PhD thesis

Daniel Stiefler-Jensen

STABILITY OF HYPERTHERMOPHILIC PROTEINS
-Post-translational modifications and other factors

Academic Supervisors: Qunxin She & Kaare Teilum
Submitted: December 2016
PhD thesis

Stability of Hyperthermophilic Proteins
- Post-translational modifications and other factors

This thesis has been submitted to the PhD School of The Faculty of Science, University of Copenhagen

By Daniel Stiefler-Jensen

Academic supervisor: Qunxin She

Co-supervisor: Kaare Teilum
Institute name: Biologisk Institut

Name of department: Department of Biology

Section: Functional Genomics

Author: Daniel Stiefler-Jensen

Titel(dansk): Hyper-thermofile proteiners stabilitet
- Post-translatoriske modifikationer og andre faktorer

Title(english): Stability of Hyperthermophilic Proteins
- Post-translational modifications and other factors

Subject description: This thesis aims to evaluate the importance of Post-translation modification as a mechanism of protein stabilisation in hyperthermophilic archaea along with the investigation of other factors involved in protein stability of hyper-thermophiles.

Main supervisor: Qunxin She

Co-supervisor: Kaare Teilum

Submitted: December 2016

Type: Synopsis
Preface

This thesis is the conclusion of my PhD. Study, started March 2012, at the Faculty of Science, University of Copenhagen. The majority of the experimental work was carried out at the Danish Archaea Centre (DAC) and Structural Biology and NMR Laboratory (SBiNLab) of the Department of Biology.

The thesis was supervised by Director of DAC and Associate Professor Qunxin She with a heavy involvement of Co-supervisor Kaare Teilum, group leader at SBiNLab and Associate Professor.

During my PhD. I had the opportunity and pleasure of joining the laboratory of Simonetta Bartolucci, at the University of Naples Federico II, where I joined the groups of Gabriella Fiorentino and Patrizia Contursi for three months.

I also had the pleasure of supervising two bachelor students and a laboratory technician through their final projects and internships, respectively. Additionally, I assisted with the teaching of Biochemistry III for two consecutive years.

Daniel Stiefler-Jensen
Copenhagen, Denmark
December, 2016
Summary

Enzymes are not only essential for the workings of every cell in our bodies. They are also becoming more important for our daily life as they play an increasing role in production of and as part of everyday items. Thus, increasing the efficiency of these enzymes can result in better and possible cheaper products. One aspect that can have a large impact on the efficiency of an enzyme is its stability. By increasing the enzyme stability production cost and time can be reduced, and consumers will have a better product with longer activity.

In the past it was only possible to increasing enzymes stability by randomly generate mutants and lengthy screening processes to identify the best new mutants. However, with the increase in available genomic sequences of thermophilic or hyperthermophilic organisms a world of enzymes with intrinsic high stability are now available. As these organisms are adapted to life at high temperatures so are their enzymes, as a result the high stability is accompanied by low activity at moderate temperatures. Thus, much effort had been put into decoding the mechanisms behind the high stability of the thermophilic enzymes. The hope is to enable scientist to design enzymes with high stability and activity at a target temperature. This thesis presents an investigation of an hyperthermophilic esterase, and how post-translational modifications (PTM) and other factors affect the stability of this enzyme. As part on the ongoing effort to understanding the mechanisms involved in the high stability of hyperthermophilic enzymes.

The thesis starts with an introduction to the field of protein and enzyme stability with special focus on the thermophilic and hyperthermophilic enzymes and proteins. After the introduction three original research manuscripts present the experimental data related to this study. In the first manuscript, the effect of lysine methylation on enzyme stability is investigated. This study makes use of two different methods to acquire enzyme without the native lysine methylation. The effect of the methylation is subsequently evaluated by testing the stability of the different versions of the esterase. Here we show that methylation plays a minimal, yet significant, role in stabilising the enzyme. We also show that if the esterase is produced from a mesophilic host, an organism that lives at moderate temperatures, in this case *E. coli*, that additional factors that serves to stabilise the enzyme is either missing or that factors from *E. coli* destabilises the enzyme. We were however unable to identify these factors. The second manuscript present an investigation of the importance of dimer formation for the stability of the esterase investigated in
manuscript one. By the introduction of a novel cross dimer disulphide bond we demonstrated a stabilising effect of dimer formation by increasing the half-life of the esterase by 1.62 fold at a temperature of 90°C. However, efforts to disrupt the native dimer formation by targeting the core elements of the dimer interface failed. The mutants were however compromised in catalytic stability at a temperature of 90°C. The third manuscript details problem area and makes suggestions for better design of esterase activity assay for use at temperature up to 90°C.

ved en temperatur på 90°C. Men bestræbelserne på at bryde dimer dannelselse ved at målrettet at mutere de centrale elementer i dimer grænsefladen mislykkedes. Mutanternes katalytiske stabilitet blev imidlertid kompromitteret ved en temperatur på 90°C. Det tredje manuskript beskriver problemområder og giver forslag til bedre udformning af esterase aktivitets assays til anvendelse ved temperaturer op til 90°C og kontinuert data indsamling.
Acknowledgements

During the last 4, almost 5, years I have learned a lot, not only in respect to my research but also about myself. I have learned that I can be my own worst enemy, and how to overcome myself. In addition to the scientific knowledge, I have gained over this period of my life, I count these experiences as the most important and I would like to thank all who have been involved in my personal and professional evolution during my PhD. If I fail to mention you by name, I do apologise. But do know that you have been a part of an important and amazing part of my life.

I would like to thanks Qunxin She and Kaare Teilum for the opportunity to conduct my PhD study in their laboratories and for the support and guidance I have received over the years. I will also like to thanks Roger A. Garrett and Xu Peng for their always enlighten input to a scientific discussion. Special thanks go to my colleague and follow PhD student Troels Schwarz-Linnet, thanks for great team work. I will also like to thanks current and former members of the Danish Archaea Centre. Also thanks to Hien Phan, Mariana Awayez, and Pia Skovgaard for their technical expertise in the laboratory, without laboratory technicians the labs would fall apart.

I will like to thank my parents and my in-laws for their support during this period of my life without your support and understanding it would not have been possible.

The conclusion of the PhD would not have been possible without my wonderful and patient family. I wished to express my deepest gratitude to my lovely wife, Pernille and my two wonderful sons, Milo and Sirius. Without their support and understanding I would not have been able to conclude my work. From my heart thanks.

-Daniel
List of manuscripts


# Table of Contents

Preface ......................................................................................................................... iii  
Summary ......................................................................................................................... iv  
Sammenfatning ............................................................................................................... vi  
Acknowledgements ....................................................................................................... viii  
List of manuscripts ....................................................................................................... ix  
Table of Contents ......................................................................................................... x  

## Synopsis .................................................................................................................... 1  
1 Aims of the thesis ...................................................................................................... 1  
2 Introduction ................................................................................................................ 2  
3 Protein stability .......................................................................................................... 3  
  3.1 Protein folding ......................................................................................................... 4  
  3.2 Stabilising factors .................................................................................................. 6  
  3.3 Non-covalent interactions ....................................................................................... 6  
    Ionic interaction ....................................................................................................... 8  
    Van der Waals forces ............................................................................................... 9  
    Hydrophobic interactions ....................................................................................... 9  
  3.4 Covalent interactions ............................................................................................ 10  
  3.5 Post-translational modifications .......................................................................... 11  
4 Protein structure determination ............................................................................... 12  
  4.1 Computational structural biology ........................................................................ 14  
    Homology modelling .............................................................................................. 14  
    Discrete state off-lattice models ............................................................................ 15  
    Structural fragment modelling ............................................................................. 16  
5 Protein engineering .................................................................................................... 17  
  5.1 Directed evolution, traditional ............................................................................. 17  
  5.2 Rational design, modern ....................................................................................... 18  
6 Protein adaptation to extreme temperatures ........................................................... 19  
  6.1 Amino acid composition ....................................................................................... 20  
  6.2 Counter selection of thermolabile amino acids .................................................... 21  
  6.3 Increased number of charged amino acids ............................................................ 21  
  6.4 Lysine methylation: .............................................................................................. 22  
  6.5 Structural optimisation ......................................................................................... 23  
  6.6 Hydrogen bonds .................................................................................................. 24  
  6.7 Hydrophobicity .................................................................................................... 24  
  6.8 Increased number of disulphide bonds ................................................................. 24
Oligomerisation ........................................................................................................................................... 25
7 Enzymes of special interest .......................................................................................................................... 25
7.1 EstA, an Archaeal esterase .......................................................................................................................... 26
7.2 aKMT, an Archaeal lysine methyltransferase ............................................................................................ 28
8 Summary of results .......................................................................................................................................... 29
8.1 Optimisation of esterase assay .................................................................................................................... 29
8.2 Expression of soluble dimeric EstA for E. coli host ...................................................................................... 30
8.3 Lysine methylation as a factor in protein stability ..................................................................................... 30
8.4 Investigation mesophilic expressed EstA increased sensitivity to detergents ............................................. 31
8.5 The Effect of dimer integrity on the stability of EstA .................................................................................. 31
9 Conclusion & Perspective .............................................................................................................................. 31
10 References ..................................................................................................................................................... 34
Chapter 1 ......................................................................................................................................................... 43
Chapter 2 ......................................................................................................................................................... 65
Chapter 3 ......................................................................................................................................................... 89
Appendix I ......................................................................................................................................................... 99
A – Introduction to Mutations Analyser ........................................................................................................ 99
B – Mutations Analyser python code ............................................................................................................. 106
Synopsis
Stability of Hyper-Thermophilic Proteins

1 Aims of the thesis

The work presented here were made in an effort elucidate the role of post-translational modifications, with special focus on lysine methylation, on the stability of proteins from hyperthermophilic archaea of the genus *Sulfolobus*. To achieve this an esterase native to *Sulfolobus islandicus*, EstA, previously showed to have compromised stability when expressed in *Escherichia coli*, used as a model for post-translational modification dependent thermostability. Through the use of this model the following is addressed:

- Characterising the difference between protein expressed from *E. coli*, a mesophile, and *S. islandicus*, a hyperthermophile.
  - Addressing key issues in previously published studies of EstA stability when expressed from *E. coli*: Oligomeric state, difference in amino acid sequence.
- Identify post-translational modifications of EstA expressed from the native host *S. islandicus*.
- Produce EstA from *S. islandicus*, the native host, without *in vivo* lysine methylation
- Evaluating the stabilising effect of post-translation modification.
- Investigate the importance of the EstA oligomeric state in relation to stability.
2 Introduction

For thousands of years’ and across the world, Mankind have exploited microorganisms, most notably for fermentation of sugar to ethanol e.g. in the making of wine and brewing of beer, and for leavening of bread. Our utilization of nature is now at the point where we can pick “parts” from all manner of life and use them as we see fit, these “parts” or genes can code for enzymes. The possible uses of enzymes are almost limit less, Enzymes can be used in the production of medicines, biofuel, chemical, food, and in paper production and recycling, as treatment for metabolic disorders in enzyme replacement therapy (Neufeld 2006), or as active ingredients in e.g. laundry detergents, toothpastes and dish washing detergents and much more. The use of enzymatic biocatalysts in an industrial production can be beneficially in several ways e.g. by lowering energy usage, increasing product yields and quality, and reducing the environmental impact. As enzymes in many cases can replace expensive and polluting production methods in the chemical industry with relative mild reaction conditions. This use of enzymes is known as white biotechnology (Frazzetto 2003). In the white biotechnology industry and in general, high enzyme stability is a valued trait for almost all application of enzymes, one exception is DNA restriction endonucleases used in molecular biology where thermal inactivation is a practical trait of the enzymes. The stability of an enzyme can have a large impact on the amount of enzyme and processing time needed and as a result on the cost of using an enzymatic catalyst. Hence maximising the stability of an enzyme, sometimes at the cost of activity, can be high beneficially (Cao et al. 2015), as an increase in stability allows for an increase in the processing temperature, which in turn result in acceleration of substrate conversion (Klibanov 1983). It is however not always possible to obtain enzymes with a high stability from organisms living at normal temperatures, known as mesophiles. However, thermophiles and hyperthermophiles, organisms living at high temperatures above 45°C and 65°C respectively, possess enzymes that through evolution have acquired high stability. It is not always possible to find a thermophilic enzyme suitable as a replacement for a mesophilic enzyme. Hence decoding the factors that make thermophilic enzymes stable, may allow us to engineer thermophilic enzymes from mesophilic enzymes. The research presented in this thesis investigate the effect of post-translational modifications and other factors on the stability of hyperthermophilic enzymes.
In the following section I will introduce determining factors in protein stability and folding along with some background information about protein structure determination and protein modelling.

3 Protein stability

Enzyme activity is essential for life, as enzymes are responsible for the cells metabolism. Enzyme functionality and activity is dependent on the active site, a micro-environment tailored for catalysis of specific chemical reactions. Thesis micro-environments are the results of the peptide-polymer of the enzyme folding to a specific three-dimensional (3D) structure where the enzyme is active. This structure is known as the native conformation. If for any reason this 3D structure is disrupted the enzyme loses it activity and functionality. Hence the longer an enzyme can maintain its native conformation the better. Both the 3D structure and its stability is dependent on the amino acid sequence of the enzyme. This was evident as early as 1961 where it was show that peptides can fold spontaneously in vitro (Anfinsen et al. 1961). The amino acid sequence of all proteins is encoded in DNA as genes. The DNA encoded amino acid sequence is transcribed from DNA to mRNA and translated from mRNA into a peptide during protein synthesis. The flow of information from DNA to protein is known as the central dogma of information flow, see Figure 1. As the amino acid sequence is the only deterministic feature for the protein structure, it must also be deterministic for its stability. How the amino acid sequence is responsible for the folding and stability, is the subject of the following sections.

Figure 1. The central dogma of information flow. 1| a DNA encoded gene or open reading frame is transcribed to mRNA by a RNA polymerase. 2| the mRNA is translated to a peptide sequence by the ribosome and tRNA’s.
3.1 Protein folding

As indicated above protein folding and protein stability are not two independent concepts, both are dependent on the amino acid sequence of the protein. The factors that determines the protein folding are also the factors that determines the stability of the fold. Proteins are synthesised as linear polymers consisting of a combination of 20 standard amino acids. Each amino acid consists of an α-amino group, α-carboxyl group, a α-carbon and a side chain, see Figure 2. The α-amino group, α-carboxyl group, and a α-carbon form the peptide back-bone, where individual amino acids are joined together with a peptide bond through a condensation reaction between the α-carboxyl group of the first amino acid and the α-amino group of the next in the poly-peptide chain. The side chain of each amino acid is unique, and the amino acids can be divided into four groups, based on chemical properties of their side chain: hydrophobic, polar uncharged, charged, and special cases. As proteins are synthesised they start interaction with its solvent, water under normal circumstances. As a result, some hydrogen bonds between water molecules will be broken and new ones will be formed between the protein and the water. Usually this will be thermodynamic unfavourable. This unfavourable interaction with water drives the folding of the poly-peptide to minimise the Gibbs free energy for the protein/water system, see BOX I. The efficiency of this minimisation determines the stability of the protein. The average difference in Gibbs-free energy, obtained through protein folding, $\Delta G_F$, is only about 40 kJ/mol (Dill et al. 2008).

![Figure 2](image.png)

**Figure 2.** The parts of the amino acid: α-amino group, α-carboxyl group, α-carbon, and R the side chain

---

**BOX I: Thermodynamic of state change**

For a state change, like that of a protein folding from a non-native state to the native state, to be spontaneously and stable the change in Gibbs-free energy, $\Delta G$, must be $< 0$ and a minimum, see Eq. 1 where, $G_N$ and $G_D$ are the Gibbs-free energy of the native fold and the non-native folds respectively.

$$\text{Eq. 1} \quad \Delta G = G_N - G_D.$$  

The Gibbs-free energy of each state can be expressed as a function of entropy ($S$) and enthalpy ($H$), at constant pressure and temperature in the following manner:

$$\text{Eq. 2} \quad G = H - TS$$

By combining Eq. 1 & Eq. 2 we get Eq. 3. From Eq. 3 it can be seen that any loss of entropy as a result of steric constrains must be compensated for by a loss in enthalpy.

$$\text{Eq. 3} \quad \Delta G = \Delta H - T \Delta S$$

A spontaneous State changes, $\Delta G < 0$, can be enthalpy or entropy driven if $\Delta H < 0$ & $\Delta S < \Delta S <= 0$ or $\Delta S > 0$ & $\Delta S < \Delta H >= 0$ respectively.
Most protein folds from a linear poly-peptide to its native conformation on a microsecond (µs) time scale. An estimation of folding speeds for single domain proteins can be made from the length of poly-peptide (N) by N/100 µs (Kubelka et al. 2004). How protein and enzymes fold to their 3D structure in such a short time scale have troubled researchers. Thus Levinthal’s paradox states that proteins can’t fold to its native conformation by random. As his thought experiment demonstrated: a 101 amino acid long peptide will take $10^{27}$ years to sample all possible structures if each peptide bond only has 3 degrees of freedom and $10^{13}$ different conformations can be sampled per second (Zwanzig et al. 1992). To overcome this problem an initial theory where protein folding through specific pathways (Levinthal 1968) has now evolved to a theory where protein folding is seen as a funnel shaped energy landscape, see Figure 3. Where intermedia structures/local energy minimums guide the conformation to the native fold through multiple pathways (Dill & Chan 1997), and thereby greatly accelerate protein folding to a microsecond timescale.

Three mechanisms, that are not mutual exclusive, are used to describe the progression through the energy landscape. The frame-work mechanism accounts on the formation of local secondary structures independent of tertiary interaction, with a subsequent conglomeration to the native conformation by random movement and collision (Karplus & Weaver 1994; Kim Baldwin, R.L. 1982), see Figure 4a. In contrast the hydrophobic collapse model relies on the initial formation of a molten globular structure as a result of

![Figure 3. Proteins have a funnel-shaped energy landscape with many high-energy, unfolded structures and only a few low-energy, folded structures. Folding occurs via alternative microscopic trajectories.](image-url)

![Figure 4. Three types of protein folding pathways.](image-url)
stacking hydrophobic residues. From the molten globular secondary and tertiary structures will emerge (Dill 1985), see Figure 4b. The hydrophobic collapse is believed to be one of the earliest events in protein folding and experiments evidence place it on the Nano-second scale (Sadqi et al. 2003). The nucleation-condensation model is based on the initial formation of local interactions which serve as a template around which the structure folds around (Fersht 1995; Nölting & Agard 2008), see Figure 4c. The next section will focus on the factors stabilising and driving the formation of these structures.

3.2 Stabilising factors

As protein folds to minimise their Gibbs free energy, interactions are formed within the protein structure and between the protein and its solvent. As a result, of the folding sterically constrains are placed on the polypeptide. This destabilises the protein fold together with factors like desolvation of polar and charged. These destabilising effects must be compensated for by the formation of stabilising interactions. These interactions takes many forms and are described in following sections.

**BOX II: Definition of the hydrogen bond**

“**The hydrogen bond is an attractive interaction between a hydrogen atom from a molecule or a molecular fragment X–H in which X is more electronegative than H, and an atom or a group of atoms in the same or a different molecule, in which there is evidence of bond formation. A typical hydrogen bond may be depicted as X–H•••Y–Z, where the three dots denote the bond. X–H represents the hydrogen bond donor. The acceptor may be an atom or an anion Y, or a fragment or a molecule Y–Z, where Y is bonded to Z. In some cases, X and Y are the same. In more specific cases, X and Y are the same and X–H and Y–H distances are the same as well leading to symmetric hydrogen bonds. In any event, the acceptor is an electron rich region such as, but not limited to, a lone pair of Y or π-bonded pair of Y–Z.**” (Arunan et al. 2011)

3.3 Non-covalent interactions.

The non-covalent interaction constitutes the vast majority of intra molecular interaction in the folded protein. Non-covalent interactions are transient in contrast to covalent bonds. **Hydrogen bonds** are the most common interactions found. Hydrogen bonds stabilises the protein through bonds between adjacent polar atoms, e.g. Oxygen and Nitrogen, and hydrogen atoms in the protein structure, as it is the case in water and especially ice, see Figure 5. Hydrogen bonds can be classified as; sidechain-sidechain, sidechain-backbone, or backbone-backbone. Hydrogen bonding from backbone-backbone interactions account for 68% of all hydrogen bonds in proteins (Stickle et al. 1992). The majority of a protein structures are made up of repeater structure motifs, known as α-helices, β-sheets, loops and turns. Here backbone-backbone hydrogen bonds play an important role. An α-helix is formed by hydrogen bonds
between the peptide backbone of two amino acids separated by three amino acids and are hence local interactions, see Figure 6A. On an α-helix the side chain of each amino acid will protrude outwards from the centre of the helix. β-sheets on the other hand are formed by hydrogen bonds between the peptide backbone of amino acids which not necessarily are close to each other in the primary structure. β-sheets can consist of parallel and/or antiparallel strands, see Figure 6B & C. Between parallel β-strands each amino acid forms a hydrogen bond with two different amino acids on the opposing strand. In antiparallel β-strands the hydrogen bonds are formed between two amino acids one on each strand. In both cases each amino acid participates in two hydrogen bonds. The side chains of a β-strand alternate between pointing up and down from a horizontal backbone. Each hydrogen bond is estimated to contribute with 1-2 kcal/mol on average and is considered a weak interaction (Myers & Pace 1996). The maximum length of a hydrogen bond in protein structures are estimated to about 3 Å (Stickle et al. 1992)

Figure 5 Hydrogen bond network of water. The hydrogen atoms form covalent bonds with the oxygen atom. This results in a dipole that allows the oxygen atom to form hydrogen bonds with it remaining electron pairs to the hydrogen’s of other water molecules. The same is true for the hydrogen’s of the same molecule. In total four hydrogen bonds can be formed per water molecule.

Figure 6 Hydrogen bonding in Secondary Structure. A| Hydrogen bonds in an α-Helix, between amino acids with 3 intervening amino acids B| Hydrogen bonds in parallel β-sheets, 1 bond between 2 amino acids on opposite strands C| Hydrogen bonds in anti-parallel β-sheets, 2 bonds between 2 amino acids on opposite strands. Side-chains not shown.
**Ionic interaction** is non-covalent interaction between positively charged side-chains of lysine, arginine, and in some cases histidine, and the negative charge of the side-chains of aspartate and glutamate. Ionic interaction has been defined in 1983 as ion pairs, when Nitrogen or Oxygen atoms from two opposing charged side-chains or termini are within 4 Å of each other. It has also been show that at low salt concentration 0.01 M weak interaction below 0.2 kcal/mol can extent more than 10 Å (Lee et al. 2002). Ionic interactions can be further strengthened, by hydrogen bonding between the two charged side-chains, hence forming a salt-bridges see Figure 7, this combination of interaction result in rather strong interaction, 4 kcal/mol (Kumar & Nussinov 1999). The pure charged based ionic interaction is sometimes referred to as coulombic interaction. The idea that proteins are stabilised by ionic interactions is challenges by desolvation and entropy cost of the. This aspect of ionic interactions is discussed in more detail in section 6.3.

As ionic interactions are dependent of the charge of the residues, this type of interactions is sensitive to charge screening by ions in the solvent. As a result, the salt concentration of solvents can affect the stability of proteins. This effect can be both positive or negative depending of the thermodynamics of the ionic interaction e.g. screening an unfavourable charge thereby increasing the stability or vice versa. Additionally, as the charge of these amino acids side-chain are the result of their basic and acidic nature, the charges are dependent on pH.

![Figure 7 Example of a possible salt bridge between Arginine and glutamate.](image-url)
Van der Waals forces (VdW) are the weakest and has the shortest range of the non-covalent interactions. VdW interaction are the result of electrostatic fluctuations between atoms that comes in to close proximity, the resulting complementary electrostatic di-pols attracts the atoms to each other, see Figure 8. Hence stabilising the protein with 0.5-1 kcal/mol per atom pair (Berg et al. 2007). However, if the atoms come to close to each other the VdW switches to a repulsive force as the electron clouds of the two atoms collide.

Hydrophobic interactions are the last of the none-covalent interaction. In contrast to other none-covalent interaction hydrophobic interaction is not based on electrostatic attractions between the residues. In contrast, hydrophobic interactions are the result of entropy loss minimisation. The solvation of hydrophobic residues is the result of rearrangement of water molecules around the hydrophobic side-chain. As the water can’t form hydrogen bonds with the side-chains and only few orientations allows for optimal hydrogen bonding, a crystal-like shell is formed around the hydrophobic side-chain (Dias et al. 2010). This state change from bulk water to shell water is thermodynamically unfavourable. As, the enthalpy of the state change can be positive, zero, or negative, but the entropy will be negative and much larger that the change in enthalpy, $\Delta H < \Delta S$. Hence per Eq. 3 in BOX I this results in a positive $\Delta G$ of the state change. However, this unfavourable Gibbs free energy can be reduced by a seconded state change. Here the desolvation of the side chains, by collapsing them in upon them self thereby reducing their surface area and the amount of shell water results in a net increase of enthalpy and entropy.

As before the change in entropy is much larger than that of enthalpy. As a result, the change
in Gibbs free energy is negative and thermodynamically favourable. This minimisation of shell water and its high cost in entropy are the force that drives hydrophobic interaction. From Eq.3 we see that since the hydrophobic interactions are entropy driven, as \( \Delta S > 0 \) and \( \Delta H > 0 \), that hydrophobic interactions are strengthened by increasing temperatures. As mentioned above collapse of a hydrophobic core, see Figure 9, is believed to be an important pathway for protein folding, and is driven by this entropy minimisation. Hydrophobic interaction is however not limited to the folding of the protein it also plays an important role in complex formation (Byun et al. 2007).

3.4 Covalent interactions

Of all the interactions observed in proteins, covalent bonds between individual amino acids, excluding the peptide bonds, are the rarest. Only 6% or 34165 of the 552884 entries in SWISS-PROT\(^1\) are annotated with disulphide bonds. Which under normal conditions are the only covalent bonds found between residues in proteins. This kind of covalent bonds are considered strong interactions due to their covalent nature, the enthalpy of such a bond is 62 kcal/mol (Jursic 1996). These covalent bonds impose severe steric constrains on the peptide-chain. This greatly reduces their contribution to the stability to the protein fold, due to a large entropy cost of the bond formation. The Chain-entropy model suggest that disulphide bonds lower the entropy of none native conformation there by decreasing their stability (Pace et al. 1988) rather than directly increasing the stability of the fold. Experimental data suggest that disulphide bonds can be both stabilizing and disruptive of both the native and denatured

---

structure (Zavodszy et al. 2001). It is estimated that disulphide bond can contribute with as much as 2-4 kcal/mol (Betz 1993).

3.5 Post-translational modifications

Inter- and intra-molecular interactions are not the only factors that can affect protein stability. Post-translational modifications of the amino acid side-chains generate an additional diversity to the peptide chain otherwise not possible with the use of only 20 unique amino acids. These protein modifications can range from small modifications like addition of a methyl group to lysine to the addition of large branched carbon hydrates also known as oligosaccharides. This type of modification is known as glycosylation. For some proteins like the human hormone Erythropoietin (EPO) the oligosaccharides added to the protein constitutes 50% of the final mass (Yamaguchi et al. 1991). Protein glycosylation comes in three forms, N-, O-, and S-glycosylation, named after the glycosyl acceptor atom of an amino acid side chain, Nitrogen, Oxygen, and Sulphur (Stepper et al. 2011) respectively. Protein glycosylation are seen in all three domains of life, Archaea, Bacteria, and Eukarya (Wieland et al. 1985; Szymanski et al. 1999; Neuberger 1938) but most common in Eukaryotes. The chemical diversity of glycosylation is enormous, and estimated to consist of over 7000 unique structures as reviewed by R. D. Cummings in 2009 (Cummings 2009). Protein glycosylation has been show to play a part in many biological functions e.g. secretion, adhesion, function etc. Glycosylation has also been linked to protein folding and stabilisation. As N- glycosylation was showed to increase protein folding rates and stability (Hanson et al. 2009). It is theorised that the addition of the oligosaccharide to the poly-peptide decreases enthalpy of the unfolded state hence stabilising the protein fold (Shental-Bechor & Levy 2008). Both these studies show that the modifications are more important than the size of the oligosaccharide.

Methylation is like glycosylation best known from eukaryotes, and its role in epigenetic regulation of gene expression, through lysine methylation of histone and DNA condensation as reviewed by (Bannister & Kouzarides 2005; Kouzarides 2007). Lysine methylation has however, been suggested as a general mechanism for stabilisation of proteins. As Crenarchaea has been shown to have wide spread (Botting et al. 2010) and in some cases temperature dependent lysine methylation (Baumann et al. 1994). The implication of methylation on protein stability will be further discussed in section 6.4. The chemical diversity of methylation is small compared to that of glycosylation. This is due to the simple chemical composition of the methyl group. Both arginine and lysine subjects of methylation can be both mono- and di-
methylation while lysine also can be tri-methylated, see Figure 10. Lysine methylation is usually catalysed by a lysine methyl transferase, EC 2.1.1, through the conversion of S-Adenosyl methionine (SAM) to S-Adenosyl-L-homocysteine (SAH) and the simultaneously transfer of the methyl group from the SAM to the N-ε-amino group of the lysine side-chain.

Another common type of PTM is proteolytic cleavage of the peptide backbone, to produce truncated proteins. It is a well-known process with a few classic examples: The removal of the N-terminal methionine (Sherman et al. 1985), and the maturation of insulin (Docherty et al. 1982).

Beside these three types of modifications, are there countless other. From the dbPTM a data base of post translation modification there are 60² different types of side-chain modifications. These 60 types are in addition to glycosylation, lipidation and protein crosslinks. Covering the complete range is outside the scope of this thesis and will not continue beyond this point.

4 Protein structure determination

With all these different forces and interactions driving and stabilising the proteins folds, the task of modelling the structure of proteins has, so far in most cases, been out of reach. The scope to the problem was highlighted by Levinthal’s paradox in section 3.1. However, the structure of a protein contains vast amount of useful information. To access this information, we need a model of the structure. Currently, are there several ways to obtain structure information from proteins each with its own strengths and weaknesses. The most common methods used are: X-ray crystallography and Nuclear magnetic resonance spectroscopy (NMR).

X-ray crystallography is the most used method, 90%³ of structures in the Protein Data Bank (PDB) are obtained by X-ray diffraction. The main challenge in X-ray crystallography usually is obtaining the necessary high amount of protein and then producing crystals. Finding the proper condition for the formation of crystals can be very time consuming, as it is a process of trial and error.

NMR has been used for solving 9%³ of the structures in PDB and is the second most used method. It utilizes the magnetic moment or spin of specific elemental isotopes e.g. $^1\text{H}$, $^{13}\text{C}$ and $^{15}\text{N}$. As the natural abundancy of $^{13}\text{C}$ and $^{15}\text{N}$ are low and they are required for solving large structures, obtaining protein containing these isotopes relies on the expression of the protein from a host grown in a defined media where the heavy isotopes, $^{13}\text{C}$ & $^{15}\text{N}$, are the only Carbon and Nitrogen source. These media are expensive and not all organisms can be grown in these sufficiently defined media, also known as minimal media. In contrast to X-ray crystallography, NMR is limited to proteins of a size below 50 kDa.

In addition to classic experimental based methods like NMR and X-ray crystallography are the computational or Computer based methods. The rapid evolution of computers is pushing us towards the point where we can model protein folding. This emerging field is described in section 4.1.

Protein structures are an important tool in research fields. In medical research structural information can be used for rational drug design where information about protein-protein interaction or ligand binding sites can be instrumental in the development of inhibitors of virulence factors and oncogenes (Guillon et al. 2014; Lessene et al. 2013). Protein structures are also widely used in the study of protein stability. Here the structures provide information on how the different intra-molecular interactions come together and stabilises the protein fold. E.g. structural information of two homologues enzymes, one mesophilic and one thermophilic, can be compared and rationalising of differences in the structures can provide insight into the basis for the difference in stability. Additionally, Large scale structural analysis projects comparing structural tendencies of many proteins might provide information of general mechanisms utilised for achieving high protein stability. Protein stability modelling

make use of experimentally determined protein structures. For more information see the following section, 4.1.

4.1 Computational structural biology

Computational structural biology (CSB) offers an alternative to X-ray crystallography and NMR for determining the structure of proteins. In CSB computer modelling an All-atom Molecular Dynamic (MD) simulation are replacing empirical experimentation and data acquisition with computation to generate structures of folded proteins. Determining a protein structure purely in silico by MD relays on the knowledge of the factors and forces driving the protein folding for the simulation. As demonstrated by Leviathans paradox, see 3.1, there is a near endless number of possible structures for even short peptides. To do modelling protein folding from linear peptides through folding to a native structure is extremely calculative demanding, and a major limiting factor for the success of MD. The process of modelling protein folding from only a known amino acid sequence is known as “Ab initio” (Bonneau & Baker 2001) from Latin for “from the beginning”. To circumvent the high demand of calculation power needed to model protein folding Ab initio several methods has been developed that simplify the protein to reduce the variables, or use data from all ready solved structures to guide the modelling of novel structures. The following is an introduction of a few general methods for in silico determination of novel protein structures.

Homology modelling (HM) is one such “short cut” that employs sequence homology of the target protein to proteins with solved 3D structure. As a result, the final 3D model produced by HM is affected by the quality, resolution(Å), of the structures used for the modelling of the novel structure. HM can be simplified to three main steps (Khan et al. 2016): 1| Identification of a suitable template(s), which is done by searching databases for homologues peptide sequences with solved structures. An identity >40% are needed for a good result. 2| Model building, based on the structure of the template a new 3D model is built e.g. by using the peptide backbone as a guide. 3| Model evaluation and refinement. Here the thermodynamic plausibility of the model is tested, if the novel structure fails, then energy minimization algorithms are used to refine/tweak the novel structure until the result is thermodynamic favourable.

---

One method that uses a simplified model of the protein to estimate the structure for *Ab initio* structure modelling is **Lattice models**. Here the peptide chain is reduced to its simplest forms, ball and chain model and placed in a lattice structure, a systematic arrangement in 2D or 3D space (Dill et al. 1995). With this model, it is possible to model interactions based on sequence, not at atomic level but at a functional level, as each amino acid is treated at a single “atom”. In its simplest form, amino acids are reduced to hydrophobic (H) or Polar (P). The ball and chain model of the peptide can now be placed in its lattice grid (Figure 11) and each possible conformation can rapidly be evaluated through a simple sum of contacts between adjacent balls as each possible case is an assigned energy contribution, e.g. -1,0,0 for HH, HP, PP respectively. In this simplified model the most likely structure, based on hydrophobic interactions is the one with the lowest total energy. Bond angles and lengths, are defined by the lattice used in the simulation and not the properties of the actual bond. Lattice based models allows only the mapping of the peptide backbone in a rapid and calculation light manner compared to complete atomic modelling.

**Discrete state off-lattice models** are a simplified versions of MD, in contrast to true MD the complexity of these models are reduced (Thirumalai & Klimov 1999). Usually by limiting the sterically conformation, and bond lengths of the side chains, while the back-bone conformation is also limited to specific phi/psi angels. As a result, the degrees of freedom in the simulation is reduced and the calculation time for sampling the conformational space in reduced. In contrast to lattice based modelling this allows for the determination of secondary structure as well as higher order structure as the amino acid side are included in the modelling.

All-atom Molecular Dynamic (MD), the pinnacle of off-lattice models. Here the complexity of the model is complete, meaning that all atoms have their movement and interaction individually modelled. Theoretically this allows for the prediction of structures at all levels of complexity. As a result of the high complexity MD is very calculating heavy, especially as the number of atoms increases, and this limits it success. In an effort to overcome this problem the distributed computing project: Folding@home was started in 2000 by Vijay Pande out of
Standford University, currently ~85000 users provide ~77000 Teraflops\(^5\). This has allowed simulations to exceed the previously short time scales of nano- and micro-seconds to milliseconds (Voelz et al. 2010)

Beside the difference in complexity between lattice and off-lattice based models they differ significantly in how they work when they predict the structure. Off-lattice based models are time dependent simulations of the protein folding while lattice based models often are Monte Carlo based. Monte Carlo based models are dependent of random sampling, in this case sampling of protein conformations. The use of this method is only possible due to the reduced complexity of the models, otherwise Leviathan’s paradox would also apply to the simulation, however here the time scale will depend on the calculating power available. The simulation of protein folding in MD is done with a force field that also account for atomic momentum. As a result, it is possible to simulate changes in the structure as a function of time (Allen 2004).

**Structural fragment modelling**, contrast the other *Ab initio* methods by not using a simplified or constrained model to simulate the folding of the protein. Structural fragment modelling is a statistical based method. For this method, structural information, of a large set of non-homologues protein structures, are divided into fragments, and sorted into recurring structural elements. These recurring elements form a library of building blocks/motifs that can be applied to a peptide sequence to build a peptide back bone. Once a back bone has been built side chain conformation can be estimated by using preferred rotamer conformations databases (Baeten et al. 2008).

A common feature for all the methods described above is in the use of energy force fields/functions. Eq. 10 is an example of such an empiric force field use to calculate the Gibbs-free energy of folding (Schymkowitz et al. 2005). In all cases but MD they are used for evaluating the quality of the structure. In MD the force field directly involved in the simulation resulting in a structure.

\[
\Delta G = W_{vdw} \cdot \Delta G_{vdw} + W_{solvH} \cdot \Delta G_{solvH} + W_{solvP} \cdot \Delta G_{solvP} + \Delta G_{wb} + \Delta G_{hbond} + \Delta G_{el} + \Delta G_{Kon} + W_{mc} \cdot T \cdot \Delta S_{mc} + W_{sc} \cdot T \cdot \Delta S_{sc}
\]

Where \(W\) are weigh\(s\). \(vdw\): Van der Wells interactions, \(solvH\): solvation of hydrophobic residues, \(solvP\):solvation of polar residues, \(wb\): Water atoms with +2 Hydrogen bonds to the protein, \(hbond\): hydrogen bonds, \(el\): electrostatic, \(Kon\): electrostatic interaction from protein-protein interaction, \(mc\): entropy of peptide bond angles (phi-psi), \(sc\): entropy side chains

---

\(^5\) Fold@home homepage: [http://folding.stanford.edu/home/](http://folding.stanford.edu/home/) Date. 09-08-2016
The use of these force fields is not limited to modelling of *ab initio* structures. They also play an important role in *in silico* protein engineering, to evaluate the effect of mutations on the thermodynamic of protein folds, see section 5.2.

## 5 Protein engineering

While nature is vast in terms of diversity, most enzymes of not optimised for industrial processes. With protein engineering we can modify the amino acid sequences of proteins, thereby changing their functionality. The aim of these modifications are usually to improve the protein function in a specific context. Hence, stability, activity and substrate specificity are common targets. While the modifications of natural enzymes are useful, the design of novel enzymes engineered for specific applications provide near endless possibilities for enzyme applications. An enzyme has already been made that catalysed reactions not previously observed in nature (Siegel et al. 2010).

Protein engineering usually made by one of two methods or a combination of the two, directed evolution or rational design Both types of protein engineering will be introduced in the following section.

### 5.1 Directed evolution, traditional

Through selective breeding, many organisms have been adapted to our preferences e.g. higher yield, temperament, physical appeal, etc. Hence, once an understanding of the functionality of DNA was obtained the next step was the application of methods to speed up the evolutionary process.

Chemical mutagenesis paved the way for modern DNA based methods, by demonstrating the acquisition of advantageous phenotypes (Cobb et al. 2013). The idea behind direct evolution is the induction of random genetic changes, diversification, followed by a selective process where improved mutants are identified by screening. This process can be repeated in a circular fashion, to further increase (Eijsink et al. 2005). There are many methods by which genetic diversification can be obtained, chemical, irradiation, error prone PCR and other PCR based methods, gene shuffling, etc. (Cobb et al. 2013). The true power of direct evolution is the vast number of different mutants generated, thereby increasing the chance of wanted improvements. The effectiveness of directed evolution relies heavily on the availability of a high-throughput screening methods for the selection of advantageous mutants. The screenings
methods will most likely vary greatly from case to case. Directed evolution can be employed at organism and protein level, e.g. improving the growth rate of industrial relevant organisms on cheap mediums, or stability of an enzyme respectively.

Modern direct evolution usually follows a twostep approach, 1| library generation and 2| library screening (Cobb et al. 2013).

5.2 Rational design, modern

Rational design contrast direct evolution, by not being random and producing a comparable small number of mutants to screen. With rational design, knowledge about the target protein or organisms is used for the selection of changes to be made to achieve the sought-after effect e.g. specific amino acid substitution to increase enzyme stability (Wijma et al. 2014) or eliminate an unwanted metabolic pathway in improve yields (Sánchez et al. 2005).

The studies by (Wijma et al. 2014) is a good case story of the use of structural information and CSB. This also highlights the importance of structure knowledge for efficient rational design. Wijma and colleges used a combination of computational methods to evaluate amino acid substitution effect on protein stability, prediction of disulphide bonds, and MD simulation to eliminate mutations that result in increased protein stability. Furthermore, they employed a manual curation of the 248 point mutations, to eliminate any mutation with features know to be destabilising. In total, 64 point mutations were selected by experimental verification. Where 21 showed an increase in stability. In the end through combining the mutations an increase of $T_M$ of 35.5°C were obtained (Wijma et al. 2014). In the case highlight above the empirical force field, Eq. 10, were used to estimate the effect of amino acid substitutions of the stability of the protein structure.

To be able to predict changes to increase the stability of an enzyme is all well and good. And it can be highly successful as with the example from (Wijma et al. 2014) to generate a hydrolase with a temperature optimum 80°C and a long half-life at temperature over 70°C. However, evolution has produced organisms that lives comfortably at temperature above 80°C, and ever higher. The subject of the next sections is the adaptation of proteins to these extreme temperatures.
6 Protein adaptation to extreme temperatures

All the before mentioned factors that works to stabilise proteins are general for all proteins whether they originated from a mesophilic or a thermophilic organism, moderate and high temperature loving organisms respectively. As a consequence of the high temperature of their habitat, thermophilic organisms have evolved proteins with an intrinsic high thermal stability. This is a result of an otherwise unviable increase in protein metabolism that the organisms would have to maintain. In order to continually replenish its pool of active enzymes, if it enzymes had stability similar to those of mesophils. Generally speaking high thermal stability is associated with stability in general e.g. also towards denaturants such as detergents, urea and organic solvents and long shelf life of purified protein (Vieille & Zeikus 2001). An example of this is seen in the rational engineering to increase thermostability of a hydrolase from Bacillus stearothermophilus. Here the half-life of the enzyme at 70°C was improved with 12.7 fold, by a single point mutation. This increase in stability was accompanied by an increase in resistance to non-ionic detergents of ~2 fold (Ece et al. 2015). Due to the high stability and resistance to e.g. detergents of their enzymes, thermophiles and hyperthermophiles are currently experiencing an increase in interest, as they are potential host for enzymes of significant industrial interest. Beside the identification and characterisation of enzymes with high stability much effort has also been put into deciphering the molecular determinants of this increased stability. The hope is to identify general mechanisms that in time can be applied in protein engineering and cell factory optimisation to improve the stability of all enzymes, hence not relaying on the identification or random generation of new enzymes with high stability. Some investigations have looked at thermophilic organisms in general, to discern how thermophilic protein achieve their stability. While other have distinguished between thermophiles and hyperthermophiles or bacteria and archaea, in order to filter out convergent evolution that might obscure individual mechanisms.

A consensus is emerging where no single mechanism is responsible for the high stability of thermophilic proteins. Instead optimisations of many different interactions and mechanisms are responsible for the increased stability (Petsko 2001). In this section I will summarise what is known about these protein traits and their association with high stability and a correlation with high optimal growth temperature. Cellular traits like DNA composition and condensation, plasma membrane chemistry, and specific enzyme activities will not be discussed here.
6.1 **Amino acid composition**

As we progress further into the post genomic age, and more genomes of thermophiles become available a logical avenue of investigation is to look for differences in the amino acid composition of mesophiles and thermophiles. The most significant tendencies that has come from these investigations indicates that thermophilic protein has an overrepresentation of charged amino acids (Szilágyi & Závodszyk 2000; Fukuchi & Nishikawa 2001; Nakashima et al. 2003; Suhre & Claverie 2003; Cambillau & Claverie 2000; Tompa et al. 2016; Taylor & Vaisman 2010).

An analysis of the localisation of the amino acids within the protein structure of 47 thermophilic, optimal growth temperatures between 65°C and 70°C, and 144 mesophilic proteins. Here they found an overrepresentation of charged residues on the surface of ~17% compared to the mesophilic proteins. This correlated with an 15% and 17% increase of charged residues in the interior and the general composition of the proteins, respectively (Fukuchi & Nishikawa 2001) indicating a general increase in charged residues. This increase in charged residues is compensated for by a general decrease of polar residues. For both thermophilic and mesophilic proteins ~45% of amino acids fall in to one of those two categories, polar or charged. Of the charged residues on the surface of thermophilic proteins, glutamate, arginine, and lysine where overrepresented while aspartate was underrepresented. This elimination of aspartate from the surface of the protein contrast the situation in the interior where aspartate is overrepresented compared to the three other residues (Fukuchi & Nishikawa 2001). An increase in the usage of branched hydrophobic residues was also observed, i.e. alanine elimination was compensated for primarily by an increase in valine but also leucine and isoleucine content (Fukuchi & Nishikawa 2001).

Many of the same tendencies have been reported in other studies, (Cambillau & Claverie 2000) compared mesophilic and hyper-thermophilic optimal growth temperatures between 83°C and 97°C proteins. (Szilágyi & Závodszyk 2000) has the same general results but fail to identify an increase in lysine usage of thermophilic enzymes of organisms with a growth temperature below 100°C. (Nakashima et al. 2003) were able to demonstrate an almost linear correlation between optimal grown temperature and the amino acid composition, in the proteome of organisms living between 10 and 103°C. The major positive contributors to this correlation were charged and hydrophobic amino acids.
To summarise, the charged residues, arginine, lysine and glutamate, and the branched hydrophobic, leucine, isoleucine, and valine have been positively selected for in thermophilic proteins. While aspartate together with the polar residues have been counter selected. Rationalisation of these differences along with other traits related to stability of thermophilic proteins will be highlighted in the following sections.

6.2 Counter selection of thermolabile amino acids

The above mentioned counter selection of aspartate, asparagine, and glutamine in thermophilic enzymes might be due to the thermolability of these amino acids, as reviewed in (Daniel et al. 1996). At high temperatures asparagine and glutamine are labile to heat induced deamination, while aspartate and asparagine also can participate in heat induced hydrolysis of the peptide back-bone. However, these temperature-driven degradations only become significant at temperatures above 80°C at neutral pH. The thermolability of these amino acids can however be reduced or prevented by their local environment e.g. sterically hindrance of the formation of intermediates, stabilisation by hydrogen bonding, or the presence of co-factors like Ca⁺ ions.

6.3 Increased number of charged amino acids

The increase in the number of lysine, arginine, and glutamate residues in thermophilic proteins, can be an indication of an increased number of ionic interaction and salt-bridges, that serve to stabilise thermophilic proteins. This correlation has been observed in some cases but is not universal (Xiao & Honig 1999). The stabilising effect of ionic interaction is challenged by a high cost upon formation, due to desolvation (Hendsch & Tidor 1994), and loss of entropy as a result of increased sterically movement (Sun et al. 1991). The formation of a salt-bridges can compensate for the cost by the formation of hydrogen bonds between the charged groups. The stabilising effect of salt bridges were evaluated in a study containing 222 salt-bridges from 36 proteins, primarily from mesophilic organisms. This showed that the majority of the tested salt bridges were stabilising (Kumar & Nussinov 1999). 32 of the 222 salt bridges or 14% were determined to be destabilising the protein fold. 30% of the salt bridges were buried and 88% of these were deemed to be stabilizing. And once the desolvation cost was payed, the average salt bridge contributes to the stability of the folding with 4 kcal/mol, but much higher contribution is also possible. In an addition study an inverse correlation between temperature and desolvation cost of charged residues, thereby making salt bridges contribute more to the stability at high temperatures and more advantages for
thermophiles then mesophiles (Elcock 1998). These observation correlates with an increase in the number of salt bridges in thermophilic proteins compared to mesophilic (Kumar et al. 2000).

Optimisation of the ionic interaction have been show to play a role in the stabilisation of thermophilic enzymes, the optimisation of ionic interaction through the elimination of unfavourable repulsion between residues with the same charge, has been proposed as a relative simple mechanism, compared to the introduction of additional stabilising ionic interaction (Perl & Schmid 2001). As the elimination of unfavourable repulsion aren’t associated with the same desolvation and entropic cost as the formation of ion-pairs and salt-bridges (Perl & Schmid 2001). The optimisation of these interactions was also seen when the comparing the contribution to stability of ionic interaction from mesophilic a thermophilic enzymes, however there were many different methods to achieve these optimisations, compensation of desolvation by polar interaction with additional non charged residues, increase in the number of ionic interaction, reduction of buried charged groups (Xiao & Honig 1999). The localisation of ionic interaction on the surface of thermophilic enzymes were found to be an identifiable trait of thermophilic enzymes (Taylor & Vaisman 2010), also indicating differences in the utilisation of ionic interactions.

6.4 Lysine methylation:

Lysine methylation has been suggested to increase the stability of protein by strengthen ionic bond formation, and preventing aggregation (Febbraio et al. 2004). Proteins from hyperthermophiles archaea of the order Sulfolobales show extensive lysine methylation (Maras et al. 1992; Botting et al. 2010; Guo et al. 2008; Baumann et al. 1994; Febbraio et al. 2004). In the case of sul7d, a histone like protein found in Crenaraeota, the degree of lysine methylation seemed to be dependent of the growth temperature (Baumann et al. 1994). The observed temperature dependent methylations seemed to be independent of the amino acid sequence around lysine’s in α-helix’s (Febbraio et al. 2004; Botting et al. 2010). These observations are in good correlation with the identification of the first crenarchaeal lysine methyl transferase, aKMT, which activity appears to be unspecific but tending towards α-helixs (Chu et al. 2012; Niu et al. 2013). The widespread lysine methylation and the unspecific nature of the aKMT, contrasts the previously observed specific methylation observed in bacteria and eukaryotes. aKMT is responsible for the vast majority of lysine methylation in *S. islandicus* (Chu et al. 2016). This has led to the hypothesis that the lysine methylation might be related to increased
thermos ability. The research presented later in the thesis show that in vivo lysine methylation has a positive, if minor, effect on the stability of thermophilic esterase EstA from *Sulfolobus islandicus* (Stiefler-Jensen et al. 2016). These data correlate with recent in vitro experimental evidence suggesting a correlation between stability and lysine methylation (Xia et al. 2015). Here in vitro lysine methylation was reported to improve the stability and activity of MCM from *S. islandicus* when purified from *E. coli*. In contrast β-Glycosidase from *S. Sulfolobus* showed no correlation between stability and lysine methylation (Moracci et al. 1995). Additionally, it has been shown that aKMT mediated lysine methylation is dispensable for growth at temperatures up to 75°C on optimal growth media (Chu et al. 2016). All the evidence above suggests that lysine methylation in *S. islandicus* only play a minor role in heat tolerance.

6.5 **Structural optimisation**

Various structural optimisations have been linked to the high stability of thermophilic proteins. These include reduced flexibility and minimisation of the protein structure along with cavity filling. The reduction of cavities within the protein structure has in some cases been demonstrated to be associated with increased stability (Vieille & Zeikus 2001; Szilágyi & Závodszky 2000; Rhee et al. 2006; Ece et al. 2015). A reduction of cavities in thermophilic proteins was indicated to be facilitated by an increase in amino acid with extended side chains (Tompa et al. 2016). Reducing the flexibility of the protein structure is believed to be a mechanism employed to increase protein stability of thermophiles (Scandurra et al. 1998). A general reduction in protein flexibility comes with a cost, the activity of enzymes is often depended upon conformational changes e.g. upon substrate binding and release. If this flexibility is not maintained to some degree at physiological temperatures the effectiveness of the enzyme will be compromised. The negative effect of low flexibility can be observed through the low activity of thermophilic enzymes at low temperatures (Závodszky et al. 1998). It has been proposed that, to counteract this negative effect on activity the reduction of flexibility is limited to structural areas that do not interfere with activity, this mechanism is observed both for thermophilic and mesophilic enzymes (Karshikoff et al. 2015).

Reduction of the peptide chain and shortening of surface loops has also been proposed as a method for increasing thermostability (Chakravarty & Varadarajan 2000; Thompson & Eisenberg 1999; Taylor & Vaisman 2010; Walden et al. 2001). Preferences in secondary structure related to increased stability in not clearly evident. Some data do suggest a weak
correlation between an increased presence of α-helix and β-strands and a reduction of irregular structures (Szilágyi & Závodszky 2000; Chakravarty & Varadarajan 2000).

6.6 Hydrogen bonds
A 2016 study did not find a significant difference in the amount of hydrogen bonding observed in a set of 373 thermophilic and mesophilic proteins (Tompa et al. 2016). It was however shown that hydrogen bonds were utilized very differently by the two classes of proteins. Thermophilic hydrogen bonds were to a higher degree used for anchoring of the protein surface to the interior. In contrast mesophilic proteins utilised hydrogen bonding for stabilising the protein core. An earlier study, on a much smaller dataset, also concluded that there was no difference in the number of hydrogen bonds. It did however, find indications that hydrogen bonding tended to be better optimised in thermophilic proteins, as unsatisfied hydrogen bond donor and acceptor minimisation seemed correlated to optimal growth temperature (Szilágyi & Závodszky 2000).

6.7 Hydrophobicity
Optimization of hydrophobic interactions has long been considered a trait of thermophilic proteins (Perutz & Raidt 1975). Examination of protein structures of both thermophilic and mesophilic proteins show that most thermophilic proteins had a better minimization of apolar atoms accessible to the solvent (Spassov et al. 1995). In a later study it was shown that thermophilic proteins are more stabilised by desolvation of side chain carbon, minimisation of unfavourable apolar interaction, atoms than homologues mesophilic proteins (Tompa et al. 2016). As well as a stronger hydrophobic core (Taylor & Vaisman 2010). That hydrophobic interactions are better optimised and more widespread in thermophilic protein correlates well with the increased strength of hydrophobic interactions at high temperatures (Baldwin 1986; Shimizu & Chan 2002).

6.8 Increased number of disulphide bonds
Disulfide bonds are generally associated with extracellular proteins due to the reducing nature of the cytoplasm. Nevertheless, it seems that thermophiles, especially Archaea, make use of a disulphide bond in cytoplasmic proteins. This is evident both in the frequency of cysteines pairs in close proximity and the presence of genes that promote cytoplasmic formation of disulphide bonds (Beeby et al. 2005; Jorda & Yeates 2011). Both native and engineered
disulphide of cytoplasmic proteins have been shown to be enhancing protein stability (Hattori et al. 2015; Yin et al. 2015; Cao et al. 2015).

6.9 Oligomerisation

Oligomerisation has been mentioned as a possible mechanism for increasing protein stability (Walden et al. 2001; Tanaka et al. 2004; Korkhin et al. 1999; Villeret et al. 1998). It is however not necessarily detectable in studies with large numbers of proteins (Kumar et al. 2000). This indicates that oligomerisation likely is a high specialised mechanism for thermophilic adaptation.

The stabilising mechanism of oligomerisation is very similar to that of hydrophobic interactions, see section 3.3. As water is perturbed from is bulk state, with only water to water hydrogen bonds, to form a hydration shell around a protein, where water to protein hydrogen bonds are formed, result in an increase in Gibbs free energy of the shell water due to loss of entropy, $S_{\text{bulk}} > S_{\text{shell}} \rightarrow 0 > \Delta S_{\text{bulk} \rightarrow \text{shell}}$, and a thermodynamic unfavourable state, $\Delta G_{\text{bulk} \rightarrow \text{shell}}$.

As the amount of shell water is determined by the solvent accessible surface area (SASA) a reduction of this surface will, like it was the case for hydrophobic interaction, result in an increase in entropy, $0 < \Delta S$, and a more favourable thermodynamic state. We know from Eq. 3 that for a state change there is an inverse correlation between $\Delta S$ and its contribution to the stability of the state change as well as a temperature dependency, $-T\Delta S$. Hence, as for hydrophobic interaction the stabilising effect of reducing the SASA is strengthened by increasing high temperatures.

Comparison of mesophilic and thermophilic protein-protein interfaces have shown that additional hydrophobic and polar interaction stabilises the interface while the number of ionic interaction are more or less similar between the two groups of proteins (Maugini et al. 2009). It has also been observed that cross-linking of the protein-protein interface via disulphide bonds increase the stability of the proteins (Cao et al. 2015; Stiefler-Jensen et al. 2016).

7 Enzymes of special interest

In the following two section, two enzymes of special interest will be introduced, as they both play a central role in the research presented in the last part of this thesis. Both enzymes are native to the hyper-thermophilic and acidophilic crenarchaeal archaea *Sulfolobus islandicus*. The first enzyme an esterase(EstA), was first characterised in 2012 by Mei et al. 2012. The
results from that study made the foundations for this thesis. The second enzyme a lysine methyltransferase (aKMT/aKMT4), was first identified as the first archaeal methyltransferase in 2012.

7.1 **EstA, an Archaeal esterase**

EstA, WP_014511906, from *S. islandicus* is an esterase with a classic alpha/beta hydrolase fold\(^6\), with a central twisted \(\beta\)-sheet of 8 strands supported by \(\alpha\)-helixes on both side (Ollis et al. 1992), see Figure 12. The EstA protein consists of 305 amino acids in its native form as encoded by the gene SIRE_RS01465. Or when expressed recombinantly as a 314 amino acid large fusion protein with a C-terminal 6x histidine tag. EstA catalyses the hydrolysis of a carboxylic ester through the consumption of a single water molecule, as per EC3.1.1.1, see Figure 13.

In 2012, EstA were reported to be exceptional stable both in regards to thermal and organic solvents inactivation (Mei et al. 2012). At 90\(^\circ\)C the recombinant enzyme remained active after more than 24 hours’ thermal treatment, with a half-life of 16 hours, when expressed from its native host, the hyperthermophilic Archaea *S. islandicus* (SiEstA). SiEstA purified as a dimer in solution, and had a temperature optimum at 90\(^\circ\)C, well above the growth temperature of *S. islandicus* at 80\(^\circ\)C. The high stability of SiEstA was also characterised by the presence of dimers during SDS-PAGE analysis. The activity of EstA was only reduced with up to 20\% in the present of 5 mM 1,4-dithiothreitol (DTT), Urea, and mercaptoethanol. Diaminoethane-tetraacetic acid

---

\(^6\) Marchler-Bauer A et al. (2015), "CDD: NCBI's conserved domain database.", *Nucleic Acids Res.* **43**(D)222-6
(EDTA) at 5 mM reduced that activity to 58% while phenylmethanesulfonyl fluoride (PMSF) was the strongest inhibitor of SiEstA and reduced activity with 90%. SiEstA was able to hydrolyse p-Nitrophenyl ethylester with ethyl chain length from two to 16 carbon atoms. The highest activity was observed with a six carbon ethyl chain, p-Nitrophenyl Hexanoate.

In the Mei et al. paper from 2012, recombinant EstA from E. coli (EcEstAI) was also analysed. In all aspects EcEstAI preformed worse than SiEstA, except in regards to resistance to Urea and 1 mM EDTA. Here EcEstAI showed similar performance to SiEstA. The optimal temperature of EcEstAI was determined to be 65°C, and the half-life at 90°C were 30 minutes, 32 fold lower than that of SiEstA. EcEstAI purified as a monomer contrasting the dimer formation by SiEstA. In the paper the differences in thermostability were hypothesized to be a result of different or absent PTM when EstA is purified from E. coli. The stability of esterase of other thermophiles and hyperthermophiles has been the subject of investigation a number of times in the past and reviewed in 2009 by (Levisson et al. 2009). In general, the stability varies a lot, half-life of esterases from the genus Sulfolobus range from approximately 40 to 6000 minutes at 80°C, half-lives of 50 hours at 90°C has also been reported. The closes homology to EstA investigated previously, with a 99% identityscore, had a half-life of 40 min at 80°C and close to 10 minutes at 90°C, when expressed from E. coli in a multimeric form (Kim & Lee 2004). Lipases and esterases share many of the same properties, the major difference is in the ability of lipases to utilize substrates not in solution. Both classes of enzymes catalyse the hydrolysis of ester bonds through similar mechanisms. Carboxylesterases are currently best described by their amino acid sequence and are divided into three families: LPL-, EST-, and HSL-family, denoted by L, C, and HSL respectively, EstA is a member of the HSL family. Historically lipases and esterase’s has been classified by the substrate specificity, however the difference is more related to the physical state of the
substrate, soluble/insoluble (Chahinian & Sarda 2009). Lipases/esterase’s have an important place in industrial use. Esterase’s are used in multiple steps during the production of recycled paper, and in the biosynthesis of organic compounds, the production of wines and juice in the food industry (Panda & Gowrishankar 2005; Hasan et al. 2006; Damhus et al. 2008).

7.2 aKMT, an Archaeal lysine methyltransferase

In 2012, the first lysine methyltransferase from a crenarchaeal, were identified and characterised from S. islandicus (Chu et al. 2012), see Figure 15 for the structure. The protein had a novel N-terminal catalytic domain, and was named aKMT for Archaeal lysine MethylTransferase, with the one-letter abbreviation of lysine: K in place of L. aKMT is encoded by the gene SiRe_1449, and consist of 161 amino acid and has a molecular mass of 18.2 kDa. aKMT methylates a broad range of proteins from Sulfolobus, when purified both from the native host and from E. coli. bovine serum albumin (BSA) was however not methylated by aKMT. Four month later in 2013 a different group reported the identification of aKMT through sequence similarities with aKMT4/Dot1 methyltransferase from eukaryotes and named it aKMT4 (Niu et al. 2013). They demonstrated aKMT/aKMT4 auto-methylation activity, which regulate its own activity. Auto-methylation was inhibited by the presence of DNA. The sequence unspecificmethylation of aKMT/aKMT4 were confirmed in a study from 2016, see Figure 16A, however a bias towards α-Helix structures were observed, see Figure 16B. The same study also investigated the effect of aKMT/aKMT4 deletion on the proteome of S. islandicus and found no correlation aKMT/aKMT4 mediated methylation and increased thermostability (Chu et al. 2016).
8 Summary of results

The aim of this project was to investigate, and if possibly, identify post-translational modifications and/or other factors, not related to the primary structure, that are responsible for the high stability of proteins expressed from the thermo- and acido-phlic crenarchaeal organism *S. islandicus*.

As a representative of the proteome of *S. islandicus* an esterase EstA, recently characterised and shown to possess host dependent structural stability, i.e. 32 fold reduced thermos stability when expressed in the mesophilic host *E. coli* were used.

As an initial step to gain insight into the post-translational modifications were the determination of the crystal structure of EstA when expressed from its native host. This was done in collaboration with Kaare Teilum group from the Structural Biology and NMR Laboratory of University of Copenhagen. The work was led by PhD. Student Troels Schwarz-Linnet. The project resulted in a structure of EstA in a double dimer conformation at a resolution of 2.6 Å, covering residue 2 to 303 of the 314 amino acid large his-tagged recombinant EstA. The structure has been submitted to The Worldwide PDB\(^7\), as 5LK6 where it awaits publication. The crystal structure did not contain any discernible post-translational modification or co-factors of EstA.

8.1 Optimisation of esterase assay

In order to assure optimal data acquisition the assay used previously to study EstA, (Mei et al. 2012), were optimised. This resulted in a robust and minimalistic assay, with a minimum of substrate auto-hydrolysis. The conditions for data collection were also improved by eliminating light scattering. The performance of the buffer at high temperatures, 90°C were also improved.

The assay has been used in volumes ranging from 250 µl to 1 ml and is suitable for real-time data acquisition in a temperature range from 25-90°C, by directly measuring the substrate turnover by absorption at 410 nm.

\(^7\) [http://wwpdb.org/](http://wwpdb.org/)
Reaction mix: 50 mM NaH\(_2\)PO\(_4\), 25 mM Sodium lauroyl sarcosinate (Sarkosyl), 10% v/v 2-propanol, 1 mM p-Nitrophenol, 0.02 µg/ml EstA. As an alternative to sarkosyl can 1% Triton X-305, Triton X-305 is a non-ionic detergent, hence less disruptive of the protein structures.

8.2 Expression of soluble dimeric EstA for *E. coli* host

One of the previously difficulties encountered in expression EstA from *E. coli* was its apparent insolubility when expressed in this host. To overcome this EstA were previously expressed and purified from *E. coli* as a fusion-peptide of EstA and thioredoxin (Trx). This resulted in difference in amino acid sequence of the mesophilic and thermophilic expressed EstA, even after proteolytic removal of the Trx-tag. Additionally, the enzyme purified from *E. coli* was monomeric and not dimeric as when purified from the native thermophilic host. In order to assure a proper comparison between the 2 expression systems, *E. coli* and *S. islandicus*, the differences between the enzymes were eliminated. Result in expression of EstA from both meso- and thermo-philic hosts with sequences identical to the dimeric recombinant EstA, previously purified from *S. islandicus* (SiEstA). This resulted in the purification of dimeric EstA from *E. coli* (EcEstA) by expression for 2 hours at 30 or 37°C by induction with 1 mM IPTG. In general, 0.9 mg/l EstA were obtained at a concentration of ~300 µg/ml.

8.3 Lysine methylation as a factor in protein stability

The initial hypothesis that lysine methylation might be a factor that positively affected the stability of EstA were investigated through 2 experimental paths; 1| Comparison between recombinant EstA expressed from, a mesophilic host, *E. coli*, without any cytoplasmic lysine methylation, and the native thermophilic host, *S. islandicus*, with cytoplasmic lysine methylation. 2| Comparison between EstA expressed from the native host, *S. islandicus*, with and without cytoplasmic lysine methylation. Both experimental paths compared methylated protein from *S. islandicus* with a none-methylated variant while the second path also eliminated host dependent differences.

The methylation states of EstA were confirmed by mass spectrometry analysis. All three variants showed similar temperature optimums at 75°C and half-life at 90°C of approximately 30 minutes. Kinetically the *Sulfolobus* expressed variants of EstA were similar while EstA from purified from *E. coli* showed a reduced kinetics. From our *in vitro* stability analysis of EstA with and without *in vivo*, methylation it is evident that lysine methylation plays a small
role in stabilising the enzyme. The methylation is however not a major determinator for thermostability. The differences in half-life of methylated and none-methylated EstA purified from the native host is only 1.08 fold in favour of the methylated enzyme. In addition, the kinetic of the enzyme does not seem to be affected by the methylation when the enzyme is expressed from the native host.

8.4 **Investigation mesophilic expressed EstA increased sensitivity to detergents.**

It was observed that EstA purified from *E. coli* had an increased sensitivity towards detergents, as the presence of 25 mM sarkosyl resulted in a 2.5 fold reduction in half-life at 90°C. Anion-exchange chromatography analysis did not yield any information about any possible factors responsible for the reduced half-life. Isoelectric focusing gel electrophoresis showed that EstA purified from *E. coli* was more homogenic compared to EstA purified from *S. islandicus*, regardless of methylation state, indicating a possible link between protein charge and detergent resistance and possibly the reduced kinetic of EcEstA.

8.5 **The Effect of dimer integrity on the stability of EstA**

Several mutations in EstA’s dimer interface were made in an effort to generate monomeric EstA, to determine that stabilising effect of the oligomeric state of EstA. None of the mutations converted EstA to a monomer, however amino acid substitution of phenylalanine 269, located in the dimer interface, severely compromised the stability of EstA.

In the reverse experiment, stabilisation of the dimer interface successful through the substitution of S295 to cysteine thereby generating an intra molecular disulphide bond. The introduction of this novel cysteine bridge increased the kinetic stability of EstA resulting in an increase of EstA half-life from 12.3 to 20 minutes at 90°C. Thereby demonstrating a stabilising effect of dimer formation of EstA. Amino acid substitution of S266 & N270 and G261 to cysteine and aspartic acid, respectively. In contrast to the substitutions of hydrophobic residues in the dimer interface the introduction of an inter-subunit di-sulphide bond did not affect the activity of the mutated esterase.

9 **Conclusion & Perspective**

From the data presented in the following chapters and highlighted above we can see that in the case of the hyperthermophilic esterase from *S. islandicus*, EstA, PTM and more specifically Lysine mono methylation, do not play a critical role in the stability of hyper-
thermophilic enzymes from crenarchaeal. It is however, evident that the esterase is stabilised by the methylation. That Lysine methylation not by itself provides a high degree of stability to crenarchaeal enzymes is not surprising. As a conglomeration of small optimisations, that are more easily acquirable overtime is the more likely explanation for the development of hyper-stable enzymes.

If this unspecific lysine methylation serves additional biological functions, beside the increase it does provide to the stability, this function remains unknown. As in the case of this esterase the methylation is not causally linked to changes in enzyme activity.

Even though the positive effect of this unspecific lysine methylation only is minimal, it would be interesting to investigate if this protective effect can be transferred.

It has already been shown that the methyltransferase aKMT4 from *S. islandicus* is active and able to methylate co-expressed *Sulfolobus* enzymes in *E. coli*. However, to my knowledge it has not been tested whether these enzymes gain any stabilising effect from the *in vivo* methylation in mesophilic host. Additionally, it would be interestingly to test whether enzymes from other thermophilic organisms, lacking this kind of lysine methylation can benefit from lysine methylation as EstA.

The best-case scenario is that enzymes co-expressed with aKMT4 gain some stability, hence engineering production stain expressing aKMT4 could improve enzyme stability in general.

It is known that expression of thermophilic and hyperthermophilic enzymes from *E. coli* often is associated with loss of stability, and true in the case of EstA. Usually the lack of prober PTM or co-factors is the expected cause. In our investigation, we did not find any evidence that pointed towards differences in modification of EstA expressed in *E. coli* or the native host lacking *in vivo* methylation. However, their stability and sensitivity to ionic detergents are vastly different. The only differences we could find were that enzyme expressed from the native host had a more heterogeneous charge distribution, while *E. coli* expressed EstA, were homogenous. The underlying cause for the observation are still unknown. Deamidation is a possible explanation.

During an investigation of the importance of dimer formation of the stability of EstA, were we able to improve EstA stability through the introduction of a novel dimer spanning disulphide bond. This single modification of the enzyme increased that stability of mesophilic esterase by 1.62 fold. Whether this effect will persist when the mutant is expressed from *S. islandicus*, remains to be tested. In any case, this result demonstrates the potential stabilising
effect of oligomerisation and the possible benefits of targeting protein interaction interfaces when engineering enzymes for stability.

Finally, the optimisation p-nitrophenol based activity assays for esterase activity can be helpful in future investigations of thermophilic esterases.

In addition to the data presented in this thesis, I would like to point out some observation on protein purification of thermophilic and hyper-thermophilic proteins. Usually it is suggested to add between 150-500 mM NaCl to IMAC buffers to avoid ionic interaction with the column. However, with the presence of an increased number of charged residues on the surface of these types of proteins, and an expectation that they provide stabilising ionic interaction. Having a high ion strength in the buffer can disrupt the positive interaction provided by the charges residues, there by potentially reducing the stability and solubility of the protein. a problem extents to the concatenation of imidazole as it under normal operation condition carry’s a positive charge. As the argument for using buffers contain a certain ionic strength is valid and the use of imidazole very convenient for IMAC purification. I would recruitment a buffer exchange to a buffer with low ionic strength, e.g. 50 mM Na/K-phosphate or similar, after the protein is purified. Personally, this approach has in two cases rectified problem with protein precipitation after purification.
10 References


Hanson, S.R. et al., 2009. The core trisaccharide of an N-linked glycoprotein intrinsically accelerates folding and enhances stability. *Proceedings of the National Academy of Sciences of the United States of America*, 106(9), pp.3131–6. Available at:


Sánchez, A.M., Bennett, G.N. & San, K.-Y., 2005. Novel pathway engineering design of the anaerobic central metabolic pathway in Escherichia coli to increase succinate yield and


Stiefler-Jensen, D., Teilum, K. & She, Q.-X., 2016. Effect of dimer interface integrity on the


Villeret, V. et al., 1998. The crystal structure of Pyrococcus furiosus ornithine carbamoyltransferase reveals a key role for oligomerization in enzyme stability at extremely high temperatures. Proceedings of the National Academy of Sciences of the United States of America, 95(6), pp.2801–6. Available at: http://www.pnas.org/content/95/6/2801.long.

Voelz, V.A. et al., 2010. Molecular simulation of ab initio protein folding for a millisecond


Chapter 1

In this Chapter is the first of three-presenting research conduction during my PhD study. Here we present an investigation into post-translation modifications role for the stability of hyperthermophilic enzymes.

The manuscript is still in preparation, but in the final state before submission to Protein science as a full-length Article of maximum 5000 words.
The extraordinary thermal stability of EstA from *S. islandicus* is independent of post translational modifications

Daniel Stiefler-Jensen²⁺⁺, Troels Schwarz-Linnet¹⁺, Casper de Lichtenberg¹, Tam T. T. N. Nguyen³, Kasper D. Rand³, Li Huang⁴, Qunxin She² and Kaare Teilum¹⁺⁺⁺

¹Structural Biology and NMR Laboratory, Department of Biology, University of Copenhagen, Denmark
²Archaea Centre, Department of Biology, University of Copenhagen, Denmark
³Department of Pharmacology, University of Copenhagen, Denmark
⁴State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China

*Correspondence: kaare.teilum@bio.ku.dk
⁺These authors contributed equally to this work.

Abbreviation:
1,4-Dithiothreitol 
Circular dichroism spectroscopy 
Electrospray ionization mass spectrometry 
EstA expressed from *S. islandicus* 
EstA expressed from *S. islandicus* with knock out of *akmt* 
EstA expressed from *E. coli Shuffle T7* 
Isoelectric focusing 
Isopropyl β-D-1-thiogalactopyranoside 
N-lauroylsarcosine 
Optical density at 600nm 
Polyacrylamide gel electrophoresis 
Post-translation modifications 
Size exclusion chromatography 
Thioredoxin

DTT
CD
ESI-MS
SiEstA
ΔSiEstA
EcEstA
IEF
IPTG
Sarkosyl
OD₆₀₀
PAGE
PTM
SEC
TrxA

Abstract

Enzymes from thermophilic and hyper-thermophilic organisms have an intrinsic high stability. Understanding the mechanisms behind their high stability will be an important tool for the engineering of novel enzymes with high stability.

Histone like proteins, sul7d, from the hyperthermophilic and acidophilic archaea have been shown to have both unspecific and temperature dependent lysine methylation. Here we test the effect of eliminating this *in vivo* lysine methylation on the stability of an esterase (EstA) purified from the native host *S. islandicus* as well as the effect of expression in a mesophilic host, *E. coli*, without the machinery for *in vivo* lysine methylation. We find that lysine mono methylation indeed has a significant positive effect on the stability of EstA, although the effect is small.

However, the effect is secondary to that of expression in *E. coli*, as the enzyme expressed in *E.
coli is compromised both on stability and activity which we cannot contribute to a covalent difference.

**Introduction**

High stability and activity are key requirements for the use of enzymes in most industrial applications.\(^1\) Extremophilic organisms are natural reservoirs for enzymes with high stability, and many enzymes with high stability characterized from extremophiles show potential as biocatalysts in industrial processes.\(^1,2\)

An esterase (EstA) isolated from the hyperthermophilic and acidophilic crenarchaeal organism *Sulfolobus islandicus* was recently characterized.\(^3\) High thermostability and the resistance to solvents makes EstA a good candidate for industrial exploitation. As protein production from the native host is expensive, and cumbersome due to its extreme growth conditions, alternative and well-established mesophilic expression organisms, such as *E. coli* are more attractive for enzyme production.\(^4,5\) Intriguingly, when EstA was purified from the native host, the half-life of the enzyme (in terms of catalytic activity) was 16 hours at 90\(^\circ\)C, whereas for the same protein recombinantly expressed in *E. coli*, the half-life reduced 32-fold.\(^3\) EstA expressed in *E. coli* was also more sensitive to organic solvents than EstA expressed in the native host.

The increased stability of EstA expressed in *S. islandicus* was suggested to be a result of post-translational modifications which are not present in the protein when it is expressed in *E. coli*. Mono methylation of the \(\varepsilon\)-amino group of lysine have been shown to be extensive and sequence unspecific in *Sulfolobus* and other archaea.\(^6,10\) In the case of the Sso7d a DNA binding protein from *S. solfataricus*, the methylation seemed to be dependent of the growth temperature of the native host.\(^7\) In addition, *in vitro* methylation of recombinant MCM purified from *E. coli* was linked with increased stability and activity.\(^11\) Lysine methylation of \(\beta\)-Glucosidase from *S. Sulfolobus* did however, not show any stabilizing effect.\(^12\)

Recently, a lysine methyl transferase, aKMT4, with broad substrate specificity from *S. islandicus* was characterized.\(^13,14\) Deleting the gene for aKMT4 in *S. islandicus* significantly reduces the level of methylated lysine, but only moderately impairs the growth of *S. islandicus*.\(^15\) The identification of this lysine methyl transferase opens an opportunity of investigating the role of lysine methylation on the stability of proteins from creanarchaeal archaea independent of mesophilic expression systems.
In this work, we have been interested in understanding the underlying reason for the previously observed difference in the stability of EstA expressed in *E. coli* and *S. islandicus*. Our hypothesis was that lysine methylation occurring in *S. islandicus* conferred stability to the enzyme. We therefore expressed and purified EstA from wild-type *S. islandicus*, *S. islandicus*(Δakmt) and *E. coli*, termed SiEstA, ΔSiEstA and EcEstA respectively. All three variants were characterized with respect to covalent modifications, enzyme kinetic activity and thermal stability. We find a large difference in the enzymatic activity and thermal stability between EstA expressed in *E. coli* and in *S. islandicus*. In contrast, there is only a minimal difference in the properties of EstA from the two *S. islandicus* strains. Our results thus show that the methylation of lysine only has a small influence on the stability of EstA and thus the difference in both activity and stability must be a result of non-covalent interactions stabilizing the wildtype enzyme from *S. islandicus*.

**Materials and methods**

*Sulfolobus* strains, growth condition, and transformation.

The *Sulfolobus* strains were propagated in a salt and vitamin base medium supplemented with 0.2% (w/v) vitamin free casamino acids (Difco Vitamin Assay, BD), 0.125% yeast extract (CVy*). As carbon source, 0.02% (w/v) sucrose were used for regular growth (SCVy*), and 0.02% (w/v) D-arabinose (Sigma-Aldrich, St. Louis, MO) (ACVy*) for expression cultures. *S. islandicus* REY15A E233S, previously transformed with pSe-estA, were used as a thermophilic host for the expression of EstA, (EstA from this host is abbreviated SiEstA). *S. islandicus* REY15A E233S with an additional knockout of the gene coding for aKMT/aKMT4, SiRe_1449, was transformed with pSe-estA, following procedures previously described, with the following modifications: After electroporation the cells were inoculated at 65°C in 227 mM (NH₄)₂PO₄, 28.7 mM K₂SO₄, 13.4 mM KCl, 93.2 mM glycine. Competent cells were prepared by growth at 70°C in sucrose media without yeast extract (SCV). EstA expressed from host line is abbreviated ΔSiEstA

*E. coli* vector modification

Based on a previous vector (p32a-estA) for expression of EstA in *E. coli* as a 50 kDa fusion protein with an N-terminal thioredoxin (TrxA)³, a new vector construct for the expression of EstA in *E. coli* with the exact same sequence as the EstA expressed in *S. islandicus* was prepared by (Figure 1). EstA expressed in *E. coli* is called EcEstA.
Expression and purification

SiEstA and ΔSiEstA were purified from 1 litre cultures of \textit{S. islandicus} and \textit{S. islandicus} (Δakmt), respectively, grown at 75°C in ACV\textsubscript{y}* until OD\textsubscript{600} ~0.8. The cells were harvested by centrifugation at 12000 rpm for 10 minutes at room temperature (centrifuge 5810R, Eppendorf AG, Hamburg). EcEstA were purified from 2 litre \textit{E. coli} SHuffle T7 grown at 30°C in LB medium supplemented with 100 µg/ml of ampicillin and 10 µg/ml chloramphenicol. Protein expression was induced by 1 mM of Isopropyl β-D-1-thiogalactopyranoside (IPTG) at OD\textsubscript{600} ~0.8 and allowed to proceed for 2 hours. The \textit{E. coli} cells were harvested by centrifugation at 12000 rpm for 10 minutes at 4°C (centrifuge 5810R, Eppendorf AG, Hamburg). EstA expressed in \textit{E. coli} is abbreviated EcEstA.

Both \textit{S. islandicus} and \textit{E. coli} cells were resuspended in immobilized metal ion affinity chromatography (IMAC) binding buffer supplemented with 1 µg/ml DNasel. The cells were lysis by three passes through a french press at 1500 psi. The cell extract was clarified by centrifugation at 12000 rpm for 20 minutes (centrifuge 5810R, Eppendorf AG, Hamburg) at room temperature or 4°C. The \textit{E. coli} supernatant was additionally heat precipitated in Falcon Tubes in a water bath at 70°C for 10 minutes, followed by an additional clarification, as above. Supernatants was filtered by 0.2 µm before chromatography.

The proteins were purified by IMAC using a HisTrap HP 1 ml column (GE Healthcare Life Science) with a 20-500mM imidazole gradient in pH 7.4, 1.37 mM NaCl, 27 mM KCl, 100 mM NaH\textsubscript{2}PO\textsubscript{4}, 18 mM KH\textsubscript{2}PO\textsubscript{4}. The main peak, at approximately 200mM imidazole, was collected before buffer exchange to 50 mM K-phosphate, pH 7.4. SDS-PAGE was used to determine the purity of the final EstA sample. Enzyme concentration was estimated by measuring absorbance at 280nm with an extinction coefficient 31900 M\textsuperscript{-1} cm\textsuperscript{-1} and a mass of 34.5 kDa.

Native PAGE analysis

6% Native PAGE was used to analyse heat treated enzyme. 50 µg/ml enzyme in 50 mM KH\textsubscript{2}PO\textsubscript{4}, pH 7.4 and 25 mM sarkosyl, were incubated at 90°C for up to 1 hour while shaking. The heat treatment was quench by dilution, 1:10, with the same buffer at room temperature. 10 µl or 50 ng enzyme were loaded to the native PAGE. The gel was stained with Coomassie Blue.

Size exclusion chromatography analysis

100 µl of purified EstA were run at 750 µl/min in 50 mM NaH\textsubscript{2}PO\textsubscript{4}, 150 mM NaCl, pH 7.8 on a Superdex 200 10/300 GL. As a standard of molecular mass 100 µl of the following proteins were used: Ribonuclease A 1.5 mg/ml, Ovalbumin 2 mg/ml, and Conalbumin 2 mg/ml.
Mass spectrometry

LC-MS was performed using a NanoAQUITY UPLC setup coupled to an ESI Synapt G2 Q-TOF mass spectrometer (Waters, Milford, USA). Protein samples (50 pmol) were loaded onto the UPLC system, then trapped and desalted on a C4 trap column (ACQUITY UPLC Protein BEH C4 VanGuard Pre-column, 300Å, 1.7 µm, 2.1 mm x 5 mm column, Waters, Milford, USA) for 3 min at 40 µL/min of 99% of mobile phase A (0.23% Formic Acid). Isocratic elution from the C4 column into the mass spectrometer was performed at a flow-rate of 40 µL/min of 95% mobile phase B (ACN, 0.23% Formic Acid). Mass spectra were acquired in positive ionization mode over an m/z range of 50-2000 and a scan time of 1 second. Each sample was run in duplicate. The acquired mass spectra were deconvoluted using the MaxEnt1 algorithm of MassLynx (Waters, Milford, USA). The three charge states with the highest intensity were used to estimate the average Full Width at Half Maximum. The deconvolution resolution was set to 0.1 Da.

Circular dichroism spectroscopy

Far- and near-UV spectra of the esterases in 50 mM K-phosphate, pH 7.4 at 25°C were recorded with a resolution of 0.2 nm on a Jasco spectropolometer J-810 equipped with a PTC-423s peltier element.

Esterase activity assay

Esterase activity was determined as the rate of hydrolysis of p-nitrophenyl esters by following the formation of p-nitrophenolate by the change in absorbance at 410 nm. The activity was assayed in 50mM Na-phosphate, 25 mM N-lauroylsarcosine (Sarkosyl) where indicated, 10% (v/v) 2-propanol, pH 7.0 with 20 ng/ml EstA. The substrate was 1 mM p-nitrophenyl hexanoate unless otherwise stated. The change in absorbance was measured with an Infinite M200 (Tecan Trading AG, Switzerland) plate reader at 40°C except for the temperature-activity profile which was measured in a Jasco spectropolometer J-810 with PTC-423s peltier element.

Isoelectric focusing analysis

Isoelectric focusing analysis was made with pH 3-10 Criterion IEF™ Gel (BIO-RAD Laboratories), and IEF Anode and Cathode and loading buffers (BIO-RAD Laboratories). 125-200 µg/ml enzyme were used and the analysis was made following the recommendation from the manufacture. The enzyme was visualized by Bio-Rad Silver stain (BIO-RAD Laboratories) by following the recommended protocol.
**Anion Exchange chromatography**

A MonoQ HR 5/5 column was used for anion exchange chromatography in 50 mM bis-tris, pH 6.5 with a gradient from 0 to 500 mM NaCl. Elution made by a shallow gradient, 0.5 %/min.

**Results**

A pET32a based vector (p32a-estA) for expression of EstA in *E. coli* as a 50 kDa fusion protein with an N-terminal thioredoxin (TrxA) and an S-tag has previously been constructed.³ Proteolytic cleavage with enterokinase releases EstA in a form, which is different at both the N- and the C- termini compared to the EstA expressed from pSe-estA in *S. islandicus*. At the N-terminal in the *E. coli* expressed EstA, Ala-Met-Ala replace the starting Met and at the C-terminal a linker sequence and a His-tag (DKSAAALEHHHHHH), which differs from the linker and His-tag used in the pSe-estA by the additional Leu and Glu residues underlined in the sequence above. A new vector construct for the expression of EstA in *E. coli* with the exact same sequence as the EstA expressed in *S. islandicus* was prepared to eliminate any difference between primary structure of the proteins purified from *S. islandicus* and *E. coli*. The sequence coding for EstA in the expression vectors resembles that of wild-type EstA (Uniprot: F0NDQ1) with the addition of a C-terminal extension with the sequence Ala₃His₆.

SiEstA was purified from *S. islandicus* REY15A E233S.¹⁸ Additionally, EstA was purified as ΔSiEstA from *S. islandicus* REY15A E233S Δakmt with a knockout of the gene SiRe_1449 coding for a lysine methyl transferase.¹⁵ *E. coli* SHuffleT7 which is engineered for improved formation of disulphide bond formation, was used for the expression of EcEstA.

The final samples of all three variants of EstA gave rise to a single band of expected 35 kDa in SDS-PAGE (Figure 2). Size exclusion chromatography shows similar retention volumes for all EstA variants with an estimated size slightly above 75 kDa corresponding to EstA being a dimer in solution (Figure 3).

**Covalent characterization**

To identify any post-translational modifications, we characterized the covalent structure of the different variants of EstA. As *S. islandicus* is able to mono-methylate the ε-amino group of lysine residues we used amino acid analysis to quantify the amount of mono-methylated Lys in SiEstA (Table 1). On average 0.4 of the 14 Lys-residues in SiEstA are mono-methylated. Mass spectrometry confirmed the presence of multiple methyl groups on SiEstA (Figure 4). For non-methylated EcEstA, the experimentally determined mass is 34367.5 Da, corresponding well with an accuracy of 0.7 Da or 20 ppm to the predicted mass of EstA without the start codon
methionine (34368.8 Da). For methylated SiEstA, besides a species with mass corresponding to EstA (34366.6 Da), additional species (34381.0 and 34395.3 Da) were observed in the deconvoluted mass spectrum (Figure 4). The mass differences of 14.4 and 28.7 Da of these species compared to the EstA correspond to the addition of one or two methyl groups, respectively. A minor component corresponding to the addition of three methyl groups can also be observed a low intensity.

EstA has three Cys residues in the sequence. The redox state of the Cys residues were analysed by SDS-PAGE in the absence and presence of DTT (Figure 5). Under reducing conditions the protein migrates slightly longer than under non-reducing conditions. EstA purified from E. coli migrate exactly the same as when it is purified from S. islandicus. These results demonstrate that EstA forms an intramolecular disulphide bond and since the protein migrates as a monomer under non-reducing conditions in SDS-PAGE the EstA dimer is not linked by a disulphide bond.

To assess if there are any differences in the non-covalent structure we compared both near- and far-UV CD spectra of EcEstA and SiEstA (Figure 6). The spectra are identical demonstrating that there are no differences in the secondary structure (near-UV CD) or in environment of the aromatic residues (far-UV CD).

Enzymatic activity

Following the previous report about the difference in temperature optimum for catalysing the hydrolysis of phenyl esters we compared the temperature-activity profiles of the EstA variants. In contrast to the previous report on EstA, the temperature optima of all the variants are approximately 75 °C (Figure 7A), which is close to the optimal growth temperature of S. islandicus at 80 °C. At all temperatures the measured activity of EcEstA, however, is about half of those measured for the EstA variants expressed in S. islandicus.

To compare the catalytic characteristics of the three EstA variants we complemented the temperature-activity profiles with an Michaelis-Menten analysis of the substrate dependence of the catalytic activity. The analysis was performed at 40 °C as the kinetic assay at this temperature can be performed a in 96-well plate format, which is considerably more simple to handle experimentally than the assays carried out in single cuvettes at elevated temperatures. p-nitrophenylhexanoate starts precipitating above 1 mM and it was therefore not possible to reach substrate saturation. The two variants of EstA expressed in S. islandicus have almost the same kinetic parameters whereas both k_{cat} and K_{M} for EstA expressed in E. coli are significantly reduced (Figure 7B and Table 2).
Thermostability of EstA

Attempts to measure the thermodynamic stability of EstA using chemical denaturants failed as the protein aggregates when it unfolds, which renders the folding irreversible. Temperature unfolding of EstA is not possible either as the unfolding transition occurs above 95 °C both for the protein expressed in *E. coli* and in *S. islandicus*. However, incubation of EstA at 90 °C leads to loss of enzymatic activity (Figure 7C&D). We therefore used the half-life of the catalytic activity after incubation at 90 °C as a measure of the thermostability of EstA. In phosphate buffer at pH 7 the half-life of the different EstA preparations are 33.9 min for SiEstA, 31.4 min for ΔSiEstA and 26.5 min for EcEstA. Although small, the differences are statistical significant (ANOVA, p < 0.05). Addition of 25 mM sarkosyl, a strong anion surfactant, amplifies the differences. The half-life of the activity of SiEstA increases slightly to 36.3 min while it is reduced to only 10.5 min for EcEstA (Figure 7D). In sarkosyl the half-life of ΔSiEstA is slightly reduced relative to SiEstA. Native-PAGE analysis of the esterase under these conditions shows changes in the electrophoretic mobility for EcEstA occurring much faster than for SiEstA (Figure 8). Whereas the methylation state of EstA only has a slight effect on the thermostability, the effect of the expression host is much more pronounced.

As we do not see any difference in the mass from ESI-MS of EcEstA and ΔSiEstA but still has a significant difference in enzymatic activity and thermostability a possible explanation could be that the protein expressed in *S. islandicus* at 75 °C was deamidated, which is a post translational modification not detectable by mass spectrometry of the intact protein. We therefore analysed the EstA preparations by anion exchange chromatography and isoelectric focusing to identify any difference in the net charge of the protein. The chromatograms from the anion exchange show that the major fraction of EstA elutes at the same concentration of NaCl for all three variants (Figure 9A). SiEstA and to some extent ΔSiEstA have additional peaks eluting at higher concentrations of NaCl. This heterogeneity of the protein purified from *S. islandicus* is also seen in IEF-PAGE (Figure 9B). Measuring the activity and thermal stability of EstA in main peaks eluting from the anion exchange column results in the same half-life of the activity as seen before. Thus, although there are clear differences in both the elution profile from the anion exchange column and in the IEF-PAGE profiles, the fractions of SiEstA that elutes at the same NaCl concentration as EcEstA still has a longer half-life of the catalytic activity and a difference in charge cannot explain the difference in activity.

**Discussion**
The thermostability of EstA depends on the organism in which it is expressed. Such behaviour is well known for proteins that become post-translationally modified. In particular glycosylations are known to increase the kinetic stability and lower the aggregation propensity. Our initial hypothesis about the poor stability of EstA expressed in *E. coli* reported previously was therefore that the native EstA was post-translationally modified, and that the lack of this modification destabilized the protein. Expression of identical constructs of EstA in *E. coli* and *S. islandicus* shows that the EstA expressed in *E. coli* is indeed lacking post-translational methylation of lysine residues. These modifications were recently found in a proteomic analysis of aKMT4 targets in *S. islandicus* and shown to occur on four lysines (K8, K21, K50 and K138). In agreement with our observation that on average only 0.4 lysine residue are methylated per EstA molecule, the modification of the four lysine residues were found to be incomplete. Comparing SiEstA and EcEstA the only covalent difference, which we have been able to find, is that lysine residues in SiEstA are methylated. Using *S. islandicus* REY15A E233S Δakmt as expression host also resulted in EstA that was unmethylated and the covalent structure of this variant of EstA is identical to that of EstA expressed in *E. coli*. It is therefore puzzling that the catalytic activity and the thermal stability of the two unmethylated variants of EstA differs. Indeed, EstA expressed in and purified from the two different strains of *S. islandicus* have much more similar properties.

One explanation for this behaviour is that some unidentified co-factor only present in *S. islandicus* stabilize EstA. We therefore tested if EstA purified from *E. coli* could gain (some of) the thermal stability that SiEstA possesses by incubating EcEstA with *S. islandicus* lysate. After purification from the lysate EcEstA retains its reduced stability (data not shown). Thus, the difference in stability remains elusive.

Our results show that the effect of lysine methylations on protein stability, at least of EstA, is small, whereas some other effect of the expression host is much more important. First of all, these observations demonstrate that proteins from hyperthermophilic organisms advantageously may be expressed in a hyperthermophilic organism. If the factor responsible for the increased stability of EstA also have an effect on a mesophilic protein is hard test by expression in *S. islandicus*, as the melting temperatures of most mesophilic proteins are below the growth temperature of *S. islandicus*. Recombinant expression in *S. islandicus* is therefore often only feasible for thermophilic proteins.
The observation that lysine methylation only has a minor effect on EstA stability adds to the speculation about the physiological role of lysine methylation. The most prevalent hypothesis is that lysine methylation increases the protein stability, and is in part based on the difference in stability of a β-glucosidase from *S. solfataricus* purified from the natural source and the protein recombinantly expressed in *E. coli*. It is not known if the underlying reason for the increased stability of this β-glucosidase is the same as for the increased EstA stability, but it will be of interest to determine if the stabilizing effect we observe on EstA is general or specific to EstA.

**References**


12.


**Figures**

*Figure legends*

**Figure 1** Schematic representation of EstA constructs. A) Previous construct used for expression in *E. coli* (EcSisEstAII) and in *S. islandicus* (SiEstA). B) Difference in protein sequence between EstA produced from EcSisEstAII and from SiEstA. The sequence of EstA produced from the new *E. coli* construct EcEstA is identical to that produced from SiEstA.

**Figure 2** SDS-PAGE analysis of purified EstA. The samples were run on a 12% acrylamide/bis-acrylamide gels.

**Figure 3** Size exclusion chromatography. The retention volumes of the three EstA variants SiEstA, ΔSiEstA, and EcEstA are similar and correspond dimers. The samples were run on a Superdex200 10/300 GL column in 50 mM NaH2PO4, 150 mM NaCl at pH 7.8. The elution
volumes of a set of standard proteins used for calibration of the column are also shown. The dashed line is a linear fit to the data for the standard proteins.

**Figure 4** Mass analysis of SiEstA and ΔSiEstA by ESI MS. (A) Representative mass spectrum of SiEstA (A) and ΔSiEstA (C) and the deconvoluted spectrum of SiEstA (B) and ΔSiEstA (D).

**Figure 5** Analysis of disulfide bond formation by SDS-PAGE analysis. Samples of purified SiEstA and EcEstA were prepared in the presence or absence of 50 mM DTT as indicated.

**Figure 6** Circular dichroism spectra of SisEstA (red) and EcEstA (blue). Far-UV (top) and near-UV spectra (bottom) were recorded at pH 7.4 and 25 °C. The far-UV spectra were recorded at approximately 0.2 mg/ml EstA in a 1 mm cuvette and the near UV spectra at approximately 7 mg/ml EstA in a 10 mm cuvette. The measured signal was normalized to the concentration and plotted as the molar ellipticity.

**Figure 7** Activity bases characterisation of EstA variation. The activity of SiEstA (red square), ΔSiEstA (green circle), and EcEstA (blue triangle) were measured by following the rate of formation of p-nitrophenylate produced by the hydrolysis of 1 mM p-nitrophenyl-hexanoate by 20 µg/ml enzyme unless otherwise indicated. A) Temperature-activity profile of EstA. The assay was performed at varying temperatures in 50mM Na-phosphate, 25 mM N-lauroylsarcosine, 10% 2-propanol, pH 7.0. B) Michaelis-Menten kinetics of EstA’s by the hydrolysis of varying concentrations of p-nitrophenyl-hexanoate as incated. The assay was performed at 40 °C in 50mM Na-phosphate, 25 mM N-lauroylsarcosine, 10% 2-propanol, pH 7.0. C) Residual activity of EstA’s. The activity of the EstA’s were measured after incubated at 90°C for 5-120 min at 90 °C in 50 mM NaH₂PO₄, pH 7. D) Residual activity of EstA’s. The activity of the EstA’s were measured after incubated at 90°C for 5-120 min at 90 °C in 50mM Na-phosphate, 25 mM N-lauroylsarcosine, 10% 2-propanol, pH 7.0

**Figure 8** Native-PAGE analysis of temperature incubated EstA. Prior to the electrophoresis, SiEstA and EcEstA were incubated for the indicated amount of time at 90 °C in 50 mM NaH₂PO₄, 25 mM Sarkosyl, pH 7.

**Figure 9** Anion exchange & IEF analysis of SiEstA (red), ΔSiEstA (green), and EcEstA (blue). A) Anion exchange chromatography of each of the three EstA variants as eluted from a MonoQ
HR 5/5 column by a 0.5%/min NaCl gradient from 0 to 500 mM. In 50 mM bis-tris buffer at pH 6.5. B) Isoelectric focusing gel electrophoresis of the three EstA variants, separated on 3-10 pH gradient.
**Figure 1**

![Diagram showing protein structures](image)

EcEstA II: 50.45 kDa  
SiEstA: 34.51 kDa

EcEstA I : N-Ala-Met-Ala-EstA-Ala-Ala-Ala-Leu-Glu-6xHis-C  
SiEstA/EcEstA : N-EstA-Ala-Ala-6xHis-C

**Figure 2**

![Western blot image](image)
Figure 5

Figure 6
Figure 7

A. Activity (AU/Sec) vs. Temperature (°C)

B. Activity (AU/Sec) vs. Substrate conc. (mM)

C. Product Yield vs. Time at 90°C (min)

D. Product Yield vs. Time at 90°C (min)

Figure 8

Gel electrophoresis images showing the effect of time at 90°C on the activity of SIEsIA and EIEsIA.
### Table 1 Amino acid analysis of SiEstA

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>mol %</th>
<th>Number of amino acid residues&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>9.9</td>
<td>31.0</td>
</tr>
<tr>
<td>Arg</td>
<td>5.1</td>
<td>16.0</td>
</tr>
<tr>
<td>Asp</td>
<td>10.9</td>
<td>34.4</td>
</tr>
<tr>
<td>Cys</td>
<td>1.0</td>
<td>3.2</td>
</tr>
<tr>
<td>Glu</td>
<td>7.4</td>
<td>23.2</td>
</tr>
<tr>
<td>Gly</td>
<td>8.5</td>
<td>26.6</td>
</tr>
<tr>
<td>His</td>
<td>2.5</td>
<td>7.8</td>
</tr>
<tr>
<td>Ile</td>
<td>6.6</td>
<td>20.7</td>
</tr>
<tr>
<td>Leu</td>
<td>10.2</td>
<td>32.0</td>
</tr>
<tr>
<td>Lys</td>
<td>4.2</td>
<td>13.1</td>
</tr>
<tr>
<td>MeLys</td>
<td>0.14</td>
<td>0.4</td>
</tr>
<tr>
<td>Met</td>
<td>0.7</td>
<td>2.3</td>
</tr>
<tr>
<td>Phe</td>
<td>5.5</td>
<td>17.4</td>
</tr>
<tr>
<td>Pro</td>
<td>6.1</td>
<td>19.2</td>
</tr>
<tr>
<td>Ser</td>
<td>6.9</td>
<td>21.5</td>
</tr>
<tr>
<td>Thr</td>
<td>2.0</td>
<td>6.4</td>
</tr>
<tr>
<td>Tyr</td>
<td>4.4</td>
<td>13.8</td>
</tr>
<tr>
<td>Val</td>
<td>7.7</td>
<td>24.2</td>
</tr>
</tbody>
</table>

<sup>a</sup>Calculated for a total of 313 amino acid residues
<table>
<thead>
<tr>
<th>Variant</th>
<th>$K_M$ [mM]</th>
<th>$k_{cat}$ [10$^2$ s$^{-1}$]</th>
</tr>
</thead>
<tbody>
<tr>
<td>SiEstA</td>
<td>0.66 ±0.04</td>
<td>192 ± 6</td>
</tr>
<tr>
<td>ΔSiEstA</td>
<td>1.00 ±0.05</td>
<td>229 ± 7</td>
</tr>
<tr>
<td>EcEstA</td>
<td>0.42 ±0.02</td>
<td>100 ± 2</td>
</tr>
</tbody>
</table>

Table 2: Enzyme kinetic parameters for EstA variants at 40°C
Chapter 2

In this manuscript, the importance of dimer formation on the stability of the hyper-thermophilic esterase EstA is investigated.

To facilitate the in silico modelling presented in this manuscript a custom python script was written by myself for the automation of the modelling. The modelling itself was made with FoldX4, http://foldxsuite.crg.eu. In appendix I you can find an introduction to the functionality of the script and a guide for use, as well as the python code.

The manuscript is still in preparation, and written as a full-length article of maximum 5000 words for Protein science.
Effect of dimer interface integrity on the stability of a hyperthermophilic esterase

Daniel Stiefler-Jensen¹, Kaare Teilm², and Qunxin She¹,*

¹Archaea Centre, Department of Biology, University of Copenhagen, Denmark
²Structural Biology and NMR Laboratory, Department of Biology, University of Copenhagen, Denmark

*Correspondence: Qunxin@bio.ku.dk

Abbreviations

Amino acid AA
Imobilized metal ion affinity chromatography IMAC
Isopropyl β-D-1-thiogalactopyranoside IPTG
N-lauroylsarcosine Sarkosyl
Protein Interaction Atom Distance Algorithm PIADA
Proteins, interfaces, Structures and Assemblies PISA
Protein Structure and Interaction Analyzer PSAIA
Site-directed ligation independent mutagenesis SLIM
Solvent accessible surface area SASA
Thioredoxin Trx
Van der Waals VdW

Abstract

Many factors influence the stability of hyper-thermophilic enzymes by different mechanisms, and protein oligomerisation is one of them. Oligomerisation increases enzyme stability by reducing the solvent-exposing surface area through protein-protein interaction, presumably by an entropy driven release of hydration water. Here we investigate the importance of dimer interface integrity for the stability of an esterase from the hyper-thermophilic and acidophilic archaea Sulfolobus islandicus (EstA). We found that compromising the interactions at the dimer interface by alanine substitution of hydrophobic residues present in the core of the interface drastically reduced the stability of mutant EstA proteins although these proteins still formed dimer at ambient temperatures, suggesting that the protein stability could be correlated with the strength of inter-subunit interactions. Indeed, strengthening the interaction at the dimer-interface by generating a novel inter-molecular disulphide bond resulted in an increase in protein stability at 90°C by 1.63 fold.

Introduction
The hunt for enzymes with the highest possible stability is an ever-ongoing process in the biotech industry. One way to obtain enzymes with a higher stability is to utilise directed or rational evolution of mesophilic enzymes to make them more thermophilic.\textsuperscript{1,2} (Hyper)Thermophilic organisms that lives at temperatures above 50°C are another source of thermostable enzymes.\textsuperscript{3} Their enzymes have evolved to cope with the high temperatures of their host organism’s natural environment and are therefore intrinsically stable. \textit{Pyrococcus furiosus}\textsuperscript{4} and \textit{Pyrolobus fumarii}\textsuperscript{5} are both examples of hyperthermophilic organisms isolated from environments where the temperature are up to and above 100°C. From organisms like these, there is a potential for isolation of enzymes with high stability. For example, A β-glucosidase from \textit{P. furiosus} showed an optimal temperature of 105°C and a half-life of 13 hours at 100°C\textsuperscript{6}. Beside the isolation of thermophilic enzymes for their use in the biotechnology industry\textsuperscript{7}, they are also studied to discern the mechanisms that provide their high stability. So far research into these mechanisms has yielded many clues as to what makes thermophilic enzymes as stable as they are. However, no general mechanism has yet, if it ever will, been identified. The information accumulated so far, points to a wide range of optimisation that converge to contribute to the high stability of thermophilic enzymes\textsuperscript{8}. These include, among others, increased number charged residues on the protein surface, an under representation of heat liable amino acids, void filling, rigidity, volume reduction and oligomerisation\textsuperscript{3,9–12}.

Oligomerisation, the formation of stable transient, protein-protein interaction is a possible mechanism for protein stability. By engaging in protein-protein interaction, the overall solvent accessible surface area (SASA) are reduced. The reduction in SASA provides a stabilising effect by reducing unfavourable interaction between water and the protein surface. Comparison of protein-protein interfaces from mesophilic, thermophilic enzymes, has shown that thermophilic interfaces are more hydrophobic and rely less on hydrogen bonding, compared to mesophilic interfaces, while the number of ionic interaction are similar.\textsuperscript{13} The stabilising effect of protein oligomerisation have been experimentally observed on several occasions\textsuperscript{14–16}.

One such case, is an esterase from the hyper-thermophilic and acidophilic archaea \textit{Sulfolobus islandicus} (EstA). Here a large difference in stability was observed between monomeric and dimeric version of EstA\textsuperscript{17,18}, suggesting the dimerization provides a stabilising effect on the esterase. Both versions of EstA has been reported to have similar, but reduced enzyme kinetics when expressed in \textit{E. coli}\textsuperscript{17,18}. Hence the reduction in the enzymatic kinetics was an effect of the expression in \textit{E. coli} and not an effect of the enzyme being monomeric. Consequently, EstA
dimer formation is not linked to function, and the dimer formation of EstA likely serve an alternative purpose e.g. stabilisation of the protein structure. As the catalytic activity is not affected by the oligomeric state of the esterase, this enzyme is a good model for studying the stabilising effect of protein oligomerisation in hyper-thermophilic enzymes.

In this study, we investigate the effect of dimer interface integrity on the stability of the hyper-thermophilic esterase EstA of *Sulfolobus islandicus*. The hypothesis is that weakening or disruption of the intra-molecular interaction will result in a reduced stability, while strengthening these interactions will improve the stability of the esterase. To accomplish this, we used *in silico* and structural analysis to pinpoint target for mutagenesis to disrupt the dimer and structural inspection for identification of likely targets for stabilising mutations, then constructed the mutants and characterised their stability *in vitro*.

**Material and methods**

*Cell lines and vectors and enzyme preparation*

All expression vectors were cloned and purified from *E. coli* DH5α. For expression and purification of protein *E. coli* SHuffleT7 was used. Expression of wildtype EstA was done from a pET32A based, p32a-estA_v2, vector containing the complete reading frame of EstA with a C-terminal 6x histidine tag. Mutants of and EcEstA were purified from 2 litre *E. coli* Shuffle T7 grown at 30°C in LB medium supplemented with 100 µg/ml of ampicillin and 10 µg/ml chloramphenicol. Protein expression was induced by 1 mM of Isopropyl β-D-1-thiogalactopyranoside (IPTG) at OD₆₀₀ ~0.8 and allowed to proceed for 2 hours. The *E. coli* cells were harvested by centrifugation at 12000 rpm for 10 minutes at 4°C (centrifuge 5810R, Eppendorf AG, Hamburg). EstA expressed in *E. coli* is abbreviated EcEstA.

Both *S. islandicus* and *E. coli* cells were resuspended in immobilized metal ion affinity chromatography (IMAC) binding buffer supplemented with 1 µg/ml DNaseI. The cells were lysed by three passes through a french press at 1500 psi. The cell extract was clarified by centrifugation at 12000 rpm for 20 minutes (centrifuge 5810R, Eppendorf AG, Hamburg) 4°C. The supernatant was heat precipitated in water bath at 70°C for 10 minutes, followed by an additional clarification, as above. Supernatants of the heat precipitation was filtered through a 0.2 µm filter before chromatography.
The proteins were purified by IMAC using a HisTrap HP 1 ml column (GE Healthcare Life Science) with a 20-500mM imidazole gradient in pH 7.4, 150 mM NaCl, 50 mM NaH₂PO₄. The main peak, at approximately 200mM imidazole, was collected before buffer exchange to 50 mM Na-phosphate, pH 7. SDS-PAGE was used to determine the purity of the final EstA sample. Enzyme concentration was estimated by measuring absorbance at 280nm with an extinction coefficient 31900 M⁻¹cm⁻¹ and a mass of 34.5 kDa.

**SLIM Mutagenesis**

Mutation of the estA gene on p32a_estA_v2 was made the Site-directed ligation independent mutagenesis (SLIM). ¹⁹,²⁰ The mutagenesis was performed as described in.²⁰ with the following modifications: template removal by DpnI treatment was performed after the SLIM hybridisation. The PCR was performed by an initial denaturing step at 98°C for 30s. Followed by 35 cycles of 98°C for 30s, 50°C for 30s, 72°C for 6 minutes. And a final elongation step of 72°C for 10min concluded the reaction. The PCR was made with 2U of Phusion Hot Start II DNA Polymerase (Thermo Fisher Scientific, USA). See Table I for primers used.

**Size exclusion chromatography analysis**

100 µl of purified EstA were run at 750 µl/min in 50 mM NaH₂PO₄, 150 mM NaCl, pH 7.8 on a Superdex200 10/300 GL. As a standard of molecular mass 100 µl of the following proteins were used: Ribonuclease A 1.5 mg/ml, Ovalbumin 2 mg/ml, and Conalbumin 2 mg/ml.

**Esterase activity/stability assays**

Esterase activity was determined as the rate of hydrolysis of p-nitrophenyl esters by following the formation of p-nitrophenolate by the change in absorbance at 410 nm. The activity was assayed in 50mM Na-phosphate, 25 mM N-lauroylsarcosine (Sarkosyl), 10% (v/v) 2-propanol, pH 7.0 with 20 ng/ml EstA. The substrate was 1 mM p-nitrophenyl hexanoate unless otherwise states. The change in absorbance was measured with an Infinite M200 (Tecan Trading AG, Switzerland) plate reader at 40°C. Thermal inactivation of EstA was performed by incubating 5µg/ml enzyme in 50mM Na-phosphate, 25 mM N-lauroylsarcosine (Sarkosyl), for the appropriate time.
**Interface prediction**

PIADA analysis was performed with PSAIA Version 1.0, using standard settings.²¹ PISA analysis was performed using standard settings from PDBePISA homepage⁸,²²

Stability modelling/ model building (FoldX script)

FoldX⁴²³ were used to repair, modify PDB files with the RepairPDB and BuildModel commands respectively. All Stability and AnalyseComplex calculation was made with the following values for temperature, pH, and ion strength: 308.15K, 7, 0.1.

**Structural inspection and image generation**

All structural inspection and imagen was made with “The PyMOL Molecular Graphics System” version 1.3.

**Results and discussion**

**In silico analysis of the dimer interface**

Recently the structure of an esterase native of hyperthermophilic archaea *Sulfolobus islandicus* (EstA) has been solved with a 2.6 Å resolution (PDB ID: 5LK6). The enzyme is a dimer in the crystal, and this is consistent with previously *in vitro* analysis, where the enzyme was purified as a dimer from *S. islandicus*.¹⁷,¹⁸ Structural analysis by PIADA and PISA algorithms revealed that the dimer interface of the enzyme is located in the C-terminal region and consist of the C-terminal β-strand (8), of an alpha/beta hydrolase fold and α-helix 12. A total of 29 amino acid (AA) residues were predicted to participate in the dimer formation (Figure 1). Of these AAs are 9 polar, 8 hydrophobic, 8 charged AAs (5 negative and 3 positive), as well as 3 proline and a single glycine (Figure 2), thus all types of AAs are represented. PIADA analysis predicted 188 interactions this represents an over estimation since each interaction was predicted twice, once for each subunit, and half of the interaction between the subunits can therefore expectedly be removed once the redundancy of the two data set was confirmed. Resulting is a total of 63 Van der Waals (VdW), 6 hydrophobic, 20 polar, 5 ionic interaction, totalling 94 unique interaction predicted by PIADA analysis. The PISA analysis predicted 30 interactions, 24 hydrogen bonds and 6 salt bridges. Of the 24 hydrogen bonds, 4 are those of the 4 salt bridges detected, hence 24 hydrogen bonds and 4 ionic interaction.

---

¹⁸ http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html
However, the PIADA analysis do not give any information about multiple interactions of the same type between two defined amino acids, e.g. both back-bone to back-bone and side-chain to side-chain hydrogen bonding between two amino acids. A curation of the polar and ionic interaction from both PIADA and PISA identified 34 unique hydrogen bonds and 5 ionic interactions. A final count of the predicted interactions yields 102 in total: 63 VdW, 6 hydrophobic interactions, 34 polar interaction/hydrogen bonds, and 5 ionic interactions, see Table II for a list of all non-VdW interactions and Figure 3 for the EstA structure with the interface highlighted. A possible explanation to the differences in the hydrogen bonds and ionic interaction predicted by PIADA and PISA can arise from differences in the criteria of detection interactions. PISA uses a maximum distance of 3.89Å for hydrogen bonds where PIADA uses 4.7Å. For ionic interactions, the distances are 4Å and 6Å for PISA and PIADA respectively. This can explain why only PIADA predicted the ionic interaction between OE2 of Glu_242 and NH1 of Arg_268, where the distance is 5.9Å, hence outside the maximum range of PISA. This also explains why PIADA do not predict the symmetric amino acid pair to form an interaction, as the distance between NH1 of Arg_268 and OE2 of Glu_242 is 6.4Å. The differences in maximum distances also explain all but one of the cases where PISA fails to predict hydrogen bonding, compared to PIADA. The exception is the interaction between Arg_287 and Arg_299, as the minimum distances here are 3.7Å and 3.8Å for the two symmetrical amino acid pairs. In contrast, why PIADA fail to detect hydrogen bonding between: Asn_255 to Asn_271, 3.04-3.05Å, Val_264 to Gln_285, 3.28-3.46Å, or Thr_265 to Asp_288, 2.44-2.62Å, which are predicted by PISA is unknown. It should be noted that, as for the polar interactions, the number of VdW interactions do not specify if an amino acid participate in more than one VdW interaction, only that two unique residues form at least one VdW interaction. As such, the 63 VdW interactions predicted are properly an underestimation. The 29 amino acids selected for further analysis, include all amino acids predicted by at least one of the methods to participate in at least one interaction regardless of type.

The thermodynamic contribution to the dimer formation of each of the 29 residues identified above, was investigated by generating 29 structural models. Each representing an EstA mutant protein with alanine substitution for each of the 29 residues. The change in thermodynamic energy of the dimer interface ($\Delta \Delta G_D$) of each alanine substitution, was as calculated as the difference in $\Delta G$ of mutant dimer interface relative to the wild-type dimer interface. The structural modelling and the thermodynamic calculation, were made in silico with FoldX4, and
residues with a $\Delta \Delta G_D$ of less than 2 kcal/mol upon substitution were discarded as possible targets for \textit{in vitro} analysis. This left only 10 of the 29 amino acids predicted with PISA and PIADA analysis, for further analysis (Table III). Unfortunately FoldX4 is unsuitable for stability modelling above a moderate temperature of 35°C, hence the increased contribution of ionic and hydrophobic interaction cannot be modelled at relevant temperatures.\textsuperscript{23} Therefore, the estimation of the thermodynamic contribution of the hydrophobic and charged amino acids, are likely underestimated.

Of the 10 residues Glu251 is the only residue predicted to engage in the dimer interface through VdW interaction alone. It is surprising that substitution of Glu251 with alanine has such a large effect (2.3 kcal/mol) on the stability of the dimer interface since Glu251 only contribute through VdW interaction. It is however, possible that this surprisingly large effect of G251A substitution on the dimer interface does not arise through its direct involvement in the dimer interface, i.e. the VdW contact with Asn271. Glu251 is part of a ionic network close to the dimer interface, where it neutralises Arg268\textsuperscript{24} (see Table IV for alignment). Arg268 is one of the 10 residues predicted to contribute with more than -2 kcal/mol to the dimer interface, it is possible that the loss of this negative charge and thus the neutralisation of Arg268, act as a secondary effect destabilising the dimer interface (Figure 4). Modifications of the interaction between the two residues, Glu251 and Arg268, and the resulting effect on dimer stability has been tested elsewhere.\textsuperscript{24} By substituting Arg268 with glutamate, glycine and lysine, it has been show that the dimer remains intact and only about 40% was disrupted by 3M urea of for the substitutions of glycine and glutamate substitution.\textsuperscript{24} Consequently, E251 and R268 was excluded from \textit{in vitro} mutational analysis.

Also among the 10 residues, are the hydrophobic amino acids Val267, Phe269, and Leu292, located in the core of the interface (Figure 5) and the polar residues Thr265, Asn270, and Asn271, which participate in side-chain to side-chain hydrogen bonds (Figure 6). As the strength of both hydrophobic\textsuperscript{25} and ionic interaction\textsuperscript{26} increases with temperatures, the hydrophobic and charged residues are prioritised over polar, as hydrogen bonds become less relevant at the physiological temperatures, 75-80°C, and above. Hence T265, 270, and 271 was eliminated for \textit{in vitro} analysis, since the importance of hydrophobic interactions in dimer interfaces of homologous thermophilic esterases have been demonstrated elsewhere. By alanine substitutions of residues homologous to Val267 and Phe269 in a homologue thermophilic esterase (EstE1), (see Table IV for alignment) the mutations were shown to convert EstE1 to a monomer,
supporting the importance of the hydrophobic core for dimer integrity. Likewise, glutamate substitution of the Phe269 homologous residue also disrupted the dimer, thereby converting EstE1 to a monomer. As a result, all three hydrophobic residues were selected for in vitro analysis.

The last two of the 10 residues, Asp288 and Arg299, form a dimer spanning salt-bridge. Arg299 was selected for in vitro analysis over Asp288, as this residue had the highest number of interface spanning interaction, 4 versus 3 (Figure 7), thereby disrupting most of the interactions in the interaction network consisting of Arg299, Thr265, Glu284, Asp288, Ser291 (Figure 7).

The 4 residues, Val267, Phe269, Leu292, and Arg299, selected for mutational analysis are involved in 17 of the 34 non-VdW interaction predicted by PIADA and PISA (Table II), hence by analysing only 4 of the 24 AAs participating in non-VdW interaction 50% of the interaction are analysed. In fact, a triple alanine substitution of Val267+Phe269+R299, disrupt 16 of 34 dimers spanning non-VdW interaction. The 17 interactions are all based on the side-chain the mutated residues i.e. no interaction with the backbone of the 4 AAs. This indicates side-chain connectivity as a possible identifier if residues important for protein-protein interface stability and as targets for disruptive mutagenesis.

A progressive mutation strategy was selected for studying function of the hydrophobic core in the dimer interface. Mutants were generated with one, two, or three alanine substitution in the hydrophobic core. A glutamate substitution of Phe269 was also included as a single point mutation in the analysis of the hydrophobic core. To analyse the effect of Arg299, the arginine was substituted with an alanine. Additionally, a triple mutation where Arg299, Val267, and Phe269 were all substituted to alanine were constructed, as well.

Additional mutants

In addition to the four residues selected for studying the effect of destabilising the dimer interface of EstA, four additional mutants were also designed. To test the importance of an intramolecular disulphide bond between Cys101 and Cys103 for stability and activity, a serine substitution of Cys101 was made. It has previously been shown that the cystine are present when EstA is expressed from E. coli Shuffle T7, a cell line engineered to facilitate cytoplasmic
cysteine formation in *E. coli*. Knowledge about functional importance of this disulphide bond can have an impact on future choice of expression systems.

As destabilisation of the dimer is expected to lead to a reduced stability of the enzyme, it is interesting to test whether further stabilisation of the dimer will result in an increase in stability of EstA. To test this hypothesis, EstA mutants was designed with novel inter-molecular interaction. An ionic interaction was introduced by substituting Gly261 with aspartate, as a partner to Arg6 (Figure 8). Introduction of two symmetrical inter-molecular disulphide bond through cysteine substitution of Ser266 and Asn270 (Figure 9) was similarly tested. A novel interface spanning disulphide bond on the symmetrical axis of the EstA dimer by substitution of Ser295 to cysteine was also hoped to stabilise the protein-protein interaction (Figure 10). These targets were each identified by visual inspection of the dimer interface. The thermodynamic effect of three mutants designed to stabilise the dimer interface was modelled similar to the other mutants (Table V). Their effect on the dimer interface, ΔΔG, is estimated to only 0.7 kcal/mol. The substitution of Ser295 is modelled to not affect to stability of the esterase significantly, while both the mutants with substitution of Ser292 and Asn270 for cysteine, and glutamate substitution for Gly261 are modelled to have a significant effect on the stability of EstA, by -1.153 kcal/mol and 7.261 kcal/mol. This suggests, making an introduction of a possible disulphide bond through cysteine substitution of Ser251 and Asn270, to be the best candidate for improved stability.

The 12 mutants, described above, were generated by SLIM mutagenesis of p32a-EstA_V2 and the enzyme was expressed and purified from *E. coli* ShuffleT7.

**Biochemical characterisation of selected EstA mutant proteins**

Wildtype EstA and the 12 mutants were successfully purified to homogeneity by IMAC from *E. coli* (Figure 11). The oligomeric state of each EstA variant was evaluated by size-exclusion chromatography (Figure 12A). All variation of EstA except those carrying substitutions of Arg299, eluted as a single peak, with a retention time, corresponding to a dimer (Figure 12B). The mutants with Arg299 substitutions elutes with an addition peak of about 20 kDa. The size of this population seems to increase with additional mutations. Generally, the EstA dimer remain intact despite the introduction of respective mutants this demonstrates interaction in the EstA dimer interface is stronger than that of EstE1, a homologous thermophilic esterase identified
from metagenomic data. In this esterase homologue, alanine substitution of Val267 and Phe269 or Phe269E substitution result in a complete disruption of the dimer at room temperature.\textsuperscript{15}

The effect of the mutations on the catalytic stability of EstA, was evaluated by heat treatment of the enzymes at 90°C up to 30 minutes. Here, in contrast to the size-exclusion chromatography, the differences between different variations of EstA are clear. As predicted by the \textit{in silico} modelling, Table V, the stability of the esterase is severely compromised as mutations accumulates in the hydrophobic core. The strength of the destabilisation for the single substitutions mutants, shows a correlation with the hydrophobicity of the mutated residue, Phe269A>Leu292>Val267. The double and triple mutants are both severely compromised and are almost completely inactive after 5 min incubation at 90°C. Their half-lives are estimated to be between 0.7-0.9±2.2 minutes (Figure 14A). The substitution of arginine 299 to alanine only had a moderate effect on EstA stability, compared to the other single point mutants. The half-life of the R299A mutant is reduced to 10.9±0.6 from 12.3±0.7 minutes of the wild-type. This indicates that the additional population, observed in size-exclusion chromatography for the R299A mutant, is unrelated to the stability.

As seen in Table VI, substitution of only one hydrophobic residues in the dimer interface, correlate with a decrease of at least 54% in half-life at 90°C. This is also in good agreement with the prediction from the \textit{in silico} modelling, which predicted a larger effect on the stability compared to the dimer formation (Table V).

Of the three mutants, with substitutions aimed to stabilise the dimer interface, only substitution of S295 by cysteine resulted in an increased catalytic stability, by increasing the half-life of the esterase from 12.3±0.68 to 20±1.64 minutes. It was shown by SDS-PAGE analysis in the presence and absence of DTT, that the mutation successfully has introduced a stable inter-molecular disulphide bond. It is therefore likely, that the increase in catalytic stability, is a result of this stabilisation of the dimer interface through this covalent bond. This demonstrates the importance of the dimer formation on the stability of EstA. Cross-dimer bonds has also previously been identified as a method for stabilising esterases, here the disulphide bond was the result of random mutagenesis.\textsuperscript{16}
The role of dimer formation in EstA

From our results, we can see that the introduction of an inter-molecular disulphide bond increases the stability of EstA. It is unlikely that this stabilisation arises from an increase in intra-molecular interactions, i.e. that cysteine substitution of Ser295, resulted in additional stabilising interaction compared to the original serine residue. The stabilisation is most likely a result of a stabilisation of the dimer formation and the protection the dimer provides. However, it is also possible, that as a result of conformational constrains placed on the unfolded state, due to the covalent bonding of the two peptides, it increased the stability of the folded structure.

As we were unable to obtain a monomeric EstA mutant protein, through mutagenesis, it is difficult to determine if there is a stabilising effect of the dimer formation. However, it is possible the mutations introduced a heat liability to the dimer interface, i.e. the dimer can remain intact at ambient temperature, 25°C, while the dimer at a higher temperature e.g. above 75°C is disrupted. If this is the case, the reduction of catalytic stability observed in vitro, can be the result of monomerization. To test the hypothesis, dynamic light scattering experiment was attempted without any conclusive results (data not shown).

In a paper from 2012, Mei et al. show data for EstA in a monomeric state when expressed as a thioredoxin (Trx) fusion protein. This version of the enzyme remains as a monomer (EcEstA-M) after removal of Trx-tag. There are however differences between the sequence of the EstA used here and EcEstA-M used by Mei et al. The differences are limited to the N- and C-terminal of the protein. But as both the termini are close to the dimer interface, it is possible these differences can interfere with the dimer formation. In any case, here the monomeric EstA was reported to have its stability significantly reduced, as the stability of the natively expressed and dimeric EstA was 32 fold higher. We have previously shown that EstA expressed from *E. coli* as a dimer, is relatively more stable compared to the monomeric EstA, while the stability of natively expressed EstA is only 1.27 fold higher than that of dimeric EstA expressed from *E. coli*.

These results point towards a significant effect of dimer formation, on the stability of EstA.

One could also argue whether disrupting the dimer, is a valid way of studying the stabilising effect of oligomerisation. E.g. if one is successfully in disrupting the dimer formation with only a few amino acid substitutions, the dimer interface will now be exposed to the solvent. The dimer interface has evolved as a “buried” surface, i.e. optimised for protein-protein interaction and not
protein-solvent interaction. It is therefore highly likely, that because of the interfaces native environment, protein-protein, that the contact between the interface and solvent, due to monomerisation/dimer disruption, will be less thermodynamic favourable, $\Delta G > 0$, than regular solvent exposed surface. Hence reducing the Gibb free energy between native and denatured conformation, $\Delta G_{\text{FOLD}}$.

As a result, the loss of the stabilising effect of oligomerisation, has to be considered in the combination of the possible destabilising effect of exposing the protein-protein interface to the solvent. An experimental design as the one used here, will be unable to provide a clear answer about the stabilising effect of the oligomerisation, due to this double destabilising effect upon conversion to monomeric state. In contrast, with an experimental design where the stabilising effect of the dimer is evaluated by constructive and not disruptive amino acid substitution, it is possible to determine if there is a stabilising effect of the oligomerisation, if the mutants are designed to only strengthen the dimer interface and not protein fold directly. As in the case of the introduction of a novel inter-molecular disulphide bone presented here.

**Conclusion**

The dimer interface of EstA is not stabilised by hydrophobic interaction to the same extent as the hyperthermophilic esterase EstE1. EstA retain it dimeric state despite elimination of the hydrophobic core and an ionic interaction network in the dimer interface. The stability of the EstA dimer interface is clearly demonstrated by the triple mutant V267A+F269A+R299A, where ~47% of the dimer spanning non-VdW interaction in the interface are disrupted, still maintain a dimeric state at ambient temperature. Regardless of the stability of the dimer interface the mutants stability was compromised when heat treated at 90°C possibly indication a stabilising effect by protein-protein interactions. We were able to demonstrate a stabilising effect of protein oligomerisation, through stabilisation of the dimer interface by a novel inter-molecular disulphide bonds resulting from a cysteine substitutions of Ser295, resulting in an increase of EstA half-life by 1.63 fold. Together these results show the important of dimer formation for the stability of EstA.

**References**


**Figures**

*Figure legends*

**Figure 1** Distribution of amino acids in dimer interface by method.

**Figure 2** Amino acids distribution by type

**Figure 3** EstA Dimer interface based on PIADA analysis. Dashed lines are polar interactions, max 4Å.

**Figure 4** Ionic interaction network of close to the dimer interface of EstA. A| Wild type EstA with Glu25 and Arg2681. The charge of Glu251 and Arg268 neutralises each other allowing the Arg268 of each subunit to remain in close proximity. B| Mutant EstA with Ala251 and Arg268. Arg268 is not neutralised causing the positive charge of Arg268 of each of the 2 subunits to repel each other destabilising the dimer interface.

**Figure 5** Hydrophobic amino acids in the EstA dimer interface, based on PIADA analysis. In the centre, the hydrophobic core consisting of V267, P269, and L292 can be seen.

**Figure 6** Dimer spanning Side-chain to side-chain interaction of Thr265, Asn270, and Asn271

**Figure 7** Polar and ionic side chain interactions on helix 12 and around R299 of EstA.
Figure 8 Glycine 261 substitution to aspartate together with other charged residues in the close vicinity. Arginine 6 and glutamate 9 of the adjacent subunit. Glycine 261 is coloured red, a D261 is pink. The four residues are represented by a stick model.

Figure 9 S266 and N270 of the EstA dimer. The four residues are represented by a stick model.

Figure 10 S295 of both subunits of EstA. The two residues are represented by a stick model.

Figure 11 SDS-PAGE analysis of purified of EstA and mutants. 12% SDS-PAGE, 0.78 µg enzyme per lane.

Figure 12 Size exclusion chromatography of EstA and mutants. Left panel| Chromatograms. Right panel| size estimation. Column: Superdex200 10/300, flow 750 ml/min. Buffer: NaH2PO4 50 mM, NaCl 150 mM, pH 7.8.

Figure 13 SDS-PAGE analysis of Disulphide bond dependent dimerization. WT:1 & 5, S295C: 2 & 6, S266CN270C:3 & 7, C101S: 4 & 8. 12 % SDS-PAGE, 2 µg protein per lane.

Figure 14 Residual activity of EstA and mutants after incubation at 90°C. 50 mM NaH2PO4, 25 mM Sarkosyl, 10% v/v 2-propanol, 1 mM p-nitrophenol hexanoate. A| Wild-type and mutants with mutation targeting the dimer-interface for disruption. B| Wild-type and mutants with mutation aimed at future stabilization of the dimer interface. • WT, ◆ V267A, ● F269A, ○ L292A, ❑ R299A, ◀ F269E, ○ VF, □ VFL, ■ VFR, ▲ G261D, □ SN, ▲ S295C.
Figures

Figure 1

![Venn diagram showing overlap between PISA and PIADA categories with specific residues highlighted.]

Figure 2

![Pie chart illustrating percentage distribution of hydrophobic, polar, charged, and special cases.]

Figure 3

![3D molecular structure image.]

81
Table I: Primer for mutagenesis of EstA by SLIM.

<table>
<thead>
<tr>
<th>Subunit</th>
<th>Table</th>
<th>Phe</th>
<th>Leu</th>
<th>Pro</th>
<th>267</th>
<th>267</th>
<th>269</th>
<th>1</th>
<th>1</th>
<th>292</th>
<th>258</th>
</tr>
</thead>
<tbody>
<tr>
<td>301-X-F</td>
<td>GCCATGATCCGATGATAATTAGGTT</td>
<td>AAACATGATCCGATGATAATTAGGTT</td>
<td>GATACGCTTACCAATAGGTT</td>
<td>ACGATGATCCGATGATAATTAGGTT</td>
<td>GATACGCTTACCAATAGGTT</td>
<td>ACGATGATCCGATGATAATTAGGTT</td>
<td>GATACGCTTACCAATAGGTT</td>
<td>ACGATGATCCGATGATAATTAGGTT</td>
<td>GATACGCTTACCAATAGGTT</td>
<td>ACGATGATCCGATGATAATTAGGTT</td>
<td>GATACGCTTACCAATAGGTT</td>
</tr>
<tr>
<td>301-Y-R</td>
<td>CCATGACGCTTACCAATAGGTT</td>
<td>ACGATGATCCGATGATAATTAGGTT</td>
<td>GATACGCTTACCAATAGGTT</td>
<td>ACGATGATCCGATGATAATTAGGTT</td>
<td>GATACGCTTACCAATAGGTT</td>
<td>ACGATGATCCGATGATAATTAGGTT</td>
<td>GATACGCTTACCAATAGGTT</td>
<td>ACGATGATCCGATGATAATTAGGTT</td>
<td>GATACGCTTACCAATAGGTT</td>
<td>ACGATGATCCGATGATAATTAGGTT</td>
<td>GATACGCTTACCAATAGGTT</td>
</tr>
<tr>
<td>301-X-R</td>
<td>ACGATGATCCGATGATAATTAGGTT</td>
<td>GATACGCTTACCAATAGGTT</td>
<td>ACGATGATCCGATGATAATTAGGTT</td>
<td>GATACGCTTACCAATAGGTT</td>
<td>ACGATGATCCGATGATAATTAGGTT</td>
<td>GATACGCTTACCAATAGGTT</td>
<td>ACGATGATCCGATGATAATTAGGTT</td>
<td>GATACGCTTACCAATAGGTT</td>
<td>ACGATGATCCGATGATAATTAGGTT</td>
<td>GATACGCTTACCAATAGGTT</td>
<td>ACGATGATCCGATGATAATTAGGTT</td>
</tr>
<tr>
<td>301-Y-F</td>
<td>ACGATGATCCGATGATAATTAGGTT</td>
<td>GATACGCTTACCAATAGGTT</td>
<td>ACGATGATCCGATGATAATTAGGTT</td>
<td>GATACGCTTACCAATAGGTT</td>
<td>ACGATGATCCGATGATAATTAGGTT</td>
<td>GATACGCTTACCAATAGGTT</td>
<td>ACGATGATCCGATGATAATTAGGTT</td>
<td>GATACGCTTACCAATAGGTT</td>
<td>ACGATGATCCGATGATAATTAGGTT</td>
<td>GATACGCTTACCAATAGGTT</td>
<td>ACGATGATCCGATGATAATTAGGTT</td>
</tr>
</tbody>
</table>

Each mutant is made the two set of primer pairs, set ‘X’ and set ‘Y’. The PCR product for the individual reaction are mixed, and the resulting hybridisation product are two self-circularisation linear vectors. The primer name indicates mutation, primer set, and direction, each separated with a ‘-’. (‘) These primers are used to generate four separate mutants, V267A, F269A, F269E, and V267A+F269A.

Table II: Non-Van der Waals interaction of the dimer Interface

<table>
<thead>
<tr>
<th>Hydrophobic (6)</th>
<th>Ionic (5) [Atoms]</th>
<th>Backbone to backbone (6)</th>
<th>Side-chain to backbone (11) (Not directional)</th>
<th>Side-chain to side-chain (17)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pro,5-Leu,258</td>
<td>Leu,258-Pro,5</td>
<td>Glu,242-Arg,268</td>
<td>Arg,299-Glu,284 Arg,299-Asp,288</td>
<td>Asp,288-Arg,299</td>
</tr>
<tr>
<td>Val,267-Phe,269</td>
<td>Phe,269-Val,267</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu,292-L,292</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table of Subunit information:

Subunit is marked by a subscript number. **Bold**: Mutagenesis targets, *underlined*: Part of salt bridge.

<table>
<thead>
<tr>
<th>Hydrophobic (6)</th>
<th>Ionic (5) [Atoms]</th>
<th>Backbone to backbone (6)</th>
<th>Side-chain to backbone (11) (Not directional)</th>
<th>Side-chain to side-chain (17)</th>
</tr>
</thead>
</table>

Each column represents a specific interaction type, and the numbers denote the number of specific interactions within that category.
Table III. Effect of alanine substitution.

<table>
<thead>
<tr>
<th>Residues</th>
<th>ΔΔG_S [kcal/mol]</th>
<th>ΔΔG_D [kcal/mol]</th>
<th>ΔΔG_D / ΔΔG_S</th>
</tr>
</thead>
<tbody>
<tr>
<td>E251</td>
<td>2.425</td>
<td>2.2992</td>
<td>95%</td>
</tr>
<tr>
<td>T265</td>
<td>6.503</td>
<td>4.4011</td>
<td>68%</td>
</tr>
<tr>
<td>V267</td>
<td>7.163</td>
<td>3.8166</td>
<td>53%</td>
</tr>
<tr>
<td>R268</td>
<td>3.934</td>
<td>2.9045</td>
<td>74%</td>
</tr>
<tr>
<td>F269</td>
<td>11.077</td>
<td>5.1797</td>
<td>47%</td>
</tr>
<tr>
<td>N270</td>
<td>2.095</td>
<td>2.2895</td>
<td><strong>109%</strong></td>
</tr>
<tr>
<td>N271</td>
<td>3.6</td>
<td>2.1606</td>
<td>60%</td>
</tr>
<tr>
<td>D288</td>
<td>1.495</td>
<td>3.7114</td>
<td><strong>248%</strong></td>
</tr>
<tr>
<td>L292</td>
<td>7.353</td>
<td>2.7275</td>
<td>37%</td>
</tr>
<tr>
<td>R299</td>
<td>1.558</td>
<td>3.0598</td>
<td><strong>196%</strong></td>
</tr>
</tbody>
</table>

Table IV. Alignment of 3 Hyper thermophilic esterases. SiRe_EstA, EstA from *S. islandicus*. St-EstA, esterase from *S. tokodaii*. EstE1, esterase from metagenomics data. **Δ:Cystine 101**, ^:Valine267, Phenylalanine269, Leusine292. *: Arginine299, ε:Serine266 & 295, Asparagine270. The alignment was generated with ClustalX 2.1. Bold residues are in the dimer interface.

Table V. Stability of mutants

<table>
<thead>
<tr>
<th>Amino acid substitutions</th>
<th>ΔΔG_S [kcal/mol]</th>
<th>ΔΔG_D [kcal/mol]</th>
</tr>
</thead>
<tbody>
<tr>
<td>C101S</td>
<td>1.958</td>
<td>-0.0068*</td>
</tr>
<tr>
<td>G261D</td>
<td>7.261</td>
<td>0.2642*</td>
</tr>
<tr>
<td>S266C+N270C</td>
<td>-1.153</td>
<td>0.7103</td>
</tr>
<tr>
<td>F269E</td>
<td>12.667</td>
<td>4.9871</td>
</tr>
<tr>
<td>S295C</td>
<td>0.425*</td>
<td>0.5702</td>
</tr>
<tr>
<td>V267A</td>
<td>7.171</td>
<td>3.8178</td>
</tr>
<tr>
<td>F269A</td>
<td>11.062</td>
<td>5.1728</td>
</tr>
<tr>
<td>L292A</td>
<td>7.293</td>
<td>2.7467</td>
</tr>
<tr>
<td>V267A+F269A</td>
<td>15.943</td>
<td>7.3473</td>
</tr>
<tr>
<td>V267A+F269A+L292A</td>
<td>21.209</td>
<td>9.4451</td>
</tr>
<tr>
<td>V267A+F269A+R299A</td>
<td>18.116</td>
<td>10.791</td>
</tr>
<tr>
<td>R299A</td>
<td>1.418</td>
<td>2.9559</td>
</tr>
</tbody>
</table>

*Not significant
Table VI. Enzymatic half-life at 90°C

<table>
<thead>
<tr>
<th>Mutant:</th>
<th>Half-life [min]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild-type</td>
<td>12.31±0.68</td>
</tr>
<tr>
<td>V267A</td>
<td>5.33±0.14</td>
</tr>
<tr>
<td>F269A</td>
<td>1.64±0.55</td>
</tr>
<tr>
<td>L292A</td>
<td>5.69±0.73</td>
</tr>
<tr>
<td>R299A</td>
<td>10.86±0.55</td>
</tr>
<tr>
<td>F269E</td>
<td>1±0.46</td>
</tr>
<tr>
<td>V267A+F269A</td>
<td>0.89±0.51</td>
</tr>
<tr>
<td>V267A+F269A+L292A</td>
<td>0.75±1.88</td>
</tr>
<tr>
<td>V267A+F269A+R299A</td>
<td>0.69±2.17</td>
</tr>
<tr>
<td>C101S</td>
<td>12.99±0.71</td>
</tr>
<tr>
<td>G261D</td>
<td>9.23±0.6</td>
</tr>
<tr>
<td>S266C+N270C</td>
<td>12.73±1.34</td>
</tr>
<tr>
<td>S295C</td>
<td>20.05±1.64</td>
</tr>
</tbody>
</table>
Chapter 3

Here, in the third and final manuscript, where we highlight some of the pitfalls associated with using the classically esterase assays at high temperature and suggest a set of optimized conditions better suited for temperatures relevant for hyperthermophilic enzymes.

The manuscript in formulates as a technical note of maximum 2400 words for submission to Analytical Biochemistry: Methods in the Biological Sciences. The manuscript is still in preparation.
An esterase activity assay suitable for use at temperatures up to 90 °C

Daniel Stiefler-Jensen¹, Troels Schwarz-Linnet², Qunxin She¹, and Kaare Teilum²,*

¹Archaea Centre, Department of Biology, University of Copenhagen, Denmark
²Structural Biology and NMR Laboratory, Department of Biology, University of Copenhagen, Denmark

*Correspondence: kaare.teilum@bio.ku.dk

Abbreviations
pNPE p-Nitrophenyl esters
pNPH p-Nitrophenyl hexanoate
SDS Sodium dodecyl sulfate

Abstract
The hydrolysis of p-nitrophenol hexanoate is often used as an assay for esterase activity. However, the common protocols are not optimized for high temperatures. Here we highlight several aspects that need consideration when designing a temperature stable p-nitrophenol hexanoate assay for esterase activity. The conditions (50 mM sodium phosphate, 1% Triton X-305, 10% 2-propanol, 1 mM p-nitrophenol hexanoate, pH 7) we found to be optimal resulted in an assay that is functional between 20-90°C and suitable for real-time data collection.

Introduction
Thermophilic and hyperthermophilic organisms are natural repositories of enzymes with high stability. As many biotechnological applications of enzymes require elevated temperatures, the interest in isolating and characterizing enzymes from these organisms is increasing. Esterases and their cousins lipases are enzymes often used in industrial application as they catalyse reactions that may replace traditional organic chemical reactions. Several reports on thermophilic esterases and lipases use activity assays developed for analysis of mesophilic enzymes, with optimal temperatures usually between 20°C and 40°C [1–3]. In contrast, hyperthermophilic enzymes may have activity maxima at temperatures above 75 °C and still maintain activity above 90 °C. For proper characterisation, especially determination of temperature-activity profiles, an assays with a stable performance over a wide temperature range e.g. 20-90°C is required. Here we will highlight some of the pitfalls associated with using the classically esterase
assays at high temperature and suggest a set of optimized conditions better suited for these temperatures.

Esterase activity is often measured by following the hydrolysis of p-nitrophenyl esters (pNPE), which results in a p-nitrophenoxide product that has an absorption maxima at 410 nm [4,5]. p-nitrophenoxide is the deprotonated state of p-nitrophenol, which has pKa of 7.15 [6]. The titration results in a spectral change with an isosbestic point of 348 nm, and at pH values below 7.15 the extinction coefficient is low. For the sensitivity of the assay it is therefore desirable to work at high pH. However not all enzymes are stable or active at this pH. In addition, spontaneous hydrolysis of pNPEs become a large problem when pH increases above 8, especially for short alkyl chains.

The pKa of all buffers are to some extent temperature dependent. The selection of a buffer must therefore not only be based on the working pH but also on the temperature range the buffer is expected to operate within. For the selection of a working pH these four factors need to be considered: the substrate stability, the signal strength, the enzyme activity, and temperature dependency of the buffer pKa.

As a consequence of the amphipathic nature of pNPE they will form micelles in aqueous solutions. Micelle formation will limit the substrate availability and increase light scattering. To counteract the micelle formation, detergents and/or organic solvents are often used to make an emulsion. In some assays a mixture of Triton X-100 and Arabic gum is used [1,7].

Some non-ionic detergents have an inverse correlating between solubility and temperature, meaning that above a temperature threshold, known as cloud point, the solution turns turbid/opaque which results in a drastically increase in light scattering. Generally non-ionic detergents have low cloud points compared to ionic detergents. Therefore, many non-ionic detergents are not suitable for assays preformed at high temperatures, e.g. Triton X-100 and Tween 20 have cloud points of 65 °C and 76 °C respectively.

Light scattering is not only arising from the detergent. As the reaction solution is an emulsion, it has a certain emulsifying capacity. If the concentration of the micelle-forming component surpasses this capacity, micelles will form and result in light scattering. Hence, the substrate

---

concentration needs to be optimised for the highest possible concentration for e.g. Michaelis-Menten kinetics while avoiding micelle formation.

**Result and Discussion**

We were motivated to optimizing the pNPE hydrolysis assay because we wished to characterize an esterase (EstA) from the hyperthermophilic and acidophilic archaea *Sulfolobus islandicus* that lives in acidic volcanic hot springs at temperatures around 75-80°C and pH ~3[8]. The optimal conditions for pNPE hydrolysis catalysed by EstA is 75°C [9] and pH 8 [1].

The initial conditions we used for assaying the activity was 50 mM TrisHCl, pH 8.0, 0.4% Triton X-100 and 0.1% Arabic gum with p-nitrophenyl hexanoate dissolved in 2-propanol as the substrate, which are identical to the conditions suggested by Pencreac’h and Barratti [5]

Although a pH-value of 8 would be desirable, p-nitrophenyl hexanoate hydrolyses spontaneously at pH 8, which at high temperatures results in a prohibitively high background. The rate of auto hydrolysis is reduced by 80% using pH 7 instead of pH 8. This comes at a cost of a 40% reduced extinction coefficient at 410 nm. (Figure S1 & Figure S2). We chose to use 50 mM sodium phosphate to buffer the assay at pH 7 as the pK$_a$ temperature coefficient (dpK$_a$/dT) for phosphate is only -0.0028 which is much smaller than for the suitable alternatives BisTris, MOPS, and HEPES that have dpK$_a$/dT between -0.011 and -0.017 [10]. In the temperature interval from 20 to 90 °C where we performed the assay, the pH of a phosphate buffer will only decrease by 0.2 pH units.

The use of Triton X-100 and Arabic gum as emulsifiers to avoid that pNPEs forms micelles is problematic at high temperatures as Triton X-100 has a cloud point of 66 °C. As alternatives to 0.4% Triton X-100 we tried another non-ionic detergent, 1% Triton X-305, and an anionic detergent, 25 mM lauroyl sarcosinate (Sarkosyl), which both have cloud points above 100 °C and make stable emulsions of 1 mM p-hydroxyl hexanoate even without the presence of Arabic gum (Figure S4), see Table 1 information on different detergents. The effect of the different detergents on the measured enzyme activity is show in Figure 1. Whereas all three detergents give useful data at 40 °C it is clear that nothing can be measured at 90 °C when Triton X-100 is used. The missing change in measured absorbance is a result of high background from light scattering when passing the cloud point for Triton X-100, and not due to inactivation or inhibition of the enzyme as the reaction still turns yellow. In chosing between Triton X-305 and
Sarkosyl it is important to consider the stronger denaturating effect of Sarkosyl, which may compromise the activity of the enzyme.

The highest quality data is obviously obtained if traces of the absorbance as a function of time can be recorded. If, however, it is necessary to stop the reaction before measuring the absorbance (e.g. if the spectrophotometer or plate reader has no temperature control) we suggest to stop the reaction by addition of 0.2-1.8% SDS, which will denature the enzyme [7,11].

In addition to the consideration about the chemical composition if the esterase assays the way the assay is mixed is also important. As both the detergents and organic solvents in the assay affect the surface tension, regular pipettes are difficult to handle reproducible. For this reason, we recommend the use of automatic dispensing pipettes that extract an excess volume.

Thermophilic and hyperthermophilic enzymes are typically restricted in their catalytic ability at moderate temperatures, e.g. EstA has a temperature optimum of 75°C and only retain about 20% activity at 40°C [9], which is believed to be the result of the evolution of a protein structure with optimal functional flexibility close to the growth temperature of the organism it originates from [12]. Working at high temperatures is usually not possible unless specialised equipment is available. The equipment we found best suited for the assay at high temperatures were a spectrophotometer with a Peltier temperature control unit. With this setup, we able to obtain reproducible data up to 90 °C. Above this temperature the buffer starts to boil. For high throughput analysis, we have used plate readers at 40°C.

In conclusion, with the optimized conditions, 50 mM Na phosphate, (25 mM Sarkosyl or 1% Triton X-305), up to 1 mM p-Nitrophenyl hexanoate, 10% v/v 2-propanol, pH 7, and careful preparation of the assay we were able to acquire high quality and reproducible activity data on EstA in the temperature interval between 20 °C and 90 °C.

**Tips & tricks**

- Use auto pipette whenever possible, due to the surfactant and 2-propanol in the buffer modifies the surface tension. As a result, we obtained more consistent results with the use of auto-pipets where effects coursed by surface tension can be omitted. i.e. with a pipet capable of dispensing, retrieve 1 ml and dispense 3x 250 µl and discard the remaining volume.
Avoided vigorous mixing of the buffer to reduce the formation of bubbles and foam, due to the presence of the surfactant.

The substrate should be prepared as late as possible. First dilute the substrate in e.g. 2-propanol and then add the dilution gently to the buffer and mix by gently inverting.

Start the reaction by adding enzyme, this allows the substrate buffer mix to become clear after mixing.

Avoided stopping reaction with base, acid, or PMSF. Use strong detergents e.g. SDS if necessary.

Use boil proof plastic ware.

**Suggested assay conditions**

- 50 mM NaH2PO4, 25 mM Sarkosyl, up to 1 mM p-Nitrophenyl hexanoate, 10% v/v 2-propanol, pH 7
- 50 mM NaH2PO4, 1% Triton X-305, up to 1 mM p-Nitrophenyl hexanoate, 10% v/v 2-propanol, pH 7

Assay volumes between 250-1000 µl. Enzyme concentrations ~500 nM. Temperatures between 20-90°C.

**References**


Figures

**Figure legends**

**Figure 1 Comparison of buffer performance at 40°C and 90°C.** The activity of EstA was tested in the three buffers 50 mM NaH2PO4, 25 mM sarkosyl, pH 7 (red), 50 mM NaH2PO4, 1% w/v Triton X-305, pH 7 (blue), 50 mM Tris, 1% Triton X-100, pH 7 (green), at 40°C (A) and 90°C (B). (A) all three buffer have similar curves, when disregarding the differences in activity. (B) Here the two buffers with a cloud point above 100°C show a nice activity curve while signal is lost in the Triton X-100 buffer due to light scatter. The flat lining of the signal is not due to reduced activity. But a result of extensive light scatter. The difference in graph shape from 40°C to 90°C, is a result of increased reaction rate resulting in none linear substrate conversion soon after measurement starts this can be postponed by increasing substrate concentrations from 0.5 mM used her to 1 mM. Here the difference serves as a reminder of the different requirements for substrate concentrations and observation at low and high temperatures.

**Figure 1**

![Figure 1](image)

Tables

**Table 1. Basic information of Surfactants.**

<table>
<thead>
<tr>
<th>Name:</th>
<th>Structur:</th>
<th>Cloud point:</th>
<th>Group:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brij 58</td>
<td><img src="image" alt="Image" /></td>
<td>&gt;100°C</td>
<td>Non-ionic</td>
</tr>
<tr>
<td>Sodium dodecyl sulfate (SDS)</td>
<td><img src="image" alt="Image" /></td>
<td>&gt;100°C</td>
<td>Anionic</td>
</tr>
<tr>
<td>Sodium lauroyl sarcosinate (Sarkosyl)</td>
<td><img src="image" alt="Image" /></td>
<td>Anionic</td>
<td></td>
</tr>
<tr>
<td>Triton X-100</td>
<td><img src="image" alt="Image" /></td>
<td>66°C</td>
<td>Non-ionic</td>
</tr>
<tr>
<td>Triton X-305 n=30</td>
<td><img src="image" alt="Image" /></td>
<td>&gt;100°C</td>
<td>Non-ionic</td>
</tr>
<tr>
<td>Tween 20</td>
<td><img src="image" alt="Image" /></td>
<td>76°C</td>
<td>Non-ionic</td>
</tr>
</tbody>
</table>
Supplementary figures

*Figure legends*

**Figure S1** pH dependent auto-hydrolysis of p-Nitrophenyl hexanoate. 1 mM pNPH in 50 mM NaH2PO4, 25mM Sarkosyl, 10% v/v 2-propanol at indicated pH’s at 40°C.

**Figure S2** pH dependent absorbance of p-Nitrophenol. 100 µM p-Nitrophenol in 50 mM NaH2PO4, 25mM Sarkosyl, 10% v/v 2-propanol at indicated pH’s at 40°C.

**Figure S3** Temperature dependency of pKa, for H2PO4 (solid) and Tris (dashed).

**Figure S4** Light scattering dependency on substrate and detergent. A| Triton X-305 B| Sarkosyl. In 50 mM NaH2PO4 10% v/v 2-propanol at pH 7.

*Figure S1*

![Figure S1](image1)

*Figure S2*

![Figure S2](image2)
Figure S3

Figure S4

A. 

B.
Appendix I

A – Introduction to Mutations Analyser

1 Disclaimer

This script is made to function with python 2.7 in a windows 10 64-bit system environment. The script has not been tested on other systems to date, 2016-09-06. The Script has been tested by myself and can be crashed, and I properly haven’t found all bugs. All credit for the FoldX utility goes to the FoldX team, please visit http://foldxsuite.crg.eu/. The FoldX utility requires a license and must not be freely distributed. The python code, as seen in Mutational_Analysis.py or in appendix IB is written in its entirety by myself, Daniel Stiefler-Jensen.

All use of the python script is at your own risk.

IMPORTANT that temperature range of the empirical data that forms the basis for the FoldX calculation is limited. As a result, data from temperatures above or below 35°C and 15°C respectively, are at best rough estimations. The Standard deviation $\Delta\Delta G$ from Stability calculations relative to experimental data is 0.46 Kcal/mol (Schymkowitz et al., 2005).

2 Purpose

This script was made to facilitate easy and fast analysis of multiple defined mutations on the stability and integrity of complex through a temperature range or at a fixed temperature. For systematic studies of mutations AlaScan and Pssm command are recommended, see http://foldxsuite.crg.eu/documentation#manual for more information on these commands and additional not utilised by the script. The script makes use of the following command from FoldX4: BuildModel, Stability, and AnalyseComplex.

The script has three parts, Generation of Mutated PDB file, calculating the $\Delta G$ of the new PDB files, calculating the stability of dimer interaction of the new PDB files. The action is performed by the FoldX program, the only ability of the script is an automation of combining the three commands into a single action and thereby avoiding manual activation of each command.

The following section is a brief description of the three FoldX command and how they are set up and works.
2.1 **BuildModel**

The BuildModel command allows for modification of peptide chains in a PDB file, (*.pdb). In the process, the conformation of the new residues is optimised. Only a single PDB file can be modified at a time.

e.g. a PDB file, EstA.pdb, contains 2 chains C & D of a homodimer hyperthermophilic esterase, Accession: ADX84384. If one wishes to mutate valine267 of chain C to alanine the following line must be in individual_list.txt: VC267A; Where V is for Valine, C is for chain C, 267 is for the residue number in the chain, A is for substitution to Alanine, and finally ‘;’ to end the line. To mutate several residues in the same PDB file simple add “,” between the target designation, e.g. “VC267A,VD267A” to mutate V267 to A267 in both chain C & D. Each line in the individual_list.txt will generate a new mutant, i.e. multiple different mutants can be generated from the same run of the script.

The BuildModel generates two files for each mutation set, one with WT_ prefix and one without. The WT_ file for control, here the same residues that are mutated in the mutated file, the one without the WT_ prefix, are mutated to the wildtype residues to eliminate any difference introduced in the structure as a result of the process. ΔΔG analysis should be made between the WT_ file and the mutant file, not the original PBD file listed in Mutational_AnalyserPDB_BM.txt.

The results if the BuildModel command is a PDB file for each mutation, they are placed in the following folder, Mutational_AnalyserPDB. Additional data files generated by the BuildModel command can be found in Mutational_AnalyserData_BM.

2.2 **Stability**

The Stability analyse command calculates the stability, ΔG, of the protein structure in the PDB files listed in Mutational_AnalyserData_BM\PdbList_New.fxout. For each protein structure in this file the script will calculate the stability in a temperature range, defined by a start and stop temperature and an ΔT. The individual data files can be found at Mutation_Analyser_KopiData_ST. The script also collects all the data into a single file at Mutation_Analyser_KopiResults\Result_Stability.dat. This file is inspected with any text editor. The results are tabulated and each line contains the results of the stability analysis of a single structure. The first line of the file is a header identifying each value in the lines below. A
brief description of each value can be found in the FoldX manual at
http://foldxsuite.crg.eu/documentation#manual. For general use only the three first values are of
interest: Temperature, Pdb, and Total energy.

The Temperature is the temperature of the analysis in Kelvin. Pdb contains the path to the PDB
subject to the stability analysis. The Total Energy in the ΔG of the structure in Kcal/mol.
Calculations of ΔΔG\text{Mut} as a result of mutagenesis should be made between the Total Energy of
the mutant and the WT_ prefixed file of the mutant at the same temperature see Table 1. The
remaining values in the Result_Stability.dat is the individual components of the empirical force
field used for calculating the stability these includes Electrostatic interaction and Van der Waals
interaction and many more. For an example of a ΔΔG\text{Mut} see Table 1.

<table>
<thead>
<tr>
<th>Table 1 Results_Stability.dat example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature [K]:</td>
</tr>
<tr>
<td>293.15</td>
</tr>
<tr>
<td>293.15</td>
</tr>
</tbody>
</table>

Example of the first three lines of a Result_Stability.dat. ΔΔG calculation should be made between such a file
pair, WT_FILENAME.pdb and FILENAME.pdb. In this case ΔΔG_{\text{Mut}}=(-107.133 \text{ Kcal/mol})-(-100.376
\text{ Kcal/mol}) = 6.757 \text{ Kcal/mol}. The Data shown here is the different in stability of the mutation of Valine 267 in
both subunits of an hyperthermophilic esterase of \textit{S. islandicus} calculated at 293.15\textdegree K or 20\textdegree C

2.3 AnalyseComplex

AnalyseComplex calculates the contribution of intermolecular interaction for the stability of the
complex, in Kcal/mol. For the script AnalyseComplex is treated in the same manner as the
Stability analysis. The data files are located in Analyses specific folder,
\Mutation_Analyser_Kopi\Data_AC, and the Result file is named Result_AnalyseComplex.dat.
For more information about the folder structure see the Stability section.

The result file for data analysis is however different from that of the Stability analysis.

The results are tabulated and each line contains the results of the stability analysis of a single
structure. The first line of the file is a header identifying each value in the lines below. A brief
description of each value can be found in the FoldX manual at
http://foldxsuite.crg.eu/documentation#manual. For general use only the following values are of
interest: Temperatur, Pdb, Interaction Energy. Value 1,2, and 7 in the tabulated data set. For an
example of $\Delta \Delta G_{\text{inter}}$ calculation see Table 2. The Script as only been optimised for dimer analysis.

<table>
<thead>
<tr>
<th>Temperature [K]</th>
<th>Pdb</th>
<th>Interaction Energy [kcal/mol]</th>
</tr>
</thead>
<tbody>
<tr>
<td>293.15</td>
<td>\Mutation_Analyser_KopiPDB\EstA_R_VC267A_VD267A.pdb</td>
<td>-20.2628</td>
</tr>
<tr>
<td>293.15</td>
<td>\Mutation_Analyser_KopiPDB\WT_EstA_R_VC267A_VD267A.pdb</td>
<td>-23.9401</td>
</tr>
</tbody>
</table>

Example of of a data extract Result_AnalysisComplex.dat. $\Delta \Delta G_{\text{inter}}$ calculation should be made between such a file pair, WT_FILENAME.pdb and FILENAME.pdb. In this case $\Delta \Delta G = (-20.2628 \text{Kcal/mol}) - (-23.9401 \text{Kcal/mol}) = 3.6773 \text{ Kcal/mol}$. The Data shown here is the different in interaction energy of the mutation of Valine 267 in both subunits of an hyperthermophilic esterase of *S. islandicus* calculated at 293.15°K or 20°C

2.4 RepairPDB

Before using a PDB file in combination with FoldX4 it is recommended to ensure that the structure is optimised. FoldX4’s RepairPDB command do just that, optimising rotamer conformation and eliminating Van der Waals clashes. The original pdb files must be repaired before using the script. pdb files generated by the script through the BuildModel command can be repaired during the run, by pressing y to the question is prompted after selection the BuildModel command. Repairing pdb files takes a long time so if you are generating a large number of mutant structures, expect long processing times.

**Using the Mutation_Analyser**

3 Setup before run

3.1 Installing FoldX4

Simply copy the following files to the Mutation_Analyser folder: foldx.exe, .DS_Store, rotabase.txt

3.2 Files and paths

Before running the script, the following must files must be modified for your personal setup, all can be found in \Mutational_Analyser:

PDB_BM.txt
Write the name of the PDB file to be mutated. The script can only handle 1 file at the time, as a consequence this file shall only contain a single entry in the following format: FILENAME.pdb

PDB_Dir_path.txt

Write the complete path to the folder containing the PDB file to be used for BuildModel. The script can only handle 1 file at the time, as a consequence this file shall only contain a single entry in the following format: C:\PDB_folder

Individual_list

In this file write a new line for each modified structure following this syntax unit: \( A_n C_n n M_n \). Where \( A_n \) is the 1-letter abbreviation of the residue in the poly-peptide chain, \( C_n \) is the Chain name where \( A_n \) is located, \( n \) is the residue number of \( A_n \) in \( C_n \), \( M_n \) is the 1-letter abbreviation of the residue \( A_n \) in \( C_n \) shall be mutated to. Several mutation can be introduced to the structure at the same time simply by adding additional syntax units after the first each separated by “,”. finally end the line with a semi-colon, “;” the final result of a double substitution can look like this: \( A_1 C_1 n_1 M_1, A_2 C_2 n_2 M_2 \);

3.3 Conditions

The script has been coded with standard values for the following FoldX settings:

<table>
<thead>
<tr>
<th>Table 3 Standard setup for config files of FoldX used by the script</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BuildModel:</strong></td>
</tr>
<tr>
<td>NumberOfRuns</td>
</tr>
<tr>
<td>Stability:</td>
</tr>
<tr>
<td>ionStrength [M]</td>
</tr>
<tr>
<td>pH</td>
</tr>
<tr>
<td>AnalyseComplex:</td>
</tr>
<tr>
<td>ionStrength [M]</td>
</tr>
<tr>
<td>pH</td>
</tr>
</tbody>
</table>

These values can be changes by modified the Mutant_Analyser.py in a text editor, see Table 3 for linenumbers. The temperature is defined when the script in initiated, see below.
4 Running the script

Run the script by executing Mutant_Analyser.py. This can be done both from Window and the command prompt.

4.1 Selection FoldX commands

The script will first ask you three yes/no question that must be answered by typing “y” or “n” followed by hitting the enter key. These question ask what analysis should be run. Any combination of BuildModel, Stability, AnalyseComplex, can be made. However, Stability and AnalyseComplex requires that BuildModel has been run from the same folder at least once. Each time one of the three FoldX commands are used the previously results from the command will be overwritten.

4.2 Selecting temperature range

After selecting the wanted FoldX commands to be used, you will be asked if you want to use a temperature gradient. By selecting yes the following standard temperature range for Stability and AnalyseComplex analysis: 293.15°K to 373.15°K in 10°K intervals is suggested. Answer “y” or “n” as before, Yes will start the script. answering No will prompt new question for you to setup a temperature range. Start by inputting the new start temperature in kelvin, press enter and input the new end temperature, press enter and enter the new temperature interval. The temperature interval has to be number that results in an integer when the dividing the different between start and end temperature with the interval. Saying no to a temperature grading will allow you to input a single temperature for calculation. Depending of the mutations and to length of the temperature range the execution time will vary, generally continual feedback will be given in the prompt. If the script does not end with an overview message to “Press any key to continue to run overview” an error has occurred and you have to troubleshoot. Troubleshooting can sometime be made easier by executing the script from a command prompt.

4.3 Output/results

Once the FoldX commands you selected has been completed, you can inspect an overview of the tasks performed and folder location by pressing any key. Exit the script by pressing any key a final time.

The Calculations of ΔΔG must be made from the data provided by the script elsewhere.
5  Known bugs and errors:

The file path the Mutation_analyser.py MUST NOT contain any BLANK spaces " " use _ instead.

Running the script with for the first time in a new folder without model builder option will crash the script, warnings are in place.

6  References

B – Mutations Analyser python code

```python
import subprocess
import os
import glob
import shutil
import sys

# constants
T=293.15
T_max=373.15
T_int=10

# configuration of the FoldX command to be used
RP=False

def intro():
    # A simple intro message to be displayed
    print('This script is for automatization of mutagenesis and analysis of the stability of the protein fold and dimer interface through a temperature range.

    def setup():
        # configuration of the FoldX command to be used
        global RP
        global RP_C
        if os.path.exists(D+'\Results\'): # resets Result folder.
            shutil.rmtree(D+'\Results')
        print('--------- Analysis Selection ------

        while True:
            ask2=raw_input('Do you want to repair the generated pdb files after mutagenesis? This can take a long time but gives the best results. press \"y\" for yes and \"n\" for no: \n\n')
            if ask2=='Y' or ask2=='y':
                Modes[3]='RP'
                break
            elif ask2=='N' or ask2=='n':
                Modes[3]='x'
                break
            else:
                print('Please use \"y\" or \"n\"')
                break

        if Modes[0]=='BM':
            os.path.exists(D+'\Data_BM'):
                shutil.rmtree(D+'\Data_BM')
            if not os.path.exists(D+'\Data_BM\PdbList_.fxout'):
                print('--------- IMPORTANT -------\nif BuildModel has not been run before in this folder, Stability and AnalysComplex will not work.\nStop script by \"Ctrl+C\"')
            ask2=raw_input('Do you still want to skip Buildt Model, this may crash the script. press \"y\" for yes and \"n\" for no: \n\n')
            if ask2=='Y' or ask2=='y':
                Modes[0]='x'
                break
            elif ask2=='N' or ask2=='n':
                Modes[0]='BM'
```

106
54    break
55    else:
56        break
57    else:
58        print('Please use \nYes or \nNo.\n')
59    while True:
60        ask=raw_input('\nShall we run Stability ?, press \nYes or \nNo: \n')
61        if ask=='y' or ask=='Y':
62            Modes[1]='ST'## only usable if BM has already been run once
63            if os.path.exists(D+'\Data_ST'):
64                shutil.rmtree(D+'\Data_ST')#resets data files
65                break
66            elif ask=='N' or ask=='n':
67                Modes[1]='x'
68                break
69            else:
70                print('Please use \nYes or \nNo.\n')
71    while True:
72        ask=raw_input('\nShall we run AnalysisComplex ?, press \nYes or \nNo: \n')
73        if ask=='y' or ask=='Y':
74            Modes[2]='AC'## only usable if BM has already been run once
75            if os.path.exists(D+'\Data_AC'):
76                shutil.rmtree(D+'\Data_AC')#resets data files
77                break
78            elif ask=='N' or ask=='n':
79                Modes[2]='x'
80                break
81            else:
82                print('Please use \nYes or \nNo.\n')
83            if Modes[1]=='ST' or Modes[2]=='AC':
84                if os.path.exists(D+'\PDB_RP\Repaired_pdb.txt')==True or
85                    Modes[3]=='RP':#checks Results folder and creates if necessary
86                    while True:
87                        ask=raw_input('\nDo you want to use repaired PDB files for calculating?\n')
88                        if ask=='y' or ask=='Y':
89                            RP=True
90                            break
91                        elif ask=='N' or ask=='n':
92                            RP=False
93                            break
94                        else:
95                            print('Please use \nYes or \nNo.\n')
96            def Temp(T,T_max,T_int):## configuration of the temperature range used in Stability and AnalysisComplex calculations
97                TS=0
98                if Modes[1]=='ST' or Modes[2]=='AC':
99                    print('\n\n--- Temperature Setup ------')
100               while True:
101                   ask=raw_input('\nUse temperature gradient ? press \nYes or \nNo: \n')
102                   if ask=='N' or ask=='n':
103                       while True:
104                           try:
105                               T0=float(raw_input('\nAt what temperature ? temperatures between 288.15 and 308.15, are suggested ? use \n as decimal separator:'))
106                               if T0>0:
107                                   break
108                               else:
109                                   print('\nTemperature must be positive number.\n')
110                               except ValueError:
111                                   print('\nPlease type in a number.\n')
break
if ask=='Y' or ask=='y':
    while True:
        ask=raw_input('\nUse standart Temperature setting, \$s K to \$s K in \$s K intervals? press \'y\' for yes and \'n\' for no: \n' %(T,T_max,T_int))
        if ask=='Y' or ask=='y':
            break
        if ask=='N' or ask=='n':
            while True:
                # input start temperature
                try:
                    T=float(raw_input('\nNew Start\(low) temp: \n'))
                except ValueError:
                    print('\nTemperature must be positive.'\n')
                if T>0:
                    break
                else:
                    print('\nTemperature must be positive.'\n')
                except ValueError:
                    print('\nTemperature must be positive.'\n')
                except ValueError:
                    print('\nThe number must be higher that \$s.\n' \n')
                except ValueError:
                    print('\nThe interval dose not fit in between the start and stop values as a interval.'\n')
                except ValueError:
                    print('\nThe interval temperature is dose not fit between start and end values.'\n')
                except ZeroDivisionError:
                    print('\nThe temperature intervals can\'t be zero.'\n')
                except ValueError:
                    print('\nThe temperature intervals can\'t be zero.'\n')
                if test[1]=='0' and int(test[0])>0:
                    break
                if int(test[0])<0:
                    print('\nThe interval must be a positive number.'\n')
                if test[1]!=='0':
                    print('\nThe interval dose not fit in between the start
and stop values as a integer.'\n')
                except ValueError:
                    print('\nThe interval temperature is dose not fit between start and end values.'\n')
                except ZeroDivisionError:
                    print('\nThe temperature intervals can\'t be zero.'\n')
                except ValueError:
                    print('\nThe temperature intervals can\'t be zero.'\n')
                if con=='y' or con=='Y':
                    break
                if con=='n' or con=='N':
                    Temp(T,T_max,T_int)
                    break
                else:
                    print('\n\nPlease confirm by \'y\' or \'n\': \n' %\n(T,T_max,T_int))
                    if con=='y' or con=='Y':
                        break
                    if con=='n' or con=='N':
                        Temp(T,T_max,T_int)
    break
return T,T_max,T_int,TS
def PDB_Path():#readig path of PDB folder
    Dir_path=open(PDB_Dir_path.txt','r+')
    for line in Dir_path:
def Files(FILE):  # Reading file names of PDB from BuildModel
    size = file_len(FILE)
    list = []
    i = 0
    while i < size:
        list.append('x')
        i = i + 1
    list = line.split('.') for line in l
    return list

def Config_ST(T):  # Config file for stability analysis
    config = open('config.cfg', 'w')
    config.write('command=Stability\n')
    if RP == False:
        config.write('pdb-list=' + T + D + T_ata_BM + T + PDB_RP + T + Repair_PDB.txt + T + PDBDIR + T + PDB\n')
        config.write('command=RepairPDB\n')
        config.write('temperature=%s\n % (T)
        config.write('ionStrength=0.1\n')
        config.write('pH=7\n')
        config.close()
    elif RP == True:
        config = open('config.cfg', 'w')
        config.write('command=RepairPDB\n')
        config.write('temperature=%s\n % (T)
        config.write('ionStrength=0.1\n')
        config.write('pH=7\n')
        config.close()

def Config_AC(T):  # Config file for AnalysisComplex analysis
    config = open('config.cfg', 'w')
    config.write('command=AnalysisComplex\n')
    if RP == False:
        config = open('config.cfg', 'w')
        config.write('command=AnalysisComplex\n')
        config.write('temperature=%s\n % (T)
        config.write('ionStrength=0.1\n')
        config.write('pH=7\n')
        config.close()
    elif RP == True:
        config = open('config.cfg', 'w')
        config.write('command=RepairPDB\n')
        config.write('temperature=%s\n % (T)
        config.write('ionStrength=0.1\n')
        config.write('pH=7\n')
        config.close()

def Config_BM():  # Config file for BuildModel, generation of PDB file of mutants
    config = open('config.cfg', 'w')
    config.write('command=BuildModel\n')
    config.write('numberOfRuns=1\n')
    config.write('mutant-file=individual_list.txt\n')
    config.write('pdb-list=PDB_BM.txt\n')
    config.write('temperature=%s\n % (T)
    config.write('ionStrength=0.1\n')
    config.write('pH=7\n')
    config.close()

def Config_RP(FILE_RP):
    config = open('config.cfg', 'w')
    config.write('command=RepairPDB\n')
    config.write('pdb-list=PDB_RP.txt\n')
    config.write('pH=7\n')
    config.close()

def file_len(fname):
    with open(fname) as f:
        for i, l in enumerate(f):
            pass
    return i + 1

def Collect_ST(T):  # Collection of Data files
    FILENAME = 'output_' + T + 'K_ST.dat'
    out = open(FILENAME, 'w')
    out.write('Temperatur Pdb Total Energy Backbone Hbond Sidechain Hbond Van der Waals Electrostatics Solvation Polar Solvation Hydrophobic Van der Waals clashes Entropy Side Chain Entropy Main Chain Sloop Entropy Mloop Entropy Cis Bond Torsional Clash Backbone Clash Helix Dipole Water Bridge Disulfide Electrostatic
Kon Partial Covalent Bonds Energy Ionisation Entropy Complex
Residue Number

\[ i = 0 \]
\[ \text{if } RP == \text{True:} \]
\[ \text{D1} = '\text{PDB_RP}\text{\backslash}\text{Repaired_PDB.txt'} \]
\[ \text{elif } RP == \text{False:} \]
\[ \text{D1} = '\text{Data_BM}\text{\backslash}\text{PdbList_New.fxout'} \]
\[ \text{File_names} = \text{Files(D1)} \]
\[ \text{while } i < \text{file_len(D1)}: \]
\[ \text{File} = \text{File_names[i][0]} + '_0_ST.fxout' \]
\[ f = \text{open(File, 'r+')} \]
\[ \text{for line in f:} \]
\[ \quad \text{out.write('\%s\t' \%(T))} \]
\[ \quad \text{out.write(line)} \]
\[ f.close() \]
\[ i = i + 1 \]
\[ \text{out.close()} \]
\[ \text{def Collect_AC(T):} \# \text{Collection of data files} \]
\[ \text{FILENAME} = '\text{output_}\%s' + '_K_AC.dat' \]
\[ \text{out}\_\text{open}\text{(FILENAME, 'w')} \]
\[ \text{out.write('Temperature Pdb Group1 Group2 IntraclashesGroup1} \]
\[ \text{IntraclashesGroup2 Interaction Energy StabilityGroup1} \]
\[ \text{StabilityGroup2\n')} \]
\[ i = 0 \]
\[ \text{if } RP == \text{True:} \]
\[ \text{D1} = '\text{PDB_RP}\text{\backslash}\text{Repaired_PDB.txt'} \]
\[ \text{elif } RP == \text{False:} \]
\[ \text{D1} = '\text{Data_BM}\text{\backslash}\text{PdbList_New.fxout'} \]
\[ \text{File_names} = \text{Files(D1)} \]
\[ \text{while } i < \text{file_len(D1)}: \]
\[ \text{File} = '\text{Summary_}' + \text{File_names[i][0]} + '_AC.fxout' \]
\[ f = \text{open(File, 'r+')} \]
\[ \text{LineNr} = 10 \]
\[ \text{c = 1} \]
\[ \text{for line in f:} \]
\[ \quad \text{if } c == \text{LineNr:} \]
\[ \quad \text{out.write('\%s\t' \%(T))} \]
\[ \quad \text{out.write(line)} \]
\[ \text{c = c + 1} \]
\[ f.close() \]
\[ i = i + 1 \]
\[ \text{out.close()} \]
\[ \text{def Collect_BM():} \# \text{Collection of data files} \]
\[ \text{if not os.path.exists(D+}'\text{\backslash}\text{Data_BM}') : \# \text{checks for and if needed creates folder} \]
\[ \text{for BuildModel data} \]
\[ \text{os.makedirs(D+}'\text{\backslash}\text{Data_BM}') \]
\[ \text{print('Folder for Stability results deposit created at:\n' + D+}'\text{\backslash}\text{Data_BM}\n')} \]
\[ \text{if not os.path.exists(D+}'\text{\backslash}\text{PDB}') : \# \text{checks for and if needed creates folder for} \]
\[ \text{PDB files one level down} \]
\[ \text{os.makedirs(D+}'\text{\backslash}\text{PDB}') \]
\[ \text{print('Folder for Stability results deposit created at:\n' + D+}'\text{\backslash}\text{Data_BM}\n')} \]
\[ \text{for file in glob.glob(D+}'\text{\backslash}\.\text{fxout}') : \# \text{Move data files from BuildModel to} \]
\[ \text{...}\text{\backslash}\text{Data_BM} \]
\[ f = \text{os.path.split(file)} \]
\[ \text{shutil.move(file,D+}'\text{\backslash}\text{Data_BM}' + F[1]) \]
\[ \text{for file in glob.glob(D+}'\text{\backslash}\.\text{pdb}') : \# \text{Move PDB files to PDB folder one level down.} \]
\[ f = \text{os.path.split(file)} \]
\[ \text{shutil.move(file,D+}'\text{\backslash}\text{PDB}' + F[1]) \]
```python
File2 = open(D+'\Data_BM\PdbList_New.fxout', 'w')

File3 = open(D+'\PDB_BM.txt')
for line in File3:
    Prefix = line
File3.close()
Prefix = Prefix.split('.pdb')
Prefix = Prefix[0]
Mut = 1
cc = 0
for file_name in glob.glob(D+'\PDB\*\pdb\'):
    cc = cc + 1
cc = 1
while cc <= c/2:
    Mut = cc
    for file_name in glob.glob(D+'\PDB\*\pdb\'):
        F = os.path.split(file_name)[1] # tuple
        Fl = str.split(F[1], str(Mut) + '.pdb')
        F0 = str(F[1])
        F0 = F0.split(Prefix + '.pdb')
        F0 = F0[1].split('.pdb')
        try:
            int(F0[0])
            Test = True
        except ValueError:
            Test = False
        if int(F0[0]) == Mut:
            Test2 = True
        else:
            Test2 = False
if F0[0] != str('') and Test2 == True and Test == True:
    F0 = int(F0[0])
    LN = 1
    File = open(D+'\individual_list.txt', 'r+')
    for line in File:
        if LN == Mut:
            New_name = line.replace(',', '
')
            New_name = New_name.replace(' ','_')
            New_name = New_name.replace('
', '_')
        shutil.move(file_name, D+'\PDB\'+F[1][0]+New_name + '.pdb')
    File2.write(F1[0][0]+New_name + '.pdb
')
    line = None
    File.close()
    break
    LN = LN + 1
cc = cc + 1
def Collect_RB():
    if not os.path.exists(D+'\Data_BM\RP\'):
        os.makedirs(D+'\Data_BM\RP\')
        print('Folder for Stability results deposit created at:
')
    if not os.path.exists(D+'\PDB\RP\'):
        os.makedirs(D+'\PDB\RP\')
        print('Folder for Stability results deposit created at:
')
    for file in glob.glob(D+'\Repair.fxout\'):
        F = os.path.split(file)
        shutil.move(file, D+'\Data_BM\RP\'+F[1])
```
for file in glob.glob(D+"\Repair.pdb"): #Move PDB files to PDB folder one level down.
    F=os.path.split(file)
    shutil.move(file,D+"\PDB_RP\"+F[1])
    #Generate list of new PDB files
    out=open('Repaired_PDB.txt','w')
    for file in glob.glob(D+"\PDB_RP\*Repair.pdb"): 
        F=os.path.split(file)
        out.write(F[1]+"\n")
    out.close()
    shutil.move('Repaired_PDB.txt',D+"\PDB_RP")

def Collect_Result_TS(TS,Mode):
    if Mode=='ST':
        out=open('Result_Stability.dat','w')
        i=1
        for line in f:
            if i > 1 and i < file_len(File)+1:
                out.write(line)
                i=i+1
        f.close()
    if Mode=='AC':
        out=open('Result_AnalyseComplex.dat','w')
        out.write("Temperatur Pdb Total Energy Backbone Hbond Sidechain Hbond Van der Waals Electrostatics Solvation Polar Solvation Hydrophobic Van der Waals clashes Entropy Side Chain Entropy Main Chain Sloop Entropy Mloop Entropy Cis Bond Torsional Clash Backbone Clash Helix Dipole Water Bridge Disulfide Electrostatic Kon Partial Covalent Bonds Energy Ionisation Entropy Complex
Residue Number'n")
        File='output_\'+str(TS)+"K_ST.dat'
        f=open(File, 'r+')
        i=1
        for line in f:
            if i > 1 and i < file_len(File)+1:
                out.write(line)
                i=i+1
        f.close()

def Collect_Result(t,t_max,T_int,Mode): # Mode ST=stability AC=AnalyseComplex Data collection and generation of Result files
    if Mode=='ST':
        out=open('Result_Stability.dat','w')
        out.write("Temperatur Pdb Total Energy Backbone Hbond Sidechain Hbond Van der Waals Electrostatics Solvation Polar Solvation Hydrophobic Van der Waals clashes Entropy Side Chain Entropy Main Chain Sloop Entropy Mloop Entropy Cis Bond Torsional Clash Backbone Clash Helix Dipole Water Bridge Disulfide Electrostatic Kon Partial Covalent Bonds Energy Ionisation Entropy Complex
Residue Number'n")
        while t<t_max+1:
            File='output_\'+str(t)+"K_ST.dat'
            f=open(File, 'r+')
            i=1
            for line in f:
                if i > 1 and i < file_len(File)+1:
                    out.write(line)
                    i=i+1
            f.close()
            t=t+T_int


if Mode=='AC':
    out=open('Result_AnalyseComplex.dat','w')
    out.write('Temperatur Pdb Group1 Group2 IntraclashesGroup1 StabilityGroup2 Interaction Energy StabilityGroup1
    Folder for AnalyseComplex result deposit created
    while t<t_max+1:
        File='output_'+str(t)+'K_AC.dat'
        f=open(File, 'r+')
        i=
        for line in f:
            if i > 1 and i < file_len(File)+1:
                out.write(line)
                i=i+1
        f.close()
    t=t+i

def foldx(t,t_max,t_int,Mode):
    if Mode == 'BM':
        Config_BM()
        cmd='foldx.exe -f config.cfg'
        subprocess.call(cmd, shell=True)
        subprocess.call('exit 1', shell=True)
        Collect_BM()
        if Modes[3]=='RP':
            File_names=Files(D+\Data_BM\PdbList_New.fxout)
            i=0
            while i < len(File_names):
                Config_RP(File_names[i][0])
                cmd='foldx.exe -f config.cfg'
                subprocess.call(cmd, shell=True)
                subprocess.call('exit 1', shell=True)
                i=i+1
        Collect_RP()
    if TS!=0:
        if Mode == 'ST':
            Config_ST(TS)
        if Mode == 'AC':
            Config_AC(TS)
            cmd='foldx.exe -f config.cfg'
            subprocess.call(cmd, shell=True)
            subprocess.call('exit 1', shell=True)
        if Mode == 'ST':
            Collect_ST(str(TS))
        if not os.path.exists(D+\Data_ST'):
            Folder for Stability result deposit created
            os.makedirs(D+\Data_ST')
            print('Folder for AnalyseComplex result deposit created
            at:\n'+D+\Results\')
            for file in glob.glob(D+\"\ST.fxout"): #Move files from Stability to data
                F=os.path.splitext(file)
                shutil.move(file,D+\Data_ST\"+str(TS)+'K'+F[1])
        if Mode == 'AC':
            Collect_AC(str(TS))
            if not os.path.exists(D+\Data_AC'):
                Folder for AnalyseComplex result deposit created
                os.makedirs(D+\Data_AC')
            os.makedirs(D+\\Results\')
            print('Folder for AnalyseComplex result deposit created
            at:\n'+D+\Results\')
            for file in glob.glob(D+\AC.fxout): #Move files from AnalyseComplex to
                os.path.splitext(file)
                shutil.move(file,D+\Data_AC\"+str(TS)+'K'+F[1])
```python
# T=str(t)
else:
    if t!=t_max:
        x=(t_max-t)/t_int
    else:
        x=1
    i=0
while i <= x:
    if Mode == 'ST':
        Config_ST(t+(t_int*i))
    if Mode == 'AC':
        Config_AC(t+(t_int*i))
    if Mode == 'ST' or Mode == 'AC':
        subprocess.call(cmd, shell=True)
    T=str(t)
    if Mode == 'ST':
        Collect_ST(str(t+i*T_int))
        if not os.path.exists(D+'\Data_ST'): # checks for and if needed creates
            folder to ST data files
            os.makedirs(D+'\Data_ST')
            print('Folder for Stability result deposit created
at:
D+\Results\n')
    for file in glob.glob(D+'\*ST.fxout'): # Move files from Stability to
        data folder and overwrite files in folder with same name, no promt.
        F=os.path.split(file)
        shutil.move(file,D+'\Data_ST\'+str(t+(t_int*i))+'K'+F[1])
    if Mode == 'AC':
        Collect_AC(str(t+i*T_int))
        if not os.path.exists(D+'\Data_AC'): # checks for and if needed creates
            folder to AC data files
            os.makedirs(D+'\Data_AC')
            print('Folder for AnalyseComplex result deposit created
at:
D+\Results\n')
    for file in glob.glob(D+'\*AC.fxout'): # Move files from AnalyseComplex
to data folder and overwrite files in folder with same name, no promt.
        F=os.path.split(file)
        shutil.move(file,D+'\Data_AC\'+str(t+(t_int*i))+'K'+F[1])
        i=i+1
    else:
        break

def Cleanup(): # Moving file in to folder
    print('----------------------------------------
')
    if Modes[0] == 'BM':
        print('New PDB build with the following Mutation:
')
        F=open('PDB_BM.txt', 'r+')
        for line in F:
            SPLIT = str.split(line, '\n')
            print(SPLIT[0]+':
')
        F.close()

        print('')
    if Modes[1] == 'ST':
        print('Stability analysis preformed of the following PDB files:
')
```

114
```python
#open(D+'\Data_BM\PdbList_New.fxout','r+')
for line in F:
    SPLIT = str.split(line, '\n')
    print('\t'+SPLIT[0])
F.close()
print('')
if Modes[2]=='AC':
    print('AnalyseComplex analysis preformed of the following PDB files:')
F=open(D+'\Data_BM\PdbList_New.fxout','r+')
for line in F:
    SPLIT = str.split(line, '\n')
    print('\t'+SPLIT[0])
F.close()
print('')

print('----- File Information -----')
if Modes[1]=='ST' or Modes[2]=='AC':
    if not os.path.exists(D+'\Results\'): #checks Results folder and creates if necessary
        os.makedirs(D+'\Results\')
        print('Folder for result deposit created at:\n'+D+'\Results\')
    if Modes[0]=='BM':
        print('PDB files of Mutants are plased in the following folder:\n'+D+'\PDB\')
    if Modes[1]=='ST':
        if os.path.exists(D+'\Result_}\stab.doc.dat\'): #checks for and if present move
            shutil.move(D+'\Result_}\stab.doc.dat\'),D+'\Result_\stab.doc.dat\'
            print('The result from Stability analysis are available in the following folder as Result_}\stab.doc.dat:\n'+D+'\Result_\stab.doc.dat\n')
            else:
                print('No Result file for Stability analysis\n')
        if Modes[2]=='AC':
            if os.path.exists(D+'\Result_}\analyse.doc.dat\'): #checks for and if present move
                shutil.move(D+'\Result_}\analyse.doc.dat\'),D+'\Result_\analyse.doc.dat\'
                print('The result from AnalyseComplex analysis are available in the following folder as Result_}\analyse.doc.dat:\n'+D+'\Result_\analyse.doc.dat\n')
                else:
                    print('No Result file for AnalyseComplex analysis\n')
        for file in glob.glob(D+'\*.dat\'): #remove temp files,*.dat
            os.remove(file)
        for file in glob.glob(D+'\*.cfg\'): #remove temp files, *.cfg
            os.remove(file)
        for file in glob.glob(D+'\*.pdb\'): #remove temp files, *.pdb
            os.remove(file)
        for file in glob.glob(D+'\*.fxout\'): #remove temp files, *.fxout
            os.remove(file)
    def Credits(): #Credits to the FoldX team
        print('')
        print('FoldX 4 (c)')
        print('code by the FoldX Consortium')
        print('Jesper Borg, Frederic Rousseau')
        print('Joost Schymkowitz, Luis Serrano')
        print('Peter Vanhe, Erik Verschueren')
        print('Lien Beaten, Jevier Delgado')
        print('And Francois Stricher')
        print('And any other if the 9! permutations')
```
Based on an original concept by Raphael Guerois and Luis Serrano

raw_input('\n\n----- Press any key to continue to run overview. -----')

PDBDIR=PDB_Path()
intro()
setup()
T,T_max,T_int,TS = Temp(T,T_max,T_int)
i=0
if TS!=0:
    while i< len(Modes):
        foldx(T,T_max,T_int,Modes[i])
        Collect_Result_TS(TS,Modes[i])
        i=i+1
    while i< len(Modes):
        foldx(T,T_max,T_int,Modes[i])
        Collect_Result(T,T_max,T_int,Modes[i])
        i=i+1
else:
    while i< len(Modes):
        foldx(T,T_max,T_int,Modes[i])
        Collect_Result(T,T_max,T_int,Modes[i])
        i=i+1
Credits()
Cleanup()

print('\n---------------------------------------------------------------------\n')
print('Script by Daniel Stiefler-Jensen, e-mail: Stiefler.jensen@gmail.com\n')
Quit=raw_input('----- Press any key to Exit -----')