured, motile molecules. *Biochemistry*, 21(22), 5509-5515.

Meshalkina,L., Nilsson,U., Wikner,C., Kostikowa,T., & Schneider,G. (1997). Examination of the thiamin diphosphate binding site in yeast transketolase by site-directed mutagenesis. *Eur.J.Biochem.*, 244(2), 646-652.

Miranda,A.A., Le Borgne,Y., & Bontempi,G. (2008). New Routes from Minimal Aproximation Error to Principal Components. *Neural Processing Letters*, 27(3), 197-207.

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

Miyazaki,K. & Arnold,F.H. (1999). Exploring nonnatural evolutionary pathways by saturation mutagenesis: rapid improvement of protein function. *J.Mol.Evol.*, 49(6), 716-720.

Morley,K.L. & Kazlauskas,R.J. (2005). Improving enzyme properties: when are closer mutations better? *Trends Biotechnol.*, 23(5), 231-237.

Morris,K.G., Smith,M.E.B., Turner,N.J., Lilly,M.D., Mitra,R.K., & Woodley,J.M. (1996). Transketolase from Escherichia coli: A practical procedure for using the biocatalyst for asymmetric carbon-carbon bond synthesis. *Tetrahedron: Asymmetry*, 7(8), 2185-2188.

Muller,Y.A., Lindqvist,Y., Furey,W., Schulz,G.E., Jordan,F., & Schneider,G. (1993). A Thiamin Diphosphate Binding Fold Revealed by Comparison of the Crystal-Structures of Transketolase, Pyruvate Oxidase and Pyruvate Decarboxylase. *Struct.*, 1(2), 95-103.

Murphy,R.M. (1997). Static and dynamic light scattering of biological macromolecules: What can we learn? *Curr.Opin.Biotechnol.*, 8(1), 25-30.

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

Nath,D. & Rao,M. (2001). pH dependent conformational and structural changes of xylanase from an alkalophilic thermophilic Bacillus sp (NCIM 59). *Enzyme and Microbial Technology*, 28(4-5), 397-403.

Nelson, D. and Cox, M. Lehninger: Principles of biochemistry. [4th edition], 1-1119. 2004. W.H.Freeman.
Ref Type: Serial (Book,Monograph)

Ness,J.E., Welch,M., Giver,L., Bueno,M., Cherry,J.R., Borchert,T.V., Stemmer,W.P., & Minshull,J. (1999). DNA shuffling of subgenomic sequences of subtilisin. *Nat.Biotechnol.*, 17(9), 893-896.

Nielsen,J.E. (2007). Analysing the pH-dependent properties of proteins using pKa calculations. *Journal of Molecular Graphics and Modelling*, 25(5), 691-699.

Nikkola,M., Lindqvist,Y., & Schneider,G. (1994). Refined structure of transketolase from *Saccharomyces cerevisiae* at 2.0 A resolution. *J.Mol.Biol.*, 238(3), 387-404.

Nilsson,U., Meshalkina,L., Lindqvist,Y., & Schneider,G. (1997). Examination of substrate binding in thiamin diphosphate-dependent transketolase by protein crystallography and site-directed mutagenesis. *J.Biol.Chem.*, 272(3), 1864-1869.

Nist/Sematech (2008). *e-Handbook of Statistical Methods*.

O'Fagain,C. (2003). Enzyme stabilization--recent experimental progress. *Enzyme and Microbial Technology*, 33(2-3), 137-149.

Olsen,M., Iverson,B., & Georgiou,G. (2000). High-throughput screening of enzyme libraries. *Curr.Opin.Biotechnol.*, 11(4), 331-337.

Pace,C.N. (1990). Measuring and increasing protein stability. *Trends Biotechnol.*, 8(4), 93-98.

Pace,C.N., Shirley,B.A., McNutt,M., & Gajiwala,K. (1996). Forces contributing to the conformational stability of proteins. *FASEB J*, 10(1), 75-83.

Pace,C.N., Trevino,S., Prabhakaran,E., & Scholtz,J.M. (2004). Protein structure, stability and solubility in water and other solvents. *Philos.Trans.R.Soc.Lond B Biol Sci*, 359(1448), 1225-1234.

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

Perez-Jimenez,R., Godoy-Ruiz,R., Ibarra-Molero,B., & Sanchez-Ruiz,J.M. (2004). The efficiency of different salts to screen charge interactions in proteins: a Hofmeister effect? *Biophys J*, 86(4), 2414-2429.

Perutz,M.F. & Raidt,H. (1975). Stereochemical basis of heat stability in bacterial ferredoxins and in haemoglobin A2. *Nature*, 255(5505), 256-259.

Peterson,M.E., Eisenthal,R., Danson,M.J., Spence,A., & Daniel,R.M. (2004). A New Intrinsic Thermal Parameter for Enzymes Reveals True Temperature Optima. *Journal of Biological Chemistry*, 279(20), 20717-20722.

Pharkya,P., Nikolaev,E.V., & Maranas,C.D. (2003). Review of the BRENDA Database. *Metabolic Engineering*, 5(2), 71-73.

Philippov,P.P., Shestakova,I.K., Tikhomirova,N.K., & Kochetov,G.A. (1980). Characterization and properties of pig liver transketolase. *Biochim Biophys Acta*, 613(2), 359-369.

Pohl,M., Grotzinger,J., Wollmer,A., & Kula,M.R. (1994). Reversible dissociation and unfolding of pyruvate decarboxylase from *Zymomonas-mobilis*. *Eur.J.Biochem.*, 224(2), 651-661.

Puchkaev,A.V., Koo,L.S., & 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.

Querol,E., Perez-Pons,J.A., & Mozo-Villarias,A. (1996). Analysis of protein conformational characteristics related to thermostability. *Protein Engineering Design and Selection*, 9(3), 265-271.

Reetz,M.T. & Carballeira,J.D. (2007). Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat.Protocols*, 2(4), 891-903.

Richardson,T.H., Tan,X., Frey,G., Callen,W., Cabell,M., Lam,D., Macomber,J., Short,J.M., Robertson,D.E., & Miller,C. (2002). A novel, high performance enzyme for starch liquefaction. Discovery and optimization of a low pH, thermostable alpha-amylase. *J Biol Chem*, 277(29), 26501-26507.

Robertson,D.E. & Steer,B.A. (2004). Recent progress in biocatalyst discovery and optimization. *Current Opinion in Chemical Biology*, 8(2), 141-149.

Rozzell, J. D. Biocatalysis at commercial scale: myths and realities. Chimica Oggi . 1999.
Ref Type: Magazine Article

Rufus,L., Rodney,B., & John,F.B. (1966). Validity of the "two-state" hypothesis for conformational transitions of proteins. *SO: Biopolymers*, 4(8), 917-944.

Sandberg,W.S. & Terwilliger,T.C. (1989). Influence of interior packing and hydrophobicity on the stability of a protein. *Science*, 245(4913), 54-57.

Sandgren,M., Gualfetti,P.J., Shaw,A., Gross,L.S., Saldajeno,M., Day,A.G., Jones,T.A., & Mitchinson,C. (2003). Comparison of family 12 glycoside hydrolases and recruited substitutions important for thermal stability. *Protein Science*, 12(4), 848-860.

Schenk,G., Duggleby,R.G., & Nixon,P.F. (1998). Properties and functions of the thiamin diphosphate dependent enzyme transketolase. *Int.J.Biochem.Cell Biol.*, 30(12), 1297-1318.

Schiraldi,C. & De Rosa,M. (2002). The production of biocatalysts and biomolecules from extremophiles. *Trends in Biotechnology*, 20(12), 515-521.

Schmid,A., Dordick,J.S., Hauer,B., Kiener,A., Wubbolts,M., & Witholt,B. (2001). Industrial biocatalysis today and tomorrow. *Nature*, 409(6817), 258-268.

Schmidt-Dannert,C. (2001). Directed evolution of single proteins, metabolic pathways, and viruses. *Biochemistry*, 40(44), 13125-13136.

Schmidt-Dannert,C. & Arnold,F.H. (1999). Directed evolution of industrial enzymes. *Trends Biotechnol.*, 17(4), 135-136.

Schomburg,I., Chang,A., Ebeling,C., Gremse,M., Heldt,C., Huhn,G., & Schomburg,D. (2004). BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Res.*, 32(Database issue), D431-D433.

Schomburg,I., Chang,A., Hofmann,O., Ebeling,C., Ehrentreich,F., & Schomburg,D. (2002). BRENDA: a resource for enzyme data and metabolic information. *Trends Biochem.Sci*, 27(1), 54-56.

Schorken,U. & Sprenger,G.A. (1998). Thiamin-dependent enzymes as catalysts in chemoenzymatic syntheses. *Biochim.Biophys.Acta*, 1385(2), 229-243.

Schwehm,J.M., Fitch,C.A., Dang,B.N., Garcia-Moreno,B., & 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.

Selivanov,V.A., Kovina,M.V., Kochevova,N.V., Meshalkina,L.E., & Kochetov,G.A. (2003). Studies of thiamin diphosphate binding to the yeast apotransketolase. *J.Mol.Catal.B: Enzym.*, 26(1-2), 33-40.

Serrano,L., Day,A.G., & Fersht,A.R. (1993). Step-wise Mutation of Barnase to Binase : A Procedure for Engineering Increased Stability of Proteins and an Experimental Analysis of the Evolution of Protein Stability. *Journal of Molecular Biology*, 233(2), 305-312.

Serrano,L. & Fersht,A.R. (1989). Capping and [alpha]-helix stability. *Nature*, 342(6247), 296-299.

Serrano,L., Horovitz,A., Avron,B., Bycroft,M., & 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., & Fersht,A.R. (1992). Effect of alanine versus glycine in [alpha]-helices on protein stability. *Nature*, 356(6368), 453-455.

Sheinerman,F.B., Norel,R., & Honig,B. (2000). Electrostatic aspects of protein-protein interactions. *Curr.Opin.Struct.Biol*, 10(2), 153-159.

Sprenger,G.A., Schorken,U., Sprenger,G., & Sahm,H. (1995). Transketolase A of *Escherichia coli* K12. Purification and properties of the enzyme from recombinant strains. *Eur.J.Biochem.*, 230(2), 525-532.

Stigter,D., Alonso,D.O.V., & Dill,K.A. (1991). Protein Stability: Electrostatics and Compact Denatured States. *Proceedings of the National Academy of Sciences of the United States of America*, 88(10), 4176-4180.

Straathof,A.J., Panke,S., & Schmid,A. (2002). The production of fine chemicals by biotransformations. *Curr.Opin.Biotech.*, 13(6), 548-556.

Sundstrom,M., Lindqvist,Y., & Schneider,G. (1992). 3-Dimensional structure of apotransketolase - flexible loops at the active-site enable cofactor binding. *FEBS Lett.*, 313(3), 229-231.

Suzuki,Y. (1999). The proline rule - A strategy for protein thermal stabilization. *Proceedings of the Japan Academy Series B-Physical and Biological Sciences*, 75(6), 133-137.

Sylvestre,J., Chautard,H., Cedrone,F., & Delcourt,M. (2006). Directed evolution of biocatalysts. *Organic Process Research & Development*, 10(3), 562-571.

Szilagyi,A. & Zavodszky,P. (2000). Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. *Structure.*, 8(5), 493-504.

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

Thayer,A.M. (2001). Biocatalysis. *Chemical & Engineering News*, 79(21), 27-34.

Thompson,J.D., Higgins,D.G., & 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.

Tomazic,S.J. & Klibanov,A.M. (1988). Mechanisms of irreversible thermal inactivation of Bacillus alpha- amylases. *Journal of Biological Chemistry*, 263(7), 3086-3091.

Turner,N.J. (2000). Applications of transketolases in organic synthesis. *Curr.Opin.Biotechnol.*, 11(6), 527-531.

Turner,N.J. (2003). Directed evolution of enzymes for applied biocatalysis. *Trends in Biotechnology*, 21(11), 474-478.

van den Burg,B. & Eijsink,V.G. (2002). Selection of mutations for increased protein stability. *Current Opinion in Biotechnology*, 13(4), 333-337.

Veitch,N.J., Maugeri,D.A., Cazzulo,J.J., Lindqvist,Y., & Barrett,M.P. (2004). Transketolase from *Leishmania mexicana* has a dual subcellular localization. *Biochem.J*, 382, 759-767.

Vetriani,C., Maeder,D.L., Tolliday,N., Yip,K.S., Stillman,T.J., Britton,K.L., Rice,D.W., Klump,H.H., & Robb,F.T. (1998). Protein thermostability above 100 degreesC: a key role for ionic interactions. *Proc.Natl.Acad.Sci U.S.A*, 95(21), 12300-12305.

Vieille,C. & 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. & Axelrod,B. (1971). Heptulose Synthesis from Nonphosphorylated Aldoses and Ketoses by Spinach Transketolase. *J.Biol.Chem.*, 246(10), 3126-3131.

Vogt,G., Etzold,T., & Argos,P. (1995). An Assessment of Amino Acid Exchange Matrices in Aligning Protein Sequences: The Twilight Zone Revisited. *Journal of Molecular Biology*, 249(4), 816-831.

Wahler,D. & Reymond,J.L. (2001). Novel methods for biocatalyst screening. *Curr.Opin.Chem.Biol.*, 5(2), 152-158.

Walsh,C. (2001). Enabling the chemistry of life. *Nature*, 409(6817), 226-231.

Wang,L.j., Kong,X.d., Zhang,H.y., Wang,X.p., & Zhang,J. (2000). Enhancement of the Activity of -Aspartase from Escherichia coli W by Directed Evolution. *Biochemical and Biophysical Research Communications*, 276(1), 346-349.

Warshel,A., Sharma,P.K., Kato,M., & Parson,W.W. (2006). Modeling electrostatic effects in proteins. *Biochim.Biophys Acta*, 1764(11), 1647-1676.

Williams,J.C., Zeelen,J.P., Neubauer,G., Vriend,G., Backmann,J., Michels,P.A., Lambeir,A.M., & Wierenga,R.K. (1999). Structural and mutagenesis studies of leishmania triosephosphate isomerase: a point mutation can convert a mesophilic enzyme into a superstable enzyme without losing catalytic power. *Protein Eng*, 12(3), 243-250.

Witt,A.C., Lakshminarasimhan,M., Remington,B.C., Hasim,S., Pozharski,E., & Wilson,M.A. (2008). Cysteine pKa Depression by a Protonated Glutamic Acid in Human DJ-1. *Biochemistry*.

Woodley,J.M., Mitra,R.K., & Lilly,M.D. (1996). Carbon-carbon bond synthesis - reactor design and operation for transketolase catalyzed biotransformations. *Ann.N.Y.Acad.Sci.*, 799, 434-445.

Yang,A.S. & Honig,B. (1993). On the pH Dependence of Protein Stability. *Journal of Molecular Biology*, 231(2), 459-474.

Ye,J., McGinnis,S., & Madden,T.L. (2006). BLAST: improvements for better sequence analysis. *Nucleic Acids Research*, 34(suppl_2), W6-W9.

Zhao,H. & Arnold,F.H. (1999). Directed evolution converts subtilisin E into a functional equivalent of thermitase. *Protein Engineering Design and Selection*, 12(1), 47-53.

Zhao,H., Chockalingam,K., & Chen,Z. (2002). Directed evolution of enzymes and pathways for industrial biocatalysis. *Current Opinion in Biotechnology*, 13(2), 104-110.

Zimmermann,F.T., Schneider,A., Sch÷rken,U., Sprenger,G.A., & Fessner,W.D. (1999). Efficient multi-enzymatic synthesis of -xylulose 5-phosphate. *Tetrahedron: Asymmetry*, 10(9), 1643-1646.