The effect of different rotamers of histidine-70 on the reaction rates of aqualysin I, a subtilase from the thermophilic bacterium *Thermus aquaticus*

Arnór Freyr Sævarsson

Faculty of Physical Sciences
University of Iceland
2016

The effect of different rotamers of histidine-70 on the reaction rates of aqualysin I, a subtilase from the thermophilic bacterium *Thermus aquaticus*

Arnór Freyr Sævarsson

15 ECTS thesis submitted in partial fulfilment of a *Baccalaureus Scientiarum* degree in Biochemistry

Advisor

Magnús Már Kristjánsson, Professor

Co-advisors

Kristinn Ragnar Óskarsson, M.S.c
Bjarni Ásgeirsson, Professor

Faculty of Physical Sciences
School of engineering and Natural Sciences
University of Iceland
Reykjavik, May 2016

The effect of different rotamers of histidine-70 on the reaction rates of aqualysin I, a subtilase from the thermophilic bacterium *Thermus aquaticus*
Histidine-70 rotamers in aqualysin I
15 ECTS thesis submitted in partial fulfilment of a *Baccalaureus Scientiarum* degree in biochemistry

Faculty of physical Sciences
School of Engineering and Natural Sciences
University of Iceland
Hjarðarhagi 2-6
107, Reykjavík

Telephone: 525 4700

Bibliographic information:
Arnór Freyr Sævarsson, 2016, *The effect of different rotamers of histidine-70 on the reaction rates of aqualysin I, a subtilase from the thermophilic bacterium Thermus aquaticus*, Bachelors of science thesis, Faculty of Physical Sciences, University of Iceland, pp. 48.

Printing: Háskólaprent ehf
Reykjavík, Iceland, May 2016

Útdráttur

Subtilín-líkir serín próteasar hafa lengi vel verið notaðir sem módelprótein í próteinverkfræði til þess að öðlast betri skilning á virkni og stöðugleika próteina. Þetta verkefni er hluti af stærra verkefni sem hallast að samanburðatilraunum á ensíminu aqualysin I (AQUI) úr hitaaðlöguðu bakteríunni *Thermus aquaticus* við samstofna ensím úr kuldakærri *Vibrio*-tegund (VPR). Í fyrra verkefni var framkvæmd stökkbreytingin D98S á AQUI og olli það tvöföldun í hverfitölu (k_{cat}), þreföldun í hvötunarhæfni (k_{cat}/K_M) og engu marktæku tapi á hitastöðugleika. Í þessari tilraun var N68Q stökkbreyting framkvæmd á AQUI í von um að líkja eftir D98S stökkbreytingunni að hluta til. Tölvulíkan sýndi fram á að D98S stökkbreytingin stuðlaði að einni stellingu histidíns (His70) á hvarfstöð á kostnað annarra stellinga. Tölvulíkan af stökkbrigðinu N68Q sýndi svipuð einkenni á His70 og því varð stökkbreytingin fyrir valinu. Til þess að meta afleiðingar N68Q stökkbreytingar þá var stökkbrigðið borið saman við villigerð. Hitastigsstöðugleiki N68Q minnkaði ekki marktækt samkvæmt T_M og $T_{50\%}$ mælingum. Hraðafræðiniðurstöður sýndu fram á minni breytingar en búist var við þar sem K_M fór úr 1,00 mM (± 0,05 mM) í 0,81 mM (± 0,05 mM) og hverfitalan (k_{cat}) fór úr 33,3 s^{-1} (± 4,6 s^{-1}) í 33,6 s^{-1} (± 5,7 s^{-1}), á meðan hvötunarhæfnin (k_{cat}/K_M) breyttist ekki marktækt með stökkbreytingunni. Mældir hraðafræðifastar á villigerðinni í þessu verkefni höfðu nokkur frávik frá öðrum mældum gildum frá fyrri verkefnum á sömu tilraunastofu.

Abstract

Subtilisin-like serine proteases have proven to be useful as model proteins in protein engineering to better understand catalytic activity and stability of proteins. This project is a part of a larger research project which focuses on comparison of the enzyme aqualysin I (AQUI) from the thermophilic bacterium *Thermus aquaticus* with a homologous enzyme from the psychrotrophic *Vibrio*-species (VPR). In earlier parts of the project, it was observed that catalytic activity (k_{cat}) was doubled, the catalytic efficiency (k_{cat}/K_M) tripled and no loss of thermostability was observed in the variant AQUI_D98S. Computer simulations of the D98S variant showed that one of the histidine rotomers (His70), in the catalytic site, was more prevalent compared to the wild type. The variant N68Q showed similar results in computer simulations and it was hypothesized that it might have similar characteristics. This project was an attempt to define the N68Q variant's characteristics experimentally. The AQUI_N68Q variant was produced and the Michaelis-Menten kinetics, and thermostability experiments were conducted. The observed thermostability of the variant showed no significant deviations from the wild type. The kinetic characteristics for the variant changed less than expected with K_M dropping from 1.00 mM (± 0.05 mM) to 0.81 mM (± 0.05 mM) and the turnover-number (k_{cat}) dropping from 33.3 s^{-1} (± 4.6 s^{-1}) to 33.6 s^{-1} (± 5.7 s^{-1}) with no significant change in catalytic efficiency (k_{cat}/K_M) compared to the wild type. Observed kinetic parameters of the wild type of AQUI in this project were different from the kinetic parameters measured before in other projects in the same lab.

Table of Contents

List of Figures ..... **x**

List of Tables ..... **xi**

1 Introduction **1**
1.1 Temperature adaptations in proteins 2
1.1.1 Structural stabilization with increasing temperature 2
1.1.2 Catalytic activity and flexibility relationship in enzymes 7
1.2 Proteases 10
1.2.1 Serine proteases 11
1.2.2 Subtilin-like serine proteases (Subtilases) 13
1.2.3 Aqualysin I (AQUI) is a proteinase K-like protease 16
1.2.4 Comparisons between the thermophilic AQUI and the psychrotrophic protease VPR based on thermostability and flexibility 18
1.3 The aim of the project 21

2 Materials and methods **23**
2.1 Purification of AQUI 23
2.1.1 Purification table and Bradford protein-assay 24
2.1.2 SDS polyacrylamide gel electrophoresis analysis 24
2.2 Enzyme assay 25
2.3 Thermal inactivation of AQUI ($T_{50\%}$) 25
2.4 Melting point (T_m) determination with CD 26
2.5 Enzyme kinetics 27
2.6 Structure and classification of proteins 28

3 Results **29**
3.1 Purification of AQUI and its variants 29
3.1.1 SDS-page 32
3.2 Michaelis-Menten kinetics 33
3.3 Thermostability of AQUI 35
3.3.1 Thermal inactivation of AQUI ($T_{50\%}$) 35
3.3.2 Melting point (T_m) 37

4 Conclusions **39**

5 References **43**

List of Figures

Figure 1. (A) A thermodynamic diagram between folded and unfolded state in reversible denaturation of proteins. (B) Three thermodynamic models suggested for thermostability of thermophilic proteins. 3

Figure 2. The trade-off relationship between flexibility and thermostability in homologous enzymes. 9

Figure 3 The generally accepted mechanism for serine proteases. 12

Figure 4. The His, Asp, Ser catalytic triad in chymotrypsin from a mouse (1DPO) on the left and subtilisin Carlsberg (1SCN) from B. licheniformis *on the right.* 13

Figure 5. Enzyme precursor diagram for AQUI and VPR in the proteinase K family. 14

Figure 6. Crystal structure of unautoprocessed precursor of a subtilase from the hyperthermophilic archaeon T. kodakaraensis *(2EIP).* 15

Figure 7. Three-dimensional structure of AQUI (4DZT). 17

Figure 8. VPR (1SH7) and AQUI (4DZT) superimposed on each other. 19

Figure 9. Mutations near the His70 residues of AQUI (4DZT). 22

Figure 10. Example of a hydrophobic phenyl-Sepharose column purification for AQUI. 29

Figure 11. Example of a cationic CM-Sepharose column purification for AQUI. 30

Figure 12. Example of the (latter) second phenyl-Sepharose column used to concentrate (reduce the total volume of) the AQUI containing solution. 31

Figure 13. SDS-PAGE of purified AQUI samples collected from the purification procedures that were carried out. 32

Figure 14. Example of an experimental data fitted to a Michaelis-Menten nonlinear equation. 33

Figure 15. Example of an experimental data fitted to Lineweaver-Burk linear equation. 34

Figure 16. An example of a thermal inactivation experiment conducted on $AQUI_{wt}$. 36

Figure 17. Comparison of Arrhenius plot between $AQUI_{wt}$ *and AQUI N68Q.* 36

Figure 18. Normalized melting curves for AQUI N68Q on the left and $AQUI_{wt}$ *on the right.* 37

List of Tables

Table 1. Structural comparisons of hyper- and thermophilic proteins compared to mesophilic proteins. 7

Table 2. The diversity of serine proteases and catalytic mechanisms. 11

Table 3. Buffers used in the purification of AQUI. 24

Table 4. Purification table for $AQUI_{wt}$. 31

Table 5. Purification table for AQUI N68Q. 32

Table 6. Kinetic parameters of $AQUI_{wt}$ and AQUI N68Q derived from Michaelis-Menten equation. 34

Table 7. Kinetic parameters of $AQUI_{wt}$ and AQUI N68Q derived from Lineweaver-Burk equation. 35

Table 8. The thermal stability parameters of $AQUI_{wt}$ and AQUI N68Q. 37

Table 9. The main results of the project. 39

Table 10. Different results of $AQUI_{wt}$ from several published studies. 40

1 Introduction

The conditions on different parts of the Earth can differ greatly but are almost all filled with life. Of utmost importance for life is liquid water and an energy source. Evolutionary adaptive capabilities allow living organisms to thrive and proliferate at extreme environments which require trade-offs and specific adaptation. These organisms are called extremophiles. Most extremophiles are microorganisms but few are insects where cellular functions have evolved to operate under extreme conditions. These environments can be extreme with respect to salt concentration, acidity, alkalinity and temperature to name a few [1]. Many proteins extracted from these extremophiles have a potential to play an important role in industry as they can work under conditions where most proteins/enzymes from mesophilic hosts cannot [2]. The main focus of this thesis will be on temperature adapted proteins which are originated from organisms living in extreme temperature niches. These extremophilic living conditions range from permafrost soil in Arctic regions at -10° C to steam vents where the temperature can reach higher than 100°C [3, 4]. These extreme temperatures require the organism to overcome difficult obstacles. Sub-zero adapted organisms have to overcome the detrimental effects of the freezing of water and do so by using specific antifreeze proteins or other solutes, while thermophilic (heat-loving) organisms have more rigid and compact proteins to avoid denaturation, as an example [5, 6].

This project focused on aqualysin I (AQUI) which is an alkaline serine proteinase secreted by *Thermus aquaticus*. *T. aquaticus* is a thermophilic gram-negative rod bacterium which was first isolated from a thermal spring in Yellowstone Park. *T. aquaticus* has an optimum growth temperature at 70°C and a maximum growth temperature at 79°C. One well known protein from *T. aquaticus* is a DNA polymerase which has widely been used for the polymerase chain reaction (PCR) [7, 8]. Crystal structure of aqualysin I (AQUI) is available the Protein Data Bank (http://www.rcsb.org) at 1.95 Å resolution under the code 4DZT [9]. Two structural homologues of AQUI are part of this study, proteinase K from the mesophile (moderate-loving) fungus *T. album* and *Vibrio* proteinase from the psychrotrophic (cold-loving) *Vibrio* species. These proteins are all subtilisin-like serine proteinases (subtilases). Subtilases have been extensively researched and their most common industrial use is in laundry detergents. Subtilases have active sites that are geometrically very similar, as well as in their mechanism, to chymotrypsin related proteases, although their overall protein structure and their evolutionary history is different [10].

1.1 Temperature adaptations in proteins

As mentioned, microorganisms that have adapted to live and grow under different temperature conditions have adapted so that cellular function and metabolism is functioning optimally under different temperatures [1]. Despite the extreme temperature difference between the upper and the lower limits that are able to sustain life, all the proteins persisting at those temperatures are prescribed by the same codes, composed of the four bases, A, T, C, and G in their deoxyribonucleic acid (DNA) translated into amino-acid sequences which consist mainly of the same set of 20 different amino-acids. The difference between a cold adapted and a heat adapted protein is encoded in the sequence of the 20 different amino-acids [11].

The three-dimensional structure of proteins which serve the same function in cells can often be superimposed on each other even though the bacteria which the protein originate from thrive in completely different areas of the temperature spectrum. This is because those proteins have evolved from an ancestral protein which have accumulated many mutations over millions of years which allows the proteins to adapt to the different temperatures. Although proteins of the same origin are superimposable they may differ significantly in amino acid sequence and e.g. thermophilic proteins might share only 20-30% of its amino acid sequence with their mesophilic counterparts. This accumulation of changes in the sequence makes it harder to elucidate which changes result in stabilization. This difference in sequence could have neutral effects and may not affect activity or selective evolution. Selective evolution is not only bound to thermal stabilization but also other environmental aspects such as to salt concentration, acidity, alkalinity which might have caused great variability through the evolutionary path which complicates thermostability elucidation even further [12].

1.1.1 Structural stabilization with increasing temperature

Thermodynamic stability of proteins are measured by using Gibbs free energy (ΔG). We can assume two-states of conformations for the protein, a folded (f) or an unfolded (u) state:

$$\Delta G = \Delta H - T\Delta S \qquad \text{(I)}$$

The enthalpy (ΔH) and the entropy (ΔS) changes refer to the change in those parameters between the folded and the unfolded states [13]. Thermodynamic stability of a protein can be defined as the energy released because of unfolding:

$$\Delta G = RTlnK_U \qquad \text{(II)}$$

Where K_U is the equilibrium constant ($K_U = k_f/k_u$), R is the universal gas constant, and T is the absolute temperature. In brief terms, the thermodynamic relationship between a lower energy folded state and a higher energy unfolded state is partly the causation for folding of proteins (*Figure 1A*). This equilibrium can, however, only be observed in reversible denaturation of proteins. Reversible denaturation experiments can be conducted using denaturing agents such as guanidine chloride and urea, or with thermal denaturation given that the protein has not been post-translationally altered so that the thermodynamic driving force has been changed, e.q. by removing a part of the sequence or nicking of the polypeptide chain.

Three models have been suggested for thermostabilization of thermophilic proteins (*Figure 1B*). In the first model, the thermophilic protein has a greater stability over the entire temperature range compared to its mesophilic homologue. In the second model, the protein has more stability at a higher temperature, but lesser at lower temperatures while having the same maximum free Gibbs energy for unfolding. In the third model, the protein is more stable at all temperatures but has the same peak in Gibbs free energy compared to its mesophilic homologue. All these homologues have been observed in nature [14]. It is theorized that many thermophilic proteins are more kinetically stable than their mesophilic homologous counterparts, as well as being more thermodynamically stable. This means that the energy of the transition state (TS in *Figure 1A*) increases to gain more stability. Some thermophilic proteins have even been observed to be kinetically stable but not thermodynamically. This means that the unfolded state can be energetically more favorable than the folded state, but the transition state provides a high energetic barrier which can slow the denaturation process immensely [15].

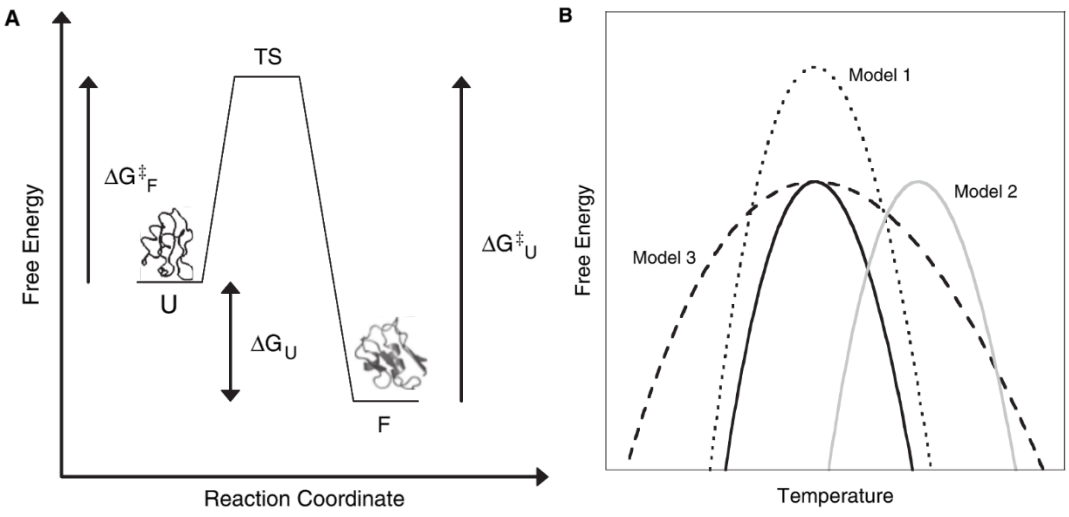

Figure 1. (A) A thermodynamic diagram between folded and unfolded state in reversible denaturation of proteins. (B) Three thermodynamic models suggested for thermostability of thermophilic proteins. The black solid line is the thermodynamic stability of a mesophilic protein which is a homologue of the three suggested models [14].

The hydrophobic effect is an important factor in conformational stability of proteins. The hydrophobic effect affects nonpolar molecules when introduced to a polar solvent such as water. Polar molecules form H-bonds with each other and water, but do not form H-bonds with nonpolar molecules. The lack of H-bonds formation between nonpolar molecules and water causes a highly ordered water structure around nonpolar molecules dissolved in water in order to fulfill as many H-bonds as possible. This formation of highly ordered water structure around nonpolar molecules causes nonpolar molecules to clump together to reduce the nonpolar surface area to reduce the energy of the system. This is thermodynamically unfavorable and causes the hydrophobicity effect [16]. Hydrophobic residues are, therefore, usually found inside the core of the protein where they can avoid interactions with the polar solvent water [17]. These multiple hydrophobic residues keep the core packed and contribute to the protein's folded state. The formed hydrophobic core is tightly packed and reaches almost full capacity in both meso- and thermophilic proteins which, therefore, excludes solvent particles such as water molecules or ions [18, 19]. A study comparing mesophilic and thermophilic homologous proteins measured no significant difference in the hydrophobic core of water soluble proteins in terms of gaps between atoms [19]. The force for hydrophobic interactions strengthens with increasing temperature at least under normal physiological temperatures (0-80 °C), but start declining after a certain maximum point and then disappears at extreme temperatures. Favorable exchange in entropy as a nonpolar molecule transfers from a polar solvent to a non-polar environment is the driving factor of the hydrophobicity effect at lower temperatures (~20°C). This is mainly due to the release of polar molecules such as water that form a highly ordered structure around the nonpolar molecule because the highly ordered water structure starts to dissolve (melt) and water starts to behave like bulk water at higher temperatures. Enthalpy becomes a greater stabilizing factor with increasing temperature. The enthalpy change for the transfer of a nonpolar molecule from water to a nonpolar environment is approximately zero at 20°C and consists of a favorable enthalpy change due to van der Waals interactions and an unfavorable enthalpy change due to disturbance of the highly ordered water structure formed around each individual nonpolar molecule. The enthalpy change due to disturbance of the water structure becomes less of a factor since water behaves more like bulk water as mentioned before, so the enthalpy change becomes more favorable with increasing temperature. This creates an enthalpy/entropy compensation where the Gibbs free energy involved in hydrophobic forces is less affected with temperature than either enthalpy or entropy are. Hydrophobic forces reach maximum strength around 100-140°C after which they start declining again. This causes hyperthermophilic proteins to resort more to other measures for stability [15]. This raises the question; what causes the great stability of hyperthermophilic proteins? This question cannot be answered in one single answer but instead in many small ones.

Disulfide bridges are a form of interactions that stabilize folded proteins by the means of decreasing the entropy of the unfolded state [20]. These bridges, however are susceptible to breaking at high temperatures of which 100°C is believed to be the absolute maximum

temperature before bond breaking. The maximum temperature of 100°C has been refuted since proteins with disulfide bridges have been found that work above that temperature. One of these proteins is a serine proteinase from the hyperthermophile *Aquifex pyrophilus*, which contains 8 cysteine residues and undergoes a much faster denaturation in the presence of a reducing agent. This suggests that the disulfide bridges are enhancing the thermostability of the proteins above 100°C. Disulfide bridges might therefore be protected by chemical shielding in thermophilic proteins [18, 21].

Salt bridges (ion bonds) are one form of a "stabilizing strategy" in the folding repertoire of proteins. Salt bridges have been shown to occur more often in certain thermophilic proteins when compared to their cold adapted counterparts. The strength of ion bonds can be estimated with Coulombs law:

$$E = k\,\frac{q_1 q_2}{D \times r} \qquad \text{(III)}$$

Where E is the energy between two charges, k is the Coulomb's constant, q is the unit charge, r is the distance between the two charges and D is the dielectric constant. The dielectric constant is different for each solvent used. The energy involved in an ionic bond decreases with increasing dielectric constant and or increasing distance. The dielectric constant is greater for polar solvents compared to nonpolar ones. The dielectric constant is quite large for water at around 80 at 25°C, but decreases with temperature to around 55 at 100°C. This opens new possibilities for stabilization at higher temperatures such as increased stabilization with salt bridges on protein surfaces [22]. Thus, even surface salt bridges may supply stabilization at high temperatures when the contribution by hydrophobic interactions is diminished [23]. The dielectric constant for salt bridges in proteins can be decreased even further and salt bridge strengthened by placing the salt bridge in lower dielectric conditions less exposed to solvent interactions [13]. Salt bridges that are buried are usually found to be more conserved compared to salt bridges on the surface. This might be because a misplaced charge inside the protein is more detrimental to the protein tertiary structure compared to a misplaced charge on the surface. Surface charges can possibly be stabilized by solvent interactions such as water molecules or ions [24]. These surface salt bridges were also often found to be longer than 4.5 Å which was previously thought to be the upper limit of ion bond stabilization. Especially, since at greater distance than 4.5 Å, acid-acid and base-base ion interactions are more likely to occur, which would repulse each other [25]. These repulsive ion interactions could prove detrimental to the structure of proteins. These detrimental acid-acid and base-base interactions can, however, be shielded by introducing an opposite charge in-between. This means that if acid-base-acid and base-acid-base bond networks are taken into consideration then the possible detrimental base-base and acid-acid ion interactions would be shielded and not detrimental to the protein's structure [25]. The strengths of selected salt bridges from hyperthermophilic and mesophilic protein homologues were compared by Xiao and Honig [26]. In every case, the hyperthermophilic homologue had more favorable electrostatic contribution to folding and stabilization. This applied even if

the hyperthermophilic protein had fewer ionic bonds than its homologue. Extensive stabilizing salt bridge networks, containing 16 ion-pair forming members, can be seen in thermophilic enzyme glutamate dehydrogenase from the bacterium *Pyrococcus furious* which has an extreme growth temperature of 100°C [27]. Similar surface ion networks can be seen in most hyperthermophilic proteins. Surface ion networks in thermophilic proteins are electrostatically optimized and cooperate to strengthen ion bonds in a network and minimize (acid-acid and base-base) repulsive ion interactions. With this in mind, one can see that ionic bonds can be immensely important for hyperthermophilic protein folding and thermostability [28].

As mentioned before, the structures of homologous proteins which have adapted to different temperatures can often be superimposed on each other. However, when this is done, no major overall differences are commonly seen [12]. Still, changes can be seen where amino acid exchanges are observed between homologous mesophilic and thermophilic proteins. Some of the most significant changes observed in such studies as relates to thermophilic proteins is the reduction of uncharged polar residues such as glutamine, asparagine, threonine, and serine [27]. Glutamine and asparagine are prone to deamination at high temperatures in the vicinity of serine or threonine [29]. The number of charged amino acids increases, which is because of stabilization from salt bridges. In one comparative study of homologous proteins, the average length thermophilic proteins was stated to be around 268 residues while mesophilic were 310 residues on average [30]. In another comparison study it was claimed that thermophilic homologues had deletions in surface exposed loops [31]. Reduction in surface loops and in overall length of proteins should in theory flatten the curve for the free Gibbs energy of unfolding as a function of temperature, which increases thermal stability as described in model 3 in *Figure 1*B. Another reasoning behind the shorter chain was thought to be shortening of solvent exposed loops. This decrease in loop size/length should reduce the entropy gained by denaturation which makes the denaturation less favorable. This is because a shortened loop has fewer conformations after denaturation than a longer one [27]. New data claims, on the contrary to earlier hypothesis, that there is no significant difference between the loop lengths of mesophilic and thermophilic proteins [19]. Das and Gerstein [32] suggested that the partial reason why thermophilic proteins are shorter on average is because thermophiles are often from the archaeal kingdom. This was claimed in their comparative study of homologous proteins from the archaeal, the eukaryote, and the eubacteria kingdom. The study stated that thermophilic proteins from the archaeal kingdom were shorter than the thermophilic proteins from the bacterial kingdom and both were shorter than mesophilic homologues from the eukaryote kingdom. Therefore, it was thought that the length of proteins increased going from archaeal to eukaryotic proteins rather than with temperature adaptation [32]. *Table 1* shows a systematic comparison of structural parameters of proteins adapted to different temperatures. The comparison was based on a study where three-dimensional structures of 29 thermophilic and 64 mesophilic protein subunits, which represented 25 protein families, were analyzed. Significantly fewer cavities, both in number,

volume, and surface area, were observed in proteins originating from bacteria with optimum growth temperature of 100°C compared to meso- and thermophilic homologues [27]. Number of unsatisfied H-bonds slightly decreased with temperature. Ion bonds increased especially relating to long range ion bonds. The polarity of the surface for thermophilic proteins increased compared to mesophilic homologues, but the polarity of the surface did not increase with hyperthermophilic proteins compared to mesophilic proteins. As mentioned before, this might be because hydrophobicity weakens at extreme temperatures [33].

Table 1. Structural comparisons of hyper- and thermophilic proteins compared to mesophilic proteins. Arrows show if the change in structure as considered insignificant (1), moderately significant (2), or highly significant (3) [33].

Property		**Correlation with temperature**	**Change in proteins from moderately thermophiles**	**Change in proteins from extreme thermophiles**
Cavities	Number	↓↓	0	↓↓↓
	Volume	↓	↑	↓
	Area	↓	↑	↓↓
Hydrogen bonds	Number	0	0	0
	Unsatisfied	↓	↓	↓
Ion pairs	<4.0 Å	↑↑	↑	↑↑↑
	<6.0 Å	↑↑	↑↑	↑↑↑
	<8.0 Å	↑↑↑	↑↑↑	↑↑↑
Secondary structure	α	0	↑	0
	β	↑	0	↑↑
	Irregular	↓	↓	↓
Polarity of surfaces	Exposed	↓↓	↑↑↑	0
	Buried	0	↑	↑

1.1.2 Catalytic activity and flexibility relationship in enzymes

Enzymes are defined as proteins that catalyze chemical reactions in living organisms. Without them, reactions under biological condition would be too slow to maintain life. These macromolecules have specific active sites which bind to specific substrates and catalyze specific reactions [11]. Compared to the size of the active site enzymes are quite large. Enzyme structures main function is to serve as a scaffold for their active site. The active site contains a microenvironment which is highly adapted to catalyze a specific reaction. To set up and stabilize the best geometrical microenvironment for catalysis, the enzyme must be big. The first step in enzyme catalyzed reactions is the substrate binding strongly to the enzyme. The bigger the enzyme compared to the substrate result in better binding. The

enzyme undergoes conformational changes resulting from the substrate binding, which makes the enzyme bind stronger to the transition state. The transition state is a high energy state in between substrate and product and is the main reason why reactions take long time in nature without a catalyst [34, 35]. The active site consists of side-groups of amino acids which are oriented and placed in a very specific manner for catalysis. It is possible to make minor changes to these few side-groups to get a new activity which has not been seen before [36]. The active site residues are not optimized for conformational stability in enzymes. Mutations in these residues, or those nearby can cause greater thermostability but catalytic activity may be lost in a trade-off. This is in fact observed among many naturally temperature-adapted enzymes. This means that thermophilic enzymes in some way trade off their catalytic activity for stability and vice-versa based on the active site alone [37]. Thermophiles are thought to have been the first life on Earth. If that conclusion is correct, then meso- and psychrophilic enzymes have most likely evolved from their thermophilic counterparts. In that process they traded off their stable active site for catalytic activity at lower temperatures [38]. The active sites are not the only thing to change with temperature adaptation. The whole structure of the enzyme has to adapt as well and that does not come without a cost. Enzymes need to fulfill the laws of physics where structural thermostability and rigidity may transform into less catalytic activity at lower temperatures. Cold adapted enzymes might need to trade-off their rigidity and thermostability to be able to catalyze reactions at colder temperatures. This is because flexibility and catalytic activity for enzymes are in direct correlation-relationship with each other. This means that heat-adapted enzymes are rigid and lose their flexibility and, therefore, their catalytic activity at lower temperatures. Even though enzymes generally undergo loss of thermostability when adapting enzymatic activity at lower temperatures and vice-versa it does not give credit to the whole story. Temperature adaptation of meso- and psychrophilic enzymes in laboratories has been carried out successfully and cold-active thermophilic variants have also been generated through directed evolution. Therefore, thermophilic enzymes in nature might lose their thermostability when adapting for activity at lower temperatures simply because there is no selective pressure for thermostability at lower temperatures and vice-versa (*Figure 2*) [12].

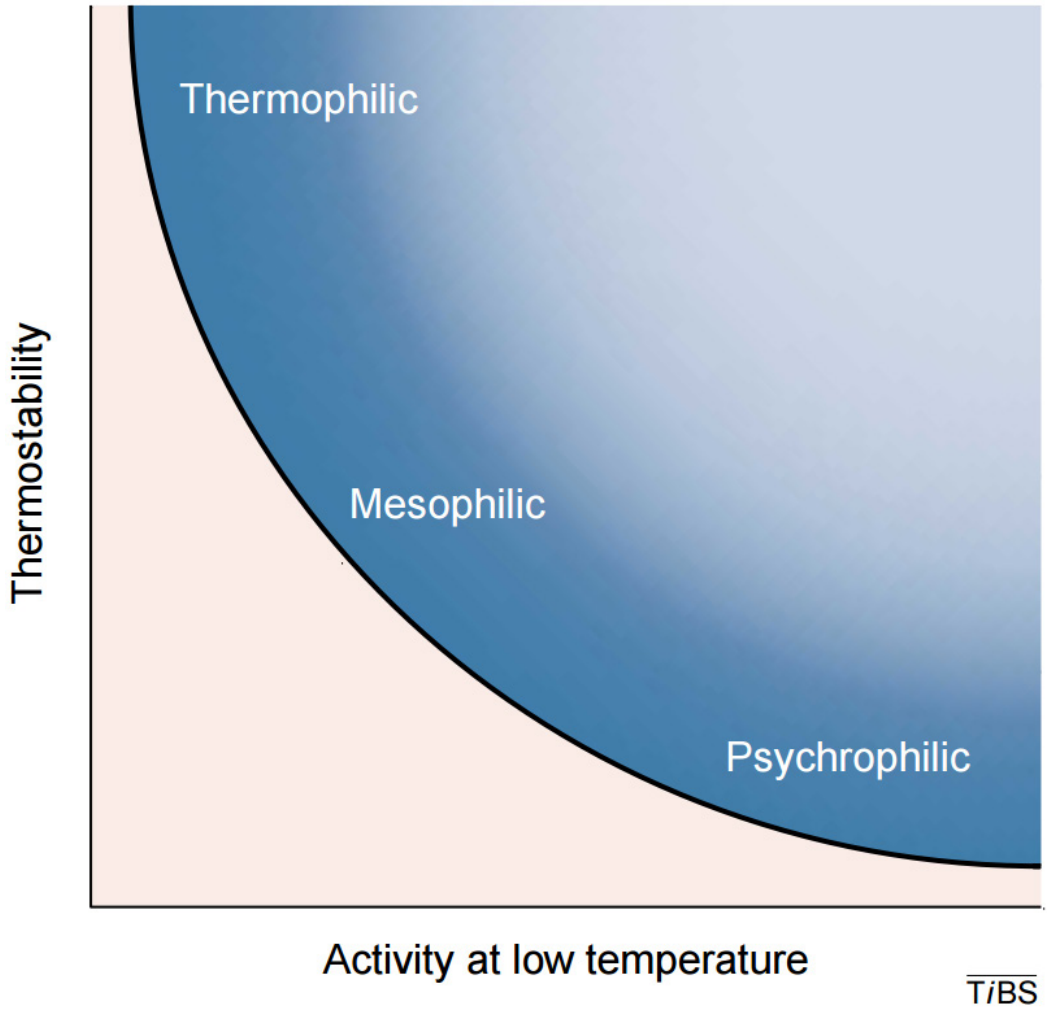

Figure 2. The trade-off relationship between flexibility and thermostability in homologous enzymes. More flexibility gives more catalytic activity at lower temperatures but correlates with lower thermostability. Natural enzymes lie in the darker blue shaded area whereas the pink shaded area isn't biologically viable since it is not active and or stable enough to support life. Enzymes that are both thermostable at high temperatures and catalytic at lower temperatures (lighter blue shaded) might be physically viable but aren't biologically relevant [12].

Although the flexibility is traded off at higher temperatures, the flexibility of thermophilic and mesophilic proteins is similar at their optimum activity temperature [39, 40]. Enzyme reactions occur on a millisecond timescale. Molecular motions associated with weak forces, which are in the most part responsible for maintaining protein structure and thermostability, occur in the pico- to nanosecond range. These forces are referred to as weak because their corresponding energy is similar to the environmental thermal energy (~k_BT). Although the difference between the timescales is great, these weak molecular motions in the protein are suggested to serve as a lubricant for the slower movements important for enzyme catalysis. The global flexibility for the thermophilic and mesophilic homologous enzymes lactate dehydrogenase from a rabbit and malate dehydrogenase from the hyperthermophile *M.*

jannaschii was observed to be essentially identical at their respective optimum temperatures [39]. It was suggested that the global flexibility is adjusted to the optimum working temperature. The flexibility of the mesophilic enzyme is, therefore, increased compared to the thermophilic enzyme to allow the weak forces, which are weaker at lower temperatures, to work more efficiently as lubricant for catalysis at lower temperatures [39]. Experimental results on adenylate kinase support this hypothesis. Adenylate kinase (Adk) from the mesophilic *E. coli* and thermophilic *A. aeolicus* are homologous. The active site of Adk is closed by two lids which require large domain motions. Navigation between a closed and an open state of the lids is the rate limiting factor for Adk. The energy involved in rotating between a closed lid and an open one is much greater than the molecular motions associated with the weak forces, closely connected to the global stability and the flexibility of proteins, as mentioned before. The smaller molecular motions are however the physical root of these larger domain motions. This was observed in nuclear magnetic resonance (NMR) results that showed similar atomic fluctuations and movement of hinges, responsible for the lid movement, in the mesophilic Adk at 20°C and the thermophilic Adk at 80°C. The mesophilic Adk has a comparable activity at 20°C to the thermophilic Adk activity at 80°C, thus the activity of the protein is directly connected to these weak atomic fluctuations [40].

1.2 Proteases

Proteolytic enzymes that cleave peptide bonds via hydrolysis are commonly termed proteases, proteinases or peptidases. Differences between these terms are historic since they have been forgotten and are often interchangeable. Peptidases are termed exo- or endopeptidases depending on where the hydrolysis takes place on the peptide. Endopeptidases, also termed proteinases, cleave internally and exopeptidases cleave externally [41]. Proteases serve a huge role in cellular functions in all three kingdoms of life, from viruses and bacteria to multicellular organisms such as humans. They are often very selective to which bond they break and when. This is used in many vital pathways such as in (pro)insulin activation by cleaving it into its active derivatives (mature descendant) [42] and blood clotting pathways to name a couple [43]. Proteases are classified by their mechanisms [44]. Seven major types are known; aspartic, cysteine, glutamic, metallo-, serine, asparagine, and threonine. The MEROPS peptidase database (http://merops.sanger.ac.uk) is useful in protease research and classification where the statistics on the site show that serine proteases are the most common with metallo- and cysteine proteases coming in second and third [45].

1.2.1 Serine proteases

More than one third of all known proteases are serine proteases, they are grouped into 13 clans and 40 families (*Table 2*) by MEROPS, based on statistically significant similarities [45]. The family name stems from the nucleophilic serine residue in the active site of all serine proteases [46] These proteases serve important biological functions such as blood clotting, digestion, fibrinolysis, development, fertilization apoptosis and immunity [47]. In most cases, the serine of the active site is made strongly nucleophilic by the means of a charge-relay system which is called the catalytic triad. In the chymotrypsin family, this happens because of an H-bond between an aspartic acid and a histidine, the histidine is also connected to the nucleophilic serine through H-bond. This creates polarity which can deprotonate and activate the serine, so it's more reactive towards esters and amides [48].

Table 2. The diversity of serine proteases and catalytic mechanisms. Where the # is the sequential number of the nucleophilic Ser [47].

Clan	Families	Representative member	Fold	Catalytic residues	#	Primary specificity	PDB
PA	12*	Trypsin	Greek-key β-barrels	His, Asp, Ser	195	A, D, F, G, K, Q, R, W, Y	1DPO
SB	2	Subtilisin, sedolisin	3-layer sandwich	Asp, His, Ser	221	F, W, Y	1SCN
SC	2	Prolyl oligopeptidase	α/β hydrolase	Ser, Asp, His	554	G, P	1QFS
SE	6	D-Ala–D-Ala carboxypeptidase	α-helical bundle	Ser, Lys	62	D-A	3PTE
SF	3	LexA peptidase	all β	Ser, Lys/His	119	A	1JHH
SH	2	Cytomegalovirus assemblin	α/β Barrel	His, Ser, His	132	A	1LAY
SJ	1	Lon peptidase	α+β	Ser, Lys	679	K, L, M, R, S	1RR9
SK	2	Clp peptidase	αβ	Ser, His, Asp	97	A	1TYF
SP	3	Nucleoporin	all β	His, Ser	na	F	1KO6
SQ	1	Aminopeptidase DmpA	4-layer sandwich	Ser	250	A, G, K, R	1B65
SR	1	Lactoferrin	3-layer sandwich	Lys, Ser	259	K, R	1LCT
SS	14	L,D-Carboxypeptidase	β-sheet+β-barrel	Ser, Glu, His	115	K	1ZRS
ST	5	Rhomboid	α-barrel	His, Ser	201	D, E	2IC8

The catalytic triad in the trypsin-clan (PA) is geometrically the same as in three other clans (SB, SC, and SK) although the folds of the proteins belonging to the four clans are entirely different. This suggests four different evolutionary paths that gave the same catalytic triad structure. Other serine proteases use other ways to enhance the serine residue nucleophilicity, some of which use a simpler dyad system where lysine and histidine is routinely used [46].

Chymotrypsin is a part of the PA clan and was the first serine protease that was explored in detail. Chymotrypsin is related to trypsin and elastase which are all proteases used in digestion and are produced in the pancreas [49]. The trypsin fold consists of two six stranded β-barrels which can be seen on the left in *Figure 4*. The active site is positioned between the two barrels and is decorated by 8 surface loops. Connected to the active site is the primary specificity pocket which binds the side chain closest to the cleavage site of the peptide

undergoing cleavage. Within the same fold the trypsin primary specificity pocket preferably binds positively charged side chains of arginine and lysine residues, but chymotrypsin is specific to aromatic side chains of phenylalanine, tyrosine, and tryptophan residues. Both trypsin and chymotrypsin are produced as inactive precursors and undergo several cleavages that cause the activation of the protease. The peptide bond joining arginine 15 and isoleucine 16 is cleaved and the new N-terminus forces a conformational change that activates the enzyme [46]. The active protease then undergoes further cleavage to remove dipeptides 14-15 and 147-148, which results in the stable form of the protease. The final mature protease consists of three separated chains that are linked together through two disulfide bonds [50].

All proteases have to overcome three obstacles via hydrolysis of peptides: (1) Amide bonds are stable due to their double-bond character. Proteases overcome this by activating the carbonyl oxygen with a general acid, this also distorts the resonance stabilization (2). Water is a poor nucleophile; proteases activate water, usually with a general-base (3). Amines are usually poor leaving groups but are protonated before expulsion. The rate of peptide hydrolysis by serine proteases is ~10^{10}-fold greater than for the uncatalyzed reaction [51]. The generally accepted mechanism of peptide hydrolysis by serine proteases can be seen in *Figure 3*. The side group of the active site serine makes a nucleophilic attack on the carbonyl group carbon involved in the peptide bond and creates a short-lived tetrahedral intermediate which expels the N-terminal part of the cleaved peptide chain and forms an acylenzyme. The serine is made more nucleophilic by the histidine which acts as a general-base and removes the proton that leaves with the first product. The acylenzyme then undergoes a nucleophilic attack via a hydroxide ion from a water which is activated by protonation of the same histidine. This creates a short-lived tetrahedral intermediate and the rest of the peptide is expelled. The oxyanion created in the tetrahedral intermediates is stabilized by two NH groups from the main peptide chain. Again, on expulsion of the second tetrahedral intermediate, the His residue acts as a general-acid and donates a proton to the leaving group. The protonated histidine in the two tetrahedral intermediates is stabilized by aspartic acid, thus explaining how the catalytic triad works [51].

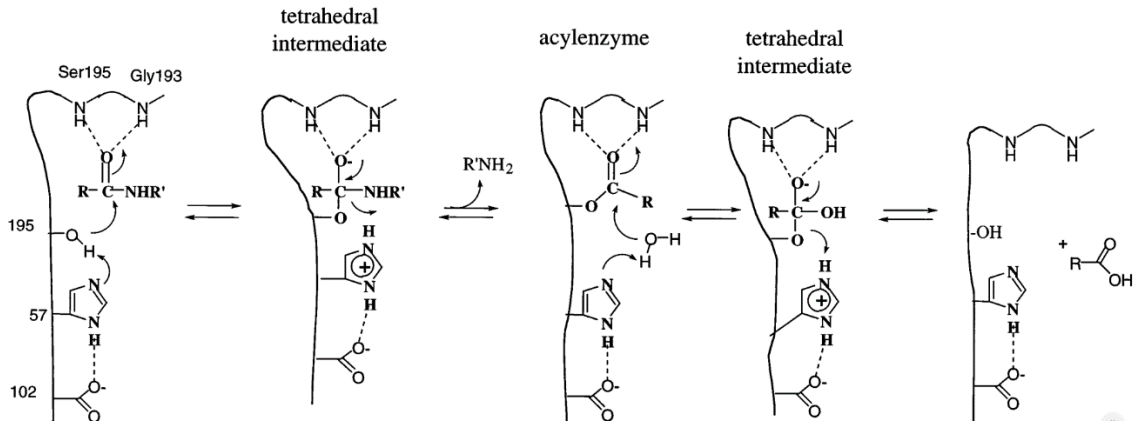

Figure 3 The generally accepted mechanism for serine proteases [51].

1.2.2 Subtilin-like serine proteases (Subtilases)

Subtilisin Carlsberg is a protease that was originally discovered in the Gram-positive bacterium *Bacillus subtilis*. Subtilisin represents the more common S8 family (subtilases) of two families in the SB clan. Sedolisin is the other family (S53) under SB clan. Proteases in the S53 family use glutamic acid instead of the common histidine in their catalytic triad and are active at low pH. The SB clan is prevalent in bacteria and plant genomes with few representatives in the animal kingdom in contrast to the PA clan which is more prevalent in eukaryotes [45]. Subtilisin along with chymotrypsin from the PA clan were among the first protein crystal structures to be determined [52, 53]. The His, Asp, Ser catalytic triads for subtilases and chymotrypsin-like proteases are exactly the same geometrically, although the protein structure is entirely different. The order of the residues of triad is also different, in the subtilases the order is Asp-His-Ser, whereas in chymotrypsin it is His-Asp-Ser. *Figure 4* shows the difference between the folds of the two proteases with the same catalytic triad. Subtilases tend to be nutrition oriented where they are secreted outside the cell for breakdown of extracellular proteins. Ten proteins from the SB clan have been identified in humans. Proprotein convertase subtilisin-like kexin type 9 (PCSK9) is one of them and it regulates levels of low-density lipoproteins (LDL). Mutation in PCSK9 that increase its activity may lead to coronary heart disease and hypocholesterolemia. Inhibiting PCSK9 may therefore be promising as a treatment for patients carrying this mutation [47].

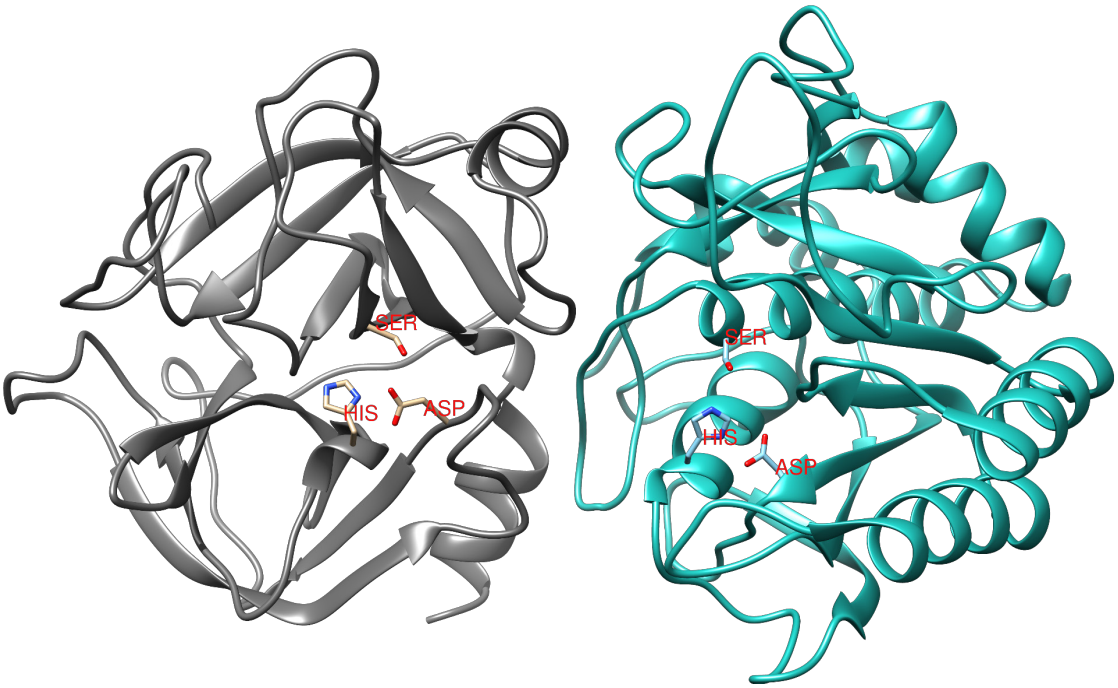

Figure 4. The His, Asp, Ser catalytic triad in chymotrypsin from a mouse (1DPO [52]) on the left and subtilisin Carlsberg (1SCN [53]) from B. licheniformis *on the right. The folds are drastically different but the catalytic triad is geometrically identical.*

Subtilases have an α/β fold where the fold alternates between α-helices and β-strands. In the middle of the fold, there is a parallel β-sheet which is surrounded by α-helices to create a three layer αβα sandwich. The parallel β-sheet consists of 7 strands in the order 2314567 with a rare left-handed crossover connection between strand 2 and 3 which is necessary in subtilases to form a proper active site. The helices and strands are connected via loops on the surface of the enzyme. According to the CATH database [54] (http://www.cathdb.info/), the topology is consistent with the Rossmann-fold. Two loops originating from β-strands turn away from each other in the parallel β-sheet which forms a switch-point. The switch-point forms a crevice which hosts the active site [55].

All prokaryotic subtilases are synthesized as inactive precursor proteins. In the case of aqualysin I (AQUI) and VPR from the proteinase K (sub)family of subtilases, the synthesized precursor protein contains four domains (*Figure 5*). There is one protease domain, two prodomains, and a signal peptide apart from the subtilase domain. The signal peptide allows the precursor protein to be transported through the inner membrane. The N-terminal prodomain works as an intramolecular chaperone (IMC) which is essential for the correct folding of the enzyme. After folding it is cleaved off by autocatalysis. The N-terminal IMC and the signaling peptide are found in all subtilases but the C-prodomain found in AQUI and VPR is not as prevalent in all subtilases like the other two prodomains [55]. The C-terminal prodomain helps with extracellular secretion. Deletion of the domain negatively affects secretion of the protein in its originating thermophilic organism, *T. aquaticus,* but had no effect on the cloned protease production in the mesophilic bacterium *E. coli* [56].

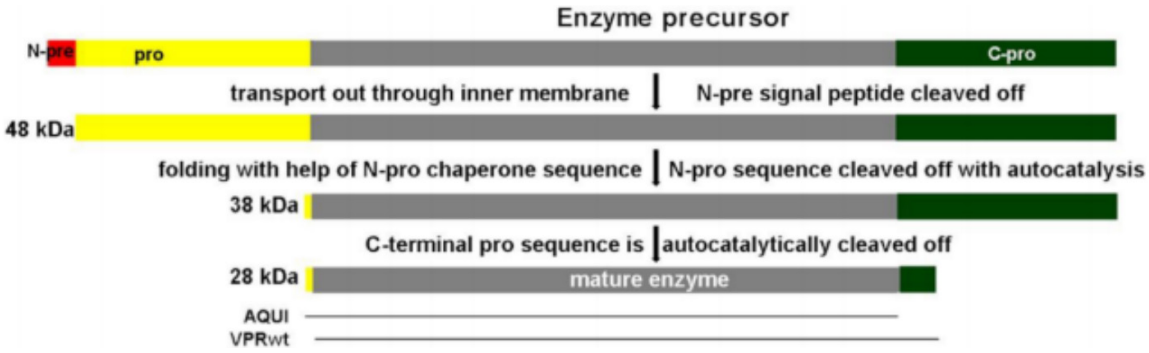

Figure 5. Enzyme precursor diagram for AQUI and VPR in the proteinase K family. The signaling peptide (N-pre) is shown in red. The folding N-terminal prodomain is shown in yellow and is found in all subtilases. The C-prodomain from AQUI and VPR shown in green. The final mature protease domain is shown in gray [55].

The IMC prodomain can assist with folding even though it is not linked to the protease chain. When bound it acts as an inhibitor. The prodomain consists of a four-stranded antiparallel β- sheet and two three turn helices. The IMC domain binds tightly to the protease to form a complex. The β-sheet of the prodomain is tightly bound to an α-helix on the protein (*Figure*

6). Two acidic residues in the prodomain form caps on the N-terminal of two helices. The prodomain mediates a transition from an unfolded state to a folded state through a high energy transition state. This high energy transition state effectively becomes a kinetic barrier which keeps the protein folded after the prodomain has been cleaved off. Once the prodomain is cleaved off subtilases become thermodynamically unstable but very kinetically stable. Such kinetic stabilization is characteristic among other secreted proteases such as proteases from the PA clan. The energy barrier corresponding to the kinetic stability makes it unfeasible for equilibrium constant determination under practical experimental conditions. It would simply take too long to wait for equilibrium between folded and unfolded state. The energy barrier along with the thermodynamics of the folding causes the denaturation of the folded state to be practically irreversible [55].

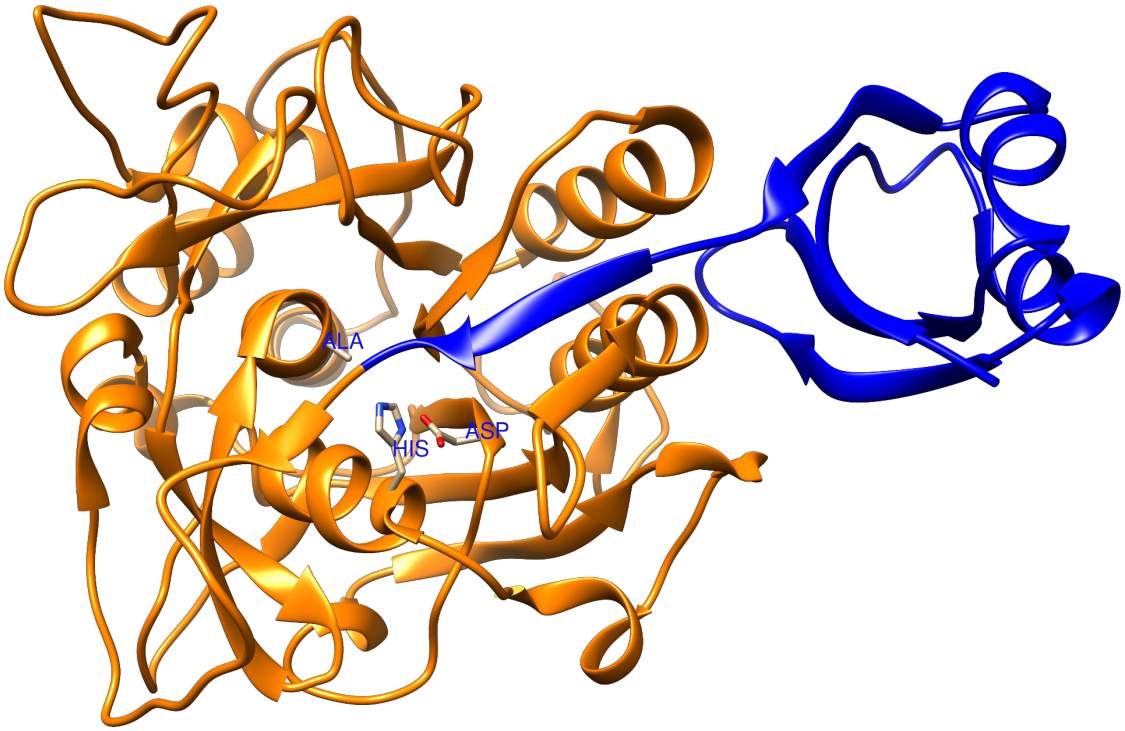

Figure 6. Crystal structure of unautoprocessed precursor of a subtilase from the hyperthermophilic archaeon T. kodakaraensis *(2EIP). The serine residue in the catalytic triad has been mutated to alanine to remove most of the catalytic activity. The mature domain is colored orange and the N-terminal IMC chaperone is colored blue. It lies over the active site and undergoes auto proteolytic cleavage during activation of the protease [57].*

The native state of these proteases is furthermore stabilized by calcium binding. Calcium binding sites can be found in nearly all subtilases and two sites are classified as being present in all *true* subtilases [58]. These binding sites are termed Ca1 and Ca2, where Ca1 binds the calcium ion stronger. The calcium binding energy can support conformational stabilization for the native state but not the high-energy transition state. Proline isomerization between

cis- and trans-isomers is a major rate limiting factor in folding of many proteins. There are three prolines that are in close contact to Ca2 binding site in subtilisin and they all need to be in the correct isomer conformation for calcium binding to Ca2. With the prodomain present those isomers are stabilized but not without it. This raises the energy barrier between the high energy transition state and native state which stabilizes the protein drastically [56].

Subtilases (S8) are subdivided into six subfamilies based on their sequence homology. The subfamilies are subtilisin, thermitase, proteinase K, lantibiotic peptidase, kexin and pyrolysin, termed for historical reasons for being the first protease identified in the subfamily [59]. The proteinase K subfamily will be of special interest since the main protease in this research, AQUI is classified to that family. Among other related proteins of interest in the same subfamily is the mesophilic proteinase K and a psychrotrophic protease, VPR, from the *Vibrio* species PA44 [10].

The proteinase K family consists of mostly secreted endopeptidases which are only found in fungi, yeast, and gram-negative bacteria as of yet. The peptidases from this family share a relively high sequence identity of over 37% [59]. The representative member of this family is the proteinase K from the fungus *Tritirachium album* with a alkaline pH optimum of 7.5-12.0 [60].

1.2.3 Aqualysin I (AQUI) is a proteinase K-like protease

Aqualysin I (AQUI) is an alkaline subtilase secreted by the thermophilic bacterium *T. aquaticus*. It was first characterized in Japan in 1983 [61, 62]. It is produced as a precursor protein with an intramolecular chaperone which is cleaved off via autoproteolytic cleavage (*Figure 5*). The C-terminal prodomain is also part of the precursor which helps with extracellular secretion [55, 56]. AQUI contains two calcium binding sites [9]. Experimental results show that one has a much stronger association constant of more than 10^{10} M^{-1} compared to the weaker one of 3.1×10^{3} M^{-1}. The weaker calcium binding site is crucial for thermal stability. Removing of the calcium from the weaker binding site with gel filtration results in a half-life of 31 minutes at 80°C, whereas with addition of calcium to the weaker site, no thermal inactivation was observed over a 3 hour period at 80°C [63]. The stabilization of the weaker site is not limited to calcium, as it can bind a variety of metal cations (Na^{+}, Nd^{3+}, La^{3+} and Sr^{2+}) and in some cases it suffers no more destabilization than if calcium was bound. This suggested that the weak binding site is flexible. It was suggested that the stronger binding site corresponds to the stronger Ca1 site of proteinase K which corresponds to the weaker Ca2 site in all *true* subtilases [63]. However, newer more credible evidence suggests that it is the other way around. Calcium binding sites Ca3 and Ca1, termed after VPR (*Figure 8*), in a psychrotrophic proteinase K- like protease (SPRK) from a psychrotrophic *Serratia* sp. are analogous to the binding sites found in AQUI. The calcium

binding site Ca3 is novel in these proteases as it is not present in proteinase K or other subtilases. The Ca1 site corresponds to the reserved weaker Ca2 site in all *true* subtilases. Crystal structure of SPRK suggests the presence of three calcium binding sites but only the Ca3 site was occupied. The fact that the Ca3 site was the only calcium binding site occupied strongly suggests that it is the strongest calcium binding site in these proteases and not the reserved Ca1 site as suggested before. [64]. *Figure 7* shows the structure of AQUI with calcium bound which was obtained from the Protein Data Bank with the pdb code 4DZT [9, 55].

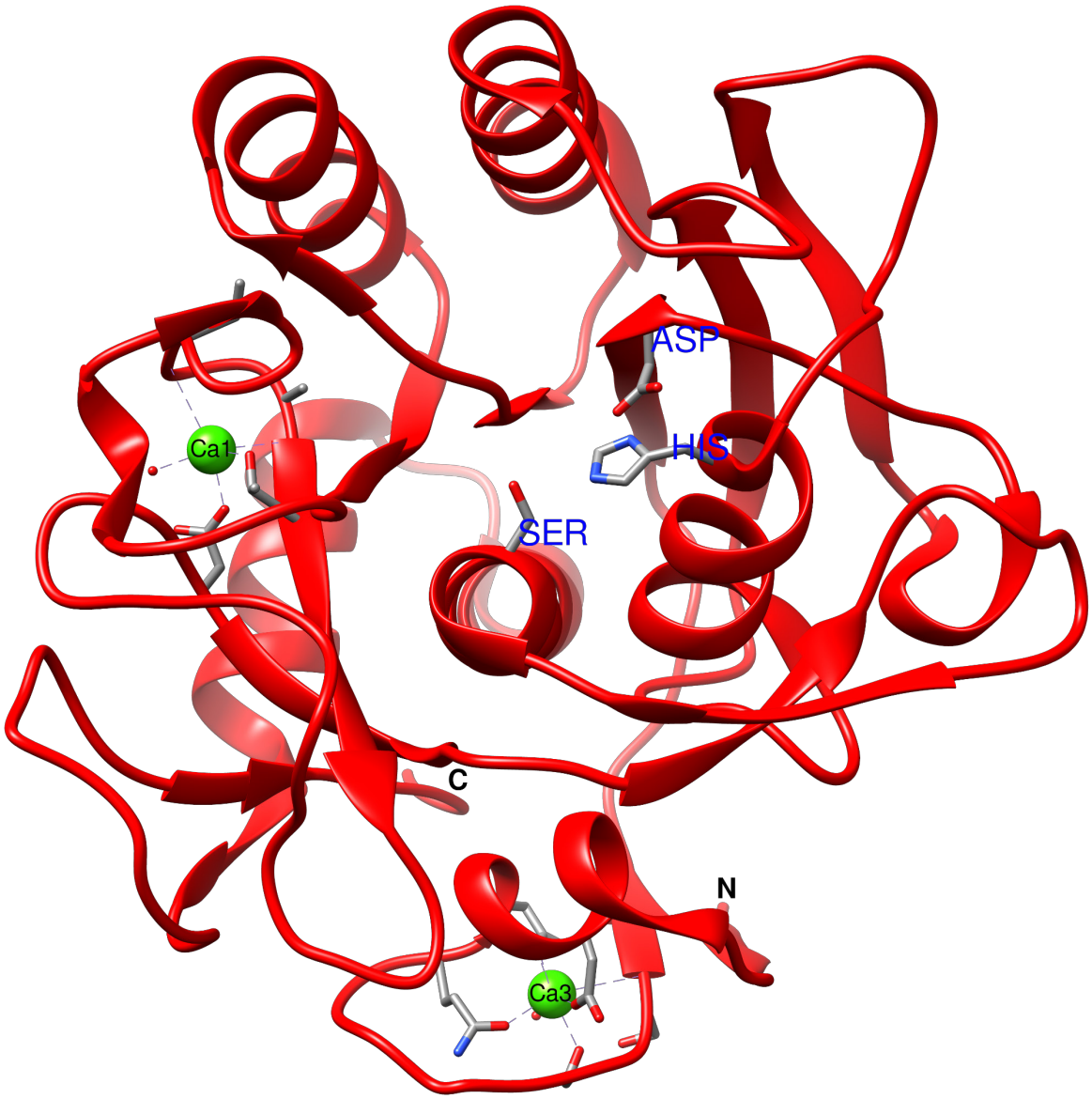

Figure 7. Three-dimensional structure of AQUI (4DZT [9]). The catalytic triad Asp39, His70, and Ser222 along with the two calcium binding sites can be seen. The N-terminal (N) (can be seen right of Ca3 binding site) and the C-terminal (C) are marked after their initials.

The catalytic triad Asp39, His70, and Ser222 of the active site are positioned in a crevice in the middle of the protein (*Figure 7*). The crevice is formed by a switch-point from two loops, originating from β-strands 1 and 4 in the β-sheet, turning away from each other. This topology is consistent with the Rossmann fold. The oxyanion hole consists of Asn157 which stabilizes the tetrahedral intermediate during hydrolysis of peptides (*Figure 3*). The Asn residue is positioned on a loop, which originates from β-strand 5 in the β-sheet, overlaying the active site (*Figure 7*). Two β-strands, number 2 and 3, form a left-handed crossover where we can see the second β-strand crossing over and transition to the α-helix containing His70. This left-handed crossover is a characteristic of subtilases and is quite rare in other proteins. The protein has a αβα sandwich fold like all subtilases, the β-sheet, which is surrounded by α-helices, can be seen in the middle. The stronger calcium binding site near the N-terminal, termed Ca3 after VPR protease terminology [10], links the first α-helix and the succeeding loop together in a bundle. The weaker Ca1 binding site can be shown in *Figure 7* between two loops. This calcium binding site has been proven to be crucial for the stability of the protein [64]. It corresponds to the weaker Ca1 binding site in proteinase K and Ca2 binding site in all *true* subtilases. The calcium is bound through negatively charged atoms in residues, most often by aspartic acid or the carbonyl oxygen of asparagine. Binding site Ca1 is also partly coordinated by a threonine sidechain [55].

1.2.4 Comparisons between the thermophilic AQUI and the psychrotrophic protease VPR based on thermostability and flexibility

For greater insight into the structural origin of thermostability and activity at lower temperatures it can prove useful to compare homologous proteins that have adapted to different temperatures. Homologous proteases which are comparable with AQUI are the mesophilic proteinase K and the psychrotrophic protease VPR. These proteases are all a part of the same proteinase K subfamily in the SB clan. VPR is a subtilase originating from *Vibrio* species strain PA-44. The bacterium has an optimum growth temperature at 19°C [10]. The protease structure has been determined and is available in the Protein Data Bank at a 1.84 Å resolution with the pdb code 1SH7. VPR can be distinguished from its heat-adapted relatives by its strong anionic character. The anionic character is caused by high occurrence of uncompensated negatively charged residues on the surface. Both the thermophilic AQUI and the psychrotrophic VPR were found to have more extensive hydrogen bonds and ion pair interactions compared to the mesophilic proteinase K which supports a dual role of electrostatic interactions in adaptations to both temperature extremes [65]. VPR has three calcium binding sites [55, 65]. Two of them are the same as in AQUI and is seen in *Figure 7*. *Figure 8* shows VPR and AQUI superimposed on each other with the additional calcium binding site labeled as Ca2. The superimposed overall structures of VPR and AQUI are almost identical. VPR has an extra calcium binding site (Ca2) which is neither seen in AQUI

nor proteinase K but has been observed in other close relatives [55]. Some loops are differently coordinated, or are shorter in AQUI. The C-terminal is extended (*Figure 5*) and, therefore, appears to point more outwards on VPR as compared to AQUI [9, 55, 65].

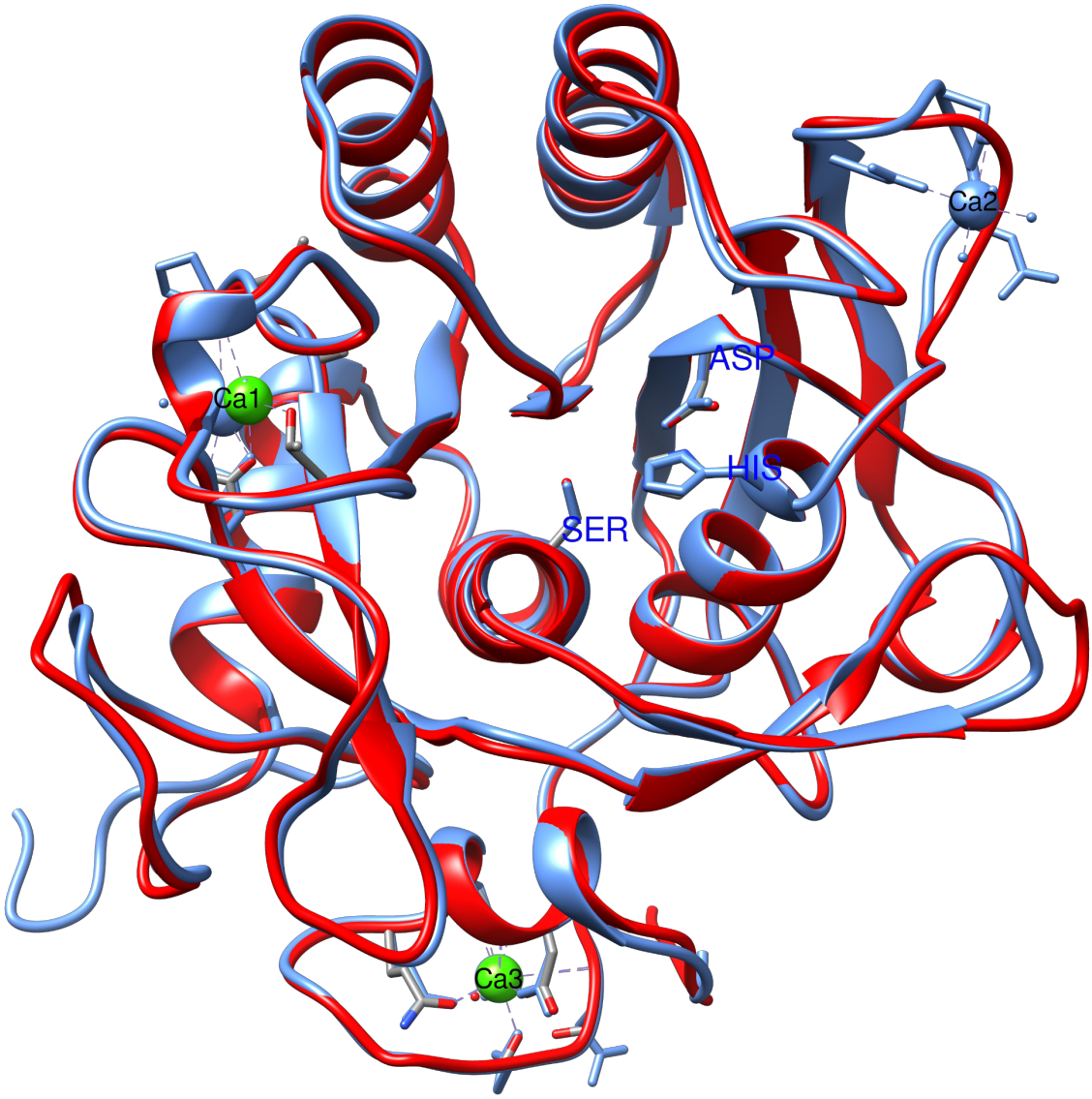

Figure 8. VPR (1SH7 [65]) and AQUI (4DZT [9]) superimposed on each other. VPR is colored blue while AQUI is colored red with sticks colored after their corresponding element. The calcium binding site Ca2 in VPR is not found in proteinase K or AQUI.

Extensive comparative studies of VPR and AQUI have been conducted to gain a better understanding of thermostabilization. One of the most crucial mechanism of thermostabilization is through salt bridges, often in networks or clusters. One salt bridge of interest which may elucidate thermostabilization is the salt bridge Asp17-Arg259 that is present in AQUI but not in VPR. The salt bridge links a loop near the N-terminal Ca3 binding

site to a loop much closer to the C-terminal. Mutation of the negatively charged aspartic acid (Asp17) to an uncharged asparagine (Asn17) (as present in VPR) of a similar size was predicted to destabilize the protein by 1.6 kcal/mol by computer calculations [66]. The observed $T_{50\%}$ (the temperature at which half of the protein activity is lost in half an hour) of the variant (D17N) was approximately ~9°C lower compared to the wild type of AQUI ($AQUI_{wt}$). No significant difference in catalytic activity was measured in the same variant [66].

The concept of the stabilizing Asp17-Arg259 salt bridge led to an attempt to try to introduce it into the structure of VPR. This was done with an Asn15Asp (N15D) mutation, and the resulting variant was studied with respect to stability and activity. The mutation proved to be successful in creating a new salt bridge as the melting point (T_m) of an inhibited VPR N15D variant along with the $T_{50\%}$ increased ~3°C. The introduction of the salt bridge into VPR doubled the catalytic efficiency, although the increased activity was not significant in a double variant that had also been stripped of the last 15 residues from the C-terminus. These 15 residues are not present in AQUI as it is shorter (*Figure 5*), so removing them makes comparison between the two homologues more reliable. Two prolines were also introduced near the N-terminal of VPR through a mutation of residues N3P/I5P. It has been suggested that the N-terminus is more flexible in VPR and is also cleaved two residues shorter as compared to AQUI. The N3P/I5P variant extended the N-terminus to the length of AQUI [67]. A possible formation of a β-sheet at the terminus in the variant might also have decreased the flexibility of the N-terminus. The side effect, however, was a decrease in activity. The variant N3P/I5P with the C-terminal truncated was approximately ~4°C more stable than the C-terminal truncated variant ($VPR_{\Delta C}$) without the proline mutations in terms of T_m. The variant $VPR_{\Delta C}$/N3P/I5P/N15D had approximately ~5°C higher T_m than the $VPR_{\Delta C}$ variant without the three mutations. The VPR triple variant had half of the activity compared to the $VPR_{\Delta C}$ [67]. Added thermostability through direct rationalized changes by comparing cold adapted proteins to their thermophilic homologues can, therefore, be a remarkably good tool in protein engineering.

Comparing VPR and AQUI can also help in understanding new ways of increasing the catalytic activity of AQUI without affecting the thermostability. Mutating the Asp98 residue to Ser98 (D98S) (*Figure 9*) in AQUI was aimed at deleting a salt bridge that was proposed to be present in AQUI but not in VPR. Computer simulations of the area around Asp98 suggested a network of salt bridges in AQUI. The Asp98 site in AQUI is occupied by Ser in VPR. The D98S variant was therefore thought to cause a reduction in the number of H-bonds that would lead to more conformational flexibility around the active site more comparable to VPR. The D98S variant was a success with no significant reduction in thermostability and doubled turnover number (k_{cat}) and tripled in catalytic efficiency (k_{cat}/K_M) compared to the wild type. The variant affinity for the substrate also increased as the Michaelis-constant, K_M, decreased [68].

1.3 The aim of the project

This study is a part of a larger research project being conducted in Magnús Már Kristjánsson's laboratory that focuses on subtilisin-like (subtilases) serine proteases from the proteinase K subfamily; the homologous proteases aqualysin I (AQUI) originating from thermophilic *Thermus aquaticus* and a protease originating from a psychrotrophic *Vibrio* species (VPR). The two proteases have been well defined with crystal structures available for both enzymes. The research team has conducted a number of site directed mutagenesis experiments and observed changes in properties in the resulting variants [64-69]. The mutations are usually chosen with both the psychrotrophic VPR and the thermophilic AQUI structures in mind when hypotheses are set for the expected results. The structure of AQUI gives insight into thermostability which might be applicable to improve the properties of VPR. Certain characteristics from VPR structure on the other hand, might be suitable for higher activity at lower temperatures for AQUI. Both these cases have been carried out successfully with experiments.

Mutation of AQUI in Asp98 to Ser (*Figure 9*) as seen in VPR (AQUI_D98S) resulted in doubled turnover number k_{cat} and it almost tripled the catalytic efficiency k_{cat}/K_M. The mutation did not however have significant effects on the stability of the protease. The mutation causes reduction in the number of H-bonds positioned around the active site of the protease. This H-bond reduction may cause more conformational flexibility at this site which may result in more activity. Computer simulations of molecular dynamics and metadynamics in AQUI have estimated that His70 residue, which is a part of the catalytic triad found in serine proteases, has two possible rotomers. One of the rotomers is perceived to be more active because it is more prevalent in a computer simulation of the active variant AQUI_D98S. The hypothesis of this project was directed at molecular attributes that may contribute to the more active rotomer while reducing the less active rotomer. It was proposed that factors which contribute to more prominence of the active rotomer should increase the activity of the protease. The reduction in the number of H-bonds in the variant AQUI_D98S causes a conformational change in the nearest vicinity of the active site that supports the active rotomer. Computer simulations conducted by a collaborator in this project Dr. Elena Papaleo showed that the active rotomer can be promoted by localized changes in the nearest proximity of His70. If the computer simulations are correct then the mutation of the residue Asn68 to the larger Gln68 should in theory reinforce the more active His70 rotomer (*Figure 9*). Computer simulation of a mutation to the smaller Ser68 residue suggested in contrast to the N68Q mutation a decrease in protease activity.

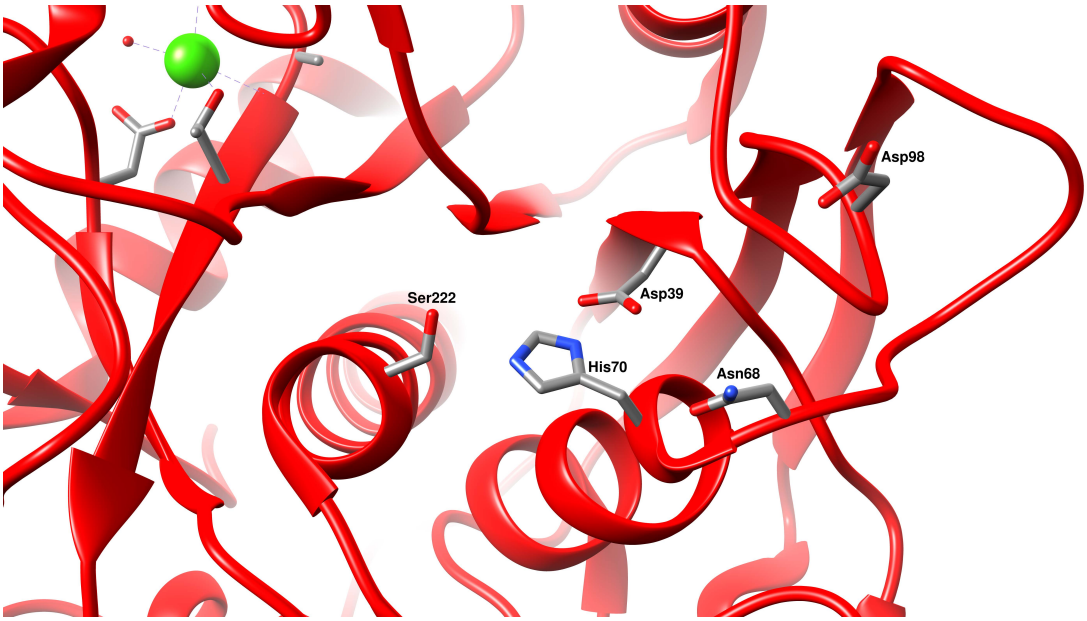

Figure 9. Mutations near the His70 residues of AQUI (4DZT [9]). The Asp98Ser variant had double activity compared to wt. In this project we characterized the Asn68Gln variant which has shown promise to do the same in computer simulations.

In this project, the Asp68Gln (AQUI_N68Q) variant was produced and purified and its characteristics compared to wild type enzyme ($AQUI_{wt}$) with respect to stability and activity. Kinetic characteristics were observed using succinyl-AlaAlaProPhe-p-nitro-anilide as a substrate. Denaturation of the proteases with temperature was observed using a CD-spectrometer to estimate the melting point (T_m). $T_{50\%}$ was observed by thermal inactivation of the proteases at different temperatures.

2 Materials and methods

2.1 Purification of AQUI

AQUI was synthesized in *E. coli* strain BL21 by Kristinn Ragnar Óskarsson according to Ellertsson [68]. Frozen pellets containing AQUI in *E. coli* cells were thawed in 50 ml of buffer A (25 mM Tris, 10 mM $CaCl_2$, pH 8) containing one mg/ml of lysozyme and one µg/ml of DNA-ase. The cell mixture was then shaken for two hours at room temperature, quick-frozen in liquid nitrogen (N_2), thawed and quick-frozen in liquid nitrogen again, this cycle was done three times. After the third quick-freeze in nitrogen, the mixture was allowed to thaw and was shaken overnight at 4°C. The crude cell extract was incubated at 70°C for an hour to allow the *E. coli* proteins to denature and AQUI to become free from the inner cellular membranes. The heating process also promotes the autocatalytic cleavage off of the C-terminal and, therefore, activates the enzyme. The mixture was centrifuged at 10000 x*g* for 20 minutes and the precipitate was discarded. The remaining supernatant was made one molar in ammonium sulfate (buffer B; *Table 3*) and immediately spun again at 10000 x*g* for 15 minutes. The insoluble impurities were once again discarded before column purification.

The ammonium sulfate protein solution was applied to a phenyl-Sepharose hydrophobic column (70 ml) that had been equilibrated with buffer B (1 M ammonium sulfate, 25 mM Tris, 10 mM $CaCl_2$, pH 8). Impurities were eluted off the column stepwise; first with buffer B, followed by 0.5 M ammonium sulfate in buffer A and finally with buffer A, without ammonium sulfate. The volume for each eluate mixture was at least two column volumes and were run until no further peaks were observed at 280 nm which was continuously monitored using BioLogic-LP workstation from Bio-Rad. Fractions containing AQUI were eluted from the column with 50:50 ethylene glycol:buffer A and collected in ~4 ml fractions in test tubes. Test tubes containing over 0.2 U/ml were collected and dialyzed overnight at 4°C in buffer C (25 mM MES, 1 mM $CaCl_2$, pH 6).

The dialysate containing AQUI activity in buffer C was applied to a CM-Sepharose ion-exchange column (70 ml) that had been equilibrated with buffer C. The column was eluted with buffer C to get rid of loosely bound proteins then with a salt gradient from 0 to 500 mM NaCl to elute AQUI from the column. The eluate was collected in ~4 ml fractions and fractions containing more than 0.2 U/ml were collected. The collected sample was equilibrated with one molar ammonium sulfate and loaded onto another smaller phenyl-Sepharose column (10 ml). AQUI was then eluted off the column in the same way as before only now in smaller quantities containing more concentration of purified AQUI. The eluate was collected in ~2 ml fractions and fractions containing more than 1.5 U/ml were collected. The protein peak can be split into stronger concentration and weaker concentration aliquots

as different concentration are useful to conduct different experiments. Table 3 summarizes the buffer compositions.

The collected eluate from the later phenyl-Sepharose column was diluted with 50:50 buffer A to get 25% ethylene glycol solution containing purified AQUI. The solution was then split into 1 ml aliquots and quick-frozen in liquid nitrogen and stored at -20°C for later use.

Table 3. Buffers used in the purification of AQUI.

Buffer A	**Buffer B**	**Buffer C**	**Buffer D**
25 mM Tris	25 mM Tris	25 mM MES	25 mM MES
10 mM $CaCl_2$	10 mM $CaCl_2$	1 mM $CaCl_2$	1 mM $CaCl_2$
pH 8.0	pH 8.0	pH 6.0	pH 6.0
	1 M $(NH_4)_2SO_4$		0.5 M NaCl

2.1.1 Purification table and Bradford protein-assay

Aliquots (0.5 ml) were collected at each purification step for measuring protein concentration and enzyme activity. Protein concentration measurements were conducted with the Bradford protein assay using protein assay according to Zaman-Verwilghen [70]. In a standard assay, 2.75 ml of Coomassie Brilliant Blue G250 solution was mixed with 0.25 ml of protein-solution, or 0.25 ml of MilliQ water for blank. The absorbance of the solution was then measured at 620 nm and protein concentration estimated after a 15 minute incubation period.

Enzyme activity was measured using 0.5 mM sAAPF-pNA at 40°C (chapter 2.3). The specific activity along with the protein-assay was then used to calculate yield and purification for the protease.

2.1.2 SDS polyacrylamide gel electrophoresis analysis

In order to estimate the degree of purification of AQUI in the purification scheme, SDS PAGE electrophoresis was carried out of the purified protein fractions. The discontinuous electrophoresis was conducted using 12.5% polyacrylamide vertical slab gels according to Laemmli [71]. Before electrophoresis, the enzyme solutions were inhibited with 25 mM phenylmethylsulfonylfluoride (PMSF) solution to a final concentration of 5 mM PMSF concentration to prevent auto-proteolytic cleavage. The solution was boiled in a water-bath for five minutes.

After electrophoresis the gel was fixed for 30 minutes using 50% EtOH:3.5% H_3PO_4:46.5% MilliQ water, washed for 20 minutes twice with MilliQ water and then stained overnight

using Blue-Silver Coomassie stain. The gel was rinsed with water several times until unbound color dissipated.

The Blue-Silver Coomassie solution was made by mixing enough phosphoric acid to 100 ml of water to get a 10% concentration of phosphoric acid in one liter of the final solution. Hundred grams of ammonium sulfate were mixed in with the phosphoric acid solution along with 1.5 g of powdered G250 Coomassie Blue. The solution was then filled up to 800 ml using MilliQ water and finally 200 ml of methanol were added [72].

2.2 Enzyme assay

The specific activity of the protease was measured with succinyl-AlaAlaProPhe-p-nitro-anilide (sAAPF-pNA) (Bachem) as a substrate at concentrations between 0.1 – 2 mM. The activity of the protease was measured at 40°C in 100 mM Tris, 10 mM $CaCl_2$ at pH 8.6 (pH 9.06 at 25°C). The diluted substrate solutions were made from a stock solution of 25 mM sAAPF-pNA dissolved in dimethyl sulfoxide (DMSO) which was stored at 4°C. The protease cleaves the amide bond of sAAPF-pNA that results in free p-nitroaniline [73]. The molar absorbance coefficient for p-nitroaniline (ε) is 8480 M^{-1} cm^{-1} at 410 nm. The enzyme activity was measured by observing the change of absorbance at 410 nm over 30 seconds which correlates to change in free p-nitroaniline. The rate of the enzyme reaction was then calculated by using the molar absorbance and absorbance at 410 nm with Beer's law:

$$\text{Rate of reaction} = \frac{\Delta Absorbance\ at\ 410\ nm/sec}{8480\ M^{-1}\ cm^{-1} \times 1\ cm} \times 1000 \frac{mM}{M} = mM/sec \qquad \text{(IV)}$$

2.3 Thermal inactivation of AQUI ($T_{50\%}$)

The thermal inactivation of proteins is most often caused by denaturation that affects the active site of the protein. This inactivation with temperature can be used as an evaluation of thermal stability and denaturation of proteins.

Protein sample containing ~1 U/ml was thawed and dialyzed overnight at 4°C in 25 mM Tris, 1 mM $CaCl_2$, 100 mM NaCl, and pH 8.95. Approximately 350 µl aliquots of the enzyme solution were measured into test tubes and then incubated at different temperatures. The thermal inactivation was measured at six selected temperatures at regular time intervals. These temperatures were 87°C, 89°C, 91°C, 93°C, 95°C and 97°C. The temperature was maintained constant by the use of a water bath. The exact temperature of the heat bath was measured by the use of digital thermometer (HI935005 thermocouple thermometer from

Hanna). The loss of enzyme activity was then measured at regular time intervals with 0.5 mM sAAPF-nPA at 40°C in 100 mM Tris, 10 mM $CaCl_2$ at pH 8.6 (410 nm).

The thermal inactivation of AQUI follows a first order reaction with respect to time according to the following equation [74]:

$$[E] = [E]_0 \times e^{-kt} \qquad \text{(V)}$$

Where [E] is the concentration and t is time and k is the first order rate constant which is different for each temperature. Equation V can be rearranged to equation VI where relative reduced enzyme activity [E] with time yields a slope which gives negative k constant.

$$\ln \frac{[E]}{[E]_0} = -kt \qquad \text{(VI)}$$

The rate constants (k) determined at each temperature (T) can then be plotted into a linear derivation of Arrhenius equation:

$$\ln k = \ln A - \frac{Ea}{R} \times 1/T \qquad \text{(VII)}$$

Where A is the pre-exponential constant and R is the universal gas constant (8.314 J/mol*K). Activation energy (Ea) can be obtained from the slope of the linear plots of lnk and the reciprocal of the absolute temperature (T). The temperature at which the enzyme loses 50% of its activity in 30 min, $T_{50\%}$, can then be obtained from the half-life equation for first order reaction (eq. v) by rearranging the first order rate equation for $T_{50\%}$:

$$k\ (T_{50\%})\ s^{-1} = \frac{\ln 2}{30\ \text{min} \times 60\ sek/min} \qquad \text{(VIII)}$$

The temperature associated with the k ($T_{50\%}$) can then be calculated from the Arrhenius plot. At least three $T_{50\%}$ measurements were taken for the variant along with the wild type and the average along with standard deviation were calculated.

2.4 Melting point (T_m) determination with CD

Denaturation of AQUI with temperature can be measured by monitoring changes in circular-dichroism of the proteins in the far UV region. Circular dichroism spectrometers use left- and right-polarized light and measure the difference of absorbance between the two (mdeg). This is especially handy with proteins since they are chiral molecules, denaturing proteins with heat reduces this chiral property [75].

Protein samples with absorbance of ~0.3 at 280 nm, were inhibited with 5 mM phenylmethanesulfonyl fluoride (PMSF) for ~5 minutes and complete inhibition was

confirmed by assaying the enzyme using sAAPF-pNA. The inhibited enzyme solution was dialyzed against 25 mM glycine, containing 1 mM EDTA, pH 8.6 at 4°C.

Absorbance (mdeg) was monitored at 222 nm using a JASCO J-810 CD spectropolarimeter, this wavelength has the greatest difference in absorbance between native and denaturated state. Temperature gradient 1°C/min from 25°C to 90°C was used and measurements were carried out in a 0.2 cm cuvette. The temperature gradient was kept with a PTC-423S Peltier type single cell holder and temperature control system. Assuming that the protein exists in either of two state, unfolded or folded, the following applies:

$$Total\ protein = 1 = f_u + f_f \qquad \text{(IX)}$$

Where f_u is fraction unfolded and f_f is the fraction folded. The absorption (mdeg) from the solution against temperature resulting from denaturation of the protein can be written in the following equation:

$$y = f_u y_u + f_f y_f \qquad \text{(X)}$$

The raw data from the spectrometer has to be normalized because the absorbance changes with temperature even if no conformational changes between the two states occur. Linear extrapolation against temperature is made from both the native (y_u) and denaturated (y_f) state of the protein where y is the measured absorbance:

$$f_u = \frac{y_f - y}{y_f - y_u} \qquad \text{(XI)}$$

Equation XI gives the fraction of the protein that is unfolded at any temperature measurement on a normalized scale between 0-1. The data was then fitted to a sigmoidal curve using KaleidaGraph. The melting point of the protein was then calculated from the sigmoidal fit ($f_u = 0.5$).

2.5 Enzyme kinetics

Protein samples were dialyzed overnight at 4°C in assay buffer (100 mM Tris, 10 mM $CaCl_2$, and pH 8.6 at 40°C). Protein concentration for the samples were determined by using absorbance measurements at 280 nm. The molar absorbance coefficient for AQUI (ε: 34,630 M^{-1} cm^{-1}) was obtained from Expasy Bioinformatics Resource Portal using ProtParam (http://web.expasy.org/protparam/). The samples were then diluted to a final concentration of around ~1 U/ml where the dilution was measured exactly with a scale (Mettler AE100). Seven different concentrations of substrate were prepared ranging from 0.1 mM to 2 mM sAAPF-pNA by dilution in the assay buffer. The rate of reaction was measured using HeΛios α spectrophotometer from Thermo Spectronics. The temperature was kept steady at 40°C by

using Peltier heating device from Thermo Spectronics. The measurements were made over 30 seconds by using 950 µl of substrate mixture against 50 µl of enzyme, the final solution was collected into test tubes then incubated for 2 hours at 40°C. The rate of reaction was calculated from equation IV. For more accurate measurements of substrate concentrations, the substrate and enzyme reaction mixtures were allowed to stand for the reaction to go to completion. Absorbance was then measured at 410 nm using assay buffer as a reference and using the molar absorbance coefficient of 8480 M^{-1} cm^{-1} for sAAPF-pNA (eq. XII). The sAAPF–pNA solution was diluted appropriately so the absorbance of the solution would not exceed a value of 1 (10 to 20-fold dilution).

$$[sAAPF - pNA]\,(mM) = \frac{A\ at\ 410\ nm}{1\ cm \times 8480\ M^{-1}\ cm^{-1}} \times 1000\ mM/M \qquad \text{(XII)}$$

The data for rate of reaction and substrate concentration were then fitted to Lineweaver-Burk and Michaelis-Menten equation using the software KaleidaGraph. Equation XIII is Michaelis-Menten and Lineweaver-Burk is the inverse of that in equation XIV.

$$V = \frac{V_{max} \times [S]}{K_M + [S]} \qquad \text{(XIII)}$$

$$\frac{1}{V} = \frac{K_M}{V_{max} \times [S]} + \frac{1}{V_{max}} \qquad \text{(XIV)}$$

The maximum velocity (V_{max}) and Michaelis constant (K_M) were calculated from both methods where V is velocity of the reaction at that substrate concentration and K_M is substrate concentration needed to fulfil half V_{max}. The turnover number (k_{cat}) can be calculated from V_{max} where k_{cat} is a unit that reflects how many moles of substrate one mole of enzyme catalyzes in a second:

$$k_{cat} = \frac{V_{max}}{[E] * 60\ s/mín} \qquad \text{(XV)}$$

Where [E] is the enzyme concentration and V_{max} is the maximum velocity observed using [E]. The catalytic efficiency for the protein can be calculated by dividing k_{cat} with K_M which will be shown in M^{-1} s^{-1} units dependent on time and substrate concentration.

2.6 Structure and classification of proteins

Three-dimensional protein structures in this project were visualized with the software UCSF Chimera (http://www.cgl.ucsf.edu/chimera/) [76]. The structures for Chimera were obtained from the protein databank (http://www.rcsb.org/) where AQUI is coded 4DZT [9] and VPR 1SH7 [65]. MEROPS the Peptide Database (https://merops.sanger.ac.uk/) [45] was used for the classification of proteases and CATH (http://www.cathdb.info/) [54] was used for protein structure classification.

3 Results

3.1 Purification of AQUI and its variants

$AQUI_{wt}$ and the variant AQUI N68Q were purified according to chapter 2.1. The purification began by denaturating most of the *E. coli* proteins and rupturing the cell membranes by incubating the cell lysate at 70°C for 60 minutes. The cell lysate was then spun and the insoluble impurities were discarded, the remaining solution will be referred as cell lysate spin. The cell lysate was made 1 M $(NH_4)_2SO_4$, spun and applied to the first hydrophobic phenyl-Sepharose column (*Figure 10*). Other proteins were washed off with and without $(NH_4)_2SO_4$ and appeared around 200 ml and 500 ml. The AQUI eluate was washed off with 50% ethylene glycol in buffer A and appeared around 650 ml.

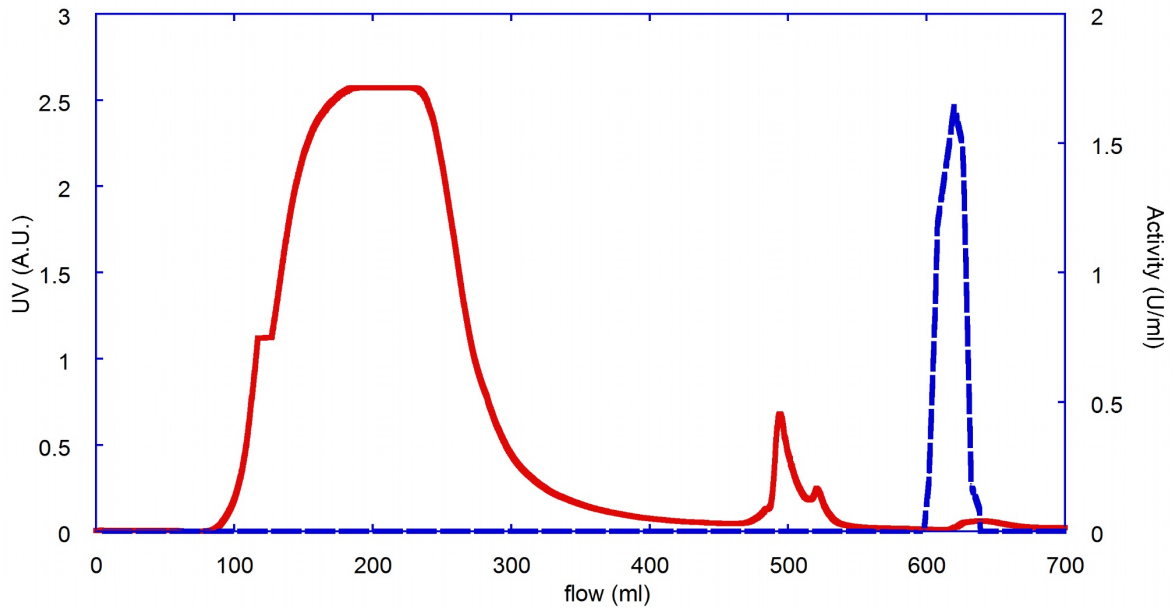

Figure 10. Example of a hydrophobic phenyl-Sepharose column purification for AQUI. The red solid line is absorbance at 280 nm and the blue dotted line is activity assayed with 0.5 mM sAAPF-pNA at 40°C in 100 mM Tris, 10 mM $CaCl_2$, and pH 8.6. Majority of impurities, containing no AQUI activity, were washed off at ~200 ml and ~500 ml by going from 1 M $(NH_4)_2SO_4$ to none incrementally in buffer A. The eluate containing AQUI activity was washed off with 50% ethylene glycol and appeared at ~620 ml.

The eluate from the first phenyl-Sepharose column was checked for activity. The largest peak observed at ~200 ml and the smaller peaks apparent at 500 ml had negligible activity

(Figure 10). The eluate was collected in fractions beginning at 600 ml. The fractions were checked for activity and fractions containing acceptable activity (>0.1 U/ml) were pooled.

The pooled solution containing AQUI activity from the phenyl-Sepharose column was prepared for the next column by making them to 25 mM MES, 1 mM $CaCl_2$, pH 6 (buffer C) via dialyzation. The buffer C dialysate containing AQUI activity was applied to a CM-Sepharose ion-exchange column and impurities were washed off with buffer C without salt (*Figure 11*). AQUI was eluted from the column, by a linear gradient of increasing salt concentration to 0.5 M NaCl in buffer C. The eluate was collected in fractions after the salt concentration gradient was applied and the fractions containing enzyme activity were eluted from the column at about 250 ml (*Figure 11*). The enzyme activity of the fractions was estimated and fractions containing acceptable activity (>0.1 U/ml) were pooled. The pooled solution was then applied to another phenyl-Sepharose column (*Figure 12*). The second phenyl-Sepharose column has the purpose to (reduce the volume of solution containing) concentrate fractions containing AQUI activity. The eluate fractions containing acceptable activity (>0.5 U/ml) AQUI activity from the second phenyl-Sepharose column were pooled and made 25% ethylene glycol with buffer A and then quick-frozen in 1 ml aliquots for later experiments.

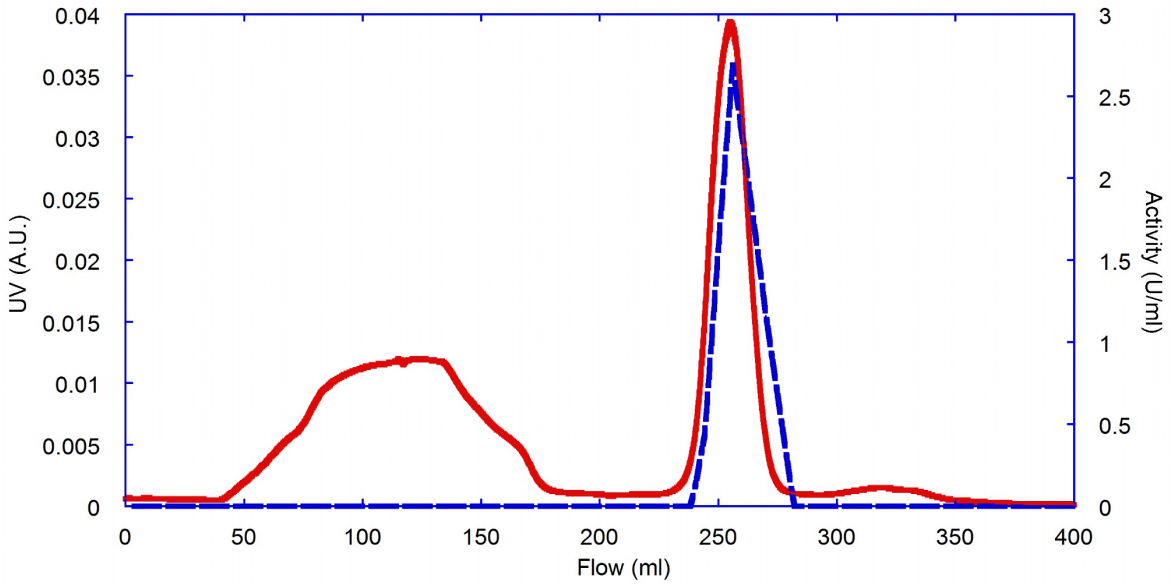

Figure 11. Example of a cationic CM-Sepharose column purification for AQUI. The red solid line is absorbance at 280 nm and the blue dotted line is activity assayed with 0.5 mM sAAPF-pNA at 40°C in 100 mM Tris, 10 mM $CaCl_2$, and pH 8.6. Impurities containing no AQUI activity were washed off around 100 ml with no salt. AQUI was washed off with a salt gradient, which started at 200 ml, going from 0 to 0.5 M NaCl in buffer C. The eluate containing AQUI activity appeared around 260 ml.

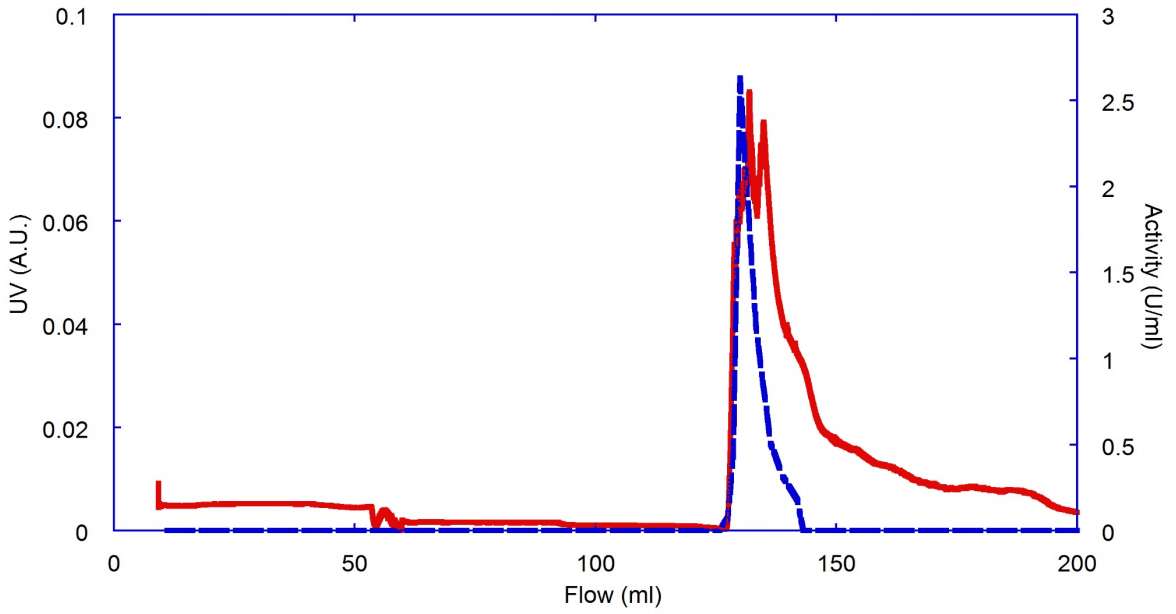

Figure 12. Example of the (latter) second phenyl-Sepharose column used to concentrate (reduce the total volume of) the AQUI containing solution. The red solid line is absorbance at 280 nm and the blue dotted line is activity assayed with 0.5 mM sAAPF-pNA at 40°C in 100 mM Tris, 10 mM $CaCl_2$, and pH 8.6. Going from 1 M $(NH_4)_2SO_4$ to zero stepwise in buffer A where the interchange can be seen around 50 ml. The eluate containing AQUI activity appeared around 140 ml and was washed off with 50% ethylene glycol in buffer A.

Purification was successful and purification tables were made for both $AQUI_{wt}$ (*Table 4*) and AQUI N68Q (*Table 5*). The observed specific activity for the variant N68Q was halved compared to $AQUI_{wt}$.

Table 4. Purification table for $AQUI_{wt}$. The activity was measured after each step with 0. mM sAAPF-pNA at 40°C in 100 mM Tris, 10 mM $CaCl_2$, and pH 8.6. Protein concentrations were estimated with a Bradford-assay according to chapter 2.1.1. PS is an abbreviation for phenyl-Sepharose.

Step	Volume (ml)	Activity (U/ml)	Total activity (U)	Protein conc. (mg/ml)	Total protein (mg)	Specific activity (U/mg)	Yield (%)	Purification (x)
Cell lysate spin	88	6.3	555	0.46	40.3	13.7	100	1.0
PS-column (1)	151	3.3	500	0.13	19.0	26.3	90	1.9
CM-sepharose column	31	7.3	226	0.21	6.6	34.2	41	2.5
PS-column (2)	28	8.2	230	0.21	5.8	39.5	41	2.9

Table 5. Purification table for AQUI N68Q. The activity was measured after each step with 0.5 mM sAAPF-pNA at 40°C in 100 mM Tris, 10 mM $CaCl_2$, and pH 8.6. Protein concentrations were estimated with a Bradford-assay according to chapter 2.1.1. PS is an abbreviation for phenyl-Sepharose.

Step	volume (ml)	Activity (U/ml)	Total activity (U)	Protein conc. (mg/ml)	Total protein (mg)	Specific activity (U/mg)	Yield (%)	Purification (x)
Cell lysate spin	98	1.0	94	0.37	36.5	2.6	100	1.0
PS-column (1)	190	0.4	83	0.06	11.0	7.5	88	2.9
CM-sepharose column	39	1.7	65	0.13	5.1	12.7	69	4.9
PS-column (2)	38	1.1	43	0.06	2.3	18.9	45	7.3

3.1.1 SDS-page

SDS-page was carried out for the final purified products in order to estimate the degree of purity for the AQUI purifications (*Figure 13*). The SDS page showed one intense protein band in all cases. The purification of $AQUI_{wt}$ had faint protein bands that were smaller than the intense AQUI band which suggests that a portion of the sample might have undergone autoproteolytic cleavage during SDS preparation. The second unspecified $AQUI_{wt}$ (U) in the SDS page was from a purification procedure which was carried out in this project but without establishing a purification table. The AQUI protein bands appeared just above 25 kDa band in the protein ladder. Which is consistent with the published AQUI mass of 28.5 kDa [61].

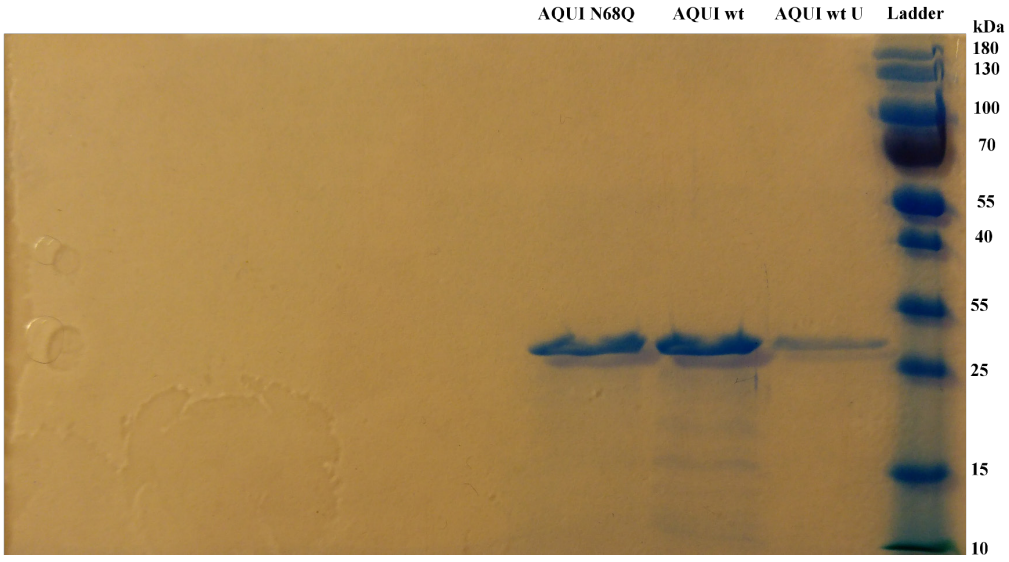

Figure 13. SDS-PAGE of purified AQUI samples collected from the purification procedures that were carried out. From right: PAGE ruler™ prestained proteinladder (26616), unspecified $AQUI_{wt}$ (U), $AQUI_{wt}$ as in Table 4, and AQUI N68Q as in Table 5.

3.2 Michaelis-Menten kinetics

Michael-Menten kinetic experiments were carried out on $AQUI_{wt}$ and AQUI N68Q to estimate the changes in the variant. Reaction rates were observed at 40°C with selected substrate concentrations from 0.1 mM to 2 mM sAAPF-pnA in 100 mM Tris, 10 mM $CaCl_2$ and pH 8.6. Resulting data was fitted to nonlinear Michaelis-Menten equation and the Michaelis-Menten parameters were estimated (*Figure 14*).

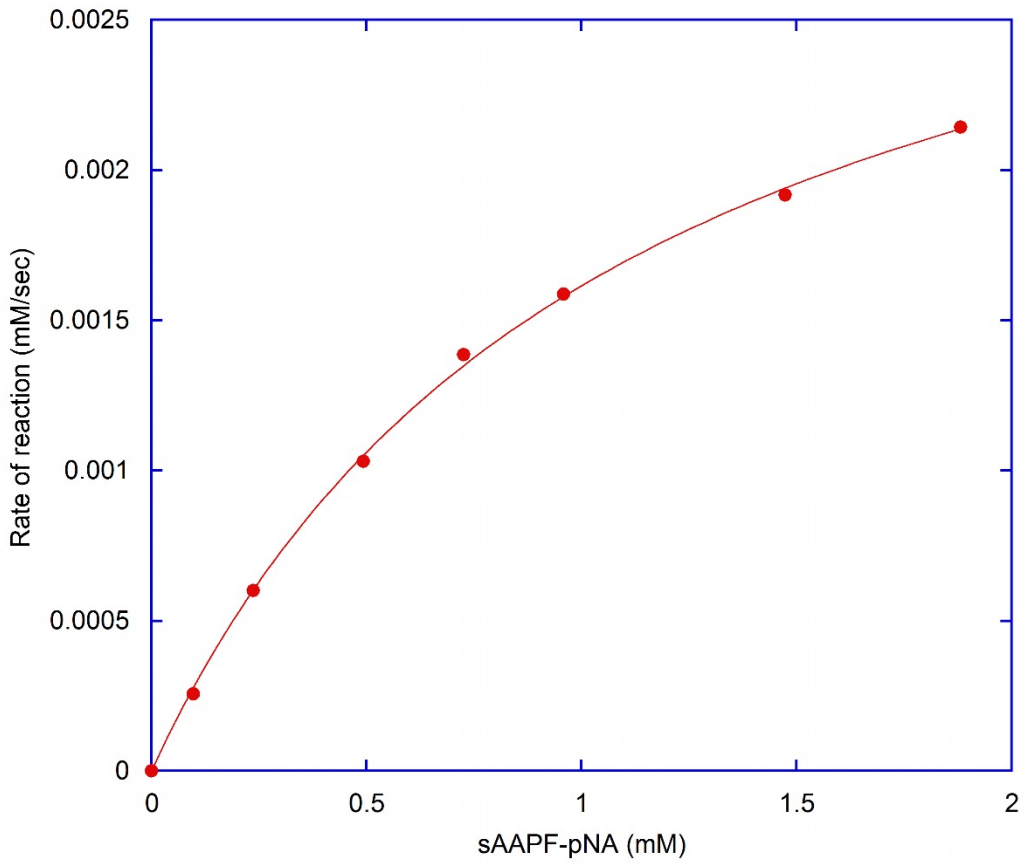

Figure 14. Example of an experimental data fitted to a Michaelis-Menten nonlinear equation. The red dots are measured reaction rates from an $AQUI_{wt}$ experiment and the red line is the fitted Michaelis-Menten equation. The parameters in this case were observed to be K_M: 1.09 mM, V_{max}: 0.00337 mM/sec, k_{cat}: 32.8 sec^{-1}, and k_{cat}/K_M: 30.1 $s^{-1}mM^{-1}$.

Michaelis-Menten parameters for both the $AQUI_{wt}$ and N68Q were measured from at least three experiments each and the average calculated. The standard deviation of the mean was calculated and shown with the average in *Table 6*.

Table 6. Kinetic parameters of $AQUI_{wt}$ and AQUI N68Q derived from Michaelis-Menten equation. Expressed as mean values ± standard deviation of the mean. The activity was measured with sAAPF-pNA as a substrate at concentrations from 0.1 mM to 2 mM. The assay was made at 40°C in 100 mM Tris, 10 mM $CaCl_2$, and pH 8.6.

Mutation	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M ($s^{-1}mM^{-1}$)
$AQUI_{wt}$	33.3 ± 4.6	1.0 ± 0.05	32.4 ± 3.9
AQUI N68Q	23.6 ± 5.7	0.81 ± 0.05	29.2 ± 7.7

The variant had a significantly lower K_M and k_{cat} according to a two tailed Student's T-test with a significance level of 0.05 and 14 degrees of freedom (kinetic parameters were measured 8 times for both variants) but the k_{cat}/K_M did not show significant changes. Measured data were also fitted to the linear Lineweaver-Burk equation for comparison (*Figure 15*).

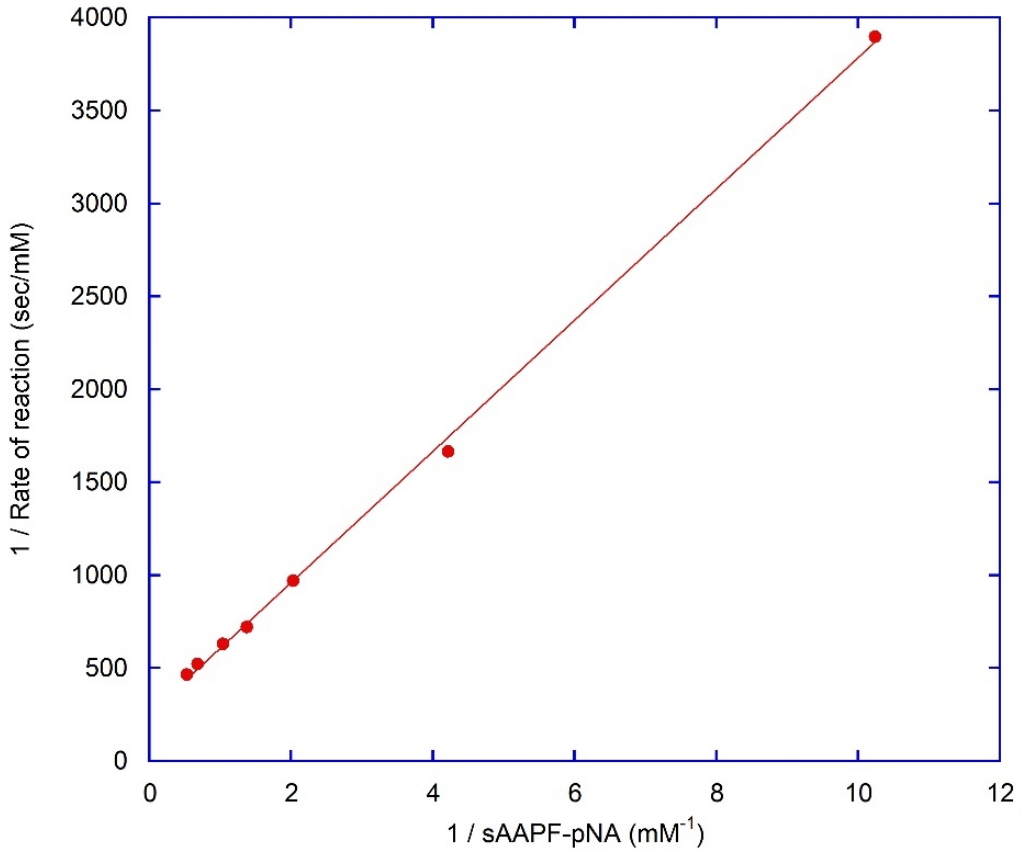

Figure 15. Example of an experimental data fitted to Lineweaver-Burk linear equation. The red dots are measured reaction rates from an $AQUI_{wt}$ experiment (same as figure 14) and the red line is the fitted Lineweaver-Burk equation. The parameters in this case were observed to be K_M: 1.39 mM, k_{cat}: 34.4 sec^{-1}, and k_{cat}/K_M: 24.7 $s^{-1}mM^{-1}$.

The kinetic parameters for AQUI calculated using the Lineweaver-Burk equation are shown in *Table 7*. The K_M for $AQUI_{wt}$ and AQUI N68Q was significantly higher when using the Lineweaver-Burk method (*Table 7*) compared to the Michaelis-Menten method (*Table 6*). The k_{cat} was also higher for $AQUI_{wt}$ compared to AQUI N68Q when using the Lineweaver-Burk method, resulting in a significant difference in k_{cat} between the N68Q variant and $AQUI_{wt}$. However, the catalytic efficiency was very similar for both methods in both $AQUI_{wt}$ and AQUI N68Q. The Michaelis-Menten method was deemed more reliable because all measured data weighs the same when determining the kinetic parameters and thus is more widely used, so the results for discussion were rather derived from it. The Lineweaver-Burk method weighs the measured variables in an inordinately fashion where the rate of reaction at lower substrate concentrations affects the determined kinetic parameters more than the rate of reaction at higher substrate concentrations. Similar effect on kinetic parameters determination is also observed in other linear methods such as the Eadie-Hofstee plot. Of all the linear transformations used to estimate kinetic parameters, the Lineweaver-Burk method is considered to be the least reliable [77].

Table 7. Kinetic parameters of $AQUI_{wt}$ and AQUI N68Q derived from Lineweaver-Burk equation. Expressed as mean values ± standard deviation of the mean. The activity was measured with sAAPF-pNA as a substrate at concentrations from 0.1 mM to 2 mM. The assay was made at 40°C in 100 mM Tris, 10 mM $CaCl_2$, and pH 8.6.

Mutation	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M ($s^{-1}mM^{-1}$)
$AQUI_{wt}$	38.8 ± 5.3	1.15 ± 0.04	33.7 ± 5.2
AQUI N68Q	24.5 ± 5.7	0.90 ± 0.06	27.2 ± 5.7

3.3 Thermostability of AQUI

3.3.1 Thermal inactivation of AQUI ($T_{50\%}$)

Thermal inactivation of AQUI was observed by incubating samples at six selected temperatures and measuring the relative loss of activity at regular time intervals (*Figure 16*). The derived first order rate constants at each temperature were then fitted to an Arrhenius equation (*Figure 17*). The Arrhenius equation was then used to find a reaction rate constant which corresponded to halved activity in half an hour and the activation energy of the Arrhenius plot. The final results were compiled together with the melting point of AQUI (*Table 8*).

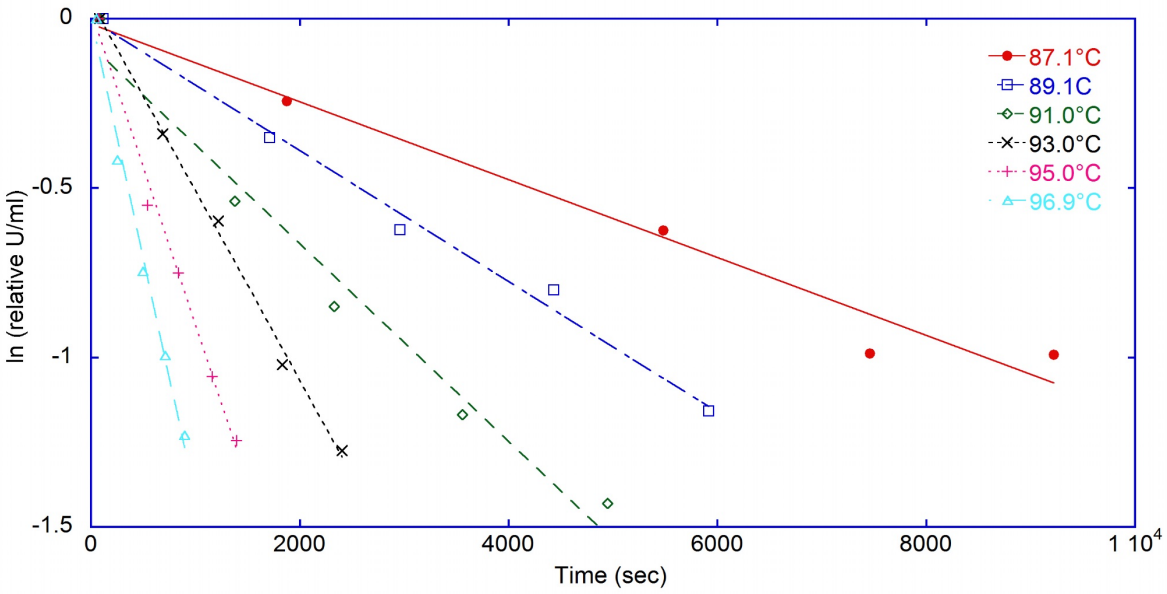

Figure 16. An example of a thermal inactivation experiment conducted on $AQUI_{wt}$. AQUI was incubated at selected temperatures in 25 mM Tris, 100 mM $NaCl_2$, 1 mM $CaCl_2$, and pH 8.95. Activity was measured at regular intervals with 0.5 mM sAAPF-pNA at 40°C in 100 mM Tris, 10 mM $CaCl_2$, and 8.6 pH. The resulting slope is the negative reaction rate constant (-k) and is used for the Arrhenius equation plot (Figure 17).

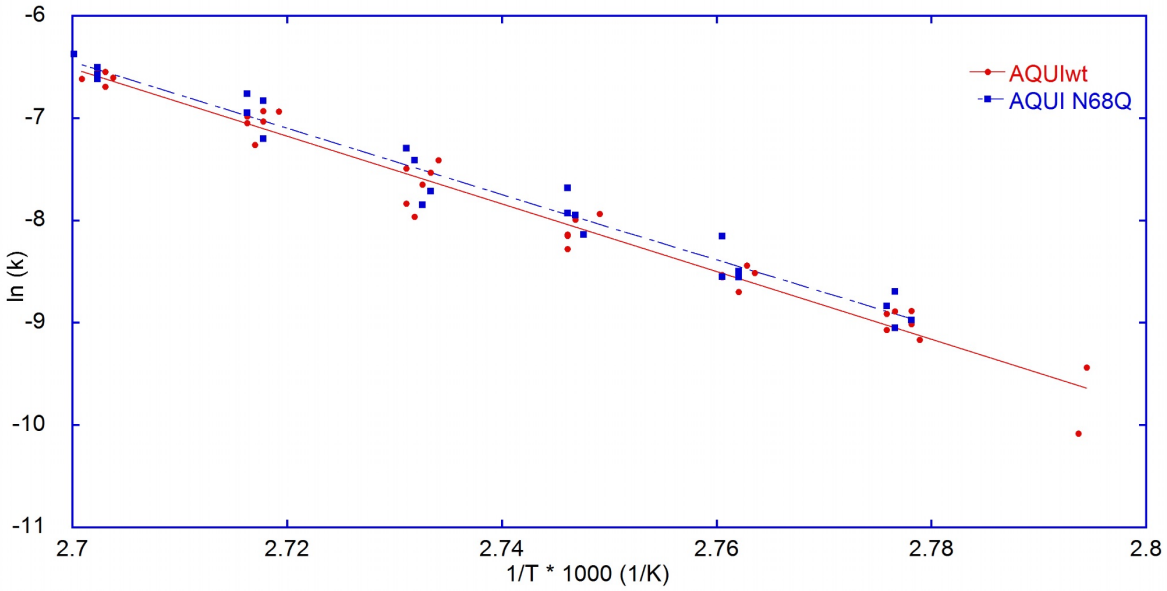

Figure 17. Comparison of Arrhenius plot between $AQUI_{wt}$ and AQUI N68Q. The data shown is thermal inactivation rate (k) at selected temperatures in 25 mM Tris, 1mM $CaCl_2$, and pH 8.95. The activity was measured at 40°C with sAAAPF-pNA as substrate in 100 mM Tris, 10 mM $CaCl_2$, and pH 8.6.

3.3.2 Melting point (T_m)

The melting point (T_m) of $AQUI_{wt}$ and the variant AQUI N68Q were determined in the absence of calcium using a CD-spectrometer according to chapter 2.4. The raw data was normalized so that the vertical axis showed the denaturated fraction (*Figure 18)*. A sigmoidal curve was fitted to the measured data and the melting point was determined from the curve when half of the protein is in a denaturated form (i.e. Fu = 0.5). The results obtained are shown with the $T_{50\%}$ results in *Table 8*.

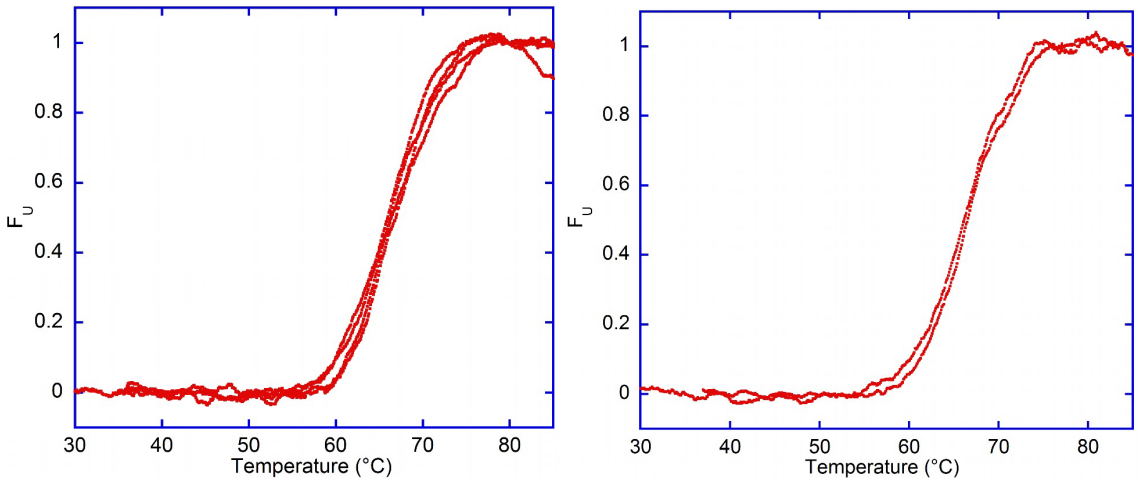

Figure 18. Normalized melting curves for AQUI N68Q on the left and $AQUI_{wt}$ on the right. The enzymes were inhibited with PMSF and the melting curve was measured at 222 nm with a temperature gradient of 1°C/min. The melting procedure was carried out in 25 mM glycine, 1 mM EDTA (without Ca^{+2}), and pH 8.6.

Table 8. The thermal stability parameters of $AQUI_{wt}$ and AQUI N68Q.

Variant	$T_{50\%}$ (°C)	Ea (kJ/mol)	T_m (°C)
$AQUI_{wt}$	91.7 ± 0.3	276 ± 27	66.4 ± 0.4
AQUI N68Q	91.3 ± 0.6	267 ± 9	66.4 ± 0.3

The results showed that there were no significant changes observed in the stability of the variant AQUI N68Q according to the results in activation energy of denaturation (Ea), the melting point (T_m), and the $T_{50\%}$ (*Table 8).* This was expected, since the N68Q change was not expected to affect the stability of the protein.

4 Conclusions

The aim of this project was to increase the catalytic activity of AQUI while maintaining the thermostability of $AQUI_{wt}$. This was attempted by introducing the mutation N68Q based on the premise that an earlier D98S mutation doubled the turnover number and tripled the catalytic activity. It was suggested that the D98S mutation led to a reduced number of H-bonds near the active site and computer simulations suggested that the mutation could enforce one rotomer over other rotomers in the His70 residue [68]. Based on computer simulations, it was hypothesized that due to the larger side chain resulting from the N68Q mutation that the imidazole side group of His70 residue in the catalytic triad would be shifted towards a more active rotomer, as found in the D98S variant. Thus the objective was to test this hypothesis and if successful it should have made the enzyme more active while maintaining the thermostability of $AQUI_{wt}$ (*Figure 9*). The observed thermostability was not altered in the N68Q variant so in that respect it was successful. This result was anticipated since the mutation was only supposed to affect the structure in a localized area near the His70 residue but not the whole structure. However, the main goal of increased catalytic activity was not observed in the variant. The only significant change was the drop of K_M in the AQUI N68Q compared to the wild type. The k_{cat} dropped as well but the variant and the wild type were within one standard deviation of the each other which resulted in very similar k_{cat}/K_M for both cases. *Table 9* summarizes the main results of this project.

Table 9. The main results of the project. Kinetic parameters were derived from a nonlinear Michaelis-Menten plot and two indicators of thermostability were observed.

Variant	$T_{50\%}$ (°C)	T_m (°C)	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M ($s^{-1}mM^{-1}$)
$AQUI_{wt}$	91.7 ± 0.3	66.43 ± 0.38	33.3 ± 4.6	1.15 ± 0.04	33.7 ± 5.2
AQUI N68Q	91.3 ± 0.6	66.38 ± 0.32	23.6 ± 5.7	0.90 ± 0.06	27.2 ± 5.7

What is perhaps most peculiar is the fact that the catalytic activity of $AQUI_{wt}$ observed in this project was not comparable with the $AQUI_{wt}$ previously observed in the same [64, 66, 69]. *Table 10* shows the catalytic activity results for AQUI observed in this project compared to published results from other studies. The published results in catalytic activity for $AQUI_{wt}$ seem to vary significantly between different labs. Results from other research groups have measured the AQUI activity under slightly different circumstances but they were still comparable to $AQUI_{wt}$ k_{cat} results from the Kristjánsson's lab but had a much lower K_M [78, 79]. One older study observed similar catalytic activity in $AQUI_{wt}$ as in this project [80].

Table 10. Different results of $AQUI_{wt}$ from several published studies. The most significant changes in method is the use of 100 mM HEPES, 1 mM $CaCl_2$, pH 7.5 in assays for Sakaguchi et al. and Tanaka et al. [78, 80]. Magnús Már Kristjánsson's lab used 100 mM Tris, 10 mM $CaCl_2$, pH 8.6. All projects used sAAPF-pNA as a substrate and did their assays at 40°C [64, 66, 69].

Origin of $AQUI_{wt}$ results	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M ($s^{-1}mM^{-1}$)
This project	33.3 ± 4.6	1.0 ± 0.05	32.4 ± 3.9
Kristjánsson´s group [64, 66, 69]	71.7 ± 5.8	1.35 ± 0.1	48.2 ± 2.0
Sakaguchi et al. (2007) [78]	98.1 ± 2.75	0.69 ± 0.04	141.6 ± 9.3
Sakaguchi et al. (2008) [79]	91.6 ± 2.75	0.79 ± 0.04	116
Tanaka et al. (1998) [80]	33	1.2	27
Lin et al. (1997) [81]	84	1.3	65

The $AQUI_{wt}$ results in this project seems to be most similar those of Tanaka et al. (1998) [80]. Most significant changes between methods used in this project and those of Tanaka and coworkers is use of a different buffer in the assays and the protein was synthesized in *T. aquaticus* in the case of Tanaka et al. compared to *E. coli* in present results. The Lineweaver-Burk derived results (*Table 7*) as compared to the results obtained from nonlinear regression analysis of the Michaelis-Menten plots (*Table* 6), observed in this study were also more similar to results published by Kristjánsson's lab. This observation might be the partial reason for the kinetic parameters deviation from the Kristjánsson's lab published work in this project if the method of kinetic parameters derivation used by Kristjánsson's lab was the linear Lineweaver-Burk method or other methods such as the Eddie-Hofstee method. Comparing kinetic parameters derived from linear-transformation methods such as the Lineweaver-Burk plot or the Eddie-Hofstee plot with kinetic parameters derived from a nonlinear Michaelis-Menten plot might partly explain unexpected deviations because linear-transformation methods weigh the data unevenly.

With the variability in kinetic parameters for the catalytic activity of $AQUI_{wt}$ published works one might think that comparing any AQUI variant to its wild type might involve a lot of inaccuracy. To abstain from most variables other than the mutation being researched, it should be important to use exactly the same method throughout all the experiment to achieve the most accurate results. So the observed kinetic catalytic activity of AQUI N68Q in this project are most accurate when compared to catalytic activity $AQUI_{wt}$ results from this project because they were analyzed in parallel alongside each other. It should prove useful to find the root of the variability in measured results for the catalytic activity of $AQUI_{wt}$ before continuing further in catalytic activity engineering of AQUI. In this project it was attempted to troubleshoot these variables both with sequencing the $AQUI_{wt}$ plasmid used and by probing the catalytic activity of other $AQUI_{wt}$ purifications. The plasmid sequence came back without any new unknown mutations and other $AQUI_{wt}$ purifications varied in

catalytic activity and at a quick glance were not easily comparable with earlier results by Kristjánsson's team (*Table 10*).

On the topic of AQUI engineering and the N68Q variant compared to the D98S variant. The N68Q mutation lengthened the side-chain of an asparagine by one carbon to glutamine in the corresponding variant and in theory positively influenced the active rotomer of His70 as suggested by MD simulations due to steric hindrance. For greater insight on the role of the His70 rotomers in catalysis it might prove informative to mutate the asparagine into a smaller side-chain such as seine for an example to allow for greater movement of the His70 residue. If catalytic activity of AQUI is reduced from the N68S mutation and later successful N68Q experiments show increased activity compared with $AQUI_{wt}$, then the His70 active rotomer hypothesis is verified.

5 References

1. Rothschild, L. J. & Mancinelli, R. L. (2001) Life in extreme environments, *Nature.* **409**, 1092-1101.

2. Demirjian, D. C., Morís-Varas, F. & Cassidy, C. S. (2001) Enzymes from extremophiles, *Curr. Opin. Chem. Biol.* **5**, 144-151.

3. Bakermans, C., Tsapin, A. I., Souza-Egipsy, V., Gilichinsky, D. A. & Nealson, K. H. (2003) Reproduction and metabolism at − 10°C of bacteria isolated from Siberian permafrost, *Environ. Microbiol.* **5**, 321-326.

4. Kashefi, K. & Lovley, D. R. (2003) Extending the upper temperature limit for life, *Science (New York, N.Y.).* **301**, 934.

5. Berezovsky, I. N. & Shakhnovich, E. I. (2005) Physics and evolution of thermophilic adaptation, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 12742-12747.

6. D'Amico, S., Collins, T., Marx, J.-C., Feller, G. & Gerday, C. (2006) Psychrophilic microorganisms: challenges for life, *EMBO Rep.* **7**, 385-389.

7. Brock, T. D. & Freeze, H. (1969) *Thermus aquaticus* gen. n. and sp. n., a Nonsporulating Extreme Thermophile, *J. Bacteriol.* **98**, 289-297.

8. Holland, P. M., Abramson, R. D., Watson, R. & Gelfand, D. H. (1991) Detection of specific polymerase chain reaction product by utilizing the 5' -> 3' exonuclease activity of *Thermus aquaticus* DNA polymerase, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 7276-7280.

9. *Aqualysin I: The crystal structure of a serine protease from an extreme thermophile,* Thermus aquaticus *YT-1*, http://www.rcsb.org/pdb/explore/explore.do?structureId=4DZT.

10. Kristjánsson, M. M., Magnússon, Ó. T., Gudmundsson, H. M., Alfredsson, G. Á. & Matsuzawa, H. (1999) Properties of a subtilisin-like proteinase from a psychrotrophic *Vibrio* species, *Eur. J. Biochem.* **260**, 752-760.

11. Nelson, D. L., Cox, M. M. & Lehninger, A. L. (2013) Lehninger Principles of Biochemistry 6th edn., pp. 189-241; 501-542, W.H. Freeman and Company, New York.

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

13. Kumar, S. & Nussinov, R. (2001) How do thermophilic proteins deal with heat?, *Cell. Mol. Life Sci.* **58**, 1216-1233.

14. Luke, K. A., Higgins, C. L. & Wittung-Stafshede, P. (2007) Thermodynamic stability and folding of proteins from hyperthermophilic organisms, *FEBS J.* **274**, 4023-4033.

15. Daniel, R. M., Danson, M. J., Hough, D. W., Lee, C. K., Peterson, M. E. & Cowan, D. A. (2008) Enzyme stability and activity at high temperatures, *Protein Adaptation in Extremophiles*, 1-34.

16. Lins, L. & Brasseur, R. (1995) The hydrophobic effect in protein folding, *FASEB J.* **9**, 535-540.

17. Nicholls, A., Sharp, K. A. & Honig, B. (1991) Protein folding and association: Insights from the interfacial and thermodynamic properties of hydrocarbons, *Proteins: Struct., Funct., Bioinf.* **11**, 281-296.

18. Vieille, C. & Zeikus, G. J. (2001) Hyperthermophilic enzymes: Sources, uses, and molecular mechanisms for thermostability, *Microbiol. Mol. Biol. Rev.* **65**, 1-43.

19. Meruelo, A. D., Han, S. K., Kim, S. & Bowie, J. U. (2012) Structural differences between thermophilic and mesophilic membrane proteins, *Protein Sci.* **21**, 1746-1753.

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

21. Huber, R., Wilharm, T., Huber, D., Trincone, A., Burggraf, S., König, H., Reinhard, R., Rockinger, I., Fricke, H. & Stetter, K. O. (1992) *Aquifex pyrophilus* gen. nov. sp. nov., Represents a Novel Group of Marine Hyperthermophilic Hydrogen-Oxidizing Bacteria, *System. Appl. Microbiol.* **15**, 340-351.

22. Loladze, V. V., Ibarra-Molero, B., Sanchez-Ruiz, J. M. & Makhatadze, G. I. (1999) Engineering a thermostable protein via optimization of charge−charge interactions on the protein surface, *Biochemistry.* **38**, 16419-16423.

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

24. Matsui, I. & Harata, K. (2007) Implication for buried polar contacts and ion pairs in hyperthermostable enzymes, *FEBS J.* **274**, 4012-4022.

25. Alsop, E., Silver, M. & Livesay, D. R. (2003) Optimized electrostatic surfaces parallel increased thermostability: a structural bioinformatic analysis, *Protein Eng.* **16**, 871-874.

26. Xiao, L. & Honig, B. (1999) Electrostatic contributions to the stability of hyperthermophilic proteins1, *J. Mol. Biol.* **289**, 1435-1444.

27. Sterner, R. & Liebl, W. (2001) Thermophilic adaptation of proteins, *Crit. Rev. Biochem. Mol. Biol.* **36**, 39-106.

28. Karshikoff, A. & Ladenstein, R. (2001) Ion pairs and the thermotolerance of proteins from hyperthermophiles: a 'traffic rule' for hot roads, *Trends Biochem. Sci.* **26**, 550-557.

29. Wright, H. T. (1991) Sequence and structure determinants of the nonenzymatic deamidation of asparagine and glutamine residues in proteins, *Protein Eng.* **4**, 283-294.

30. Chakravarty, S. & Varadarajan, R. (2000) Elucidation of determinants of protein stability through genome sequence analysis, *FEBS Lett.* **470**, 65-69.

31. Thompson, M. J. & Eisenberg, D. (1999) Transproteomic evidence of a loop-deletion mechanism for enhancing protein thermostability, *J. Mol. Biol.* **290**, 595-604.

32. Das, R. & Gerstein, M. (2000) The stability of thermophilic proteins: a study based on comprehensive genome comparison, *Functional & Integrative Genomics.* **1**, 76-88.

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

34. Britt, B. M. (1997) For enzymes, bigger is better, *Biophys. Chem.* **69**, 63-70.

35. Srere, P. A. (1984) Why are enzymes so big?, *Trends Biochem. Sci.* **9**, 387-390.

36. Toscano, M. D., Woycechowsky, K. J. & Hilvert, D. (2007) Minimalist active-site redesign: Teaching old enzymes new tricks, *Angewandte Chem. Intl.* **46**, 3212-3236.

37. Mukaiyama, A., Haruki, M., Ota, M., Koga, Y., Takano, K. & Kanaya, S. (2006) A hyperthermophilic protein acquires function at the cost of stability, *Biochemistry.* **45**, 12673-12679.

38. Pace, N. R. (1991) Origin of life-facing up to the physical setting, *Cell.* **65**, 531-533.

39. Tehei, M. & Zaccai, G. (2007) Adaptation to high temperatures through macromolecular dynamics by neutron scattering, *FEBS J.* **274**, 4034-4043.

40. Henzler-Wildman, K. & Kern, D. (2007) Dynamic personalities of proteins, *Nature.* **450**, 964-972.

41. Barrett, A. J., Rawlings, N. D., Salvesen, G. & Fred Woessner, J. (2013) Introduction in *Handbook of Proteolytic Enzymes* pp. li-liv, Academic Press.

42. Nishigori, T., Yanagita, M. & Takeuchi, T. (1996) Proinsulin cleaved by furin is processed to chromatographically mature insulin by carboxypeptidases in nonneuroendocrine cells, *Peptides.* **17**, 789-796.

43. Coughlin, S. R. (2000) Thrombin signalling and protease-activated receptors, *Nature.* **407**, 258-264.

44. Shen, H.-B. & Chou, K.-C. (2009) Identification of proteases and their types, *Anal. Biochem.* **385**, 153-160.

45. Rawlings, N. D., Barrett, A. J. & Finn, R. (2016) Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors, *Nucl. Acids Res.* **44**, D343-D350.

46. Di Cera, E. (2009) Serine Proteases, *IUBMB life.* **61**, 510-515.

47. Page, M. J. & Cera, E. (2008) Serine peptidases: Classification, structure and function, *Cell. Mol. Life Sci.* **65**, 1220-1236.

48. Blow, D. M., Birktoft, J. J. & Hartley, B. S. (1969) Role of a buried acid group in the mechanism of action of chymotrypsin, *Nature.* **221**, 337-340.

49. Neitzel, J. (2010) Enzyme catalysis: the serine proteases, *Nat. Educ.* **3**, 21-25.

50. Berg, J. M., Tymoczko, J. L. & Stryer, L. (2002) Section 10.5, Many enzymes are activated by specific proteolytic cleavage, *Biochemistry 5th edn.,* New York.

51. Hedstrom, L. (2002) Serine Protease Mechanism and Specificity, *Chem. Rev.* **102**, 4501-4524.

52. Earnest, T., Fauman, E., Craik, C. S. & Stroud, R. (1991) 1.59 Å structure of trypsin at 120 K: comparison of low temperature and room temperature structures, *Proteins.* **10**, 171-187.

53. Steinmetz, A. C., Demuth, H. U. & Ringe, D. (1994) Inactivation of subtilisin Carlsberg by N-((tert-butoxycarbonyl)alanylprolylphenylalanyl)-O-benzoylhydroxyl- amine: formation of a covalent enzyme-inhibitor linkage in the form of a carbamate derivative, *Biochemistry.* **33**, 10535-10544.

54. Sillitoe, I., Lewis, T. E., Cuff, A., Das, S., Ashford, P., Dawson, N. L., Furnham, N., Laskowski, R. A., Lee, D., Lees, J. G., Lehtinen, S., Studer, R. A., Thornton, J. & Orengo, C. A. (2015) CATH: Comprehensive structural and functional annotations for genome sequences, *Nucl. Acids Res.* **43**, D376-D381.

55. Kristjánsson, M. M. (2011) Thermostable subtilases (subtilisin-like serine proteinases), *Thermostable Proteins: Structural Stability and Design*, pp. 67-104.

56. Lee, Y.-C., Koike, H., Taguchi, H., Ohta, T. & Matsuzawa, H. (1994) Requirement of a COOH-terminal pro-sequence for the extracellular secretion of aqualysin I (a thermophilic subtilisin-type protease) in *Thermus thermophilus*, *FEMS Microbiol. Lett.* **120**, 69-74.

57. Tanaka, S., Saito, K., Chon, H., Matsumura, H., Koga, Y., Takano, K. & Kanaya, S. (2007) Crystal structure of unautoprocessed precursor of subtilisin from a hyperthermophilic archaeon: evidence for Ca2+-induced folding, *J. Biol. Chem.* **282**, 8246-8255.

58. Alexander, P. A., Ruan, B. & Bryan, P. N. (2001) Cation-dependent stability of subtilisin, *Biochemistry.* **40**, 10634-10639.

59. Siezen, R. J. & Leunissen, J. A. M. (1997) Subtilases: The superfamily of subtilisin-like serine proteases, *Protein Sci.* **6**, 501-523.

60. Ebeling, W., Hennrich, N., Klockow, M., Metz, H., Orth, H. D. & Lang, H. (1974) Proteinase K from *Tritirachium album Limber*, *Eur. J. Biochem.* **47**, 91-97.

61. Matsuzawa, H., Tokugawa, K., Hamaoki, M., Mizoguchi, M., Taguchi, H., Terada, I., Kwon, S.-T. & Ohta, T. (1988) Purification and characterization of aqualysin I (a

thermophilic alkaline serine protease) produced by *Thermus aquaticus* YT-1, *Eur. J. Biochem.* **171**, 441-447.

62. Matsuzawa, H., Hamaoki, M. & Ohta, T. (1983) Production of thermophilic extracellular proteases (Aqualysins I and II) by *Thermus aquaticus* YT-1, an extreme thermophile, *Agric. Biol. Chem.* **47**, 25-28.

63. Lin, S.-J., Yoshimura, E., Sakai, H., Wakagi, T. & Matsuzawa, H. (1999) Weakly bound calcium ions involved in the thermostability of aqualysin I, a heat-stable subtilisin-type protease of *Thermus aquaticus* YT-1, *Biochim. Biophys. Acta.* **1433**, 132-138.

64. Arnorsdottir, J., Magnusdottir, M., Fridjonsson, O. H. & Kristjansson, M. M. (2011) The effect of deleting a putative salt bridge on the properties of the thermostable subtilisin-like proteinase, aqualysin I, *Protein Pept. Lett.* **18**, 545-551.

65. Arnorsdottir, J., Kristjansson, M. M. & Ficner, R. (2005) Crystal structure of a subtilisin-like serine proteinase from a psychrotrophic *Vibrio* species reveals structural aspects of cold adaptation, *FEBS J.* **272**, 832-845.

66. Jónsdóttir, L. B., Ellertsson, B. Ö., Invernizzi, G., Magnúsdóttir, M., Thorbjarnardóttir, S. H., Papaleo, E. & Kristjánsson, M. M. (2014) The role of salt bridges on the temperature adaptation of aqualysin I, a thermostable subtilisin-like proteinase, *Biochim. Biophys. Acta.* **1844**, 2174-2181.

67. Óskarsson, K. R. (2015) Rational design of the cold active subtilisin-like serine protease VPR towards higher activity and thermostability, University of Iceland (Master's thesis), Reykjavík, Iceland.

68. Ellertsson, B. Ö. (2012) Áhrif saltbrúa og yfirborðshleðslna á hitastigsaðlögun subtilísin-líkra serín próteinasa, University of Iceland (Master's thesis), Reykjavík, Iceland.

69. Sigtryggsdottir, A. R., Papaleo, E., Thorbjarnardottir, S. H. & Kristjansson, M. M. (2014) Flexibility of cold- and heat-adapted subtilisin-like serine proteinases evaluated with fluorescence quenching and molecular dynamics, *Biochim. Biophys. Acta.* **1844**, 705-712.

70. Zaman, Z. & Verwilghen, R. L. (1979) Quantitation of proteins solubilized in sodium dodecyl sulfate-mercaptoethanol-Tris electrophoresis buffer, *Anal. Biochem.* **100**, 64-69.

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

72. Candiano, G., Bruschi, M., Musante, L., Santucci, L., Ghiggeri, G. M., Carnemolla, B., Orecchia, P., Zardi, L. & Righetti, P. G. (2004) Blue silver: A very sensitive colloidal Coomassie G-250 staining for proteome analysis, *Electrophoresis.* **25**, 1327-1333.

73. DelMar, E. G., Largman, C., Brodrick, J. W. & Geokas, M. C. (1979) A sensitive new substrate for chymotrypsin, *Anal. Biochem.* **99**, 316-320.

74. Keith J. Laider, J. H. M. a. B. C. S. (2003) Physical chemistry 4th edn., pp. 362-417, New York.

75. Kelly, S. M., Jess, T. J. & Price, N. C. (2005) How to study proteins by circular dichroism, *Biochim. Biophys. Acta.* **1751**, 119-139.

76. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis, *Journal of computational chemistry.* **25**, 1605-1612.

77. Dowd, J. E. & Riggs, D. S. (1965) A comparison of estimates of Michaelis-Menten kinetic constants from various linear transformations, *J. biol. Chem.* **240**, 863-869.

78. Sakaguchi, M., Matsuzaki, M., Niimiya, K., Seino, J., Sugahara, Y. & Kawakita, M. (2007) Role of proline residues in conferring thermostability on Aqualysin I, *J. Biochem.* **141**, 213-220.

79. Sakaguchi, M., Takezawa, M., Nakazawa, R., Nozawa, K., Kusakawa, T., Nagasawa, T., Sugahara, Y. & Kawakita, M. (2008) Role of disulphide bonds in a thermophilic serine protease Aqualysin I from *Thermus aquaticus* YT-1, *J. Biochem.* **143**, 625-632.

80. Tanaka, T., Matsuzawa, H., Kojima, S., Kumagai, I., Miura, K. & Ohta, T. (1998) P1 specificity of aqualysin I (a subtilisin-type serine protease) from *Thermus aquaticus* YT-1, using P1-substituted derivatives of Streptomyces subtilisin inhibitor, *Biosc. Biotech. Biochem.* **62**, 2035-2038.

81. Lin, S. J., Kim, D. W., Ryugo, Y., Wakagi, T. & Matsuzawa, H. (1997) Increase of the protease activity of aqualysin I, a thermostable serine protease, by replacing Asn219 near the catalytic residue Ser222, *Biosc. Biotech. Biochem.* **61**, 718-719.