Developing Methods to Assess Enzyme Stability in Liquid Laundry Detergents

AINSWORTH, NIAMH,MARIE

Developing Methods to Assess Enzyme Stability in Liquid Laundry Detergents

Niamh Ainsworth

A Thesis Presented for the Degree of
Doctor of Philosophy

Department of Chemistry
University of Durham
March 2018

Abstract

This thesis presents work focused on developing established protein analysis methods for use in studying enzyme inactivation in laundry detergent systems. In a multi-billion dollar per year industry, basic, labour intensive procedures still dominate commercial stability studies, with extensive storage tests and activity assays remaining the industry standard. These methods are both inefficient and provide little insight into inactivation processes, leading to a 'trial and error' approach to product development. This slows the introduction new formulations and enzyme variants to the market. Furthermore, a valuable opportunity is being missed, harnessing available resources in the detergent industry to advance both protein analysis technologies and understanding of protein denaturation processes.

Transfer from these basic, low throughput methods to those favoured by other protein-focused industries has been hindered by sample complexity and the presence of high concentrations of the surfactant, LAS. In this work, two novel approaches to enzyme analysis in LAS-rich media will be presented. The first employing an analogous surfactant, SDS, which yields similar effects on protein stability but does not affect UV detection, and the second, exploiting the irreversible nature of detergent enzyme unfolding to enable manipulation of formulations to within instrument specifications. These approaches will allow for incorporation of ultra-high throughput screening methods, such as DSF, as well as techniques which provide further insight into protein unfolding processes, such as CD, to the available suite of analytical techniques.

Thermal data arising from this work were compared with rates of degradation obtained through conventional storage tests. Empirical fittings suggest a linear relationship between T_m values and long-term storage stability, enabling the use of thermal analysis as a tool for prediction of degradation rates. Further work is required to refine these models, however, before expanding to more complex systems.

Acknowledgements

A huge thank you to both of my fantastic supervisors, Drs AnnMarie O'Donoghue and David Hodgson. Your support and advice, both in Durham and on my Belgian holiday are greatly appreciated. As are all the pizza and BBQs. Thanks also to Jean Francois Bodet, the MICSED supervisory team, John Evans, Euan McGennis and Julie Mcgloughin for keeping us all on track for the past four years. I hope the Belgian hotel food and bar were worth all the early morning flights! Special thanks also to Mark Forrest and Michelle Jackson for all their help and invitations to Longbenton, and to my fellow MICSED students, sudoku champs, Anna and Benjamin and german Diaper Duo, Elise and Salvatore. All the best for graduation and the future!

Another massive thank you to everyone from the Hodgeson,O' Donoghue and Mosely groups, past and present. I'm hugely grateful to Stefanie Freitag-Pohl, who did so much work on this project. Thanks for all your help, commitment and especially patience! Special thanks as well to Andy, Lottie and Vian, who never missed a Food Friday, and Peter, Neshat and Chris who always made time for a quiz. To Becky and Naomi, Lunchtime Funtimes were often the only thing keeping me in the building. All the group venting is very much appreciated. Also immense thanks to my other half, Vicki Linthwaite. You deserve a medal for all your proof reading and flawless science support. Not to mention maintaining my sanity for the last 6 months. Can't wait for our next adventure/trip to Passing Clouds.

Any residual health and fitness I owe to Kathleen Reynolds. Thanks for dragging me out of bed at unearthly hours, I swear I never resented it. And secondly, to the DUMC kids for luring me climbing with the promise of the Colpitts, especially Steph, for South Africa hand in motivation, and my fellow moron and sister, Hannah. I'm also forever indebted to my long-suffering housemate Andrei Markin, the world's most persistent alarm clock. Thanks for always being up for a strategic Baileys. I promise to resume regular sleeping patterns as soon as possible. (+500 housemate points.)

I owe so much gratitude to all my friends and family for their long-ish distance support. Colleen, Tofie, Orla, Beccah, Rachel and Lexy, thanks for all the adventures and never losing touch. Thanks also to my brothers, Rob and Niall for always making it to Brady's and picking up the slack on every last-minute birthday present for the past four years. Finally, a massive thank you to my parents, who never let a little bit of terrorism get in the way of visiting their daughter, and always made sure to have some Tayto waiting in the press.

Declaration

The work presented within this thesis was carried out at the Department of Chemistry at Durham University between October 2014 and December 2017. All of the work presented in this thesis was carried out by the author unless otherwise stated and has not previously been submitted for a degree at this, or any other university.

Statement of Copyright

Table of Contents

List of Figures

1.0 An Introduction to Enzyme Analysis in Complex Detergent Formulations

3.0 The Challenge of Enzyme Analysis in Laundry Formulations

4.0 Alternative Approaches to Enzyme Analysis in LAS-Rich Environments

5.0 Predicting Storage Stability of Detergent Enzymes by Thermal Analysis

Abbreviations

1,8 ANS	1-Anilinonaphthalene-8-sulfonic Acid
AE	Alkyl ethoxylates
AEO	Alcohol ethoxylates
CCVJ	9-(2-carboxy-2-cyanovinyl)julolidine
CD	Circular dichroism
CMC	Critical micelle concentration
Cp	Heat capacity
DSC	Differential scanning calorimetry
DSF	Differential scanning fluorimetry
EDTA	Ethylenediaminetetraacetic acid
EPS	Ethylidene-paranitrophenol-glucose-7
HDL	Heavy duty liquid (detergent)
HEDP	1-hydroxyethane 1,1-diphosphonic acid
ITC	Isothermal titration calorimetry
LAS	Linear alkylbenzene sulfonates
L-CPL	Left-handed circularly polarised light
mdeg	Millidegrees (of CD)
MEA	Monoethanolamine
MST	Microscale thermophoresis
PCA	Principle component analysis
PMSF	Phenylmethylsulfonyl fluoride
PNA	Para-nitroaniline
q-PCR	Quantitative polymerase chain reaction
ITC	Isothermal titration calorimetry
LAS	Linear alkylbenzene sulfonates
L-CPL	Left-handed circularly polarised light
mdeg CD	Millidegrees of circular dichroism
MEA	Monoethanolamine
MST	Microscale thermophoresis
PCA	Principle component analysis

PMSF	Phenylmethylsulfonyl fluoride
PNA	Para-nitroaniline
q-PCR	Quantitative polymerase chain reaction
R-CPL	Right-handed circularly polarised light
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
suc-AAPF-pNA	*N*-succinyl-Ala-Ala-Pro-Phe-*p*-nitroanilide
$T_{1/2}$	Half life
TICTs	Twisted Intramolecular Charge Complexes
TAGG	Aggregation temperature
T_m	Melting temperature

1 An Introduction to Enzyme Analysis in Complex Detergent Formulations

Modern laundry detergents are complex systems consisting of a broad range of compounds including surfactants, bleaching agents, chelators, builders, structurants, dyes and perfumes. Each additive plays a different role in the cleaning action, aesthetics or stability of the formulation. Since the introduction of enzymes in the 1960s, biological laundry products have been market leaders in the detergent industry, comprising more than 80% of sales in the US, Europe and Japan.[1] The popularity of these formulations stems from their ability to remove stains without the need for high temperatures or severe agitation, which are expensive, and reduce the lifetime of fabrics. Biological components account for less than 0.1% of commercial products by weight, however, they are responsible for much of the cleaning power of the detergent.

Protein interactions with other laundry additives can have adverse effects on enzyme structure and activity, however, as they can freely associate under both storage and wash conditions. In traditional powdered detergents, this effect was reduced by keeping the formulation dry and packing enzymes in separate granules, protected by sugars, salts and waxy hydrophobic builders. Liquid laundry formulations, however, are becoming increasingly popular, accounting for 25% of global detergent sales and dominating North American and European markets.[2,3] The success of these products is mainly due to reasons of end-user convenience and ease of transportation. As enzymes are free to interact with other detergent components throughout the product lifetime, further demands are placed

on protein stability, necessitating greater understanding of interactions with each additive.

Recent growth of laundry markets in less developed parts of the world has also highlighted the need for improved stability in biological formulations.[2] In more established markets, products tend to be transported and stored in mild conditions, ensuring formulations are kept dry and cool. Product turnover is high, resulting in relatively short periods of storage. Developing markets are associated with stressed conditions with high humidity and extended storage times as such detergents are considered luxury items.

Catering to a global market demands product performance at a high standard across a range of wash conditions. A typical wash operates at high temperatures of between 30 ^{0}C and 60 ^{0}C and at alkaline pH of between 8 and 12, which are non-optimal for many enzymes. Furthermore, these conditions can vary greatly between countries and individual users, emphasising the importance of detailed understanding of enzyme structure and activity.

The study of enzymes in laundry systems presents an analytical challenge, however, due to the number of freely interacting components, and the highly coloured and viscous nature of formulations, which interfere with many analytical methods. The current standard for monitoring enzyme stability in fully formulated detergents involves extensive storage tests combined with enzyme activity assays. These experiments are time consuming, taking up to 12 weeks at accelerated rates, and provide little insight into the mechanisms of inactivation. This lack of understanding has resulted in a 'trial and error' approach to product development. This thesis presents options, relevant to the detergent industry, for improved methods of rapidly determining enzyme stability. This would reduce screening times to hours, rather than the weeks necessitated by current storage tests. Furthermore, techniques selected for use in this study have

the potential to provide greater insight into denaturing processes. This would enable transfer to a 'stability by design' approach to formulation development, rather than after the fact testing currently employed. This approach reduces both time and financial investment associated with establishing new commercial products. An overview of available methods will be presented in *Section 1.3*.

1.1 Enzyme Structure and Function

Enzyme inactivation in detergent systems is caused by changes to native protein structure. This can occur in the form of conformational unfolding or fragmentation of the peptide chain. Unfolding is generally the result of temperature or chemically-induced stress, whereas fragmentation is due to proteolysis. To understand these processes, we must first have a basic understanding of native protein folding.

Protein folding is divided into four levels of increasing structure, from primary to quaternary. The most basic; primary structure, refers to the order of amino acids in the peptide chain as coded by DNA and RNA through transcription and translation. Today, this can easily be determined using mass spectrometry. The protein sequence dictates the eventual native conformation, as well as protein function. It is, however not currently possible to predict structure or function by studying the primary structure alone.[4]

Secondary structures refer to folding which occurs due to interactions between residues positioned locally in the peptide sequence. The most commonly observed motifs are α-helices and β-pleated sheets. Helices are produced through hydrogen bonding formation between the carboxyl oxygen of one amino acid and the amide hydrogen of the amino acid three positions away. Repetition of this pattern creates a right-handed helix with 3.6 amino acids per turn and a pitch of 5.6Å. The side chains of the helix point outwards and down, towards the N-terminus. Helices are generally

unstable in aqueous solution as the high entropic cost of formation outweighs the weak hydrogen bonding. In native protein conformations, however, the exclusion of water overcomes this barrier by facilitating the formation of stabilising crosslinks between proximal side-chains, making alpha helices a highly stable fold.[5,6]

β-strands form along extended portions of polypeptides of 3-10 amino acids in length. Multiple β-strands existing in parallel are linked by hydrogen bond interactions between carboxyl and amide groups of the opposite chain, creating the β-sheet. Chains can run parallel or antiparallel, with the latter being more stable as groups fit more closely together. Side chains point straight up and down from the plane of the sheet, creating the 'pleat'.[5,7] Combinations of secondary structures occurring within a protein are known as motifs. Often, a motif will be replicated across a range of proteins from the same family or which preform similar functions. Examples of these include TIM barrels, found in 10% of all enzymatic proteins, and Greek key motifs, found more specifically in amylases. Both will be discussed in *Section 1.1.2*.[8,9]

Further folding, known as tertiary structure, occurs between side chains of residues which are removed from each other in the protein sequence. These interactions include salt bridges formed between acidic and basic residues, hydrophobic interactions and di-sulphide bridges formed between cysteine residues. This level of folding is required to orientate residues into active sites and binding pockets, providing protein function. Some proteins also have quaternary structures, referring to the combination of two or more folded peptide chains to form a single protein, such as that seen in haemoglobin.

Common folding patterns are observed within protein classes, often despite very different primary sequences. As a result, genetic engineering of detergent enzymes focuses on improving binding or structural stability without dramatically altering the protein fold. We will now look at the

general structures of the detergent enzyme classes, proteases (subtilisin-based), amylases (1,4 glucan hydrolase based) and lipases derived from pancreatic lipase.

1.1.1 Proteases

Proteases were the first enzymes to be introduced to commercial laundry formulations, with an aim to improve the cleaning capabilities of surfactants on protein-based stains such as sweat and sebum. Today, proteases alone are responsible for a 20% increase in stain removal over the same, non-biologic formulation.[10] The first biological detergents, developed by Jaag *et al*[11] used porcine pancreatic trypsin, however, this was quickly superseded by microbial enzymes of the class 'Serine Protease II', also known as 'Subtilisin-like proteases'. In turn, these were replaced by Savinase, a highly alkaline peptidase, engineered to have a lower isoelectric point, thus making it more compatible with basic detergents.[11–13]

Savinase, and its modern derivatives are single domain enzymes of approximately 27 kDa. The folded globular protein exists as a hemisphere with a diameter of 40 Å. The active site is accessed via the flat side of the hemisphere, within a substrate binding channel of parallel β-strands. Binding of a polypeptide or inhibitor creates an anti-parallel β-sheet, which is more stable than its parallel counterpart, driving substrate binding. Binding sites along the channel interact with polypeptide side chains to give substrate specificity.[12,14]

The catalytic triad is identical to that of trypsin, though the two are structurally and evolutionarily different. This is a popular example of convergent evolution. The active site residues are situated at the carboxy terminus of the protein, buried in the core of the protein. Catalysis is achieved through a serine residue in the protein chain, which acts as a nucleophile and binds the peptide at its carbonyl carbon to form the enzyme-substrate complex. This intermediate is subsequently hydrolysed

and the products are released, freeing the binding site for further catalysis, as illustrated in *Figure 1*.[14,15]

Protein engineering has improved the stability and functionality of Savinase-based proteases in laundry systems. The pH profile was changed by modifying the protein surface charge. Hydrophobic residues were introduced to the binding loop to increase alkalinity for detergent conditions of pH 8-12. Thermal stability was improved by increasing the number of salt bridges to 7, from the 3-5 found in subtilisins, despite having fewer basic residues. Yet more of these interactions can occur on speciation at high pH..[13,16] Thermal stability was improved further through the deletion of single residues from unstructured portions of the protein. This promoted the formation of small β-sheets which bind the vulnerable loose N-terminus to the body of the protein via the Ca^{2+} binding loop.[12,17]

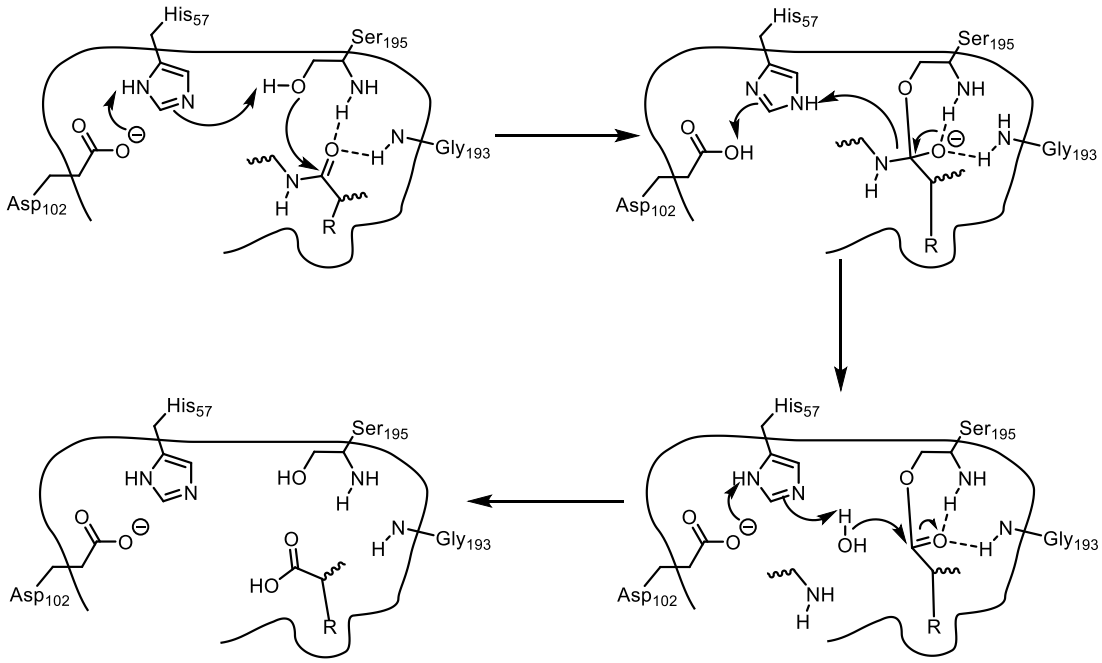


Figure 1: Hydrolysis of a peptide by a subtilsin-based protease.

1.1.2 Amylases

The success of early biological detergents led to the inclusion of amylases and lipases to facilitate removal of a broader range of stains. Amylases, of

the class α-amylase and derived from 1,4-glucosidase, target starch-based stains such as amylopectin. The mechanism of action involves the breakdown of long polysaccharides into shorter chain molecules and single sugar monomers. These subunits are more soluble under wash conditions.

Termamyl was the first thermophilic α-amylase used in commercial formulations. It was developed from a chimeric bacterial glucanhydrolase; a combination of the highly thermostable regions of B. amyloliquefaciens and B. licenformis strains. Modern developments, as with detergent proteases, increased the pH profile of the enzymes, but also lowered the temperature range, as the advent of cool wash cycles has reduced the need for such high thermophilicity.[18–20]

The α-amylases consist of three distinct protein domains. Domain A incorporates the N-terminus of the enzyme, as well as the residues of the catalytic triad; two aspartic acid residues and a glutamic acid. This region takes the structure of a centralised α/β barrel *(Figure 2a),* situated at the core of the enzyme. Domain B is a loop domain and provides the greatest source of variability between α-amylases. The ends of the complex loop are linked to the third β-strand and the third α-helix of the Domain A barrel. The final Domain, C takes the form of a Greek key motif, containing the C-terminus *(Figure 2b).*[21–23]

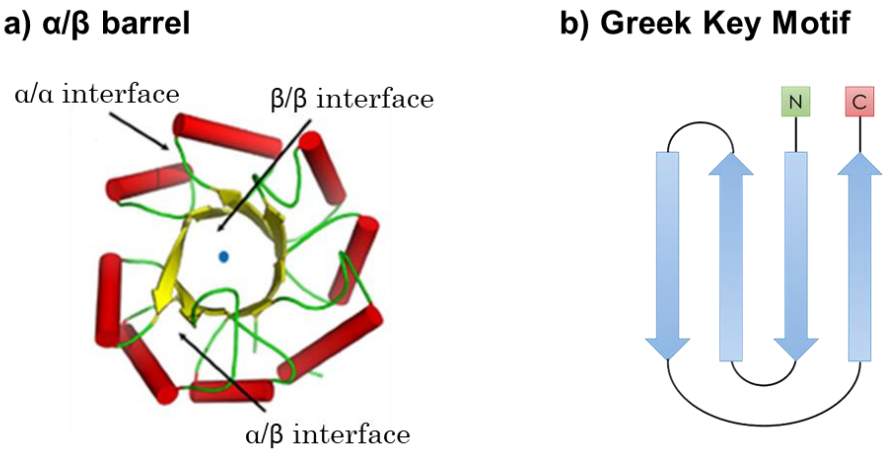


Figure 2: Protein motifs of α-amylases. a) α/β barrel active site, reproduced with permission from Vijayabaskar and Vishveshwara[24]. b) Greek key motif of Domain C.

Catalysis is achieved through two acidic residues, aspartic acid (Asp) and glutamic acid (Glu). These promote hydrolysis of the glycosidic bonds which link monomers into a polysaccharide chain *(Figure 3).* The nucleophilic charged oxygen of Asp attacks at C_1 of a glucose subunit in the chain, making a leaving group of the remainder of the polysaccharide. Glu acts as a proton donor to the newly electrophilic oxygen. As the protonated glucose leaves the active site, a water molecule enters and is activated by the deprotonated Glu. The bond between Asp and C_1 is then hydrolysed to release the remaining product. The third residue of the catalytic triad, a second Asp (not pictured), stabilises the formation of the intermediates through the formation of hydrogen bonds with proximal glucose hydroxyl residues. The presence of anionic surfactants in the wash aids the catalytic action of lipase by reducing the surface tension of stains through electrostatic interactions, exposing the stain to the active site.[25–27]

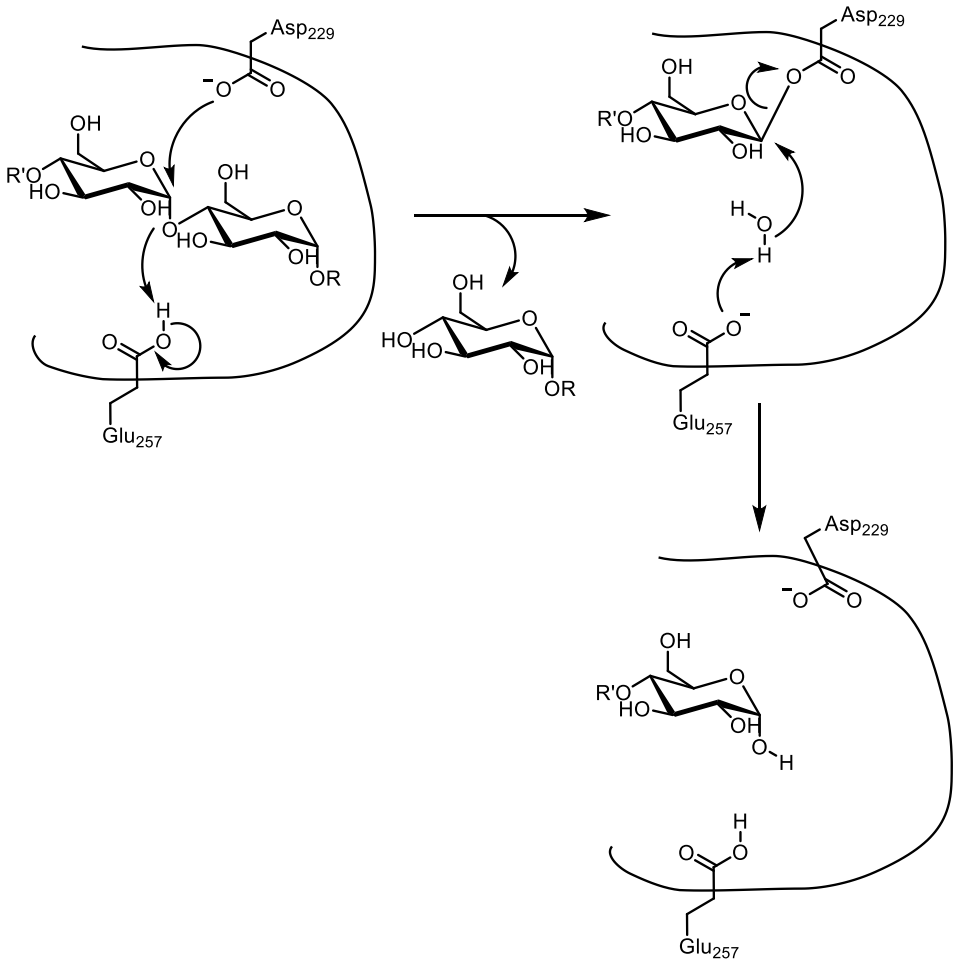


Figure 3: Hydrolysis of a glycosidic bond by an alpha-amylase.

1.1.3 Lipases

Detergent lipases combat oily stains such as those which arise from food stains, cosmetics and sebum from the skin. This is achieved through the degradation of triglycerides to the more hydrophilic compounds, glycerol and fatty acid *(Figure 4)*. The increase in hydrophilicity facilitates removal without the high temperatures necessary to liquidize triglycerides in the original stain.[28] The first use of lipases in a laundry context was recorded in 1913, however, the idea was not developed until the digestive ability of lipases in laundry was described in the 1970s. The first commercial detergent with a lipase enzyme was released in 1987 by a Japanese company, Lion.[29] Initially mammalian pancreatic lipases were used before the development of recombinant Lipex and Lipolase from Thermomyces

lanuinosu by Novozymes in the 1990's. These are naturally thermally stable and highly alkaline.[30]

True lipases are distinguished from general esterases by the existence of an interfacial activation at the oil-water interface. This aids the accessibility of an oily stain to the aqueous detergent components in a similar fashion to surfactants. Detergent lipases are from the α/β-hydrolase fold family and generally have two distinct protein domains. The N-terminal domain, which is the larger of the two, takes on an α/β fold, with a large central β-sheet featuring seven parallel and two antiparallel strands. The smaller c-terminal domain consists of a β-sandwich, two layers of β-sheets containing four antiparallel strands each.[31–33]

Serine, histidine and aspartic acid make up the lipase catalytic triad, similar to that of serine proteases and chemotrypsin (*Section 1.1.1*). The active site is covered by an α-helix lid. This rigid body moves on two 'hinges'. The 'closed' state is strained due to distortion of the alpha helix. On substrate binding, the lid is opened through a translocation of disulphide bridges in the presence of a reducing agent. The reaction mechanism then follows in a similar fashion to subtilisin *(*Figure 4: Hydrolysis of a triglceride by lipase.*Figure 4)*.[30,34]

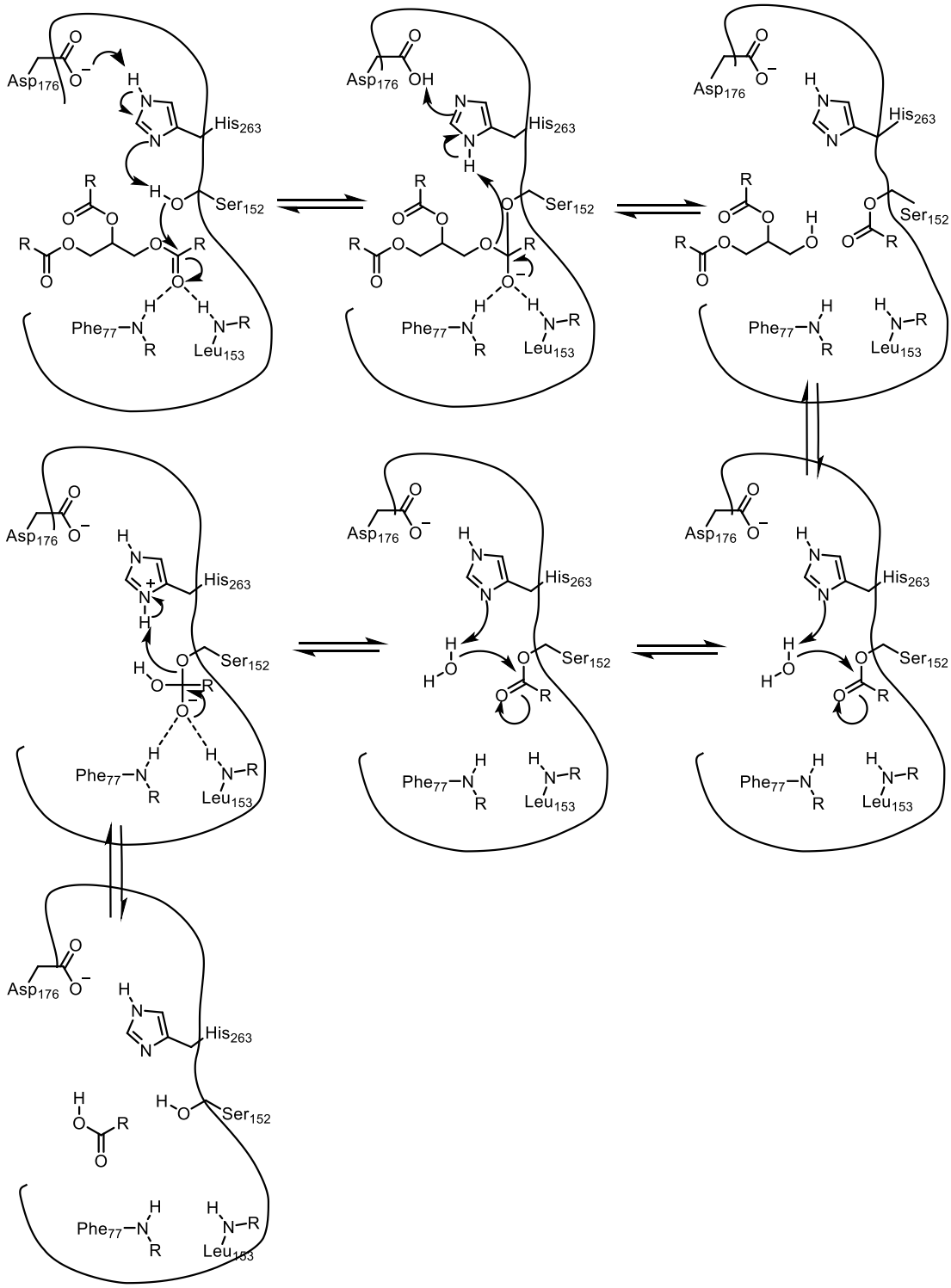


Figure 4: Hydrolysis of a triglceride by lipase.[31–33]

The introduction of lipases was slowed due to issues of instability in the presence of surfactant and oxidation of crucial disulphide bridges by bleach. Enzyme production was also expensive, driving up the cost of biological detergents until 1988 when Novozyme began the mass-production of lipases through genetic engineering of *Humicola* fungi

strains expressed in *Aspergillus oryzae*.[35,36] Following this, commercial laundry detergents containing lipases could be found in markets worldwide.[29] Bleaches, which target cysteine residues, necessary for lid-opening, are also less commonly in modern liquid formulations, and are instead sold as separate laundry additives, reducing the strain on lipase stability.

Optimal conditions for the activity of lipases consist of temperatures of 40 °C, a water concentration of 10-40% and a pH of 6-10. Low water concentration requirements result in the majority of lipase activity occurring during fabric drying. Stains are then removed in the subsequent wash. Multiple washes can therefore be required before the effects of the enzyme can be seen. At lower pH, the liberated fatty acids remain in a solid hydrophobic state with similar characteristics to the original triglyceride stain, preventing any observable improvement in stain removal. Fortunately, the majority of detergents operate in the high pH region, enabling efficient removal under wash conditions. [28,29]

1.2 Mechanisms of Protein Denaturation

Loss of protein structure, or denaturation, leads to enzyme inactivation as crucial residues are no longer in fixed proximity in active and binding sites. Unfolding is caused when the non-covalent interactions along the peptide chain, responsible for maintaining secondary and tertiary structures are interrupted. This can be caused by heat, pH or chemical factors. Destabilisation tends to occur in two stages. The first, reversible, stage involves unfolding of tertiary structure, while the second, irreversible, stage results in loss of secondary structure. Although the latter leads to protein aggregation and precipitation from solution, it is possible that some enzyme activity may be recoverable in cases where only tertiary degradation has occurred. These processes are described by the Lumry-Eyring model of unfolding $N \rightleftharpoons \mathrm{U} \rightarrow \mathrm{F}$ (Eq. 1***Error! Reference source not found.***).[37]

$$N \rightleftharpoons U \rightarrow F \qquad \textbf{(Eq. 1)}$$

where N is the native state, U is the reversible unfolded state and F, the final irreversibly denatured state.

The interactions of detergent enzymes with several other laundry components can accelerate the rate of denaturation. Surfactants, chelators, builders and proteases are particularly destabilising.

1.2.1 Contribution of Surfactants to Protein Instability

Surfactants are amphiphilic molecules, consisting of a hydrophilic head and a hydrophobic tail. Before the introduction of enzymes to detergents, surfactants provided the primary mechanism for the removal and suspension of stains. In biological detergents, surfactants are still used to aid the solubilization of oils and to lift particulates from fabrics. Due to the amphiphilic nature of the molecules, they also help to ensure phase stability within formulations.

In aqueous solution, at high surfactant concentrations, non-polar groups aggregate to form assemblies with hydrophobic interiors and hydrophilic shells. These aggregates, known as micelles, are responsible for the cleaning action of the surfactants. Micelles form around dirt particles, encapsulating them within the hydrophobic interiors and allowing the stain to be suspended in the wash.

Micelles and monomers interact with detergent enzymes in very different ways. Skin irritation complaints are also linked only to the monomeric species. As a result, it is important to understand the critical micelle concentration (CMC) of the surfactant being used. This is the concentration above which micelle formation is favoured. A low CMC is preferable in detergent formulations.[1,38]

The CMC is governed by the zeta potential (ζ), an indicator of the net electrostatic repulsion between charged monomers in a colloidal dispersion, such as a surfactant-rich solution. This value can be estimated using electrophoresis. High values indicate a stable dispersion, with

sufficient repulsion between particles to prevent aggregation, resulting in a high CMC. In solutions with a low ζ-potential, attractive forces (hydrophobic interactions in surfactant) exceed head group repulsion.[39,40]

CMC values are not constant, and change based on the conditions of the solution. The main factors affecting aggregation in detergent formulations are electrostatic interactions, protein concentration and temperature. The ionic strength of the aqueous media and the charges on the surfactant molecules influence the repulsive forces which must be overcome for surfactants to aggregate, as indicated by the ζ-potential. The presence of protein results in sequestration of surfactant molecules, which reduces the effect concentration of 'free' monomers, leading to elevated observed CMC values.[10] Temperature can also influence surfactant aggregation. At very low temperatures, surfactants assume a crystalline state and do not form micelles. The point at which monomers are solubilised and can begin to aggregate is known as the Krafft temperature. Below this point, no CMC exists.[41]

As temperature and surfactant concentration is increased, further aggregation states are observed. In place of spherical micelles, a range of aggregates including long rod-like structures, lamellar states and bilayers. The conditions under which each of these aggregates are formed differ between surfactants. In *Figure 5*, below we can see that in the anionic surfactant SDS, these phases are well defined at each concentration and temperatures point in aqueous solution. In LAS, the predominant surfactant in laundry formulations, however, overlapping phases of various lamellar states are observed at concentrations above the CMC. This is due to the variety in chain lengths found in commodity grade LAS solutions, which provide greater flexibility in the formation of aggregates.[42,43]

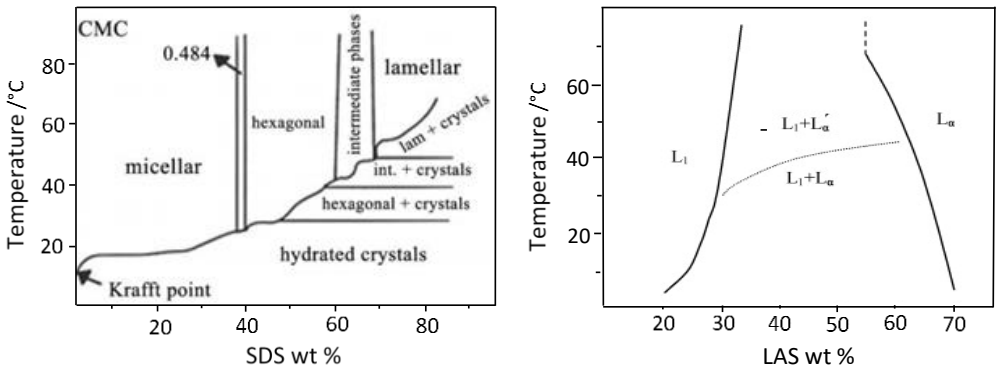


Figure 5: Phase Diagrams of SDS (left) and LAS (right) in water at various temperatures and concentrations. Figures reproduced with permission from Rossi et al[42] and Stewart et al[43].

The presence of surfactants in laundry formulations can put strain on enzyme stability, as they promote greater conformational freedom within the protein structure. This creates low-energy pathways between the natured and denatured states.[44] The extent of destabilisation is dependent on both the structure of the enzyme and the type of surfactant in question. Non-ionic surfactants tend to bind without interrupting even the tertiary structures of proteins, and in some cases have been found to promote enzyme stability. In contrast, some ionic surfactants have been found to destabilize enzyme structure even at low concentrations.[1,45,46]

As well as thermodynamic destabilisation of enzyme structures, surfactants can also act as competitive or non-competitive inhibitors of active sites. These inhibitory effects cause a reductions in the activities of the enzymes and slow the rate of hydrolysis of substrates, resulting in poor stain removal.[44]

The most common anionic surfactants are linear alkylbenzene sulfonates (LAS) and alkyl ethoxy sulfates (AES) while alcohol ethoxylates (e.g. AE) are common non-ionic surfactants *(Figure 6).* LAS is known to be particularly detrimental to protein structure. In spite of this, it is a popular choice for detergents due to its low cost, good biodegradability and the fact that it has higher thermal and chemical stability than other detergents such as soaps and sulfates.[47]

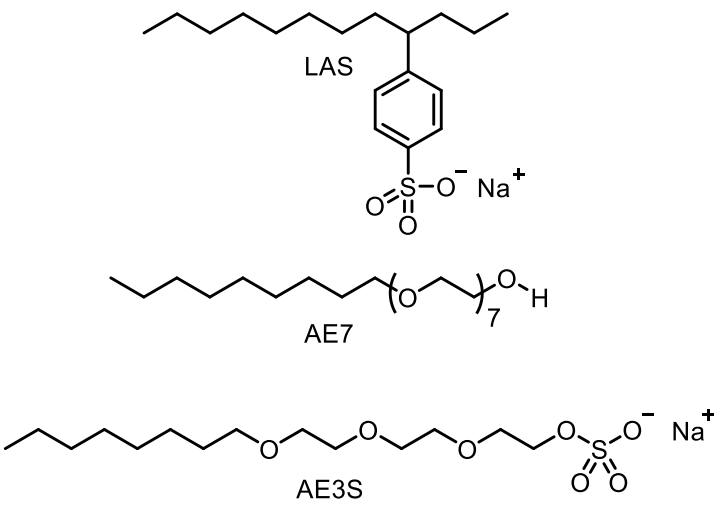


Figure 6: Structures of common surfactants; AE3S, AE7 and LAS.

Surfactant-induced enzyme unfolding is caused by conformational stress induced by protein-bound surfactant. A study conducted by Otzen[47] found that anionic surfactants were interacting with proteins through the basic amino acid residues, histidine, lysine and arginine, at their positively charged side chains. The binding of surfactant molecules to proteins occurs in a co-operative fashion, with each additional surfactant molecule making the binding of the next more favourable. This is in accordance with the Wyman Linkage Relation, which states that the presence of the surfactant stabilises the bound conformation, in which the protein is unfolded, making the unfolding thermodynamically favourable.[4] As more surfactant molecules bind, the protein becomes increasingly denatured and so exposes more surface area to facilitate further binding. This leads to hundreds of surfactant molecules surrounding the protein creating a complex which coils around the micellar aggregates.[48]

Amino acid substitutions can be used to improve the stabilities of protein structures by promoting stabilising interactions. For example, the inclusion of acidic and basic residues, which carry charges, maintains protein structure through the formation of salt bridges. Hydrogen bonding and Van der Waals forces also strengthen folding interactions. Di-sulphide bridges are not found in laundry enzymes as cysteine residues are avoided due to the oxidation they experience in the presence of bleach.

1.2.2 Effects of Chelators and Builders to Protein Instability

Chelators and builders are both added to sequester unwanted metal ions from the wash. Chelators such as EDTA *(Figure 7)* primarily target transition metals which often anchor stains to clothes, allowing them to be removed more easily from fabric. Builders such as citric acid and stearate (fatty acid) are used to sequester excess calcium and magnesium ions which are prevalent in hard water areas. These ions cause limescale build-up and the precipitation of anionic surfactant from solution.

Certain enzymes, particularly proteases and amylases, rely on Ca^{2+} ions for structural support. If the concentration of calcium is too low in the wash, this structural calcium can be sequestered by both chelators and builders, leading to unfolding.[10,49]

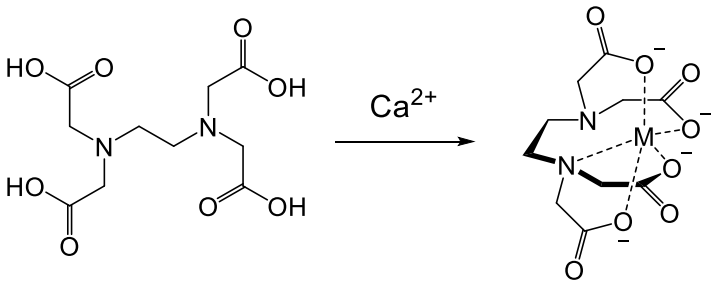


Figure 7: EDTA binding Ca^{2+}.

Subtilisin proteases have been found to be more resilient than detergent alpha-amylases to chelant-induced destabilisation due to the presence of a second calcium ion binding site. This is located at the *N*-terminus of the protein and forms an octahedral complex with Ca^{2+}. All six of the contributing oxygen atoms in the octahedron come from the subtilisin peptide chain, giving this site a higher binding affinity.[50]

The weaker binding site is found closer to the *C*-terminus and consists of the carbonyl oxygen of a glutamic acid residue and the two oxygens of an aspartic acid side chain. The geometry of the binding site resembles a 'distorted pentagonal bipyramid[51], with the four water molecules comprising the rest of the coordination sphere. As only three of the formal ligands come from the protein itself, the binding affinity of this site is much

lower. Dialysis in an excess of EDTA shows 50% removal of bound calcium from the protein, indicating that just one binding site, the weaker of the two, is affected. The remaining Ca^{2+} ion offers a degree of stabilisation that improves the resilience of the enzyme to chelant-induced denaturation. As a result, mutations aimed at improving protease stability often focus on the weaker binding site.[52]

1.2.3 Contribution of Proteases to Protein Instability

Detergent proteases are specifically selected for their high activity and low substrate specificity to improve their stain removing capabilities across a broad range of proteinaceous stains. These traits also improve their activity towards other detergent enzymes, however, increasing the susceptibility of detergent enzymes towards proteolysis. When this occurs between proteases, the process is known as proteolytic autolysis, but other classes of enzyme can also be affected. According to Lalonde[53] and Stoner[54], proteolysis is the main cause of loss of enzyme activity in liquid detergents, particularly in the case of subtilisins which are prone to autolysis. Amylases, however, are considered to be relatively resistant. This is thought to be due to their intrinsically high thermal stabilities, as partially unfolded proteins are more susceptible to protease attack.[55]

1.2.4 Methods of Improving Enzyme Stability

To date, there have been numerous patents granted dealing with the subject of stabilizing laundry formulations.[21,23-24] The use of chemical additives is a popular method of improving enzyme stability. Compounds including polyols, such as glycol and sorbitol, boric acids and borate salts and carboxylic acids have all been shown to reduce the rate of protein unfolding. However, the addition of these compounds increases the cost of production and takes up valuable space in the formulation.

Autolysis has been found to be the primary cause of irreversible denaturation of detergent proteases. Lowering water concentration in

liquid formulations slows this process, as high levels of 'free water' promote autolysis.[58] This has driven the move towards concentrated HDL products, which also have lower carbon footprints.

LAS-induced unfolding can be reduced in the presence of non-ionic surfactants. Anionic surfactants exhibit a stronger ability to degrade enzymes than their non-ionic counterparts as initial protein binding occurs via electrostatic interactions. The addition of less aggressive, ethoxylated co-surfactants sequesters LAS molecules into micelles by minimising repulsive effects between the anionic head groups.[58] These micelles are less reactive towards basic protein residues. Lowering CMC values also prevents precipitation through calcium ion binding, reducing the required concentration of chelating agents which cause further destabilisation.

To ensure high levels of enzyme activity in formulation, intrinsic protein stability is key. Modern detergent enzymes are based on naturally thermophilic proteins which are then modified to improve resistance denaturing laundry conditions.

1.3 Techniques for Monitoring Protein Stability

The study of protein stability is well established across several industrial applications, providing numerous techniques for the analysis of structure and activity. Each method provides a unique viewpoint on stability and exhibits both benefits and limitations over the others. A selection was chosen to be screened for their capabilities in assessing protein stability in HDL, as will be discussed in *Chapter 3*. The theory behind these techniques will be outlined in the following sections.

Storage tests, the industry standard for determining long-term protein stability in formulation, will be discussed first. This method delivers accurate rates of enzyme inactivation on a real-time scale; however, tests require up to 12 weeks to complete and provide little insight into the sources of instability.

To improve screening efficiency, protein-based industries have turned to thermal methods as alternative measures of stability. Several authors have reported linear correlations between melting temperatures identified by DSC and observed rates of degradation in pharmaceutical samples.[59–62] Application of these models to detergent formulations has been hindered by sample complexity which masks small protein unfolding signals. Recent work by Lund,[49,55] however, has shown that nano-DSC has the sensitivity to deliver T_m values for proteins in laundry systems. Following from this success, work in this thesis aims to reassess the capabilities of both modern and long-established techniques in determining protein melting temperatures in these complex systems. Alongside nano-DSC, an overview of isothermal scanning calorimetry (ITC), fast pulse proteolysis (FastPP), differential scanning fluorimetry (DSF), circular dichroism (CD) and microscale thermophoresis (MST) will be given in the following sections. Each method accesses unfolding parameters through different enzyme properties, yielding new insight in protein interactions and stability.

1.4 Activity Assays

Biological function assays are the most commonly used methods of monitoring activity levels of enzymes in formulations. These assays are often carried out in conjunction with storage experiments to investigate long-term stability. Original storage experiments were conducted in real-time with periodic assessments of stability levels. These have come to be replaced by accelerated stability studies which are far more efficient. In accelerated tests, samples are stored under stress conditions such as higher temperatures in order to increase the rate of denaturation. This reduces the length of time required before a significant level of activity loss is identifiable. Greater stresses will further reduce the time required; however, there is a limit to the extent to which the temperature can be raised. At very high temperatures, the effects of thermal unfolding may

interfere with the kinetic rate of degradation.[63,64] Real-time rates can then be calculated using the Arrhenius equation *(Equation 2)*.

$$k = Ae^{-E_a/RT} \quad \textbf{(Eq. 2)}$$

where k is the rate of inactivation, A is the pre-exponential factor, E_a is the activation energy, R is the universal gas constant and T is temperature

The activity of an enzyme is measured by recording either the appearance of product or the loss of substrate. The most convenient method utilises a labeled substrate which produces a colour change on conversion to products. The appearance of colour is measured by UV-Vis and plotted against time. Initial rates, observed when the concentration of surfactant far exceeds that of the enzyme, are zero-order and used to assign a value to the enzyme activity. Recording enzyme activity over the course of several weeks shows the loss in the presence of active enzyme with storage time. A half-life for the enzyme in the given formulation can then be determined using first-order kinetics *(Figure 8)*.

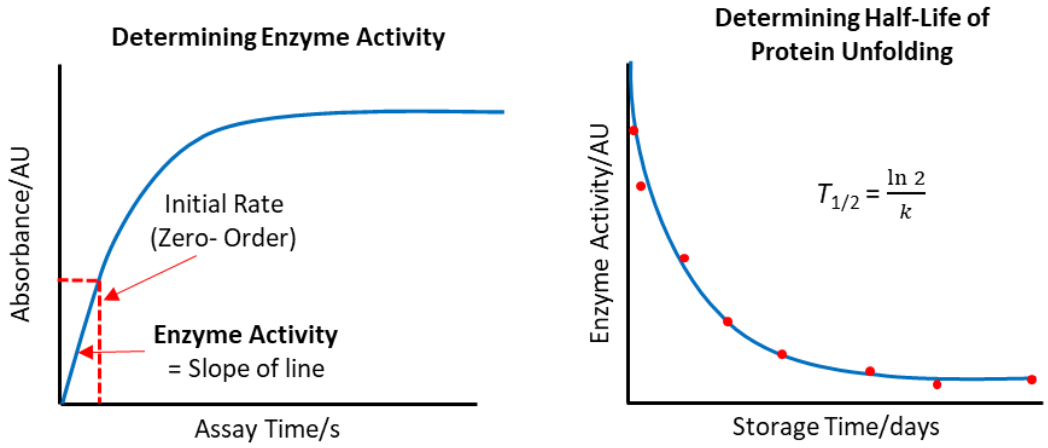


Figure 8: Sample traces of enzyme activity assays. A plot showing the change in absorbance with time can be used to calculate a value for enzyme activity. A second plot of enzyme activity against formulation storage time gives the half-life of protein denaturation at accelerated rates.

Assays must be tailored to the enzyme being studied, as many enzymes will only act on very specific substrates. Detergent enzymes, however, are designed to be active on a broad range of substrates and so there are multiple assay options. A commonly used assay for amylases involves the use of the labelled substrate ethylidene-*para*nitrophenol-glucose-7 (EPS),

which is cleaved into small glucose fragments. A second enzyme, α-glucosidase, then separates the sugars from the *p*-nitrophenol *(Figure 9)*. The liberation of the chromophore then produces the colour change.[65]

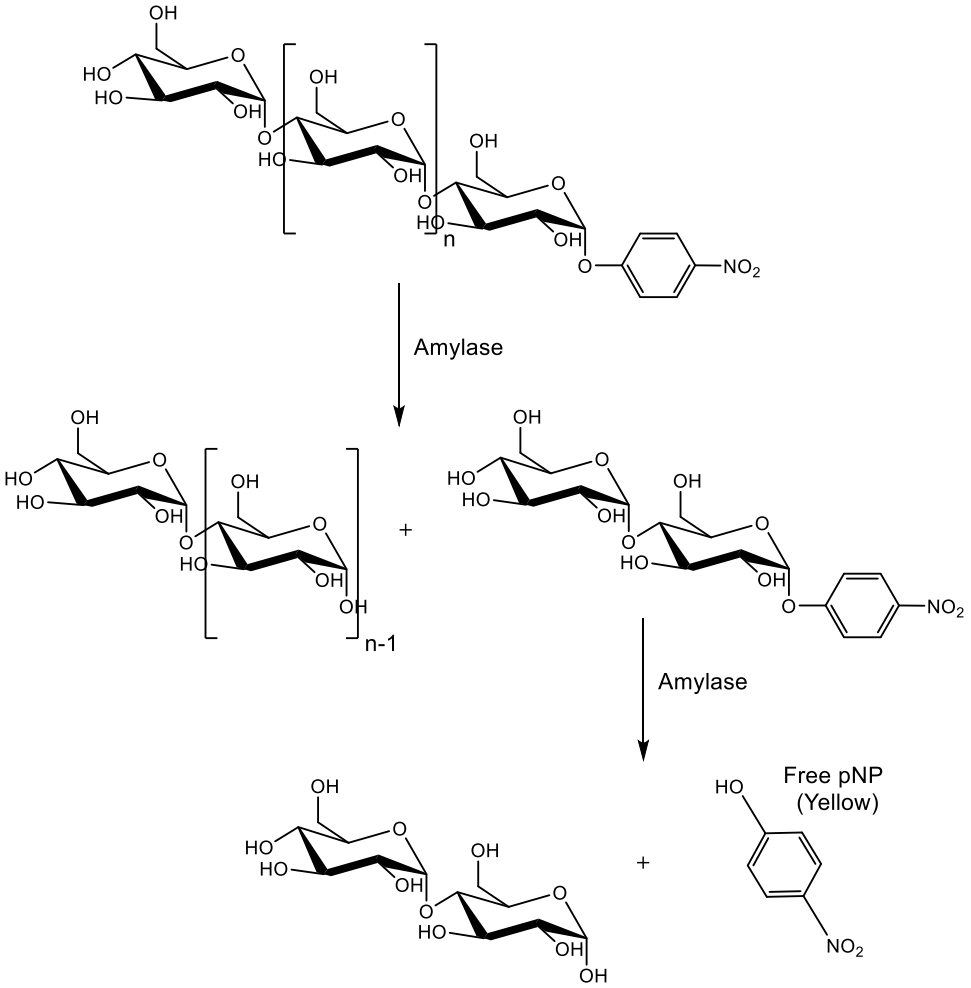


Figure 9: The use of EPS as a substrate for an amylase assay. The chromophore, pNP is liberated causing a colorimetric change.

Para-nitroaniline (PNA) is another commonly used chromophore. In protease assays, the compound is linked through an amide bond to a short peptide sequence, which acts as the substrate. The protease enzyme then breaks the amide bonds between the amino acids, including that of the chromophore *(Figure 10)*. This produces a yellow colour which can be detected at 405 nm.[20]

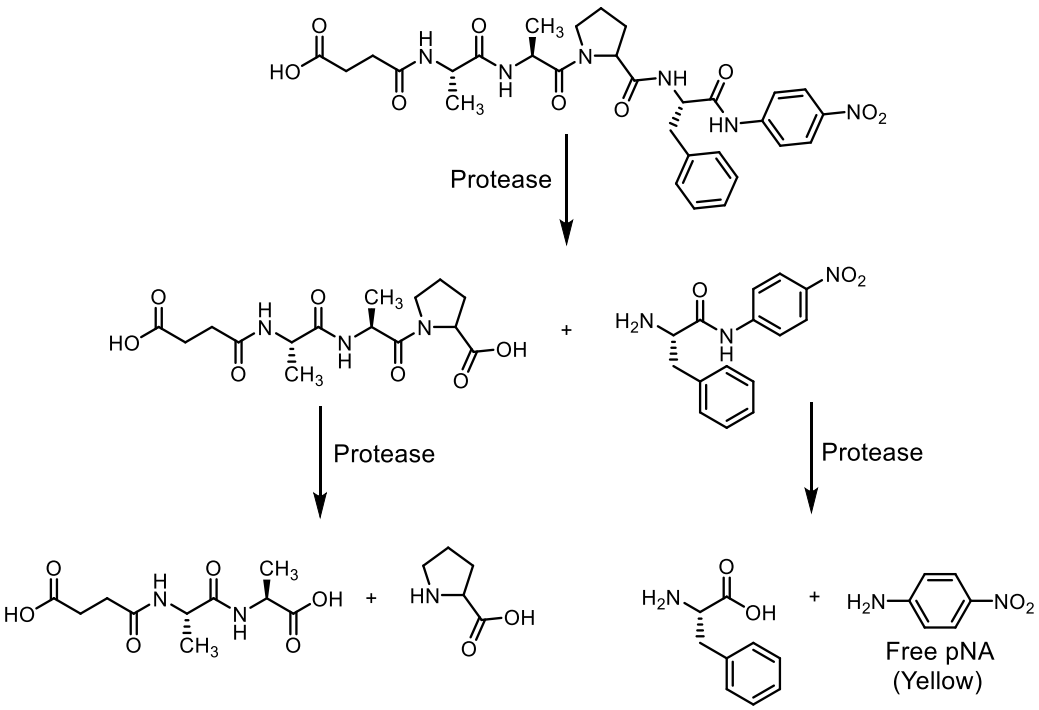


Figure 10: The use of p-nitroaniline reagent as a substrate for a protease assay. P-Nitroaniline is liberated to produce a yellow colour. The peptide chain continues to be degraded.

1.4.1 Advantages and Limitations

Chromogenic assays provide sensitive and precise data on the change in enzyme activity in various environments. The spectrophotometric measurements enable simple and accurate data analyses and the method can also be conducted with high throughput, allowing for the screening of large numbers of samples. Data from storage experiments can also be directly related to the shelf life of the product.

The assays do not provide any information on degradation processes which do not directly affect the activity, however. Small conformational changes can often be early indicators of instability and also highlight regions susceptible to unfolding. This presents limited scope to probe mechanisms behind inactivation. Storage experiments, even at accelerated rates are also lengthy and require large volumes of sample and expensive reagents. As a result, recent work in protein stability has leaned towards the use of thermal denaturation as a measure of stability, with a view to reduce the

need for routine storage tests. Examples of these alternative methods will be provided in the coming sections.

1.5 Differential Scanning Calorimetry

The first thermal technique which will be discussed is differential scanning calorimetry (DSC). T_m values are determined by measuring changes in sample heat capacity as thermal processes occur during heating. The area under the peak also provides enthalpy values for each transition. As unfolding signals are often interrupted by protein precipitation, melting temperatures are assigned to the temperature at which maximum peak height is recorded, $T_{max.}$

Heat capacity is defined as the observed enthalpy increase of a system with increasing temperature *(Eq. 3).*[66] In protein samples, changes in heat capacity (ΔC_p), with respect to a reference, indicate unfolding. Reference samples are identical to that of the analyte, but with the protein omitted. ΔC_p arises due to the latent heat associated with unfolding, which increases the amount of energy required to maintain the reference temperature.

$$C_p = \frac{\Delta H}{\Delta T} \qquad \textbf{(Eq. 3)}$$

DSC instruments can operate via 'heat flux' or 'power compensated' mechanisms. A schematic representation of heat flux DSC is presented in *Figure 11*. The instrument consists of two sample 'pans', one for the analyte sample, and one for the reference. These are heated in the same furnace, ensuring conditions are identical.[67]

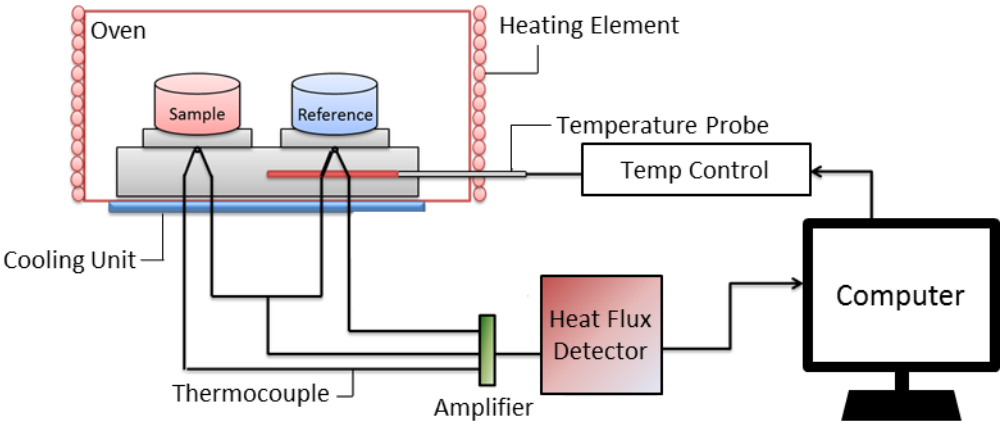


Figure 11: Schematic of a Heat Flux DSC.

As the furnace is heated, variation in the temperatures of the two pans will arise due to the latent heat associated with unfolding processes in the analyte sample. Thermocouples, attached to each pan, convert this temperature gradient to a voltage, which is used to monitor the heat flux in the cells. In a power-compensated DSC, pans are separately heated according to identical temperature gradients. The difference in energy required to raise the temperature in each pan is recorded. This is known as the difference in thermal power and is plotted against temperature in a thermal analysis curve.[68]

Both methods should give rise to a peak in ΔC_p as a function of temperature for each thermal transition *(Figure 12)*. Integration of peak area gives the enthalpy of that transition *(Eq. 4)*. The melting temperature of a protein is represented in DSC by the maximum point of the curve, as unfolding is associated with an increase in heat capacity. This is reported as the T_{max}.

$$\Delta H = \int_{T0}^{T1} C_p \, dT \qquad \textbf{(Eq. 4)}$$

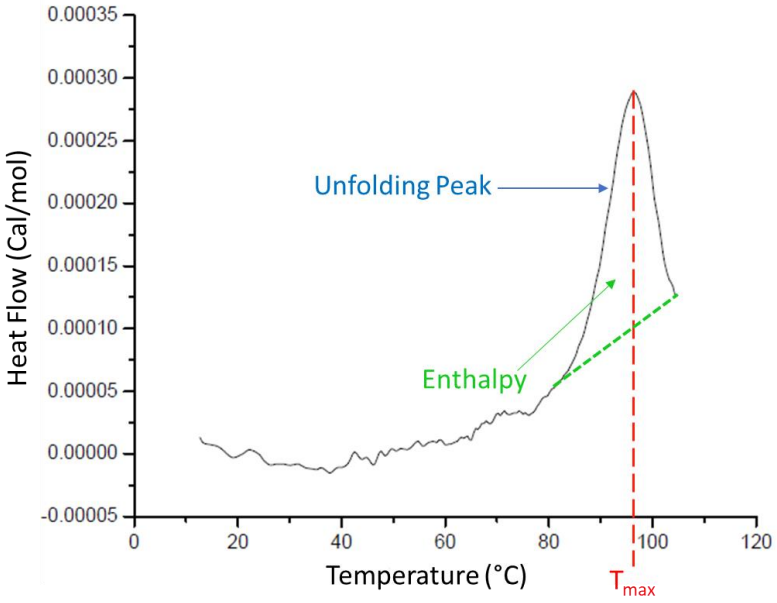


Figure 12: Sample DSC trace taken from Everest in 0.1% LAS, showing the unfolding peak, T_{max} and enthalpy (area under the peak).

1.5.1 Advantages and Limitations

The key advantage of DSC in the detergent industry, is the capability to handle complex, highly coloured and opaque formulations. Which can often present difficulties for methods using optical detectors operating in the UV-Vis region (This will be discussed in detail in *Chapter 3*). Pressurisation of cells further provides for assessment of detergent enzymes engineered for thermostability. Amylases, in particular, can maintain structure up to 90 °C in some cases. This can be beyond the scope of many other instruments due to errors arising from solvent evaporation. Small enthalpy changes associated with protein unfolding can be difficult to detect over thermal transitions occurring in the bulk, however, necessitating specialist nano-DSC equipment.[69,70]

DSC outputs provide direct access to T_{max} values of biomolecules such as proteins in a range of detergent conditions. These values have been reported by several authors[55,60,69] Further insight into intermolecular interactions can also be gained, through comparison of protein, ligand and combined protein/ligand DSC traces. Ligands with a high affinity for the native state protein will produce a different thermal analysis curve to that

of the protein and ligand individually, as the complex will have a new heat capacity. Further mechanistic insight can be obtained from parallel experiments using a second calorimetric method, ITC, which will be discussed in the following section.

Low throughput capabilities are the greatest limitation for the introduction of DSC for formulation screening. The instrument cannot run multiple samples concurrently, and each run takes approximately 2-3 hours. Run-time is dictated by the rate of heating, which is limited by slow energy transfer which can skew thermal peaks. Furthermore, the high viscosity of detergent samples prevents the incorporation of autosamplers.

1.6 Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC) is a second calorimetric technique commonly used in measuring the thermodynamics of molecular interactions in solution. ΔH is again used to identify thermal processes, in this case, however, these are associated with ligand binding, rather than unfolding.

ITC works in a similar fashion to power-compensated DSC, measuring the difference in heat energy required to maintain the sample cell at a constant temperature with respect to a reference cell. The protein analyte sample, known as the titrate, is placed into the sample cell and the second, reference cell is filled with buffer. The ligand solution, known as the titrand, is automatically titrated into the sample cell in small volumes, such as 10 µl, aliquots. After each injection, the difference in heat flow between the sample cell and reference cell is recorded, as in DSC. Cell feedback and calibration heaters keep the temperatures of the two cells constant. An adiabatic jacket prevents heat transfer to the surrounding atmosphere *(Figure 13)*.[71–73]

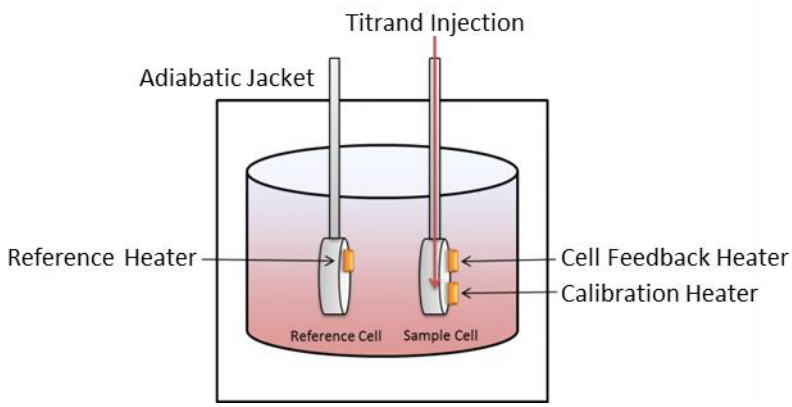


Figure 13: Schematic of Isothermal Scanning Calorimetry.

Each injection of ligand produces a thermal spike which is translated to a peak on the ITC thermal analysis curve. Peak height reduces with each injection as analyte binding sites become saturated. The heat flux after each injection is found by integrating under each peak. This is known as the apparent heat change between addition i-1 and i (Δq_i). As a titration experiment is integrated into the procedure through the addition of aliquots of ligand, the stoichiometry and equilibrium constant (K_a) of the reaction can also be calculated.

Data from ITC analysis is presented as plots of ΔH as a function of time, with spikes in enthalpy at each ligand injection *(Figure 14)*. Peak height denotes the enthalpy of binding. A plot of accumulated heat energy after each injection as a function of total accumulated ligand concentration yields a sigmoidal curve which can be used to calculate the total heat change per mole (ΔH_{app}).

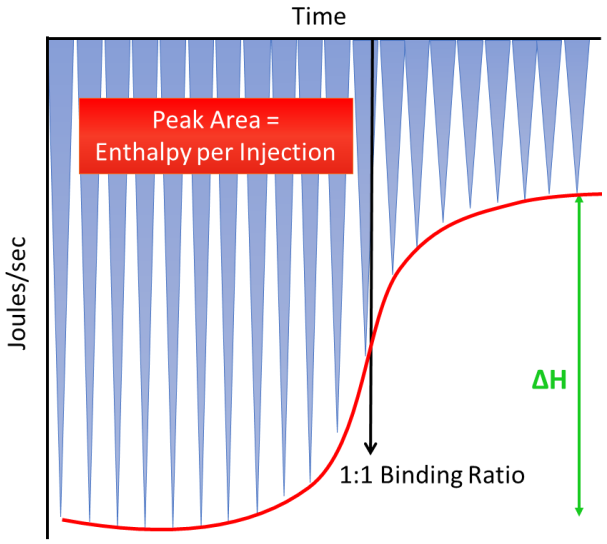


Figure 14: Sample ITC trace showing enthalpy of injection, enthalpy change (ΔH) and the binding ratio.

In contrast to DSC, thermal processes are measured at a constant temperature, with increases in the concentration of reactant, rather than across a temperature gradient. Experiments can be repeated at different set temperatures to determine the effects on binding. Parallel use of ITC with DSC presents the opportunity to study both the binding of proteins with various detergent additives and the effects of this binding on melting temperatures, providing more insight than simple T_{max} determination alone.

ITC also provides scope to probe contributing factors to ΔH_{app}. For example, determining changes in entropy using Gibbs free energy equations gives insight into the hydration shell of a protein, with high ΔS values attributed to water binding. Adjustments in buffer pH resulting in a change in the ΔH_{app} also indicates a protonation/deprotonation event is involved in the process.[74]

1.6.1 Advantages and Limitations

ITC, used in conjunction with DSC, provides the opportunity to focus on both the binding of enzymes with various laundry additives and to measure the effects of this binding on a stability scale via T_{max} values.[74]

The use of calorimetric methods removes the issue of opaque and UV-active components which interfere with optical analysis methods. As calorimetric methods relate to entire systems, the technique may not be capable of deconvoluting ligand binding of multi-component or fully formulated HDL's.

The technique is inexpensive due to lack of external elements and low sample consumption. Throughput is also low, however, as samples must be analysed individually with little scope for automation or scale up in efficiency.

1.7 Pulse Proteolysis

Low throughput and the need for sensitive 'nano' equipment limits accessibility of the above-mentioned calorimetric methods. Fast pulse proteolysis (FastPP), described by Park *et al.*, employs common, non-specialist lab equipment to determine protein stability under various conditions. The procedure combines proteolytic degradation with electrophoretic separation to identify destabilising conditions.[75,76] Due to the absence of optical detectors, interference from opaque and UV-active components should not affect analysis.

The assay employs a thermophillic protease, thermolysin, which preferentially cleaves unfolded protein. Proteolytic activity is focused on the hydrophobic residues, phenylalanine, leucine, isoleucine and valine which are generally internalized in native protein. Under destabilising conditions, these amino acids are exposed by protein unfolding, providing access to thermolysin active sites.

Fragmented protein is separated from intact, native protein using SDS PAGE *(Figure 15)*. The denaturant, SDS, removes any residual structure and coats proteins and peptide fragments with negatively charged surfactant monomers. Samples are then loaded onto an acrylamide gel to migrate towards a positive pole. The rate of migration is dictated by

molecule size. Smaller molecules, such as the protein fragments, will be free to move more quickly through the gel matrix and separate from larger, intact molecules. A molecular weight marker is also loaded onto the gel for reference. Denaturing conditions are identified from loss of band intensity at its molecular weight. T_m determination requires incubation of samples in the presence of thermolysin at a range of temperatures. Disappearance of the protein band then indicates temperatures at which protein is unfolded.

In order to prevent autoproteolysis interfering with the action of thermolysin, protease samples can be inhibited with PMSF, a common serine protease inhibitor, which acts by binding covalently the hydroxyl group of the serine protease active site. As thermolysin is a metallo-protease, it is unaffected by the inhibitor.[77]

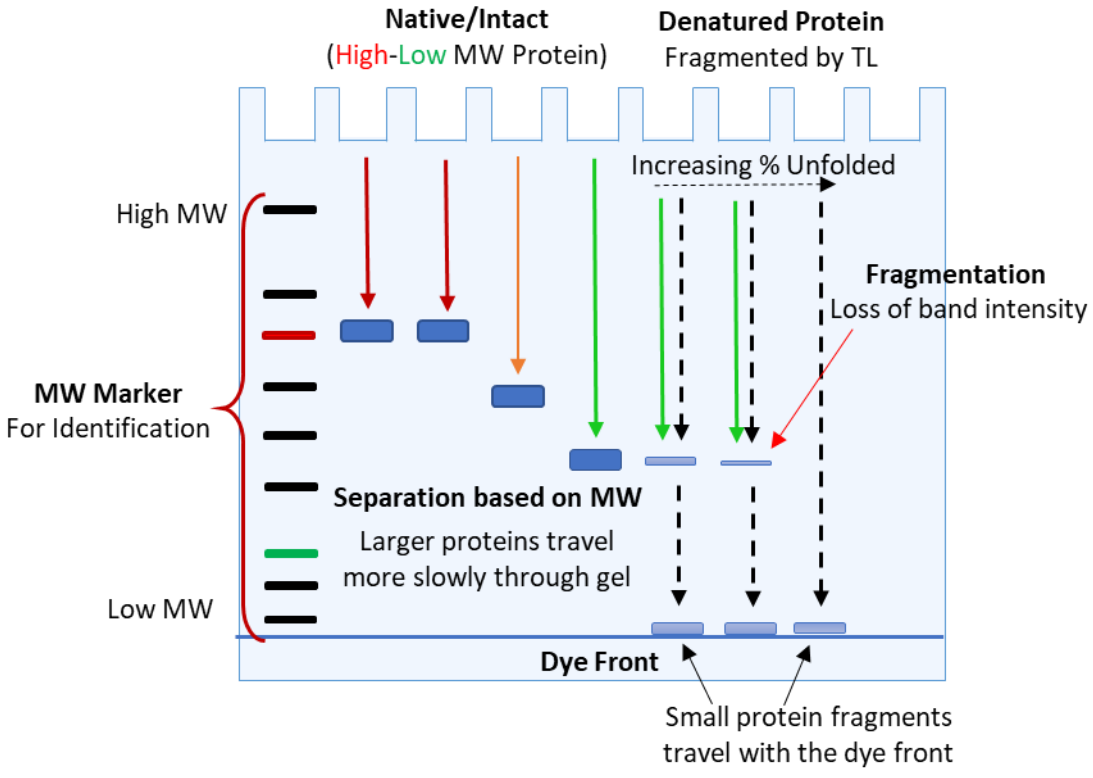


Figure 15: Schematic of SDS page analysis of Themolysin assay showing loss of protein band intensity with fragmentation.

1.7.1 Advantages and Limitations

The appeal of FastPP lies in its use of basic equipment, making it an ideal technique for the non-specialist lab. Replacing thermolysin with a detergent protease, also provides scope for exploring the effects of proteolysis in commercial formulations.

As analysis is reliant on the activity of a second enzyme, the thermolysin, extra care must be taken to ensure proteolytic rates remain constant. Conditions which influence folding and confirmation of detergent enzymes may also affect observed proteolysis rates by thermolysin. Reduced rates of proteolysis would result in an apparent increase in T_m values.

Aside from variation in thermolysin activity, multiple factors can result in variation in band intensities across a single gel. Differences are intensified when comparing results from several, independently cast and run gels. This can result in poorly defined T_m values and high margins of error which would present difficulties in routine analysis. Furthermore, sample preparation, incubation and gel loading are time consuming and labour-intensive which further hinders routine analysis.

As formulation screening requires analysis of vast numbers of samples, throughput and automation are key factors to be considered in choosing analytical techniques. The methods described above, while compatible with high detergent concentrations, are limited in this capacity. For this reason, focus was turned to optical methods which can provide several times the throughput offered by DSC. Namely, these were differential scanning fluorimetry (DSF), circular dichroism (CD) and microscale thermophoresis (MST).

1.8 Differential Scanning Fluorimetry

Differential scanning fluorimetry (DSF) was developed specifically as a high throughput method of probing changes in thermal stability of proteins as a result of small ligand binding. The technique relies on fluorescent dyes

which respond to changes in the sample environment as a result of protein unfolding.

A common DSF dye, SYPRO Orange, has been popularized due to its relatively high signal-to-noise ratio, exhibiting up to 5 times the background intensity in hen egg-white lysosome unfolding studies. Its high excitation energy of 490 nm is also removed from that of most molecules, reducing the risk of interference from absorption of sample components.[23] Fluorescence properties of the dye are suppressed in the aqueous bulk of the sample. On protein unfolding, however, interactions with exposed hydrophobic residues that were internalized in the native state, cause an increase in fluorescence. Signal intensity increases as further residues become exposed to the dye. A plot of fluorescence intensity as a function of temperature yields the T_m value *(Figure 16)*.

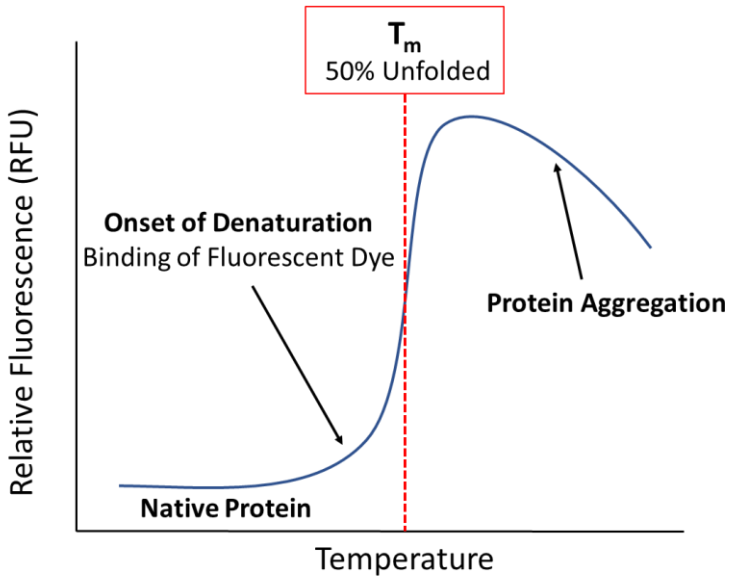


Figure 16: DSF sample trace of protein unfolding with temperature

Accurate analysis can be hindered by the requirement for these directly associating dyes, as the effects of protein binding can influence unfolding processes. Furthermore, interference has been observed in the case of surfactants, which interact with the dye through their non-polar groups. Molecular rotors have therefore been proposed as alternative sources of fluorescence to reduce these effects. CCVJ (9-(2-carboxy-2-

cyanovinyl)julolidine) is one such example, outlined by Hawe *et al.* for the detection of protein aggregation in surfactant-rich formulations.[78]

CCJV is part of a group of dyes which form twisted intramolecular charge-transfer complexes (TICTs). These consist of an electron acceptor group and an electron donor group, typically nitrogen incorporated into a pi-system, and a nitrile group respectively. The two groups are connected by a conjugated chain, which allows for rotation around single bonds when excited by light *(Figure 17)*. Rotation determines the fluorescent properties of the dye and, if hindered, changes will be detected in emission intensity.[79–81] Molecular rotors are therefore sensitive to sample viscosity, rather than polarity. Changes in viscosity are provided by protein aggregation on denaturation. In this case the output represents the temperature of aggregation (T_{AGG}) rather than the melting temperature (T_m). Aggregation tends to occur at higher temperatures and so T_{AGG} and T_m values may not be directly comparable. Furthermore, enzyme activity is generally lost long before the protein begins to aggregate, and it would not be possible to study denaturation which does not result in aggregation, as can be the case in surfactant-rich samples.

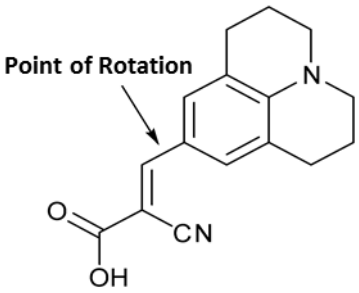


Figure 17: Structure of CCJV showing the point of rotation in the excited state.

1.8.1 Advantages and Limitations

The popularity of DSF in the study of protein stability stems from its high throughput capabilities and the rapid availability of results. Analysis is conducted in microplate format, enabling simultaneous screening of almost 400 samples.[82,83] Testing is usually conducted in triplicate within the same

plate, demonstrating reproducibility of results without the need for multiple runs. Sample loading is also low at ~20 µl of 0.1 mg/ml protein concentration per well. Using a qPCR machine, a temperature ramp of 20-100 °C generally takes 90 minutes. Data analysis can be automated using software such as NAMI[84] to generate T_m values. The most labour-intensive step involves the preparation and pipetting of the required samples. For routine industrial analysis, this can be reduced through automation or the use of multichannel pipettes.

The use of external dyes increases the cost of analysis. Furthermore, interactions between dyes and detergent components such as surfactants and interference with unfolding processes need to be considered. Molecular rotors provide an alternative to directly-associating dyes and can reduce this problem. The use of T_{AGG} in place of T_m, however, is dependent on protein aggregation, which can be limited in surfactant-rich formulations. The technique also provides little insight into unfolding mechanisms.

1.9 Circular Dichroism

Circular dichroism spectroscopy (CD) is commonly used in the determination of protein structure and stability. As well as determination of T_m values, the technique provides scope for the determination of secondary structural features. This presents the opportunity to probe unfolding mechanisms through observations of changes in α-helix and β-sheet content under various detergent conditions.

CD is based on the concept of elliptically polarised light. Plane polarised light consists of both right and left handed, circularly polarised components (R-CPL and L-CPL), which are absorbed to different extents by chiral molecules. The difference in the absorption of the two is measured in CD as millidegrees of circular dichroism (mdeg CD).[85]

A circularly polarised beam of light can be produced by passing monochromatic linearly polarised light through a specific polarising filter known as a quarter wave plate *(Figure 18)*. The filter will only allow light waves oscillating in two specific perpendicular planes to pass through. One of the waves will also be slowed by a quarter wavelength with respect to the other. This causes the electromagnetic field, which is associated with the two light waves, to rotate in a helix. The propagation of this wave can be calculated by taking the dot product of the two composite waves. The helix can be right or left handed (L-CPL or R-CPL) depending on which of the light waves has been delayed. Superimposing R-CPL and L-CPL on top of each other results in a cancellation of the rotation of the electromagnetic field, returning the light to its plane polarised form, as the field will revert to propagating as a sine wave.

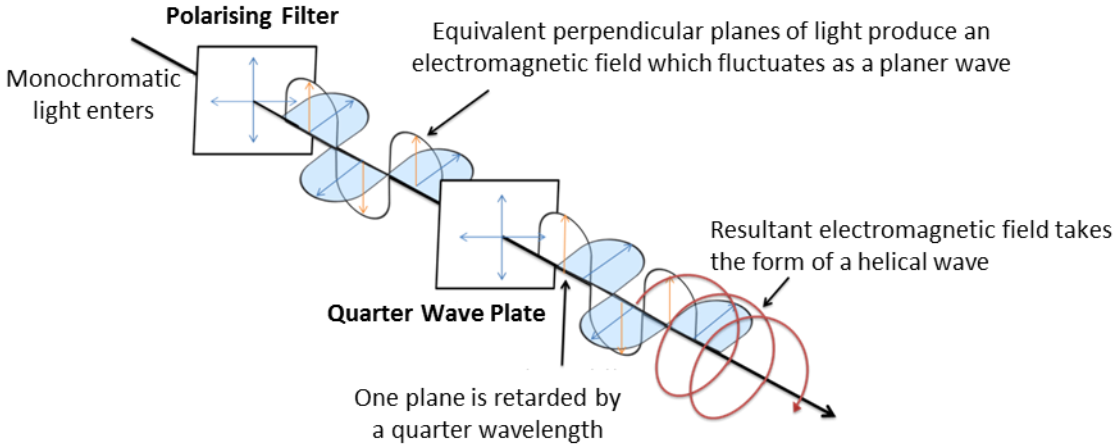


Figure 18: Plane polarisation of perpendicular waves of natural light after passing through a filter. Followed by circular polarisation with the use of a quarter wave plate.

Asymmetric compounds are known to rotate plane polarised light, and by measuring the degree of this rotation, we can calculate the optical purity of a compound. In a similar fashion, chiral molecules will absorb R-CPL and L-CPL to different extents. The level of absorption is measured by the wave's extinction co-efficient (ε). The component with the higher value for ε in a given protein will have a greater absorbance and so will continue to propagate with a weaker intensity than the other. This disrupts the plane oscillation of the electromagnetic field and when the dot product of R-CPL

and L-CPL is calculated, we find that the field oscillates in an elliptical wave, rather than the circular or planar fields described previously. CD spectroscopy measures the extent of ellipticity and records any changes.

Circular dichroism is compatible with protein molecules as the amino acid chain is chiral along its backbone at each α-carbon centre. Proteins also exhibit a native secondary structure, causing the chain to fold into periodic helices and sheets. These structures absorb as a single unit or exciton, which each give characteristic signals in CD spectra. Signals arising from secondary structural features usually appear below 230 nm, while tertiary structures can be probed by investigating the near UV region (250-300 nm). Within the far UV region, signals arising due to various specific elements of the secondary structure can be distinguished further. Both α-helices and β-sheets can be identified by the positive peaks below ~210 nm, which then become negative troughs with maximum intensity at 222-226 nm for helices and 215 nm for β-sheets. Variation in the shape of the two signals peak is evident due to the higher degree of structural freedom associated with β-sheets. This also accounts for the lower stability than associated with more rigid helices, as well as the lower intensity at both the positive and negative signal peaks *(Figure 19).*[86–88]

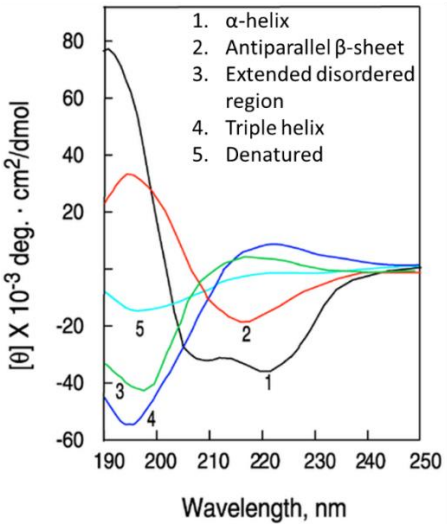


Figure 19: Spectra showing characteristic CD traces of various protein conformations. 1. α-helix, 2. Antiparallel β-sheet, 3. Extended disordered region, 4. Triple Helix, 5. Denatured protein. Reproduced with permission from Greenfield 2006.[89]

As a protein denatures asymmetric structure is lost, reducing the difference in absorbance of R-CPL and L-CPL. This manifests as reduced CD intensity, as illustrated in *Figure 20*. Protein stability can also be determined using the technique, by plotting CD intensity at a single wavelength as a function of temperature. The inflection point of the resultant sigmoidal curve denotes the T_m. Such analysis is generally conducted at 222 nm, where there is a strong signal for α-helix absorbance.

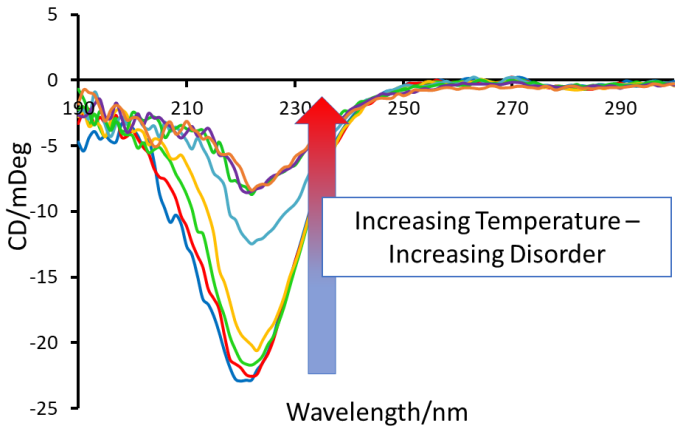


Figure 20: Sample trace of circular dichroism temperature/wavelength scans showing loss of regular asymmetric structure with increasing temperature. Each trace represents a spectrum taken at a given temperature point.

1.9.1 Advantages and Limitations

The benefit of CD lies in its use of intrinsic enzyme properties (asymmetry), making it independent of external components such as dyes, which increase the cost of analysis and can interfere with unfolding processes. Furthermore, CD provides a degree of mechanistic insight which was not possible using storage assays. Spectra can be used to estimate relative proportions of α-helix, β-sheets and random coils, improving understating of the unfolding processes behind the observed T_m values.[89,90] Comparison of enzyme profiles in the presence of various detergent components should highlight conformational changes and regions of instability in the protein. Analysis of both spectra and melting curves can also reveal if unfolding occurs as a single process or if intermediates are produced.

Sample consumption can be quite high, but this is dependent on cuvette volume and the pathlength required to refine analyte signals. Sample size can vary from 3 ml to less than 100 µl, at concentrations of less than 0.1 mg/ml, in otherwise optically silent samples. Fortunately, in the detergent industry, bulk manufacture of biological detergents lowers the cost of enzyme samples. Relative to more high-throughput methods such as DSF and MST, CD is quite time intensive as samples are analysed individually over approximately 90 minutes. This can be improved by incorporating a sample changer, however, and due to simple sample preparation and fully automated sample heating and analysis, labour is quite low.

1.10 Microscale Thermophoresis

The final optical technique which will be discussed, microscale thermophoresis (MST), is a relatively new method developed by Nanotemper Technologies, providing a relatively high-throughput means of determining protein-ligand binding constants, and subsequent effects on protein structure. Thermophoresis, the rate of molecular flow across a temperature gradient, is used to monitor small changes in the shape, size and hydration shell of molecules as a result of ligand binding, which indicate unfolding.

Analysis by MST is conducted by specifically heating a small area of a capillary tube using an infrared laser with a wavelength of 1480 nm. This creates a temperature gradient in the tube of between 2 and 6 K. The homogeneity of the solution will be interrupted as molecules will move either towards or away from the heat, along the temperature gradient *(Figure 21).*[91,92] Samples are detected by means of a fluorescent dye, bound to the analyte prior to sample loading. Alternatively, proteins with a high proportion of tryptophan residues can be detected by means of intrinsic fluorescence.

Interactions between an enzyme and other molecules in solution will alter its size, conformation, hydration shell or charge. Even small changes to a molecule will affect the progression of thermophoresis and can be detected by the instrument. High sensitivity enables detection of binding between molecules with large differences in molecular weights to be detected. This makes it ideal for protein analysis as difficulty often arises in detecting the effect of small ligands on large biomolecules.

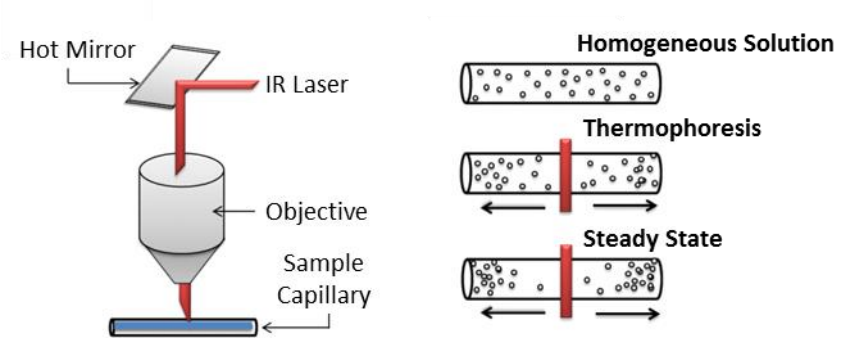


Figure 21: MST detects changes in the movement of molecules down a temperature gradient induced by an IR laser.

Progression of thermophoresis is recorded through fluorescence. One of the reactants, typically the enzyme, is labelled with a fluorescent dye appropriate to the available filters. When the IR laser is applied, changes in concentration of the fluorophore at the point of heating are monitored. A ligand-free sample is used to establish the base rate of thermophoresis and any changes in this rate can be attributed to binding interactions.

Disposable glass capillary tubes are used for sample analysis, resulting in extremely low sample consumption and reducing contamination between runs. Capillary tubes must have constant inner and outer dimensions as glass thickness dictates temperature gradients and overall sample heating. Thicker glass will have an associated UV absorbance which may interfere with results on such small scales. Further considerations to reduce noise include treatment of the inner surface of the tubes to ensure surface homogeneity and prevent samples adherence.

Up to 16 samples, including a ligand-free blank, are run in parallel. Before applying heat, initial fluorescence is recorded and should be constant

across all samples. Application of the laser causes a temperature spike and rapid increase in fluorescence known as a 'T-jump'. The jump lasts just several hundred milliseconds, allowing it to be easily distinguished from the thermophoresis signal. The change in protein concentration at the point of heating is subsequently monitored through fluorescence intensity to determine the thermophoretic rate. Comparison of rates between parallel samples gives the degree of structural change in a given environment. Titration of a ligand against a constant concentration of labelled protein enables calculation of binding constants, as described by Seidel *et al (Figure 22).*[93]

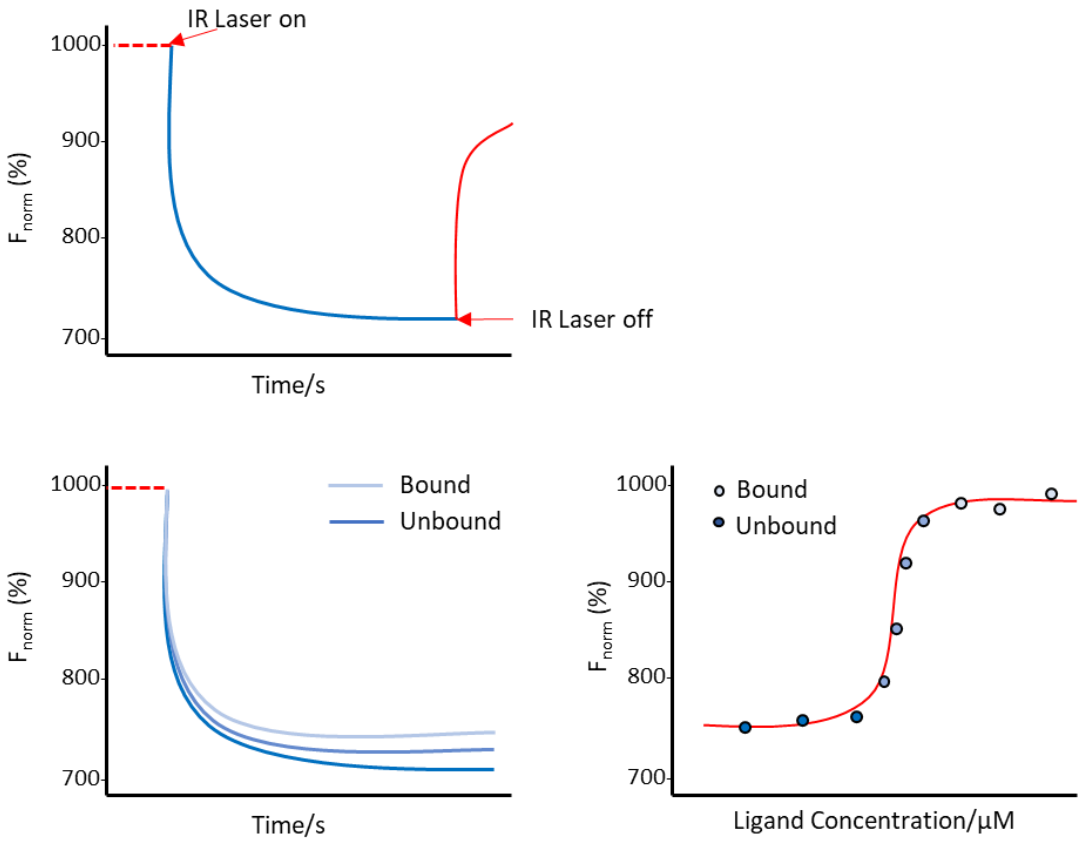


Figure 22: Sample traces of MST showing changes in fluorescence as a function of time, for bound (high concentrations) and unbound (low concentrations) ligands. A plot of fluorescence against ligand concentration gives a dose-response curve, from which binding affinity can be determined.

1.10.1 Advantages and Limitations

MST presents an opportunity to gain insight into protein-ligand binding interactions for various detergent components, highlighting key

components in enzyme inactivation processes. As analysis is conducted at a single temperature direct determination of T_m values is not possible. Parallel experiments using DSF, for example, should be run to quantify the effects of observed ligand binding. Outputs from these studies are analogous to experiments conducted using DSC and ITC based on calorimetric data but throughput is vastly improved.

1.11 Summary of Techniques

The techniques outlined above represent a range of both well-established and more recently developed methods of protein analysis. The optical-based methods, CD, DSF and MST tend to have higher throughput and offer more insight into unfolding mechanisms. Interference from opaque and UV active components of HDL formulations, however, can prevent accurate determination of enzyme properties. Alternative methods, DSC, ITC and FastPP, avoid the issue of detector saturation through the use of calorimetry or SDS PAGE. DSC and ITC show the most promise for analysis in the presence of HDL's, however, high viscosities and large numbers of freely interacting components may present issues in fully formulated samples.

The primary outputs of CD, DSF and DSC, are T_m values, based on various indicators of unfolding. Through very different approaches, each technique provides T_m values for ranking various detergent conditions for their effects on protein stability. In contrast, MST and ITC, present different approaches to determining ligand binding affinities. Thus, the use of multiple techniques in parallel provides much deeper insight into protein interactions in these complex formulations.

In the following chapters, use of these methods to develop understanding of the effects of various HDL components on stability will be discussed. *Chapter 3* presents preliminary data from screening of the capabilities of the above methods for enzyme analysis in laundry systems. Alongside

these data, several challenges which arose due to the complexity of detergent formulations will be highlighted. The presence of the surfactant LAS was particularly problematic for analysis. Methods devised to handle these issues are detailed in *Chapter 4*. Finally, *Chapter 5* will cover storage experiments conducted in parallel with thermal denaturation analysis. Rates of degradation will be compared to thermal data to construct simple models for the prediction of enzyme stability. Further work using the procedures established in this thesis to expand these datasets should provide scope to develop predictive models across a range of fully formulated laundry systems and detergent enzymes. These ideas will be summarised in the 'Conclusions' and 'Future Work' sections in the closing chapters of the thesis.

1.12 References

1. D. Otzen, *Biochim. Biophys. Acta*, 2011, **1814**, 562–591.
2. *Laundry Detergent Market Size, Share, Analysis: Industry Report, 2025*, Grand View Research, 2017, Retrieved from: https://www.grandviewresearch.com/industry-analysis/laundry-detergent-market
3. M. R. Stoner, D. A. Dale, P. J. Gualfetti, T. Becker, M. C. Manning, J. F. Carpenter and T. W. Randolph, *Enzyme Microb. Technol*, 2004, 114–125.
4. B. Wathen and Z. Jia, *Int. J. Mol. Sci.*, 2009, **10**, 1567–1589.
5. A. Fersht, *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*, W. H. Freeman, 1999.
6. L. Pauling, R. B. Corey and H. R. Branson, *PNAS*, 1951, **37**, 205–211.
7. W. Kabsch and C. Sander, *Biopolymers*, 1983, **22**, 2577–2637.
8. R. Weirenga, *FEBS Lett.*, 2001, **492**, 193–198.
9. E. G. Hutchinson and J. M. Thornton, *Protein Eng. Des. Sel.*, 1993, **6**, 233–245.
10. J. Zhang, Y. Zhang, W. Li, X. Li and X. Lian, *J. Surfactants Deterg*, 2014, **17**, 1059–1067.
11. R. Greenshields, *Res. and Appl. of Biotechnol.*, Springer, 1989.
12. C. Betzel, S. Klupsch, G. Papendorf, S. Hastrup, S. Branner and K. S. Wilson, *J. Mol. Biol.*, 1992, **223**, 427–445.

13. M. J. E. Sternberg, F. R. F. Hayes, A. J. Russell, P. G. Thomas and A. R. Fersht, *Nature*, 1987, **330**, 86–88.

14. G. Lange, C. Betzel, S. Branner and K. S. Wilson, *Eur. J. Biochem.*, 1994, **224**, 507–518.

15. M. L. Bender and F. J. Kezdy, *Annu. Rev. Biochem.*, 1965, **34**, 49–76.

16. A. V. Teplyakov, I. P. Kuranova, E. H. Harutyunyan, B. K. Vainshtein, C. Frömmel, W. E. Höhne and K. S. Wilson, *J. Mol. Biol.*, 1990, **214**, 261–279.

17. O. Almog, D. T. Gallagher, J. E. Ladner, S. Strausberg, P. Alexander, P. Bryan and G. L. Gilliland, *J. Biol. Chem.*, 2002, **277**, 27553–27558.

18. C. Andersen, Novozyme, *US Pat.*, 20170002340, 2017

19. Z. Li, X. Duan, S. Chen and J. Wu, *PLoS ONE*, , DOI:10.1371/journal.pone.0173187.

20. A. M. Brzozowski, D. M. Lawson, J. P. Turkenburg, H. Bisgaard-Frantzen, A. Svendsen, T. V. Borchert, Z. Dauter, K. S. Wilson and G. J. Davies, *Biochemistry*, 2000, **39**, 9099–9107.

21. C. Andersen, Novozyme, *US Pat.*, 7432099, 2008.

22. F. N. Niyonzima and S. S. More, *Appl. Biochem. Biotechnol.*, 2014, **174**, 1215–1232.

23. W. A. Offen, A. Viksoe-Nielsen, T. V. Borchert, K. S. Wilson and G. J. Davies, *Acta Crys. F*, 2015, **71**, 66–70.

24. M. S. Vijayabaskar and S. Vishveshwara, *PLOS Comput. Biol.*, 2012, **8**, DOI:10.1371/journal.pcbi.1002505.

25. E. Jurado, *Food Hydrocoll.*, 2010, **24**, 647–653.

26. C. Nielsen, PhD Thesis, Technical University of Denmark (DTU), 2012.

27. M. J. E. C. van der Maarel, B. van der Veen, J. C. M. Uitdehaag, H. Leemhuis and L. Dijkhuizen, *J. Biotechnol.*, 2002, **94**, 137–155.

28. E. G. Dorrit Aaslyng, *J. Chem. Technol. Biotechnol.*, 2007, **50**, 321–330.

29. J. H. van Ee and O. Misset, *Enzy. in Det.*, CRC Press, 1997.

30. K.-E. Jaeger, S. Ransac, B. W. Dijkstra, C. Colson, M. van Heuvel and O. Misset, *FEMS Microbiol. Rev.*, 1994, **15**, 29–63.

31. Y. Bourne, C. Martinez, B. Kerfelec, D. Lombardo, C. Chapus and C. Cambillau, *J. Mol. Biol.*, 1994, **238**, 709–732.

32. K. K. Kim, H. K. Song, D. H. Shin, K. Y. Hwang and S. W. Suh, *Struct. Lond. Engl. 1993*, 1997, **5**, 173–185.

33. F. K. Winkler, A. D'Arcy and W. Hunziker, *Nature*, 1990, **343**, 771–774.

34. J. Skjold-Jørgensen, J. Vind, O. V. Moroz, E. Blagova, V. K. Bhatia, A. Svendsen, K. S. Wilson and M. J. Bjerrum, *Biochim. Biophys. Acta*, 2017, **1865**, 20–27.

35. D. Aaslyng, E. Gormsen and H. Malmos, *J. Chem. Technol. Biotechnol.*, 1991, **50**, 321–330.

36. M. Chauhan, R. S. Chauhan and V. K. Garlapati, *Biomed. Res. Int.*, 2013, DOI: 10.1155/2013/374967

37. R. Lumry and H. Eyring, *J. Phys. Chem.*, 1954, **58**, 110–120.

38. A. Crutzen, M.L. Douglass, *Handbook of Detergents: Part A, Properties*, CRC Press, 1999.

39. A. Graciaa, P. Creux, J. Lachaise and J.L. Salager, *Ind. Eng. Chem. Res.*, 2000, **39**, 2677–2681.

40. M.-C. Michalski, F. Michel, D. Sainmont and V. Briard, *Colloids Surf. B*, 2002, **23**, 23–30.

41. K. Tsujii, N. Saito and T. Takeuchi, *J. Phys. Chem.*, 1980, **84**, 2287–2291.

42. C. O. Rossi, M. Macchione and L. Filippelli, *Mol. Cryst. Liq. Cryst.*, 2011, **549**, 160–165.

43. J. A. Stewart, A. Saiani, A. Bayly and G. J. T. Tiddy, *Colloids Surf. Physicochem. Eng. Asp.*, 2009, **338**, 155–161.

44. R. Schomaecker, B. H. Robinson and P. D. I. Fletcher, *J. Chem. Soc. Faraday Trans.*, 1988, **84**, 4203–4212.

45. M. N. Jones, *Surf. Act. Proteins*, 1996, 237–284.

46. L. Kravetz and K. Guin, *J. Am. Oil Chem. Soc.* 1985, **62**, 943–949.

47. D. E. Otzen, L. Christiansen and M. Schülein, *Protein Sci. Publ. Protein Soc.*, 1999, **8**, 1878–1887.

48. E. Kissa, *Pure Appl. Chem.*, 1981, **53**, 2255–2268.

49. H. Lund, S. Kaasgaard, P. Skagerlind, L. Jorgensen, C. Jørgensen and M. van de Weert, *J. Surfactants Deterg.*, 2012, **15**, 265–276.

50. W. Bode, E. Papamokos and D. Musil, *Eur. J. Biochem.*, 1987, **166**, 673–692.

51. M. Pantoliano, M. Whitlow, J. Wood, M. Rollence, B. Finzel, G. Gilliland, T. Poulos and P. Bryan, *Biochemistry*, 1988, **27**, 8311–8317.

52. P. A. Alexander, B. Ruan, S. L. Strausberg and P. N. Bryan, *Biochemistry*, 2001, **40**, 10640–10644.

53. J. Lalonde, E. J. Witte, M. L. O'Connell and L. Holliday, *J. Am. Oil Chem. Soc.*, 1995, **72**, 53–59.

54. M. R. Stoner, D. A. Dale, P. J. Gualfetti, T. Becker, M. C. Manning, J. F. Carpenter and T. W. Randolph, *Enzyme Microb. Technol.*, 2004, **34**, 114–125.

55. H. Lund, S. G. Kaasgaard, P. Skagerlind, L. Jorgensen, C. I. Jorgensen and M. van de Weert, *J. Surfactants Deterg.*, 2012, **15**, 9–21.

56. M. C. Crossin, *J. Am. Oil Chem. Soc.*, 1989, **66**, 1010–1013.

57. J. Boskamp, Lever Brothers Company, *US Pat.*, US4532064 A, 1985.

58. J. Lalonde, E. J. Witte, M. O'Connell and L. Holliday, *J. Am. Oil Chem. Soc.*, 1995, **72**, 53–59.

59. L. Burton, R. Gandhi, G. Duke and M. Paborji, *Pharm. Dev. Technol.*, 2007, **12**, 265–273.

60. D. S. Goldberg, S. M. Bishop, A. U. Shah and H. A. Sathish, *J. Pharm. Sci.*, 2011, **100**, 1306–1315.

61. T. Menzen and W. Friess, *J. Pharm. Sci.*, 2013, **102**, 415–428.

62. A. G. Quezada, A. J. Díaz-Salazar, N. Cabrera, R. Pérez-Montfort, Á. Piñeiro and M. Costas, *Structure*, 2017, **25**, 167–179.

63. R. T. Magari, K. P. Murphy and T. Fernandez, *J. Clin. Lab. Anal.*, 2002, **16**, 221–226.

64. T. B. L. Kirkwood, *J. Biol. Stand.*, 1984, **12**, 215–224.

65. F. Diagnostics, *Infinity™ Amylase Liquid Stable Reagent*, 2008, Retrieved from: https://assets.thermofisher.com/TFS-Assets/CDD/manuals/Infinity-Amylase-Reagent-EN.pdf

66. N. V. Prabhu and K. A. Sharp, *Annu. Rev. Phys. Chem.*, 2005, **56**, 521–548.

67. R. L. Danley and P. A. Caulfield, *Thermochim. Acta*, 2003, **395**, 201–205.

68. S. Tanaka, *Thermochim. Acta*, 1992, **210**, 67–76.

69. P. Gill, T. T. Moghadam and B. Ranjbar, *J. Biomol. Technol.*, 2010, **21**, 167–193.

70. N. Garbett, L. DeLeeuw and J. B. Chaires, *TA Instrum. – Appl. Note.* Retrieved from: http://www.tainstruments.com/wp-content/uploads/Nano_DSC.pdf

71. E. Freire, O. Mayorgo and M. Straume, *Anal. Chem.*, 1990, **62**, 950–959.

72. A. Velázquez-Campoy, H. Ohtaka, A. Nezami, S. Muzammil and E. Freire, in *Curr. Protocols in Cell Biol.*, John Wiley & Sons, Inc., 2001.

73. S. Leavitt and E. Freire, *Curr. Opin. Struct. Biol.*, January 9, **11**, 560–566.

74. I. Jelesarov and H. R. Bosshard, *J. Mol. Recognit.*, 1999, **12**, 3–18.

75. C. Park and S. Marqusee, *Nat. Meth.*, 2005, **2**, 207–212.

76. Y.-R. Na and C. Park, *Protein Sci.*, 2009, **18**, 268–276.

77. I. V. Demidyuk, A. E. Kalashnikov, T. Y. Gromova, E. V. Gasanov, D. R. Safina, M. V. Zabolotskaya, G. N. Rudenskaya and S. V. Kostrov, *Protein Expr. Purif.*, 2006, **47**, 551–561.

78. A. Hawe, V. Filipe and W. Jiskoot, *Pharml. Res.*, 2010, **27**, 314–326.

79. E. Ablinger, S. Leitgeb and A. Zimmer, *Int. J. Pharm.*, 1, **441**, 255–260.

80. S. Shi, A. Semple, J. Cheung and M. Shameem, *J. Pharm. Sci.*, 2013, **102**, 2471–83.

81. M. A. Haidekker, T. P. Brady, D. Lichlyter and E. A. Theodorakis, *Bioinorg. Chem.*, 2005, **33**, 415–425.

82. M. W. Pantoliano, E. C. Petrella, J. D. Kwasnoski, V. S. Lobanov, J. Myslik, E. Graf, T. Carver, E. Asel, B. A. Springer, P. Lane and F. R. Salemme, *J. Biomol. Screen.*, 2001, **6**, 429–440.

83. F. Vollrath, N. Hawkins, D. Porter, C. Holland and M. Boulet-Audet, *Sci. Rep.*, 2014, **4**, DOI: 10.1038/srep05625

84. M. K. Grøftehauge, N. R. Hajizadeh, M. J. Swann and E. Pohl, *Acta Crys. Sect. Biol. Crys.*, 2015, **71**, 36–44.

85. R. W. Woody, *Meth. in Enzymol.*, 1995, **246**, 34–71.

86. S. M. Kelly, T. J. Jess and N. C. Price, *Biochim. Biophys. Acta*, 2005, **1751**, 119–139.

87. S. M. Kelly and N. C. Price, *Biochim. Biophys. Acta*, 1997, **1338**, 161–185.

88. N. J. Greenfield, *Nat. Prot.*, 2006, **1**, 2527–2535.

89. N. J. Greenfield, *Nat. Prot.*, 2006, **1**, 2876–2890.

90. L. Whitmore and B. A. Wallace, *Biopol.*, 2008, **89**, 392–400.

91. M. Jerabek-Willemsen, T. André, R. Wanner, H. M. Roth, S. Duhr, P. Baaske and D. Breitsprecher, *J. Mol. Struct.*, 2014, **1077**, 101–113.

92. C. J. Wienken, P. Baaske, U. Rothbauer, D. Braun and S. Duhr, *Nat. Commun.*, 2010, **1**, 100.

93. S. A. Seidel, P. M. Dijkman, W. A. Lea, G. van den Bogaart, M. Jerabek-Willemsen, A. Lazic, J. S. Joseph, P. Srinivasan, P. Baaske, A. Simeonov, I. Katritch, F. A. Melo, J. E. Ladbury, G. Schreiber, A. Watts, D. Braun and S. Duhr, *Meth.*, 2013, **59**, 301–15.

2 Materials and Methods

Commercial detergent enzymes were provided by Proctor and Gamble and sourced from Novozyme. These included Natalase, Termamyl, Everest (amylases), Savinase, V42, FNA (proteases) and Lipex (Lipase). Laundry additives were also provided by Proctor and Gamble and were of commodity grade (surfactants: linear alkylbenzene sulphonate (LAS), Alkyl ethoxy sulphate (AE3S) and alcohol ethoxylate (AE7); chelating agents: HEDP, citric acid and fatty acid). SDS, EDTA, monoethanolamine and PMSF were all of analytical grade and purchased from Sigma Aldrich.

SYPRO™ Orange, Thermolysin and SDS PAGE buffer components were purchased from Sigma Aldrich. Coomasie Instant Blue and Precision Plus Prestained Protein Dual Colour standard were sourced from BioRad. PNA and EPS assay substrates were from Proctor and Gamble stocks, sourced from Thermofischer. Enzymes were dialysed using Slide-a-Lyzer™ cassettes from Thermofischer.

Fluorescence measurements and heating of DSF samples was conducted using an Applied Biosystems AB 7500 Real-Time qPCR, data analysis was completed using NAMI software *(Section 2.02.2)*. Circular dichroism was conducted using a Jasco 1500 with sample-heating for surfactant removal studies using a Thermocycler, also from Thermofischer *(Section 2.02.3.2)*. A Microcal VP-DSC was used for calorimetric measurements, samples were degassed prior to analysis with a Microcal Thermovac *(2.02.4)*. SDS-PAGE for FastPP was carried out with a BioRad MiniProtean tetra cell system *(Section 2.02.5)*. Activity assays were automated using Gallery™ Automated Photometric Analyzer by Thermo Fisher Scientific. Protein concentration was measured using a ThermoScientific NanoDrop™ while

surfactant concentration was determined by UV-Vis spectroscopy on a Cary 100 from Agilent.

2.1 Sample Preparation for Analysis of Thermal Denaturation

Liquid enzyme stocks provided by Proctor and Gamble were dialysed overnight into 0.1 M MEA at pH 8 using Slide-A-Lyzer™ dialysis cassettes. Protein concentration was determined by absorbance at 280 nm using a NanoDrop™. Stock solutions were then diluted to 2 mg/ml in 0.1 M MEA (pH 8) and stored below 5 °C.

For analysis, a range of surfactant and chelant solutions, listed in *Table 1* below, were prepared as representative mock detergent formulations. Concentrations are listed in % w/v and were freshly prepared from stocks prior to analysis. Stocks of 10, 15 and 40% w/v of each surfactant, prepared in 0.1 M MEA and adjusted to pH 8 with MEA, were stored at room temperature to facilitate dispensing of highly viscous media. All chelating agents were prepared from 5% w/v stocks in 0.1 M MEA and stored below 5 °C.

Enzymes were added to these solutions at concentrations of 0.1 mg/ml or 1 mg/ml (dependant on analysis and listed for individual methods below) and incubated in formulation at room temperature for 30 minutes prior to analysis. Protease samples were inhibited with PMSF prior to incubation to prevent autoproteolysis.

Table 1*:* List of Detergent Conditions Analysed

Anionic Surfactants		
0.1% LAS	0.1% SDS	0.1% AE3
1% LAS	1% SDS	1% AE3S
5% LAS	5% SDS	5% AE3S
10% LAS	10% SDS	10% AE3S
20% LAS	20% SDS	20% AE3S

Non-Ionic Surfactant	Chelating Agents	Multi-Component
0.1% AE7	2% EDTA	0.1% LAS, 2% EDTA
1% AE7	2% HEDP	10% LAS, 5% AE3S, 5% AE7
5% AE7	5% Citric Acid	Fairy™ (commercial non-bio)
10% AE7	5% Fatty Acid	Dash™ (commercial bio)
20% AE7		

2.2 Differential Scanning Fluorimetry

DSF analysis was conducted using a qPCR (Applied Biosystems 7500) with SYPRO™ Orange dye. Stock solutions of the dye were prepared to 4× concentration in MEA buffer. Analysis was run in 96-well plate format with 20 µl of sample (0.1mg/ml starting concentration of enzyme) and 5 µl of dye stock solution added to each well. Samples were run in triplicate.

The plate was heated on a gradient of 1 °C per minute between 20 and 100 °C with fluorescence readings taken every minute. Samples were excited at 470 nm and detection was at 510 nm. Signal intensity was plotted as a function of temperature and the data fitted to a sigmoidal curve. T_m was determined by the inflection point of the curve and was reported as the average value of triplicate results. Data fitting and T_m determination were completed using NAMI[1] software, developed by Groftehague *et al*. Error is listed as the standard error from the mean of the three analyses.

2.3 Circular Dichroism

Analysis was conducted on a Jasco 1500 CD spectrometer using a 0.01 cm pathlength cuvette. The short pathlength lowered the total absorbance of UV light by the surfactant LAS, enabling protein analysis in the presence of higher surfactant concentrations to be conducted. Enzyme samples and detergent additives were obtained from Proctor and Gamble. Enzymes

were dialysed into 0.1 M MEA at pH 8 and diluted to concentrations of 1 mg/ml. Detergent formulations were prepared as in *Section 2.1.*

2.3.1 LAS-Free samples:

Thermal denaturation analysis was conducted between 20 and 100 °C using the temperature ramp function of the instrument. Spectra were collected between 400 nm and 190 nm at 5 °C intervals. Samples were held at the appropriate temperature for 2 mins before scanning. Spectral collection was increased to 2 °C intervals across the expected protein melting temperature range to improve the accuracy of T_m values. Melting curves were fitted to a sigmoid with T_m values reported as the inflection point of the curve. Samples were run three times, with T_m listed as the average value of the triplicate scans. Data fitting and error analysis were completed using JMP software.

2.3.2 LAS-Rich Samples

LAS-based samples were purified prior to CD analysis to reduce UV absorption and stray light. Sample were prepared as described for LAS-free samples *(Section 2.1)* at LAS concentrations of 20%, 10%, 5%, 1%, 0.1% w/v. Enzymes stocks *(Section 2.1)* were added to each sample at 1 mg/ml.

Samples were heated prior to CD analysis using a LifeEco Thermocycler. 200 µl aliquots were removed at 5 °C intervals between 20 and 100 °C. Aliquots were cooled to room temperature and transferred by micropipette into 200 µl of 0.5 M $CaCl_2$ solution. Calcium stocks were prepared in identical MEA buffers to the dialysed proteins. The resultant precipitate of calcium and surfactant was removed by centrifugation at 2.0 rcf (relative centrifugal force) for 5 minutes. Faster centrifugation resulted in damage to the protein structure.

Individual spectra were collected for purified aliquots between 300 nm and 190 nm at room temperature using a 0.01 cm pathlength cuvette. CD intensity at 222 nm from each spectrum was recorded and plotted against

respective aliquot temperature to construct a melting curve. Analysis was conducted in triplicate with T_m values reported as the average of three inflection points. Data fitting and error analysis were conducted using JMP software.

2.3.3 Deconvolution of CD Spectra

Structural analysis of CD spectra was conducted using the Dichroweb website. Reference set 7 was applied to the data using both CONTIN and CDSSTR programmes. Secondary structural estimates for each sample are reported as the average of Dichroweb analysis spectra from three independent analyses. Error is listed as the standard error of the mean ratios reported for each structural feature.

2.3.4 Error Analysis

Analysis was repeated three times for each sample and stability was reported as the average of these three T_m values. Outlying data points and melting curves were omitted from fitting and calculations to prevent skewing of T_m.

Error in T_m is generally listed as the standard error from the mean of the three observed T_m values. In samples where only one T_m value was obtained, or where scatter from the constructed melting curve is higher than that of the standard error of the mean, error is reported as the standard error of the fit. Temperature control on the Jasco-1000 is correct to within 0.1 °C.

2.4 Differential Scanning Calorimetry

DSC analysis was conducted on an external contract basis by Dr Iain Manfield at the University of Leeds. Samples were prepared as described in *Section 2.1*. Reference buffers were identical to analyte samples, with the omission of enzyme.

The calorimeter was cleaned before each set of analyses by soaking the cells with Decon 90 (5% v/v) for 1 hour at 50 °C and then rinsed thoroughly with multiple changes of water. Protein samples and buffer (2 ml) were degassed using a Microcal Thermovac device for 10 minutes with the vacuum applied progressively to avoid samples bubbling too vigorously. The calorimeter (Microcal, VP-DSC) was set to heat from 10 °C to 90 – 120 °C, at 90 °C/hr, depending on where transitions were observed. An initial 2-3 scans were performed with water and buffer but not used because of "thermal history" effects arising from small differences between each cell. Cell contents were changed as the temperature cooled to 25 °C and then re-pressurised to ~29 psi, with a 15 minute equilibration step to 10 °C before the scan started.

Buffer-buffer scans were subtracted from protein-buffer scans as a baseline. Data was analysed using Origin 5.0. T_{max} values were reported as the temperature at the maximum peak height. Enthalpy was calculated by estimation of the area under the curve using the Trapezium method of approximation. As only single scan were conducted error has been listed as the average intrinsic error of the instrument (>0.1 °C)

2.5 Fast Pulse Proteolysis

Required formulations were prepared as described in *Section 2.1*. Enzymes were added to give 10 mg/ml concentrations from freshly dialysed stocks. Thermolysin was stored at 10 mg/ml in 20 µl aliquots of 50 mM Tris with 0.5 mM $CaCl_2$ (pH 8) at -20 °C. Enzyme concentration was determined by A280 using a ThermoScientific NanoDrop™. Samples were heated in 10 µl for 2 minutes at the required temperature. Centrifuge tubes containing 0.1 M MEA buffer alongside samples to ensure temperature equilibration.

While being held at the temperature point, the aliquots (10 µl) were diluted with 70 µl of heated MEA to reduce surfactant concentration. A further

aliquot (1 µl) of this dilute solution was then transferred by micropipette to thermolysin (12 µl, 0.025 mg/ml) solution, giving a final dilution of 1/96, for ease of pipetting. This allowed for suitable dilution with loading dye, while sufficiently lowering surfactant concentration (~1:100). The resultant sample was incubated for 30 seconds before quenching the reaction with 100 mM EDTA solution (2 µl).

Analysis was conducted on 15% acrylamide gels. Samples were incubated for 10 minutes at room temperature in loading dye (15 µl) before loading on the gel. Gels were run at 140 V and stained overnight in Coomasie Instant Blue. A Precision Plus Prestained Protein Dual Colour standard molecular weight ladder was used to identify intact and fragmented protein.

2.6 Activity Assays

Two sets of identical formulations were prepared as described in *Section 2.1*. Samples were spiked with either 4.5 mg/ml of V42 or 0.75 mg/ml of Everest. Enzyme concentrations were in line with commercial formulations for compatibility with existing assay procedures. Inhibitors were not included in the protease samples to prevent interference with the assay. All formulations, excluding 20% LAS, were prepared in bulk (100 g) with 1 g aliquots removed at each analysis point. Preliminary studies highlighted difficulties in homogeneous sampling of 20% LAS formulations due to high viscosities. Samples were instead stored in pre-prepared aliquots of 1 g, with appropriate concentrations of enzymes added individually to each aliquot. Samples were stored at 35 °C over a 6 month period.

Analysis was conducted using a Thermoscientific Gallery Automated Photometric Analyser™ with the synthetic substrate N-succinyl-Ala-Ala-Pro-Phe-p-nitroanilide (PNA) (protease) or Infinity™ Amylase Liquid

Stability Reagent kit (amylase). PNA was prepared as a stock solution of 10% PNA in DMSO, which was subsequently diluted 1 in 100 to a working solution in Tris Buffer, Infinity™ reagent was used neat. 1 g aliquots of each formulation were weighed into 100 ml volumetric flasks and diluted with 0.1 M sodium thiosulphate buffer with calcium. Formulations were analysed daily for the first 5 days and then weekly for a period 8 weeks. A final analysis was conducted after 6 months.

Activity was determined through the rate of appearance of substrate. These rates were plotted for each timepoint to determine initial rates enzyme degradation in each formulation. First order kinetics were used to calculated half-lives for the accelerated degradation. Curve fitting was completed using JMP software. The error is based on the standard deviation of the data from the fit.

3 The Challenge of Enzyme Analysis in Laundry Formulations

This chapter presents the results of preliminary work in screening the capabilities of popular analytical techniques for studying protein stability in high density liquid laundry formulations (HDL). The complexity of modern laundry detergents provides an opportunity to explore the scope of established methods in probing the structure and activity of proteins in industrially-relevant environments. Large numbers of freely interacting components, and high sample viscosity and opacity all present challenges to the detection and analysis of the small changes associated with protein inactivation.

ΔT_m will be used throughout the chapter as a measure of change in protein thermal stability. This parameter is defined as the temperature at which 50% of protein in a given sample is in an unfolded state. To reduce sample complexity in early screening stages, HDL formulations were separated into constituent parts. Those deemed to be most relevant to protein structure and stability were prioritised to determine key factors governing enzyme stability. These were categorised as anionic surfactant, non-ionic surfactant, builders and chelants.

Within this work, the capabilities of a range of common techniques in handling complex, multi-component samples were evaluated. These included differential scanning fluorimetry (DSF), circular dichroism (CD), differential scanning calorimetry (DSC), pulse proteolysis (FastPP) and microscale thermophoresis (MST). Each method presented a unique set of advantages and limitations, and outputs provided a different viewpoint on T_m. A summary of each technique will be given, followed by the results collected using each method.

DSF was selected for initial sample analysis as thermal shift assays operating at high throughput are already well-established. This enables the determination of T_m values for up to 96 samples in parallel. Expensive external dyes are required to produce fluorescence, however, which can influence both ligand binding and unfolding processes due to protein-dye interactions. SYPRO-orange was selected for use in this work as it is reported to be highly sensitive and minimise interactions with sample components[1,2]

T_m is determined by means of a fluorescent signal, produced in response to interactions between the dye and the increasingly hydrophobic sample media. Changes in the sample hydrophobicity are the result of non-polar residues, normally internalised in the core of native proteins, becoming exposed due to unfolding. This results in an increase in the fluorescence of the dye. A sigmoidal plot of fluorescence, as a function of temperature, yields the T_m value, at the inflection point of the curve.

A second optical method, CD, avoids the issues associated with the use of dyes as protein conformation is determined through the intrinsic property of regular structural asymmetry. Another key advantage is the level of information on secondary structural motifs that can be obtained. Deconvolution software can be applied to collected spectra to determine the ratios of α-helices, ß-sheets and random coils. This provides an added level of insight into the mechanisms of unfolding, as changes in specific types of structures can be identified. Melting temperatures are obtained by recording loss of signal intensity at 222 nm, which corresponds to the unfolding of helical regions in the protein.[3] The main disadvantage of this method is the throughput, as using a single cuvette system makes it far lower than the 96-well plate format of DSF. This can be improved somewhat, however, with the use of a sample changer.

DSC presents similar advantages to CD, in term of throughput and the avoidance of optical dyes. The method is more robust in the analysis of

highly coloured or opaque samples, however, due to the use of thermocouples which record changes in enthalpy, rather than optical detectors. DSC is limited by very low throughput, which cannot be improved with autosampling due to the viscosity of HDL solutions. Furthermore, the method lacks the structural insight provided by CD, and does not distinguish between signals associated with protein unfolding, and those of other thermal transitions in the bulk material, such as surfactant phase changes. For this reason, precisely matched reference samples are required to reduce noise due to secondary processes.

Similar opportunities and limitations are found in the case of FastPP, a relatively new method described by Park and Marqusee.[4] This technique is also unaffected by opaque or highly coloured samples, but lacks structural and mechanistic insight offered by CD. The method is based on the preferential cleavage of unfolded over native protein by a thermophilic protease, thermolysin. This provides a means of distinguishing denaturing conditions from those which have little or no effect on stability. Denaturing conditions are identified using SDS PAGE to separate protein which remains intact from fragmented samples. Inconsistencies can arise, however, if the activity of thermolysin is affected by sample components, reducing the rate of protein digestion. The method is also labour intensive, but can be conducted using simple, non-specialist lab equipment.

Finally, preliminary experiments were run using MST during a promotional demonstration by the company. The method provides very high throughput, similar to that of DSF, with up to 16 samples run in parallel. Sample consumption is also very low due to the use of capillary tubes. It is compatible with both the use of external dyes and with intrinsic fluorescence, produced primarily by tryptophan residues. The use of use of an optical detector, however, prevents the analysis of highly coloured and UV active samples which block protein unfolding signals. For this reason,

further analysis was not pursued, and the limited data obtained will not be discussed.

The results of thermal denaturation analysis, as determined by each method, are reported in the following sections. T_m values are not expected to be identical across the range of techniques, due to the different thermal processes the parameter represents in each case, however, stability rankings and trends should be reproducible between datasets. Observed limitations and obstacles to analysis will also be discussed and summary provided in the conclusions of the chapter. Each method was found to be suited to a specific type of sample analysis – high throughput, high surfactant concentrations, or structural insight. This highlights the need to expand the selection of analytical technologies available in the detergent industry, to maximise understanding and efficiency.

3.1 Analysis of Melting Temperature by Differential Scanning Fluorimetery

Preliminary DSF experiments focused on establishing baseline T_m values for a range of detergent enzymes under nil-detergent, control conditions. MEA at pH 8 was used to buffer all samples (preparation detailed in *Chapter 2.1*). Subsequent work built on formulation complexity by including LAS and EDTA as examples of surfactant and chelant respectively. These initial tests highlighted the limitations of DSF in analysis of surfactant-rich samples and so further components were not analysed using this method.

The complete dataset collected using DSF is presented in *Table 1.* Experimental T_m values are reported as an average of triplicate repeats with error listed as the standard error from the mean of the three analyses. A representative melting curve of the Natalase control, presented in *Figure 23,* illustrates the increase in fluorescence intensity in response to protein unfolding with temperature.

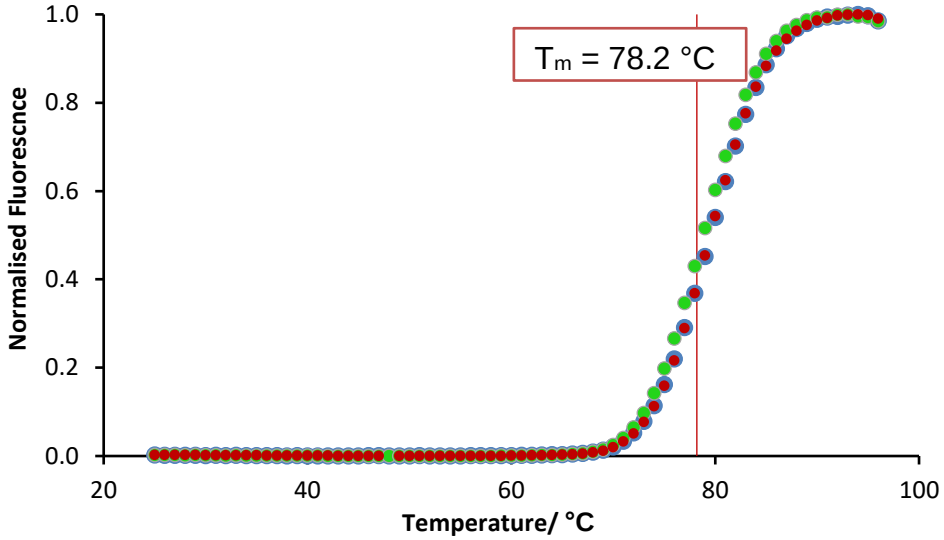


Figure 23: Representative DSF melting curve of Natalase in buffer only conditions. Experimental T_m value (78.2 °C) is an average of triplicate analyses.

Table 1: Experimental T_m values (°C) determined by DSF[a,b]

Enzyme	Class	Control	0.1% LAS	2% EDTA	LAS & EDTA
Termamyl	Amyl.	90.3 (± *0.2*)	89.7 (± *0.2*)	53.3 (± *0.0*)	51.8 (± *0.1*)
Everest	Amyl.	87.3 (± *0.2*)	89.0 (± *0.0*)	63.2 (± *0.2*)	63.4 (± *0.1*)
Natalase	Amy.	78.2 (± *0.1*)	81.7 (± *0.2*)	48.1 (± *0.2*)	45.7 (± *0.2*)
Lipex	Lip.	66.2 (± *0.2*)	54.1 (± *1.1*)	67.0 (± *0.0*)	56.1 (± *0.1*)
V42	Prot.	54.3 (± *0.1*)	53.3 (± *0.2*)	56.2 (± *0.1*)	54.0 (± *0.1*)
FNA	Prot.	61.3 (± *0.0*)	- [c]	43.7 (± *4.4*)	- [c]
Savinase	Prot.	71.6 (± *1.4*)	45.6 (± *0.1*)	40.5 (± *0.0*)	39.5 (± *0.2*)

[a]Reported T_m is an average of triplicate repeats [b]Values in brackets represent standard error of triplicate results. [c]Unclear transition, T_m value not obtained.

The most stable class of enzyme under control conditions was found to be the amylases, with an average T_m of 85.3 °C, indicating protein stability up to relatively high temperatures. The lipase, Lipex gave a T_m value of 66 °C, and proteases exhibited the lowest T_m values averaging 62.4 °C. As lipases are relatively new additions to commercial formulations, only a single example was sourced for analysis. Proteases exhibited the lowest T_m values of the group, with the exception of Savinase, which reported a T_m value higher than that of lipase. As shown in *Figure 24,* however, poor

signal to noise ratio was reported for the enzyme, resulting in a larger than average error and a poorly defined T_m.

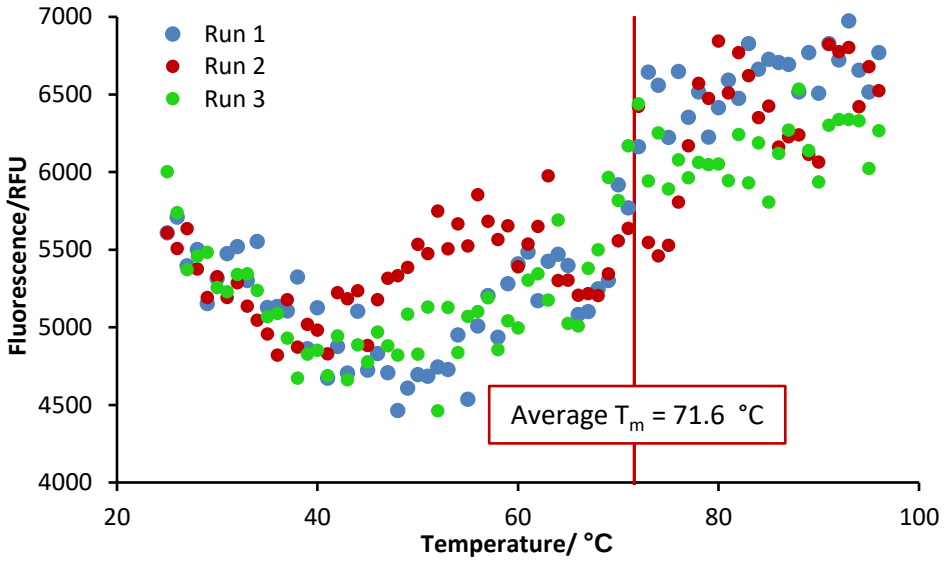


Figure 24: DSF melting curve of Savinase in buffer only conditions. Experimental T_m value (71.6 °C) is an average of triplicate analysis.

3.1.1 The Effects of Surfactant and Chelant on Enzyme Stability

The effects of surfactant and chelant on stability varied with enzyme class. Amylases experienced a significant loss in stability in the presence of chelant, with T_m reduced by 25-35 °C. Stability was maintained or even slightly improved, however, in the presence of LAS at 0.1% w/v, with an average increase in T_m of 2 °C. The opposite was true for Lipase, with surfactant reducing T_m by 11 °C and chelant having no significant effect.

The proteases were found to be significantly destabilised under both conditions, although no clear transition was detected in the presence of LAS for FNA. The more destabilising condition was not consistent across the class. V42 exhibited a greater loss in stability in the presence of LAS with T_m decreasing by 4.5 °C compared to 2 °C for EDTA. Savinase also showed a reduction in T_m of 4 °C for LAS, however a significantly larger loss in stability as observed for EDTA with a loss of almost 10 °C. Although no T_m data was collected for FNA in LAS, it showed the poorest resilience to EDTA of the group with a decrease in T_m of 17.5 °C.

The stabilisation of amylase by LAS was contrary to effects commonly reported for the surfactant. Consulting the literature, however, demonstrated several similar examples for other anionic surfactants. For example, SDS, which bears a close resemblance in structure to LAS, is a surfactant most commonly used as a denaturant in SDS PAGE. At low concentrations, however, It has been found to improve the thermal stability of BSA and even induce refolding of previously denatured protein.[5] This will be discussed in further detail in *Chapter 4.*

No synergistic effects were observed between chelant and surfactant. T_m values of multi-component formulations were consistent with the lower of the two single-component values. As a result, further experiments prioritised the collection data in single component systems, at commercially relevant concentrations. It is unlikely that multi-component effects will be absent for all combinations of excipients. Therefore, once robust analytical methods have been established, focus should shift to more complex formulations.

3.1.2 The Stabilisation of Protein Structure through Calcium Binding

The absence of chelant effects on Lipex was unsurprising as lipases do not require calcium for structural stability.[6] Both amylases and proteases incorporate calcium into their structures with loss of the metal ion inducing unfolding and significantly lower T_m values. From the data in *Table 1,* it is clear that proteases exhibit greater resilience to chelant effects than amylases. This is most likely a result of a second Ca^{2+} binding site found in subtilisin based proteases that is absent in α-amylases. Literature reports loss of calcium is only from the weaker site, with retention at the higher affinity site stabilising protein structure.[7,8]

As a result of the observations described above, mutations aimed at improving enzyme stability often focus on this weaker binding site.[8] This was evident from the improved resistance to EDTA induced denaturation

in the more newly developed V42 enzyme when compared to Savinase and FNA. A loss of 2 °C was observed in V42 compared to 10 °C and 17 °C in Savinase and FNA respectively.

3.1.3 Analysis of Enzyme Stability in the Presence of High LAS Concentrations

As illustrated in *Figure 25,* it was not possible to determine T_m values in samples of higher LAS concentrations (> 0.1%). Literature suggests that this is most likely a result of dye-surfactant interactions. Long hydrophobic chains bind the dye, producing high baseline fluorescence. This encapsulation of fluorophores has been reported across the range of hydropobe-binding dyes.[9]

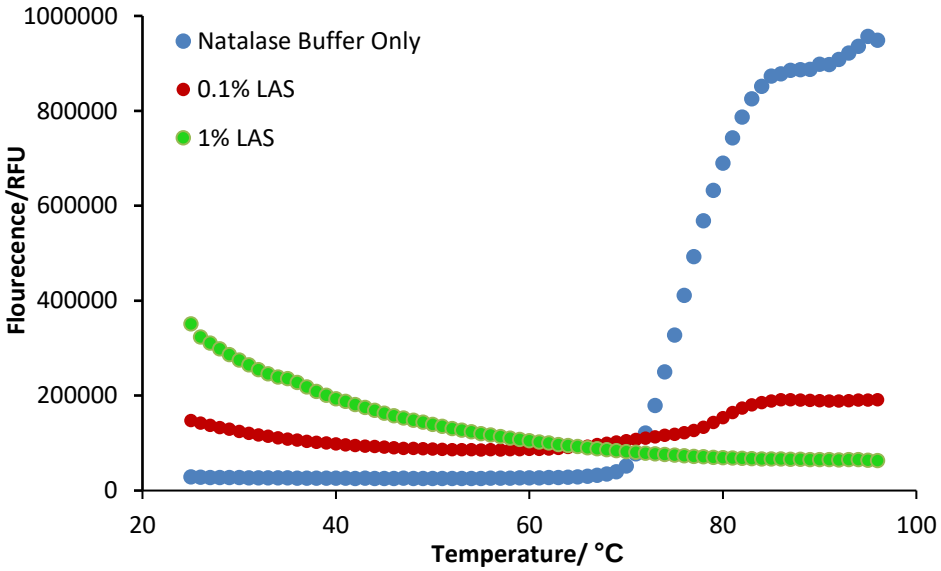


Figure 25: Melting curves of Natalase in the presence of LAS at 0.1% and 1% w/v, compared to a control.

In an attempt to produce results at more commercially relevant surfactant concentrations, the above analysis was repeated using a second optical method, CD spectroscopy. This technique is also well-established in the study of the thermal denaturation of proteins but does not require external dyes for analysis.

3.2 Analysis of Melting Temperature by Circular Dichroism

As with DSF, enzymes were first analysed under control conditions to establish baseline stability. Various detergent components (surfactant, chelants and builders) were then added to assess effects on thermal unfolding. $T_{m\ CD}$ values were determined from the inflection point of sigmoidal melting curves. As demonstrated for Natalase in *Figure 26*Figure 27, curves were generated by plotting CD intensity at 222 nm as a function of temperature. Measurements were repeated in triplicate, with average T_m values reported in *Table 2*. Error is reported as the standard error of the three results.

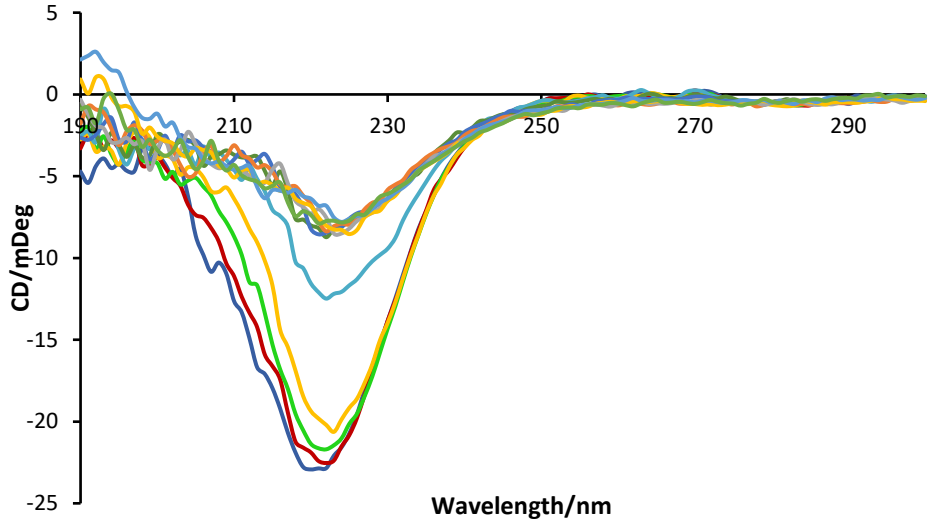


Figure 26: Denaturation analysis of the detergent amylase, Natalase by CD showing loss of enzyme structure with increasing temperature. Each scan represents an increase of 5 °C in temperature.

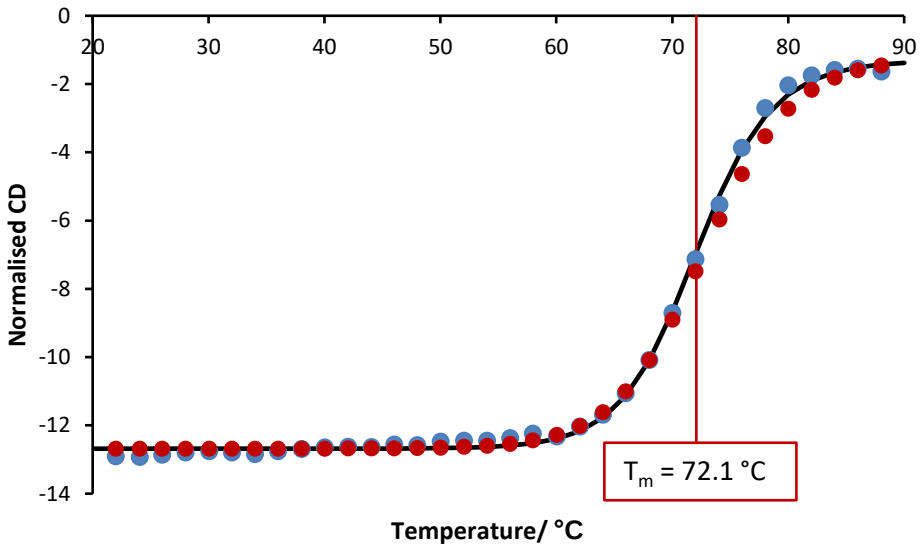


Figure 27: Normalised melting curves of Natalase determined by CD at 222 nm across two independent runs with an average T_m of 72.1 °C (± 0. 1 °C).

To inhibit protease activity and reduce the risk of autoproteolysis affecting T_m values, the protease inhibitor, phenylmethane sulfonyl fluoride (PMSF), was added to samples at 1 mM concentrations. This prevents proteolytic activity by forming an irreversible covalent bond between the sulphate group of the inhibitor and the hydroxyl group of the serine in the active site of the protease.[10]

It had been assumed that inhibition would increase apparent T_m values by preventing loss of enzyme structure through proteolysis, ensuring only thermal processes were in effect. As illustrated in *Figure 28*, however, T_m values were found to be lower in the presence of PMSF at 63.9 °C compared to 67.5 °C, suggesting that the enzyme-PMSF complex may be more susceptible to denaturation than the unbound enzyme. It has been reported in the literature that the rate of protease autolysis is far lower than that of thermal unfolding and so may be negligible with respect to the effects of temperature.[11] Regardless, analysis was conducted in the presence of PMSF to ensure consistency between measurements. This may be a result of destabilisation of secondary structural features at the binding site, however, further study into unfolding mechanisms is required to fully understand these observations.

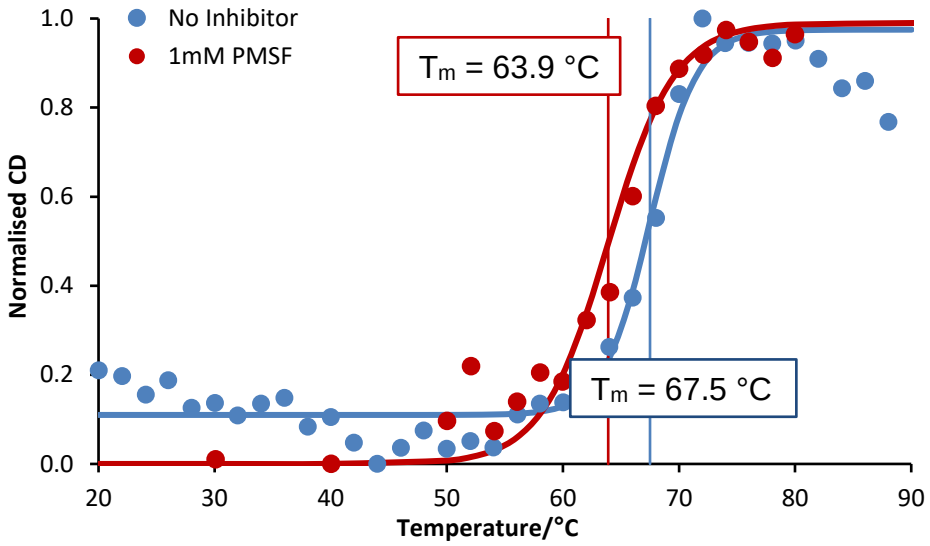


Figure 28: Melting curves of V42 control in the presence and absence of the protease inhibitor PMSF. Binding of the inhibitor was found to reduce the T_m of the protease by 3.5 °C.

Table 2: Experimental T_m values for a range of detergent enzymes as determined by CD[a]

Enzyme	Class	Buffer Only/ °C	Error in T_m[b]
Termamyl	Amylase	95.5	± *0.4*
Everest	Amylase	86.7	± *0.6*
Natalase	Amylase	72.1	± *0.1*
Lipex	Lipase	68.1 (90.2)[c]	± *0.4*
V42	Protease	63.9[d]	± *0.5*
Savinase	Protease	61.6[d]	± *0.4*

[a]Reported T_m values are an average of triplicate repeats. [b]Error values are listed in °C and represent the standard error from the mean of three independent analyses. [c]Two unfolding transitions were detected for Lipex. The T_m has been attributed to the first as the second is likely the result of protein precipitation. [d]Analysis conducted in the presence of PMSF.

As observed with DSF, the amylases, although covering a broad range of T_m values, were the most stable class of enzyme. The more recently developed amylases Everest and Termamyl, engineered for improved stability, gave T_m values 14 °C and ~ 23 °C higher than that of Natalase respectively. It was not possible to obtain a definitive value for the T_m of Termamyl as it had not completely unfolded at 100 °C, the upper temperature limit of the instrument. Unfortunately, it has been found that

in engineering stability at such extremes of temperature, the operational range of Termamyl has been raised beyond that of typical modern wash cycles, resulting in no activity towards stain removal.[12]

Two separate phases of unfolding were observed for Lipex, which had not been evident in DSF data (*Figure 29*). The first transition was used to establish the T_m value and occurred at 68 °C. The second transition was observed at 90 °C and has been attributed to protein precipitation. Finally, the proteases, again, exhibited the lowest stability values of the group with an average T_m of 62.5 °C.

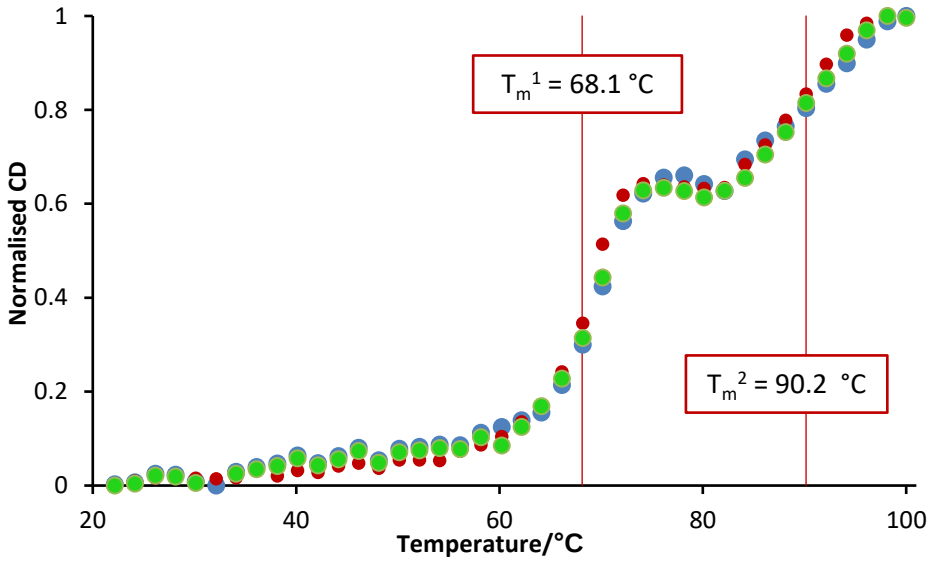


Figure 29: Melting curve of Lipex shows two transitions in unfolding, the first at 68 °C, followed by a second at 90 °C.

3.2.1 Comparison with DSF Data

Table 3 provides a comparison of T_m values determined via CD, which refer to loss of helical content and those from DSF which represent the exposure of hydrophobic residues from the core of an unfolding protein.

Both DSF and CD assigned the same stability ranking to the enzymes, with amylases observed to be the most stable class, followed by the lipase and lastly the proteases. Variation in T_m values reported by each technique was found to be surprisingly small, 2-5 °C for the majority of enzymes

(Table 3, Figure 30). Savinase and Termamyl exhibited the greatest differences, with ~10 °C and ~7 °C respectively. The high degree of scatter in the Savinase data may account for the unexpectedly high $T_{m\ DSF}$ value. Apparent T_m values were, in general, lower with DSF analysis than those determined by CD. This suggests that exposure of hydrophobic residues occurs at lower temperatures than unfolding of more structured helices. The lower T_m values also provide scope for analysis of more stable enzymes, as evidenced here with Termamyl.

Table 3: Comparison of experimental T_m values (°C) obtained through CD and DSF using SYPRO Orange dye.[a,b]

Enzyme	Class	$T_{m\ CD}$/°C	$T_{m\ DSF}$/°C
Termamyl	Amylase	95.0 (± *0.4*)	90.3 (± *0.2*)
Everest	Amylase	86.7 (± *0.6*)	87.3 (± *0.2*)
Natalase	Amylase	76.2 (± *0.1*)	78.2 (± *0.1*)
Lipex	Lipase	68.1/90.2 (± *0.4*)[c]	66.2 (± *0.2*)
V42[d]	Protease	63.2 (± *0.5*)	54.3 (± *0.1*)
Savinase[d]	Protease	62.0 (± *0.4*)	71.6 (± *1.4*)
FNA[d]	Protease	-[e]	61.3 (± *0.0*)

[a]Samples were prepared in MEA buffer at pH 8. [b]Values in brackets represent the standard error in the T_m across triplicate analyses. [c]Two transitions were observed in Lipex, T_m values reported represent the inflection points of each transition. [d]In both CD and DSF, protein.

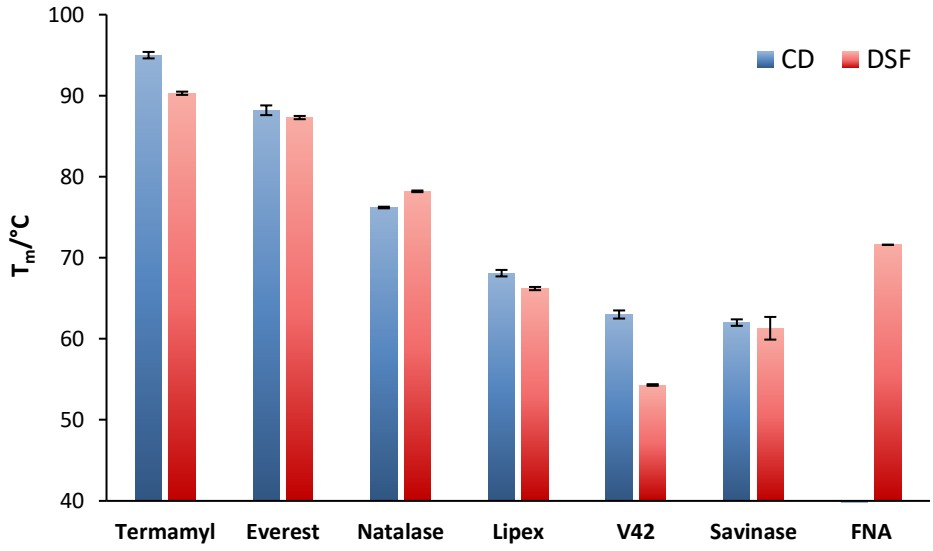


Figure 30: Comparison of experimental T_m values obtained by CD with those of DSF.

3.2.2 Building Formulation Complexity

Following successful determination of T_m values in buffered solutions, various concentrations of chelating agents and surfactants were added to study the effects of laundry additives on thermal stability. Chosen concentrations are in line with those of commercial formulations. The protease V42 was selected for in depth analysis as its lower initial T_m value provides scope to study increases, as well as decreases in stability. All samples were run in triplicate and the reported T_m represents the average of three analyses. Results have been summarised in *Table 4*. The average T_m for the V42 control sample was 63.9 °C. This will be used as a reference point when referring to the effects of various detergent components.

Table 4: Experimental T_m values (°C) determined by CD for the protease V42 in a range of laundry based formulations.[a,b]

Excipient	Additive Class	T_m / °C	Error in T_m / °C[c]
Control	No Additive	63.9	± *0.6*
AE3S 0.1%	Anionic Surfactant	66.6	± *0.9*
AE3S 1%	Anionic Surfactant	61.1	± *0.9*
AE3S 5%	Anionic Surfactant	65.5	± *0.2*
AE3S 10%	Anionic Surfactant	60.3	± *0.1*
AE7 0.1%	Non-Ionic Surfactant	64.8	± *0.1*
AE7 1%	Non-Ionic Surfactant	59.1	± *0.5*
AE7 5%	Non-Ionic Surfactant	68.5	± *0.3*
AE7 10%	Non-Ionic Surfactant	60.4	± *0.3*
EDTA 2%	Chelant	54.8	± *0.4*
HEDP 2%	Chelant	58.5	± *0.6*
Citric Acid 5%	Builder	58.1	± *0.3*
Fatty Acid 5%	Builder	59.8	± *0.1*
Mixed Surfactant[d]	Multi-Component	-	-
Fully Formulated[d]	Multi-Component	-	-
Fairy[d]	Multi-Component	-	-

[a]All analyses were conducted in the presence of PMSF. [b]T_m is recorded as the average of triplicate results. [c]Error represents the standard error of the T_m over triplicate results [d]No T_m values were obtained due to interference of laundry additives.

3.2.3 V42 and Secondary Surfactants – AE3S and AE7

No clear trend between T_m values and surfactant concentration could be determined for the anionic surfactant AE3S. Increases in T_m value were observed for V42 in the presence of both very low concentrations of AE3S (0.1%) and at 5% w/v of ~2 °C Samples at 1% and 10%, however, reduced protein stability by ~4 °C with respect to the control.

As illustrated in *Figure 31*, similar effects were observed for equivalent concentrations of the non-ionic, AE7 samples. A small stabilisation of V42 at 0.1% w/v, and a stabilisation of ~4.5 °C at 5%. Destabilisation of 4-5 °C was observed for both 1% and 10% samples. Further study into the aggregation states of these surfactants over concentration gradients, as well as protein-surfactant interactions may provide insight into the mechanisms behind these trends. AE3S and AE7 are generally present at 5% w/v in modern detergents, which was found to produce the greatest improvement in stability. The significance of these results, in relation to resultant enzyme storage stability, will be discussed in *Chapter 5.5*.

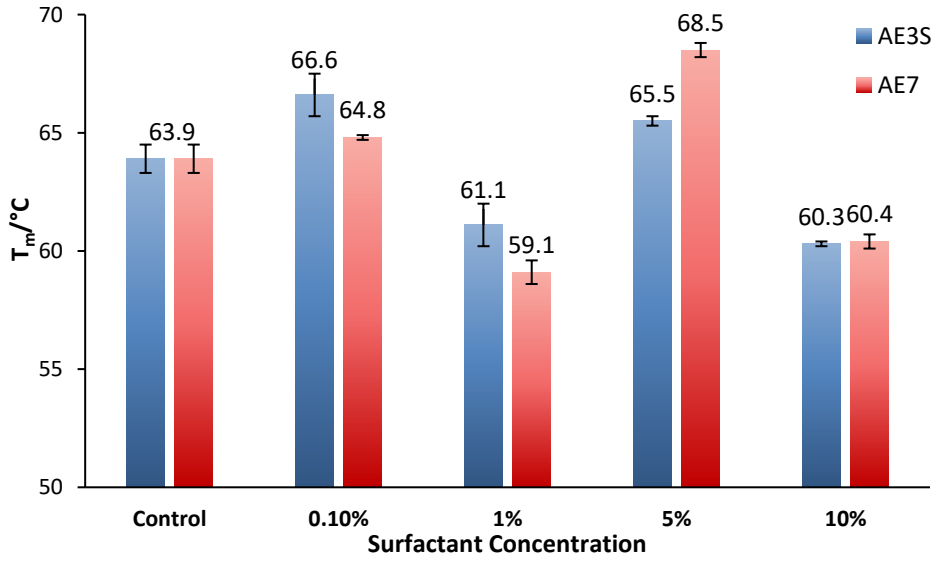


Figure 31: Chart comparing T_m values of V42 for various concentrations of surfactants, AE3S and AE7, as determined by CD.

3.2.4 V42 and Chelating Agents

Destabilising effects were observed in the presence of all chelating agent with a reduction in T_m values by 4-9 °C (*Figure 32*). The most prominent effects were induced by EDTA. This is contrary to indications from DSF analysis which found that V42 stability was maintained under these conditions. The concentration of EDTA was, however, increased from 5 mM in DSF to commercially relevant levels of 2% w/v (70 mM) for CD analysis. This evidently increased the ratio of free chelant, resulting in further sequestration of Ca^{2+} from protein binding sites, causing the observed increase in destabilisation.

Greater loss in stability was observed with EDTA than the other excipients tested. This is in line with higher dissociation constants reported in the literature for the chelant. 'Builders', which exhibit a weaker association with calcium, establish an equilibrium with protein bound calcium, thereby reducing the degree of destabilisation.[13–15] This will be discussed in further detail alongside equivalent Everest samples in *Section 3.2.7*.

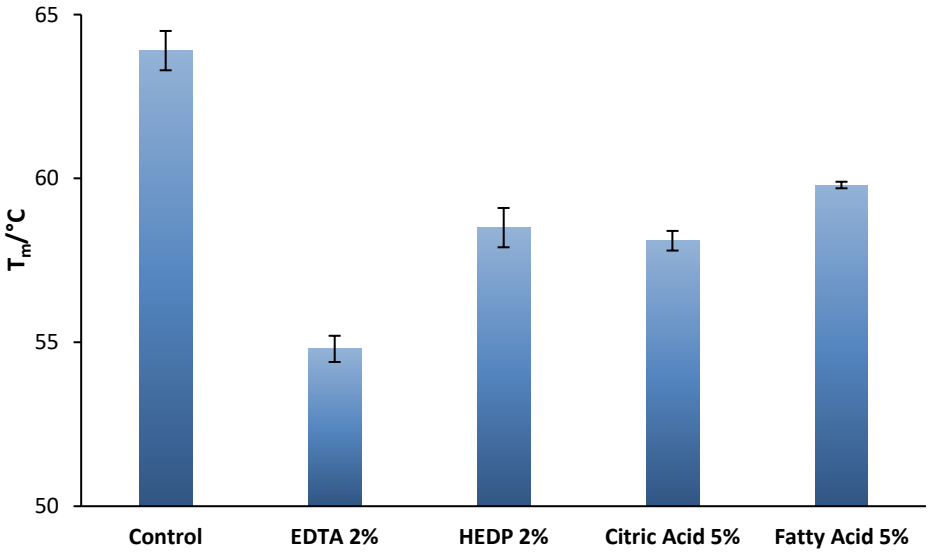


Figure 32: Chart comparing T_m values of V42 in the presence of various chelating agents (EDTA, HEDP) and builders (citric acid, fatty acid), as determined by CD.

Analysis was also attempted in the presence of LAS, however, detector saturation was evident above 0.1% surfactant. Further work to determine the extent of detector interference at lower concentrations will be discussed

in *Chapter 4.2.5*. As LAS accounts for such a large proportion of detergent formulations, analysis of multi-component systems was rendered redundant until the surfactant could be incorporated into systems. Further CD analysis instead focused on replicating the above work with a representative amylase, Everest, to determine if excipient effects were consistent across enzyme classes.

3.2.5 Effects of Various Detergent Components on the Stability of Amylases

Everest was selected an amylase of mid-range stability for CD analysis. Observed T_m values for equivalent conditions to those above, are listed in *Table 5*. The high baseline T_m of Everest (86 °C) compared to that of V42 presented difficulties in accurate determination of stability values. Under several stabilising conditions, sufficient data could not be collected to generate a complete melting curve, as they were beyond the heating capabilities of the instrument. Available data was fitted to a sigmoid to produce a T_m value *(Figure 33)*, however, errors in T_m values for Everest are higher than seen previously as a result.

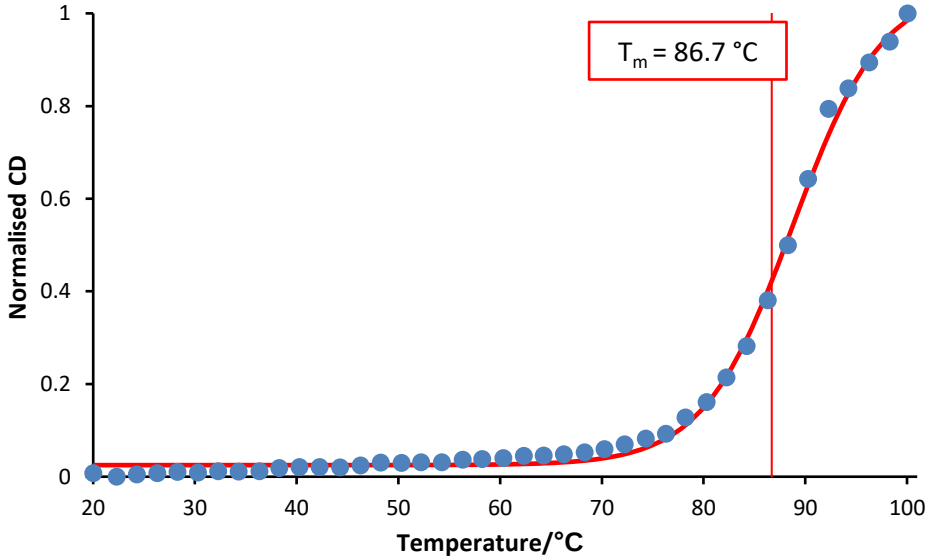


Figure 33: Melting curve of Everest under control conditions. The sigmoidal melting curve could not be completed as the stability of the enzyme was greater than the heating capabilities of the instrument. T_m was estimated by fitting to a sigmoid curve and was recorded as 86.7 °C.

Table 5: Experimental T_m values (°C) determined by CD for the amylase Everest in a range of laundry based formulations.[a]

Excipient	Excipient Class	T_m / °C	Error in T_m / °C[b]
Control	Nil Detergent	86.7	± *0.6*
AE3S 0.1%	Anionic Surfactant	94.1	± *0.1*
AE3S 1%	Anionic Surfactant	95.4	± *1.2*
AE3S 5%	Anionic Surfactant	95.6	± *0.5*
AE3S 10%	Anionic Surfactant	83.9	± *1.1*
AE7 0.1%	Non-Ionic Surfactant	94.1	± *1.9*
AE7 1%	Non-Ionic Surfactant	87.1	± *0.5*
AE7 5%	Non-Ionic Surfactant	90.0	± *0.4*
AE7 10%	Non-Ionic Surfactant	88.2	± *1.9*
EDTA 2%	Chelant	70.2	± *1.1*
HEDP 2%	Chelant	88.3	± *2.2*
Citric Acid 5%	Builder	83.1	± *1.7*
Fatty Acid 5%	Builder	83.0	± *3.0*

[a]T_m is recorded as the average of triplicate results. [b]Error represents the standard error of the T_m over triplicate results [c]No T_m values were obtained due to interference of laundry additives.

3.2.6 Everest and Secondary Surfactants AE3S and AE7

Up to 5% w/v, a stabilising effect was observed for AE3S, with increases in T_m of ~8 °C in each case. This effect dropped off at higher concentrations, however, as the 10% solution produced a 3 °C reduction in T_m compared to the control. In the non-ionic surfactant AE7, a stabilisation was also observed at 0.1% w/v from 86.7 °C to 94.1 °C. Above this concentration, T_m values of 87-90 °C were reported. These are within error of the control, suggesting that above 0.1% AE7, there is little effect on stability *(Figure 34)*.

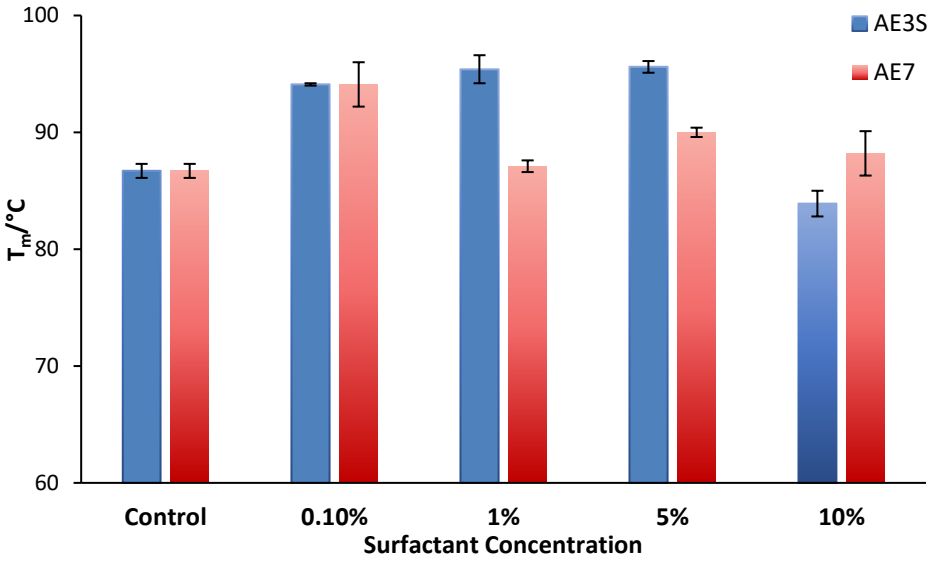


Figure 34: Chart comparing T_m values of Everest for various concentrations of surfactants, AE3S and AE7, as determined by CD.

3.2.7 Everest and Chelating Agents

Surprisingly, Everest samples exhibited greater resilience to chelating agents than V42. As shown in *Figure 35*, T_m values were within error limits of the control for HEDP, citric acid and fatty acid, suggesting low sequestration of amylase-bound calcium under these conditions. EDTA, however, induced a significant decrease in T_m values of almost 20 °C. This is in line with the chelants high association constant of ~$4.4 \times 10^{7,}$ several orders of magnitude higher than that of the other excipients.

No definitive relationship between T_m values and K_a was evident however, as illustrated in *Figure 36*. It is difficult to determine, without definitive metal ion binding constants for the detergent enzymes, how chelants values for *K*a will affect enzyme stability. Furthermore, literature values quoted here may not reflect speciation under detergent conditions. Future work may focus on developing understanding of these phenomena, but in this thesis, only empirical observations will be considered.

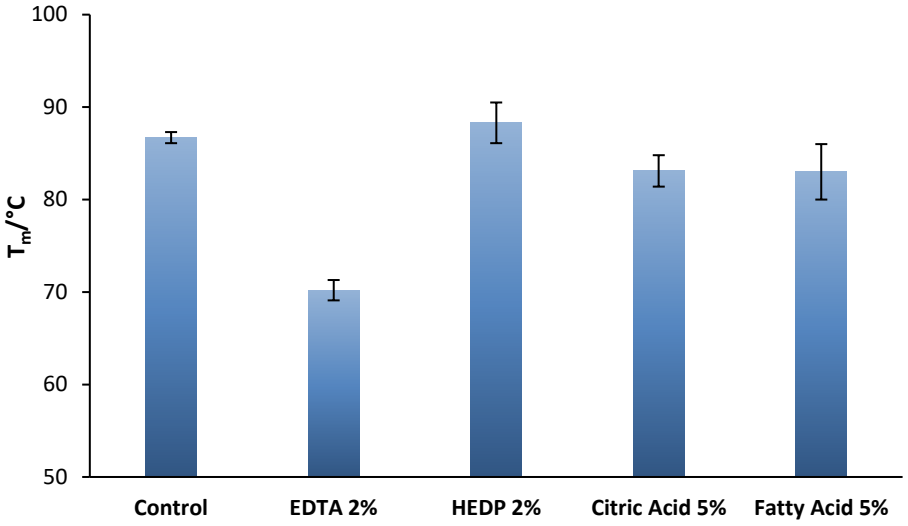


Figure 35: Chart comparing T_m values of Everest in the presence of various chelating agents (EDTA, HEDP) and builders (citric acid, fatty acid), as determined by CD.

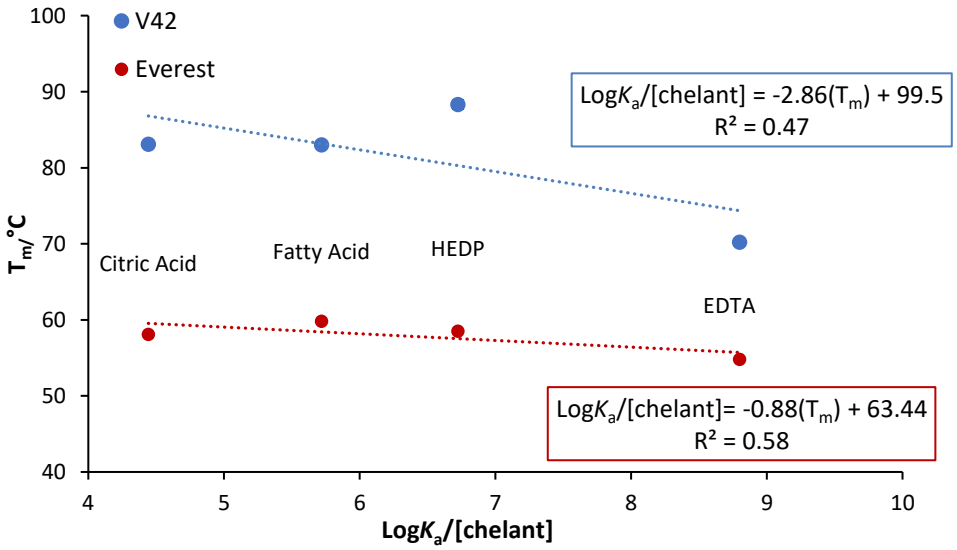


Figure 36: Relationship between chelant K_a values for calcium and observed T_m values for each enzyme.

3.2.8 Comparison of CD and DSF Analysis in the Presence of EDTA

As discussed for control samples, T_m values obtained by DSF and CD, although similar in several cases, were non-identical due to the different measures of stability employed. As detailed in *Figure 37,* variation between the two methods increased on addition of chelants to the analysis, however, a plot of T_m values obtained through DSF ($T_{m\ DSF}$) as a function of those obtained by CD ($T_{m\ CD}$), shows that this deviation is insignificant with

respect to empirical fitting of control samples. This indicates that analysis methods can be used interchangeably to determine protein stability for various conditions. This will be explored further in relation to storage stability in *Chapter 5*.

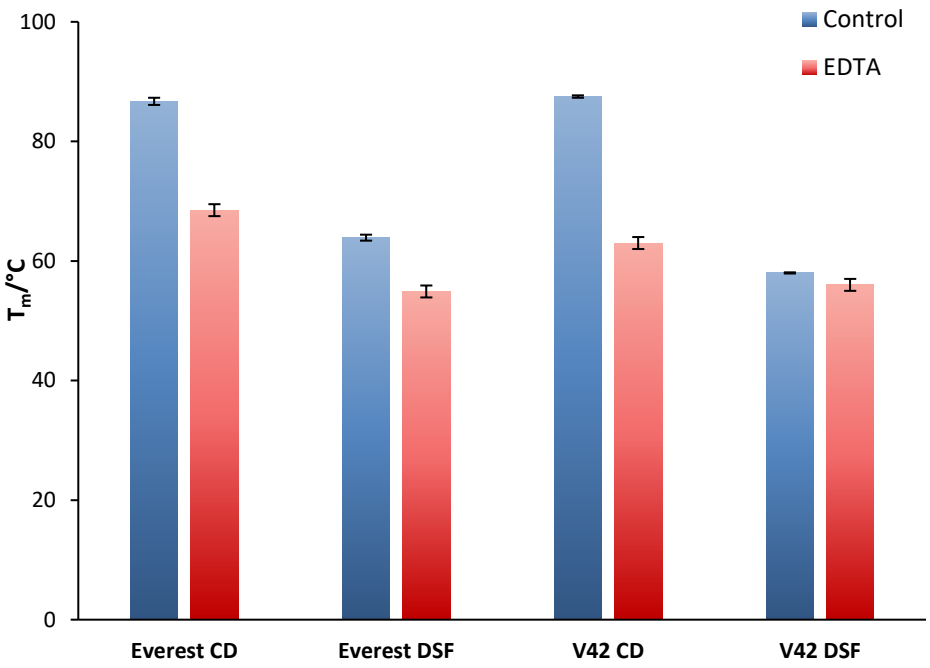
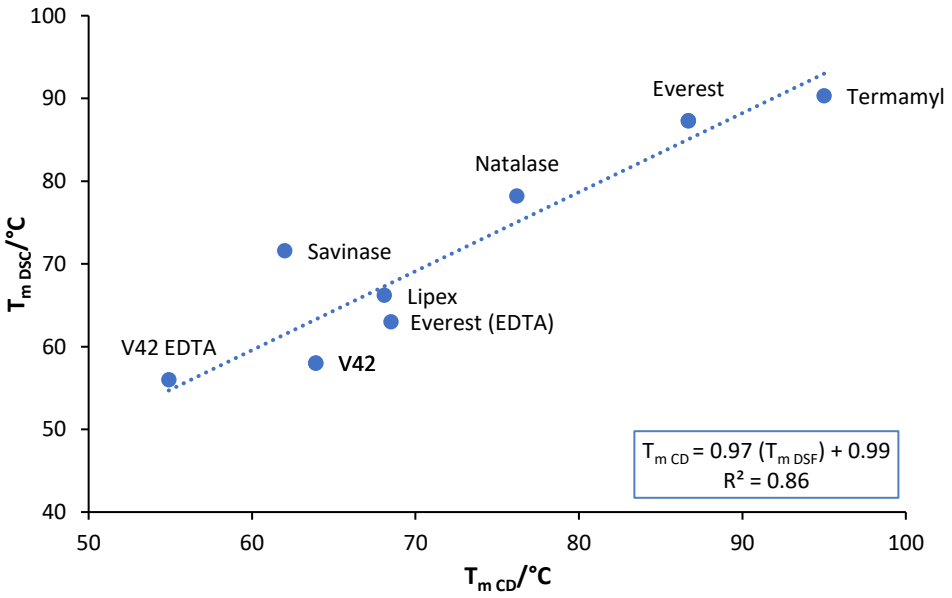


Figure 37: Comparison of T_m values obtained by CD and DSF for Everest and V42 in the presence of EDTA.



3.2.9 LAS Rich Samples

In the case of both enzymes, it was apparent that the CD instrument was not capable of detecting protein unfolding signals in LAS rich samples. The structure of this surfactant includes an aromatic ring and the associated conjugation leads to high UV activity, problem intensified by the high concentrations. Detector saturation was observed above 0.1% w/v, confirmed by UV spectroscopy *(Figure 38-Figure 39)*. Despite data towards mechanism collected at lower concentrations, it was not possible to analyse commercially relevant samples. For this reason, it was decided to explore the scope of non-optical methods, namely calorimetry and pulse proteolysis, for further analysis of these systems.

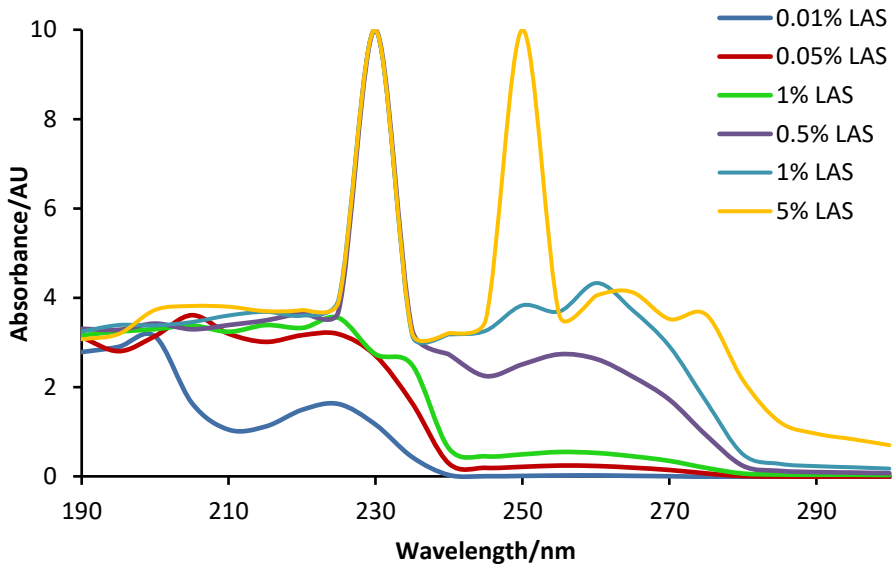


Figure 38: UV spectra of increasing concentrations of LAS. Detector saturation is apparent above 0.1% in the region commonly used in enzyme analysis (222 nm).

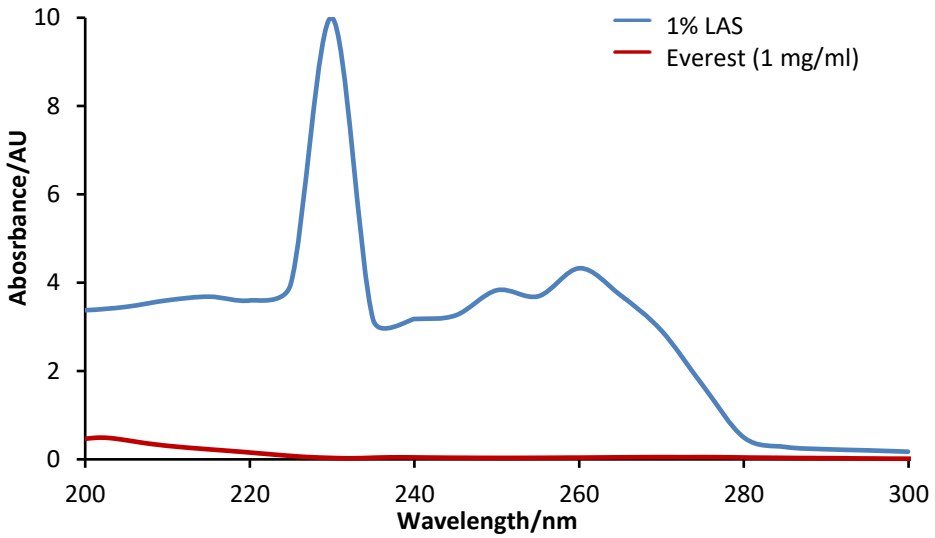


Figure 39: Absorbance spectra of Everest at 1mg/ml and LAS at 1% w/v.

3.3 Non-optical methods of T_m Analysis

Several well-established options for non-optical methods of enzyme analysis are available. Nano-DSC was selected for screening as it had been reported previously to be successful in detergent analysis.[11] Although capable of handling surfactant concentrations far higher than optical based methods, development of this technique was limited by the cost of running samples externally on a contract basis as the instrument was not available onsite in Durham.

Pulse proteolysis is a relatively new technique, first described by Park *et al*, as a method of indicating the effects of ligand binding on stability at a given temperature.[4,16] This was selected as a novel method using only common, non-specialist lab equipment. Detergent conditions being analysed, were also found to influence the thermophilic protease required for the assay, however, preventing analysis from being conducted reproducibly.

3.3.1 Measurements of T_m using DSC

Lund *et al*, have previously demonstrated that T_{max} values could be obtained for detergent enzymes in simplified HDL systems. These values also correlated linearly with observed rates of degradation in equivalent formulations, providing an efficient means of predicting storage stability. Replication of this work was initially attempted using traditional DSC, but the small energy changes associated with enzyme unfolding were indistinguishable from transitions in the buffer at this level of sensitivity. Instead samples were sent for analysis using nano-DSC at the University of Leeds. The complete report can be found in *Appendix 1*. *Figure 40* shows a representative scan of the unfolding transitions of Natalase under various conditions of LAS.

Table 6: Experimental T_{max} values (°C) as determined using nano-DSC.

Enzyme	Class	No LAS	0.1% LAS	1% LAS	5% LAS	10% LAS
Everest	Amylase	95	96.5	94.6	90.2	88.1
Natalase	Amylase	82	85.6	89.9	90.1	88.5
Lipex	Lipase	71.1	55.6	51.2	48.3	-[a]
Savinase	Protease	68	34.1	33.7	48	-[a]
FNA	Protease	65	57	53	51	-[a]
V42[b]	Protease	63.6	-	-	-	-

[a]Poor signal to noise ratio prevented determination of T_{max}. [b]No thermal transitions were detected for the protease V42 in LAS.

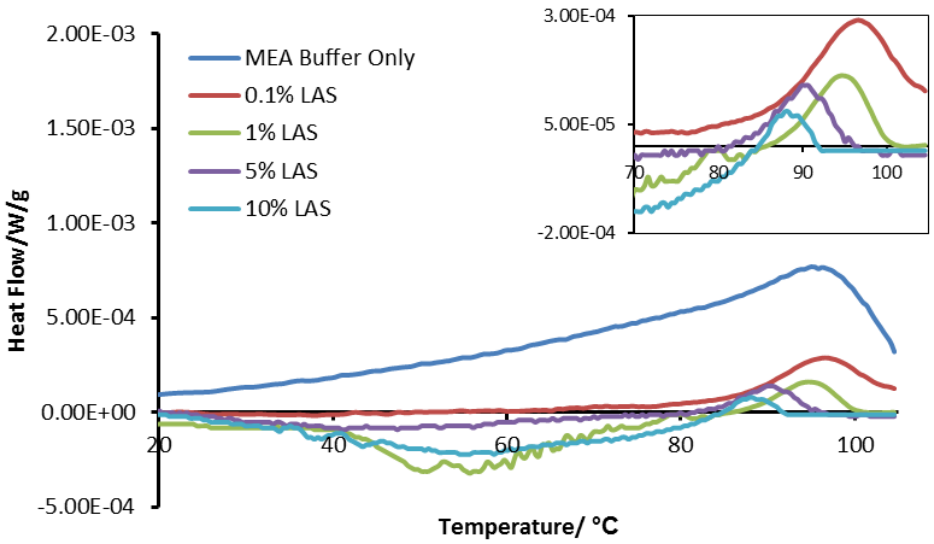


Figure 40: Representative DSC trace of Natalase in varying concentrations of LAS as detected by nano-DSC.

The order of stability of the detergent enzymes was in line with rankings determined with previous methods. The amylases were the most stable, with an average T_{max} of 88.5 °C, followed by Lipex at 68 °C, and lastly, the proteases showed an average T_{max} of 65.3 °C. Due to financial constraint, values are the results of a single analytical run. Instrumental error is listed as correct to within 0.1 °C[17] for dilute, aqueous solutions. Error in more vicious formulations, however, is likely to be much higher.

As evidenced by results presented in *Table 6*, insight into the effects of LAS was provided by nano-DSC at concentrations not possible with optical methods. Amylase samples were stabilised by lower concentrations of LAS (0.1%-1%). In Everest, the effect dropped off at 5% LAS, with the T_{max} reduced from 95 °C to 90 °C and 88 °C for 5% and 10% LAS respectively. Natalase on the other hand, continued to experience an improvement in stability up to 10% LAS, from 82 °C in the control sample, to approximately 90 °C at higher surfactant concentrations. This variation is likely a result of differences in CMC values arising from the protein chains sequestering surfactant monomers from solution.

No stabilisation was observed in the lipase or protease samples, which were far more susceptible to denaturation, even at 0.1% w/v. Savinase showed the least resistance with a decrease in T_{max} from 68 °C to 34 °C. A similar T_{max} of 33 °C was observed at 1% LAS, but at 5% an improvement in stability was observed, with T_{max} recorded at 48°C. FNA was initially less resistant to thermal denaturation when compared to Lipex and Savinase under buffer only conditions. It proved more stable to the presence of surfactant, however, with an 8-14 °C decrease in stability across the concentration range 0.1%-5% w/v. A value for 10% LAS is not listed as no clear transition was detected. No thermal transitions were detected for V42 in the presence of LAS. It is assumed this is an artefact of DSC analysis techniques, rather than the result of exceptionally high or low T_{max} values. As V42 has the lowest T_m value of the enzymes tested, its lower stability may mean that the thermal energy required for unfolding transitions may be too small to detect.

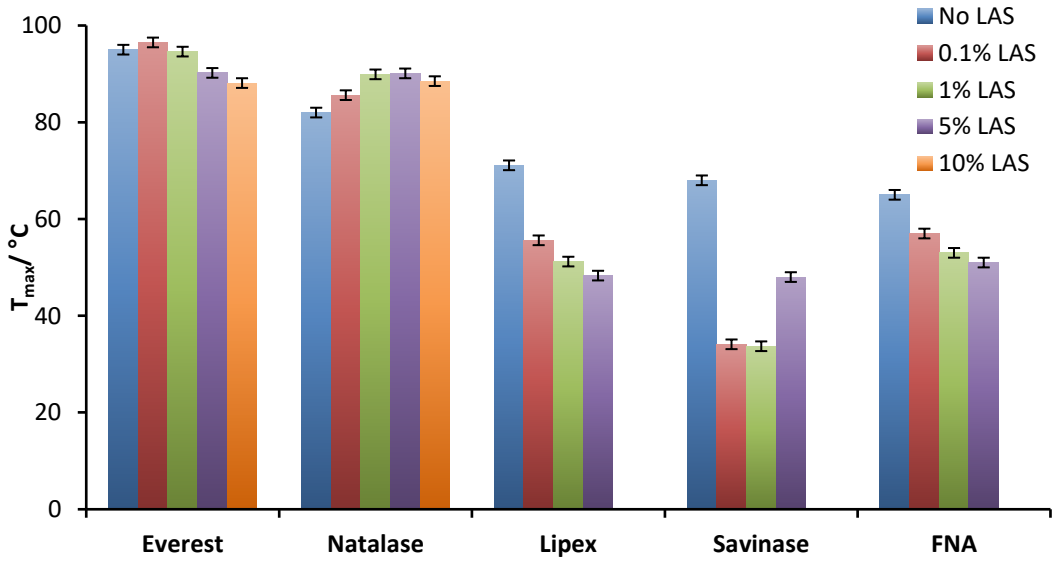


Figure 41: Experimental T_{max} values for a range of detergent enzymes in various concentrations of LAS as determined by nano-DSC. Absent bars are due to incomplete data sets, where thermograms were too noisy to clearly determine a melting temperature. Error in T_{max} estimation is based on reported precision of the instrument and is within 1°C.

3.3.2 Comparison of T_m and T_{max} values from DSF, CD and DSC Analysis

The rank order of enzyme stability was near-identical for DSF, CD and DSC. DSF tended to produce the lowest T_m values and DSC the highest due to the respective parameters used to monitor unfolding. Such observations are in line with similar work described elsewhere in the literature.[18] These different approaches to T_m determination present the opportunity to develop understanding of unfolding mechanisms. Lower T_m values in DSF, for example suggest that loss of hydrophobic interactions occurs at lower temperatures than unfolding of the more structured helices analysed by CD.

Figure 42 illustrates the clear trend between T_m values and T_{max} values obtained using each method. A linear correlation was observed between both $T_{max\ DSC}$ values $T_{m\ CD}$ values as functions of $T_{m\ DSF}$ values. The ability to directly relate techniques provides the opportunity to choose methods appropriate for the required analysis. Each technique provides a different set of advantages and limitations, for example DSF supports high throughput and DSC handles higher surfactant concentrations. Combining data from a range of methods should also produce more robust stability models.

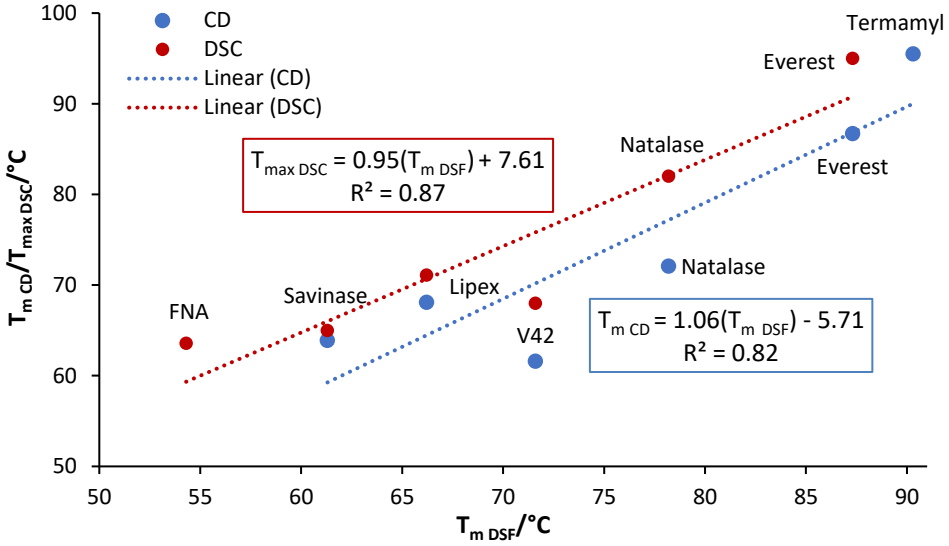


Figure 42: Plot of values obtained for T_m by CD and T_{max} by DSC, as a function of T_m values from DSF, for various detergent enzymes in the presence of MEA buffer only (controls).

3.3.3 The effects of LAS on Unfolding Enthalpy

Alongside changes in T_{max}, enthalpy values for unfolding transitions are provided by nano-DSC, represented by the area under the peak. These values were calculated using the trapezium method of estimation and results are listed in *Table 7*. In each case, the addition of 0.1% LAS reduced the unfolding enthalpy of the enzyme by over 50%, with higher concentrations amplifying the effect (*Figure 43*). The literature reports suggest that this is due to preferential binding of LAS to the unfolded state of the protein which lowers the energy barrier between the native and denatured states. Unfolding then occurs via a co-operative mechanism, resulting in lower values observed at higher concentrations.[19,20]

Table 7: Unfolding enthalpy of detergent enzymes as determined by nano-DSC[a]

Enzyme	No LAS	0.1% LAS	1% LAS	5% LAS	10% LAS
Everest	1.02E-02	3.50E-03	1.20E-03	9.00E-04	4.60E-04
Natalase	4.44E-01	1.70E-04	3.50E-04	4.30E-04	2.60E-04
Lipex	1.41E-02	8.50E-03	8.40E-04	1.91E-03	-
Savinase	4.90E-04	5.50E-03	5.60E-03	9.00E-05	-
FNA	2.08E-04	1.07E-04	5.66E-05	3.16E-05	-

[a]*Enthalpy values calculated from area under curve using 'trapeze' method of approximation. Values are listed in Calories.*

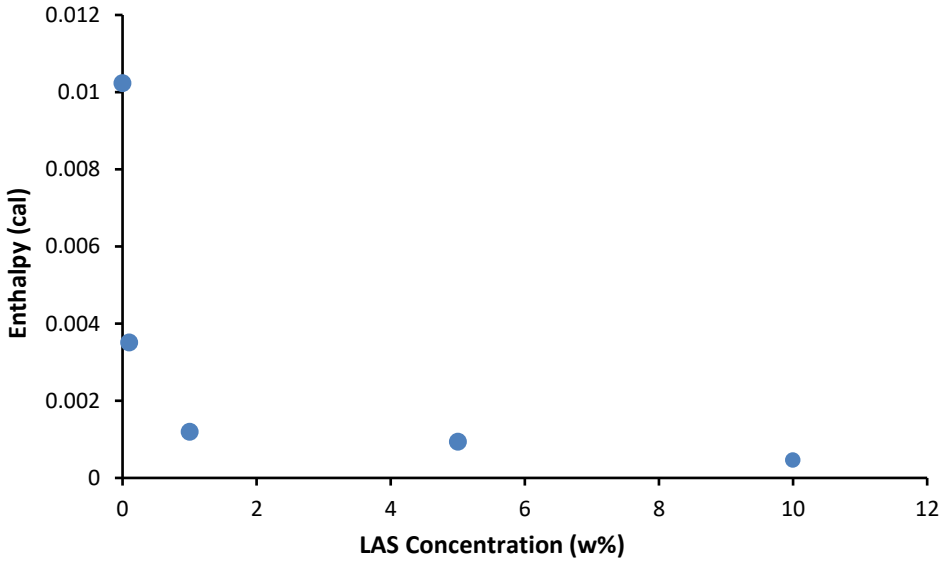


Figure 43: Plot of enthalpy values (cal) as a function of surfactant concentration for the amylase Everest as determined by area under DSC peaks.

This screening experiment emphasises the advantages of nano-DSC analysis in the detergent industry. The technique provided the scope to study protein unfolding at LAS concentrations two orders of magnitude higher than was possible using optical methods. The amylases, in particular, produced strong, clear thermal peaks up to 10% LAS which is in line with commercial HDL.

It was reported in the literature that thermal transitions were detectable at 10% w/v LAS for protease as well as in mixed surfactant samples (10% LAS, 5% AE3s and 5% AE7).[11] This in in contrast with the variable data quality obtained in this study. For example, no clear transitions could be

identified over the noise of the buffer/buffer scan in V42 samples containing surfactant, and unfolding was undetectable for several proteins at 10% LAS. Data was also not obtained for samples at higher LAS concentrations (15-20%) or in multicomponent samples, as the contributions of multiple thermal transitions, combined with small signal intensity of proteins caused poor signal to noise ratios. Furthermore, high sample viscosity presented handling issues, with loading, instrument cleaning and degassing of the sample. Automation was also incompatible with the media.

3.3.4 The Effects of Chelant and Builder on Enzyme Stability

Table 8, gives results from a brief analysis of chelant conditions. Citric acid and the chelant, HEDP were selected as representative examples of their respective excipient classes. The effects of these chelating agents on the stability of enzymes from various classes is illustrated in *Figure 44*.

Table 8: Experimental T_{max} values (°C) for various detergent enzymes as determined by nano-DSC in the presence of citric acid and HEDP

Enzyme	Class	Control	Citric Acid (Builder)	HEDP (Chelant)
Everest	Amylase	95.0	87.1	88.6
Natalase	Amylase	82.0	73.9	70.2
FNA	Protease	65.0	63.6	53.0
V42	Protease	63.6	65.5	66.4

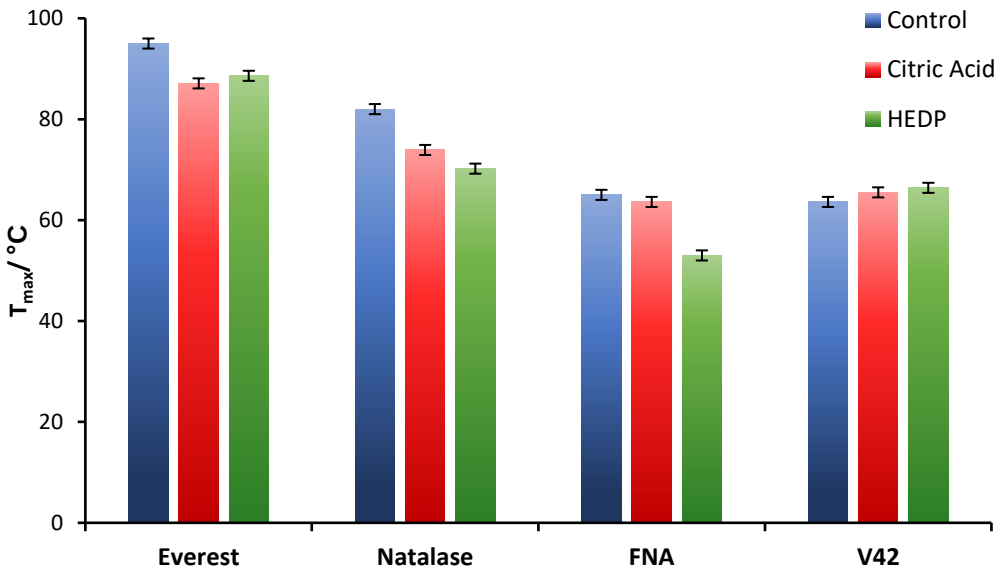


Figure 44: T_{max} values for various detergent enzymes in the presence of HEDP and citric acid as determined by DSC.

With the exception of the protease, V42, both builder and chelant induced destabilising effects across the range of detergent enzymes analysed, as evidenced by decreasing T_{max} values. Everest, which exhibited the highest stability under control conditions, saw a reduction in thermal stability of 8 °C and 7 °C in citric acid and HEDP respectively. Natalase was more vulnerable to HEDP than citric acid, however, with a decrease in T_{max} of 8 °C for the builder and 12 °C for the chelant. No change in T_{max} was observed for FNA presence of citric acid, while HEDP caused a reduction of 12 °C. In contrast, V42 showed an increase in stability in the presence of both chelating agents.

From *Table 9 and Figure 45* we see that studying chelants by two different methods (CD and DSC) gives very different values, which are unpredictable, with no definitive trend between the two datasets. Incorporating these into the plots from *Section 3.3.2*, comparing values from the two methods increases scattering from the linear fit previously observed *(Figure 46)*. The R^2 value is still quite high (0.80), however, suggesting that these changes may not be significant. Further study into the effects of various chelants on protein unfolding, as detected by each technique, may provide further insight into these observations.

Table 9: Comparison of experimental T_m[a] and T_{max}[b] values (°C) in the presence of Chelants/Builders

Excipient	V42 (DSC)	V42 (CD)	Δ T_{max} to T_m[c]	Everest (DSC)	Everest (CD)	ΔT_{max} toT_m[c]
Control	63.6	63.9	+ 0.3	95	86.7	- 8.3
Citric Acid	65.5	58.0	- 7.5	87.1	86.3	- 0.8
HEDP	66.4	57.6	- 8.8	88.6	90	+1.4

[a]T_m values obtained by CD. [b]T_{max} values obtained by DSC. [c]T_m subtracted from T_{max} value for respective samples.

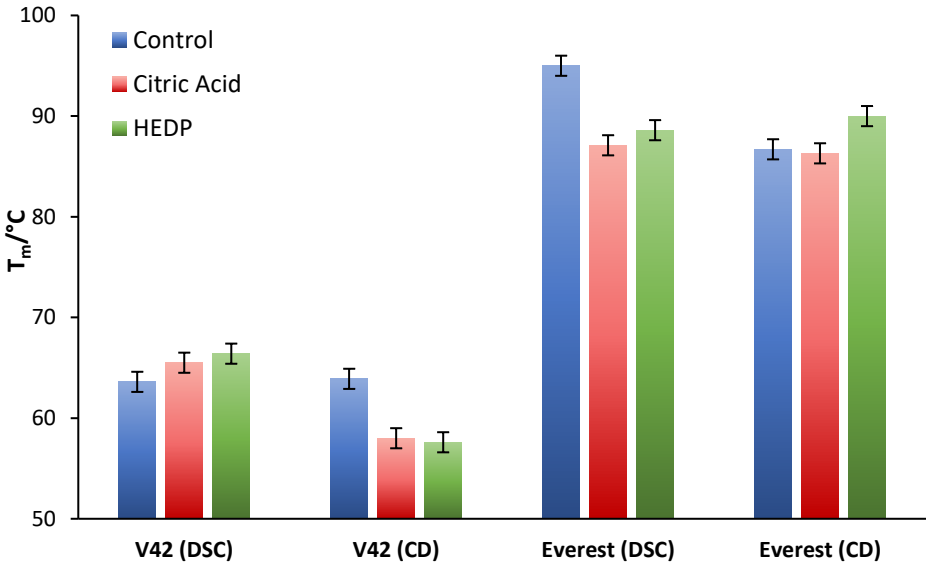


Figure 45: Comparison of T_{max} values obtained by CD and DSC for Everest and V42 in the presence of HEDP and Citric Acid.

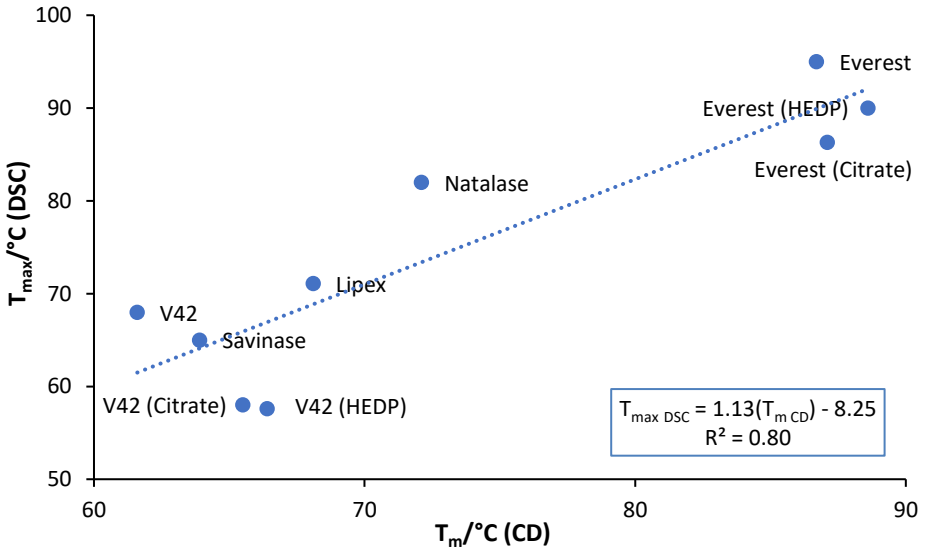


Figure 46: Plot of DSC T_{max} values as a function of CD T_m values for V42 in the presence of HEDP and citric acid.

3.3.5 Reversibility of Unfolding

Alongside improved understanding of the effects of varying LAS concentration on thermal stability, DSC also offered insight into the reversibility of the unfolding process. As illustrated in *Figure 47*, unfolding was found to be irreversible at the T_{max}. Following heating and analysis, samples were allowed to cool in the instrument, before reheating and rescanning. No repeat thermal processes were identified in this second scan, suggesting that structural features are not regained on cooling. Observations were later confirmed using CD spectroscopy to identify refolding at various temperatures. The lack of refolding indicated that samples could be cooled after incubation at required temperatures, and manipulated to facilitate analysis. This will be addressed in the following chapter.

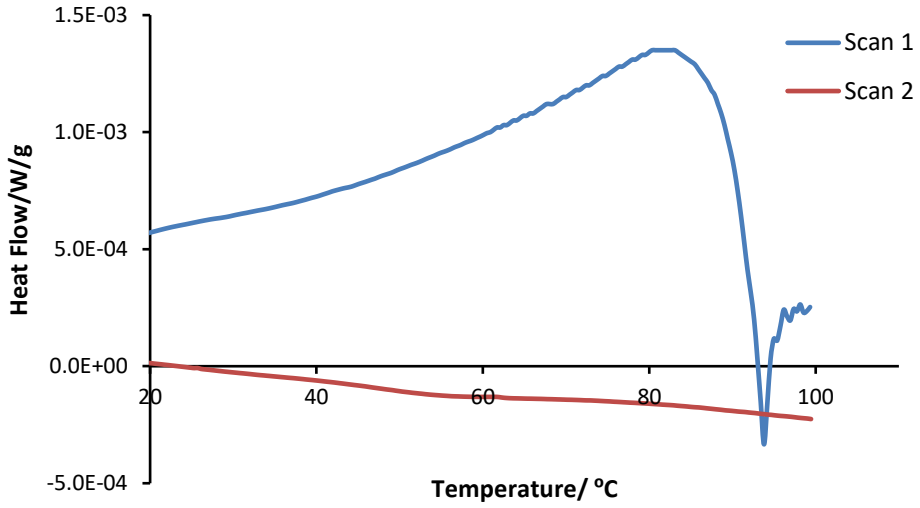


Figure 47: Melting curve of the amylase Natalase overlaid with its repeat scan (Scan 2) as observed using nano-DSC. The repeat scan was conducted by cooling the original sample back to 20 °C after the first run and then reheating to 109 °C. The absence of the initial melting curve indicates that unfolding is irreversible. This was seen across all enzymes and conditions analysed.

Due to the financial constraints involved in analysing samples with a third party, and the limitations of DSC in handling complex multicomponent and highly viscous samples, further analysis by nano-DSC was not continued.

3.4 Protein Analysis by Pulse Proteolysis (FastPP)

The key challenge in the development of FastPP procedures was to ensure the activity of thermolysin remained constant under the stresses of detergent conditions being tested (surfactant, chelant etc.). Reduction of thermolysin activity would affect the level of protein fragmentation, causing enzymes to appear more stable. To avoid this, samples were diluted following denaturation steps to reduce effective surfactant concentration. Refolding analysis by DSC indicated that this would not compromise the state of denaturation of the enzyme *(Section 3.3.5)*.

Samples were assayed at temperature intervals of 5 °C. T_{max} values estimated by SDS-Page analysis *(Figure 48)* of fragmented samples are summarised in *Table 10*. T_{max} values are significantly higher than those observed with other techniques, as loss of a protein band requires up to 100% fragmentation. Gel scanning methods, where available, can be used to estimate 50% fragmentation, as an equivalent for the T_m values of other methods.

Table 10: Estimation of Experimental T_{max} values (°C) from SDS-PAGE analysis of thermolysin assay

Enzyme	Class	Control
Natalase	Amylase	>80
Termamyl	Amylase	>80
Everest	Amylase	>80
Lipex	Lipase	~80
V42	Protease	>70
FNA	Protease	>60

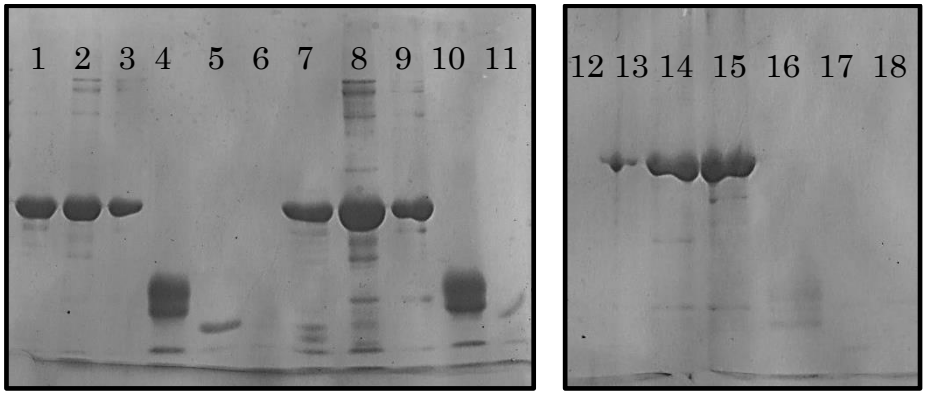


Figure 48: SDS PAGE analysis of FastPP Assay. Lane numbers correspond as follows; Lane 1: Natalase 60 °C, Lane 2: Termamyl 60 °C, Lane 3: Everest 60 °C, Lane 4: V42 60 °C, Lane 5: FNA 60 °C, Lane 6: Termamyl 80 °C, Lane 7: Natalase 70 °C, Lane 8: Termamyl 70 °C, Lane 9: Everest 70 °C, Lane 10: Lipex 70 °C, Lane 11: V42 70°C, Lane 12: FNA 70°C, Lane 13: Natalase 80°C, Lane 14: Termamyl 80 °C, Lane 15: Everest 80 °C, Lane 16: Lipex 80 °C, Lane 17: V42 80 °C, Lane 18: FNA 80 °C

Ranking of enzymes based on observed stability was identical to that of other methods, with amylases and lipases at the higher end of the scale, and proteases 10-20 °C lower. It was not possible to pinpoint exact T_{max} values however, without laborious screening of incremental temperature samples. The numerous steps required for sample preparation and analysis also present multiple opportunities for the introduction of experimental error. This creates variation in band intensity, hindering accurate interpretation of results. Due to these limitations, screening focused LAS handling capability, as other detergent components could be easily analysed with other methods.

It was found that even with sample dilution prior to the addition of thermolysin, the protein was rendered inactive at 1% LAS. As illustrated in *Figure 49*, fragmentation was not detected in any of the samples, including those known to be beyond the T_{max} of the protein such as FNA at 80 °C.

Poor reproducibility and incompatibility with high LAS concentrations, alongside the labour-intensive nature of FastPP, lead to its ultimate

abandonment in favour of more accurate and efficient technologies. Further work in the development of pulse proteolysis for use with HDL, should focus on more effective methods of purification, such as those discussed in the following chapter. Where available, the use of HPLC to separate and identify fragmented and intact protein would streamline this method and improve accuracy.

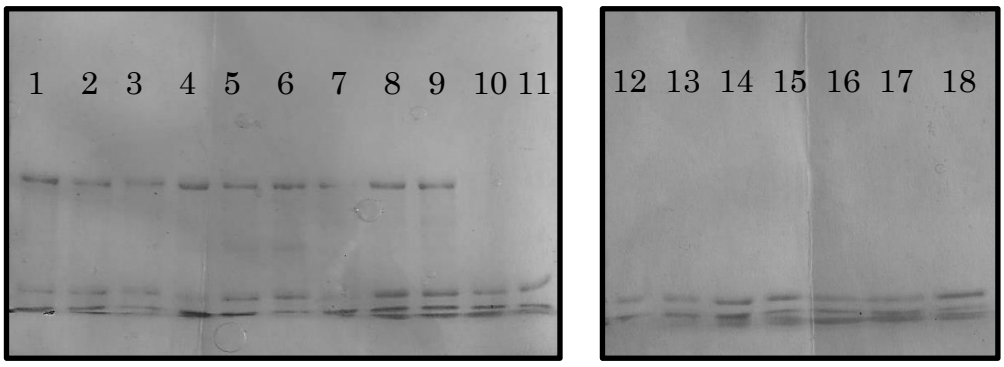


Figure 49: SDS PAGE indicating inactivation of thermolysin by LAS. Lane numbers correspond as follows; Lane 1: Natalase 60 °C, Lane 2: Natalase 70 °C, Lane 3: Natalase 80 °C, Lane 4: Termamyl 60 °C, Lane 5: Termamyl 70 °C, Lane 6: Termamyl 80 °C, Lane 7: Everest 60 °C, Lane 8: Everest 70 °C, Lane 9: Everest 80 °C, Lane 10: Lipex 60 °, Lane 11: Lipex 70°C, Lane 12: Lipex 80 °C, Lane 13: V42 60 °C, Lane 14: V42 70 °C, Lane 15: V42 80 °C, Lane 16: FNA 60 °C, Lane 17: FNA 70 °C, Lane 18: FNA 80 °C.

It was evident from the screening described in this chapter, that the majority of currently available technologies struggle to cope with high concentrations of the surfactant LAS, both due to its UV absorbent properties and high viscosity. To further develop the prediction of enzyme stability in HDL, it is necessary to incorporate multi-component and fully formulated systems, all of which incorporate LAS. The following chapter will, therefore, focus on developing alternative means of determining T_m values for detergent enzymes in the presence of the surfactant, with an aim to expand these techniques to more commercially relevant formulations.

3.5 Conclusions

The primary focus of work detailed in this chapter was to screen a range of methods for capabilities in determining enzyme stability parameters in HDL. These complex formulations present several obstacles to analysis, including the large number of freely interacting components and high levels of viscosity and opacity, which hinder analyte detection. A summary of key traits of each technique are presented in *Table 11*. The various advantages and limitations associated with each of the above methods suggests that a suite of stability analysis techniques is key to obtaining optimal results in HDL.

Table 11: Summary of the advantages and limitations of a range of protein analysis techniques when applied under HDL conditions.

Technique	Output	Advantages	Limitations
DSF	T_m – Δ Hydrophobicity	• High throughput • Automated temperature ramp with qPCR	• Dye-surfactant interactions • Max 0.1% LAS (UV Detector)
CD	T_m – Δ Helicity & α-helix/β-sheet content	• Provides structural info • Automated temp. ramp	• Max 0.1% LAS (UV Detector)
Nano-DSC	T_{max} – Δ Enthalpy	• T_{max} analysis at up to 10% LAS • Automated temp. ramp	• Requires specialised equipment • Poor results at high viscosities
FastPP	T_{max}– Resilience to proteolysis (based on unfolding)	• No specialist equipment required	• Low precision and accuracy • Max 0.1% LAS • Manual heating to each temp. point (Labour-intensive)
MST	T_m – Δ Rate of Thermophoresis	• High Throughput • Can detect intrinsic fluorescence	• Max 0.1% LAS (UV Detector)

The optical methods, CD and DSF, provide precise, robust determination of T_m values, outside of high surfactant conditions. Both techniques offer

higher throughput than DSC, with DSF providing far superior rates by incorporating 96-well plate formats. Throughput can also be improved tenfold with CD, through use of a sample changer, however, the key benefit of this technique lies in the level of structural information it provides.

FastPP, although convenient for the non-specialist lab due to its use of generic equipment, lacks the robustness, accuracy and efficiency required for routine analysis. Conversely, precise T_m determination was possible at in LAS at concentrations up to 10% w/v using nano-DSC. Reports in the literature[11] indicate the method should be compatible with multi-component samples, however this success was not replicated here.[11] Improved detection may be achieved through the use of more rigorous procedures for degassing and baseline subtraction.

The second set of conclusions drawn from this chapter relate to the observed effects of various detergent components on protein stability, as determined using the methods described above. Effects on protein stability varied between the two representative enzymes, indicating trends are not common to both amylases and proteases. Further analysis is required to determine if trends are consistent within an enzyme class.

The effects of AE3S and AE7 on thermal stability are not clearly defined by surfactant concentration. The significance of fluctuations in T_m will be discussed with respect to storage stability *Chapter 5.7*. Chelant effects varied with concentration and enzyme class. Proteases were more resilient to unfolding than amylases due to stabilisation from a second calcium binding site. EDTA induces greatest losses in stability in line with its reported higher association constant with Ca^{2+}. HEDP and the 'builders' (citric acid and fatty acid) also lower T_m values, but to lesser degrees. These weaker chelating agents establish an equilibrium for calcium binding with proteins, reducing loss of structural ions compared to EDTA.[15]

Due to challenges associated with LAS-based samples, only $T_{max\ DSC}$ data was available for concentrations above 0.1% w/v. LAS is generally

considered to be a destabilising element of detergent enzymes. In contrast, stabilising effects were observed among amylase samples at low concentrations, accessible with DSF and CD. These were later confirmed by DSC and have been attributed, by Takeda and co-workers, to surfactant interactions bridging proximal regions of folded protein.[5]

Multi-component analysis was also hindered by the presence of LAS. Brief DSF studies of chelants/surfactant combinations at low concentrations (5 mM of chelant and 0.1% w/v LAS respectively) were not found to have a synergistic effect on protein destabilisation. T_m values of proteins incubated with more than one excipient were found to be identical to samples containing only the more destabilising component. This is unlikely to be true of all excipient combinations, or in complete HDL formulations, however, for simplicity, the following work in this thesis focused on single component systems.

3.6 Future Work

This preliminary work highlighted two clear avenues for further development of protein stability models. The first involves expansion of the above methods to incorporate more complex samples and fully formulated HDLs. Secondly, a focus on the mechanisms behind observed effects of various excipients would support the transition to a 'by design' approach to formulation stability, rather than the 'trial and error' method currently employed.

Challenges associated with determining T_m values of enzymes in LAS-rich environments will be addressed in the following chapter, providing scope to study more relevant levels of the surfactant. This should also support expansion of these methods to multicomponent HDL systems, where LAS is a key component.

CD data will also be probed further to assess the degree of structural information that can be gained from the technique. This should improve understanding of conformational changes induced by ligand binding. Further mechanistic insight can be gained by probing excipient properties. This should include analysis of the effects of detergent-like conditions and the presence of enzyme on surfactant CMC values, protein-surfactant interactions, and the metal ion association constants for both enzymes and chelating agents.

3.7 References

1. S. Shi, A. Semple, J. Cheung and M. Shameem, *J. Pharm. Sci.*, 2013, **102**, 2471–83.

2. F. Vollrath, N. Hawkins, D. Porter, C. Holland and M. Boulet-Audet, *Sci. Rep.*, 2014, **4**, 5625, DOI: 10.1038/srep05625

3. N. J. Greenfield, *Nat. Prot.*, 2006, **1**, 2527–2535.

4. C. Park and S. Marqusee, *Nat. Meth.*, 2005, **2**, 207–212.

5. Y. Moriyama, Y. Kawasaka and K. Takeda, *J. Colloid Interface Sci.*, 2003, **257**, 41–46.

6. D. Aaslyng, E. Gormsen and H. Malmos, *J. Chem. Technol. Biotechnol.*, 1991, **50**, 321–330.

7. W. Bode, E. Papamokos and D. Musil, *Eur. J. Biochem.*, 1987, **166**, 673–692.

8. P. A. Alexander, B. Ruan, S. L. Strausberg and P. N. Bryan, *Biochem. (Mosc.)*, 2001, **40**, 10640–10644.

9. A. K. Bhuyan, *Biopolymers*, 2010, **93**, 186–199.

10. D. Kraut, H. Goff, R. K. Pai, N. A. Hosea, I. Silman, J. L. Sussman, P. Taylor and J. G. Voet, *Mol. Pharmacol.*, 2000, **57**, 1243–1248.

11. H. Lund, S. G. Kaasgaard, P. Skagerlind, L. Jorgensen, C. I. Jorgensen and M. van de Weert, *J. Surfactants Deterg.*, 2012, **15**, 9–21.

12. A. Crutzen, M.L. Douglass, *Handbook of Detergents: Part A, Properties*, CRC Press, 1999.

13. M. T. Henzl, J. D. Larson and S. Agah, *Anal. Biochem.*, 2003, **319**, 216–233.

14. R. P. Enever and N. Pilpel, *Trans. Faraday Soc.*, 1967, **63**, 1559–1566.

15. H. Lund, S. Kaasgaard, P. Skagerlind, L. Jorgensen, C. Jørgensen and M. van de Weert, *J. Surfactants Deterg*, 2012, **15**, 265–276.

16. Y.-R. Na and C. Park, *Protein Sci.*, 2009, **18**, 268–276.

17. N. Garbett, L. DeLeeuw and J. B. Chaires, *TA Instrum. – Appl. Note.*, Retrieved from: http://www.tainstruments.com/wp-content/uploads/MCDSC_Broch_2012.pdf

18. D. S. Goldberg, S. M. Bishop, A. U. Shah and H. A. Sathish, *J. Pharm. Sci.*, 2011, **100**, 1306–1315.

19. T.W. Randolph, M. N. Jones, *Surf. Act. Prot.*, 1996, **13**, 237–284.

20. D. Otzen, *Biochim. Biophys. Acta*, 2011, **1814**, 562–591.

4 Alternative Approaches to Enzyme Analysis in LAS-Rich Environments

Surfactants have been at the centre of the laundry industry since the first synthetic detergents were developed by Proctor and Gamble in the 1940s. These amphiphilic molecules provide the majority of cleaning power in both powdered and liquid formulations, by lifting stains from fabrics, and encapsulating dirt particles in micelles.

Modern detergents consist of up to 40% surfactant by weight, with a combination of both non-ionic and anionic compounds. Non-ionic surfactants such as alcohol ethoxylates (AE) exhibit a greater ability to lift stains from fabrics and are less sensitive to precipitation in hard-water environments. The use of the anionic surfactant LAS, however, is still far more prevalent across the industry, at up to 15% of total HDL formulations by weight, due to its ease of manufacture, low cost and biodegradability.[1–3] Interactions between LAS and detergent enzymes are more destabilising than those of AE due to the high affinity of the anionic head-group for the basic residues arginine, lysine and histidine. Thus, the ability to study LAS-rich samples is crucial to understanding enzyme inactivation in HDL.

The inclusion of LAS in enzyme samples presents a specific analytical challenge, due to high levels of UV absorptivity and viscosity. DSC is the most commonly used method of T_m analysis in HDL,[4–6] however, the technique has very low throughput and provides little mechanistic insight into protein unfolding processes.[7,8] LAS-rich formulations are also above the viscosity limit for the use of autosamplers and are difficult to degas, resulting in noisy thermal profiles which can mask small protein unfolding signals. It was therefore necessary to consider other means of exploring enzyme inactivation in these systems. Thus, two alternative routes of

analysis were developed; the first involved the use of another common anionic surfactant, SDS as a structural analog for LAS while the second removed LAS from protein samples prior to analysis. In both cases CD was used in T_m determination, enabling comparison of effects on secondary structure and avoiding the need for external dyes.

This chapter is divided into three sections. The first describes T_m data collected for the enzymes V42 and Everest in the presence of the analog SDS. The second details validation of various LAS removal methods, including precipitation with $CaCl_2$ and the resultant T_m values obtained from purified samples. The graphs from which these values were calculated can be found in *Appendix 2*. These data will also be compared with SDS data from the previous section to demonstrate the relationship between observed protein-surfactant interactions. Finally, spectra obtained from these analyses will be analysed using Dichroweb deconvolution software and principal component analysis (PCA). Although full structural determination was not possible with the available data, further support of the use of SDS in stability models for enzymes in detergent formulations was provided.

4.1.1 SDS as an Analog for LAS

Comprehensive descriptions of the anionic surfactant SDS are available in the literature due to its common usage as a denaturant in SDS PAGE. Structural similarities with LAS led us to predict that the two would elicit similar effects on detergent enzymes. Alongside the alkyl chains, common to all surfactants, both the head groups of both LAS and SDS contain sulphonate groups. The absence of an aromatic ring in SDS avoids problems of UV detector saturation which undermined experiments in LAS. SDS also dissolves more readily in aqueous solutions and is less viscous at room temperature, which reduces handling issues.

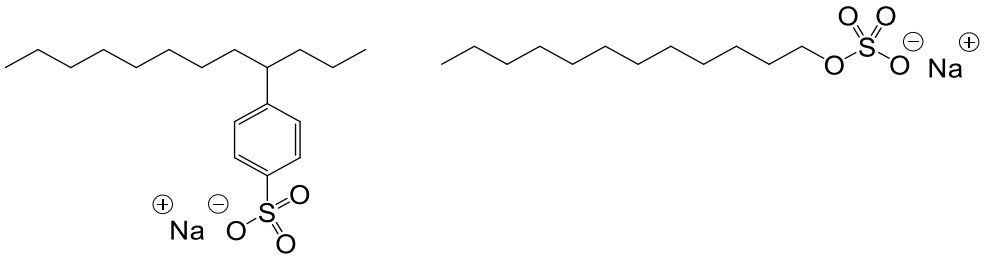


Figure 50: Structures of LAS (left) and SDS (right). The two surfactants contain a sulphonate group and alkyl chains, however, the UV active aromatic ring is absent in SDS.

The mechanism of SDS-induced unfolding is far better defined in the literature than that of LAS. It has also been reported that the two surfactants induce comparable levels of instability in proteins.[9] Although this proves only an empirical relationship, the corresponding similarities in charge and structure would suggest that further understanding of LAS binding can be gained by studying the interactions of SDS with detergent proteins.

SDS-induced unfolding of proteins is highly concentration-dependant.[2,10] At low concentrations, two binding phases are observed. The first occurs at sub-micellar concentrations (below ~1.5 mM). This involves monomeric binding, which leads to loss of tertiary structure, but retention of the majority of secondary structure. Monomers bind protein through a combination of electrostatic and hydrophobic interactions. Initial binding occurs between sulphonate groups of the anionic surfactant and basic side chains with further monomers binding in a co-operative fashion, as internalised hydrophobic residues are exposed to surfactant.

Following this initial stage, a plateau, with no increase in levels of denaturation, is observed around the point of the CMC. Once surfactant concentration exceeds the CMC, micelle formation begins along the protein chain, seeded by bound monomers. This produces a 'necklace' model of unfolded protein. As the micelles become large enough to interact, they repel each other, resulting in further loss of protein structure.

Although SDS and LAS are very structurally similar, several key differences have been noted in the literature which prevents direct application of the process described above to LAS and detergent proteins. Commodity grade LAS, used in preparation of detergent samples, consists of a range of different conformers, differing in the positioning of the aromatic ring. Chain lengths also vary between 10 and 13 carbon atoms. SDS, on the other hand, consists of a single conformer, shown in *Figure* with 12 carbons in each hydrophobic chain. These differences effect observed surfactant aggregation. For example the CMC of SDS in pure water is 8 mM, while that of LAS is generally listed as 6.9 mM, but can be less than 1 mM depending on the ratios of chain length (longer chains have lower CMC values).[11] SDS also tends to form micelles composed of 40 monomers with a smaller diameter at 3.5 – 4 nm. In contrast, LAS aggregates contain an average of 27 monomers which are less tightly packed, with a diameter of 12.5 nm.[12–15]

SDS tends to maintain this micellar form up to approximately 40 wt % at room temperature. Phasing is also clearly defined across the concentration and temperature profile. LAS, however, exhibits several lamellar phases, which tend to overlap and co-exist due to the presence of different chain lengths. Phase diagrams constructed by Rossi *et al*[16] and Stewart *et al*[17] *(Figure 51)* show these aggregation states in water, however the presence of proteins, which sequester surfactant molecules, may alter the observed phases.

To determine if these differences in surfactant properties would affect the interactions with protein chains, enzymes were studied across a range of concentrations from 0.1% to 20% w/v surfactant. Thermal denaturation and structural analyses were conducted using CD spectroscopy. In the absence of a comparable LAS data set, thermal denaturation T_{max} values obtained by DSC, as described in the previous chapter, were used to confirm common trends in the effects of the two surfactants on proteins.

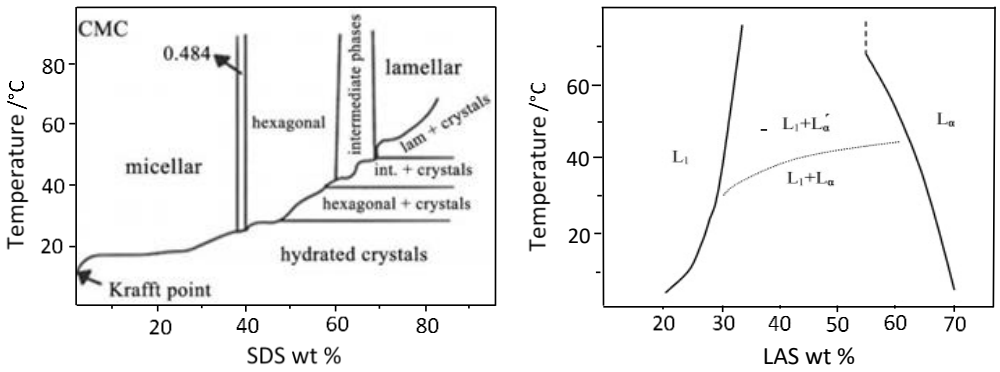


Figure 51: Phase Diagrams of SDS (left) and LAS (right) in water at various temperatures and concentrations. Figures reproduced with permission from Rossi et al[16] and Stewart et al[17].

4.1.2 Removal of LAS prior to analysis.

The second approach to probing LAS-induced denaturation in detergent systems necessitated the removal of LAS from formulations following sample incubation. In certain cases, protein structure and activity can be regained following unfolding. This process usually requires extensive dialysis or immediate reduction in temperature after heating, however and often only occurs in cases where temperatures do not exceed the T_m.[10,18] The detergent enzymes used in this work were not found to return to the native state during analysis by DSC *(Figure 52)*, in line with finding by Lund for similar proteins.[5] This property was employed in the development of surfactant removal methods for studying LAS-rich samples.

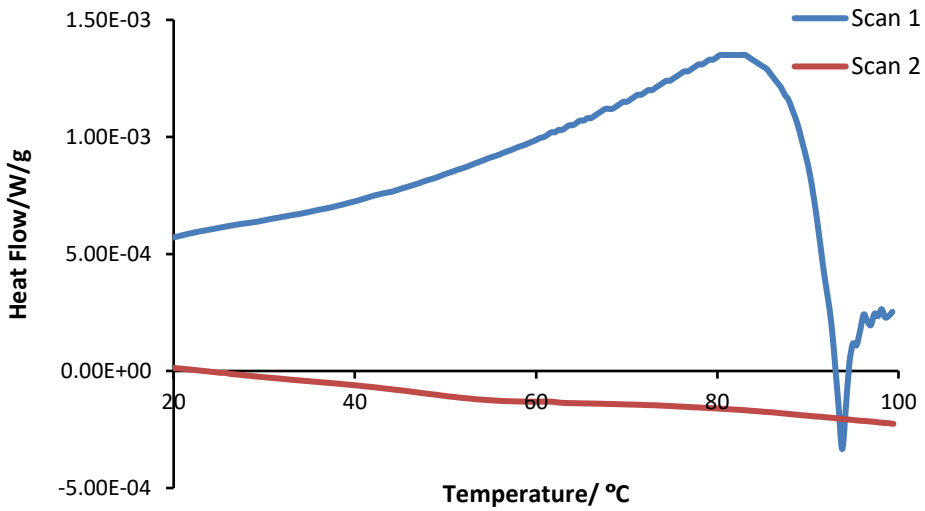


Figure 52: Irreversible unfolding of Natalase by DSC. The sample was cooled in the sample port after reaching its T_{max} and then reheated. No thermal transitions were detected in the second scan.

In order to accurately assess levels of protein denaturation, it was crucial that surfactant removal processes did not interfere with structural features retained following incubation. Several commonly used purification methods were attempted including dilution, size exclusion and ion exchange chromatography *(Section 4.4.1)*. Ultimately purification was conducted by precipitating LAS from the solution using $CaCl_2$. A common problem with detergents is the presence of Ca^{2+} ions in hard-water areas, as two anionic surfactant monomers can bind to the metal, forming a precipitate on fabrics. To prevent this, chelants and builders are added to laundry formulations.

This process was not found to affect the structure of detergent enzymes, however, as conversely, they require calcium for structural stability *(Section 4.5.1)*. This provided a means of removing the surfactant from solutions through centrifugation of the precipitate, without altering the level of protein denaturation.

The procedures described above were used to collect T_m values for two detergent enzymes, the protease, V42, and the amylase, Everest. These representative examples provided empirical proof of concept for each approach to analysis of surfactant rich samples. Further validation was

conducted by probing protein structures using CD deconvolution software and principal component analysis (PCA). These techniques provide greater insight into the effects of protein-surfactant interactions on enzyme structure and stability. In addition, the presented results demonstrate the potential to use these methods in developing mechanistic understanding of enzyme inactivation in HDL.

4.1.3 Dichroweb and PCA

The 'Dichroweb' library of deconvolution programmes, developed by Wallace and Whitmore employs existing databases of CD spectra for proteins with known crystal structures to estimate proportions of secondary structures in unknown proteins. Several methods of deconvolution are available, with each programme being tailored to specific protein structures. For example, CDSSTR is most accurate at predicting helical content of proteins, while CONTIN provides better accuracy in proteins with more β-sheet character. Both of these programmes will be applied to CD data in this chapter to assess which is most appropriate to the detergent enzymes.

Assigning proportions of secondary structural features produced some unexpected results which were inconsistent with known properties of similar enzymes. A second method, principal component analysis, was therefore applied to the spectra to verify these observations. This multivariate technique is often incorporated into deconvolution software as a tool for comparing complex data sets with multiple points of variation, such as CD spectra. Reference spectra are not applied in this analysis, therefore assigning specific structural features is not possible. Instead, proteins of similar structure are identified using scoreplots.

Principal components are values which describe multiple points of concurrent variation in a spectrum. The first principal component (PC1) describes the majority of variation in a dataset, with additional PC values representing lesser degrees, observed simultaneously along a different

axis. These values can be used comparatively, as markers of overall change in protein structure. Plotting two principal components against each other, produces the scoreplot, with proximal points indicating spectra of similar structure. Generally, PC1 and PC2 are used to generate this plot, as they describe the greatest proportion of variation. In this manner, a clear visual comparison of large numbers of spectra is achieved. These methods will be applied in *Sections 4.7-4.8* to validate novel approaches to protein analysis in LAS-rich systems presented earlier in the chapter.

4.2 Results - SDS as a Substitute for LAS in Detergent Formulations

The data reported in *Table 12* below, detail T_m values obtained by CD for the enzymes V42 and Everest in the presence of varying concentrations of SDS. These results will be compared with trends observed for corresponding LAS samples using nano-DSC, to determine if SDS is a suitable analog for optical analysis. Protease analyses were conducted in the presence of PMSF for consistency with previous experiments.

Table 12: T_m values determined by CD for the enzymes V42 and Everest in varying concentrations of SDS.[a,b]

SDS concentration	V42 T_m[c]	Everest T_m
Control	63.9 *(± 0.6)*	86.7 *(± 0.6)*
0.1%	62.3 *(± 0.9)*	89.9 *(± 2.3)*
1%	57.4 *(± 1.8)*	87.7 *(± 1.5)*
2.5%	55.4 *(± 0.6)*	88.9 *(± 0.6)* [d]
5%	53.1 *(± 0.3)*	83.8 *(± 1.2)*
7.5%	53.3 *(± 0.3)*	87.5 *(± 1.6)* [d]
10%	51.6 *(± 1.6)*	84.6 *(± 1.2)*
20%	50.5 *(± 1.8)*	80.1 *(± 0.0)* [d]

[a] T_m values determined using temperature ramp function of CD spectrometer. [b] T_m values are reported as averages of triplicate repeat scans. Values in brackets represent error in T_m, reported as the standard error of these triplicate results. [c] Analyses on the protease V42 were conducted in the presence of PMSF. [d] T_m values are an average of two analyses.

4.2.1 The Effect of Various Concentrations of SDS on the Protease, V42

At all concentration points analysed, a decrease in V42 T_m values was observed in the presence of SDS when compared to the control. The destabilisation became more significant with increasing concentrations of surfactant. At 0.1% SDS, a reduction in T_m of 1.5 °C was observed, this increased to a loss of 13.4 °C at 20% SDS. *Figure 53*Figure 54 illustrates the shift of V42 melting towards lower temperatures with increasing SDS concentrations.

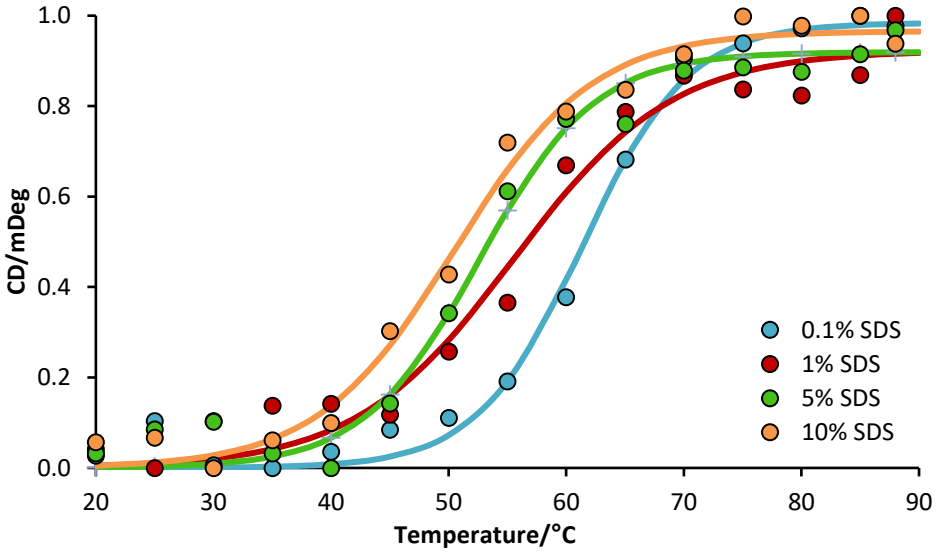


Figure 53: Melting curves of V42 in the presence of various concentrations of SDS.

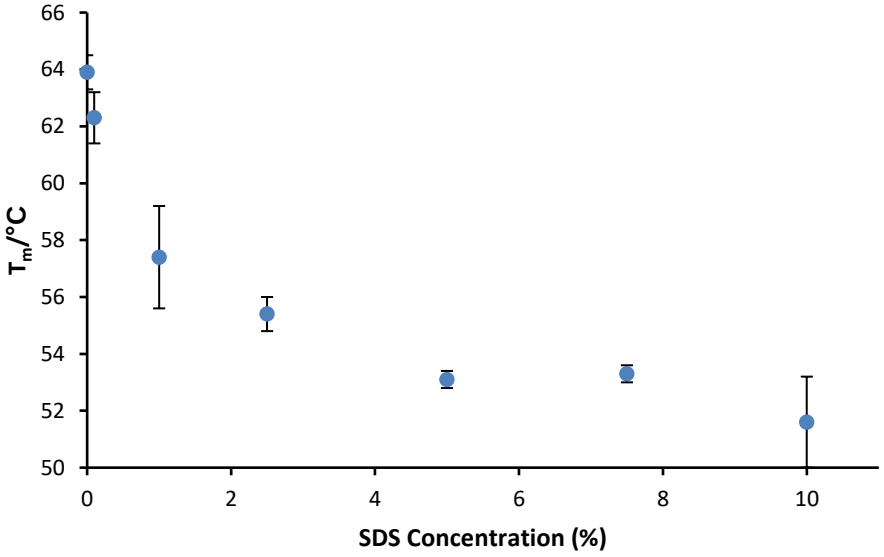


Figure 54: Plot of V42 T_m values as a function of sample SDS concentration.

4.2.2 The Effect of SDS on the Amylase, Everest

A similar trend of decreasing stability with increasing SDS concentration was observed among the amylase samples between 0.1% and 10% w/v. When compared to the control sample, however, addition of 0.1% SDS induced a stabilising effect, with a 3 °C increase in T_m. A plot of T_m values against surfactant concentration showed a generally downward trend *(Figure 55)*. Both data scatter and standard error of triplicate repeats, however, are greater than that of V42. As discussed in *Chapter 3*, this is likely due to higher T_m values which approach the limit of the instruments heating function.

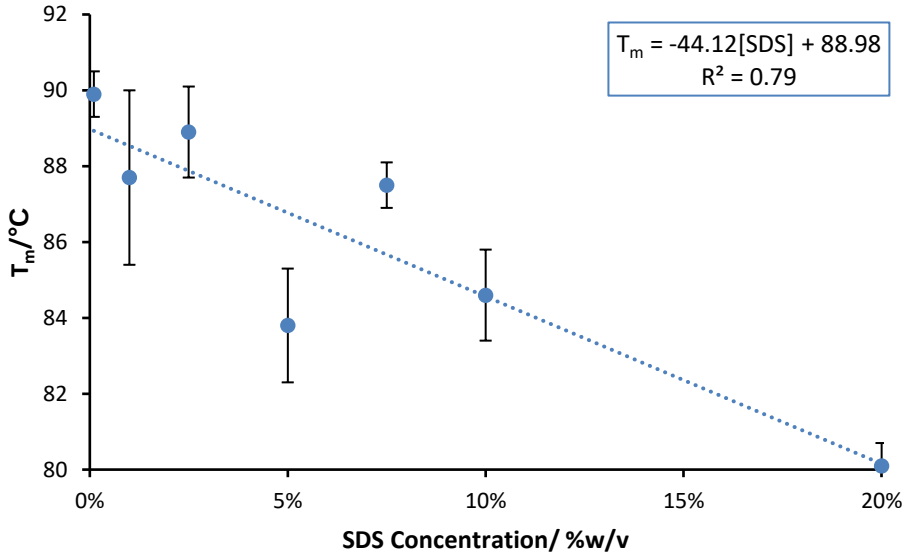


Figure 55: T_m as a function of SDS concentration for the amylase, Everest.

4.3 Comparison of SDS and LAS data collected using CD, DSC and DSF for Everest

The successful collection of T_m values, reported above, demonstrate the improved compatibility of SDS-based mock formulations with CD analysis. To establish the relationship between the effects of SDS on T_m with those of LAS, a comparable dataset based on the latter was required. Corresponding CD data were not available at concentrations beyond 0.1% w/v., therefore observations from nano-DSC experiments, described in *Chapter 3*, were compared with CD data for SDS *(Table 13)*.

Table 13: T_m (°C) and T_{max} (°C) values for Everest in SDS from CD and nano-DSC measurements.[a]

Surfactant Concentration	SDS T_m[b]	LAS T_{max}[c]
Control	86.7 (± *0.6*)	95.0 (± *1.0*)
0.1%	89.9 (± *2.3*)	96.5 (± *1.0*)
1%	87.7 (± *1.5*)	94.6 (± *1.0*)
5%	83.8 (± *1.2*)	90.2 (± *1.0*)
10%	84.6 (± *1.2*)	88.1 (± *1.0*)

[a]Nano-DSC experiments were conducted on a contract basis by Dr Manfield at the University of Leeds. [b]Reported T_m values are an average of triplicate analyses. Values in brackets represent the standard error of the three runs. [C]T_{max} values are the result of a single run, error is an approximation based on reported instrumental error values.

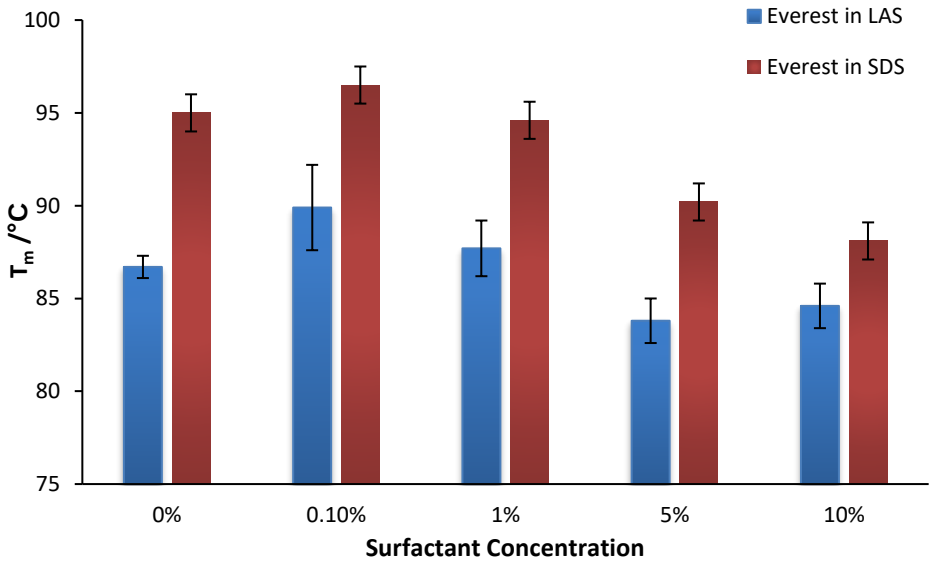


Figure 56: Comparison of Everest T_m values in SDS and T_{max} values in LAS determined by CD and nano-DSC respectively.

The two surfactants exhibited similar trends in T_m/ T_{max} values across the range of concentrations analysed *(Figure 56)*. Both LAS and SDS were found to induce a stabilising effect on Everest at 0.1%, when compared to the control. The effect was more pronounced in SDS, with an increase in T_m of 3 °C compared with 1.5 °C in LAS.

The stabilisation of Everest at low surfactant concentrations was also observed in the presence of LAS by DSC and DSF *(Chapter 3.3.2)*. This effect has been reported in the literature among anionic surfactants.[19] The improvement in stability is thought to be caused by the surfactant molecules linking basic residues with proximal hydrophobic regions,

promoting retention of folding in the native state. Stabilisation is lost at the CMC, as charges on the exterior of these aggregates repel each other causing strain on protein structure.

Following this initial stabilisation, a reduction in both T_m and T_{max} values was observed with increasing surfactant concentrations. At 1%, both surfactants gave T_m values approximately equal to those of their respective control samples. This downward trend continued at 5% and 10%, with further decreases in stability. LAS induced a decrease in T_{max} of 5 and 7 °C respectively when compared with the control samples, whereas the destabilisation effect appeared to tail off in SDS with both 5% and 10% T_m values being the same, showing a 2-3 °C decrease relative to the control.

As illustrated in *Figure 57,* a plot of T_m values in the presence of SDS against T_{max} values with LAS suggests a linear correlation may exist between the two. This supports the use of SDS in place of LAS where analysis was hindered by the latter. Conclusive determination of the relationship between the two surfactants was not possible, however, due to the limited number of samples and outlying data point at 10%. Due to the financial costs and slow turnaround time associated with external DSC analysis, it was not possible to run more samples to confirm the trend. Thus, alternative methods of direct analysis of LAS-induced unfolding were required, leading to the use of sample purification to facilitate the use of conventional techniques, available at Durham University. These methods will be described in detail in the next section.

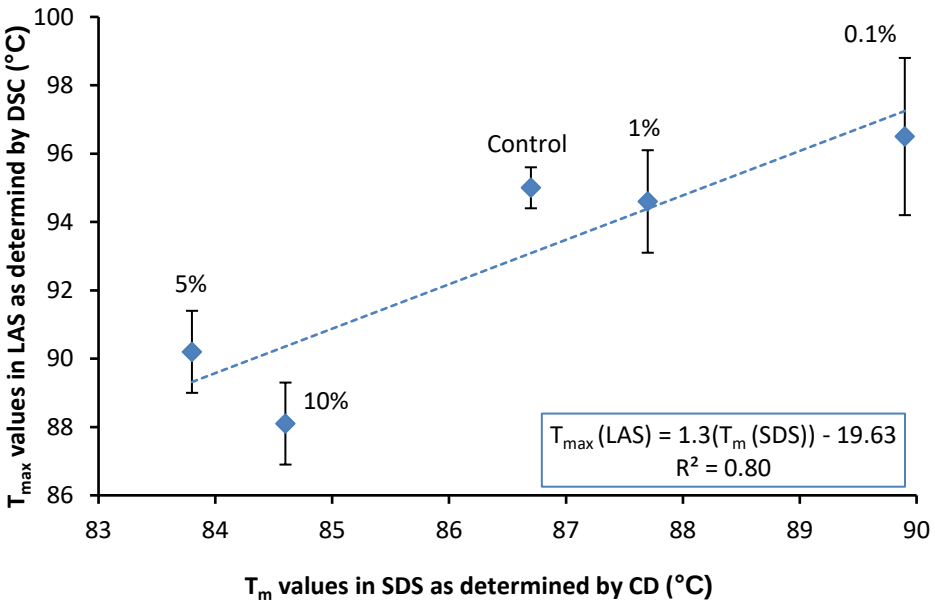


Figure 57: T_m values of Everest as determined by CD in the presence of SDS plotted against T_{max} values determined by nano-DSC in LAS.

4.4 Analyses of LAS-Induced Denaturation by CD using Surfactant Removal Methods.

As previous attempts at direct analysis of LAS-induced unfolding had been hindered by detector saturation, further work focused on reducing this interference. It was observed that concentrations above 0.1% LAS resulted in detector saturation in the case of both CD and DSF. Reduction of surfactant concentration to below this level should facilitate detection of unfolding signals. As CD provides information on protein structure in the form of ratios of α-helical and β-sheet content, it was selected as the technique for further T_m determination. This presented an opportunity to further probe the process of unfolding in terms of specific structural motifs.

4.4.1 Methods of Surfactant Removal

As detector saturation was evident in the presence of LAS 0.1% w/v, the chosen method of purification needed to be capable of reducing surfactant concentration from 20% to less than 0.1%. Furthermore, enzyme concentration must be maintained at or above the detection limit of ~0.1 mg/ml. The degree of unfolding must remain unchanged by the removal

process. Four different techniques for reducing surfactant concentration were explored, namely dilution, size exclusion, ion exchange chromatography and precipitation. As will be discussed below, precipitation using $CaCl_2$ was the only method capable of successfully separating surfactant from the sample without affecting protein analysis.

The first method involved simple sample dilution to reduce surfactant concentrations rapidly and without damaging protein structure. At 5% surfactant and above, however, the ratio of surfactant to analyte was too high to detect protein signals following dilution.

Surfactant removal was then attempted using size exclusion spin columns and ion exchange FPLC. Size exclusion columns with a molecular weight cut off of 7000 g/mol were selected.[28] This value is above the molecular weight of LAS, which varies based on chain length with an average of 326 g/mol in commercial formulations. The proteins, at 20 – 55 kDa, should pass freely through the column. On analysis of the flowthrough by UV-Vis, however, LAS was found to co-elute with the protein. We believe this was a result of the formation of micelles in the solution, increasing the molecular mass of the surfactant aggregate above that of the MWCO. The CMC of LAS in aqueous solution is 0.1 g/L or 0.01% w/v. This value is slightly higher in the presence of protein, as binding to the peptide chain reduces the availability of the monomer for micellation. Despite this, micelles were expected to be formed in all samples analysed (0.1% - 20% LAS). LAS has an average aggregation number of 27[15] and an average molecular weight of ~300 $gmol^{-1,}$ dependant on ratios of chain length. The micellar mass is therefore above that of the MWCO, resulting in the observed co-elution with the protein.

Ion exchange FPLC also proved unsuccessful. Using anionic exchange resin, the high levels of surfactant overloaded the column, causing breakthrough of LAS into the eluted sample. Multiple passes through the regenerated columns and the use of multiple columns in series were

limited by time and cost. The use of preparative columns with a higher binding capacity may facilitate sufficient removal and provide an automatable means of purification. As demonstrated in *Figure 58*, removal using cationic resin was also attempted. Efficient binding of protein to the column was not possible, however, with breakthrough of the sample on loading. This has been attributed to LAS monomers binding to the surface of the protein and interfering with resin interactions. Sufficient concentrations of the enzyme could not be purified to continue with this method.

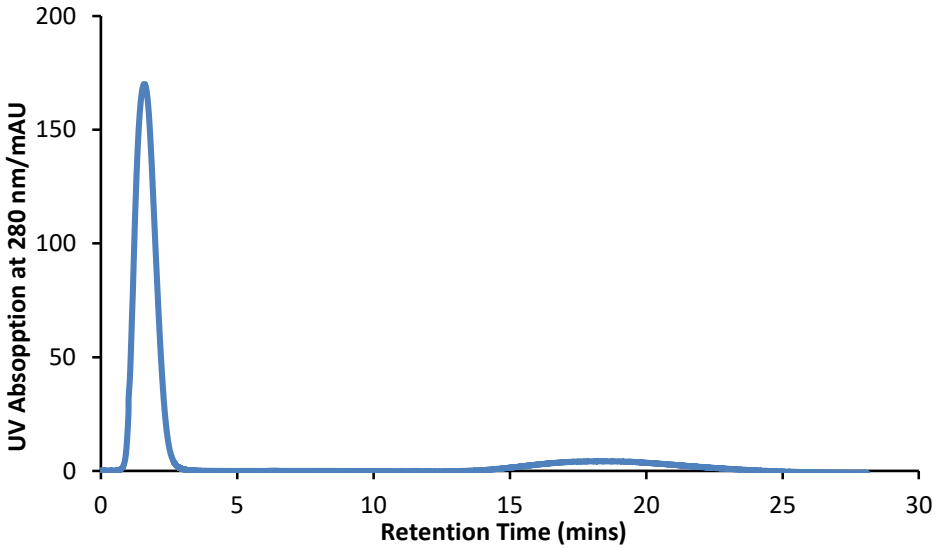


Figure 58: Chromatogram from the attempted purification of V42 using a cationic column. Co-elution of the protein with the surfactant was observed due to poor retention on the column.

As conventional methods of protein purification described above were ineffective in the removal of LAS, a final alternative was considered. It is commonly known that LAS precipitates in the presence of calcium ions, necessitating the use of chelants and 'builders' in detergent formulations desired for use in hard water areas. On addition to LAS-rich formulations, it was found that the surfactant could be precipitated from solution without affecting the solubility of the enzyme. The precipitate was then removed by low speed centrifugation *(Chapter 2.3.2)*. Analysis by CD indicated that the level of denaturation of the sample was maintained following the procedure. This validation will be described in the following

section, along with methods for T_m determination in the purified solution. T_m data relating to LAS formulations will then be compared to those for SDS reported in *Section 4.2* to determine the correlation between the effects induced by the two surfactants.

4.5 Validation of Calcium Ion-Based LAS Removal Method and CD Analysis for Assessment of Denaturing Effects of LAS

4.5.1 Analysis of the effects of Surfactant Removal on Protein Structure.

As discussed in *Chapter 3.2.9*, it was possible to determine melting temperatures of enzymes in the presence of LAS up to 0.1 wt%. This provided some comparable data for *in situ* analysis and samples purified via $CaCl_2$ precipitation to ensure the process was not affecting protein structure.

Samples analysed *in situ* at 0.1% LAS were found to give spectra with broader peaks and more noise than those of the surfactant removed samples *(Figure 59)*. These peaks were initially assumed to identify changes in protein structure at low temperatures due to binding of the surfactant. Analysis of the purified sample, however, produced cleaner spectra, similar in structural features to that of the control sample, as would be expected at 20 °C. This indicates that LAS interference with UV detection is still prevalent at low concentrations.

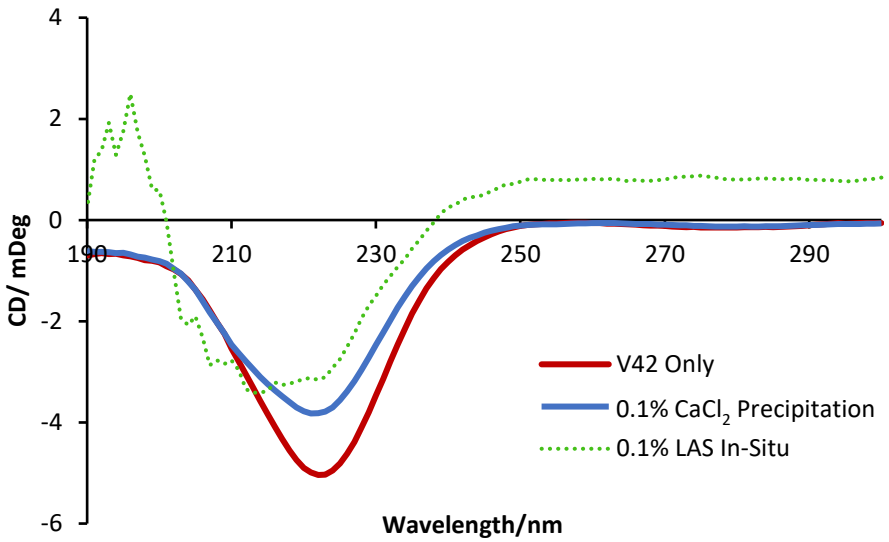


Figure 59: CD Spectra of V42 in the presence of LAS, SDS and in MEA buffer only.

The spectrum of the purified sample showed a reduction in intensity of approximately 20% when compared to that of the control, and also a slightly lower high tension (HT), *(Figure 60-*

Figure ***61****).* Loss of CD peak intensity is normally attributed to unfolding, however, the concurrent loss in HT indicates a reduction in protein concentration arising from the purification process. This was confirmed by normalising spectra, producing two curves of almost identical shape *(Figure 62).* We have assumed that both native and denatured protein are lost during sample purification in equal amounts, preventing bias in the detection of either state.

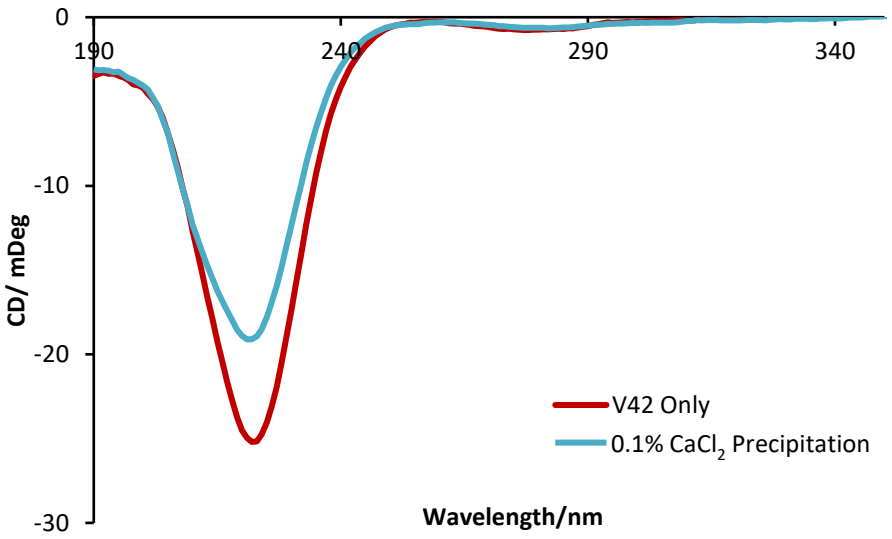


Figure 60: CD spectra of V42 incubated at 20 °C under nil detergent, control conditions (MEA buffer at pH8, red) and in 0.1% LAS, with surfactant removal by $CaCl_2$ precipitation and centrifugation (Blue).

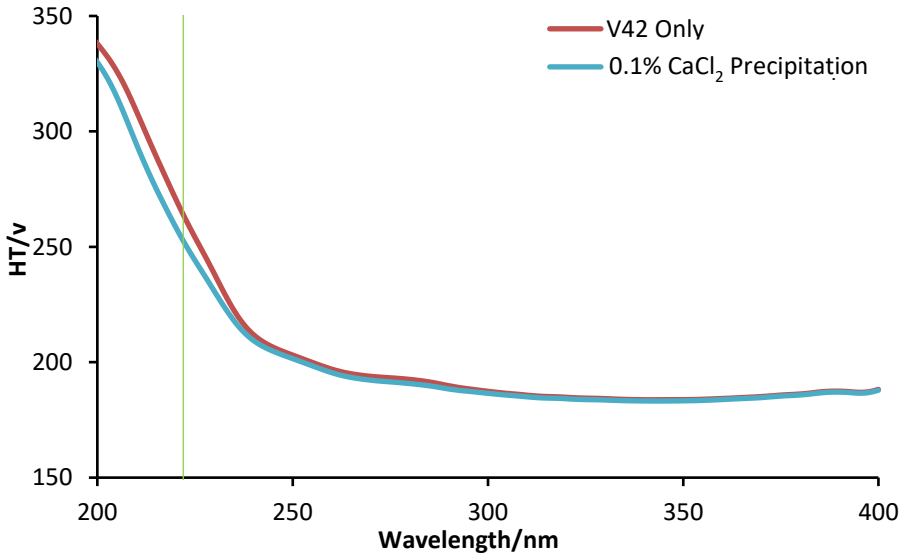


Figure 61: The HT voltage of V42 with surfactant removed (initial concentration 0.1% LAS) for spectra reported in Figure 60 (above).

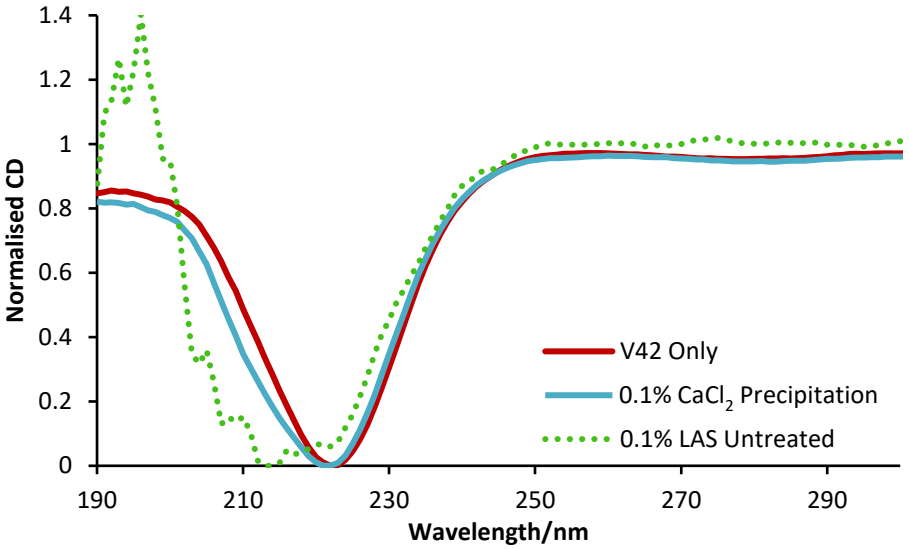


Figure 62: Normalised CD spectra showing V42 incubated at 20 °C for the control sample, 0.1% LAS in situ and 0.1% LAS, with surfactant $CaCl_2$ precipitation.

4.5.2 Validation of Melting Curves produced from Surfactant Removed Samples by Circular Dichroism

The second stage of validation of this purification method ensured that refolding was negligible, or at least constant, across all temperature ranges. Inconsistent levels of refolding following incubation and cooling would prevent the collection of coherent data required to produce a melting curve. This validation was conducted by incubating both V42 and Everest samples at temperature points between 20 °C and 100 °C at 5 °C intervals, before removing surfactant and measuring each CD spectrum. These individual spectra were then combined to generate a melting curve which was compared with melting curves obtained at 222 nm through *in situ* analysis of the 0.1% LAS sample, as illustrated in *Figure 63*. The cumulative data are presented in *Table 14*.

A degree of variance between the two methods were anticipated due to the detector interference of surfactant present in the *in-situ* sample, as noted previously *(Chapter 3.2.9)*. When compared to *in-situ* CD analysis, the surfactant removed curve showed similar levels of data scatter from the sigmoid curve. There was, however, a difference in the average T_m value of ~4 °C. This increase was constant across three independent repeat

experiments at 0.1% LAS. This has been attributed to a degree of refolding occurring at temperatures approaching the T_m. As the level of refolding is consistent for this enzyme, this can be incorporated into any models developed based on this method.

Table 14: Experimental T_m values for the enzyme V42 in 0.1% LAS as determined by CD by *in situ* analysis and following surfactant removal via $CaCl_2$ precipitation.[a,b,c]

Analysis	T_m (°C)	Error in T_m (°C)
In Situ 1	56.2	± *1.1*
In Situ 2	59.2	± *1.6*
In Situ 3	58.7	± *1.9*
Average *In-Situ*[d]	58.0	± *0.8*
LAS Removal 1	62.6	± *0.9*
LAS Removal 2	61.7	± *1.6*
LAS Removal 3	61.5	± *1.1*
Average LAS Removal[d]	61.9	± *0.5*

[a]Analysis conducted in the presence of PMSF. [b]Reported T_m values are an average of three independent analyses. [c]Error in T_m is calculated based on the standard error in T_m from fitting data to a sigmoid. [d]Error in average T_m is reported as the standard error from the mean.

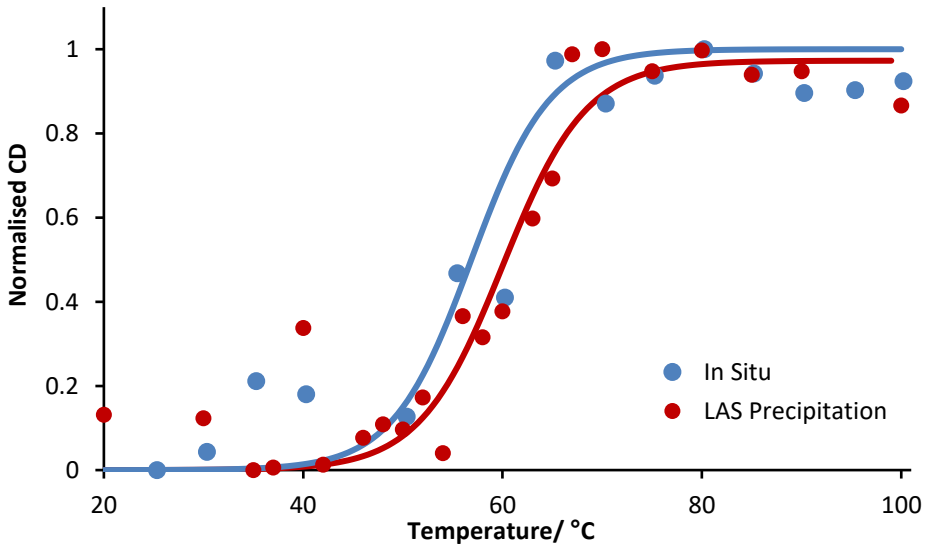


Figure 63: Comparison of melting curves of V42 in 0.1% LAS obtained *in situ* (blue) and through construction from individual temperature measurements following surfactant removal (red).

This validation was repeated for the amylase, Everest at 0.1% LAS. In this case, however, analysis was hindered by the high stability of the enzyme.

Based on observations using nano-DSC, T_m values were expected to be higher in the presence of 0.1% LAS than those of the enzyme in buffer only (~87 °C). As Everest was stable at temperatures above 90 °C, sufficient data could not be collected to produce a complete melting curve. As a result, the error in data fitting was too high to conclusively determine T_m values *(Table 15, Figure 64).* Data collected for the samples purified through $CaCl_2$ precipitation did, however, show sufficient curvature to fit to a sigmoid and an estimate of T_m. The error due to scatter from the fit was less than 1 °C and the standard error of two independent analyses was also ~1 °C *(Figure 65).*

Table 15: Experimental T_m values for the enzyme V42 in 0.1% LAS as determined by CD by *in situ* analysis and following surfactant removal via $CaCl_2$ precipitation.[a,b,c]

Analysis	T_m[d]	Error in T_m
In Situ 1	*T_m above instrument range*	
Average	>95.0	-
LAS Removed 1	89.8	± *0.8*
LAS Removed 2	92.0	± *0.7*
Average	90.9	± *1.1*

[a]Analysis conducted in the presence of PMSF. [b]Reported T_m values are an average of three independent analyses. [c]Error in T_m is calculated based on the standard error in T_m from fitting data to a sigmoid. [d]Error in average T_m is reported as the standard error from the mean.

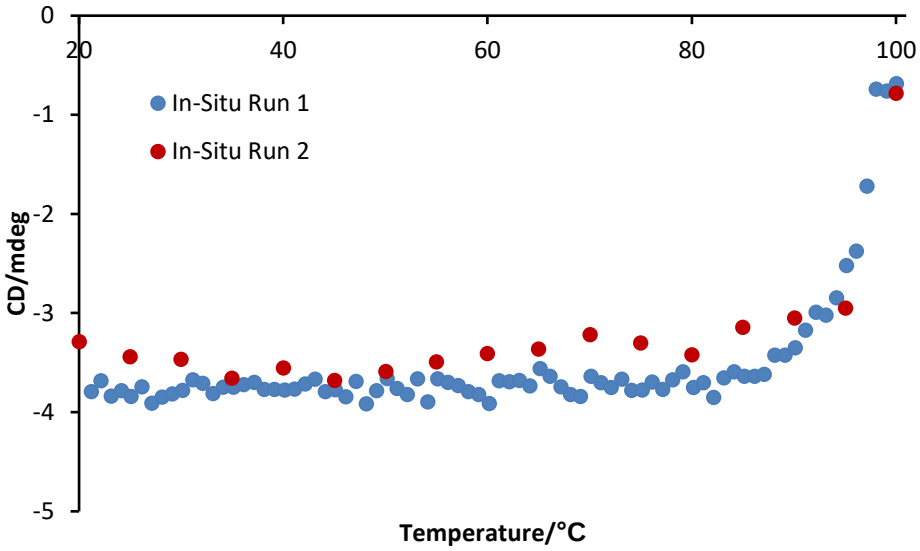


Figure 64: Melting curves for Everest in the presence of 0.1% LAS determined by CD analysis at 222 nm. Incomplete sigmoids at higher end of the temperature range prevent accurate determination of T_m values.

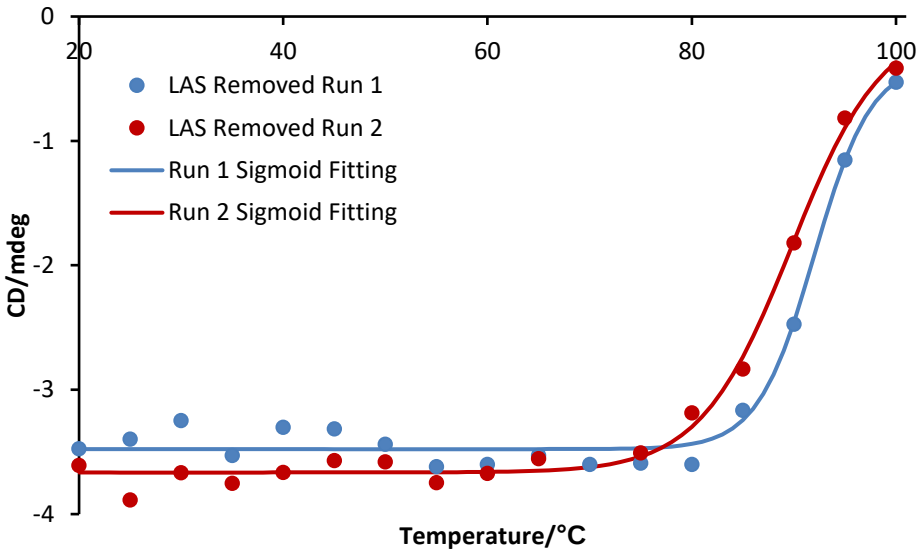


Figure 65: Melting curves for Everest in the presence of 0.1% LAS determined by CD analysis at 222 nm. Surfactant was removed via $CaCl_2$-based precipitation prior to data collection.

These experiments demonstrated that removal of surfactant prior to analysis was a precise method of estimating T_m values in the presence of low surfactant concentrations. The degree of error in curve fitting and the deviation from the mean was equal to or lower than observed when using the instrumental temperature ramp. The small difference in values reported for T_m between the two procedures can be incorporated into the development of stability models due to the consistency of the data. For this

reason, it was decided to proceed with CD analysis across a range of surfactant concentrations using the $CaCl_2$ precipitation method. This ensured the method was robust at high surfactant levels consistent with high surfactant concentrations found in commercial HDL. The resultant data set also provided a means of exploring the link between SDS-based protein interactions and those of LAS.

4.6 T_m analyses of Surfactant Removed Samples by Circular Dichroism

LAS-precipitation by $CaCl_2$, as described above, was applied to a full range of LAS concentrations. Formulations were comparable to SDS samples run previously *(Section 4.2)*. In all cases it was possible to sufficiently purify the sample using $CaCl_2$ precipitation to prevent detector saturation. Melting curves were then constructed from the resultant spectra for all formulations (*Appendix 2*). T_m values obtained from these curves are reported in *Table 16*.

Table 16: Experimental T_m values for the enzymes V42 and Everest in the presence of various concentrations of LAS, as determined by CD following surfactant removal.[a,b]

LAS Concentration	V42 T_m / °C	Everest T_m / °C
Control	63.9 (± *0.7*)	86.7 (± *0.6*)
0.1% LAS	62.0 (± *0.3*)	90.9 (± *1.1*)
1% LAS	57.3 (± *0.9*)[c]	86.9 (± *0.8*)
5% LAS	54.8 (± *1.3*)	79.0 (± *0.6*)
10% LAS	52.9 (± *1.8*)[c]	80.9 (± *5.4*)
20% LAS	52.1 (± *1.9*)[c]	79.4 (± *2.6*)[d]

[a]Reported T_m values are an average of three scans, measured at 222 nm. [b]Error values are based on the standard error of three analyses unless noted. [c]Error values based on the error in the inflection point when fitted to a sigmoid curve as this value was greater. [d]T_m value is the result of single run, error based on scatter from the fit.

4.6.1 Effects of LAS on the Stability of the Protease V42

A decrease in stability with increasing concentrations of LAS was observed for V42. In a similar manner to the SDS experiments, no stabilisation with respect to the control was observed in the concentration range analysed *(Figure 66)*.

As no data for V42 in the presence of LAS were obtained using DSC *(Chapter 3.3),* it was not previously possible to compare the effects of the two surfactants. T_m values for V42 from CD analysis in the presence of both LAS and SDS are listed in *Table 17* below. Comparable effects on the T_m of V42 were observed for both surfactants and a plot of the two datasets yields a linear correlation with an R^2 value of 0.96 *(Figure 67Figure* ***68****).* These findings support the use of SDS as an analog for LAS in studies where surfactant removal is not possible, or to reduce sample preparation and analysis times.

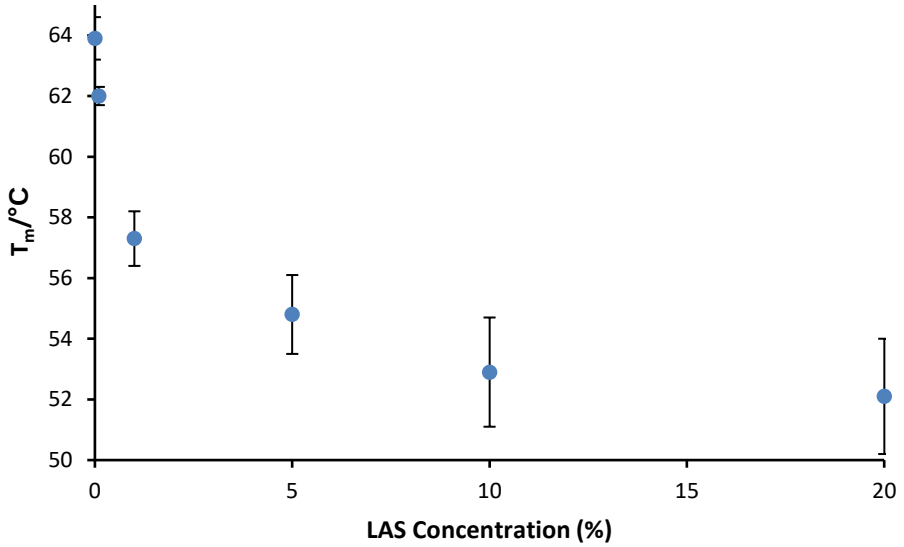


Figure 66: Reduction in T_m of V42 in increasing concentrations of LAS.

Table 17: T_m values for the enzyme V42 in the presence of LAS and SDS as determined by CD.[a,b,c]

LAS Concentration	V42 LAS T_m / °C	V42 SDS T_m / °C
Control	63.9 *(± 0.7)*	63.9 *(± 0.6)*
0.1%	62.0 *(± 0.3)*	62.3 *(± 0.9)*
1%	57.3 *(± 0.9)*[d]	57.4 *(± 1.8)*
5%	54.8 *(± 1.3)*	53.1 *(± 0.3)*
10%	52.9 *(± 1.8)*[d]	51.6 *(± 1.6)*
20%	52.1 *(± 1.9)*[d]	50.5 *(± 1.8)*

[a]T_m values in the presence of LAS based on analysis of samples with LAS removed via precipitation method. [b]T_m values in the presence of LAS collected using in-situ heating an analysis. [c]Error values are listed in brackets and reported as the standard error from the mean of three independent tests unless otherwise noted. [d]Error is the standard deviation from the curve fit as this was a greater source of error

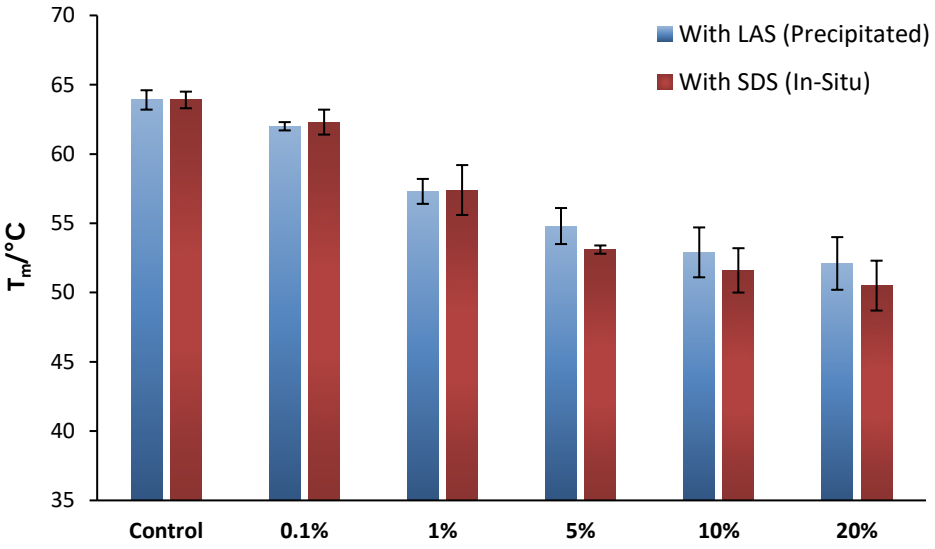


Figure 67: Trends in T_m values for V42 in the presence of LAS (blue) and SDS (red).

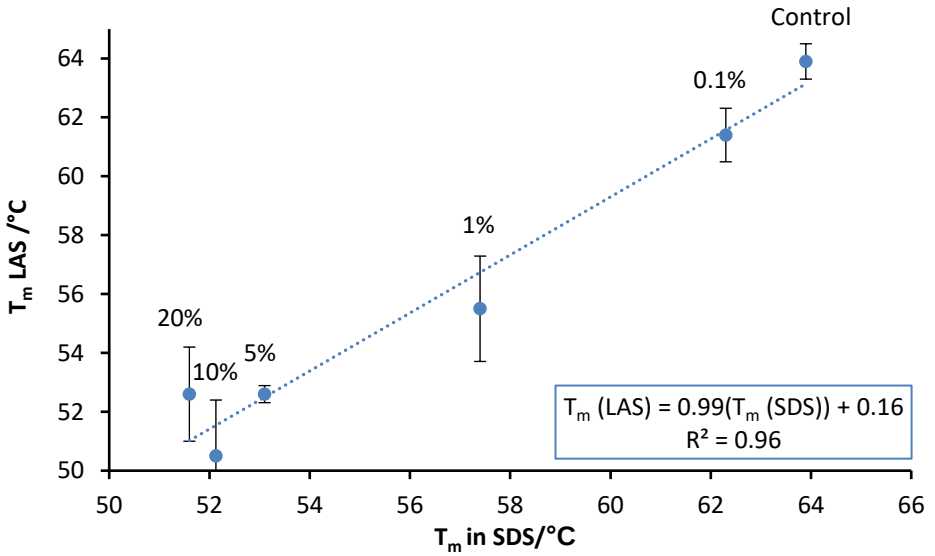


Figure 68: T_m values for V42 in the presence of LAS as a function of T_m values in SDS.

4.6.2 Effects of LAS on the Stability of the Amylase, Everest

A general downward trend in stability with increasing LAS concentration was observed for Everest. Comparison with the buffer only sample, however, showed an initial stabilisation at low surfactant levels (0.1%) of ~4 °C. This was in line with SDS trends *(Section 4.2.2)* and those obtained by nano-DSC for LAS *(Chapter 3)*. At 1% LAS, the T_m value of Everest was approximately equal to that of the buffer only sample at ~87 °C. Samples in 5%, 10% and 20% LAS were all reported to have T_m values of 79-81 °C *(Table 16)*.

It should be noted that Everest samples showed a much higher scatter from the sigmoid fitting than V42, therefore outlying data points were omitted from fittings for estimation of T_m values. A representative example of a melting curve generated at 10% LAS is shown below in *Figure 69,* with omitted outliers included as open circles. Reported values for Everest at 1% and 10% are also an average of 2 runs, as the third did not produce data of sufficient quality to determine T_m. The T_m value for 20% LAS are the result of a single run for the same reason, error has therefore been calculated based on the standard error of the fit. Difficulties in T_m determination for Everest have been experienced across all excipient classes. This was originally thought to be an artefact of high T_m values, approaching the limits of instrument capabilities. At such temperatures, evaporation of solvent may also affect results. To clarify these theories, and ensure other factors need not be considered, this work should be repeated with an amylase of lower stability, such as Natalase. Alternatively, nano-DSC using pressurised chambers may be more appropriate for enzymes which are exceptionally thermally stable.

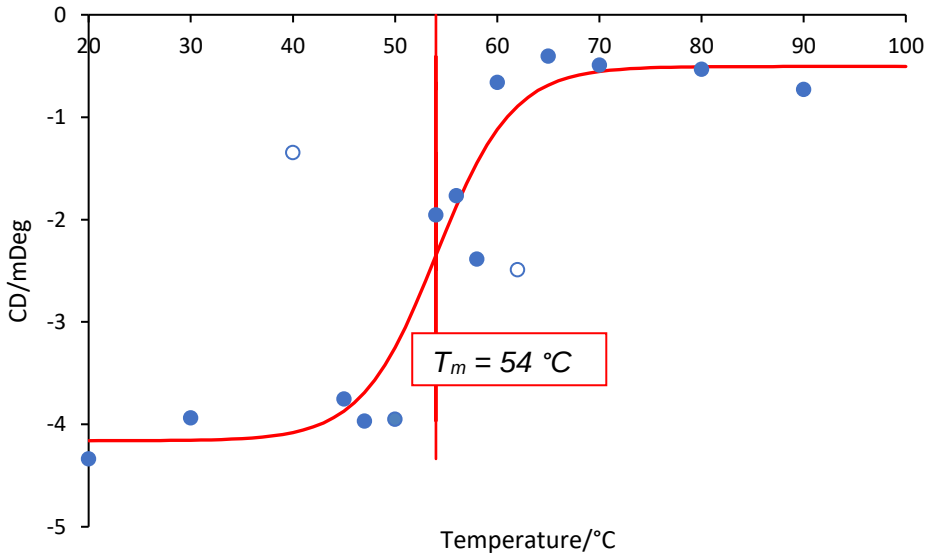


Figure 69: Representative melting curve of a single analysis run for Everest in 10% LAS, showing a Tm values of 54 °C. Points omitted from sigmoid fitting are highlighted using open circles.

4.6.3 The relationship between LAS and SDS-Induced Unfolding of Everest

In a similar manner to V42, T_m values for Everest in the presence of LAS were compared to those of SDS *(*

Table **18***)*. Although errors in T_m values were higher for the amylase samples, similar trends in the observed effects of LAS and SDS were evident, as demonstrated in *Figure 70*. A plot of these data yields a linear correlation between the two surfactants with an R^2 value of 0.97 *(Figure 71)*.

Table 18: T_m values for the enzyme Everest in the presence of LAS and SDS as determined by CD.[a,b]

Concentration	T_m / °C (LAS)	T_m / °C (SDS)
Control	86.7 *(± 0.6)*	86.7 *(± 0.6)*
0.1%	90.9 *(± 1.1)*	89.9 *(± 2.3)*
1%	86.9 *(± 0.8)*	87.7 *(± 1.5)*
5%	79.0 *(± 0.6)*	83.8 *(± 1.2)*
10%	80.9 *(± 5.4)*	84.6 *(± 1.2)*

[a]T_m values in the presence of LAS determined from melting curves following sample surfactant precipitation. [b]Values in brackets are error estimates based on standard deviation from the mean.

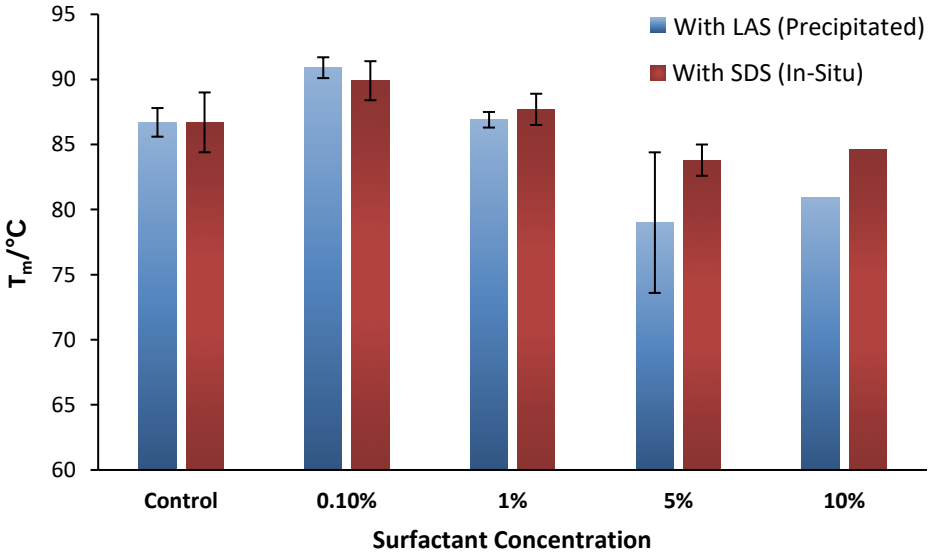


Figure 70: Trends in T_m values for Everest in the presence of LAS (blue) and SDS (red).

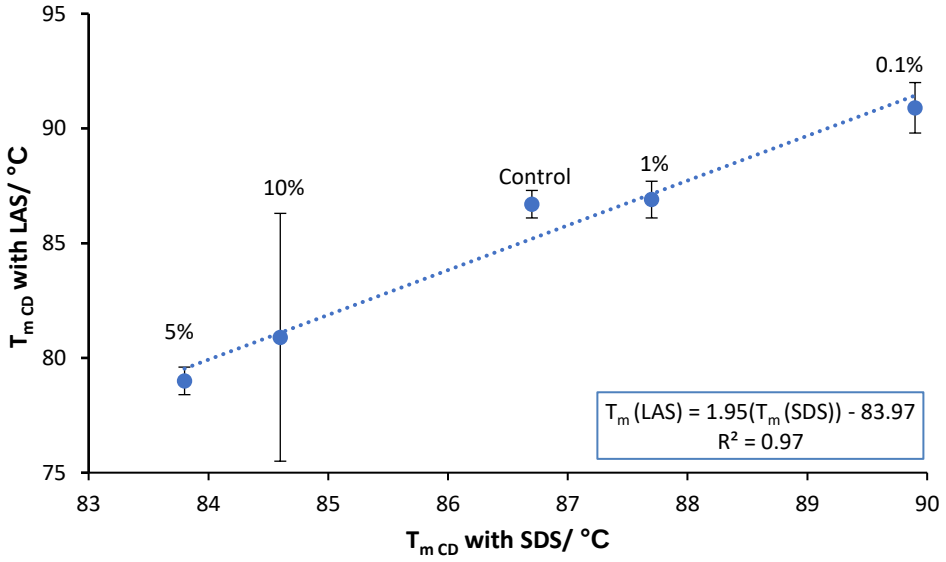


Figure 71: T_m values for Everest in the presence of LAS as a function of T_m values in SDS.

As demonstrated in *Figure 72,* a linear correlation was also observed between melting temperatures obtained from CD analysis ($T_{m\ CD}$), and those from DSC ($T_{max\ DSC}$). The R^2 value for this line was 0.85, with DSC values tending to be 7-10 °C higher than those of CD. $T_{max\ DSC}$ has been shown by Lund *et al* to be predictive of storage stability. The direct correlation with $T_{m\ CD}$ values for both SDS mock formulations and surfactant precipitation methods indicates that either could fulfil our requirements for stability modelling. This will be verified in the following chapter.

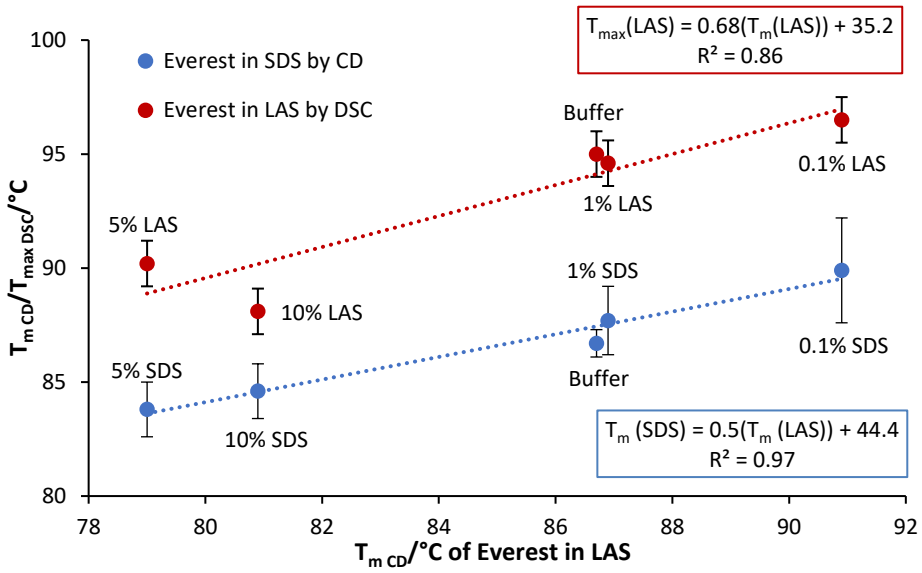


Figure 72: Comparison of data obtained for Everest in LAS at various surfactant concentrations using DSC and for Everest in SDS using CD with those of Everest in LAS using CD.

4.7 The effects of LAS and SDS on observed Structural Features in Detergent Proteins.

Empirical T_m data indicated that SDS acts as a valid analog for LAS in the determination of thermal stability. The use of CD presented an opportunity to further explore this relationship through deconvolution of CD spectra.

Overlaying the CD spectrum of V42 at 20 °C with those collected in the presence of surfactant indicates variation between both the control and each surfactant *(Figure 73Figure 74)*. Differences are more pronounced in the case of Everest, with LAS inducing a far greater effect on protein structure than SDS, when compared to the control. Further analysis was necessary to determine the significance of these effects and was conducted using Dichroweb deconvolution software and PCA.[20]

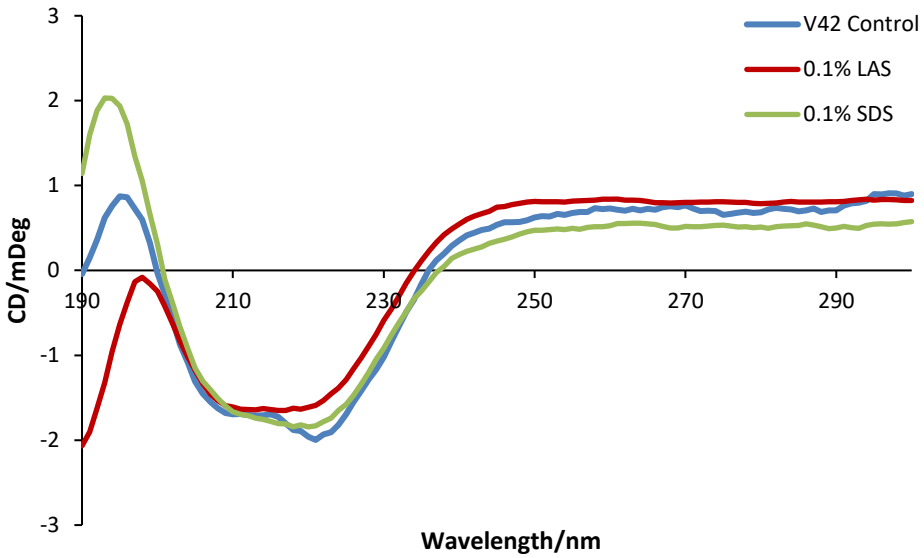


Figure 73: V42 CD spectra under control conditions and in the presence of 0.1% SDS and 0.1% LAS.

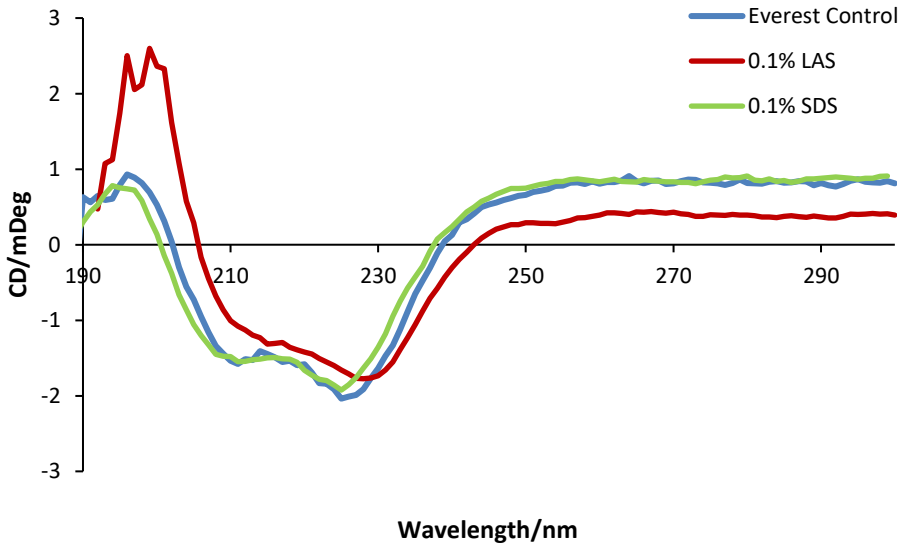


Figure 74: Everest CD spectra under control conditions and in the presence of 0.1% SDS and 0.1% LAS.

4.7.1 Dichroweb Structural Analysis

V42 was chosen as a representative example as these data, obtained through $CaCl_2$ precipitation of LAS, had been previously observed to be of higher quality than Everest. Several deconvolution programmes are available in the Dichroweb library, each tailored to a specific type of protein. Both CONTIN and CDSSTR methods were applied to the spectra to determine which best evaluated protein structure. CDSSTR is best suited to proteins with a high helical content, while CONTIN provides

better accuracy in proteins with more β-sheet character. Both CONTIN and CDSSTR methods were applied to the spectra to determine which programme was most appropriate to the data, as the prevalence of each structural feature was not known.

As illustrated in *Figure 75-Figure* 76 below, CDSSTR gave a lower standard error between repeat scans than CONTIN. Little variation with increasing concentrations of surfactant was reported by the programme, however, with all samples estimated to have ~20% helical content. This is contrary to the loss in CD intensity evident from the spectral overlay, shown in *Figure 77*. As a result, the CONTIN method, which identified greater differences in structural features between spectra, was selected for further structural determination.

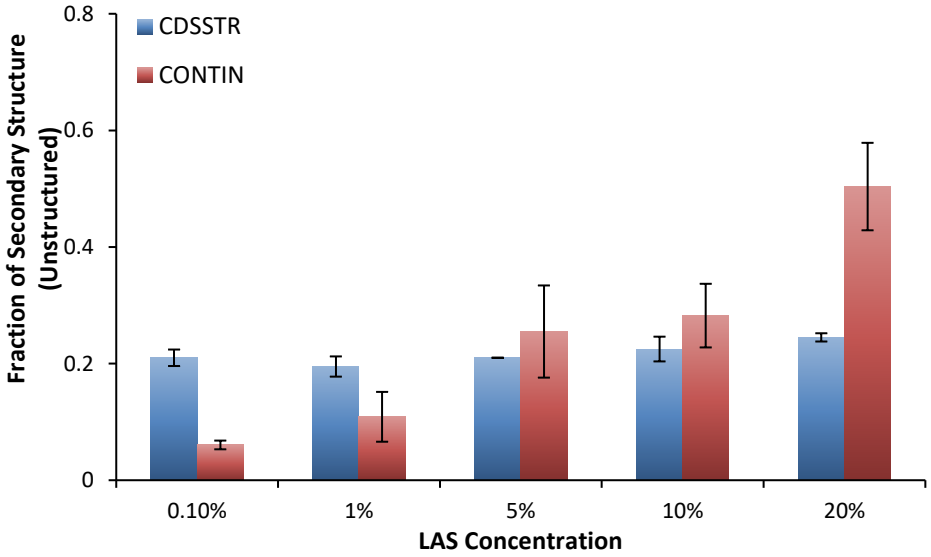


Figure 75: Proportion of V42 in unstructured state as determined by CDSSTR and CONTIN deconvolution methods in increasing concentrations of LAS.

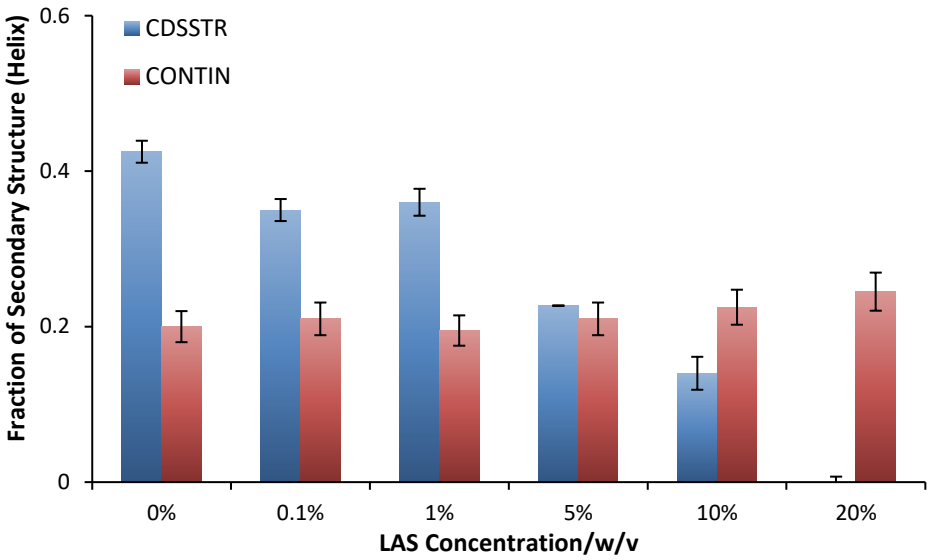


Figure 76: Proportion of V42 in helical state as determined by CDSSTR and CONTIN deconvolution methods in increasing concentrations of LAS.

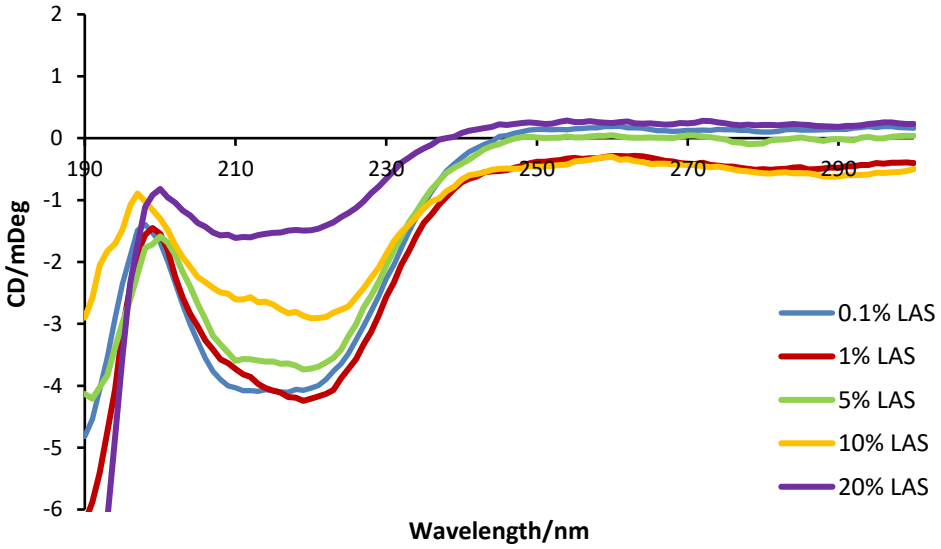


Figure 77: V42 spectra showing loss of protein structure with increasing concentrations of LAS for the enzyme V42 at 20 °C.

The figures above demonstrate the impact that the choice of deconvolution software has on structural determination. Reports in the literature also question the accuracy of these programmes, particularly in identifying non-helical structures or partially denatured proteins.[21,22] Furthermore, bound surfactant may be causing interference with the spectra, in spite of purification.

Analysis of V42 by CONTIN reported approximately 90% helical content, which is not in line with existing structural knowledge of other subtilisin based proteases. These consist of, on average, 30% helical structures and 20% β-sheets.[23] For this reason, definitive conclusions on protein structure could not be drawn. The software still served as a tool for probing relative loss of protein structure in changing environments, however, enabling comparisons to be drawn between unfolding processes in LAS and those in SDS. The description of relative loss in structure by CONTIN was confirmed by plotting estimated proportions of disordered protein against V42 CD intensity at 222 nm This yielded linear correlations for each surfactant, indicating estimations of relative levels of unfolding were in line with genuine trends *(Figure 78)*.

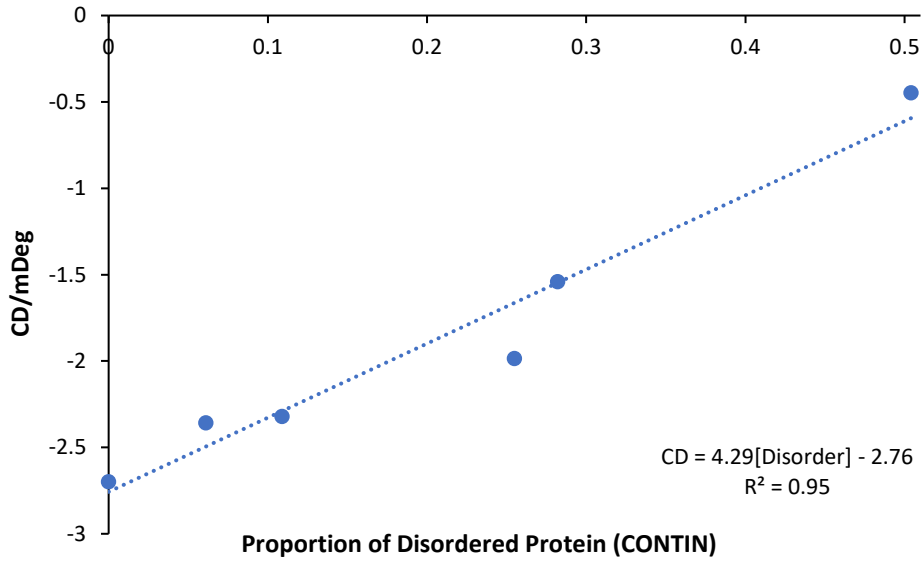


Figure 78: Plot of CD intensity against the proportion of disordered protein calculated by CONTIN structural analysis for V42 in various concentrations of LAS (0.1%-10%); showing a linear relationship.

4.7.2 The effects of LAS on protein structure as determined using the CONTIN method.

Table 19 lists the ratios of α-helix, β-sheet, random turns and disordered protein observed for V42 under a range of surfactant conditions. In the nil-detergent, control sample, the protein spectrum was found to primarily consist of helices, which accounted for over 90% of the structure. Although

this may not be representative of the genuine prevalence of helices, consistent trends were observed across the sample sets, enabling comparison of relative values. In the following section, spectra will be discussed with reference to increasing degrees of disorder which indicate loss of overall structure, due to the uncertainty in assigning proportions of specific structures.

Structural disorder increased on titration of LAS, with the 0% disorder reported in the V42 control increasing to over 50% at 20% LAS. A linear relationship was established between LAS concentration and the amount of disorder *(Figure 79).* This indicates that even in the absence of long-term storage or increased temperature, surfactant binding induces destabilisation.

The trend is also in line with the linear decrease in T_m values with increasing concentrations of LAS, observed during thermal analysis *(Section 4.6.1).* The plot of T_m values as a function of observed protein disorder, shown in *Figure 80,* shows that the two parameters are inversely proportional. This demonstrates that early indications of instability at low temperatures can be predictive of reduced T_m values. This provides a measure of instability with potential application in rapid screening of surfactant conditions.

Table 19: Structural analysis of V42 in the presence of using the CONTIN method of deconvolution.[a]

LAS	Helix1	Helix2	Strand1	Strand2	Turns	Disordered
0%	0.43	0.52	0.00	0.00	0.05	0.00
0.1%	0.35	0.52	0.00	0.00	0.07	0.06
1%	0.36	0.44	0.00	0.00	0.09	0.11
5%	0.23	0.38	0.00	0.00	0.05	0.26
10%	0.14	0.30	0.09	0.05	0.10	0.28
20%	0.00	0.44	0.00	0.00	0.05	0.50

[a]*Via Dichroweb website, values are listed as the fraction of protein secondary structure they represent. Helix 1, Helix 2, Strand 1 and Strand 2 indicate a regular α-helix, distorted α-helix, regular β-strand and distorted β-strand respectively.*

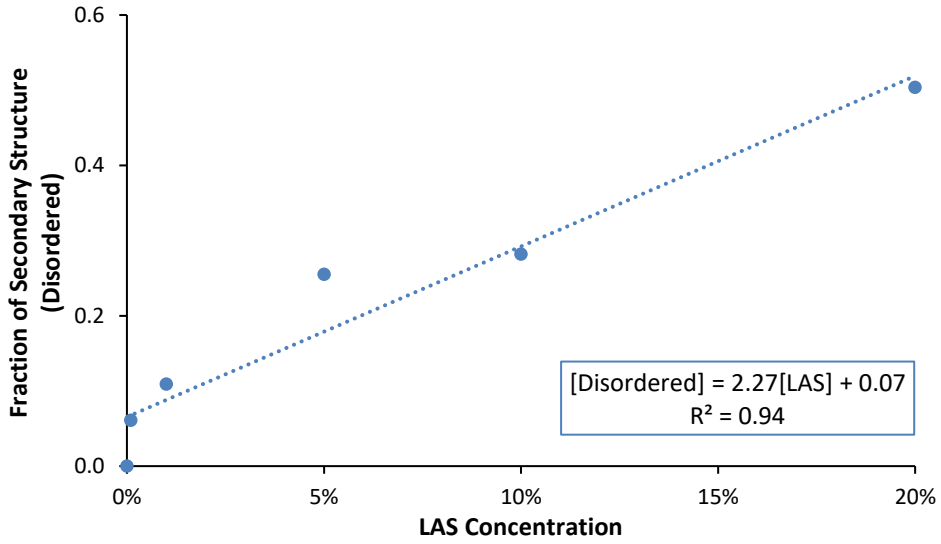


Figure 79: Proportion of disordered protein determined by CONTIN deconvolution for V42 in the presence of varying concentrations of LAS.

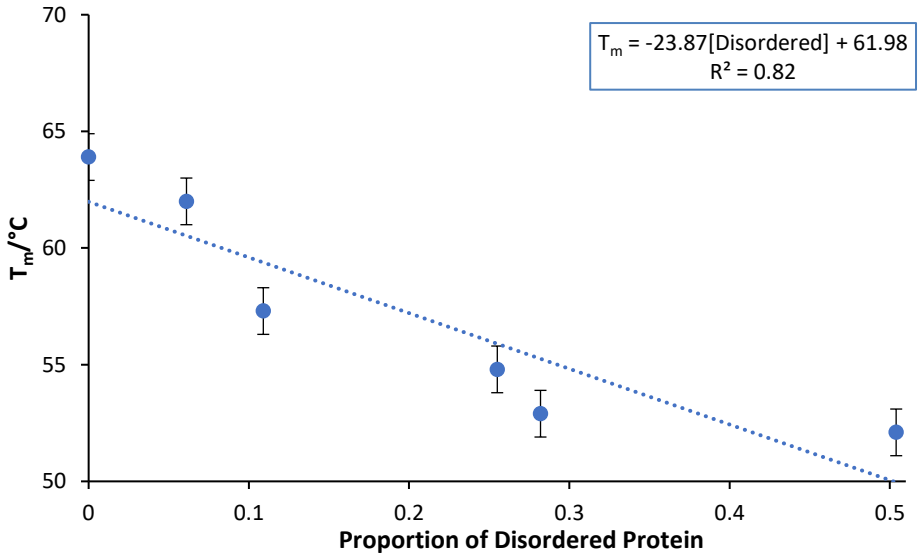


Figure 80: The relationship between the proportion of disordered protein estimated for CD spectra by CONTIN deconvolution and subsequent T_m values from thermal denaturation.

4.7.3 The Effects of SDS on Protein Structure as determined using the CONTIN Method

A similar trend was observed among SDS samples; a linear correlation between increasing surfactant concentration and the ratio of disordered to ordered protein *(Table 20, Figure 81)*. Here, the error between repeat spectra for the same condition is significantly higher than those in th presence of LAS. This was attributed to the presence of surfactant during

analysis, which may cause interference with the spectra, affecting comparisons with existing reference sets. As LAS samples were purified of surfactant prior to collection of CD spectra, the effect was reduced.

Table 20: Structural analysis of V42 in the presence of SDS using the CONTIN method of deconvolution[a].

SDS	Helix1	Helix2	Strand1	Strand2	Turns	Disordered
0%	0.43	0.52	0.00	0.00	0.05	0.00
0.1%	0.19	0.17	0.01	0.05	0.12	0.38
1%	0.25	0.22	0.00	0.03	0.08	0.43
2.5%	0.19	0.19	0.01	0.02	0.16	0.43
5%	0.23	0.22	0.00	0.03	0.13	0.40
7.5%	0.19	0.19	0.00	0.01	0.14	0.47
10%	0.11	0.13	0.00	0.03	0.13	0.60
20%	0.17	0.17	0.03	0.06	0.21	0.37

[a]*Via Dichroweb website, values are listed as the fraction of protein secondary structure they represent. Helix 1, Helix 2, Strand 1 and Strand 2 indicate a regular α-helix, distorted α-helix, regular β-strand and distorted β-strand respectively.*

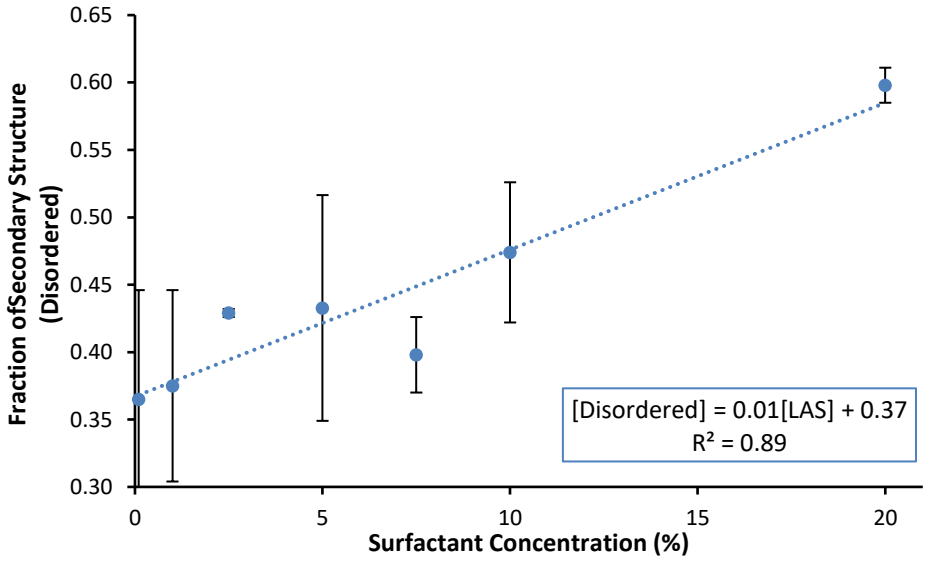


Figure 81: Proportion of V42 protein structure in unfolded state in the presence of varying concentrations of LAS as determined by CONTIN.

4.7.4 Comparison of the Structural Effects of LAS and SDS Binding on V42

Trends in protein destabilisation caused by LAS and SDS for V42 were compared to explore the similarities between the two processes *(Table 21)*. Analysis by CONTIN showed both exhibited a linear increase between the

proportions of protein disorder and increasing surfactant concertation. The level of disorder was found to be higher, however, in the presence of SDS than in LAS. This difference is most significant at low concentrations, with the two samples converging towards 20% surfactant *(Figure 82)* A plot of the relative disorder induced by each surfactant at equivalent concentrations yielded a linear relationship with an R^2 of 0.97 *(Figure 83).* This suggests that SDS could be incorporated into a model of protein stability, in place of LAS, to improve ease of analysis of commercial formulations. Further work, however, may be necessary to account for the observed differences in induced disorder, as this was not reflected in T_m values. This will be explored further in the following chapter in light of enzyme activity data for each of these conditions.

Table 21: Proportion of disorder in V42 at 20 °C in the presence of various concentrations of SDS and LAS, determined by CONTIN analysis.[a]

Surfactant Concentration	LAS Disordered	SDS Disordered	Difference
0%	0.00	0.00	0.00
0.1%	0.06	0.37	0.30
1%	0.11	0.38	0.27
5%	0.26	0.43	0.18
10%	0.32	0.47	0.16
20%	0.50	0.60	0.09

[a]*Via Dichroweb website, values are listed as the fraction of protein secondary structure they represent*.

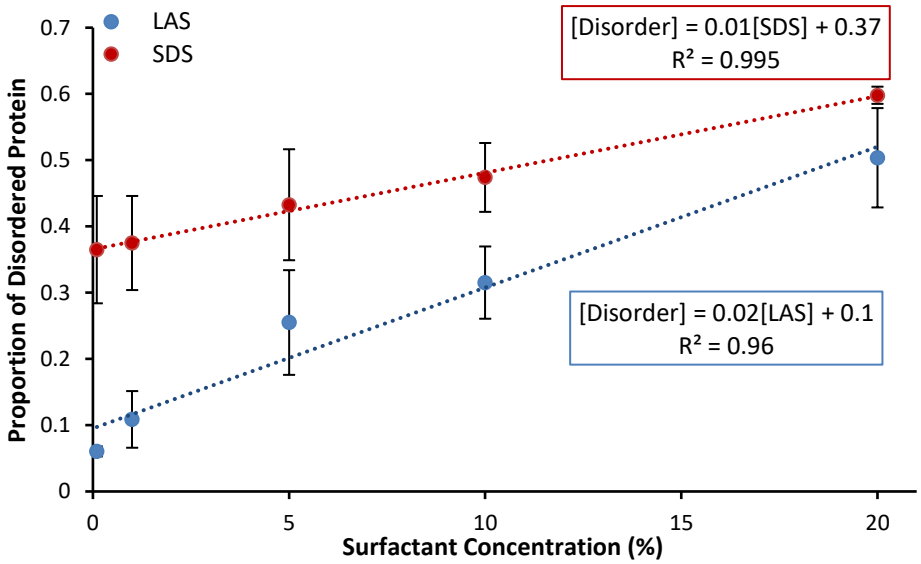


Figure 82: Proportions of disordered protein for V42 in varying concentrations of LAS and SDS as determined using CONTIN deconvolution.

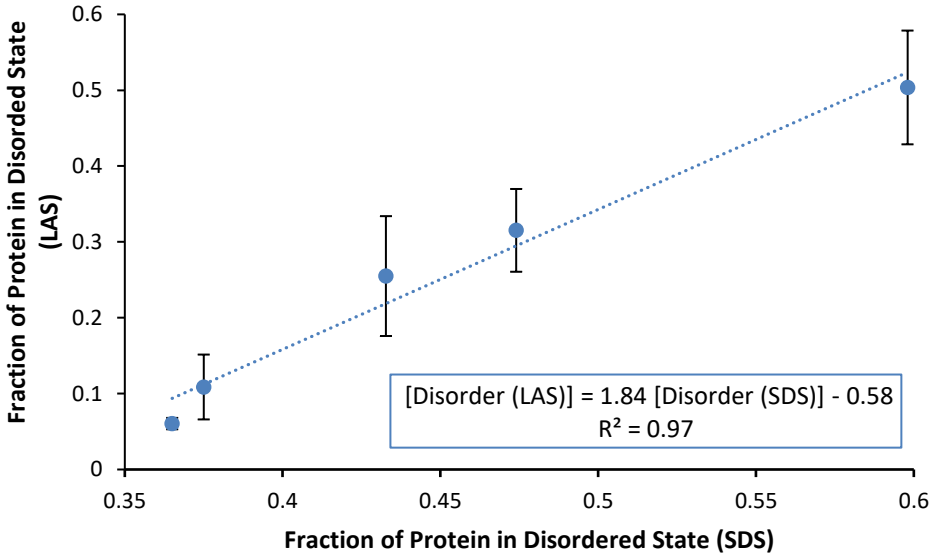


Figure 83: The degree of disorder in V42 induced by LAS as a function of that induced by SDS.

As structural estimations for V42 were not in line with those common to subtilisin-based proteases, it was necessary to determine if relative proportions of disorder, described above, could be related to genuine trends in protein unfolding. This was achieved though comparison with standardised CD intensity at 222 nm, the wavelength generally used for monitoring loss of secondary structure. These values were plotted against surfactant concentration, shown below in *Figure 84*. LAS exhibited similar trends to those determined by CONTIN, however SDS samples, were found

to have little variation with titration of surfactant at 20 °C. As both surfactants yielded similar T_m values at equivalent concentrations, initial destabilisation may not be indicative of thermal stability, as suggested above. The variation also questions the accuracy of absolute T_m values of LAS samples collected following $CaCl_2$ precipitation, when compared to *in situ* values.

A linear correlation between CD intensity at 222 nm, and protein disorder estimated by CONTIN deconvolution, is evident from data in *Figure 85* for V42 in LAS. This indicates that observed trends are, indeed, reflective of genuine unfolding effects. In contrast, trends in SDS did not reflect results from CONTIN analysis. As discussed previously, this may be a result of detector interference preventing spectra from being accurately related to reference proteins. These data suggest that the CDSSTR method of protein analysis, which estimated similar levels of disorder for each surfactant concentration, may be more appropriate. To resolve these uncertainties, the above analysis should be repeated using a similar protein with available crystal structures such as subtilisin Carlsberg. This would enable validation of such trends and identify the most suitable deconvolution methods.

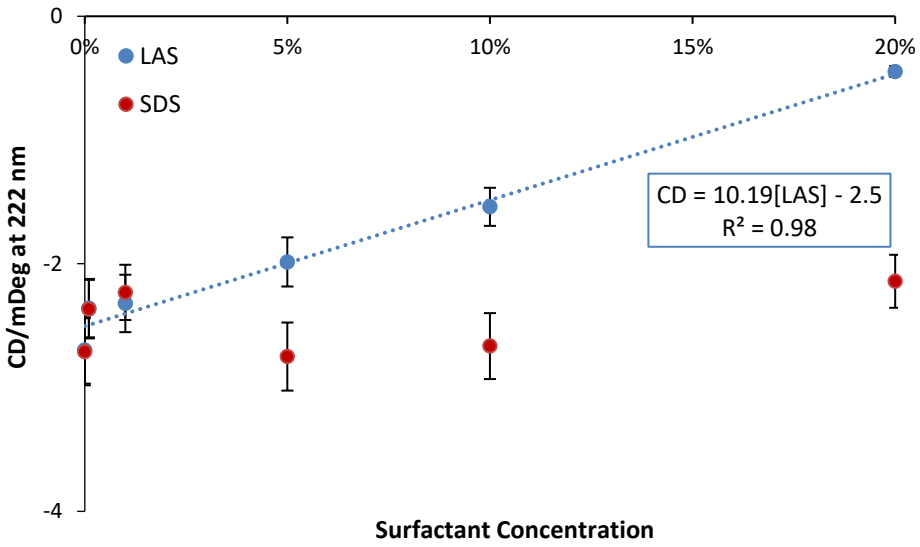


Figure 84: CD intensity at 222 nm for V42 in the presence of varying concentrations of LAS and SDS.

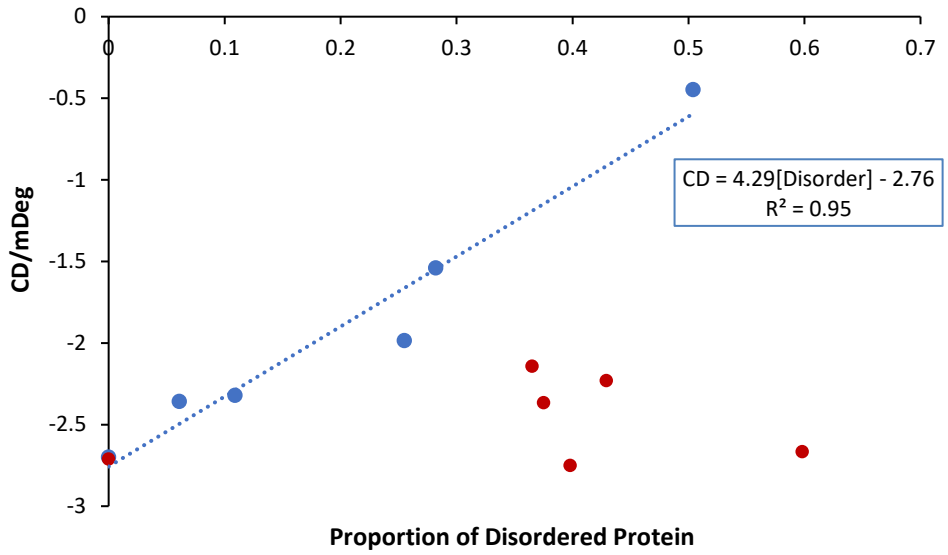


Figure 85: CD intensity at 222 nm as a function of the proportion of disordered protein

In lieu of supporting crystal structures, principal component analysis was applied to the data in an attempt to conclusively define the relationship between LAS and SDS induced unfolding. Such analyses are incorporated into deconvolution programmes, but independence from reference spectra should avoid issues of deviations as a result of ligand binding.

4.8 Principal Component Analysis of V42 Spectra in LAS and SDS

Principal components were assigned to the spectra used in the above deconvolution tests. A plot of the first principal component (PC1), as a function of surfactant concentration yielded a linear relationship for V42 in both LAS and SDS sample sets *(Figure 86)*. In this case, PC1 accounts for 50% of structural variation. Samples in SDS fit better to the trendline than those of LAS with R^2 values of 0.87 compared to 0.68. The lower values for LAS are a product of the significant change in protein structure observed between the nil detergent control and the first concentration point (0.1% LAS). Exclusion of the control increases the linear correlation to approximately that of SDS. Changes in PC1 values have been attributed to unfolding, in accordance with observations from previous analyses.

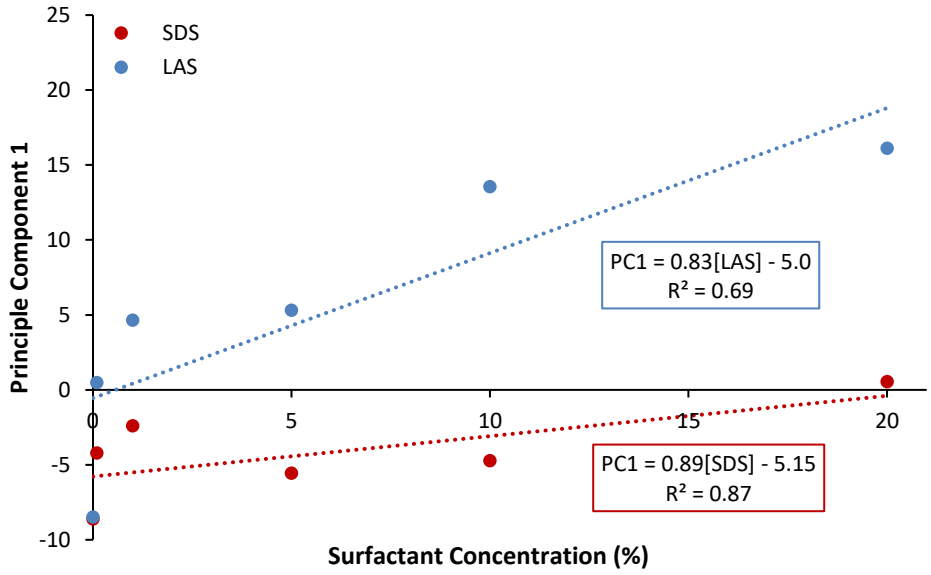


Figure 86: Plot of PC1 as a function of surfactant concentration for V42 at 20 °C in various concentrations of LAS and SDS.

Differences between structural effects induced by SDS and LAS were confirmed with the scoreplot of PC2 and PC1, shown in *Figure 87*. A clear distinction between spectra arising from the presence of each surfactant can be seen from clustering in the data. Samples in various concentrations of SDS fall closer to the nil-detergent control than those in LAS, indicating higher levels of stability. This does not reflect significant increases in unstructured protein with SDS titration which was reported by CONTIN programme *(Figure 88),* but is in line with CD intensity values at 222 nm *(Figure 84)*. Principal components for LAS, in contrast, align with both CONTIN and CD intensity trends, supporting the theory that the presence of SDS is interfering with the application of reference spectra to the data. Removal of surfactant prior to collection of CD spectra appears to reduce this effect, as seen here for samples in LAS, which yielded results in line with comparisons drawn from other methods.

As PCA does not provide estimations of structural features, the discrepancy between estimations for the detergent protease and known properties of common subtilisins cannot be explored further. Deviations may be a result of the limitations of deconvolution software in assigning

partially unfolded proteins, which has been discussed in the literature.[20,24,25]

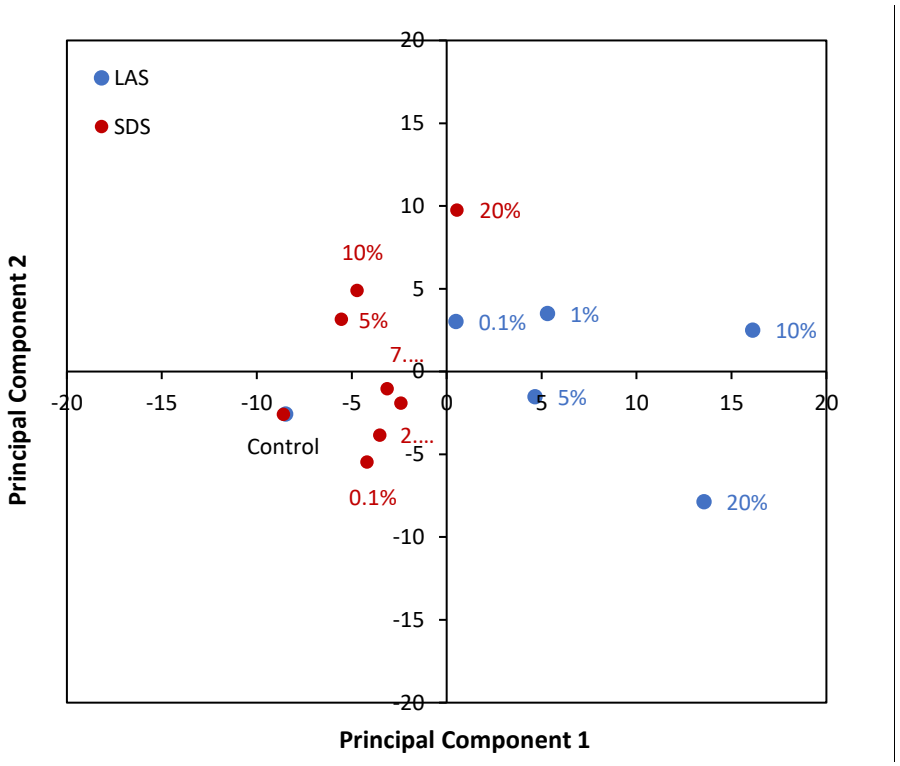


Figure 87: Scoreplot for V42 spectra in the presence of various concentrations of LAS and SDS.

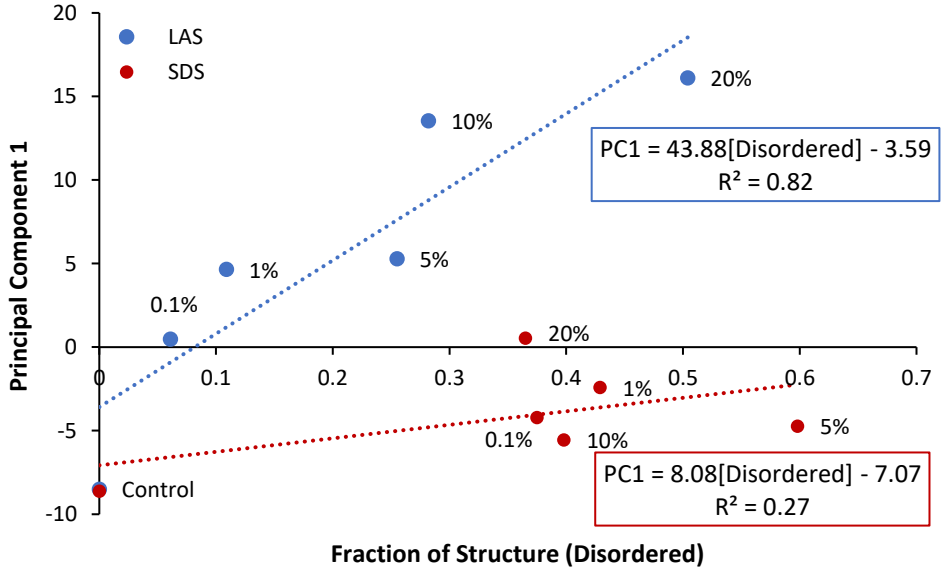


Figure 88: PC1 of V42 spectra at 20 °C for various concentrations of LAS and SDS as a function of respective proportions of disordered protein structure, estimated using the CONTIN deconvolution programme.

4.8.1 Monitoring of Thermal Denaturation using PCA.

In order to monitor changes in protein structure with temperature, PCA was applied to CD spectra collected at 5 °C intervals for V42 under control conditions. This had not previously been possible using deconvolution software due to the limited availability of partially unfolded protein in the database. PC1 values for each spectrum were plotted against respective temperatures to produce a melting curve. In *Figure 89* this curve has been compared with the standard melting curve of CD intensity at 222 nm, used in determining T_m. Similarities between the two indicate that PC1 values are a viable measure of denaturation. This also demonstrates how overall loss of structure can be simplified to CD intensity at 222 nm for straightforward calculations.

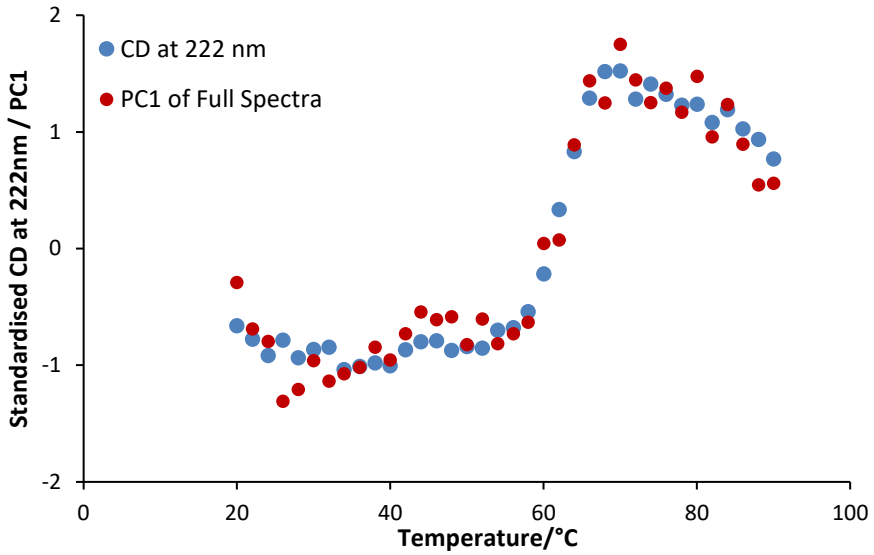


Figure 89: Comparison of the V42 melting curve under control conditions, constructed from CD intensity data at 222 nm with that of PC1 values.

To compare unfolding of V42 under control conditions to that in presence of surfactant, a scoreplot, shown in *Figure 90*, was prepared from thermal denaturation spectra. This consisted of PC1 values as a function of PC2 values for V42 samples under control condtions, 0.1% LAS, and 0.1% SDS. Spectra were collected at the same temperature intervals as in the melting curve above, with the gradient running from 20-100 °C. The resultant

scoreplot facilitated comparison of structures at various temperatures for each condition.

Figure 91 demonstrates similarities in spectra of proximal points on the scoreplot. Highlighted points represent two clusters of samples indicating similar structures. The respective spectra displayed in the inset exhibit the same clustered effect. Use of scoreplots simplifies identification of conditions producing similar conformational changes. For example, V42 in 0.1% LAS at 20 °C shows similar structural features to the V42 control and 0.1% SDS at 50 °C and 60 °C respectively, indicating that initial unfolding process occur at far lower temperatures in LAS.

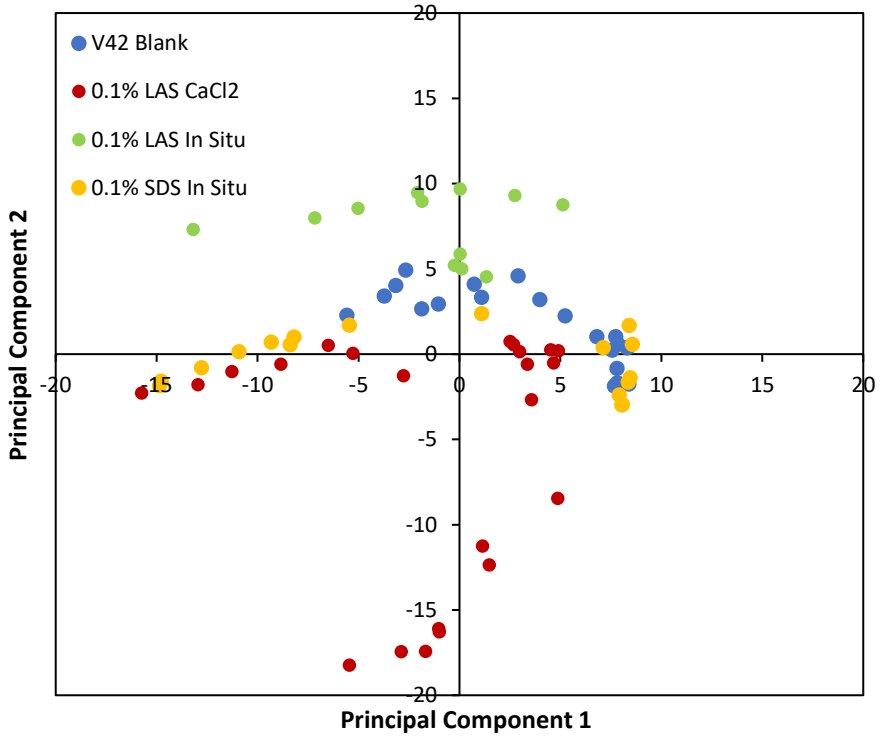


Figure 90: Scoreplot of PC1 and PC2 for V42 CD spectra collected under various surfactant conditions, control (blue), 0.1% LAS from $CaCl_2$ precipitation (red), 0.1% SDS in situ (green) and 0.1% LAS in situ (yellow).

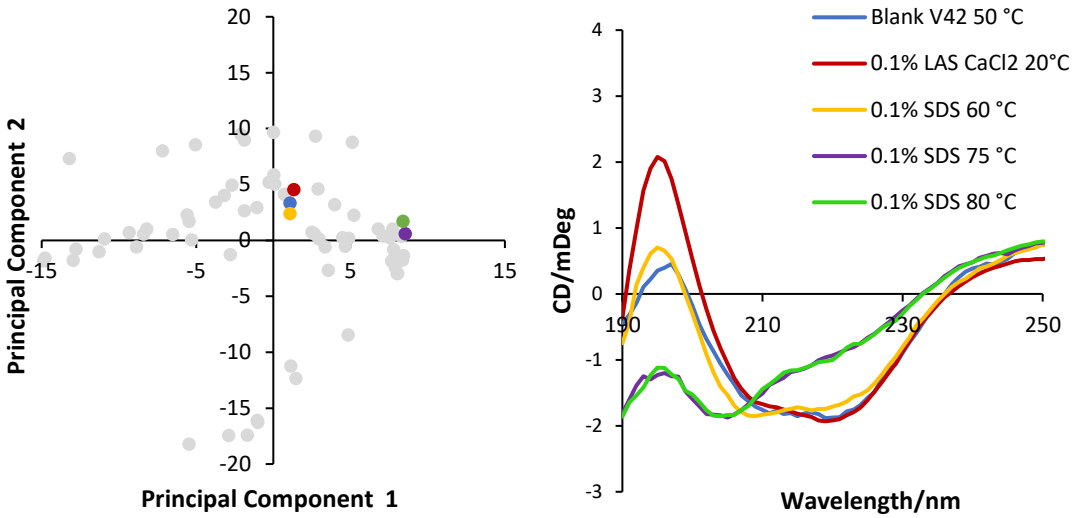


Figure 91: Examples of CD spectra for proximal points on the scoreplot of PC1 and PC2 for thermal denaturation of V42 under various surfactant conditions. Given spectra are represented by the points highlighted in blue (V42 control at 50 °C), red (0.1% LAS $CaCl_2$ precipitation °C) and yellow (0.1% SDS 60 °C), green (0.1% SDS 70 °C) and purple (0.1% SDS 80 °C).

Points representing the spectra of V42 in LAS and SDS appeared close proximity to one another on the scoreplot. This shows similar conformational changes occurring with rising temperatures in the presence of each surfactant, supporting the use of SDS as an analog for LAS in stability studies. The structural similarities described by principal components are evident up to ~60 °C. After this point the two clusters diverge, with spectra for LAS shifted significantly along the PC2 axis. As this deviation occurs at approximately the T_m of V42 in each of these conditions (~62 °C), the effect is likely an artefact of the $CaCl_2$ purification process. Removal of the $Ca(LAS)_2$ precipitate involves a centrifugation step, which may also draw any precipitated denatured protein into the pellet. This would affect observed CD spectra, resulting in the variation highlighted by PCA. The result of removal of precipitated protein from solution is illustrated in *Figure 92,* for spectra of V42 at 100 °C for each condition. Protein signals arising from the presence of unstructured protein, evident up to ~230 nm, are absent in the purified LAS sample.

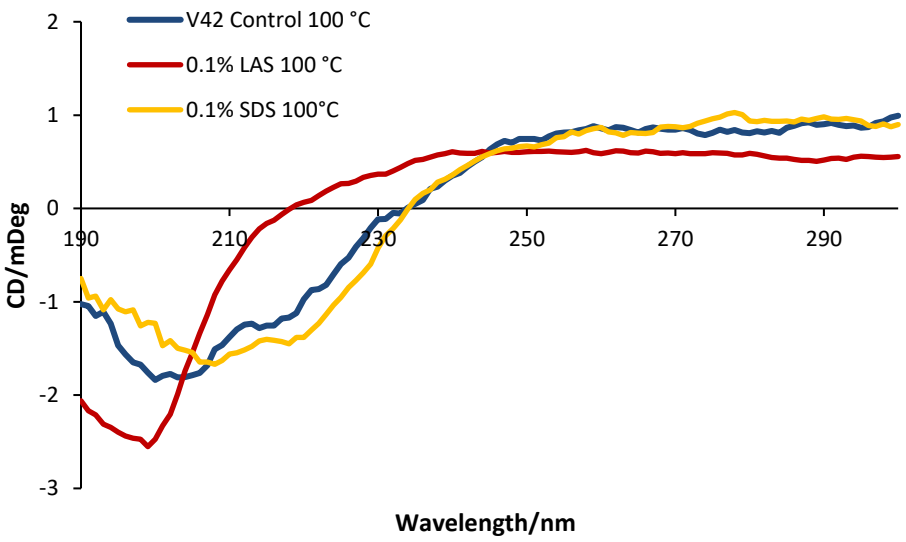


Figure 92: CD spectra of V42 in various detergent conditions at 100 °C.

Comparison of V42 thermal denaturation was conducted by PCA for each surfactant concentration. These yielded similar results to those described at 0.1% w/v, with close resemblance between LAS and SDS data up to the point of protein precipitation. These structural similarities suggest that mechanisms of SDS induced unfolding, summarised at the beginning of this chapter, can be applied to protein-LAS interactions. Furthermore, T_m determination of LAS-rich formulations can be conducted using SDS mock formulations, to reduce laborious sample preparation and analysis. In addition to the improvements in efficiency offered by this approach, artefacts of the purification process, such as removal of denatured protein during centrifugation, are avoided. As the primary use for these T_m values is the generation of procedures for the prediction of long-term stability, both SDS and LAS formulations will be included in subsequent storage tests. These will be described in the following chapter. The final validation of SDS mock-formulations will involve determining if the analogous nature of LAS and SDS on protein thermal stability observed in this chapter, is reproduced in storage stability.

4.9 Conclusions

Work in this chapter addresses the limitations of many common analytical techniques when faced with high concentrations of LAS. Two alternative approaches to T_m determination in these systems have been presented. The first involved CD spectroscopy on analogous SDS formulations to estimate T_m values under equivalent LAS conditions. The second exploited the irreversible nature of protein unfolding to facilitate surfactant removal prior to analysis.

4.9.1 Validity of New Approaches to T_m Determination in the Presence of LAS

Both approaches to exploring protein-LAS interactions have been shown to produce consistent T_m values which can be directly related to equivalent analyses by DSC for different surfactant concentrations. Results therefore also confirm trends observed by DSC, with significant destabilisation of proteases at all analysed surfactant concentrations, and Everest at concentrations above 0.1% LAS. As discussed in the previous chapter, data for V42 was of better quality than that of Everest. This is attributed to lower T_m values for the protease which fall well within temperature limits of the instrument.

Findings in the literature indicate that this correlation with DSC should ensure $T_{m\ CD}$ values determined using these methods can be directly related to long term storage stability.[5,7] This will be verified in *Chapter 5*.

4.9.2 Structural analysis of CD Data

The use of SDS as an analogue of LAS was further supported by analysis of CD data. Both PCA and deconvolution with the CONTIN programme showed comparable conformational effects for the two surfactants. While both methods provide valuable tools for the identification of similar structures, neither successfully delivered accurate data on secondary structure

CONTIN has been shown to be a valuable tool in the comparison of relative structural effects induced by various detergent conditions, however, absolute values are inconsistent with known protein properties. This is assumed to be a result of surfactant interference and limitations reported in the literature associated with comparing partially unfolded proteins with available reference spectra.[24] In light of these findings, PCA presented a more robust tool for evaluating relative effects of various excipients on protein structure.

4.10 Future work

Future work based on these conclusions should focus on further validating the above methods of evaluating protein-LAS interactions. Empirical CD data, indicating near-identical T_m values for both SDS analogue and $CaCl_2$ precipitation samples, require verification using an independent analytical method. Nano-DSC would provide a suitable platform for this work as it supports thermal analysis at up to 10% w/v LAS. More independent data would also allow the drift in T_m values observed between spectra collected *in situ* and using $CaCl_2$ purification for 0.1% LAS to be assessed. T_m values obtained via each method could then be adjusted to more accurately represent true thermal stability.

Based on successful validation of the above, we recommend focusing on the use of SDS as an analog for LAS as opposed to $CaCl_2$ purification methods for simplicity and efficiency. Further work will be required to establish these systems for multi-component samples, as interactions with other excipients will need be considered. Once fully defined, however, these formulations will open opportunities to incorporate previously inaccessible methods such as MST and DSF to the analytical suite.[2,10,13,19,26,27]

Finally, PCA should be applied to data collected for other detergent conditions, such as non-ionic surfactants and chelators, to explore the effects of these components on unfolding processes. Compiling these data

would present a global view of detergent enzyme stability, providing a basis for improved understanding of complex formulations.

Thermal denaturation data arising from this body of work, are intended for use in the development of predictive models of enzyme storage stability in HDL. Enzyme half-lives have been determined through accelerated storage tests under the detergent conditions explored in the previous two chapters. The relationship between these two measures of stability, and preliminary modelling of detergent systems will be discussed in the following chapter.

4.11 References

1. Y. Yu, J. Zhao & A.E. Bayly, *Chin. J. Chem. Eng.,* 2008, **16**, 517–527.
2. D. Otzen, *Biochim. Biophys. Acta*, 2011, **1814**, 562–591.
3. A. Crutzen, M.L. Douglass, Handbook of Detergents: Part A, Properties, CRC Press, 1999.
4. C.M. Johnson, *Arch. Biochem. Biophys.*,2013, **531**, 100–9.
5. H. Lund *et al.*, *J. Surfactants Deterg.*, 2012, **15**, 9–21.
6. H. Lund *et al.*, *J. Surfactants Deterg.,* 2012, **15**, 265–276.
7. D.S. Goldberg, S.M. Bishop, A.U. Shah & H.A. Sathish, *J. Pharm. Sci.*, 2011, **100**, 1306–1315.
8. Menzen, T., Friess, W. *J., Pharm. Sci.*, 2013, **102**, 415–428.
9. K. Miyazawa, M. Ogawa & T. Mitsui, *Int. J. Cosmet. Sci.,* 1984, **6**, 33–46.
10. A.K. Bhuyan, *Biopolymers,* 2010, **93**, 186–199.
11. E. Samper, M. Rodríguez, M.A. De la Rubia & D. Prats, *Sep. Purif. Technol.*, 2009, **65**, 337–342.
12. E. Samper, M. Rodríguez, I. Sentana & D. Prats, *Sep. Purif. Technol.*, 2010, **208**, 5–15.
13. Rafati, A. A., Gharibi, H. & Rezaie-Sameti, *J., Mol. Liq.,* 2004, **111**, 109–116.
14. M. Lindberg & A. Gräslund *FEBS Lett.,* 2001, **497**, 39–44.
15. T. Satsuki, Umehara, K. & Yoneyama, Y., *J., Am. Oil Chem. Soc.*, 1992, **69**, 672–677.

16. C.O. Rossi, M. Macchione & L. Filippelli, *Mol. Cryst. Liq. Cryst.*, 2011, **549**, 160–165.

17. J.A. Stewart, A. Saiani, A. Bayly & G.J.T. Tiddy, *Colloids Surf. Physicochem. Eng. Asp.,* 2009, **338**, 155–161.

18. D.E. Otzen, L. Christiansen & M. Schülein, *Protein Sci. Publ. Protein Soc.*, 1992, **8**, 1878–1887.

19. Y. Moriyama, Y. Kawasaka & K.J. Takeda, *Colloid Interface Sci.*, 2003, **257**, 41–46.

20. L. Whitmore & B. Wallace, 2008, **89**, 392–400.

21. N. Sreerama, S. Yu Venyaminov & R.W. Woody, *Anal. Biochem.,* 2001*,* **299**, 271–274.

22. P. Evans, O.A. Bateman, C. Slingsby & Wallace, *Exp. Eye Res.,* 2007*,* **84**, 1001–1008.

23. W. Kabsch & C. Sander, *Biopolymers,* 1983*,* **22**, 2577–2637.

24. S. Khrapunov, *Anal. Biochem*., 2009, **389**, 174–176.

25. N.J. Greenfield, *Nat Prot,* 2006, **1**, 2876–2890.

26. J. Xiang, J.B. Fan, N. Chen, J. Chen & Y. Liang, *Colloids Surf. B Biointerfaces,* 2006*,* **49**, 175–180.

27.Q. Xu & T.A. Keiderling, *Protein Sci. Publ. Protein Soc.,* 2004, **13**, 2949–2959.

28. B.S. Antharavally *et al*, *Anal. Biochem.,* 2011*,* **416**, 39–44.

5 Predicting Storage Stability of Detergent Enzymes by Thermal Analysis

Analysis of the thermal denaturation of detergent enzymes, outlined in the previous two chapters, was conducted with a view to develop predictive models of storage stability in HDL. The current industry standard for stability analysis involves lengthy storage tests and assays which are labour-intensive and offer little insight into the causes of protein instability. Several authors, however, have reported direct correlations between thermal denaturation parameters and storage stability values of various proteins.[1–4] This indicates the potential to model long-term protein stability on thermal analysis, reducing the need for storage tests. T_m values can generally be determined for a given sample within hours, in contrast to the numerous weeks required to establish degradation rates. Furthermore, high throughput techniques such as DSF facilitate the analysis of up to 96 samples simultaneously.

Application of these models to the detergent industry has been hindered, however, due to the complexity of T_m determination in HDL formulations. As described in previous chapters, the large number of sample components, high viscosity and UV active nature of these systems prevent direct application of many popular thermal analysis methods.

Recent work by Lund *et al*[5], however, has demonstrated the capabilities of nano-DSC in identification of melting temperatures under surfactant-rich conditions. Improvements in sensitivity over traditional DSC facilitates the detection of small enthalpy changes associated with unfolding, following careful subtraction of a nil-enzyme reference (*Chapter 3.3.1*). The authors prepared several simplified detergent

systems consisting of various combinations of the surfactants LAS, AE and AEO. Enzymes were stored in these formulations, under accelerated denaturation conditions of 35 °C, for a 14-day duration. Daily assays of residual enzyme activity were conducted to establish rates of degradation in each system. These data were plotted as a function of T_{max} values determined by DSC, yielding a linear correlation, as illustrated in *Figure 93*.

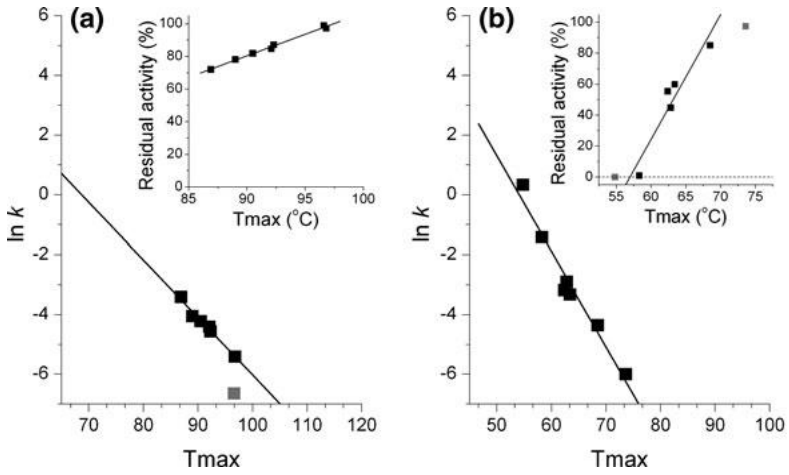


Figure 93: Data reproduced with permission from Lund[5], demonstrating the relationship between rate constant (k) and T_{max} for: (a) amylase and (b) protease. R^2 values for the linear regression plots are 0.99 and 0.98 respectively. Inset: Correlation between residual activity after storage for 14 days at 35 °C and T_{max} with R^2 values of 0.99 and 0.95 respectively.

This work demonstrates that established methods of predicting rates of denaturation from thermal stability data can be applied to HDL systems and detergent enzymes, reducing reliance on storage tests. Despite the strong linear relationship reported between rates of degradation and DSC data, however, the method offers very low throughput and little insight into unfolding processes. The high viscosity of surfactant-rich samples also prevents automation, further adding to analysis times. To address these issues, this chapter aims to replicate work on monoclonal antibodies conducted by Goldberg *et al*[2], which found a linear relationship between $T_{max\ DSC}$ data and $T_{m\ DSF}$ data. Previous work in this thesis (*Chapter 3.3.2*) indicates that the same is true of T_m values obtained by CD, allowing for the substitution of these methods in place of nano-DSC.

Thermal analysis in high surfactant conditions, described in *Chapter 5*, focused on CD methods to avoid complications arising from dye-surfactant interactions. For this reason, $T_{m\ CD}$ values will be used throughout this chapter, as representative datasets from techniques using optical detection. These data will be fitted as functions of enzyme half-lives, calculated from rates observed using industry standard accelerated stability studies. An overview of these storage experiments will be detailed in the coming chapter. To support work described by previous authors, a direct correlation between T_m and $T_{1/2}$ should be evident, providing the basis for the development of predictive long-term stability models for HDL formulations. Correlations between CD and DSF data, described in *Chapter 4.3*, should ensure straightforward transfer of the method to higher throughput techniques.

5.1 Accelerated Storage Tests of V42 and Everest

Fitting of T_m data necessitated collection of a parallel dataset detailing storage stabilities. In line with methods reported by Lund, these tests were conducted under accelerated conditions, commonly used in the protein industry to reduce the time span of experiments. Details of procedures followed can be found in *Chapter 2.1*. Accelerated tests involve storage of samples under stress conditions, typically elevated temperatures, to increase rates of denaturation. The original, unstressed values can then be calculated based on the Arrhenius model of reactions rates. The level of acceleration achievable for these tests is limited by the thermal stability of given proteins. Excessively high temperatures result in thermal unfolding which interferes with apparent storage stability. To avoid this, samples are commonly stored at 35 °C.[6,7]

At predetermined timepoints, over the course of a study, enzyme activity measured is measured using photometric assays. As detergent enzymes are designed for low specificity, a majority of available assay substrates are appropriate. In this work, *N*-succinyl-Ala-Ala-Pro-Phe-*p*-nitroanilide

(suc-AAPF-*p*NA) and ethylidene-*para*nitrophenol-Glucose-7 (EPS) were selected for the protease and amylase samples respectively *(Figure 94-Figure 95).*[8,9] Enzyme activity is determined at each time point by following the rate of substrate to product conversion based on the concentration of a released chromophore over time. Activity levels at each time point are then plotted to determine the rate of protein inactivation.

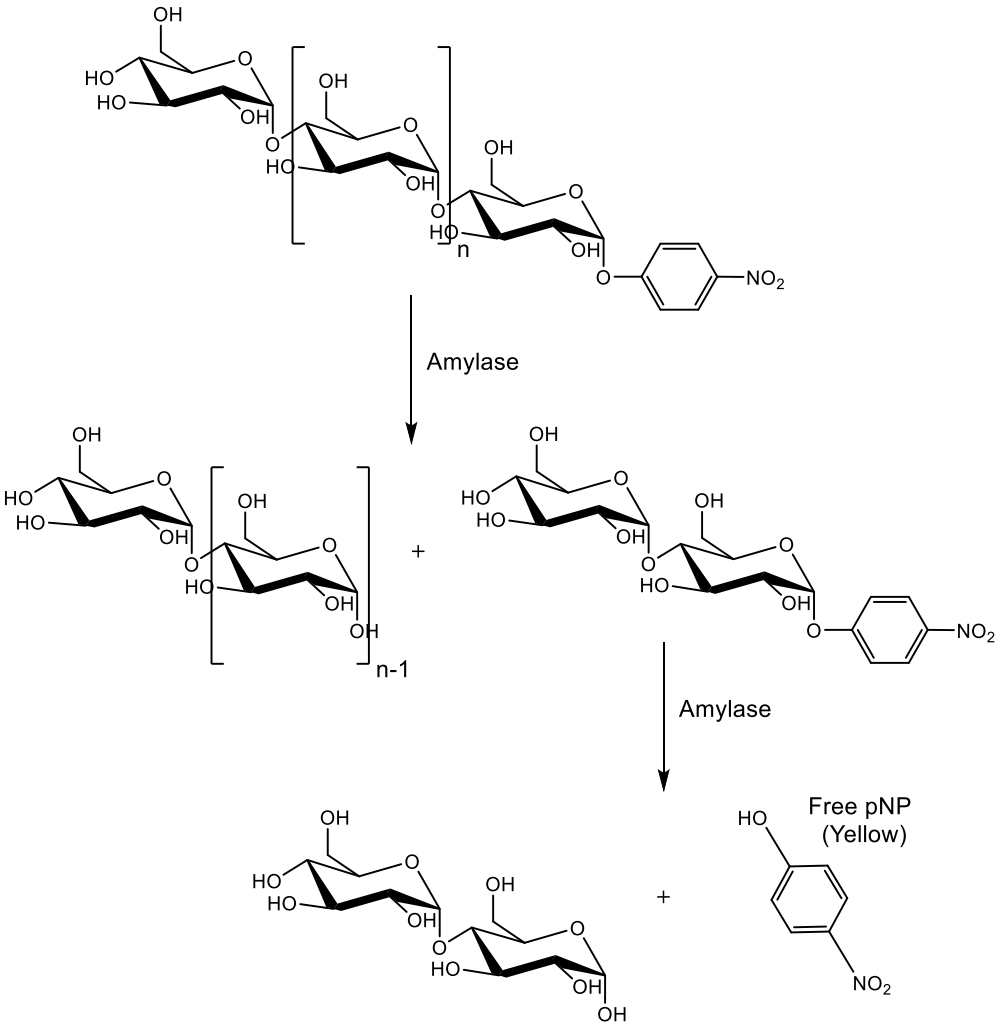


Figure 94: The use of EPS as a substrate for an amylase assay. The chromophore, pNP is liberated to produce colorimetric signal.

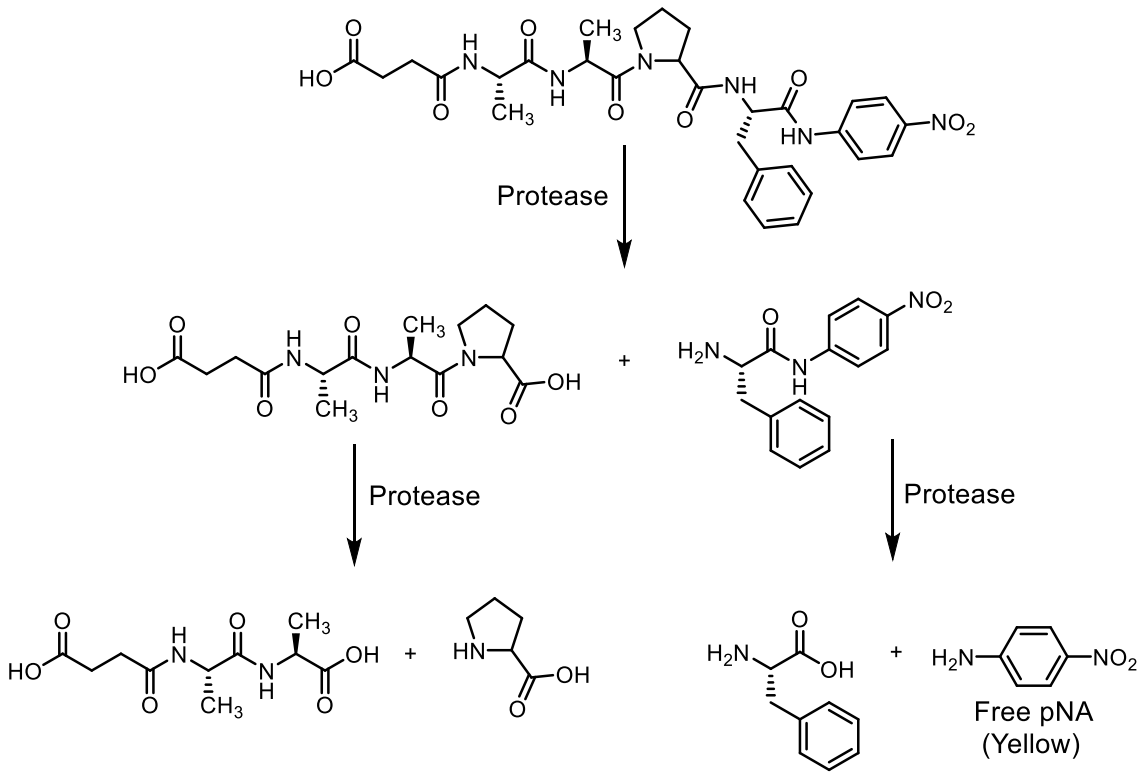


Figure 95:The use of artificial substrate with p-nitroaniline to assay protein activity. P-nitroaniline is liberated to produce a yellow colorimetric signal.

Rates of denaturation are calculated by plotting observed activity levels at each timepoint as a function of total storage time, as illustrated in *Figure 96*. These data can be fitted to an exponential curve for calculation of both the rate of degradation (k) and the half-life ($T_{1/2}$) under the given conditions, using *Equations 3-4* below.

$$\boldsymbol{y = ae^{kx}} \qquad \textbf{(Eq. 3)}$$

$$\boldsymbol{T_{\frac{1}{2}} = \frac{\ln(2)}{k}} \qquad \textbf{(Eq. 4)}$$

where a is the scale factor, k is the rate constant of degradation and $T_{1/2}$ is the half-life.

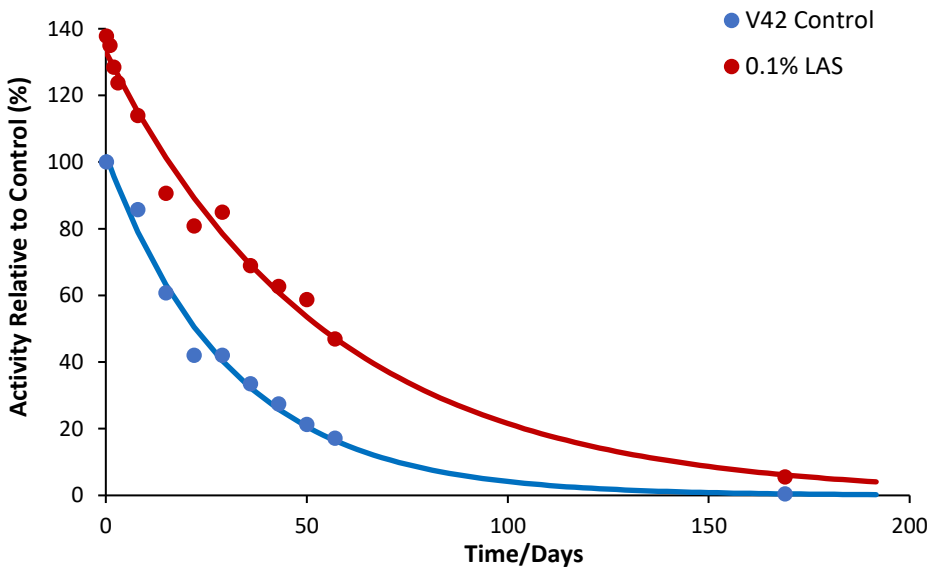


Figure 96: Degradation profile of V42 (control) and V42 (0.1% LAS) fitted to an exponential curve.

V42 (protease) and Everest (amylase) were selected as representative enzymes for storage testing to match available CD data. Enzymes were added individually to a range of formulations, corresponding to those analysed using CD (*Chapter 3.2)*. These included both anionic and non-ionic surfactants at a range of concentrations (0.1%-20%), and chelants and builders at commercial levels. Both SDS and LAS have been included in storage tests to probe the effects of observed unfolding on enzyme activity. The following results sections will detail degradation rates and half-lives obtained for each component class. These data will be discussed with reference to thermal denaturation analysis from the *Chapters 3-4* in order to evaluate the scope of predictive models. Finally, data from all anionic surfactant, non-ionic surfactant and chelating agents will be combined in an attempt to construct a single model for the prediction of enzyme stability in these simplified formulations. Future work will focus on developing these models to encompass multi-component samples and fully formulated detergents.

Due to time constraints, multiple samples were not analysed in storage tests. In the following sections, a measure of error has been provided for rates and $T_{1/2}$ based on the error arising from data fitting to an exponential. This may underestimate the systematic error of the

experiments as photometric analyses are generally considered to be accurate to within 10%. The high viscosity of surfactant-rich formulations and UV interference caused by LAS may contribute to further variation in the data. This has been considered for discussion of results in the coming sections.

5.2 Results – Accelerated Storage Stability Study of V42

The experimental values for the half-life of V42 under various detergent conditions are listed in Table 22: Experimental values for the half-life of V42 under a range of detergent conditions.[a,b,c] *Table 22* below. Plots of enzyme activity as a function of storage time, used to calculate these values, can be found in *Appendix 3.2*. For simplicity, results will be discussed in terms with respect to observed accelerated values, rather than the long-term storage stabilities they represent.

Table 22: Experimental values for the half-life of V42 under a range of detergent conditions.[a,b,c]

Formulation	Half-Life/Days	Error (Fitting)[d]	Error (Systematic)[e]
Control	22.2	± *1.3*	± *2.2*
0.1% LAS	38.0	± *2.2*	± *5.7*
1% LAS	31.7	± *2.0*	± *4.8*
5% LAS	24.9	± *0.9*	± *3.7*
10% LAS	31.8	± *2.0*	± *4.8*
20% LAS	39.3	± *2.4*	± *5.9*
0.1% SDS	12.6	± *1.4*	± *1.3*
1% SDS	9.6	± *0.8*	± *1.0*
5% SDS	3.6	± *0.6*	± *0.4*
10% SDS	1.2	± *0.1*	± *0.1*
20% SDS	0.8	± *0.1*	± *0.1*
0.1% AE3S	14.7	± *0.7*	± *1.5*
1% AE3S	17.3	± *0.8*	± *1.5*
5% AE3S	24.1	± *1.6*	± *1.6*
10% AE3S	28.3	± *0.9*	± *1.6*
20% AE3S	23.1	± *1.3*	± *1.8*
0.1% AE7	15.1	± *0.8*	± *1.5*
1% AE7	15.1	± *0.8*	± *1.7*
5% AE7	15.8	± *1.0*	± *2.4*
10% AE7	15.8	± *1.0*	± *2.3*
20% AE7	18.1	± *1.5*	± *2.8*
EDTA	1.8	± *0.3*	± *2.9*
HEDP	28.8	± *2.1*	± *0.2*
Citric Acid	12.7	± *1.3*	± *1.3*
Fatty Acid	33.4	± *2.2*	± *3.3*

[a]Activity determined using para-nitroaniline assays to assess rate of substrate conversion at each time point. [b]Activity monitored at accelerated rates over an 8-week period. [c]Rate of decay and half-life calculated based on initial rates. [d]Error values calculated based on the standard error of curve fitting. [e]Systematic error of photometric assays estimated at 10% for non-viscous solutions and 20% for viscous LAS.

5.3 The Effects of LAS on V42 Storage Stability

No clear trend in the effects of LAS concentration and storage stability could be determined from this data. Due to the high viscosity and optical

opacity, uncertainty in reported $T_{1/2}$ values has been estimated at 10-20%, as shown in *Figure 97*. As a result, the significance of observed changes in $T_{1/2}$ are unclear.

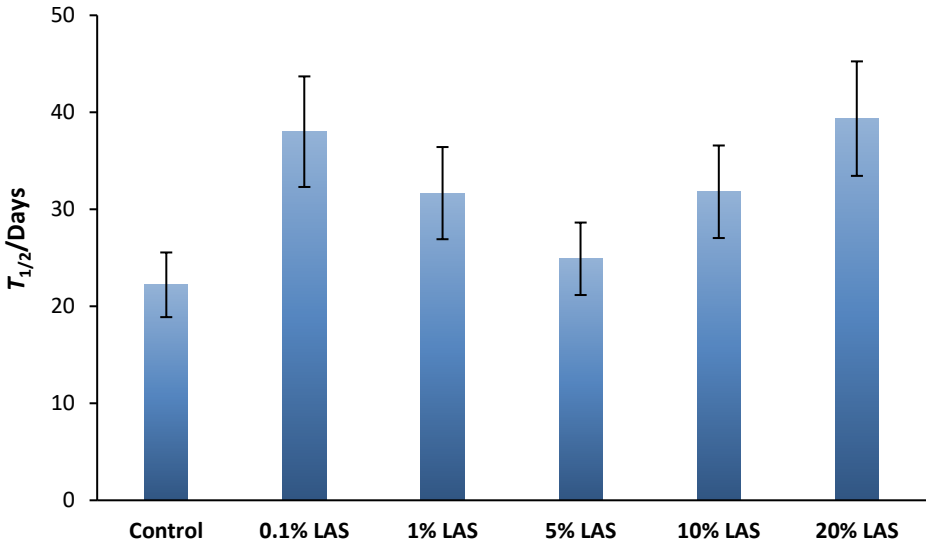


Figure 97: Chart of $T_{1/2}$ values for V42 in the presence of various concentrations of LAS as determined by accelerated storage tests.

Furthermore, no direct correlation with respective T_m values, required for predictive modelling, could be defined (*Figure 98*). Storage stability in the presence of other excipients was found to have stronger correlations with T_m values, as reported in the literature. This will be demonstrated in the coming section. These fittings indicate error may be result of LAS interference, reported across analytical methods in this thesis, rather than the absence of a link between thermal unfolding and shelf-life. Detector interference may be an artefact of either the high viscosity or the UV absorbance of the surfactant. These issues emphasise the need for more efficient and insightful methods of protein analysis, as re-analysis and confirmation of results were too time intensive to be conducted.

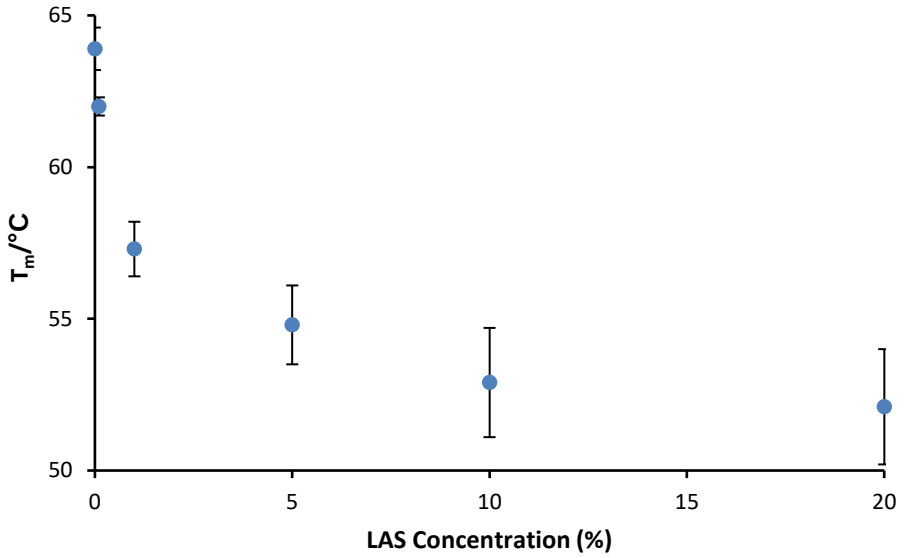


Figure 98: Reduction in T_m with increasing concentration of surfactant, observed for V42 by CD.

A second data set describing storage stability of V42 in LAS had previously been collected, however, and was available for further comparison of surfactant effects *(Table 23, Appendix 3.1)*. This dataset consisted of the results of a smaller study, conducted at the beginning of the project, to establish protocols which were in line with industry-standard methods. Although the data set is incomplete, with poor quality data at 10% LAS, available $T_{1/2}$ values evidenced the downward trend in stability which had been predicted by thermal analysis (*Figure 99*). A plot of these data as a function of respective T_m values obtained by CD yielded a linear correlation with an R^2 value of 0.92 *(Figure 100)*.

The two data sets report similar trends for 0.1-5% LAS, with deviations arising due to the more viscous, high concentration samples (10-20%) This has been attributed to detector interference and non-homogeneous sampling. The original, 'Data Set 1' reports far higher absolute values for $T_{1/2}$ than the preliminary storage tests results. This may be an artefact of different instruments and enzyme batches used between the two experiments. Variations will be considered further on preparation of the complete V42 model at the end of this section to ensuring comparability with other detergent conditions.

Table 23: Comparison of experimental T_m and $T_{1/2}$ values for V42 in varying concentrations of LAS.

LAS Concentration	'Data Set 1' $T_{1/2}$ (Days)[a]	'Data Set 2' $T_{1/2}$ (Days)[a]	V42 $T_{m\ CD}$ (°C)[b]
Control	22.2 (± *1.3*)	22.2 (± *1.3*)	63.9 (± *0.6*)
0.1%	12.6 (± *1.4*)	24.0 (± *0.8*)	62.3 (± *0.9*)
1%	9.6 (± *0.8*)	17.3 (± *1.2*)	57.4 (± *1.8*)
5%	3.6 (± 0.6)	15.2 (± *0.6*)	53.1 (± *0.3*)
10%	1.2 (± *1.1*)	-	51.6 (± *1.6*)
20%	0.8 (± *0.1*)	5.2 (± *0.7*)	50.5 *(± 1.8)*

[a]*$T_{1/2}$ values determined through storage tests and PNA activity assays. Values are listed in days.* [b]*T_m values determined using in situ heating and CD analysis. Values are listed in °C.*

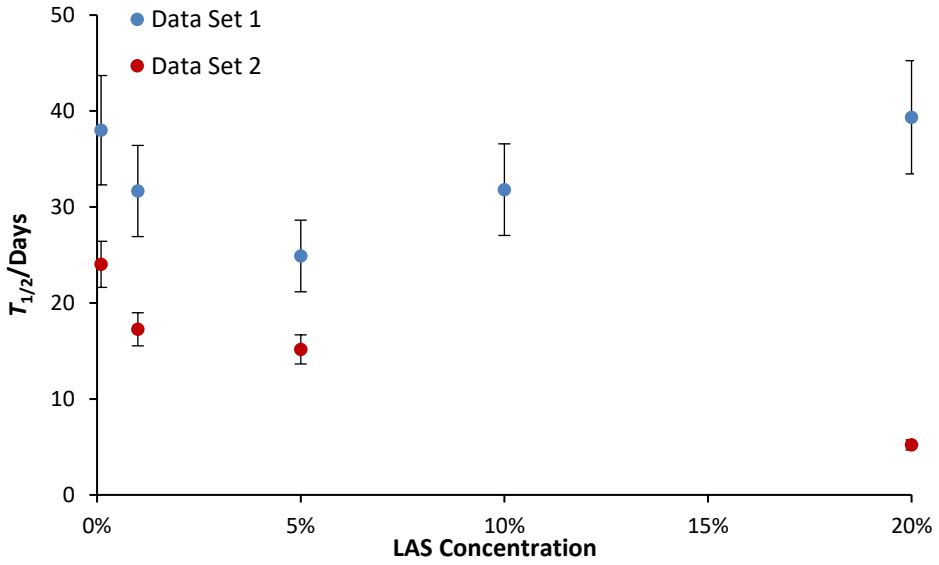


Figure 99: $T_{1/2}$ values obtained for V42 in the presence of varying concentrations of LAS over two independent storage tests. Error bars represent estimated systematic error of 20%.

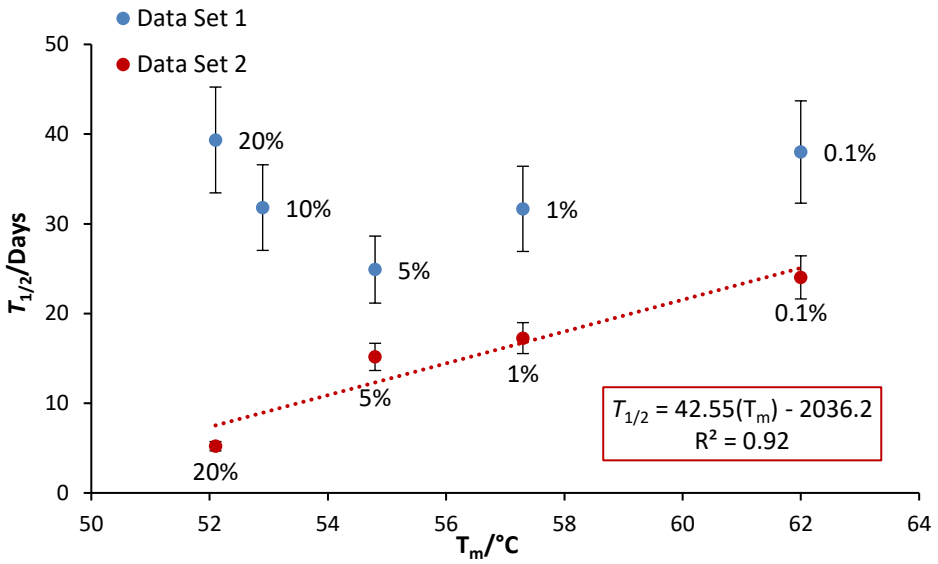


Figure 100: Half-lives of V42 in various concentrations of LAS over two independent tests as a function of T_m values obtained by CD. Error bars represent estimated systematic error of 20%.

5.4 The Effects of SDS on V42 Storage Stability

SDS, in contrast, exhibited a clear and significant downward trend in $T_{1/2}$ between surfactant concentration and the half-life of V42. At 0.1% SDS, half-life was reduced from ~22 days in the control to 12.6 days. This trend continued across the concentration gradient with $T_{1/2}$ values of 9.6, 3.6, 1.2 and 0.8 days reported for 1%, 5%, 10% and 20% SDS respectively, as illustrated in *Figure 101*. Observations were in line with predictions from thermal denaturation analysis (*Chapter 4.2*). The reduced opacity and viscosity of the SDS samples, in comparison to LAS, may account for the improved data quality.

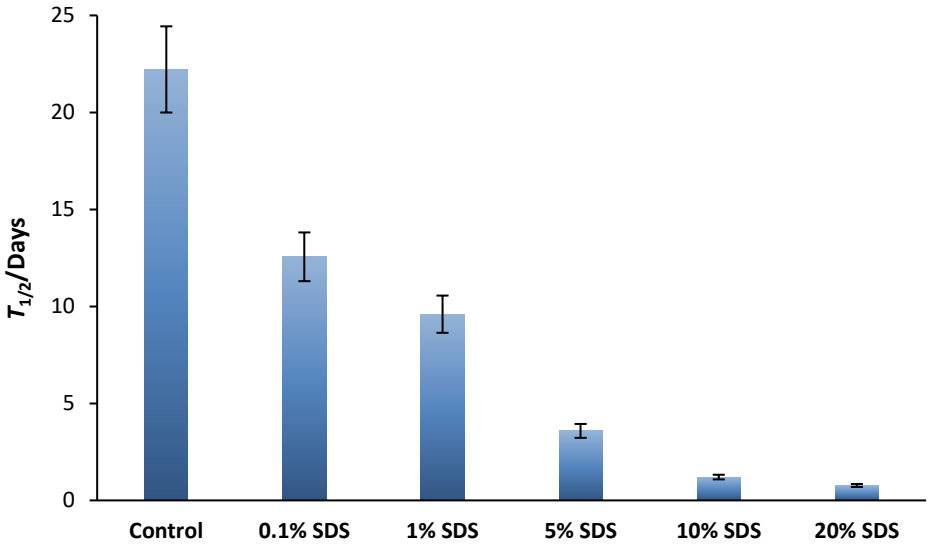


Figure 101: Chart of $T_{1/2}$ values for V42 in the presence of various concentrations of SDS as determined by accelerated storage tests. Error bars represent estimated systematic error of 10%.

During CD analysis, the effects of SDS and LAS on the thermal denaturation of V42 were observed to be almost identical (*Chapter 4.7*). Similar effects of the two surfactants on V42 storage stability values were therefore expected, supporting the above use of 'Dataset 2' for $T_{1/2}$ values in LAS. This use of thermal stability data to validate inconsistent assay results, highlights the need to expand the range analysis methods available to the detergent industry. Error in storage tests is consistently higher than that of thermal analysis and extensive work is required to confirm outlying trends in the data. The provision of a secondary, high throughput method would greatly reduce both variability in results and the efficiency of collection and reassessment of data.

5.4.1 The Relationship between Thermal Stability and Storage Stability of V42 in the presence of SDS

Thermal data were compared with respective $T_{1/2}$ values to verify T_m values as indicators of long-term stability *(Table 24)*. Evidence in *Figure 102* shows a close relationship between trends observed for the two parameters across the selected range of SDS samples.

Table 24: Comparison of experimental T_m and $T_{1/2}$ values for V42 in varying concentrations of SDS.

SDS Concentration	V42 $T_{1/2}$ (Days)[b]	V42 T_m (°C)[a]
Control	22.2 (± *1.3*)	63.9 (± *0.6*)
0.1%	12.6 (± *1.4*)	62.3 (± *0.9*)
1%	9.6 (± *0.8*)	57.4 (± *1.8*)
5%	3.6 (± *0.6*)	53.1 (± *0.3*)
10%	1.2 (± *1.1*)	51.6 (± *1.6*)
20%	0.8 (± *0.1*)	50.5 (± *1.8*)

[a]T_m values determined using in situ heating and CD analysis, listed in °C. [b]$T_{1/2}$ values determined through storage tests and PNA activity assays, listed in unit of days.

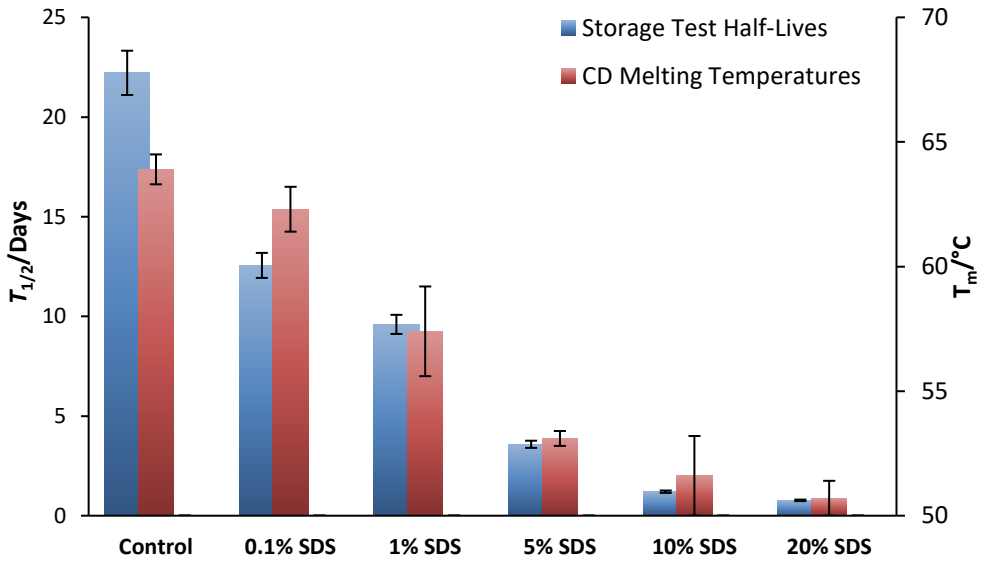


Figure 102: Chart of $T_{1/2}$ values (blue) and T_m values (red) of V42 in SDS as determined by storage tests and CD respectively. Error bars represent estimated systematic error of 10% for storage tests and standard error of three replicates for T_m values.

Empirical fitting of $T_{1/2}$ as a function of T_m for each concentration point is shown in *Figure 103*. The plot yielded a linear correlation with an R^2 value of 0.92, demonstrating the proportional increase in storage stability with thermal stability, as suggested by LAS data from 'Data Set 2' *(Figure 100)*. Through linear regression, T_m values can therefore be used to estimate the half-lives of enzymes in known concentrations of SDS. This provides the required support for the use of $T_{m\ CD}$ values to build predictive models of enzyme storage stability in the presence of anionic surfactant. These fittings had previously only been achieved using

calorimetric methods. Transfer to CD offers improved throughput using sample changers. This type of autosampling is unaffected by the viscosity issues encountered with DSC, as sample changing is by means of a moving rack of manually filled cuvettes, rather than in injection into a single port. Structural analysis can also be used to monitor changes in configuration leading to inactivity, as demonstrated in *Chapter 4.7-4.8*. Validation of T_m values, as opposed to $T_{max\ DSC}$, also introduces the potential for ultra-high throughput methods via DSF.

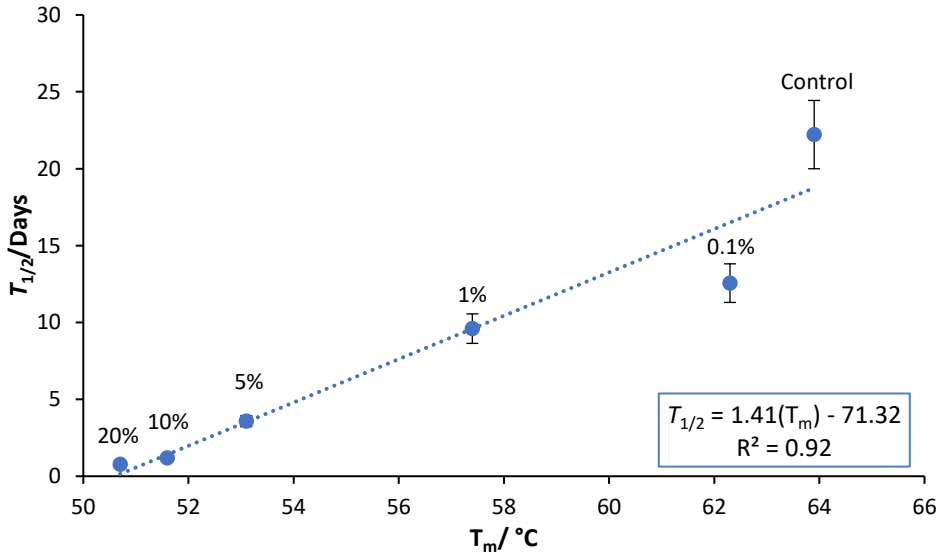


Figure 103: Correlation between experimental values for half-life ($T_{1/2}$) and melting temperature (T_m) of V42 in varying concentrations of SDS as determined by storage experiments and CD respectively.

5.4.2 The use of SDS as an Analog for the Prediction of $T_{1/2}$ values in LAS

The primary anionic surfactant found in commercial laundry formulations, however, is LAS rather than SDS. The role of SDS in this study, therefore, was to act as an analog for LAS in analytical samples to reduce detector interference. This theory was dependant on SDS producing comparable effects on protein structure and function and thus stability in formulation. In the previous chapter, almost identical T_m values were obtained for V42 in the presence of SDS and LAS. These are reported in *Table 25* along with respective $T_{1/2}$ values. Furthermore, PCA and CONTIN methods of comparison indicated similar structural effects

on V42 at equivalent temperatures and concentrations for each of the two surfactants.

A linear relationship was observed between $T_{1/2}$ values in SDS and those in LAS, with an R^2 value of 0.86. Data from preliminary studies were used for this comparison due to inconsistencies in $T_{1/2}$ values from the full stability study. Just four concentration points were available, however, and so further experiments should be conducted to confirm this trend (*Figure 104*).

The plot in *Figure 105,* of V42 $T_{1/2}$ values for each surfactant condition against respective T_m values in SDS, highlights the clear relationship between the two datasets. The trendline for LAS samples is offset from that of SDS by ~7 days, however, preventing a direct substitution of T_m values. This may be an artefact arising from differences in analytical procedures, as SDS spectra were collected *in situ* while LAS samples were purified with $CaCl_2$ precipitation. Lower T_m values were also observed for proteins analysed using the '*in situ*' method at 0.1% LAS, (*Chapter 4.5*). Values in both cases were consistent and have been attributed to a degree of protein refolding during cooling and purification. This may account for the difference in storage stabilities in LAS and SDS samples, which were not reflected by T_m values. These trends should be considered in the development of predictive models and will be discussed further in *Section 5.7.*

Table 25: Experimental $T_{m\ CD}$ and $T_{1/2}$ values for V42 in the presence of various concentrations of SDS and LAS

Surfactant Concentration	SDS		LAS	
	T_m (°C)	$T_{1/2}$ (Days)	T_m (°C)[a]	$T_{1/2}$ (Days)
Control	63.9	22.2	63.9	22.2
0.1%	62.3	12.6	62.0	24.0
1%	57.4	9.6	57.3	17.3
5%	53.1	3.6	54.8	15.2
10%	51.6	1.2	52.9	-
20%	50.5	0.8	521	5.2

[a]*Data taken from preliminary stability study (Data Set 2).*

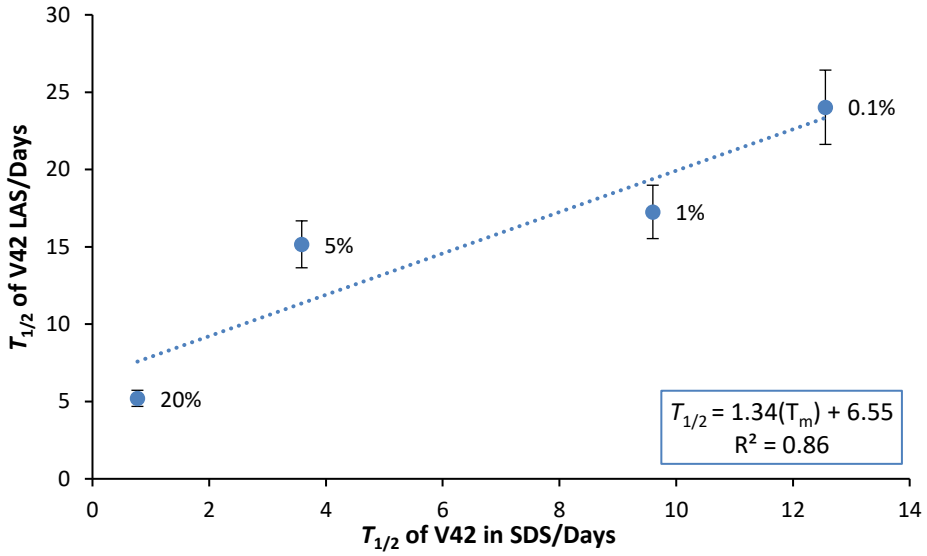


Figure 104: Correlation between $T_{1/2}$ values for Everest in the presence of LAS and those in SDS. Error bars represent estimated systematic error of 10%.

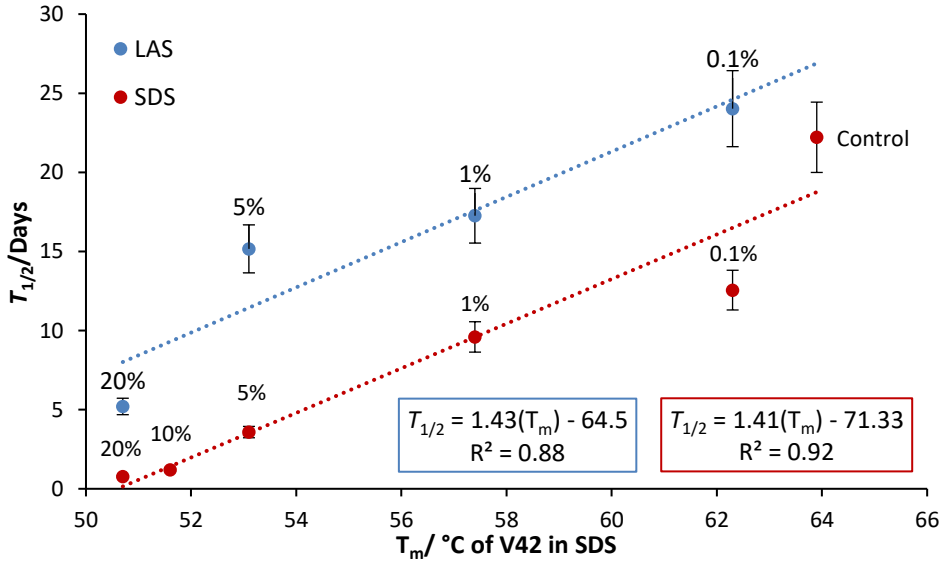


Figure 105: $T_{1/2}$ values for V42 in LAS (blue) and SDS (red) as a function of T_m CD of V42 in SDS for respective surfactant concentrations. Error bars represent estimated systematic error of 10%.

5.5 The Effects of Secondary Surfactants on V42 Storage Stability

Secondary surfactants AE3S (anionic) and AE7 (non-ionic) are included in laundry formulations to aid cleaning power and reduce the destabilising effects of LAS on detergent enzymes. Concentrations of these compounds tend to be lower at approximately 5% w/v in HDL. A summary of $T_{1/2}$ values observed under a range of these surfactant conditions is presented in *Table 26* and *Figure 106*.

Table 26: $T_{1/2}$ values for V42 in various concentrations of AE3S and AE7.[a]

Surfactant concentration	$T_{1/2}$ (AE3S)	$T_{1/2}$ (AE7)
Control	22.2 *(± 1.3)*	22.2 *(± 1.3)*
0.1%	14.7 *(± 0.7)*	15.1 *(± 0.8)*
1%	17.3 *(± 0.8)*	15.1*(± 0.8)*
5%	24.1 *(± 1.6)*	15.8 *(± 0.1)*
10%	23.1 *(± 1.3)*	15.8 *(± 0.1)*
20%	28.3 *(± 1.9)*	18.1 *(± 1.5)*

[a] *$T_{1/2}$ values determined through storage tests and PNA activity assays. Values are listed in days.*

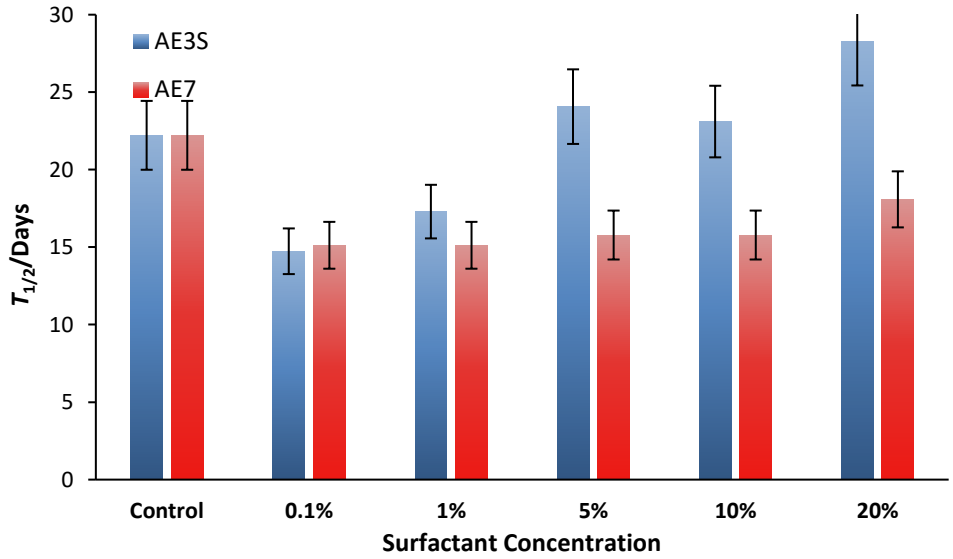


Figure 106: Chart of $T_{1/2}$ values for V42 in the presence of various concentrations of AE3S and AE7 as determined by accelerated storage tests. Error bars represent estimated systematic error of 10%.

Both surfactants induced similar levels of destabilisation at 0.1% w/v, with a reduction in $T_{1/2}$ of ~7 days. With increasing surfactant concentration from this level, significant increases in $T_{1/2}$ were observed

for AE3S. Above 5% w/v, stability was restored to that of the control sample and at 20%, an increase in $T_{1/2}$ of 6 days was recorded. In contrast, stability levels were maintained at the level of the 0.1% samples across the range of AE7 conditions analysed. A gradual positive drift in $T_{1/2}$ values, from 15 to 18 days between 0.1% and 20% surfactant, was observed, but not found to be significant with respect to the error. It was not possible to extrapolate from the data at what level of AE7, if any, a stabilising effect would be expected. As a lone additive, the non-ionic surfactant does not appear to have a positive impact on stability. Multi-surfactant systems studied by Lund, however, found the effects of LAS-induced denaturation were reduced in the presence of AE7 when compared to lone LAS formulations.[5] This highlights the need to expand this work into multicomponent systems to probe the synergistic effects of various excipients.

5.5.1 The Relationship between Thermal Stability and Storage Stability of V42 in the Presence of AE3S and AE7

Comparison of storage and thermal stability values, conducted for the two previous conditions, were repeated with AE3S and AE7 samples. These values are listed in *Table 27* below. As discussed in *Chapter 3.2.3*, no clear trend was observed between T_m and concentration for either of the secondary surfactants. In both cases, V42 was stabilised at 0.1% and 5% but destabilised at 1%, 10% and 20%. Although the mechanism behind this trend is not understood, its replication across both surfactants suggests that this is a true effect, rather than a result of experimental error. Stability and thermal data at a greater number of concentration points within this range would further support these observations. Mechanistic insight could also be gained through understanding of surfactant binding and aggregation states. Monomer, micellar and lamellar phases will all interact differently with enzyme structure, affecting both binding and unfolding kinetics. Literature CMC values may be altered by the protein binding and so individual analysis

for each enzyme is required. Changes in surfactant aggregation over the tested concentration gradient may be the source of this unusual trend.

T_m values also contrasted with observed storage stability. Plotting thermal denaturation data against $T_{1/2}$ for each of the secondary surfactant formulations did not yield any obvious relationship between the two parameters *(Figure 107)*. The broad range of T_m values (~60-67 °C) also does not reflect the small changes in storage stabilities observed across the concentration gradient analysed. Plotting the two surfactants together further highlights this clustering of data and the lack of significant effects across the concentration gradients *(Figure 108)*.

Table 27: Comparison of experimental T_m and $T_{1/2}$ values for V42 in various concentrations of AE3S.

Surfactant Concentration	$T_{1/2}$ (AE3S)[a]	T_m (AE3S)[b]	$T_{1/2}$ (AE7)[a]	T_m (AE7)[b]
Control	22.2 *(± 1.3)*	63.9 *(± 0.6)*	22.2 *(± 1.3)*	63.9 *(± 0.6)*
0.1%	14.7 *(± 0.7)*	66.6 *(± 0.9)*	15.1 *(± 0.8)*	64.8 *(± 0.1)*
1%	17.3 *(± 0.8)*	61.1 *(± 0.9)*	15.1*(± 0.8)*	59.1 *(± 0.5)*
5%	24.1 *(± 1.6)*	65.5 *(± 0.2)*	15.8 *(± 0.1)*	68.5 *(± 0.3)*
10%	23.1 *(± 1.3)*	60.3 *(± 0.1)*	15.8 *(± 0.1)*	60.4 *(± 0.3)*
20%	28.3 *(± 1.9)*	61.4 *(± 0.5)*	18.1 *(± 1.5)*	62.5 *(± 0.9)*

[a]$T_{1/2}$ values determined through storage tests and PNA activity assays. Results are listed in days. [b]T_m values determined using in situ heating and CD analysis. Results are listed in units of °C.

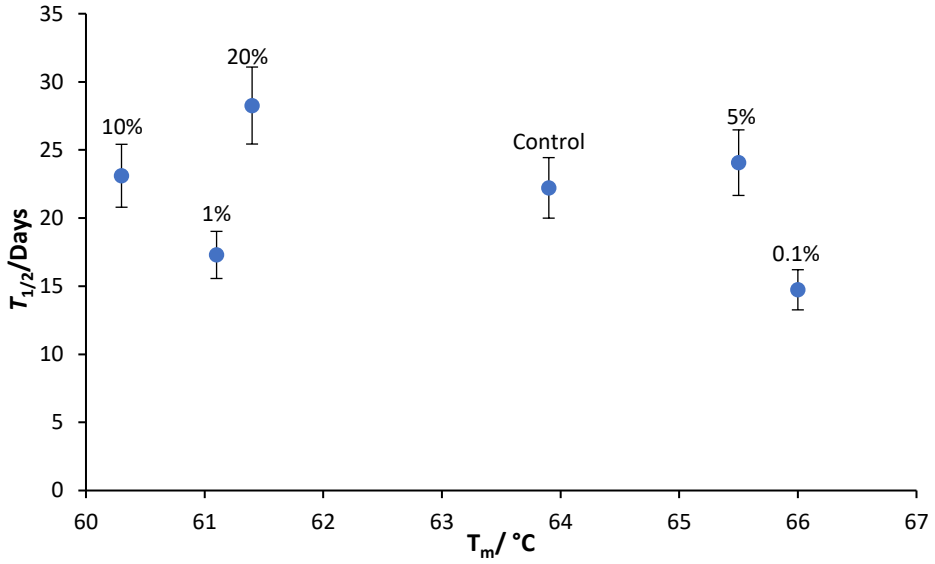


Figure 107: Plot of V42 $T_{1/2}$ in the presence of AE3S as a function of $T_{m\ CD}$ values under equivalent condition

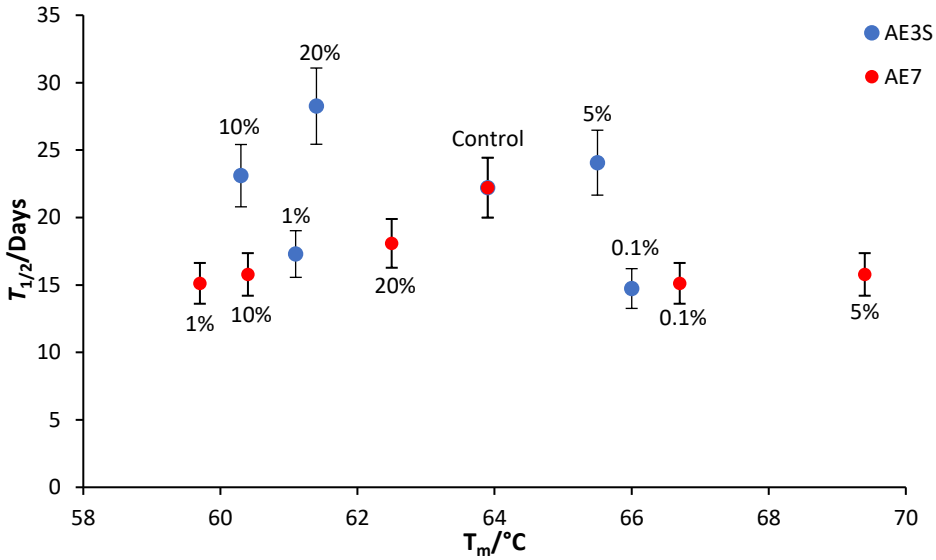


Figure 108: Plot of $T_{1/2}$ of V42 in the under various conditions of AE7 and AE3S as a function of $T_{m\ CD}$ values.

5.6 The Effects of Chelants and Builders on V42 Storage Stability

Chelating agents, added to HDL to reduce water hardness, destabilise proteins by sequestering structural calcium ions. A range of common laundry chelants and builders were analysed at commercially relevant concentrations of 2% for EDTA and HEDP (chelants) and 5% for citric acid and fatty acid ('builders'). As illustrated in *Figure 109* below, the most destabilising of these conditions was produced by EDTA with a half-

life of just 1.8 days compared to 22 in the control. This is most likely due to the high Ca^{2+} binding constant associated with EDTA, in the order of 10^{7} [10]

Citric acid was the next most destabilising, producing a $T_{1/2}$ of 12.7 days. Less significant effects were observed with HEDP, which exhibited a small stabilisation of ~6 days with respect to the control. V42 in the presence of the 'builder', fatty acid was found to increase stability with respect to the control with an improvement in $T_{1/2}$ of 11 days, however, poor solubility of the builder may have impacted these results.

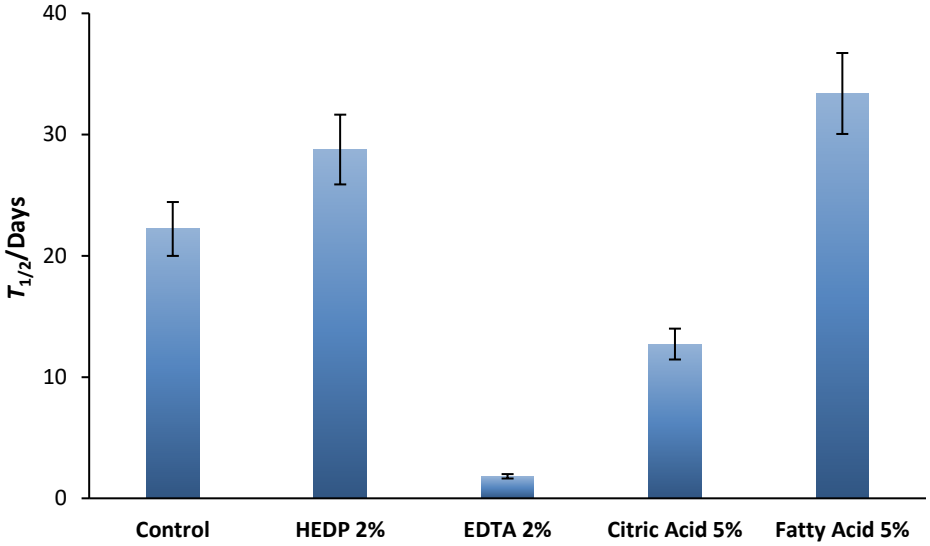


Figure 109: Chart of $T_{1/2}$ values for V42 in the presence of various chelants and builders at commercial levels, as determined by accelerated storage tests.

Observed half-lives for V42 in the presence of chelating agents were also compared with their respective T_m values *(Chapter 3.2),* however no clear correlation could be established *(Figure 110).* This suggests that multiple models may be required to fit various components, rather than a 'one size fits all model' for each enzyme as was previously postulated. Further work and data collection would be required to determine if these single component samples can contribute to modelling of fully formulated detergent samples.

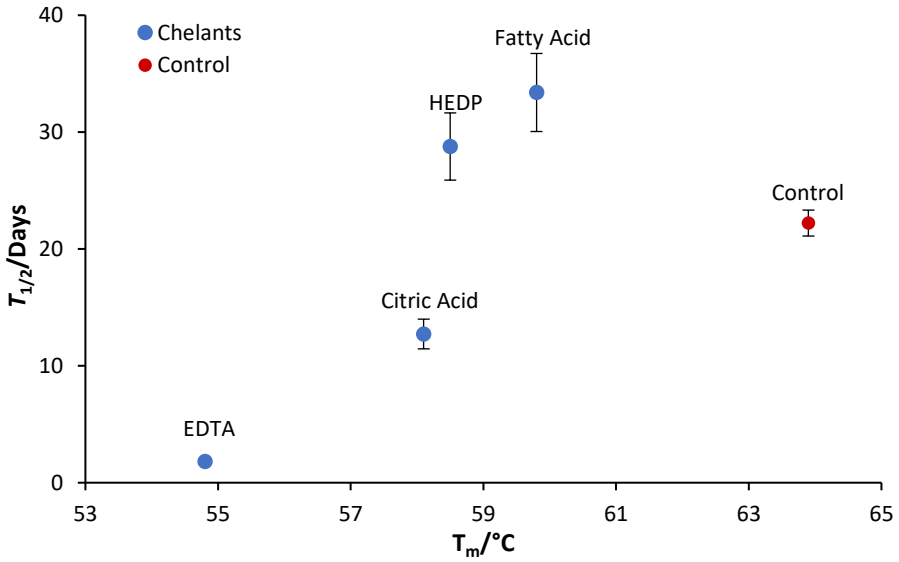


Figure 110: Plot of V42 $T_{1/2}$ values in the presence of various chelating agents as a function of their respective T_m values.

To explore the relationship between T_m and $T_{1/2}$ across all detergent components analysed the full dataset for each of the excipients discussed above will be presented in a combined plot in the following section.

5.7 Establishing T_m values as Predictive Indicators of Storage Stability

Several of the detergent conditions described in the previous sections demonstrated clear relationships between $T_{m\ CD}$ values and observed $T_{1/2}$ values. This lends support to the use of thermal denaturation analysis in prediction of long-term storage stability. The ultimate goal of this work was to establish a method of high throughput protein stability screening which was applicable to all detergent conditions. This required the generation of a single empirical fitting which mapped T_m values onto $T_{1/2}$ values for each detergent condition. The relevant thermal and storage stability data are listed in *Table* 28.

Table 28*:* Experimental values for the degradation rate and half-life of V42 under a range of detergent conditions.[a,b,c]

Formulation	Half-Life/Days	Error in $T_{1/2}$[d]	T_m/ °C[e]	Error in T_m[f]
Control	22.2	± *1.3*	63.9	± *0.6*
0.1% LAS	38.0	± *2.2*	62.0	± *0.3*
1% LAS	31.7	± *2.0*	57.3	± *0.9*
5% LAS	24.9	± *0.9*	54.8	± *1.3*
10% LAS	31.8	± *2.0*	52.9	± *1.8*
20% LAS	39.3	± *2.4*	52.1	± *1.9*
0.1% SDS	12.6	± *1.4*	62.3	± *0.9*
1% SDS	9.6	± *0.8*	57.4	± *1.8*
5% SDS	3.6	± *0.6*	53.1	± *0.3*
10% SDS	1.2	± *0.1*	51.6	± *1.6*
20% SDS	0.8	± *0.1*	50.7	± *1.8*
0.1% AE3S	14.7	± *0.7*	66.6	± *0.7*
1% AE3S	17.3	± *0.8*	61.1	± *1.5*
5% AE3S	24.1	± *1.6*	65.5	± *0.3*
10% AE3S	28.3	± *0.9*	60.3	± *0.1*
20% AE3S	23.1	± *1.3*	-	-
0.1% AE7	15.1	± *0.8*	64.8	± *0.4*
1% AE7	15.1	± *0.8*	59.1	± *0.4*
5% AE7	15.8	± *1.0*	68.5	± *0.7*
10% AE7	15.8	± *1.0*	60.4	± *0.4*
20% AE7	18.1	± *1.5*	-	-
EDTA	1.8	± *0.3*	54.8	± *0.6*
HEDP	28.8	± *2.1*	58.5	± *0.3*
Citric Acid	12.7	± *1.3*	58.1	± *0.1*
Fatty Acid	33.4	± *2.2*	59.8	± *0.1*

[a]Activity determined using EPS assays to assess rate of substrate conversion at each time point. [b]Activity monitored at accelerated rates over an 8-week period. [c]Half-life calculated based on initial rates. [d]Error values calculated based on the standard error of curve fitting. [e]T_m values determined by CD in-situ temperature ramp for all samples except LAS, determined by CD following purification. Reported values are an average of 3 independent analysis, error is listed as the standard error from the mean of these analyses.

$T_{1/2}$ values, as a function of respective $T_{m\ CD}$ values for each sample were fitted using linear regression, in line with work by Lund *et al.*[5,11] Although some scatter from the fit was observed, a linear trend was

evident between the two parameters *(Figure 111)*, indicating that the correlation between T_m and $T_{1/2}$ for V42 is constant, regardless of protein environment. The translation of T_m to $T_{1/2}$ values for a give condition is described in *Equation 5*.

$$T_{\frac{1}{2}} = 1.43(T_m) - 70.14 \qquad \textbf{(Eq. 5)}$$

As illustrated in *Figure 111*-*Figure* 112, some extreme outliers were omitted from the calculation of the line of best fit, resulting in an R^2 value of 0.79. LAS samples showed the greatest deviation from the fit, and so the entire group has been excluded. To determine whether this deviation was an artefact of the $CaCl_2$ precipitation process, or of error in storage tests, a second plot of LAS data from preliminary storage tests ('Dataset 2') was constructed *(Figure 113)*. These data showed significantly less deviation from the fit, giving an R^2 value of 0.70 with LAS data included, compared with 0.05 on inclusion of the original dataset, shown in *Figure 112*.

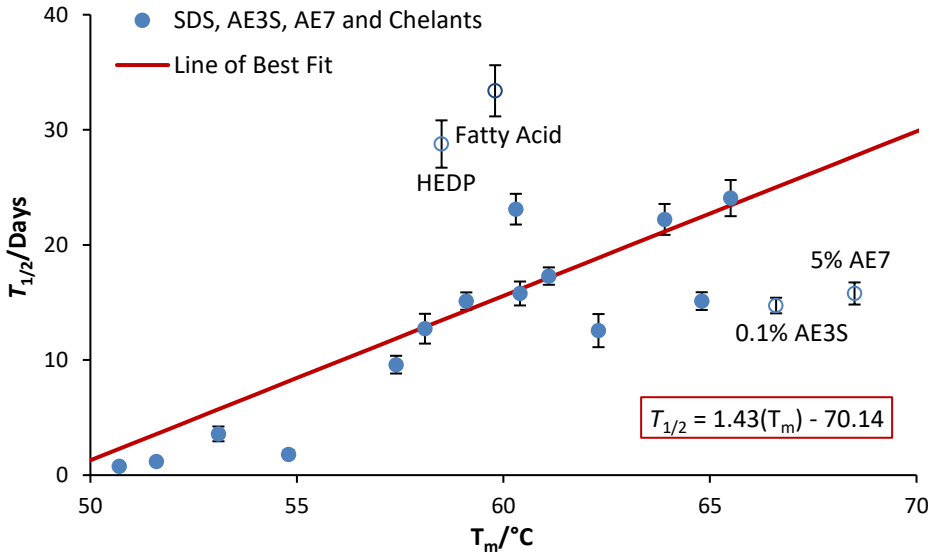


Figure 111: Plot of V42 $T_{1/2}$ as a function of T_m values for V42 determined by CD. Data points excluded from calculation of best fit line are represented by unfilled circles.

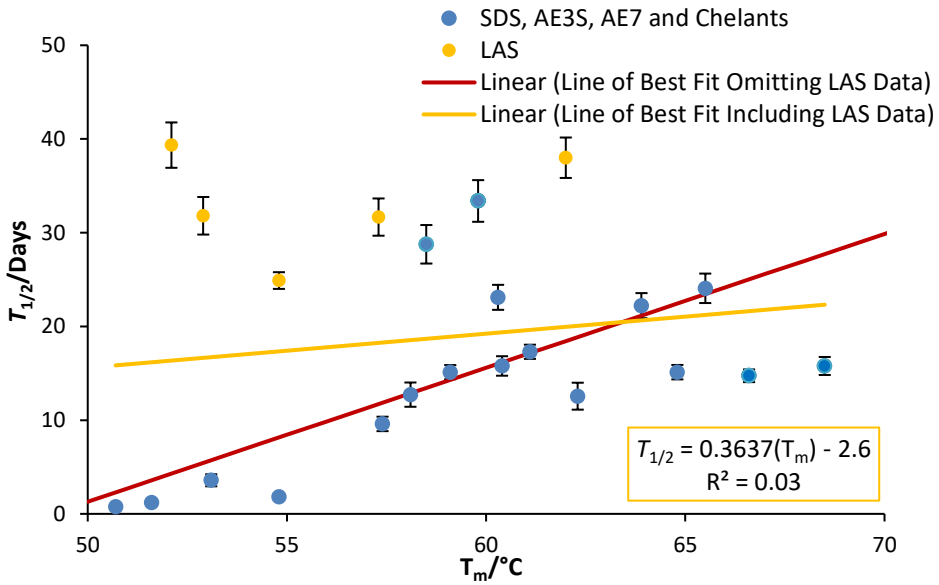


Figure 112: Plot of V42 $T_{1/2}$ as a function of T_m values for V42 determined by CD. Outlying LAS data is included in orange. Data points excluded from calculation of best fit line are represented by unfilled circles.

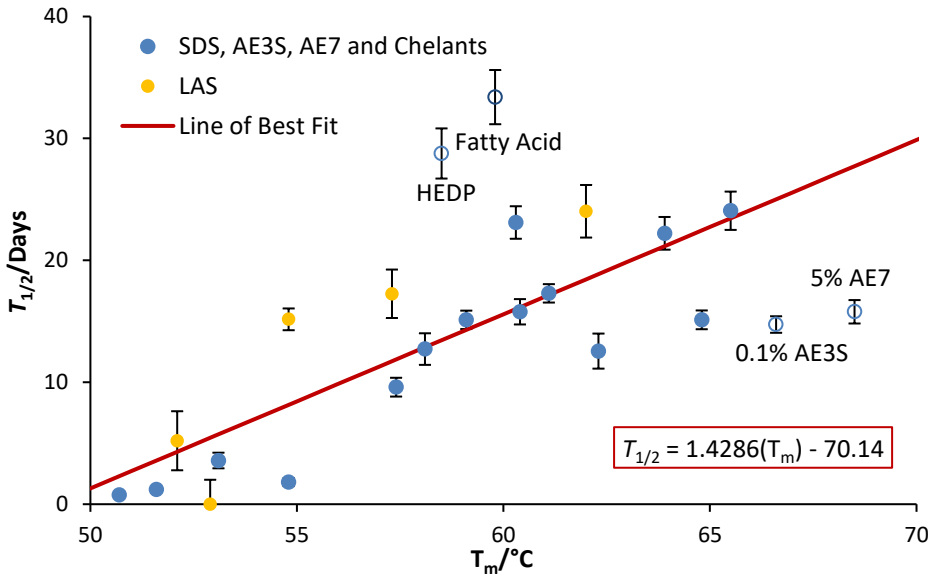


Figure 113: Plot of V42 $T_{1/2}$ as a function of T_m values for V42 determined by CD. Outlying data excluded from fitting are represented by unfilled circles. LAS data (orange) has been substituted with values obtained from preliminary storage tests.

$T_{1/2}$ values from this second LAS dataset were still found to be consistently higher than those predicted by the model fitting. These elevated values may be attributed to the differences in T_m, reported in *Chapter 4.5,* between samples cooled and purified before analysis and those analysed *in situ*. The extent of this deviation could only be estimated at 0.1% LAS, due to detector interference of *in situ* values, and

no data was obtained at higher concentrations. This drift in T_m should therefore be explored further using $CaCl_2$ precipitation analysis of SDS. The available data set of *in situ* values would aid determination of the degree of change in T_m caused by surfactant removal, enabling predictive T_m values for LAS to be adjusted accordingly.

5.7.1 Prediction of Half-Life in LAS using Analogous SDS T_m Values

Alternatively, we had hypothesised that half-life in LAS could be predicted from the T_m values of analogous SDS formulations. To validate this theory, $T_{1/2}$ values in LAS were estimated from T_m values in equivalent SDS samples using *Equation 5*. As stability in the presence of LAS was offset from that of SDS by ~7 days, T_m values were adjusted using one of two empirical relationships. The first being the correlation between T_m values of the two surfactants, reported in *Chapter 4.6.3* and the second, the relationship between the respective $T_{1/2}$ values described in *Section 5.4.2*. These correlations are described by *Equations 6-7* below respectively. Resultant values were plotted, as shown in *Figure 115*, as a function of actual recorded experimental values for $T_{1/2}$ from 'Dataset 2'. These data can be found in *Table 29*.

$$\boldsymbol{T_{m\,LAS} = 0.99(T_m) + 0.16} \qquad \textbf{(Eq. 6)}$$

$$\boldsymbol{T_{\frac{1}{2}LAS} = 1.34(T_{\frac{1}{2}SDS}) + 157.17} \qquad \textbf{(Eq. 7)}$$

Empirical relationships derived purely from stability data (*Equations 5 & 7*) produced predictive values for $T_{1/2}$ which aligned with the established model of stability from the plot in *Figure 112*. Values, unsurprisingly, are also in line with those obtained by applying the LAS T_m CD data set. Translating SDS T_m values based on the relationship between thermal stabilities of LAS and SDS (*Equation 7)*, however, gave predictions for $T_{1/2}$ which more accurately represent those observed experimentally. Comparisons of predicted trendlines from each method are presented in *Figure 115*. The green trendline represents the case where predicted values are exactly equal to those observed experimentally.

The improved precision achieved by studying trends in T_m values highlights the advantages of incorporating knowledge derived from thermal stability data in understanding and predicting enzyme inactivation processes. Furthermore, SDS analogous solutions were shown to be more effective than direct analysis of samples in LAS, supporting the replacement of labour-intensive methods of purification with *in situ* analyses.

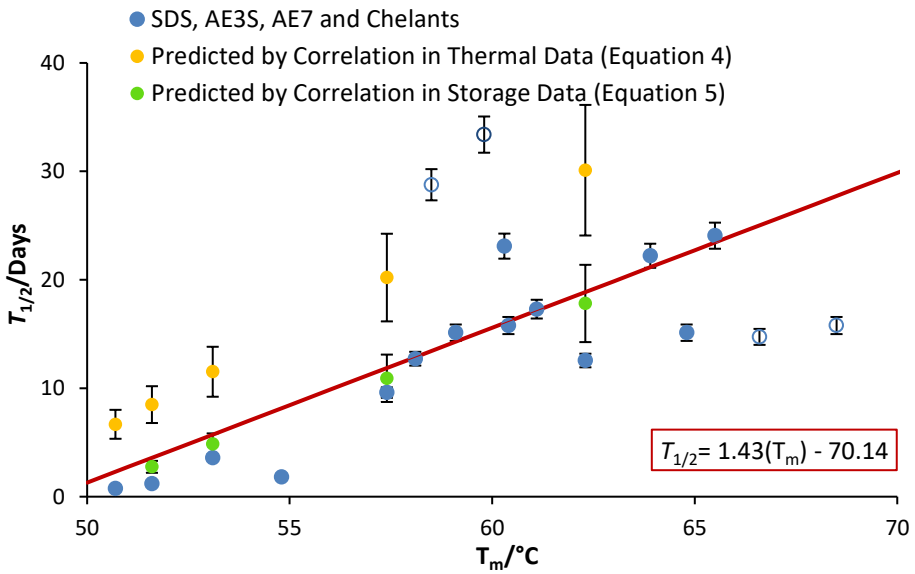


Figure 114: Empirical fitting of V42 storage stability values as a function of respective T_m CD values including predictions of $T_{1/2}$ values in the presence of LAS based on SDS T_m values and observed correlations between SDS and LAS thermal data (yellow) and SDS and LAS storage data (green)

Table 29: Comparison of experimental[a] and predicted values for $T_{1/2}$ of V42 in LAS

Surfactant Concentration	Observed $T_{1/2}$ (Days)	Predicted $T_{1/2}$ Adjusted with *Equation 6*	Δ $T_{1/2}$	Predicted $T_{1/2}$ Adjusted with *Equation 7*	Δ $T_{1/2}$
0.1% LAS	24.0	30.1	+ 5.9	17.8	- 6.2
1% LAS	17.3	20.2	+ 2.9	10.9	- 6.4
5% LAS	15.2	11.5	- 4.3	4.9	-10.3
10% LAS[b]	-	8.5	-	2.8	-
20% LAS	5.2	6.7	+1.5	1.5	- 4.7

[a]*Experimental values from preliminary study, conducted independent of other work.*
[b]*No value for the half-life of V42 in 10% LAS as data was too scattered to fit an exponential.*

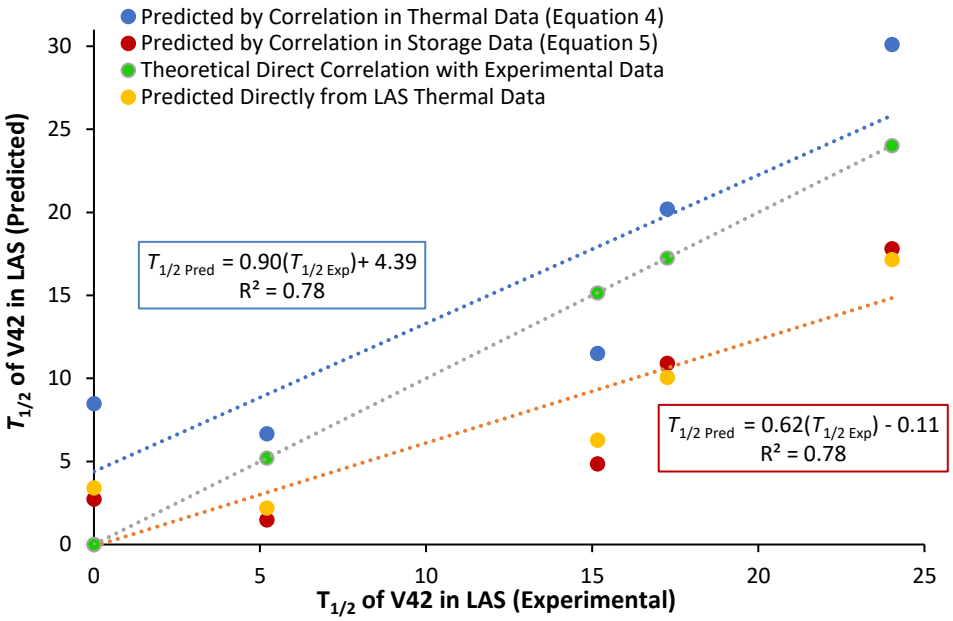


Figure 115: Agreement of V42 $T_{1/2}$ values in LAS predicted by T_m values of analogous SDS samples, with those predicted by LAS T_m values. 'Theoretical correlation' (green), representing an exact prediction of experimental $T_{1/2}$ values, has been included as a guide.

5.8 Results - Accelerated Storage Stability Study of Everest

Identical experiments were conducted in parallel for the amylase, Everest. A summary of experimental half-lives is provided in *Table 30*. Plots of enzyme activity as a function of storage time, used to calculate these values, can be found in *Appendix 3.3*. Stabilities will be discussed in terms of half-life for simplicity and rounded, where appropriate, to three significant figures.

Table 30: Experimental values for the half-life of Everest under a range of detergent conditions.[a,b,c]

Excipient	Half-Life/Days	Error (Curve Fitting)[d]	Error (Systematic)[e]
Control	470	± *23.9*	± *47.2*
0.1% LAS	146	± *33.2*	± *29.2*
1% LAS	74.6	± *9.0*	± *14.9*
5% LAS	37.3	± *0.7*	± *7.5*
10% LAS	39.0	± *1.8*	± *7.8*
20% LAS	20.6	± *2.9*	± *4.1*
0.1% SDS	290	± *11.8*	± *29.2*
1% SDS	445	± *54.8*	± *44.4*
5% SDS	635	± *62.4*	± *63.5*
10% SDS	325	± *148.2*	± *32.5*
20% SDS	325	± *8.8*	± *32.5*
0.1% AE7	420	± *20.7*	± *41.9*
1% AE7	295	± *38.9*	± *29.5*
5% AE7	300	± *10.3*	± *30.2*
10% AE7	180	± *11.1*	± *18.4*
20% AE7	165	± *10.4*	± *16.5*
0.1% AE3S	105	± *6.0*	± *10.4*
1% AE3S	45.9	± *1.6*	± *4.6*
5% AE3S	41.6	± *2.3*	± *4.2*
10% AE3S	12.0	± *0.7*	± *1.2*
20% AE3S	5.3	± *0.2*	± *0.5*
HEDP	25.3	± *1.3*	± *0.0*
EDTA	0.2	± *0.0*	± *2.5*
Citric Acid	0.1	± *0.0*	± *0.0*
Fatty Acid	2.1	± *0.2*	± *0.2*

[a]Activity determined using ethylidene-paranitrophenol-Glucose-7 (EPS) assays to assess rate of substrate conversion at each time point. [b]Activity monitored at accelerated rates over an 8-week period. [c]Rates of degradation and half-life calculated based on initial rates. [d]Error values calculated based on the standard error of curve fitting.

Storage stability was higher in general for Everest than for V42, with a half-life in buffered enzyme control samples of 470 days compared to 22 days. This corresponds with high T_m values observed in thermal degradation studies, where control samples exhibited $T_{m\ CD}$ values of over 90 °C, compared with less than 65 °C for V42. Amylase activity levels of Everest samples were, however, far lower than those of V42, as lower concentrations were used to keep in line with commercial enzyme levels (*Methods*, *Chapter 2.1*).

All formulations were found to have a negative impact on Everest stability. This is contrary to increases in $T_{1/2}$ values observed for V42 under several detergent conditions such as 0.1% LAS, high concentrations of AE3S and in HEDP. $T_{m\ CD}$ values for Everest, also predicted that several formulations would have a positive effect on stability. These included 0.1% LAS, both AE3S and AE7 and the chelator HEDP. These effects will be discussed in an in-depth analysis of each formulation in the following sections.

5.9 The Effects of LAS on Everest Storage Stability

As illustrated in *Figure 116,* trends in Everest stability in the presence of LAS were more clearly defined than those of V42. The amylase experienced a significant destabilisation in the presence of 0.1% LAS with a reduction in $T_{1/2}$ from 470 days in the control to less than 150 days. This downward trend continued with subsequent increases in surfactant, but with smaller changes in half-life between samples. At 1% LAS $T_{1/2}$ of 75 days was observed, while concentrations of 5%-10% yielded similar values of 37-39 days. A further decrease to 21 days was observed at 20%

LAS. These trends were significant with respect to the estimated error of 10-20%, as shown with error bars below.

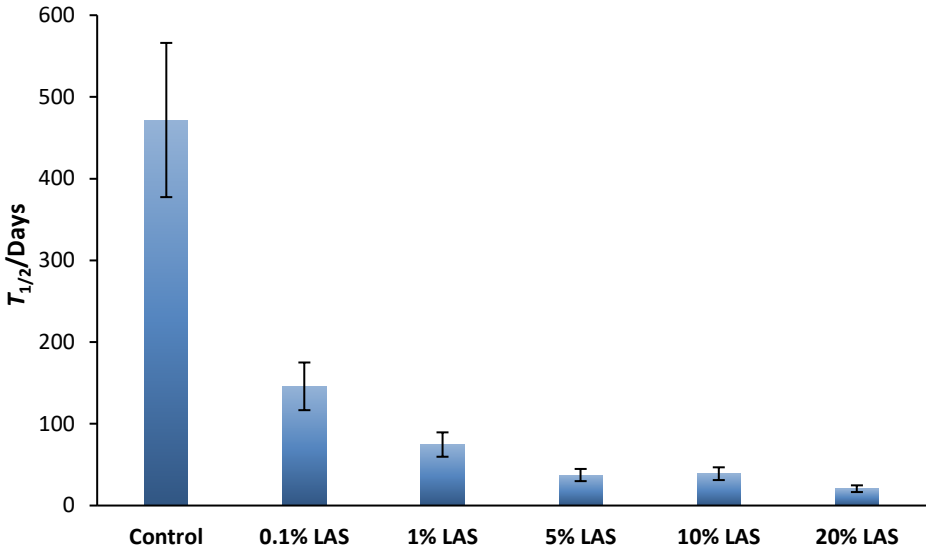


Figure 116: Chart of $T_{1/2}$ values for Everest in the presence of various concentrations of LAS as determined by accelerated storage tests. Error bars show an estimated margin of error of 20% for opaque viscous solutions analysed photometrically.

This trend is in line with predictions from CD thermal analysis. A linear correlation was initially observed between the $T_{1/2}$ and T_m values of LAS-rich samples, with an R^2 of 0.91 *(Table 31, Figure 117)*. The T_m values from *Chapter 4.6* indicated that V42 and Everest in should both show increasing LAS causing a decrease in stability. This further validated the use of 'Data Set 2' rates in place of the original dataset for V42 in LAS, as had previously been supported by similar trends induced by SDS *(Section 5.4.1)*.

The observed linear relationship between T_m and $T_{1/2}$ shown below suggests that, as demonstrated for V42, storage stability in amylases can be determined based on thermal stability data. In contrast to the protease, however, the control sample is outlying from the trend and has been omitted from the plot. Data is also more clustered than that seen previously, with much of the trend arising from the distance of 0.1% LAS from the other datapoints, resulting in lower estimations of deviation from the fit. As a result, the accuracy of $T_{1/2}$ determination, based on this

fitting, may be affected. This will be explored further, in relation to the complete amylase dataset in *Section 5.14*.

Table 31: Comparison of experimental T_m and $T_{1/2}$ values for Everest in varying concentrations of LAS

LAS Concentration	$T_{1/2}$ / Days[a]	T_m / °C[b]
Control	470 *(± 24)*	86.7 *(± 0.6)*
0.1% LAS	145 *(± 33)*	90.9 *(± 1.1)*
1% LAS	74.6 *(± 9.0)*	86.9 *(± 0.8)*
5% LAS	37.3 *(± 0.7)*	79.0 *(± 0.6)*
10% LAS	39.0 *(± 1.8)*	80.9 *(± 5.4)*
20% LAS.	20.6 *(± 2.9)*	79.4 (± 2.6)

[a]T_m values determined using in situ heating and CD analysis. Values are listed in °C. [b]$T_{1/2}$ values determined through storage tests and PNA activity assays. Values are listed in days.

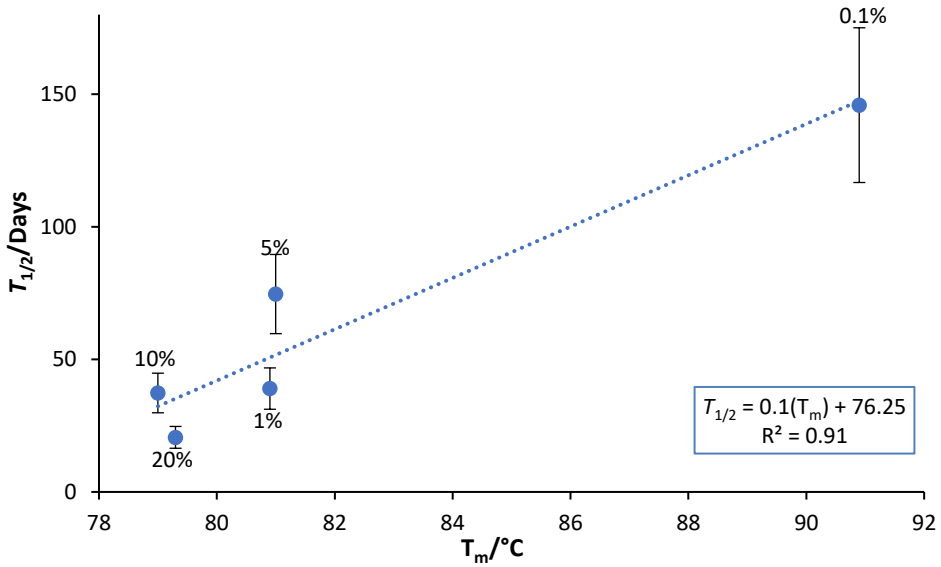


Figure 117: Plot of Everest $T_{1/2}$ values in the presence of various chelating agents as a function of their respective T_m values

5.10 The Effects of SDS on Everest Storage Stability

As demonstrated in *Figure 118*, a similar trend between half-life and surfactant concentration could not be established for SDS. Storage data also did not correlate well with respective T_m values collected for various SDS concentrations. A plot of $T_{1/2}$ as a function of T_m indicates a general upward trend, with an outlying value at 0.1% SDS. Data is too scattered,

however, to conclusively determine whether a direct correlation exists *(Table 32, Figure 119).* T_m data for Everest was found to be less consistent than that of V42 in the previous chapter, perhaps due to its high stability. It is likely that this repeated variability in storage data may also be an artefact of this property, as the timescale of the tests could be too short to be significant in the context of long stability. The lower activity levels reported due to the small concentrations of amylase used in HDL, may also have contributed to this error. Analysis of higher concentrations of enzyme, as demonstrated with V42 improves the robustness of the technique.

Table 32: Experimental values for $T_{m\ CD}$ and $T_{1/2}$ for Everest in varying concentrations of SDS.[a,b]

SDS Concentration	Everest $T_{1/2}$	Everest T_m
Control	470 *(± 24)*	86.7 *(± 0.6)*
0.1%	290 *(±12)*	89.9 *(± 2.3)*
1%	440 *(± 55)*	87.7 *(± 1.5)*
5%	635 *(± 62)*	83.8 *(± 1.2)*
10%	325 *(±148)*	84.6 *(± 1.2)*
20%	325 *(± 8.8)*	80.1 *(± 0.0)*

[a] T_m values determined using in situ heating and CD analysis. Values are listed in °C. [b] $T_{1/2}$ values determined through storage stability tests. Values are listed in days.

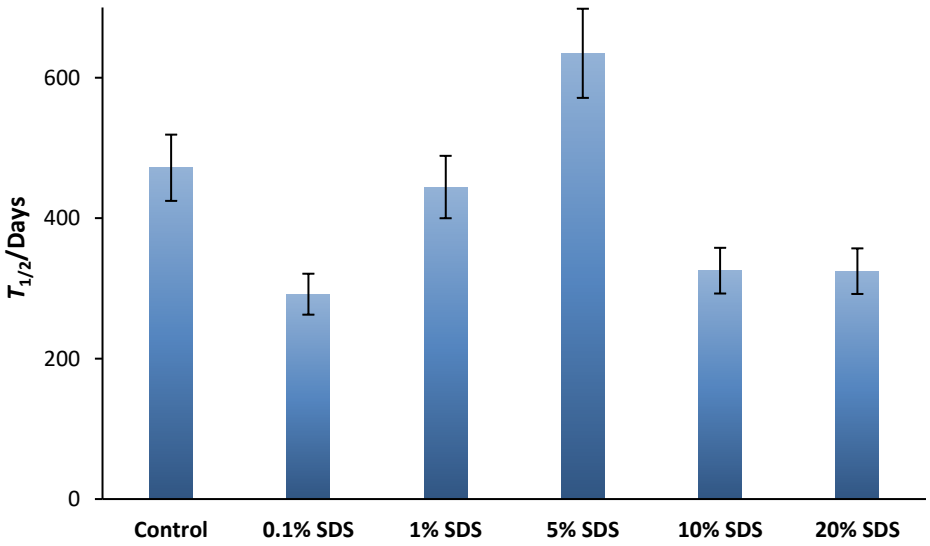


Figure 118: Chart of $T_{1/2}$ values for Everest in the presence of various concentrations of SDS as determined by accelerated storage tests. Error bars represent estimate of 10% variation on reported values.

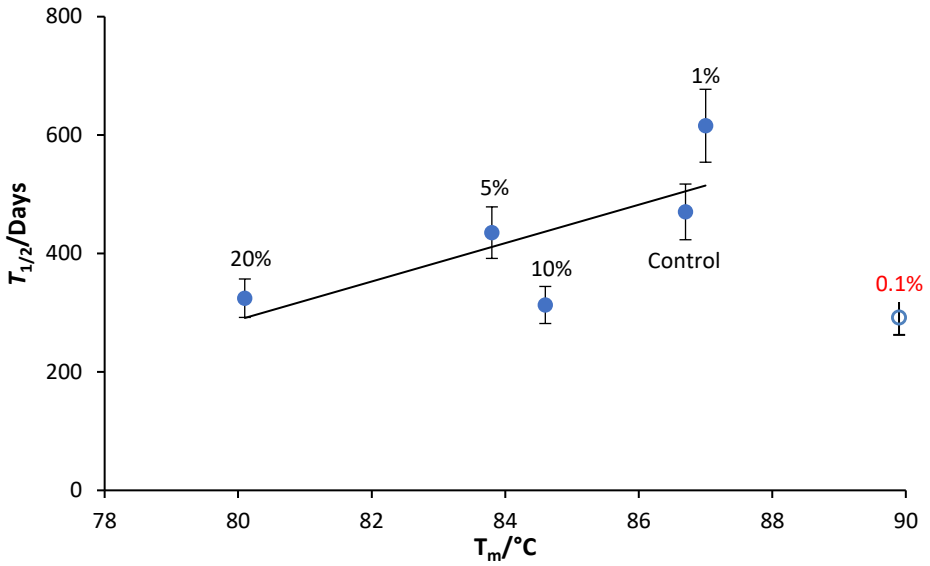


Figure 119: Plot of $T_{1/2}$ values for Everest in the presence of various concentrations of SDS as a function of their respective T_m values.

5.11 The Effects of AE3S on Everest Storage Stability

A substantial increase in the rate of degradation was observed in the presence of the third anionic surfactant, AE3S *(Table 33, Figure 120)*. At 0.1% w/v of surfactant, the half-life of Everest was reduced from 470 days in the control sample to 104 days, similar to the level of destabilisation induced by LAS. On titration of the surfactant, $T_{1/2}$ values of 46 days at 1%, 42 days at 5%, 12 days at 10%, and 5.3 days at 20% AE3S were reported. This trend reflects those of the other anionic surfactants on amylase and protease samples, though here the rates are far higher.

V42 samples, in contrast, showed a positive association between AE3S concentration and storage stability. AE3S is generally added to improve formulation stability, making the trends observed for Everest appear counter-intuitive. Weak chelating effects have, however, been reported for micelles of AE3S *(Figure 6, Chapter 1.2.1)*. As the CMC of the surfactant is quite low, the observed destabilisation may be an effect of calcium sequestration from the amylase. Subtilisins are more resilient to loss of Ca^{2+} ions than α-amylases which would account for the observed difference in effects. Partial unfolding of the protein increases the rate of

surfactant binding, resulting in co-operative degradation. This synergy may explain for the increased rate of denaturation when compare to other anionic surfactants.

Table 33: Experimental T_m and $T_{1/2}$ values for Everest in various concentrations of AE3S.

Surfactant Concentration	$T_{1/2}$ (AE3S)[a]	T_m (AE3S)[b]
Control	470 *(± 24)*	86.7 *(± 0.6)*
0.1%	105 *(± 6.0)*	95.0 *(± 0.1)*
1%	45.9 *(± 1.6)*	93.1 *(± 0.9)*
5%	41.6 *(± 2.3)*	91.4 *(± 0.5)*
10%	12.0 *(± 0.7)*	84.0 *(± 1.1)*
20%	5.3 *(± 0.2)*	-

[a] $T_{1/2}$ values determined through storage tests and activity assays. Results are listed in days. [b] T_m values determined using in situ heating and CD analysis. Results are listed in units of °C.

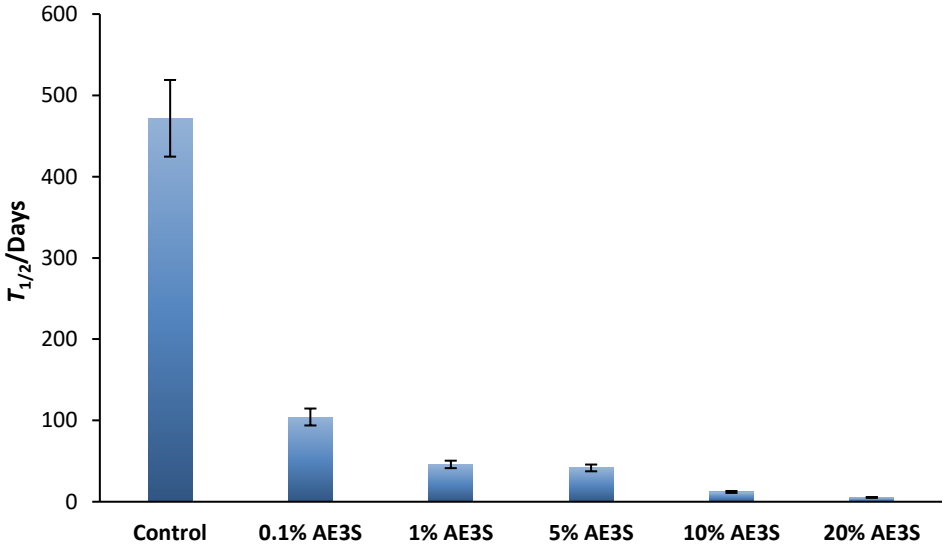


Figure 120: Chart of $T_{1/2}$ values for Everest in the presence of various concentrations of AE3S as determined by accelerated storage tests. Error bars represent 10% error estimation for lower viscosity, transparent samples.

Plotting $T_{1/2}$ as a function of respective T_m values for AE3S also yielded a linear correlation, with an R^2 value of 0.85 *(Figure 121),* supporting trends in stability data reported above. The small degree of scatter from the fit indicates that fluctuating T_m values with increasing surfactant concentration *(Chapter 3.2.3),* are insignificant in the context of long-term stability. This highlights the barriers to clear understanding of

protein inactivation processes, arising from reliance on lone analytical techniques.

The control sample was omitted from data fitting as it was an extreme outlier from the AE3S trend. This was also the case for LAS samples, suggesting that empirical fitting of T_m against $T_{1/2}$ may not be consistent for all conditions. This may limit the ability to use a single formula for conversion of T_m values to expected $T_{1/2}$ values, particularly in extrapolation to multi-component systems.

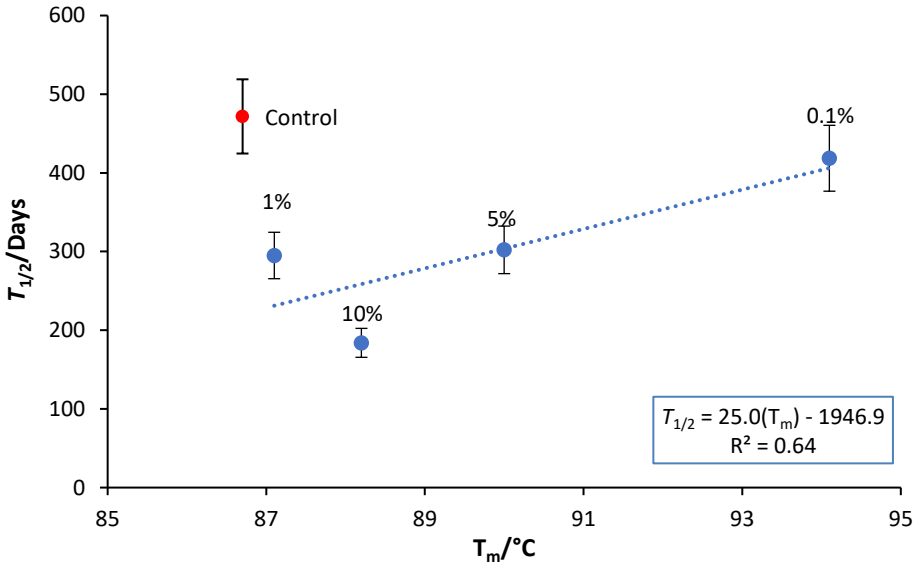


Figure 121: Plot of $T_{1/2}$ values for Everest in the presence of various concentrations of AE3S, as a function of their respective T_m values. Error is reported as the error in the T_m arising from data fitting to an exponential decay curve.

5.12 The effects of AE7 on the Storage Stability of Everest

A similar downward trend in stability was observed in the case of the non-ionic surfactant AE7 *(Figure 122)*. The extent of destabilisation was less than that induced by LAS or AE3S, however, with a maximum reduction in $T_{1/2}$ of ~290 days at 10% AE7. Regardless, the trend is significant with respect to the error estimate of 10-15%.

All concentrations of surfactant exhibited lower T_m values when compared to that of the control. At 0.1% AE7, however, this effect negligible. Both 1% and 5% AE7 samples exhibited similar levels of

stability, with values for $T_{1/2}$ of ~300 days, a reduction in $T_{1/2}$ of ~170 days, with respect to the control. 10% and 20% AE7 also exhibited similar levels of stability, with $T_{1/2}$ values of 165 and 185 days respectively. This stepwise reduction in half-life with increasing surfactant concentration may indicate that micellar arrangements influence protein unfolding, however, this is difficult to determine without full investigation into surfactant aggregation phases in protein-rich formula. Again, this was contrary to a general upward trend in stability observed in V42 for increasing AE7 concentrations.

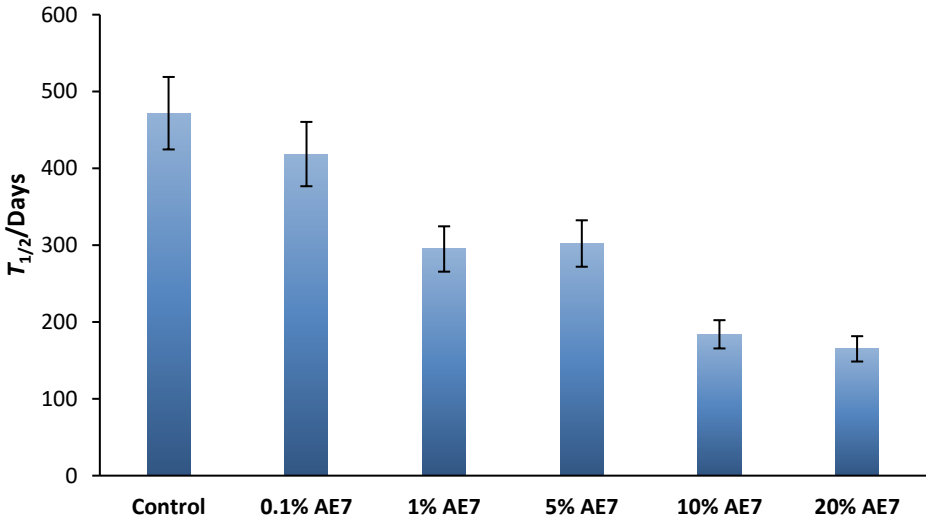


Figure 122: Chart of $T_{1/2}$ values for Everest in the presence of various concentrations of AE7 as determined by accelerated storage tests. Error bars represent 10% error estimation for lower viscosity, transparent samples.

A plot of $T_{1/2}$ values against their respective T_m values *(Table 34)* yielded similar results to that of AE3S as illustrated in *Figure 123*. Again, the control was outlying from the rest of the dataset, preventing a direct correlation from being determined.

Table 34: Comparison of experimental T_m and $T_{1/2}$ values for Everest in various concentrations of AE7.

Surfactant Concentration	$T_{1/2}$ (AE7)[a]	T_m (AE7)[b]
Control	470 *(± 24)*	86.7 *(± 0.6)*
0.1%	420 *(± 21)*	94.1 *(± 1.9)*
1%	295 *(± 39)*	87.1 *(± 0.4)*
5%	300 *(± 10)*	90.0 *(± 0.7)*
10%	185 *(± 11)*	88.2 *(± 1.9)*

[a]$T_{1/2}$ values determined through storage tests and PNA activity assays. Results are listed in days. [b]T_m values determined using in situ heating and CD analysis. Results are listed in units of °C.

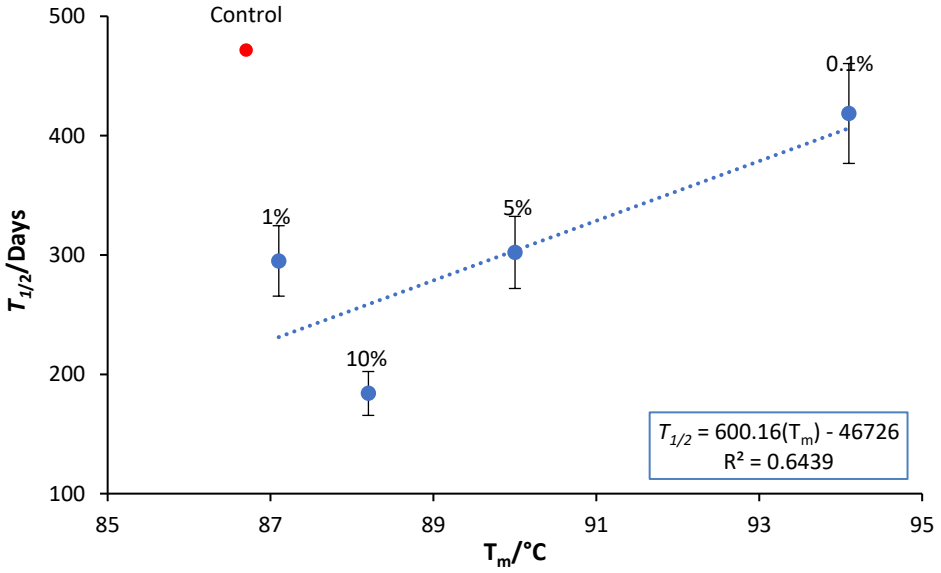


Figure 123: Plot of $T_{1/2}$ values for Everest in the presence of various concentrations of SDS as a function of their respective T_m values

5.13 The effects of Chelators and Builders on the Storage Stability of Everest

Chelants and builders induced the highest rates of inactivation for Everest of all the conditions tested, reflecting thermal observations from CD, DSC and DSF from *Chapter 3*. Everest was also far more susceptible to destabilisation by chelating agents than V42. The is a direct result of the absence of a secondary Ca^{2+} binding site in the amylase, resulting in a weak association with vital structural calcium.[12] Concentrations of chelators and builders were kept constant between the amylase and

protease samples. The resulting higher molar ratio due to low Everest concentrations, may have also contributed to denaturation rates.

Similar to observations for V42, EDTA and citric acid were the most destabilising compounds resulting in $T_{1/2}$ values of just 0.1-0.2 days *(Figure 124)*. Stability in fatty acid was higher at ~2 days, and greater again in HEDP at ~25 days. Further study into the respective metal ion binding affinities of these chelating agents and the proteins in question is required to explain these results.

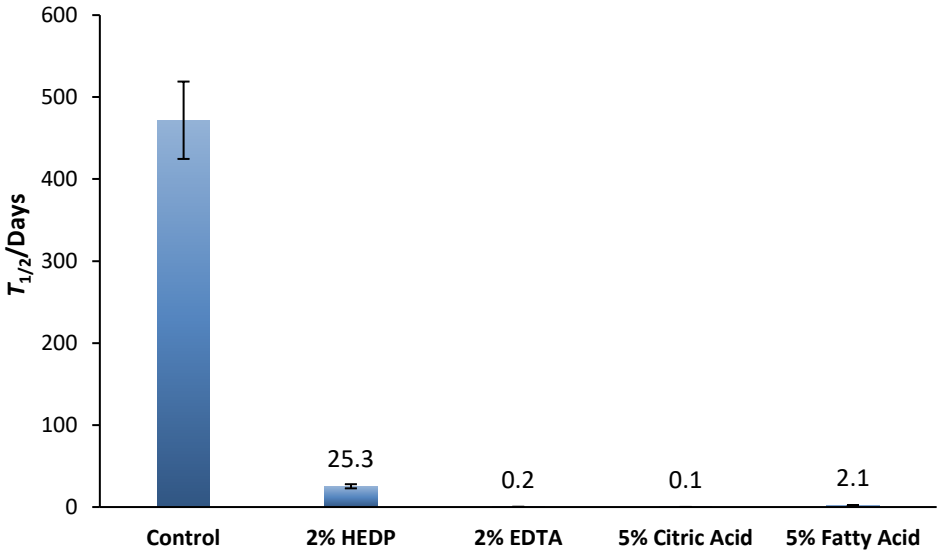


Figure 124: Chart of $T_{1/2}$ values for Everest in the presence of various chelating agents, as determined by accelerated storage tests.

A direct correlation between $T_{1/2}$ and T_m values, in the presence of chelants, could not be established. This may be an artefact of the broadly similar values observed for EDTA and the builders, particularly with respect to the control and HEDP *(Figure 125)*.

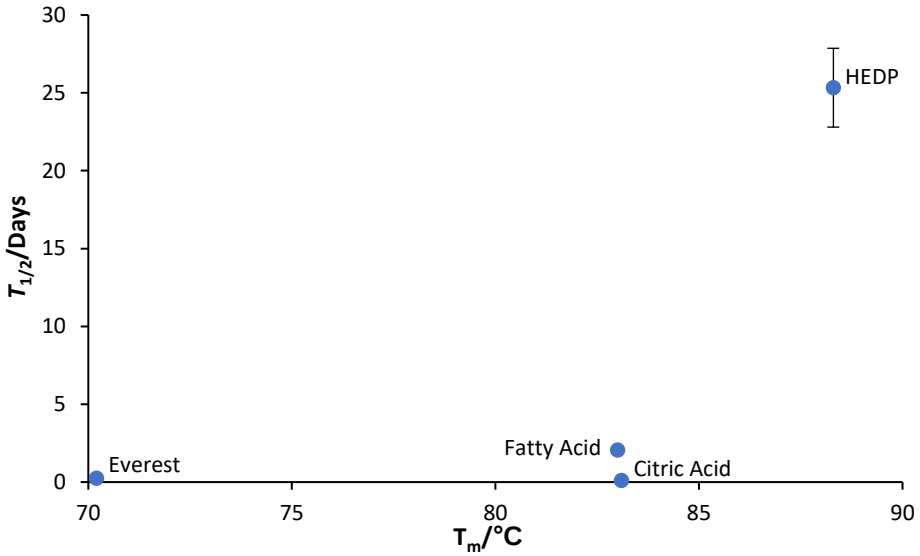


Figure 125: Plot of Everest $T_{1/2}$ values in the presence of various chelating agents against their respective $T_{m\,CD}$ values.

5.14 Establishing T_m values as Predictive Indicators of Storage Stability

An overall comparison of T_m values against $T_{1/2}$ values was conducted for Everest, in an attempt to describe the relationship between Everest thermal unfolding and storage stability across the range of detergent conditions. This mimicked empirical fitting of V42 data in *Section 5.7*. Relevant stability values are listed in *Table 35* below.

Table 35: Comparison of experimental T_m and $T_{1/2}$ values for Everest under a range of detergent conditions.

Excipient	Half-Life/Days	Error in $T_{1/2}$[d]	T_m/ °C[e]	Error in T_m
Control	11322.2	± *574.8*	86.7	± *0.6*
0.1% LAS	3500.7	± *797.9*	90.9	± *1.1*
1% LAS	1791.1	± *215.9*	81.0	± *0.8*
5% LAS	895.5	± *16.7*	79.0	± *0.6*
10% LAS	935.4	± *44.0*	80.9	± *5.4*
20% LAS	494.0	± *70.3*	79.3	± *2.6*
0.1% SDS	7001.5	± *282.6*	89.9	± *2.3*
1% SDS	15234.0	± *1497.4*	87.7	± *1.5*
5% SDS	10663.8	± *1315.5*	83.8	± *1.2*
10% SDS	7805.7	± *3557.7*	84.6	± *1.2*
20% SDS	7788.2	± *210.2*	8.1	± *0.0*
0.1% AE7	10045.6	± *496.2*	94.1	± *0.1*
1% AE7	7080.2	± *934.2*	87.1	± *0.9*
5% AE7	7251.3	± *246.8*	90.0	± *0.5*
10% AE7	4415.0	± *265.3*	88.2	± *1.1*
0.1% AE3S	2502.3	± *144.1*	95.0	± *1.9*
1% AE3S	1102.0	± *37.4*	93.1	± *0.4*
5% AE3S	998.8	± *55.1*	91.4	± *0.7*
10% AE3S	288.8	± *15.7*	84.0	± *1.9*
HEDP	5.7	± *0.1*	70.2	± *1.1*
EDTA	608.0	± *30.4*	88.3	± *2.2*
Citric Acid	2.5	± *0.2*	83.1	± *1.7*
Fatty Acid	49.5	± *4.2*	83.0	± *3.0*

[a]Activity determined using ethylidene-paranitrophenol-Glucose-7 (EPS) assays to assess rate of substrate conversion at each time point. [b]Activity monitored at accelerated rates over an 8-week period. [c]Half-life calculated based on initial rates. [d]Error values calculated based on the standard error of curve fitting. [e]T_m values determined by CD in-situ temperature ramp for all samples except LAS, determined by CD following purification. Reported values are an average of 3 independent analysis, error is listed as the standard error from the mean of these analyses.

The cumulative data for Everest under various conditions of surfactant and chelant could not be described using a single regression model. Trends within single excipient groups did not show obvious trends, such as those seen for V42 in *Sections 5.5-5.7*. The control sample also consistently presented as an outlier in these single excipient plots. The

lack of clear correlations has been attributed to systematic error arising from the elongated lifetimes and lower enzyme concentrations for amylase samples, when compared to those of the protease. Increasing the concentration of amylase in stability study samples, with respect to commercial HDL levels, may improve the robustness of the method. This is common in thermal analysis methods as higher concentrations are required to meet sensitivity limits. As detergent compounds such as surfactants are in such great excess when compared to enzyme concentration, protein-surfactant interactions should not be affected.

A consistent empirical fitting was required to translate SDS T_m values to predictive $T_{1/2}$ values in LAS. As no single function describing the relationship between $T_{1/2}$ and T_m could be established, it was not possible to further validate the use of analogous SDS formulations.

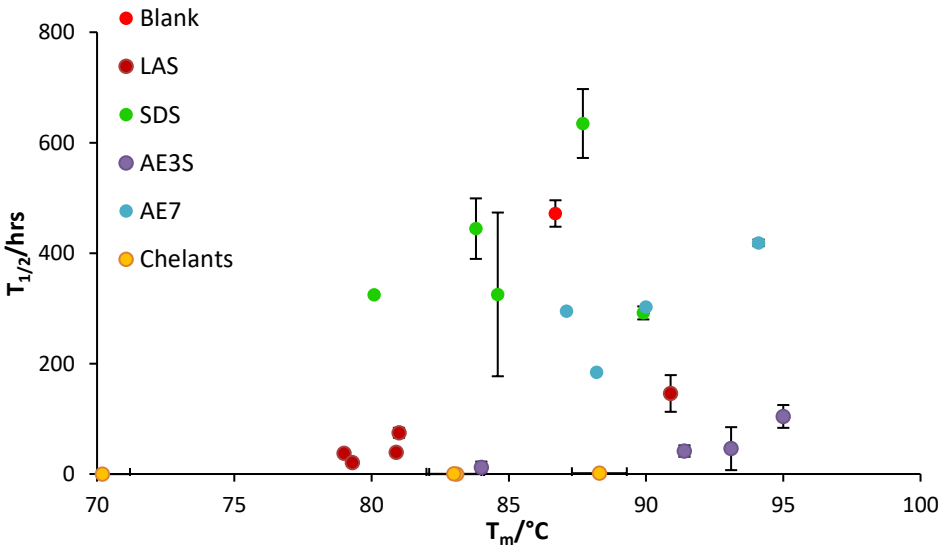


Figure 126: Attempted fitting of $T_{1/2}$ as determined through storage tests with EPS assay and T_m values determined by CD. Detergent components are grouped by colour.

High thermal stability reported for Everest is a second potential source of error, as several T_m values were beyond the temperature limit of the CD instrument. Linear correlations established by Lund avoided this issue through the use of pressurised cells in nano-DSC. Available Everest $T_{max\ DSC}$ data from *Chapter 3.3.1,* produced a linear correlation with storage values, as shown in *Figure 127*. This suggests that the lower throughput

method may be necessary in accurately determining T_m values for high stability enzymes. The reported correlation with DSC data was higher than that reported for V42 using CD data. The limited number of samples, mainly consisting of a single excipient range (LAS) contributes to reduced data scatter, however, and so further examples may be required to verify this fitting.

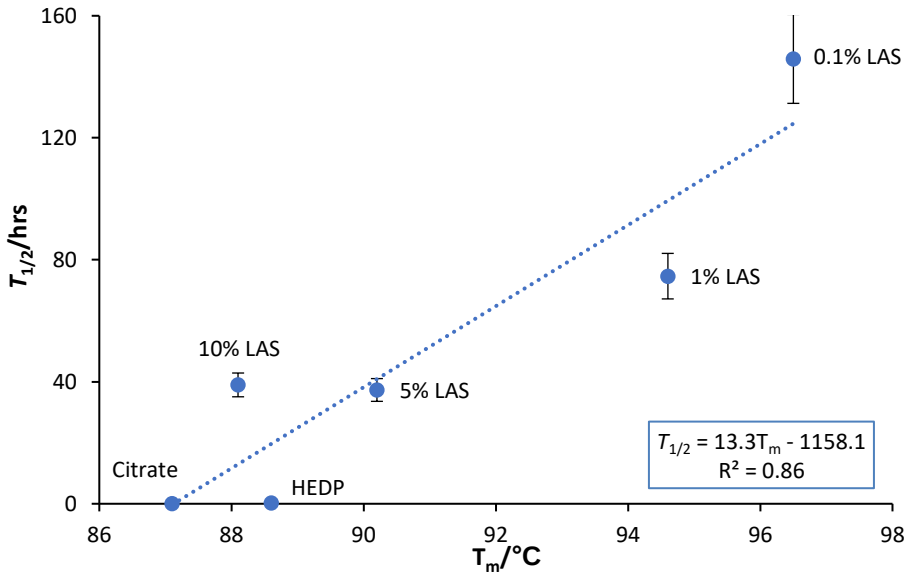


Figure 127: $T_{1/2}$ values for Everest in various concentrations of LAS as a function of T_m DSC values.

5.15 Conclusions

Work in this chapter describes the capabilities of CD as a tool for thermal denaturation analysis as a predictor of the storage stability of enzymes under simple detergent conditions. For V42, clear correlations observed between experimental T_m and $T_{1/2}$ values, in the presence of a range of detergent components, indicate a link between the two parameters. Empirical fitting of these data provided linear regression models which demonstrated the predictive ability of T_m values with respect to long-term storage stability of enzymes under given conditions.

Similar trends could not be identified among Everest data. Several factors causing variability among amylase data may have contributed to the inconclusive results. These include the intrinsic long storage stability of detergent amylases, the use of low enzyme concentrations and T_m values which approach the limit of instrumental heating ranges. Consequentially, amylase samples require further work to determine definitively if T_m values can be used as predictors of $T_{1/2}$, and subsequently to develop regression models.

Collected data for Everest did, however, highlight several key observations. Systematic error in storage tests and activity assays is far higher than that of thermal analysis methods, highlighting the need for transfer of stability testing to more precise, reliable techniques. Both thermal and stability data were required to explain several excipient effects, however, as trends were unclear using a single method in isolation. A suite of analytical techniques would therefore provide the best insight into enzyme-excipient interactions. This is likely to be emphasised on incorporation of more complex formulations.

In general, the presence of surfactants reduced enzyme shelf life, with effects intensifying at higher concentrations. V42 samples in the presence of AE3S and AE7 were the exception to this observation, with maintained or improved levels $T_{1/2}$ values on titration of surfactant. The clear destabilisation of Everest under the same conditions is likely a result of

chelant effects reported among surfactant micelles.[11] Serine proteases are more resistant to calcium chelation due to the presence of a second Ca^{2+} binding site which stabilises the native state.

These observations will be discussed in relation to those of earlier chapters in the following 'Conclusions' chapter.

5.16 Future work

Initial further work in this area should focus on improving amylase data quality to verify that the relationship between $T_{1/2}$ and T_m is consistent for all excipients. Resultant data may then be used to establish formulae for the translation of thermal stability parameters to storage stability estimates, as shown for V42. High variability in amylase data may be reduced through selection of an enzyme with a lower stability profile such as Natalase. Furthermore, these experiments should be repeated for additional proteases and amylases to determine if established regression models are relevant within an enzyme class, or if individualised fitting to of T_m to $T_{1/2}$ data is required.

Once procedures have been established in these single component systems, analysis should be extrapolated to include multi-component data, with a view to eventual application to fully formulated HDL. Emphasis should be placed on accurate thermal analysis in complex systems, as well as probing synergetic effects of multiple excipients. Successful validation would provide for the introduction of stability models to an industrial setting, reducing reliance on storage testing.

As several unexpected trends were observed among the data, further study into the mechanisms behind inactivation processes is required. This should include detailed analysis that takes account of the factors CMC values and association constants. The complex nature of these formulations alters properties generally quoted in the literature. Improved understanding of these interactions would support transfer to

more deliberate, upstream design of new formulations, rather than the trial and error methods currently employed. This would further reduce pressure on high throughput screening, switching focus to more insightful methods. Furthermore, work into the Ca^{2+} binding of specific detergent proteins may explain trends seen in the presence of the chelating agents discussed in this chapter. This should be accompanied by a study of the speciation and binding affinities of the chelants and builders in detergent formulations.

5.1 References

1. L. Burton, R. Gandhi, G. Duke and M. Paborji, *Pharm. Dev. Technol.*, 2007, **12**, 265–273.

2. D. S. Goldberg, S. M. Bishop, A. U. Shah and H. A. Sathish, *J. Pharm. Sci.*, 2011, **100**, 1306–1315.

3. A. G. Quezada, A. J. Díaz-Salazar, N. Cabrera, R. Pérez-Montfort, Á. Piñeiro and M. Costas, *Structures*, 2017, **25**, 167–179.

4. T. Menzen and W. Friess, *J. Pharm. Sci.*, 2013, **102**, 415–428.

5. H. Lund, S. G. Kaasgaard, P. Skagerlind, L. Jorgensen, C. I. Jorgensen and M. van de Weert, *J. Surfactants Deterg.*, 2012, **15**, 9–21.

6. T. B. L. Kirkwood, *J. Biol. Stand.*, 1984, **12**, 215–224.

7. R. T. Magari, K. P. Murphy and T. Fernandez, *J. Clin. Lab. Anal.*, 2002, **16**, 221–226.

8. M. C. Crossin, *J. Am. Oil Chem. Soc.*, 1989, **66**, 1010–1013.

9. F. Diagnostics, *Infinity™ Amylase Liquid Stable Reagent*, 2008, Retrieved from: https://assets.thermofisher.com/TFS-Assets/CDD/manuals/Infinity-Amylase-Reagent-EN.pdf

10. M. T. Henzl, J. D. Larson and S. Agah, *Anal. Biochem.*, 2003, **319**, 216–233.

11. H. Lund, S. Kaasgaard, P. Skagerlind, L. Jorgensen, C. Jørgensen and M. van de Weert, *J. Surfactants Deterg*, 2012, **15**, 265–276.

12. R. P. Enever and N. Pilpel, *Trans. Faraday Soc.*, 1967, **63**, 1559–1566.

13. M. W. Pantoliano, M. Whitlow, J. F. Wood, M. L. Rollence, B. C. Finzel, G. L. Gilliland, T. L. Poulos and P. N. Bryan, *Biochemistry (Mosc.)*, 1988, **27**, 8311–8317.

6 Conclusions and Future Work

The work presented in this thesis describes novel methods of probing protein-surfactant interactions in liquid laundry detergents. These approaches address the prevailing issue of high concentrations of the surfactant, LAS, which has hindered the application of high throughput analysis methods common in other protein-based industries. Furthermore, by establishing functions which relate thermal unfolding to enzyme half-life in storage, evidence to support predictive modelling of storage stability from T_m values has been provided.

6.1 Capabilities of Common Protein Analysis Methods with Detergent Samples

Results from the screening of protein analysis methods were discussed in *Chapter* 3. Each technique presents a unique set of advantages and limitations which have been summarised in *Table 1* below. To date, DSC has been the primary method for the analysis of enzymes in laundry formulations. We confirmed that this method was the most suitable for LAS-rich sample media as T_m determination was possible at up to 10% w/v of surfactant, compared with 0.1% w/v for CD and DSF. The small energy changes associated with protein unfolding, however, necessitated the use of nano-DSC to achieve the required sensitivity. Limited access to such instrumentation, coupled with low throughput of the method, drove work towards the development of optical methods as tools for stability analysis. These provide much higher sample throughput and greater insight into unfolding processes. Direct correlations observed between $T_{m\ DSF}$, $T_{m\ CD}$ and $T_{max\ DSC}$ values suggest that it should be possible to replicate successful prediction of enzyme storage stability from DSC data using CD or DSF.

Table 1: Summary of the advantages and limitations of a range of protein analysis techniques when applied under HDL conditions.

Technique	Output	Advantages	Limitations
DSF	T_m – Δ Hydrophobicity	• High throughput • Automated temperature ramp with qPCR	• Dye-surfactant interactions • Max 0.1% LAS (UV Detector)
CD	T_m – Δ Helicity & α-helix/β-sheet content	• Provides structural info • Automated temp. ramp	• Max 0.1% LAS (UV Detector)
Nano-DSC	T_{max} – Δ Enthalpy	• T_{max} analysis at up to 10% LAS • Automated temp. ramp	• Requires sensitive equipment • Poor results at high viscosities
FastPP	T_{max}– Resilience to proteolysis (based on unfolding)	• No specialist equipment	• Low precision and accuracy • Max 0.1% LAS • Manual heating to each temp. point (Labour-intensive)
MST	T_m – Δ Rate of Thermophoresis	• High Throughput • Can detect intrinsic fluorescence	• Max 0.1% LAS (UV Detector)

6.2 Analysis of Protein in LAS-rich media using Optical Detection

The limiting factor to the use of optical methods was detector saturation, caused by the high concentrations of UV-active LAS. This was overcome by developing alternative approaches to probing LAS-rich systems as described in *Chapter 4*. The first of these methods involved the use of the analogous surfactant, SDS, which has comparable effects on protein stability, but lower UV absorptivity. The second focused on the removal of LAS following sample incubation, prior to structural and T_m analysis. Protein unfolding of detergent enzymes was found to be irreversible, permitting this purification step in place of direct analysis of the heated sample. Removal of LAS through precipitation with $CaCl_2$ was found have

the lowest impact on sample structure, however, there is great scope for the use of other, higher throughput modes of purification.

Both of these methods enabled T_m determination of proteins in samples at up to 20% LAS, using CD. DSF presented further challenges to analysis in surfactant as dye-surfactant interactions induced fluorescence which masked unfolding signals. In the literature, this is addressed through the use of molecular rotor dyes, which fluoresce in response to the inhibition of free rotation of bonds in the molecule.[1] This would be triggered by the aggregation of denatured protein which forms precipitate in solution. In this case, DSF would far exceed the throughput of other methods by using 96-well plate formats. With industrial scale up, ultra-DSF has the scope to analyse almost 500 conditions in parallel.[2]

Although CD did not offer the same level of throughput as DSF, incorporating a sample-changer would improve efficiency tenfold compared to DSC. The key advantages of this technique however, lie its use of intrinsic protein properties (asymmetry) and the level of structural information provided. Intrinsic properties remove reliance on expensive and interfering dyes, while spectral deconvolution can provide insight into unfolding processes. More rigorous comparisons of detergent conditions can therefore be achieved. Spectral analysis was used in this thesis to establish SDS as an analog for LAS in stability studies. Resultant validated procedures can then be applied to higher throughput methods such as DSF.

6.3 Validation of Alternative Approaches to T_m Analysis in the Presence of LAS

The development and validation of these alternative methods described in *Chapter 4,* showed that both methods produced near identical values for T_m, which also correlated linearly with $T_{max\ DSC}$ values. This indicated that the use of SDS analogs and $CaCl_2$ precipitation could be used interchangeably for T_m determination in equivalent surfactant

formulations. Further evidence to support these approaches was provided by deconvolution and principle component analysis of CD spectra, which showed comparable levels of unfolding at each temperature point. Owing to the labour-intensive nature of surfactant removal, we suggest that the procedure be reserved for further validation of analogous SDS formulations, which will be required on expansion of these methods, both to other detergent enzymes and multi-component systems. Alternatively, as both approaches rely on indirect methods to access protein-LAS interactions, an independent method such as DSC may provide a more robust means of validation.

6.4 Structural Analysis using Circular Dichroism

Empirical validation of SDS as an LAS analog was achieved by comparing enzyme T_m values in SDS with those obtained in the presence of LAS (via DSC) and $CaCl_2$ precipitation (CD). These data were further supported by conducting protein structural analysis on CD spectra collected following incubation in each of the surfactants. This was conducted initially using the 'Dichroweb' library of deconvolution programmes, and subsequently PCA to identify common conformational changes induced by each surfactant.

Software from 'Dichroweb' proved to be a valuable tool in the comparison of relative increases in the proportion of protein in disordered conformations. Limited success was achieved in assigning specific structural features, however, as results did not align with x-ray crystal structures of similar enzymes. This is presumed to be an artefact of detergent interference with CD spectra, and partial unfolding of proteins preventing comparison with reference proteins. Structural estimates also focused only on alpha-helical content and did not accurately detect turns or β-sheets. This has been reported in the literature to be a limitation of the software.[3] Variability in assigned ratios of each structural feature was

observed between various deconvolution programmes, highlighting the importance of selecting the method more appropriate to a given dataset.

As specific structures could not be assigned, PCA presents a simplified tool for comparison of denaturation processes under various conditions. Monitoring overall loss in structure across a complete spectral range provides a more comprehensive view of denaturation than observations at a single wavelength (generally 222 nm) alone. Furthermore, scoreplots highlight conditions which induce similar conformational changes which can be used to identify equivalency between formulations. Here, such analysis has been used to demonstrate the analogous nature of SDS and LAS formulations. Relative loading values assigned to each wavelength can also be used to identify structural features which are targeted by various excipients.

6.5 Validation of SDS as an Analog of LAS for Storage Stability Tests

Thermal denaturation of proteins under various conditions of buffer and pH has been demonstrated in the literature to be predictive of associated storage stability.[3–5] Early efforts towards extrapolating this work to incorporate complex laundry formations has also been described by Lund *et al*, using DSC T_{max} values as a measure of thermal stability. In order to facilitate the use of T_m CD values, obtained via either SDS analogs or $CaCl_2$ precipitations in place of DSC, validation against storage tests of respective LAS-based formulations was required.

The half-life of V42 was found to be longer in LAS than under equivalent SDS conditions. This indicates that whilst general trends in stability can be related to those of LAS, absolute values at a given concentration are non-identical. Adjusting SDS predictions based on the relationship between T_m CD values reported for the two surfactants provided more accurate estimations of experimental values. As this adjustment was based on LAS data collected using a second novel technique, $CaCl_2$ precipitation,

further verifications of this relationship using an independently validated method such as DSC should improve these predictions.

Furthermore, both the precipitation of LAS with $CaCl_2$ and the use of SDS as an analog were validated against each other. An independent dataset, such as that from DSC would provide an external reference point to confirm the link between the two surfactants, Alternatively, the precipitation method should be repeated using SDS to account for any drift in T_m as a result of the cooling and purification steps.

6.6 Prediction of Enzyme Half-Lives under Detergent Conditions through Thermal Analysis

Establishing formulae for the prediction of enzyme half-lives across the range of detergent conditions was achieved by plotting T_m CD values for all excipients. According to similar work in the literature[4–6] a single empirical fitting should describe the relationship between the two stability parameters for each condition. The equation describing this fitting then forms the basis for the prediction of storage stability for any given detergent formulation for that enzyme.

This model was successfully applied to V42, with a distinct linear correlation observed between T_m values and respective $T_{1/2}$ values, thus providing an efficient means of predicting shelf life from rapid thermal analysis tests. The exception to this trend was the set of LAS samples which were outlying due to poor quality storage data. A second data set, generated earlier in this work, produced higher quality data for LAS samples than this study of the complete range of detergent excipients. This provided sufficient evidence for proof of concept for the use of $T_{m\,CD}$ values to predict storage stability in surfactant-rich media.

Such experimental issues highlight the need for a broader range of technologies for stability analysis as storage tests can be unreliable and require extensive work to repeat. Surfactant-rich samples introduce an

added degree of systematic error due to their opacity and handling issues. Understanding of observed stability trends required both storage and thermal stability data, however, as results from a single analysis method were often unclear in isolation. We therefore recommend employing a suite of technologies to move HDL development towards a more mechanistic, 'built-in' approach to improving stability, rather than current methods of retrospective testing.

The data presented in this thesis demonstrates the potential of applying various techniques to enzymes in detergent-based media, however, further work is required to establish procedures for fully formulated HDL systems. The key challenge in expanding the above methods of predicting $T_{1/2}$ to commercial samples will lie in the accurate determination of T_m values in the highly coloured and viscous media associated with HDL. These analyses should, however, be facilitated by the approaches to LAS-rich systems developed in this thesis. Further work is also required to establish equivalent procedures for predicting the half-life of amylase, as data collection was hindered due to exceptionally long storage stability and low analyte concentrations.

6.7 Trends in Stability for a Range of HDL Excipients.

6.1.1 LAS & SDS

LAS presents the greatest challenge for protein analysis. Both thermal and storage analyses are hindered by the UV-active and viscous nature of the compound. Through use of a range of both established and novel procedures, however, some understanding of the effects of this surfactant was achieved.

Although generally considered to be a destabilising element of HDL, LAS increased T_m values for Everest at low concentrations (0.1% w/v). This was also observed at higher concentrations for other amylases by DSC. Similar effects have been reported in the literature for SDS due to bridging

between proximal basic and hydrophobic residues in native protein. This stabilisation is lost on the formation of micelles, however, as opposing charges on the exposed exterior of the aggregate repel one another, straining protein conformations.

In contrast, no stabilisation was observed for V42, the other proteases, or Lipex. This may be a result of the shorter protein chains reducing the distance between bound micelles, or the lower effective CMC as fewer monomers are sequestered through protein binding. Stabilisation may therefore be evident at lower concentrations. At concentrations above the CMC, thermal stability is reduced with increasing surfactant concentration. This reflects $T_{1/2}$ data, which reports a linear downward trend in stability, however, further study into the effects of protein on surfactant aggregation is necessary to support these theories.

Despite not being used in commercially available detergents, SDS was included in studies of detergent excipients to determine its equivalency with LAS. Both empirical T_m values, and more in-depth structural analysis by PCA, suggested that the two surfactants induced similar levels of unfolding in V42. This was not reflected in storage tests, however, as half-lives of V42 in SDS were shorter than those of equivalent LAS formulations. As indirect methods involving surfactant removal were required to establish T_m values in LAS, a more accurate comparison of thermal stability effects of the two surfactants may be achieved via a direct analysis method such as DSC. Alternatively, T_m values in SDS can be determined using the same $CaCl_2$ precipitation method as LAS to generate an equivalent dataset.

Difficulties associated with LAS also hindered analysis of multi-component systems. Brief DSF investigations of chelant and surfactant mixtures at low concentrations (5 mM 0.1% w/v respectively), were not found to have a synergistic effect on protein destabilisation. Instead, T_m values were consistent with the lower of the two single component systems. This, however, is unlikely to be true of all excipient combinations or full HDL

formulations. As such, further work should focus on expanding these methods to industrially relevant systems.

6.1.2 AE3S & AE7

Effects of AE3S and AE7 on protein stability are not clearly defined by surfactant concentration. Consistent trends are evident between the two surfactants, however, with highest thermal stability observed at 5%. This is in line with levels found in commercial formulations. $T_{1/2}$ data for V42 indicated that the unexpected fluctuations in T_m with concentration were insignificant with respect to storage stability. Accelerated tests showed only gradual changes in half-life on titration of surfactant, which trended towards longer storage life.

Everest, on the other hand, was destabilised by both AE3S and AE7. This is thought to be induced by sequestration of structural calcium by the surfactants in micellar states. Amylases are more susceptible to unfolding under these conditions as they lack the secondary Ca^{2+} binding site found in proteases which provides added structural support. Initial unfolding caused by loss of this ion promotes surfactant binding to newly exposed residues. Greater reductions in shelf life in the presence of anionic AE3S are a result of electrostatic interactions, absent in its non-ionic counterpart.

Variation in the impact of these secondary surfactants on amylase and protease unfolding was not reflected in thermal denaturation analysis, which reported similar trends with respect to increasing surfactant concentration. Again, due to the lack of consistent trends in Everest thermal and storage stability values, it is difficult to identify the source of this deviation. $T_{1/2}$ and T_m values for V42, however, align with empirical fitting of other excipients, indicating that observations are representative of genuine processes.

6.1.3 Chelating Agents

EDTA induced the greatest effect on enzyme stability of the chelating excipients for both amylases and proteases. This reflects the high K_a values reported in the literature. HEDP, citric acid and fatty acid, despite differences in calcium association constants covering several orders of magnitude, all induced similar levels of thermal instability. These were essentially negligible in Everest, but more significant in V42 samples. Proteases are generally more resistant to calcium induced instability due to the presence of a second, more tightly bound Ca^{2+} ion, however, this is likely an artefact of the lower baseline thermal stability of V42.

The relationship between V42 values for thermal and storage stability were in line with those of other excipients. Amylase samples, in contrast, exhibited near complete loss in activity within a single day under these conditions. HEDP samples maintained activity for several weeks, however, this is still negligible with respect to the control. These effects are not reflected by T_m values and led to extreme outliers in the combined plots of all excipients.

Furthermore, literature K_a values could not be used solely to explain chelants effects, excluding the exceptionally high values for EDTA, stability trends did not align with expectations based on binding constants. Further data describing both the specific metal ion binding affinities of the detergent proteins and K_a values of chelating agents under detergent conditions are required for greater insight into these effects.

6.8 Summary

The various advantages and limitations associated with each of the above methods suggests that optimal results would be achieved through the use of a range of techniques in parallel, each focusing on different aspects of protein interactions in HDL. Comparison of datasets collected using CD, DSF and DSC in control samples yielded a linear correlation, suggesting

that datasets obtained via multiple methods could be incorporated into a single stability model. This would provide a more comprehensive and robust indication of protein structure and activity in HDL. Furthermore, once relationships between stability parameters obtained from each method were established, techniques could be chosen to fit the type of analysis required, high throughput, high surfactant concentrations, or structural analysis, for example.

Issues associated with high concentrations of LAS are crucial to address due to the prevalence of these compounds in laundry formulations. Procedures described in this thesis for approximating T_m values are robust and precise, however further work is required to confirm correlations with true values for in-situ LAS analysis. The simplest approach to this validation would be through analysis of SDS-induced unfolding by DSC.

Understanding of the complex interactions surrounding detergent enzymes in HDL is currently very limited. Empirical stability parameters collected in this work provide indications of the effects of these excipients on protein structure and function, however further insight into unfolding processes is needed. Consistent data relating to surfactant micellation, chelant binding constants and protein-ligand binding in detergent media is the first step towards mechanistic understanding of inactivation processes. This should be supported by further structural analysis using CD spectra, as demonstrated for LAS and SDS samples.

6.9 Future Work

Work in this thesis has provided proof of concept for the use of optical methods for protein stability modelling. Future work should focus on extrapolating these methods to fully formulated detergents. The primary challenge of this development will be the accurate determination of T_m values. Careful application of both SDS-mock formulations and sample purification methods should facilitate these efforts.

Indications from analysis of both empirical stability data and structural CD spectra show the link between SDS and LAS-induced protein denaturation. To establish SDS analogs as the standard for determining LAS stability, variation between the two conditions needs to be accurately determined. DSC data, capable of generating T_{max} data in the presence of high surfactant concentrations, should be employed to compare levels of destabilisation. This would provide for accurate adjustment of T_m values for the prediction of half-lives and will also be necessary for multi-component samples, to ensure interactions with other excipients are also consistent between the two surfactants.

Once empirical relationships between SDS and LAS data have been established, higher throughput methods can be employed. This will require further development of DSF procedures. Work in this thesis has focused on CD to facilitate the comparison of structural effects of various excipients. Furthermore, the use of intrinsic protein properties avoided additional complications associated with external dyes. Issues arising from DSF dye-surfactant interactions can be avoided however, through the use of molecular rotors. This would enable the incorporation of ultra-high throughput methods to the available suite of protein analysis techniques.

In the event that SDS-mock formulations are incompatible with multi-component formulations, T_m analysis can revert to surfactant removal methods. Validation of this technique may be simplified by streamlining the purification process with FPLC. Larger preparatory columns would improve on surfactant binding achieved in this work, which was insufficient the large volumes present in solutions. Efficiency can be further improved with autosampling following dilution of incubated samples.

Due to the high degrees of error in both storage testing and thermal analysis, definitive conclusions could not be drawn for the amylase, Everest. This is likely an artefact of the lower enzyme concentrations used to align with commercial formulations and the exceptionally high thermal

and storage stability of the enzyme. Increasing enzyme concentration and lengthening storage times should improve the accuracy of half-life estimations, while the use of pressurised DSC may provide more accurate T_m determination. Alternatively, validation could be conducted using an amylase with a lower thermal profile, such as Natalase.

Finally, mechanisms associated with unfolding arising from excipient interactions studied in this thesis should eb further explored, Collection of comprehensive and consistent data on excipient properties under detergent conditions will provide insight into observed stability trends. This should include analysis of association constants of ligands and relevant metals, and the effects of HDL conditions of protein-surfactant interactions and aggregation states. Application of PCA to CD spectra, as demonstrated for LAS and SDS, to remaining excipient groups, would also contribute to greater understanding of structural changes with enzyme inactivation. Future attempts towards assigning specific structural features should focus on proteins with established crystal structures, enabling validation of deconvolution programmes. These procedures can them be applied to novel detergent enzymes with confidence.

6.10 References

1. E. Ablinger, S. Leitgeb and A. Zimmer, *Int. J. Pharm.*, 1, **441**, 255–260.
2. T. Menzen and W. Friess, *J. Pharm. Sci.*, 2013, **102**, 415–428.
3. S. Khrapunov, *Anal. Biochem.*, 2009, **389**, 174–176.
4. H. Lund, *et al*, *J. Surfactants Deterg.*, 2012, **15**, 9–21.
5. H. Lund *et al*, *J. Surfactants Deterg.*, 2012, **15**, 265–276.
6. D. S. Goldberg, S. M. Bishop, A. U. Shah and H. A. Sathish, *J. Pharm. Sci.*, 2011, **100**, 1306–1315.

Appendix 1

1.1 Nano-DSC Report

UNIVERSITY OF LEEDS

Differential scanning calorimetry (DSC)

Analysis Report.

Customer;

Niamh Ainsworth, Stefanie Freitag-Pohl, David Hodgson

Department of Chemistry, University of Durham, Durham, DH1 3LE.

Analyses performed and report written by;

Iain Manfield, Astbury Centre for Structural Molecular Biology, University of Leeds, LS2 9JT.

Date; 12th May 2015

Introduction

Differential scanning calorimetry (DSC) is one of the techniques used to measure the thermal stability of proteins. As it is not an optical technique, DSC complements circular dichroism and fluorescence spectroscopy which are sensitive to the absorbance and fluorescence of other solutes and to sample turbidity.

DSC tracks the amount of power required to heat two cells, containing a protein solution and a reference buffer, at the same rate. Unfolding a protein uses power which would otherwise be heating the cell contents, giving a temperature differential, so extra power must be supplied to the sample cell. Conversely, protein aggregation is typically exothermic and this is often seen in DSC data, giving negative values. The temperature at which the extra power is required, and how much, gives us information about the melting temperature of a protein and the amount of power required to unfold the protein. However, water has a high heat capacity and the protein occupies only a small proportion of the volume of the cell. Therefore, for the best data, a well-matched buffer in the reference cell and subtraction of buffer-buffer scans are required.

Reversibility of unfolding is important for a thermodynamic analysis of DSC data, as this demonstrates that there are no additional reactions or interactions which would complicate the analysis of the power required to unfold the protein. This problem is clearest where there is a strong exothermic signal as the protein unfolds, usually taken to indicate aggregation. Without refolding, it is not appropriate to analyse the data to give T_m and enthalpy change, ΔH.

However, without matched buffers or reversible unfolding, DSC data can be analysed to compare the melting temperatures of different proteins in different buffer formulations, giving a T_{max}, a temperature of maximum heat capacity change.

Materials and Methods

Materials

Protein samples (6 ml) and LAS-free buffers (50 ml) were supplied frozen on dry ice and were stored in a cold room at 5 °C. Samples were warmed to room temperature and mixed gently before taking samples, especially for the samples with higher concentrations of LAS, showing some precipitation. For the two protease samples, PMSF (10 mM, 6 µl) was added to the 6 ml protein stock and mixed gently before a 2 ml sample was taken for analysis. The time taken for degassing and temperature

equilibration in the calorimeter is likely to be sufficient for PMSF binding to reach equilibrium.

Methods

Differential scanning calorimetry

The calorimeter was cleaned before each set of analyses by soaking the cells with Decon 90 (5% v/v) for 1 hour at 50°C and then rinsed thoroughly with multiple changes of water. Protein samples and buffer (2 ml) were degassed using a Microcal Thermovac device for 10 minutes with the vacuum applied progressively to avoid samples bubbling too vigorously. The calorimeter (Microcal, VP-DSC) was set to heat from 10 °C to 90 – 120 °C, at 90 °C/hr, depending on where transitions were observed. An initial 2-3 scans were performed with water and buffer but not used because of "thermal history" effects arising from small differences between each cell. Cell contents were changed as the temperature cooled to 25 °C and then re-pressurised to ~29 psi, with a 15 minute equilibration step to 10 °C before the scan started.

Data analysis

Scans are presented in two ways for this report. For the first replicate of scans, a repeat scan of each sample was available and as this showed no evidence of refolding it was taken to approximate a buffer-buffer scan and subtracted from the protein-buffer scan. For the second replicate, where a repeat scan of each sample was not collected, the scans are normalised on the signal intensity at 20 °C (or 30 °C for one protein). To compare peak areas, where a pre- and post-transition baseline were visible, a baseline was selected either side of the peak and the area integrated above the baseline. As buffer-buffer scans were not available and as refolding was not observed, these peak areas are presented in arbitrary units. Data were analysed using the Microcal version of Origin 5.0.

Results

DSC of a positive-control protein - lysozyme

For small, monomeric, single-domain proteins the unfolding transition often gives a peak over around 20 °C. This is flanked by flat or gently-sloping baselines, which can be used for integration of peak area. Re-scanning the same sample indicates whether re-folding has occurred. These properties can be seen with scans of unfolding lysozyme (1 mg/ml in glycine.HCl, 20 mM, pH 2.5) in Figure 1. The protein unfolds between 50 – 70 °C with a maximum at 63 °C. Re-scanning this sample shows peaks of reducing height at 6 3°C but increasing signal around 50 °C, indicative of refolding but with accumulation of some mis-folded protein.

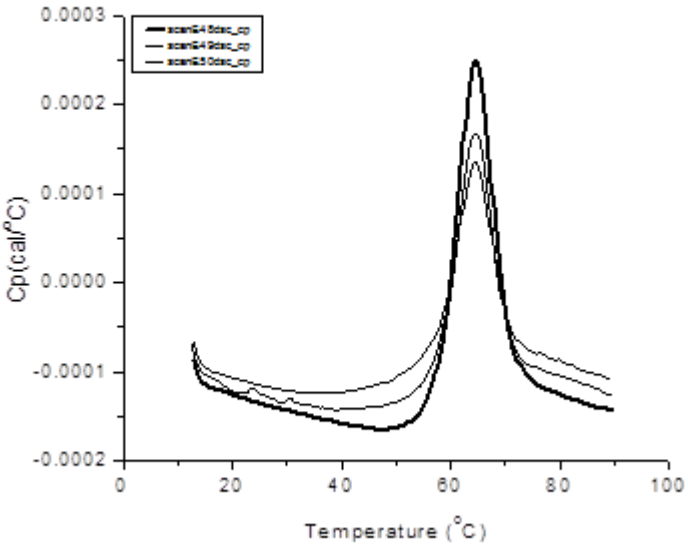


Figure 1: *Differential scanning calorimetry of hen egg white lysozyme. Protein was dissolved in glycine.HCl buffer (20 mM, pH 2.5), degassed and heated from 10 to 90 °C. The sample was re-scanned twice to assess refolding efficiency.*

DSC of commercial, washing-powder enzymes

The molecular weight and multimerisation of lysozyme may be very different from commercial, washing powder enzymes, however, the shape of the DSC trace gives us a benchmark against which to compare the DSC traces of the commercial enzymes.

Protein 1

In the absence of any LAS, a maximum heat capacity change was seen at 81 °C (Figure 2). This was followed by a steep drop in the signal which typically indicates protein aggregation; the sample showed clear aggregation when removed from the calorimeter, after a repeat scan. The gradual sloping signal from 20 °C is unusual in protein DSC but may indicate some part of the protein unfolding even at these lower temperatures, or a buffer mismatch. The absence of any flat baseline pre- or post-transition means it is not possible to calculate the area of the peak reliably; however, it is clearly much larger than in the presence of LAS.

In the presence of LAS, more typical protein unfolding peaks are observed with more-or-less flat baselines pre- and post-transition, with T_{max} increasing to around 90°C with increasing LAS concentration (Table 1). In contrast, there is some evidence for peak area reducing with higher LAS concentration. There is no evidence for aggregation from the calorimeter signal when LAS is present. However, the absence of a peak in a repeat scan of each sample and the haziness of samples after two scans

suggests that the protein did not refold. There is some evidence for smaller peaks at 20 – 50 °C.

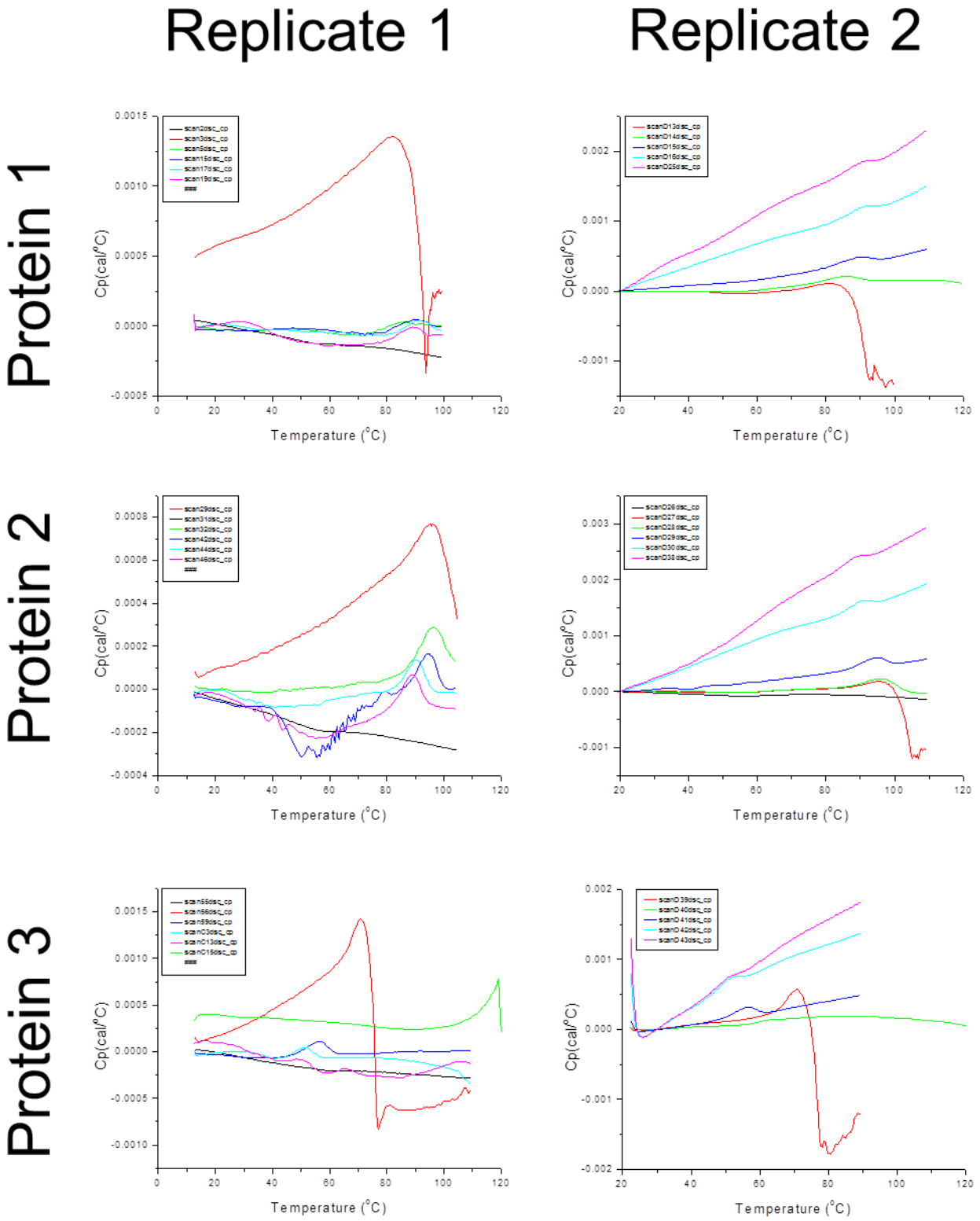


Figure 2: Differential scanning calorimetry thermograms of proteins 1, 2 and 3. Data are shown for sets of two scans of each protein, with differences in baseline slope due to differences in data analysis (see text for details). For all proteins red, green, blue, cyan and pink curves represent results for proteins without LAS and 0.1%, 1%, 5% and 10% LAS. A buffer-buffer scan is shown in black for some experiments.

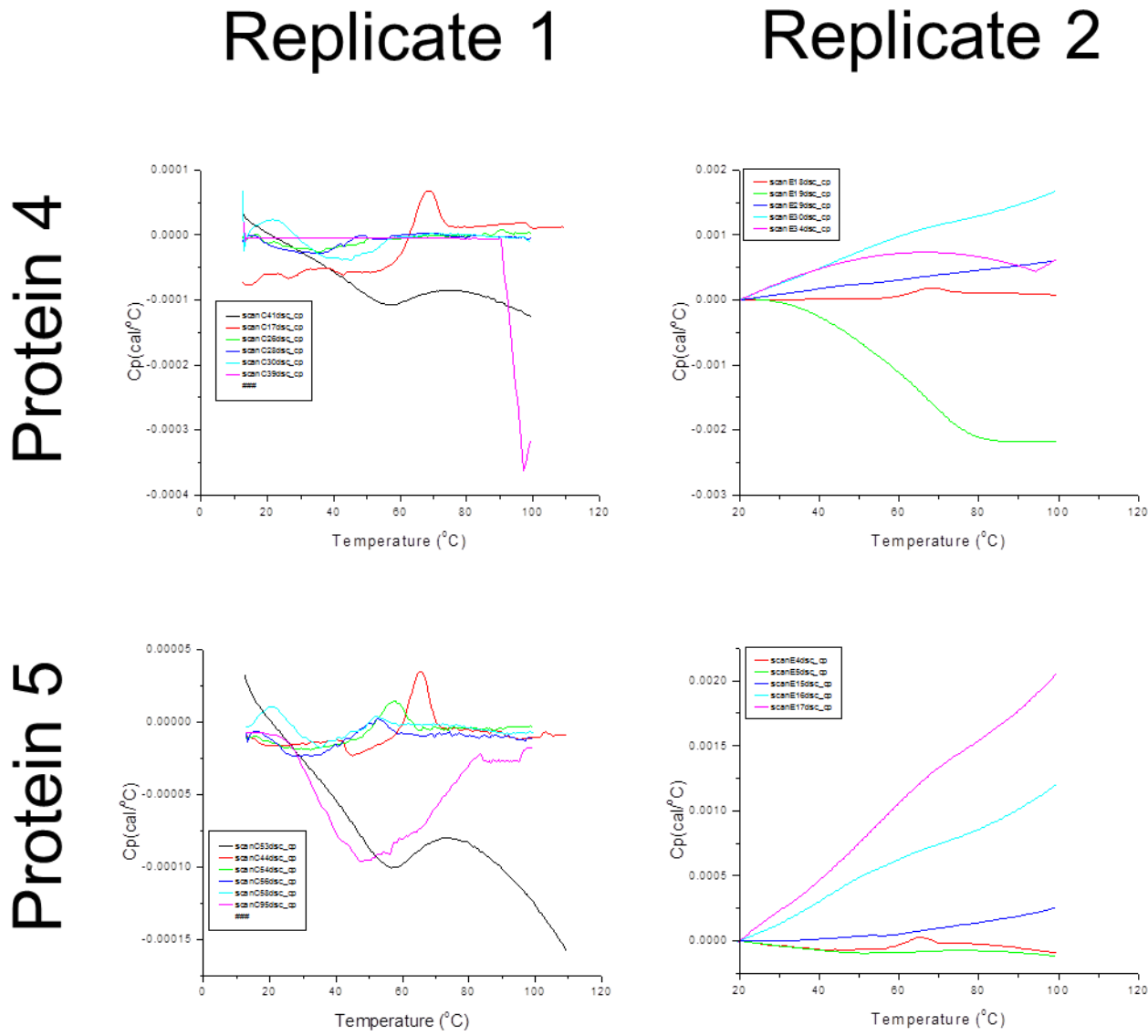


Figure 3: Differential scanning calorimetry thermograms of proteases, proteins 4 and 5. Data are shown for sets of two scans of each protein, with differences in baseline slope due to differences in data analysis (see text for details). For all proteins red, green, blue, cyan and pink curves represent results for proteins without LAS and 0.1%, 1%, 5% and 10% LAS. A buffer-buffer scan is shown in black for some experiments.

Protein 2

The maximum point of the unfolding curve for protein 1 without LAS is at 95 °C – the highest temperature for this set of proteins (Table 1). This point is preceded by a gradient from 20 °C, rather than a flat baseline. The asymmetry of the unfolding curve suggests that the protein is starting to aggregate at high temperatures, around 100 °C (Figure 2). This is clearer in the “replicate 2” scan.

Addition of LAS to 0.1% has little effect, perhaps a small stabilisation. More significantly, this low LAS concentration appears to have prevented aggregation at high temperatures. Further increases in LAS concentration destabilise the protein, but only to a T_{max} of around 90 °C. The shape of the unfolding curves suggests that there is little unfolding until >80 °C. There is some reduction in peak area as LAS concentration increases (Table 1).

Protein 3

Protein 3 without LAS shows a T_{max} of 70.8°C. However, the subsequent exothermic signal as the protein unfolds (and presumably aggregates) means that it is not clear whether 70.8°C is the maximum, or just the point where the aggregation signal becomes dominant. The curve is similar to that for Proteins 1 and 2 with a sloping baseline from 20°C.

The effect of adding LAS to 0.1% is not consistent, with only partial unfolding apparently even at 120 °C in one replicate, but a small peak at around 63 °C in the second replicate. Further additions of LAS, to 1, 5 and 10% give more consistent results with peaks at T_{max} of 50 – 60 °C. Importantly, this means that some proteins (or parts of proteins) will spend some time unfolded in the temperature range 40 – 50 °C.

Protein 4

Protein 4 without LAS showed an unfolding transition with T_{max} of 68 °C (Table 1). The calorimetry data show no evidence of aggregation after unfolding and this is supported by visual examination of the sample after repeat scans (data not shown; information recorded in “scan log” file).

In the presence of LAS, no clear unfolding peaks were seen.

Protein 5

Protein 5 without LAS showed a clear transition at around 65 °C, with no evidence of aggregation after unfolding. However there was no evidence of refolding in the repeat scan of this sample (data not shown, scan C45).

Addition of LAS reduced unfolding temperatures to around 52 °C and reduced peak areas. As with Protein 3, the destabilisation by LAS means that some molecules will spend some time unfolded at 40 – 50 ^{o}C. However, in 10% LAS, no transitions were seen.

		Scan 1				Scan 2
		T_{max} (°C)		peak area		T_{max} (°C)
Protein	LAS conc. (% w/v)	peak 1	peak 2	peak 1	peak 2	
1	0%	81.0		6.70		81
	0.10%	85.1	44.6	0.62	0.19	85.9
	1%	90.3	47.4	0.75	0.48	90.5
	5%	90.7	22.2	0.70	0.54	~90
	10%	89.2	28.6	0.53	0.38	~90
2	0%	95.0		3.83		95.5
	0.10%	96.4		1.78		95.7
	1%	94.4		1.27		94.6
	5%	90.6		1.17		92
	10%	88.7		1.47		~90
3	0%	70.8		4.60		70.8
	0.10%	>119		7.25		~63
	1%	56.0		1.39		56.8
	5%	51.0		0.87		~52
	10%	65.8	50.6	NA		~52
4	0%	67.9		0.65		68.6
	0.10%	no clear peak		NA		no clear peak
	1%	48.0		0.04		no clear peak
	5%	no clear peak		NA		no clear peak
	10%	no clear peak		NA		no clear peak
5	0%	65.2		0.28		65.4
	0.10%	57.7		0.20		no clear peak
	1%	52.5		0.08		no clear peak
	5%	51.5	21.0	0.04		no clear peak
	10%	no clear peak		NA		no clear peak

Table 1: *Melting temperatures and peak area from DSC analysis of protein unfolding. For some proteins, no clear peak was seen and therefore no peak area calculated ("NA", not analysed). Peak areas were only calculated for Replicate 1 samples. For some samples, small peaks were seen on a strongly-sloping baseline and this reduces confidence in determining T_{max}; this is indicated with the "~" symbol.*

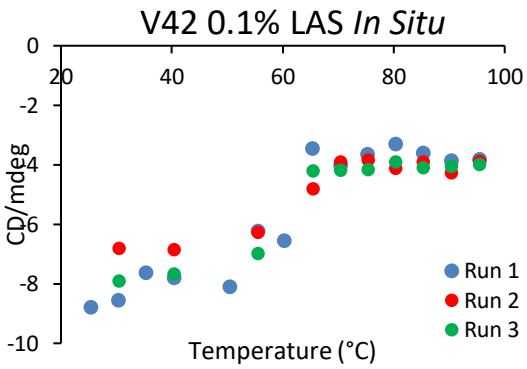
Next steps

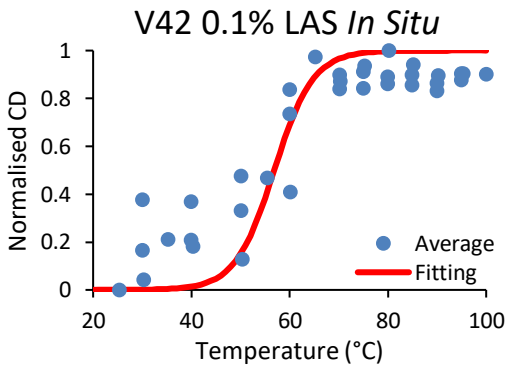
- Repeat analysis of protein 4 to determine whether the absence of a peak in the presence of any concentration of LAS is because LAS strongly stabilises or strongly destabilises the protein. Use SDS-PAGE analysis after scanning to determine if the protein is still intact, even if unfolded or inactive.
- Re-scan using matched buffers for each sample and performing buffer-buffer scans.
- Use a higher concentration of protein, e.g. 2 mg/ml, to give larger peaks, especially for the samples with higher LAS concentrations. This would likely exacerbate any aggregation, however for this project, higher protein concentrations might better reflect conditions during use.
- Add a third set of scans, allowing calculation of mean unfolding temperatures.
- Scan over a narrower range of temperatures to allow more scans to be performed per day, now we the relevant temperature range for each protein. Furthermore, starting scans from 20 °C may reduce any problem with LAS solubility at high concentrations and low temperatures, and also of high viscosity of 10% LAS solutions at low temperature, without loss of useful data.
- Test refolding by heating to temperatures which products might experience during use (e.g. 50 °C), but below full unfolding and then re-scan over the ranges tested in this report. This may be relevant for Protein 3 where unfolding was seen in the range 40 – 60 °C with LAS.

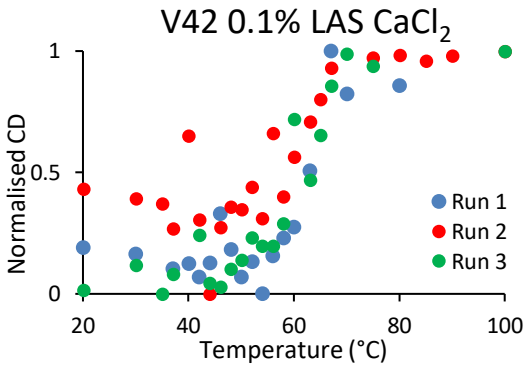
Appendix 2

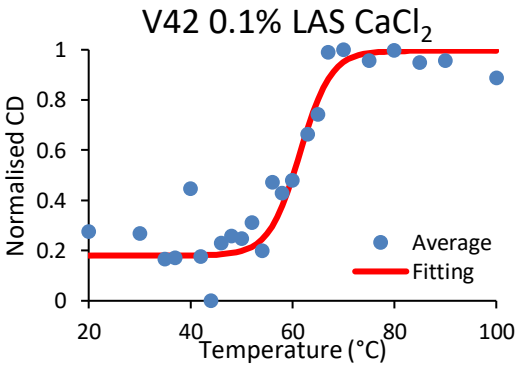
2.1 CD Thermal Denaturation Curves

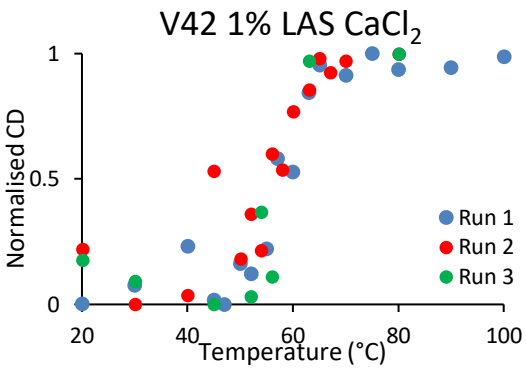
These graphs show the thermal denaturation curves for V42 and Everest in LAS, constructed from CD intensity values at 222 nm. The surfactant was removed prior to analysis by precipitating with $CaCl_2$. *'In situ'* curves at 0.1% LAS have also been included for comparison. Data and figures used to determine T_{mCD} values of other enzymes and formulations *(Chapter 3.2)* can be found in the supplementary information, along with those obtained by DSF *(Chapter 3.1)*.

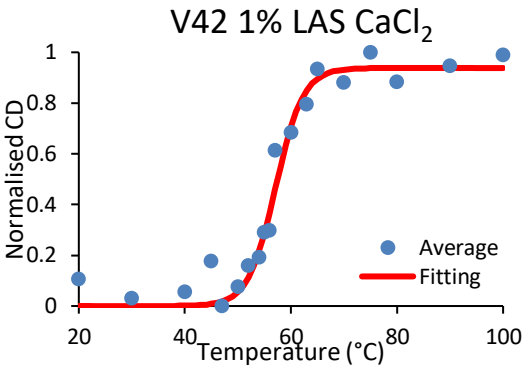


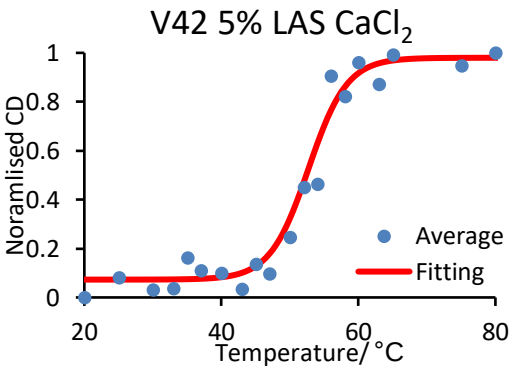


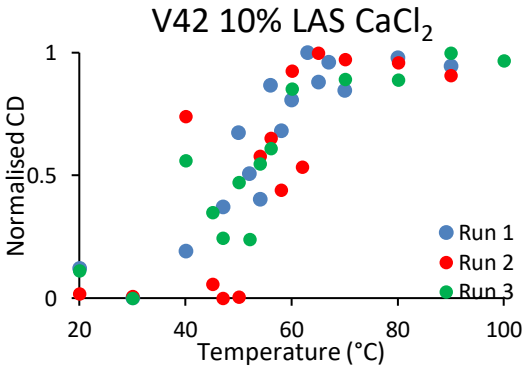


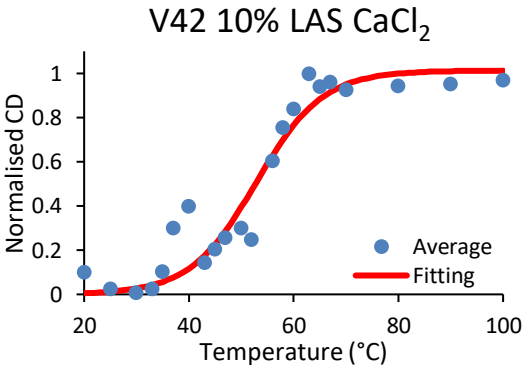


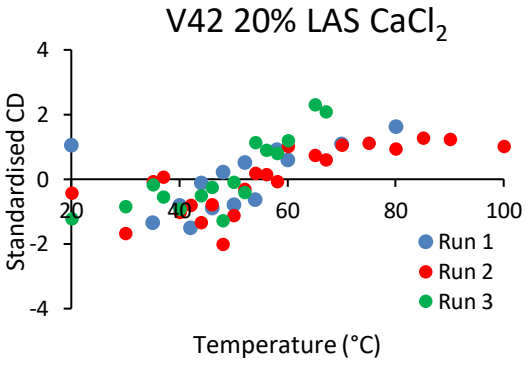

V42 1% LAS CaCl2
Normalised CD
Temperature (°C)
Run 1
Run 2
Run 3

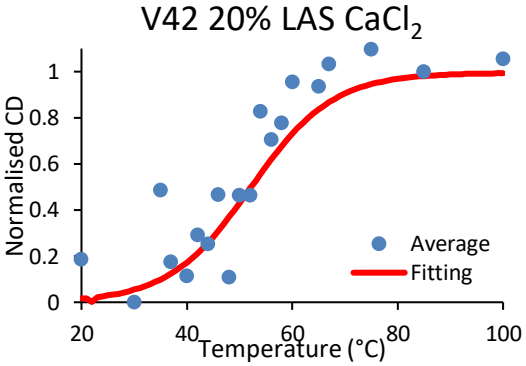

V42 1% LAS CaCl2
Normalised CD
Temperature (°C)
Average
Fitting

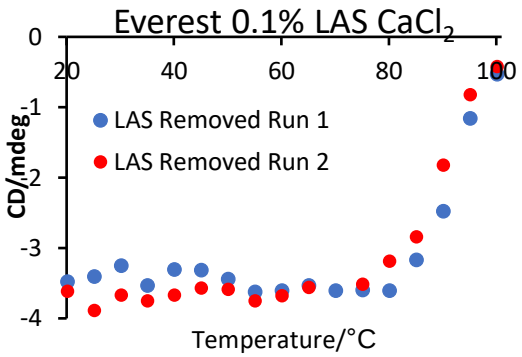

V42 5% LAS CaCl2
Noramlised CD
Temperature/ °C
Average
Fitting

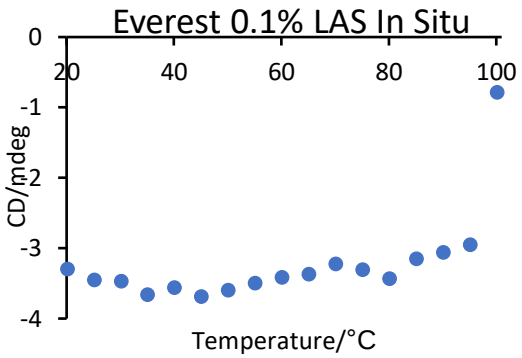

V42 10% LAS CaCl2
Normalised CD
Temperature (°C)
Run 1
Run 2
Run 3

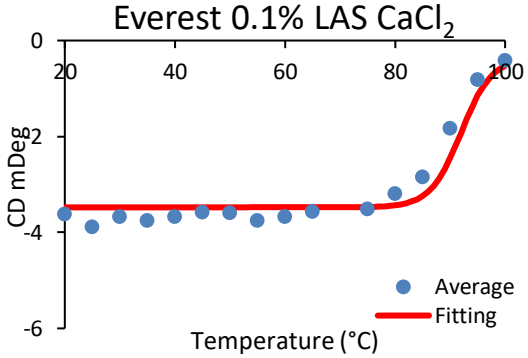

V42 10% LAS CaCl2
Normalised CD
Temperature (°C)
Average
Fitting

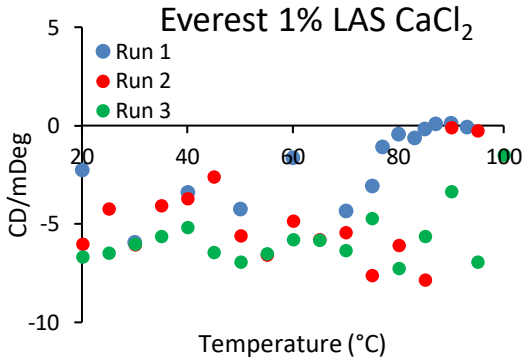

V42 20% LAS CaCl2
Standardised CD
Temperature (°C)
Run 1
Run 2
Run 3

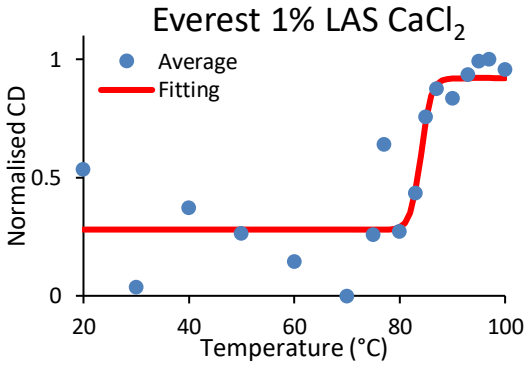

V42 20% LAS CaCl2
Normalised CD
Temperature (°C)
Average
Fitting

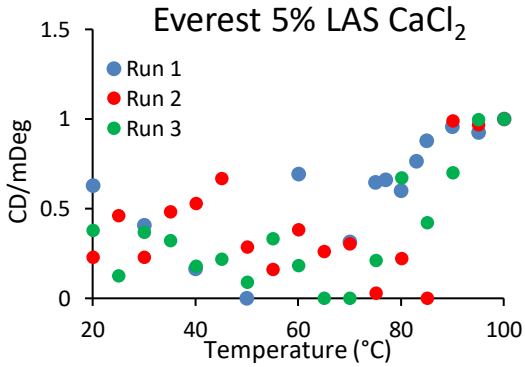

Everest 0.1% LAS CaCl2
CD/mdeg
Temperature/°C
LAS Removed Run 1
LAS Removed Run 2

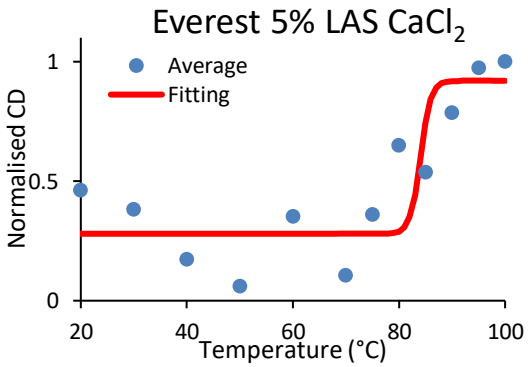

Everest 0.1% LAS In Situ
CD/mdeg
Temperature/°C

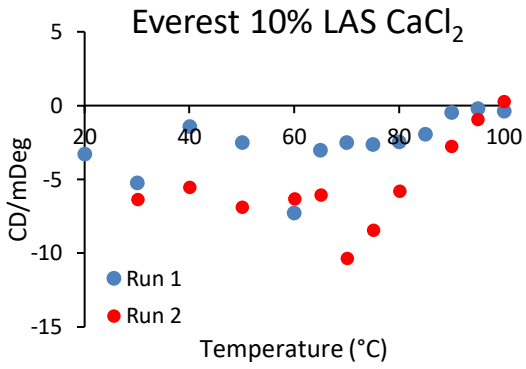

Everest 0.1% LAS CaCl2
CD mDeg
Temperature (°C)
Average
Fitting

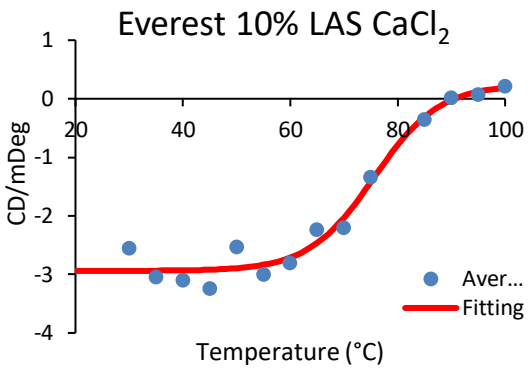

Everest 1% LAS CaCl2
CD/mDeg
Temperature (°C)
Run 1
Run 2
Run 3

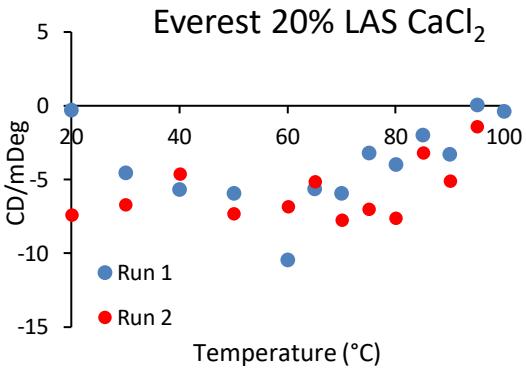

Everest 1% LAS CaCl2
Normalised CD
Temperature (°C)
Average
Fitting

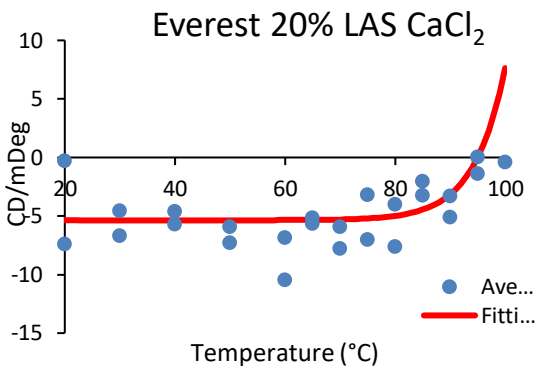

Everest 5% LAS CaCl2
CD/mDeg
Temperature (°C)
Run 1
Run 2
Run 3


Everest 5% LAS CaCl2
Normalised CD
Temperature (°C)
Average
Fitting


Everest 10% LAS CaCl2
CD/mDeg
Temperature (°C)
Run 1
Run 2


Everest 10% LAS CaCl2
CD/mDeg
Temperature (°C)
Aver...
Fitting

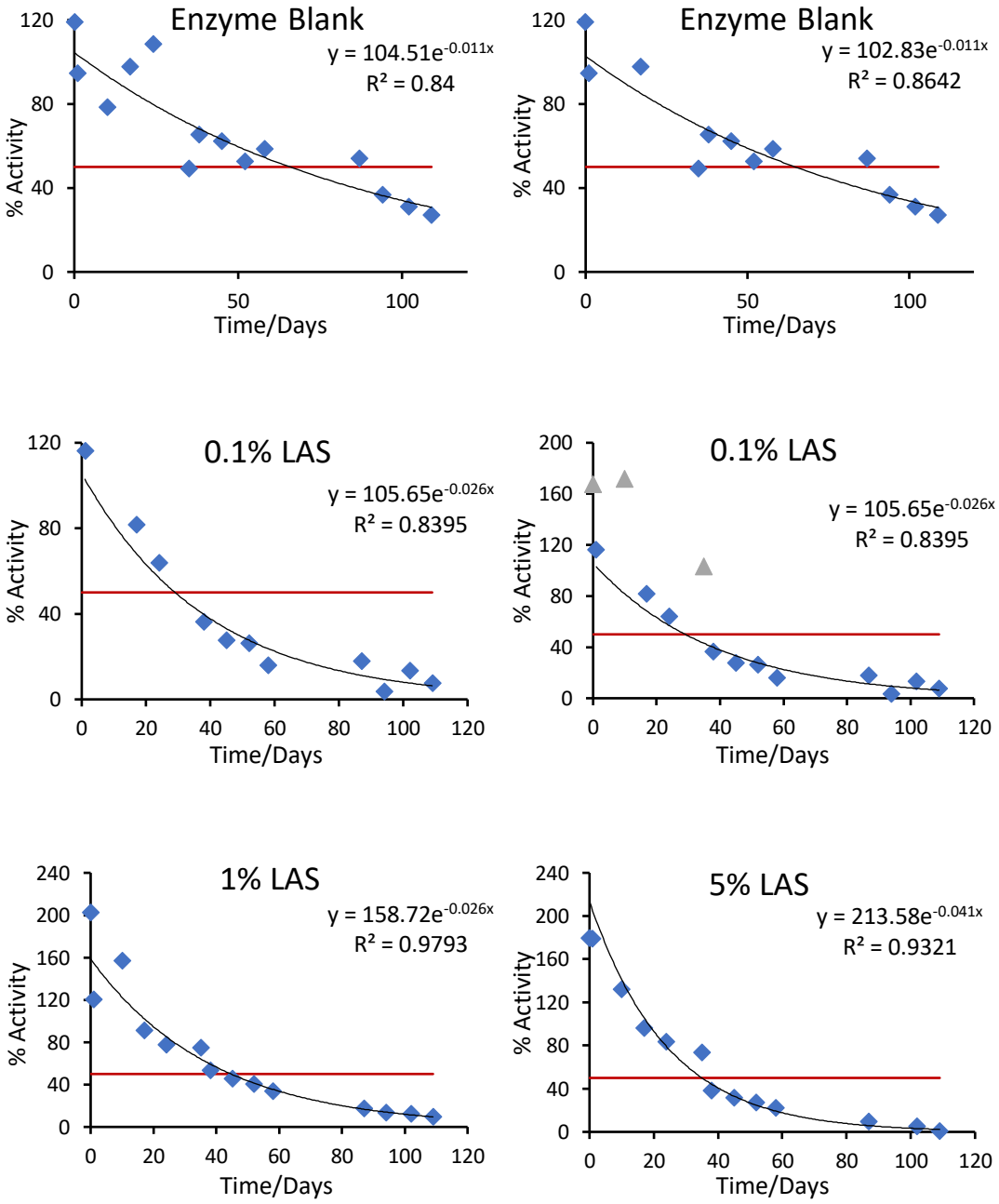

Everest 20% LAS CaCl2
CD/mDeg
Temperature (°C)
Run 1
Run 2

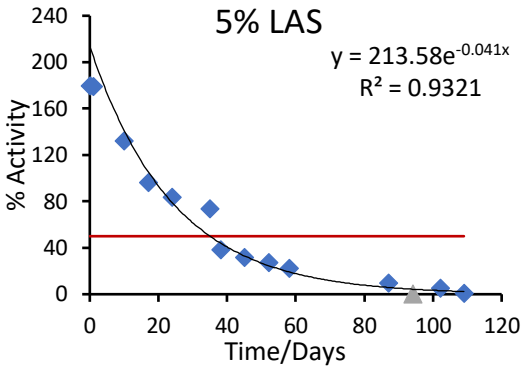

Everest 20% LAS CaCl2
CD/mDeg
Temperature (°C)
Ave...
Fitti...

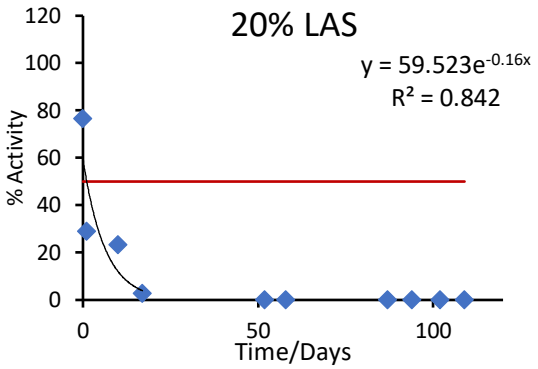
Appendix 3

3.1 Accelerated Storage Stability Tests: V42 (Dataset 2)

Samples were prepared and run as described in *Chapter 2.6.* Each formulation was run in duplicate. Results from the full study, not discussed in this thesis, can be found in the supplementary information.







Formulation	$T_{1/2}$
Buffer Blank	71.7
0.1% LAS	28.8
1% LAS	44.4
5% LAS	35.4
10% LAS	-
20% LAS	10.9
1% AE3S[1]	61.8
10% AE3S[1]	39.5
1% AE7[1]	94.3
10% AE7[1]	109.2
Citric Acid[1]	26.8
Fatty Acid[1]	47.4
HEDP[1]	30.7
EDTA[1]	1
Combined Surfactant[1]	19.4

[1]Graphs from which these $T_{1/2}$ values were obtained can be found in the supplementary information accompanying this thesis.

3.2 Accelerated Storage Stability Tests for V42 (Dataset 1)

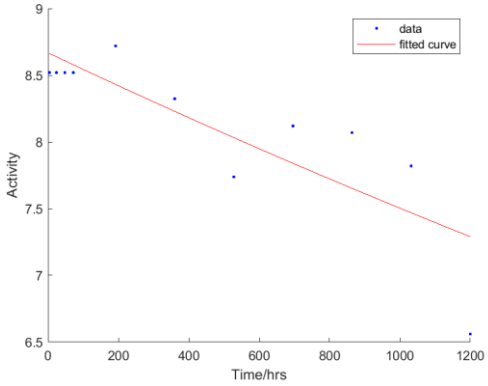


Figure 1: Blank

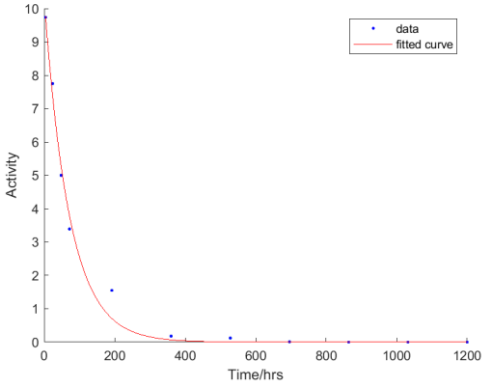


Figure 2: Stearic (Fatty) Acid

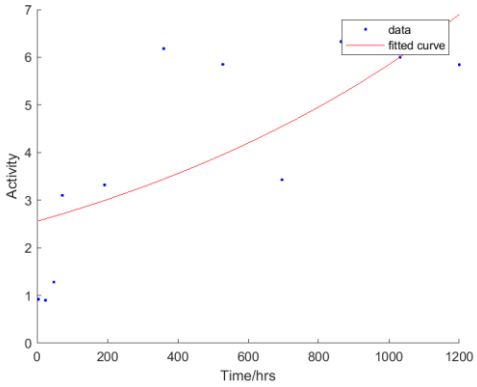


Figure 3: Mixed Surfactant

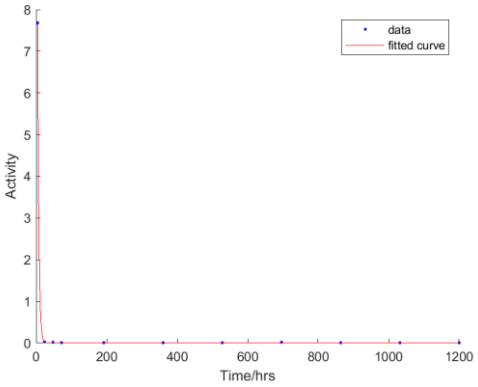


Figure 4: Citric Acid

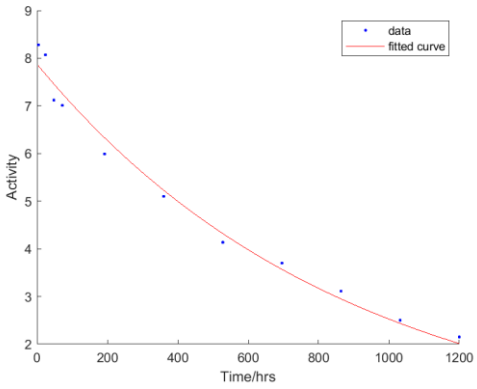


Figure 5: EDTA

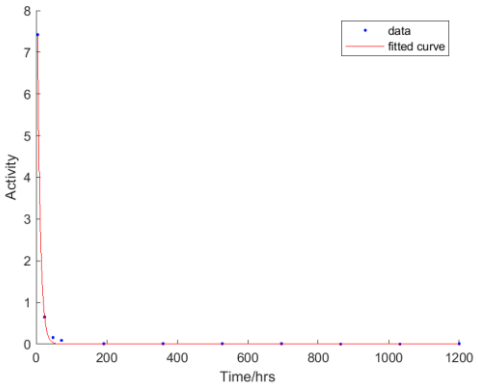


Figure 6: HEDP

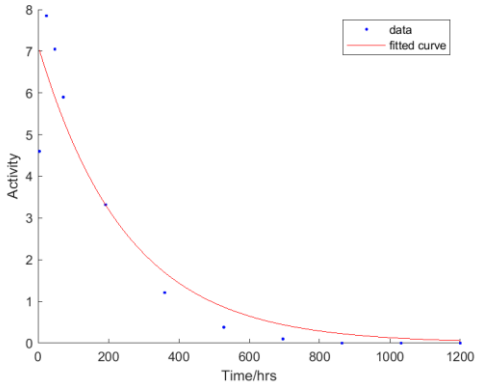


Figure 7: 20% AE3S

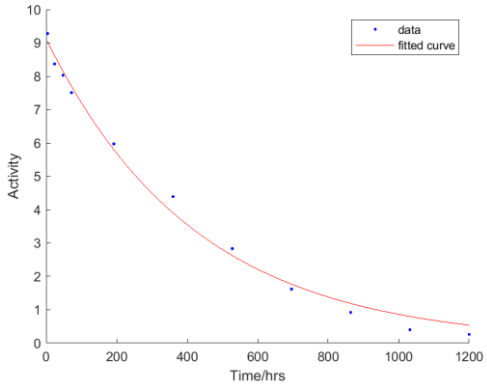


Figure 8: 10% AE3S

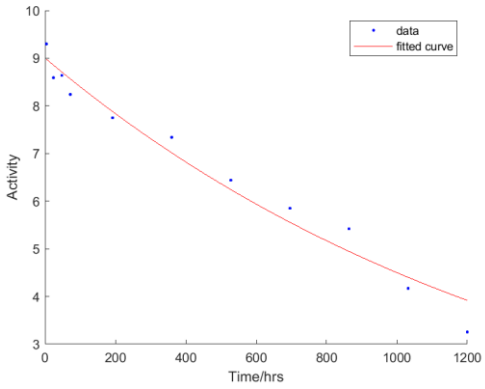


Figure 9: 5% AE3S

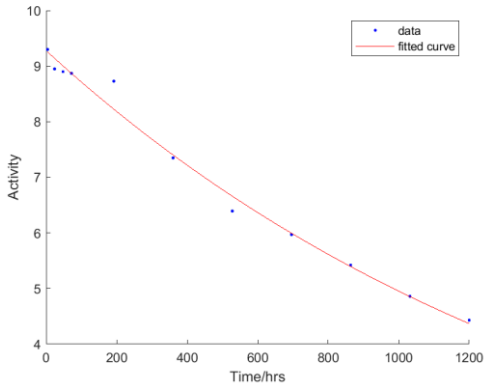


Figure 10: 1% AE3S

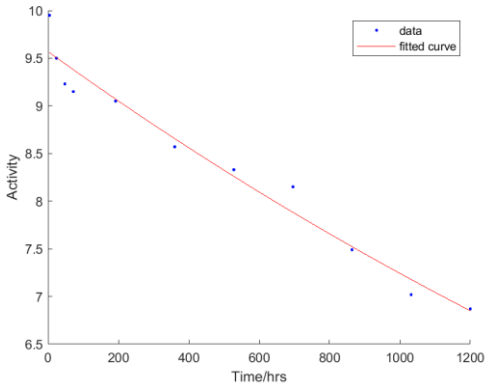


Figure 11: 0.1% AE3S

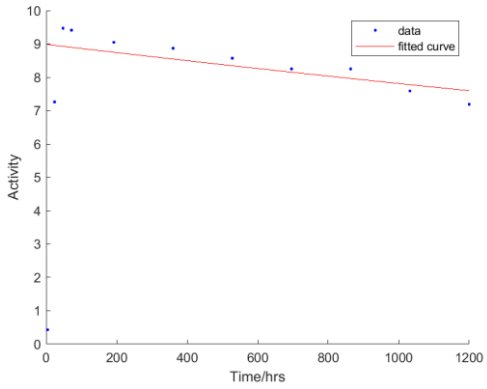


Figure 12: 20% AE7

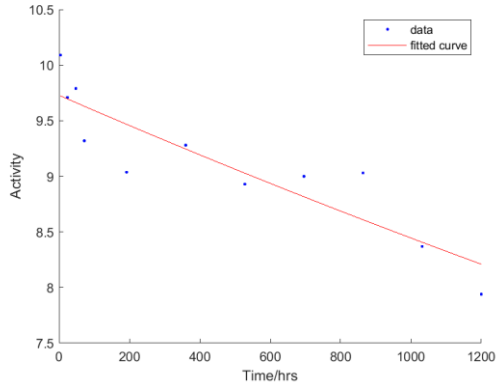


Figure 13: 10% AE7

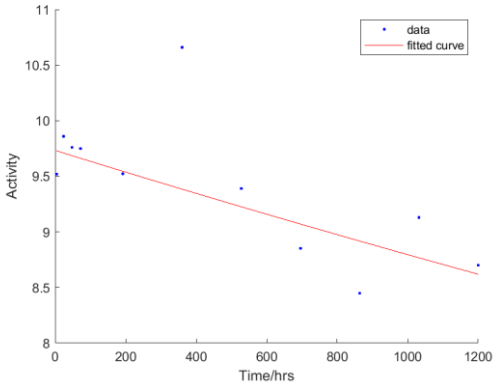


Figure 14: 5% AE7

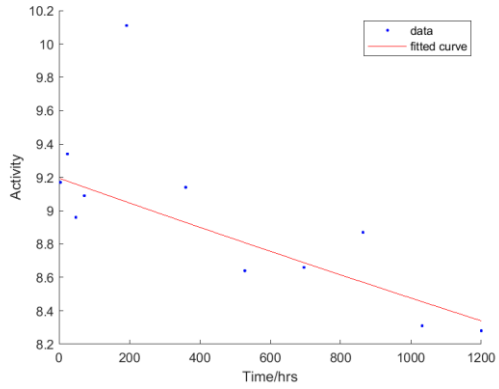


Figure 15: 1% AE7

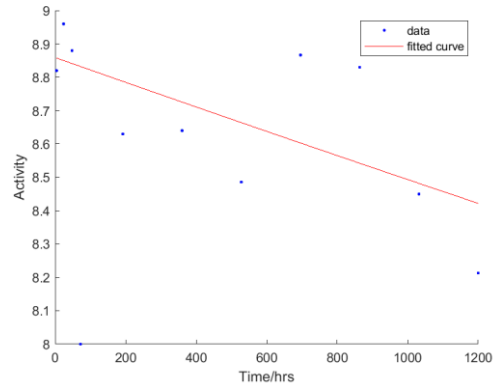


Figure 16: 0.1% AE7

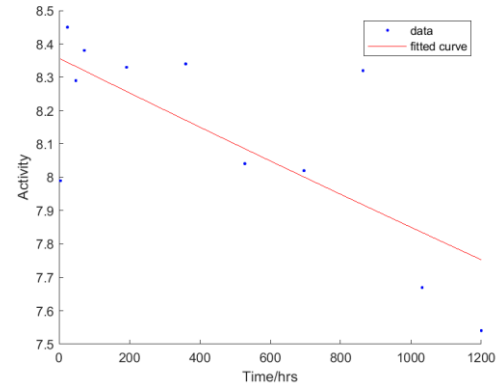


Figure 17: 20% SDS

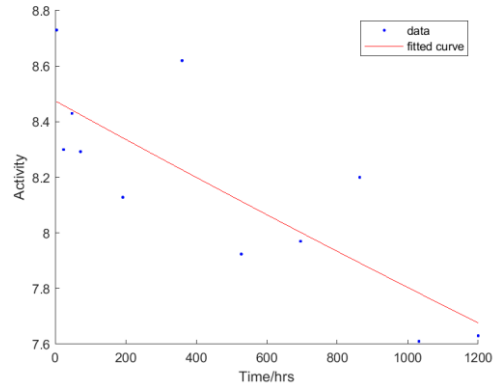


Figure 18: 10% SDS

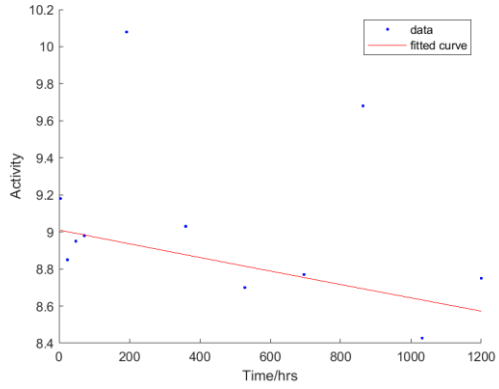


Figure 19: 5% SDS

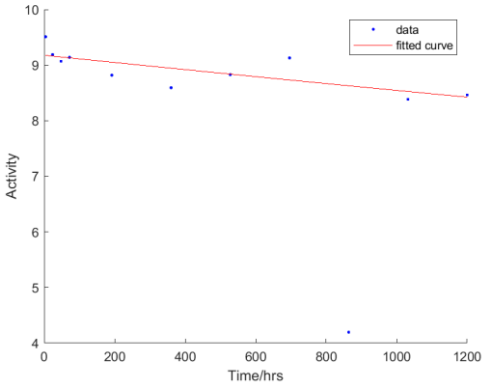


Figure 20: 1% SDS

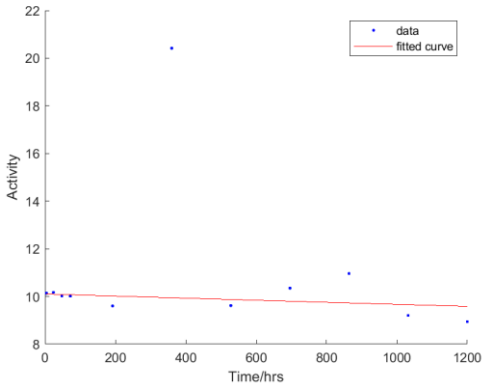


Figure 21: 0.1% SDS

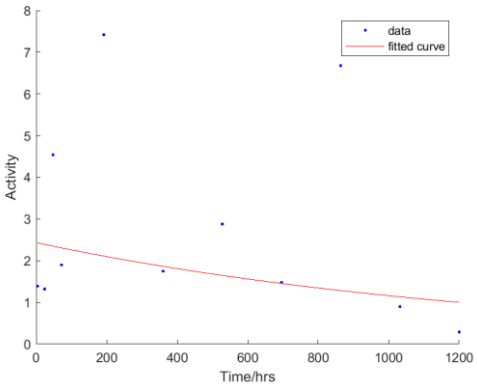


Figure 22: 20% LAS

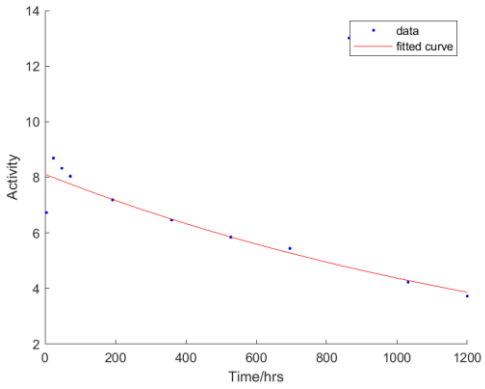


Figure 23: 10% LAS

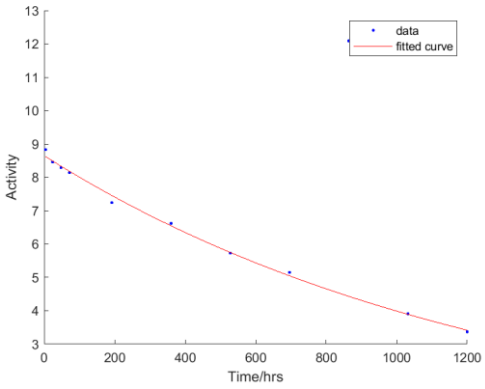


Figure 24: 5% LAS

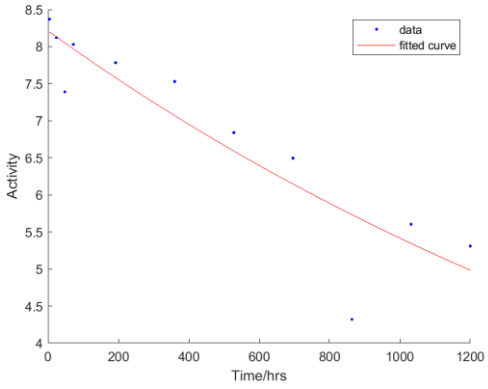


Figure 25: 1% LAS

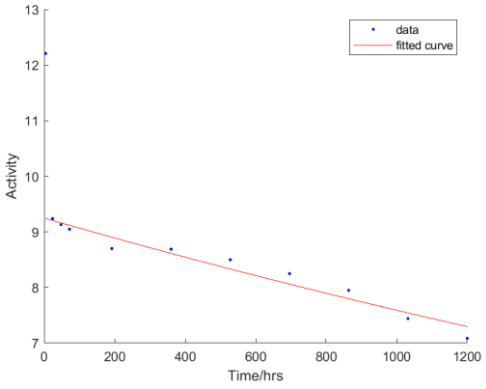


Figure 26: 0.1% LAS

3.3 Accelerated Storage Stability Tests for Everest (Dataset 1)

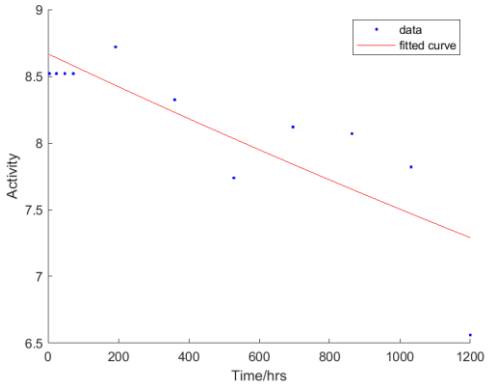


Figure 27: Blank

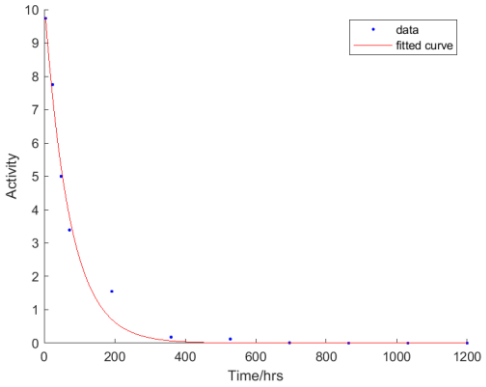


Figure 28: Stearic (Fatty) Acid

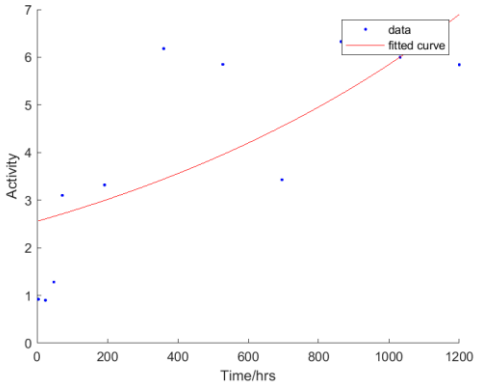


Figure 29: Mixed Surfactant

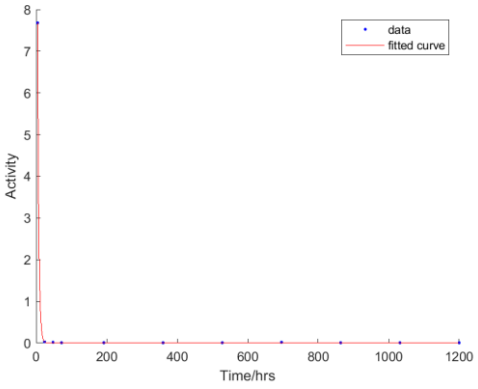


Figure 30: Citric Acid

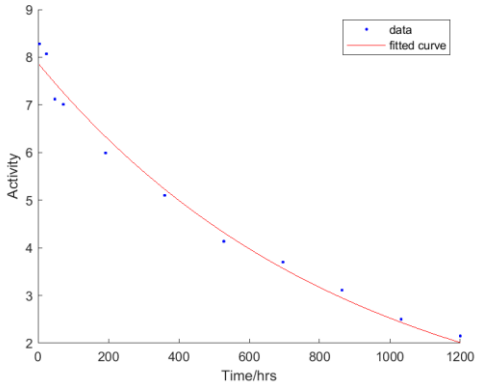


Figure 31: EDTA

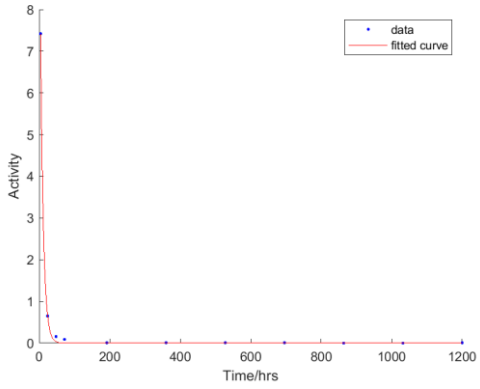


Figure 32: HEDP

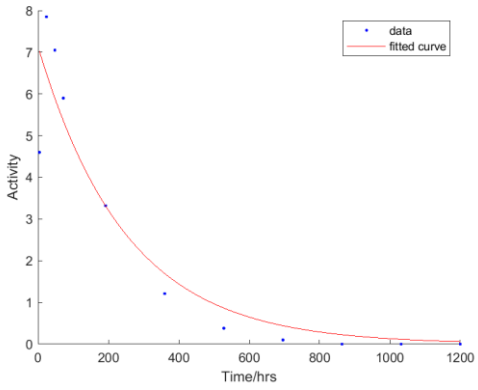


Figure 33: 20% AE3S

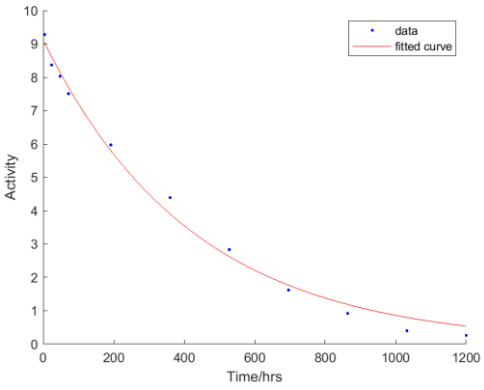


Figure 34: 10% AE3S

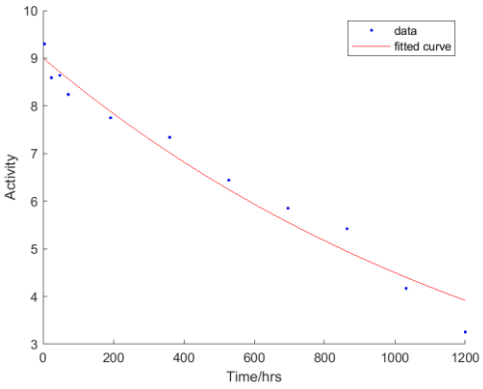


Figure 35: 5% AE3S

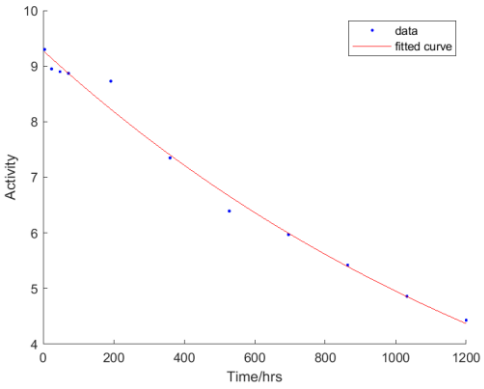


Figure 36: 1% AE3S

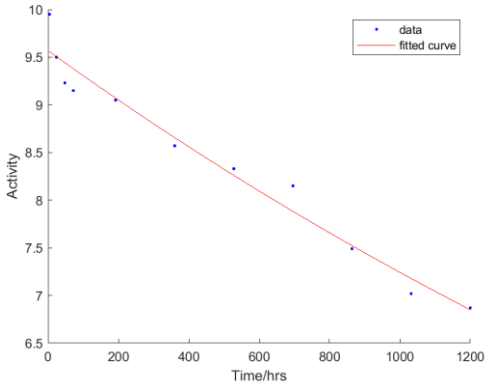


Figure 37: 0.1% AE3S

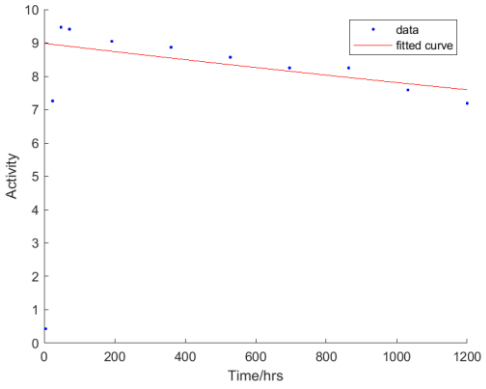


Figure 38: 20% AE7

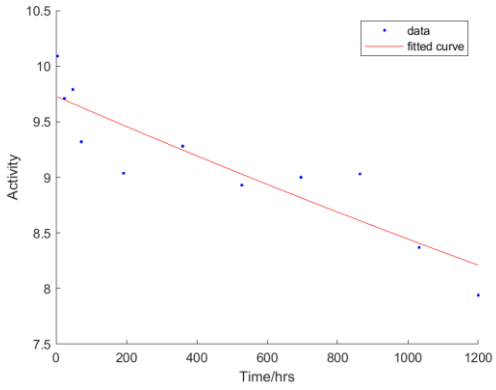


Figure 39: 10% AE7

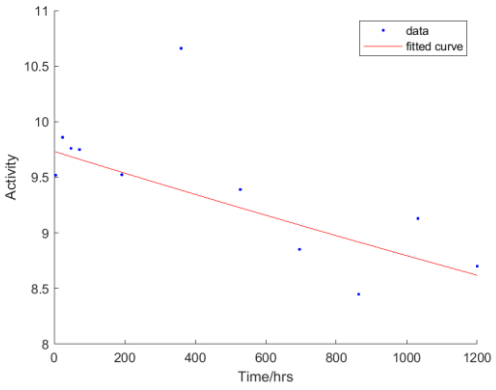


Figure 40: 5% AE7

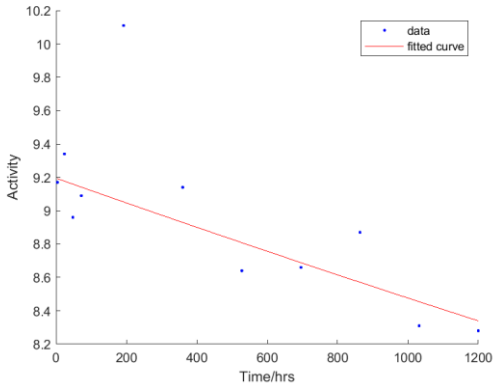


Figure 41: 1% AE7

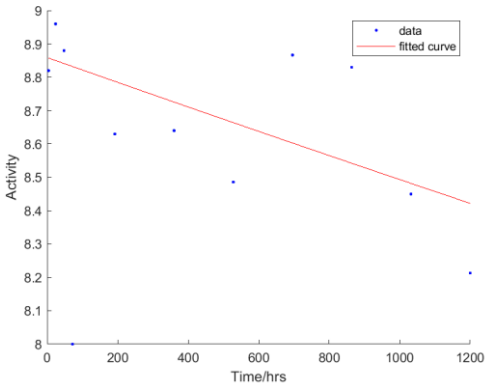


Figure 42: 0.1% AE7

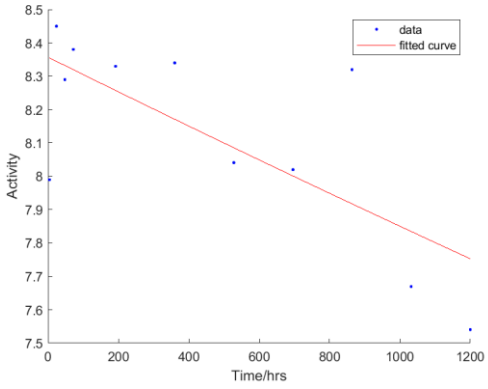


Figure 43: 20% SDS

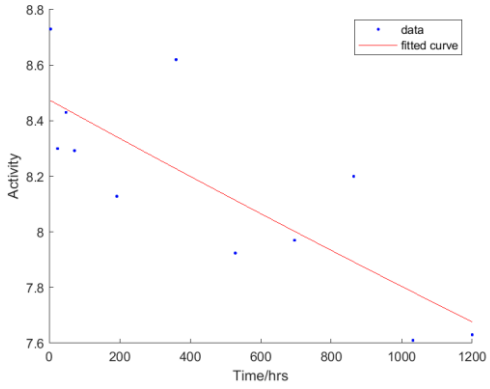


Figure 44: 10% SDS

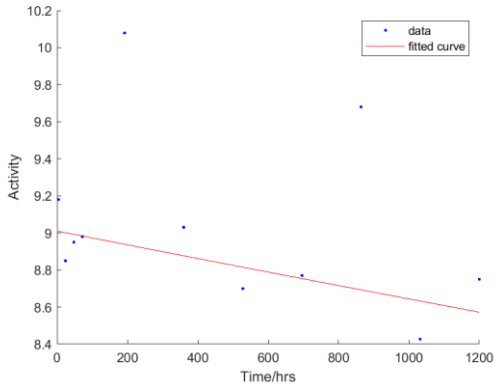


Figure 45: 5% SDS

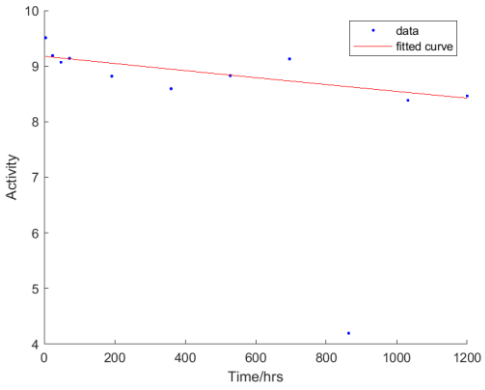


Figure 46: 1% SDS

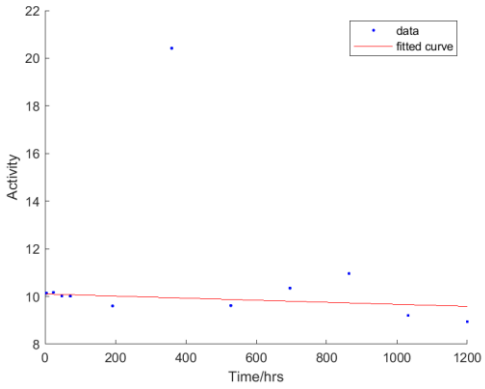


Figure 47: 0.1% SDS

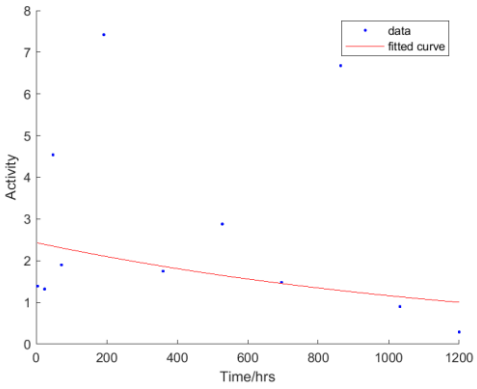


Figure 48: 20% LAS

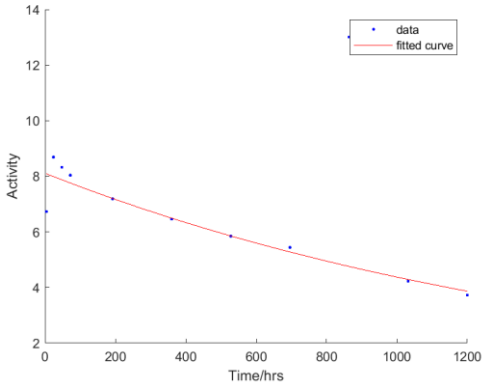


Figure 49: 10% LAS

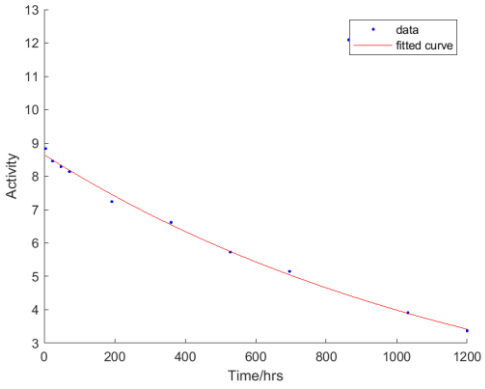


Figure 50: 5% LAS

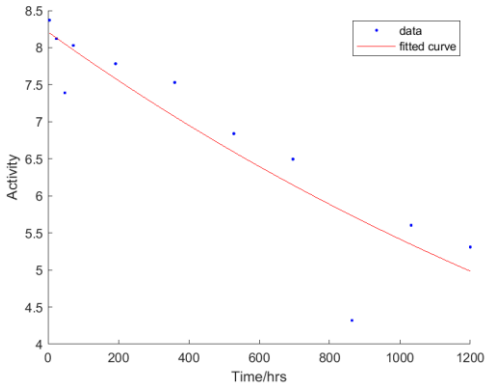


Figure 51: 1% LAS

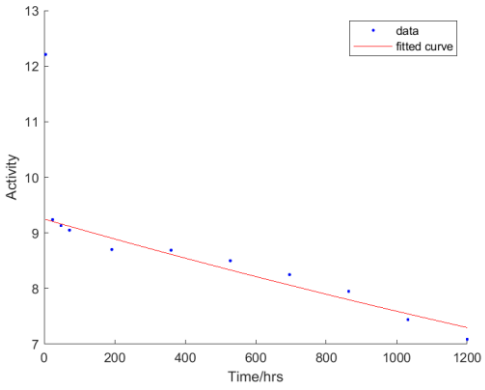


Figure 52: 0.1% LAS