Kinetic stability and temperature adaptation

Observations from a cold adapted subtilisin-like serine protease

Kristinn Ragnar Óskarsson

Faculty of Physical Sciences
University of Iceland
2020

Kinetic stability and temperature adaptation

Observations from a cold adapted subtilisin-like serine protease

Kristinn Ragnar Óskarsson

Dissertation submitted in partial fulfillment of a
Philosophiae Doctor degree in Biochemistry

Advisor
Professor Magnús Már Kristjánsson
Faculty of Physical Sciences at the University of Iceland.

PhD Committee
Professor Bjarni Ásgeirsson
Faculty of Physical Sciences at the University of Iceland.

Dr. Elena Papaleo
Associate Professor at Copenhagen University,
Head of Computational Biology at the DCRC.

Opponents
Professor Jose Manuel Sanchez-Ruiz
Faculty of Physical Chemistry at the University of Granada.

Dr. Tony Collins
Assistant Professor at the CBMA, University of Minho.

Faculty of Physical Sciences
School of Engineering and Natural Sciences
University of Iceland
Reykjavik, October 2020

Kinetic stability and temperature adaptation.
Observations from a cold adapted subtilisin-like serine protease.
Kinetic stability and temperature adaptation.
Dissertation submitted in partial fulfillment of a *Philosophiae Doctor* degree in Biochemistry

Faculty of Physical Sciences
School of Engineering and Natural Sciences
University of Iceland
Dunhaga 3
107 Reykjavik
Iceland

Telephone: +354 525 4000

Bibliographic information:
Kristinn Ragnar Óskarsson, 2020, Kinetic stability and temperature adaptation. Observations from a cold adapted subtilisin-like serine protease., Ph.D. dissertation, Faculty of Physical Sciences, University of Iceland, 208 pages.

Author ORCID: 0000-0003-0356-1770

ISBN 978-9935-9452-8-0
Háskólaprent hf.
Reykjavik, Iceland, October 2020

Abstract

Life on earth is found everywhere where water is found, meaning that life has adapted to extremely varied environments. Thus, protein structures must adapt to a myriad of environmental stressors while maintaining their functional forms. In the case of enzymes, temperature is one of the main evolutionary pressures, affecting both the stability of the structure and the rate of catalysis. One of the solutions Nature has come up with to maintain activity and stability in harsh environments over biological relevant timescales, are kinetically stable proteins. This thesis will outline work carried out on the kinetically stable VPR, a cold active subtilisin-like serine protease and discuss our current understanding of protein kinetic stability, temperature adaptation and our current hypothesis of the molecular interactions contributing to the stability of VPR. The research model that we have used to study these attributes consists of the cold active VPR and its thermostable structural homolog AQUI. The results discussed in this thesis will be on the importance of calcium, the role of prolines in loops, the role of a conserved N-terminal tryptophan residue and lastly primary observations on differences in active site dynamics between VPR and AQUI. A model is proposed of a native structure that unfolds in a highly cooperative manner. This cooperativity can be disrupted, however, by modifying calcium binding of the protein or via mutations that affect how the N-terminus interacts with the rest of the protein. The N-terminus likely acts as a kinetic lock that infers stability to the rest of the structure through many different interactions. Some of these interactions may be strengthened via proline residues, that seemingly act as anchor points that tend to maintain correct orientation between these parts of the protein as thermal energy is increased in the system. Our results give a deeper insight into the nature of the kinetic stability, the importance of cooperativity during unfolding of kinetically stable proteases, synergy between distant parts of the protein through proline mutations and how different calcium binding sites have vastly differing roles. The results provide a solid ground for continuing work in designing enzyme variants with desired stabilities and activities and improve our understanding of kinetically stable systems.

.

Útdráttur

Líf á jörðu finnst alstaðar þar sem vatn finnst og þess vegna hafa lífverur þurft að aðlagast að mjög fjölbreyttum aðstæðum. Þar af leiðandi hafa prótein lífvera þurft að þróa með sér mismunandi lausnir til að geta starfað í sínu umhverfi. Ef ensím eru skoðuð, þá er stærsti umhverfisþátturinn hitastig, sem hefur áhrif á bæði stöðugleika þeirra og hvötunargetu. Ein af þeim leiðum sem hafa þróast til þess að tryggja hvötunargetu í erfiðum aðstæðum er hraðafræðilegur stöðugleiki sameinda. Í þessari doktorsritgerð verður kynnt vinna sem hefur verið framkvæmd á hinum hraðafræðilega stöðuga, kuldavirka subtilísin-líka serín próteinasa VPR og tilgátur kynntar um hvaða millisameinda hrif kunna að stuðla að hitastigsaðlögun, virkni og stöðugleika þessa ensíms. Rannsóknarlíkanið sem hefur verið nýtt til þessara rannsókna samanstendur af VPR og hinum hitastöðuga subtilísin-líka serín próteinasa AQUI, sem hefur nánast sömu þrívíddarbyggingu og VPR. Ræddar verða niðurstöður um áhrif kalsíum jóna á stöðugleika, áhrif prólín amínósýra í lykkjusvæðum á stöðugleika, mikilvægi N-endastæðar trýptófan amínósýru fyrir bygginguna og kynning á fyrstu niðurstöðum úr tilraunum til að skoða sameindasveigjanleika innan hvarfstöðva VPR og AQUI. Líkanið sem sett er fram fyrir VPR, sýnir byggingu sem er mjög samheldin og afmyndast sem ein heild. En þessa samheldni byggingarinnar er hægt að trufla, til dæmis með því að skerða kalsíum bindingu eða með stökkbreytingum sem hafa áhrif á hvernig N-endi próteinsins hefur áhrif á aðra hluta af byggingunni. Sumir þessara innansameinda krafta geta verið styrktir með prólín stökkbreytingum, sem virðast virka sem eins konar akkeri sem skorða lykkjusvæði og viðhalda réttum sameindahrifum við hærri hitastig. Niðurstöður okkar veita dýpri sýn á eiginleika hraðafræðilegar stöðgunar, áhrifa samheldinna afmyndunarferla á stöðugleika, hvernig prólin amínósýrur hafa áhrif á sameindahrif milli fjarlægra svæða innan prótein-byggingarinnar og hvernig kalsíum binding á mismunandi bindisvæði getur haft margvísleg hlutverk. Þessi vinna skapar líka góðan grunn fyrir áframhaldandi vinnu fyrir hönnun ensímhvata með mismunandi hitastöðugleika og virkni fyrir áframhaldandi vinnu að því markmiði að auka skilning á hraðafræðilegum stöðugleika próteinsameinda.

“The first principle is that you must not fool yourself and you are the easiest person to fool.”

— Richard P. Feynman

Preface

This doctoral thesis will outline the work carried out on the kinetically stable VPR [1], a cold active subtilisin-like serine protease. This work aims to give further insights into the nature of the kinetic stability and temperature adaptation of proteins, utilizing our research model consisting of the cold active VPR and its thermostable structural homolog aqualysin I (AQUI) [2]. These enzymes are extracellular proteases that have their active structure under kinetic control, making this system a prime example to study kinetic stability and identify the structural aspects determining the ability of these enzymes to perform in their natural environments.

The first chapter will discuss the theoretical background relating to protein biophysics and in broad strokes go over the concepts used to define the thermodynamic and kinetic parameters used to quantify protein stability. Furthermore, it will be discussed how the definitions of these parameters give us the tools to gain deeper insights into how different organisms have evolved different strategies in adapting their proteomes to different temperatures, both regarding stability and activity as well as the relationship between them. The focus will then turn towards proteases, their classification and properties that will serve to paint a clear picture of the model system utilized in this work.

The second chapter will detail the methodologies used in the present study and data analysis.

The third chapter in this thesis will be an overview of the results from the different projects. There among are the improved expression of VPR in *E.coli* and the role of the different calcium binding sites for the stability of VPR, including the impact calcium has on expression [3]. Additional focus will then be directed at the effects of proline residues in loops on kinetic stability [4], the effects of a single tryptophan exchange on stability, and active site molecular movements of dansylated VPR and AQUI. Concluding the thesis will be a brief reflection on the work that has been carried out and future prospects.

Table of Contents

List of Figures

List of Tables

Abbreviations

VPR	–	Subtilisin-like serine protease from *Vibrio* sp. PA-44
AQUI	–	Aqualysin 1 from Thermus Aquaticus YT-1
LUCA	–	Last universal common ancestor
DNA	–	Deoxyribonucleic acid
ΔG	–	Change in Gibbs free energy
ΔH	–	Change in enthalpy
ΔS	–	Change in entropy
ΔC_p	–	Change in heat capacity at constant pressure
T_m	–	Melting point
T_c	–	Cold denaturing point
TST	–	Transition state theory
$TS^{\ddagger}$	–	The transition state/the activated state
$\Delta G^{\ddagger}$	–	Change in activation Gibbs free energy
$\Delta H^{\ddagger}$	–	Change in activation enthalpy
$\Delta S^{\ddagger}$	–	Change in activation entropy
$\Delta C_p^{\ddagger}$	–	Change in activation heat capacity at constant pressure
IMC	–	Intramolecular chaperone
SGPB	–	*Streptomyces griseus* Protease B
MD	–	Molecular dynamics
FRET	–	Förster resonance energy transfer
NMR	–	Nuclear magnetic resonance
ATP	–	Adenosine triphosphate
MMRT	–	Macromolecular rate theory
T_{opt}	–	Temperature of optimum activity
MEROPS	–	The database of proteolytic enzymes, their substrates and inhibitors

DFP	–	Diisopropyl-fluorophosphate
TPCK	–	Tosyl phenylalanine chloromethyl ketone
BPN´	–	A subtilisin-like protease from *Bacillus subtilis*
PCSK9	–	Proprotein convertase subtilisin-kexin type 9
IvaP	–	Subtilisin-like serine protease from *Vibrio cholerae*
IDP	–	Intrinsically disordered protein
KP-43	–	Subtilisin-like serine protease from *Bacillus* sp. KSM-KP43
Tk-SP	–	Subtilisin-like serine protease from *Thermococcus kodakaraensis*
EDTA	–	Ethylenediaminetetraacetic acid
SUB1	–	Subtilisin-like serine from protozoa of the *Plasmodium* genus
PA-domain	–	Protease associated - domain
PRK	–	Proteinase K
SPRK	–	Subtilisin-like serine protease from a psychrotrophic *Serratia* sp.
DSC	–	Differential scanning calorimetry
CD	–	Circular dichroism
$T_{50\%}$	–	The temperature which half of the activity is lost in half an hour
RMSD	–	Root mean square deviation
RMSF	–	Root mean square fluctuation
PCR	–	Polymerase chain reaction
DpnI	–	Type IIM restriction enzyme which digests methylated DNA
RNA	–	Ribonucleic acid
E_{act}	–	Activation energy
$t_{1/2}$	–	half-life
PMSF	–	Phenylmethylsulphonyl fluoride
FS	–	Fluorescence spectroscopy
AUC	–	Area under the curve
K_{sv}	–	The Stern-Volmer fluorescence quenching constant

CHARMM	–	Chemistry at Harvard Macromolecular Mechanics
NPT	–	Isothermal-isobaric ensemble
NVT	–	Canonical ensemble
LINCS	–	Linear constraint solver
PME	–	Particle mesh Ewald
DNSF	–	5-dimethylaminonaphtalene-1-sulfonyl fluoride
TCSPC	–	Time-correlated single photon counting
DAS6	–	Fluorescence decay analysis software
K_{cat}	–	Turnover number/catalytic constant
K_M	–	The Michaelis-Menten constant
K_{app}	–	Apparent equilibrium constant
MOPS	–	3-N-morpholino-propanesulfonic acid

Acknowledgements

I would like to give my sincere thanks to my supervisor Magnús Már Kristjánsson for entrusting me with this project, the guidance and enjoyable moments throughout the years.

Thanks to my doctoral committee Bjarni Ásgeirsson and Elena Papaleo for their involvement in this work.

And Bjarni Ásgeirsson thank you especially for great conversations over the years and the thorough review of this thesis.

Thanks to Elena Papaleo and Matteo Lambrughi for introducing me to ins and outs of molecular dynamic simulations and their preparations.

Jens G. Hjörleifsson, thanks for the near infinite number of conversations ranging from biochemical experimentations to the secrets of guitar playing and all the buckets of beer we have shared over the years.

Thanks to the previous master´s students of the group Arnór Freyr Sævarsson and Sveinn Bjarnason for help lifting the load this work has been.

Also, I would like to thank all past and present members of the Faculty of Physical Sciences for enjoyable times. Especially Sigurbjörn Markússon for those promising crystallization screens.

Of course, all my fellow members in our little Steak Club, Svanur Sigurjónsson, Einar Lúthersson, Jens G. Hjörleifsson, Baldur Kristjánsson, Sveinn Bjarnason and Daníel Arnar Tómasson. These cooking, eating and drinking meetups, have been the perfect way to relax.

And all my deepest thanks go to my family. Karen Lísa my dear thanks for tolerating me on a daily basis, still don’t know how you did it even I think I am boring at my crankiest. And my sincerest thanks to my parents Óskar Hjaltason and Sigríður Jónsdóttir for all the support for the last 29 years.

1 Introduction

1.1 Origins of temperature adaption

To gain an insight into the inner workings of temperature adaptation of proteins, one should look at the evolutionary history of life on Earth. Evidence for life on Earth can be traced back at least 3,400 million years as fossils of sulphur metabolizing microbes were found in Western Australia [5]. There are indications that life thrived on Earth even earlier than that, as C^{13}-isotope depleted graphite with distorted crystal structures have been observed in Western Greenland, which indicates the presence of microbial life in Earth's oceans around 3,700 million years ago [6]. The oldest possible evidence for life can however be traced as far back as 4,100 million years, where C^{13}-isotope depleted zircons from Australia potentially indicate enzymatic carbon fixation that far back [7]. This would then mean that life emerged soon after the formation of the primordial Earth around 4,550 million years ago [8]. How this primal life came to be and how it operated is a question that remains a debated topic. However, all life as we know it shares certain similarities suggesting some universal common ancestry as first suggested by Darwin [9]. This idea that life as we know it originated from some single branch of early life is a pillar of modern evolutionary theory that has been shown to stand the scrutiny of statistical models [10]. Pinpointing this last universal common ancestor (LUCA) would give us an important insight into the evolution of life and its adaptation to various conditions. One such attempt was made by looking at protein coding genes form sequenced prokaryotic genomes. From those, 355 protein families could be traced back to LUCA and painted the picture of an anaerobic thermophile that resembled modern microbes thriving in geochemically active environments such as hydrothermal vents [11]. This topic is, however, still highly debated as the origins of life and the identity of LUCA are far from being known [12].

The conditions in Earth´s oceans in early Precambrian times inferred from reconstructed ancestral proteins [13] and O^{18}-isotope analysis [14] place the temperatures of the ocean near 80°C around 3,500 million years ago. By extrapolation, the early life forms were

thermophilic organisms utilizing DNA as the genetic material, and proteins as building blocks and catalytic moieties. This would then put the starting point of temperature adaptation at high temperatures and the evolutionary process having to adapt to colder environments as Earth cooled down in the coming eons [13, 14]. However, the diversity of conditions in our biosphere may have led to highly complicated evolutionary paths for many organisms, where for instance the evolutionary path may have led to cold environments and then hot again, while also adapting to other environmental factors. This causes difficulties in distinguishing between artifacts and major contributors to temperature adaptation when comparing proteins adapted to different environments.

1.1.1 Temperature adaptation

Life has thrived under immense evolutionary pressures, under extreme conditions, at both extremes of the pH scale, at high salinity levels, bombarded by radiation and over a broad range of temperatures [15]. The temperature extremes which unicellular organisms can survive at range from around -25°C to 122°C [15-19]. Thus, at temperatures between these two extremes the molecular machinery that maintains life must function. Even more impressive are the theoretical temperature limits postulated, a lower limit of around -40°C/50°C, at which point the molecular machinery involved in the maintenance of a life form practically stops and the cells vitrify [20, 21]. The higher theoretical limit of life is set at around 140°C/150°C, where metabolic intermediates such as nicotinamide cofactors can spontaneously hydrolyze and the life-time of larger biomolecules would be so short that life could not be maintained [15, 22]. Over this wide temperature range the barriers that evolution has had to overcome to enable life forms to survive are vastly different. In the case of enzymes, temperature can be viewed as a main evolutionary driver, as temperature directly affects the rate of chemical reactions catalyzed by these proteins. At the higher end of the temperature range, enzymes would have to adapt to the extreme molecular movements and thus a very stable structure would be required to maintain the active state of the protein for a certain amount of time. On the other end of the temperature spectrum, the main challenge to overcome would be the slower reaction rates, thus a more catalytic efficient enzyme would be needed to maintain chemical reaction rates to sustain life. Many observations have been made of very active enzymes with low stabilities and extremely stable enzymes that were poor catalysts at ambient temperatures, thus suggesting an apparent trade-off between these properties, i.e. catalytic

activity and thermal stability [1, 23-25]. This prompted the activity/stability trade-off hypothesis, which states that in order to achieve high stability the molecular motions needed for rapid catalysis are sacrificed and vice versa. This would mean that during the evolutionary history of organisms, adapting to cooler environments would be achieved by a reduction in molecular contacts/interactions involved in the stability of the enzyme in order to allow for the movements needed to facilitate rapid catalysis. However, the evolutionary history is long and convoluted, evolutionary pressures do change with time. Research into adenylate kinases from organisms that have had vastly different evolutionary history showed that adenylate kinases from *B. stearothermophilus* had adapted to colder environments and then evolved again to function at high temperatures, retaining relatively high turnover rates at lower temperatures [26]. In contrast, adenylate kinases that had throughout their evolutionary history always been adapted to high temperatures, like those from *C. subterraneus* and *A. aeolicus* had a much steeper dependence on temperatures [26]. In addition, directed evolution experiments carried out on subtilisins showed that higher stability was achievable without compromising activity. Furthermore, it was shown that higher activity at lower temperatures was also achievable without stability loss [27, 28]. This indicates that the stability/activity trade-off is not a strict relationship, or at least is an oversimplification of the underlying causes, arising from the enthalpy and entropy compensations due to protein-water, protein-substrate and water-substrate interactions at various temperatures [29]. Thus, the stance can be taken that observations of activity/stability tradeoffs in nature are in part evolutionary artifacts due to selective pressures, as at lower temperatures the pressure is on maintaining catalytic rates but not on stability of the protein structure. Random mutations causing destabilizing changes could thus occur that would not be detrimental to the enzyme in the current environment, consequently, no pressure would be against it. But how do living organisms solve the problems which arise due to different temperatures? This is a question that has been on the mind of many researcher in various fields of science. In order to shed light on this topic, this thesis will transverse into the fundamentals of protein stability.

1.2 Protein stability

Protein stability describes the ability of the native/active structure of a protein to resist environmental stressors capable of inducing the transition towards a nonactive/denatured state. The stability of a protein structure can be described as having two facets, a

thermodynamic one and a kinetic one. Both aspects have an important role in the stabilization of protein structures. These two facets do however not contribute evenly to the prevalence of the native active form of proteins. Thus, proteins can be roughly divided into kinetically stable and thermodynamically stable structures.

1.2.1 Thermodynamics of protein stability

The thermodynamic part of protein stability describes the equilibrium between the states/assembles along the unfolding path. Defining and understanding these parameters can provide a wealth of information on the energetics of a system and how the underlying energetics can relate to protein stability [30]. Here, a reversible two-state unfolding/folding model shall be considered. Although being a very simplified model, the high degree of cooperativity observed in the unfolding processes of many proteins often make unfolding transitions seem two-state [31]. Here, a two-state model of a structured globular protein, in a water solution following Anfinsen´s thermodynamic hypothesis [32], will be used and serves well as an explanation model. Thus, the simplest system can be described as:

$$N \overset{K}{\rightleftarrows} D \tag{1}$$

where N represents the native state, D is the denatured assembly of states and K is the equilibrium constant. Considering a globular protein possessing a well-defined α/β tertiary structure and a hydrophobic core, a myriad of non-covalent interactions take place, such as van der Waals interactions, H-bonds, cation-π interactions, anion-π, charge-charge interactions, solvation or protein-water interactions and various ligand binding interactions, all which contribute to the stability of the protein [3, 33-42].

Enthalpy and entropy

The first factor that can be thought to be stabilizing the protein structure are the internal energies of these interactions. As the two states of the system (N and D) have vastly different physical properties and structures, most of the interactions that exist within the native state are lost during unfolding. In a biological system that is under fixed conditions, i.e. where there is no change in pressure and volume, the sum of the internal energies from the myriad of interactions that are lost in this transition are equal to the enthalpy change (ΔH) of unfolding [43]. As shown in equation 2:

$$\Delta H = \Delta U + \Delta PV \quad (2)$$

where ΔH is the change in enthalpy, ΔU is change in internal energy and ΔPV is the change in pressure and volume that is equal to zero in this case.

The second factor to be considered is the entropy change (ΔS) of the system. Entropy describes an intrinsic property of the universe detailed by the second law of thermodynamics and popularized by the saying "in an isolated system, entropy can only increase", i.e. the tendency of energy to disperse. In a system as considered here, a structured protein solvated by a water shell, one way to look at this phenomenon is through statistical mechanics, where the definition of entropy is:

$$S = k_B \ln(W) \quad (3)$$

where k_B is the Boltzmann constant and W the multiplicity of our system, the number of different arrangements the system can take. Statistical mechanics being a daunting subject in the eyes of many, is probably best laid out by the quote:

"Ludwig Boltzmann, who spent much of his life studying statistical mechanics, died in 1906, by his own hand. Paul Ehrenfest, carrying on the work, died similarly in 1933. Now it is our turn to study statistical mechanics. Perhaps it will be wise to approach the subject cautiously"
Opening lines of "States of Matter", by D. L. Goodstein.

For the purpose of gaining an insight into the nature of entropy, a simplified view of equation 3 will be considered. One simple way to look at it is to decide that all random arrangements or microstates of the system are considered equally likely. However, if all arrangements were grouped up according to similarity, those groups would be very differently populated. For example, the number of different arrangements where all units of a system are tightly clumped up in a system, is infinitesimally small against the number of different microstates where all units are spread around the system. This means that by randomly selecting a microstate from an extremely large system, the one selected will have its units spread around the system. Thus, random behavior becomes deterministic in an extremely large system. From this behavior the Boltzmann distribution naturally establishes itself [43]:

$$N_i \propto e^{-U_i/k_B T} \quad (4)$$

where N_i is the population of a state with the energy U_i, k_B is the Boltzmann constant and T is the absolute temperature. Thus, the higher the energy of the state is, the less likely it is to be occupied. However, as temperature/energy is increased the number of possible microstates of higher energies increases as well, but the most probable are still those where that energy is evenly distributed around the system. Therefore, due to randomness, every system favors its most spread energy state, or in other terms the entropically higher states, as the multiplicity of microstates with that energy distribution is higher than for the others.

In our protein and water system, the factors directly working against even energy distribution/entropy can be simplified into two main elements, the conformational entropy of the peptide chain and the conformational entropy of the water shell[44-47]. The conformational entropy of the peptide chain in a folded protein must be very low, as the multiplicity of a well-defined protein structure is way lower than the allowed arrangements of a structureless peptide chain. Thus, from that point of view the unfolded state must be much more entropically favorable. The protein, however, has a well-defined hydrophobic core and when unfolded the residues making up the hydrophobic core are more exposed. This could lead to fewer hydrogen-bonding possibilities per surface area, constricting the available arrangements of water molecules compared to the surface of the native state of the protein. Thus, at lower temperatures the entropic penalty of hydrating the unfolded state is higher than the conformational entropic gain of the unfolded state, thus favoring the native state at lower temperatures.

The established state of a system is controlled by the internal energy and the probability of a moiety within the system to absorb the energy to induce change. Thus, entropy and enthalpy of a system could be used to predict if a system will spontaneously change. The measure commonly used for this purpose is the Gibbs free energy, named after the American scientist Josiah Willard Gibbs, who showed how the first and second laws of thermodynamics could graphically be tied together [48], which leads to the correlation between entropy, enthalpy and the Gibbs free energy:

$$\Delta G = \Delta H - T\Delta S \tag{5}$$

where ΔH stands for the change in enthalpy, ΔS for the change in entropy, T for the absolute temperature and ΔG for the change in Gibbs free energy (ΔG). Under conditions where the protein favors the native state the value of ΔG is positive, as the native state is lower in

energy and thus favored. As these are state functions, the stability of the structure is completely correlated with the current conditions within the system and one may ask how stable these structures are under natural conditions in their "native" environments. The answer to that question is that they are only marginally stable (Fig. 1.1). ΔG values of unfolding for thermodynamically stable proteins have been found to be around 20 – 60 kJ/mol [49, 50], supporting the notion that an ultimate result from evolution is a structure that is just stable enough to function under their native conditions.

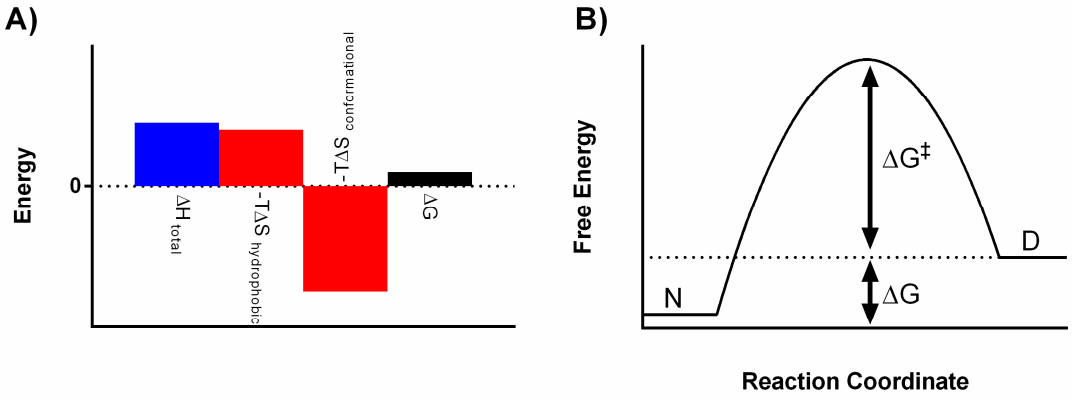


Figure 1.1. A) The contribution of enthalpy and entropy changes to the free energy change of unfolding for a hypothetical thermodynamically stable protein at a temperature where the structure is stable. B) Free energy reaction diagram of the unfolding of the hypothetical protein under the same conditions, where N stands for the native state, D for the denatured assembly, ΔG for the free energy difference between the states and ΔG‡ is the activation free energy.

As shown in (Eq. 5) the entropic term will equal the enthalpic term, so that at a certain temperature the value for ΔG equals zero. Using (Eq. 6):

$$\Delta G = -RT * \ln(K) \tag{6}$$

where R is the gas constant and T is the absolute temperature. At that temperature the equilibrium constant equals 1:

$$K = {}^{[D]}\!/_{[N]} \tag{7}$$

meaning the two states are equally populated. This temperature is defined as the melting point (T_m) of the protein. As the temperature of the system is increased further, the denatured assembly becomes dominant (Figs. 1.1 and 1.2).

The picture of protein stability described so far illustrates how the different parameters of the Gibbs equation contribute to protein stability or instability, where temperature acts as a

multiplier for the entropic component. Temperature also affects the enthalpy and entropy terms themselves. The key component which is missing in the Gibbs equation and describes this temperature dependence of ΔH and ΔS, is the change in heat capacity upon unfolding, ΔC or ΔC_p, as constant pressure mostly applies in biological systems.

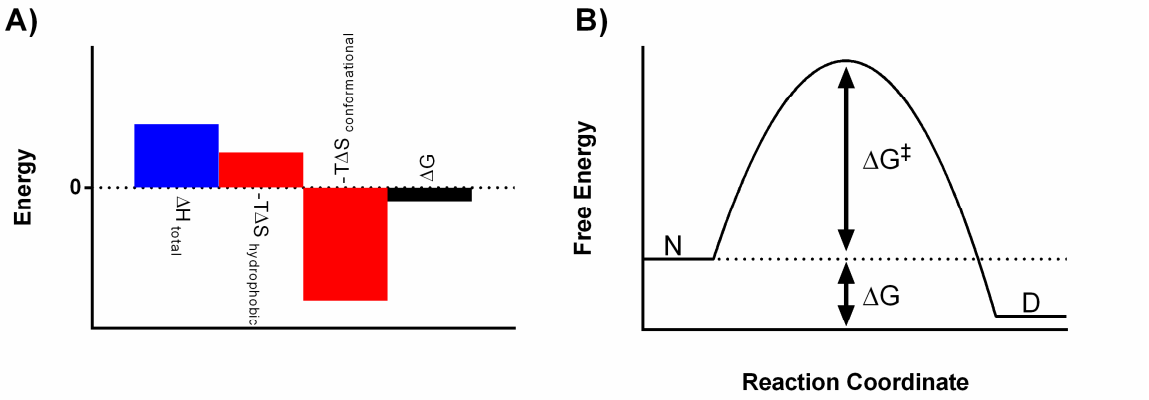


Figure 1.2. A) The contribution of enthalpy and entropy changes to the free energy change of unfolding for a hypothetical thermodynamically stable protein at high temperatures where the structure is unstable. B) Free energy reaction diagram of the unfolding of the hypothetical protein under the same conditions, where N stands for the native state, D for the denatured assembly, ΔG for the free energy difference between the states and $\Delta G^{\ddagger}$ is the activation free energy.

Heat capacity

Heat capacity is a measure of the energy required to bring about a specific temperature increase to a material or a solution. When thermal energy is applied to a solution two things happen, partly it increases the kinetic energy of molecules such as higher rotational, vibrational and translational energies, therefore increasing the temperature of the solution. The rest of the energy gets absorbed and is stored as potential energy within the molecules and their bonds. The heat capacity of a certain state within a system is essentially related to how well that state can distribute energy throughout a system [51]. In liquid water a single molecule is able to form up to four different H-bonds, breaking up and reforming H-bond with a multitude of other water molecules, all occurring within a very short timescale. Liquid water is thus able to distribute energy throughout the system via collisions and interactions with other water molecules and hence has a high heat capacity. This description essentially links this property to both the enthalpy and the entropy of a system. These relationships can be described with equations 8 and 9, where ΔC_p is regarded to be independent of temperature. This is not the case however, but is often approximated to be negligible over the temperature range where biological systems are usually observed at [52]:

$$\Delta H = \Delta H_{ref} + \Delta C_p\left(T - T_{ref}\right) \quad (8)$$

$$\Delta S = \Delta S_{ref} + \Delta C_p \; ln\left(\frac{T}{T_{ref}}\right) \quad (9)$$

where ΔH_{ref}, ΔS_{ref} and T_{ref} stand for the enthalpy change, entropy change and temperature at a convenient reference point. So, heat capacity change describes how the enthalpy and entropy change differ over a range of temperatures. From these relationships it is apparent that heat capacity is an essential thermodynamic quantity. By substituting the terms from equations 8 and 9 into equation 5 and rearrange yields:

$$\Delta G = \Delta H_{ref} - T\Delta S_{ref} + \Delta C_p\left(\left(T - T_{ref}\right) - T \; ln\left(\frac{T}{T_{ref}}\right)\right) \quad (10)$$

knowing the enthalpy change (ΔH_{ref}) and entropy change (ΔS_{ref}) at some reference temperature (T_{ref}) along with heat capacity change (ΔC_p) allows for calculations of the free energy change (ΔG) as a function of temperature. Now in practice the reference temperature is often the melting point (T_m). And as mentioned earlier, ΔG at that point is 0, so equation 5 can be rewritten as:

$$\Delta S_{T_m} = \frac{\Delta H_{T_m}}{T_m} \quad (11)$$

By exchanging the entropic term in equation 10 with equation 11 and rearranging yields a modified version of the Gibbs-Helmholtz equation:

$$\Delta G = \Delta H_{T_m}\left(1 - \frac{T}{T_m}\right) + \Delta C_p\left((T - T_m) - T \; ln\left(\frac{T}{T_m}\right)\right) \quad (12)$$

Thus, the free energy change as a function of temperature can be obtained from knowing just the melting temperature (T_m), ΔC_p and the change in enthalpy at the melting temperature. For example, all these values can be calculated from differential scanning calorimetry data of a “well behaved” protein such as shown in Fig. 1.3. The melting point is the high point of the thermogram, and ΔC_p can be estimated form the difference between the heat capacities

of the two states, before and after protein unfolding. ΔH_{Tm} is equal to the area under the curve of a baseline subtracted thermogram.

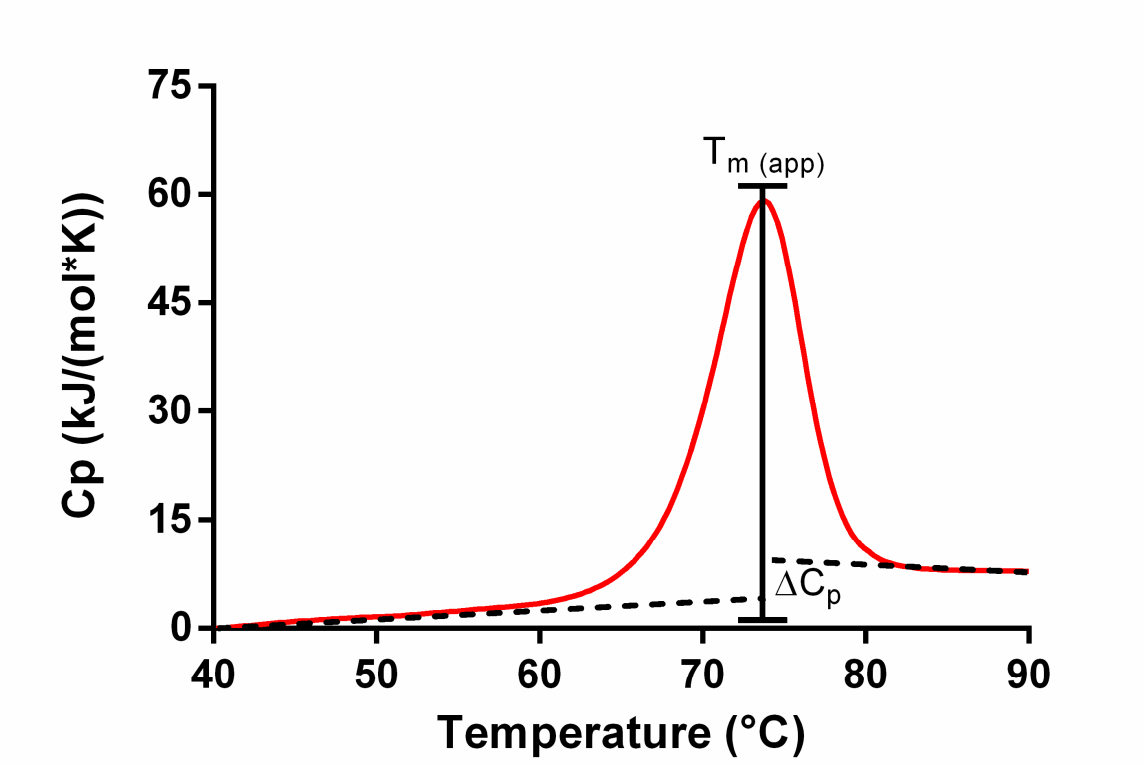


Figure 1.3. A differential scanning calorimetry (DSC) thermogram recorded of protein that undergoes reversible unfolding. The red line shows the buffer subtracted heat capacities, dotted lines are extrapolations of the heat capacities of the native and unfolded state, respectively, showing the change in heat capacity at the observed melting point.

The heat capacity change upon unfolding of proteins usually manifests itself as a positive value for globular proteins. A positive value for ΔC_p has some interesting effects on the stability curves. As if we look at equations 8 and 9, it entails that the value for ΔH increases linearly with higher temperatures from its lowest value at 0 degrees Kelvin. The term ΔS, however, starts at its lowest point at temperatures closing in on 0 degrees Kelvin and increases in a logarithmic manner. As a result of this, the term $-T\Delta S$ exhibits a sharp rise at extremely low temperatures, then reaches a plateau that gradually starts falling. At

temperatures that are biologically relevant this change looks almost linear and leads to a free energy stability diagram that has a parabolic shape (Fig. 1.4.)

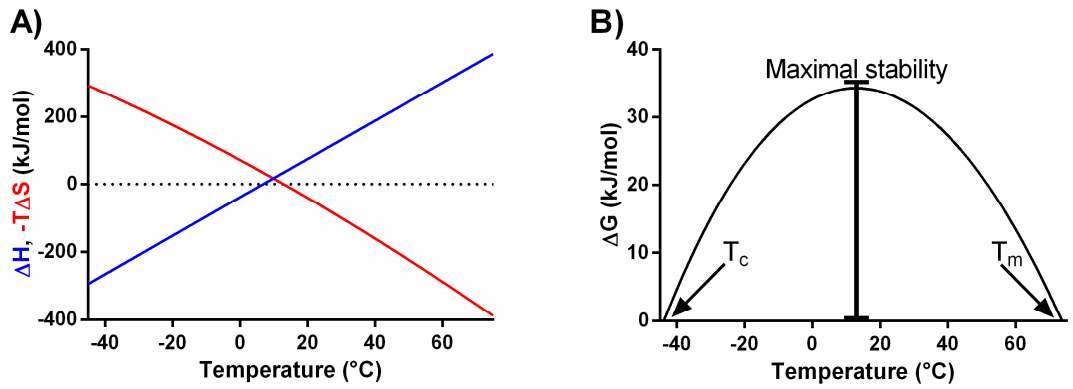


Figure 1.4. A) The effect of temperature on entropy (red line) and enthalpy (blue line) for the hypothetical protein having a positive ΔC_p. The resulting free energy curve, showing the temperature dependence of stability, showing the cold denaturation point (T_c), the point of maximal stability and the melting point (T_m).

A "parabolic" free energy diagram has the interesting consequence that it predicts that a protein molecule, such as the one depicted here, has a cold denaturation point (T_c), in addition to a temperature point of maximal stability [53]. It does not just have a melting point (T_m) and a structure that becomes progressively more stable with lower temperatures as would be the case if ΔC_p would be disregarded. Evidence for lower stability at low temperatures can be traced back to the 1930s, where it was noted that ovalbumin denatured faster at low temperatures when incubated in strong urea solutions [54]. In the decades after that, interest in low temperature denaturation and inactivation increased [55]. In addition, cold denaturation has been observed for the yeast globular protein frataxin without denaturants being added [56, 57]. The molecular mechanism for the behavior of the unfolding process has been of interest for decades and often thought to be somewhat of an anomaly, as it sounds rather counterintuitive that at higher temperatures more energy is needed to unfold a protein and that at sufficiently low temperatures exothermic unfolding could take place [51, 58]. The positive heat capacity change upon unfolding is however not a unique thing. This behavior is reminiscent to that of the melting of various organic solvents and water [59, 60]. Keeping that in mind, thermal unfolding of a protein can be thought of as an order to disorder transition where a multitude of weak bonds are broken [51, 52, 61, 62]. Looking at the whole picture, those disrupted interactions would include protein-protein, water-protein and water-water interactions in the solvation shell of the protein [63]. Water-protein interactions as a function of temperature have long been of interest in protein stability studies and those interactions

have been thought of as one of the main drivers of cold denaturation i.e. that the hydrophobic effect would be weaker at lower temperatures [55, 64]. Molecular dynamics simulations on the cold denaturing frataxin seem to support that notion [65]. There, the assemblies of the protein under cold denaturing, native and heat denaturing conditions were examined. The picture that emerged was that the cold denatured assembly was highly solvated i.e. a high number of H-bonds between the protein and the solvent existed, resulting in an expanded denatured structure. Then with higher temperatures, more protein-protein H-bonds can form as interactions between water molecules and hydrophobic surfaces become more repulsive, resulting in the formation of the native structure. The thermally denatured assembly of frataxin showed an even further reduction of water-protein H-bonds resulting in a relatively compact denatured state that was more structured than the cold denatured one [65]. An observation that coincides with the observations that intrinsically discorded proteins often show increased compactness and structure content at higher temperatures [66, 67].

Thus, it can be postulated that upon cold denaturation there is a gain in the entropy of the peptide chain as more microstates can be sampled. However, there would be an entropic loss in the solvation shell and the water solvent due to the higher recruitment of water needed to solvate the protein. The negative enthalpy change would then be related to the high number of protein-water hydrogen bonds and van der Waals interactions between water molecules and the peptide possible at low temperatures. Thermal denaturation would manifest itself as a large increase in the conformational entropy of the peptide chain and an increase in the entropy of the surrounding water. Thus, weaker protein-water interactions would become entropically unfavorable, but the total entropy would favor the denatured state. The increase in enthalpy could then be described as a lack of protein-water interactions leading to more prevalent protein-protein interactions at high temperatures.

Considering this, the enthalpy/entropy compensation observed for globular proteins seems like a logical result from the protein-water interactions and the balance between enthalpy and entropy for both the protein itself and the solvation shell [29, 68].

1.2.2 Kinetics of protein stability

All the aspects of protein stability discussed so far have been on the thermodynamic properties of a protein system at equilibrium. As ΔG, ΔH and ΔS are all state functions, they just describe the system under the given conditions, but give no information on how fast the

system can sample those states. As a consequence, the kinetic part of protein stability plays an essential part for many systems in maintaining the long term stability of some proteins [69]. A system that has no virtual kinetic barriers between functional states and non-functional states, has no protection against temporary stressors and is thus more prone to loss of function. Understanding the underlying mechanism of the kinetics governing the sampling of states is thus essential to understanding protein stability in many cases.

The history of kinetic analysis of various chemical reactions is long and an interesting one [70]. Many contributed to the development of various concepts describing chemical rates. However, around 1910s the first generally accepted way of analyzing temperature effects on chemical rates was the Arrhenius equation, developed by Svante Arrhenius based upon the work of van´t Hoff [70-72]:

$$k = Ae^{\frac{-E_a}{k_B N_A T}} \tag{13}$$

where k is the rate constant of an observed event, A is the pre-exponential factor, E_a is the activation energy, k_B is the Boltzmann constant, N_A is Avogadro´s number and T the absolute temperature. The Arrhenius equation provides an excellent empirical relationship between temperature and reaction rates. The equation states that the rate constant is determined by the minimum energy (E_a) needed to induce a state change, the energy in the system i.e. the temperature and the pre-exponential factor. The Arrhenius equation thus provides initial insights into the energetics of experimental data. The ideas of enthalpy and entropy being part of the equation, as the entropy governs the probability of a molecule reaching an energy state, embedded in the pre-exponential factor, and the enthalpy correlating with the activation energy.

As a result, a deeper understanding of the meaning behind activation energy and the preexponential factor was sought, leading to the development of the transition state theory (TST). Transition state theory presented in 1935 is arguably the most useful tool for the analysis of experimental data even though it has not delivered the best results when calculating potential energy surfaces and/or predicting them [70, 73]. The theory was developed through a series of investigations and different treatments, and it is usually contributed to H. Eyring, M. Polanyi and M. G. Evans. TST, or the activated complex theory, states that all reactants/native states are in a quasi-equilibrium with the activated state/transition state ($TS^{\ddagger}$) [74]. The reaction rate as a function of temperature is described by the Eyring equation:

$$k = \kappa \frac{k_B T}{h} e^{\left(\frac{\Delta S^{\ddagger}}{R}\right)} e^{\left(\frac{-\Delta H^{\ddagger}}{RT}\right)} \quad (14)$$

where k is the reaction rate, κ is the transmission coefficient, k_B is the Boltzmann constant, T is the absolute temperature, h is the Planck constant, $\Delta S^{\ddagger}$ is the activation entropy, $\Delta H^{\ddagger}$ is the activation enthalpy and R is the gas constant i.e. the Boltzmann constants times the Avogadro´s number. The transmission coefficient is often assumed to be equal to one, meaning that all reactants that cross the transition state will end up as a product i.e. there is no re-crossing. The equation can thus be rewritten and simplified as:

$$k = \frac{k_B T}{h} e^{\left(\frac{-\Delta G^{\ddagger}}{RT}\right)} \quad (15)$$

where $\Delta G^{\ddagger}$ is the activation free energy. The theory has been shown to be applicable to various thermally activated processes, from protein folding/unfolding and DNA translocation to the breaking of chemical bonds, processes that differ immensely in time scales [75]. For the unfolding of a protein molecule TST model can be written out by the Lumry-Eyring model [76] as:

$$N \overset{K}{\rightleftarrows} TS^{\ddagger} \overset{k}{\rightarrow} D \quad (16)$$

where N is the native state, $TS^{\ddagger}$ is the transition state, D is an inactive/unfolded state, K is the equilibrium constant between the native state and the transition state, i.e. the quasi equilibrium, and k is the rate constant of the irreversible conversion of the transition state to the unfolded state. Such an analysis provides the researcher with the thermodynamic parameters describing the difference in energetics between the native state and the $TS^{\ddagger}$. For such an analysis to provide meaningful results, there needs to be a Boltzmann distribution between the native assembly and the $TS^{\ddagger}$, an assumption that can be presumed to be true for large molecules such as proteins due to the sheer number of different microstates a protein could sample while reaching the $TS^{\ddagger}$. The assumption of a quasi-equilibrium, critical for the validity of TST, has however been debated since it was first introduced [70]. In 1940, H.A. Kramers developed his reaction-rate theory [77] for chemical reactions in solutions. His theory also considered the viscosity of a solution, thus for a unimolecular reaction in water his

theory described the crossing of a Brownian particle over an energy barrier (Fig. 1.5). This is arguably a more realistic look at a protein crossing the energy barrier in solutions compared to older theories.

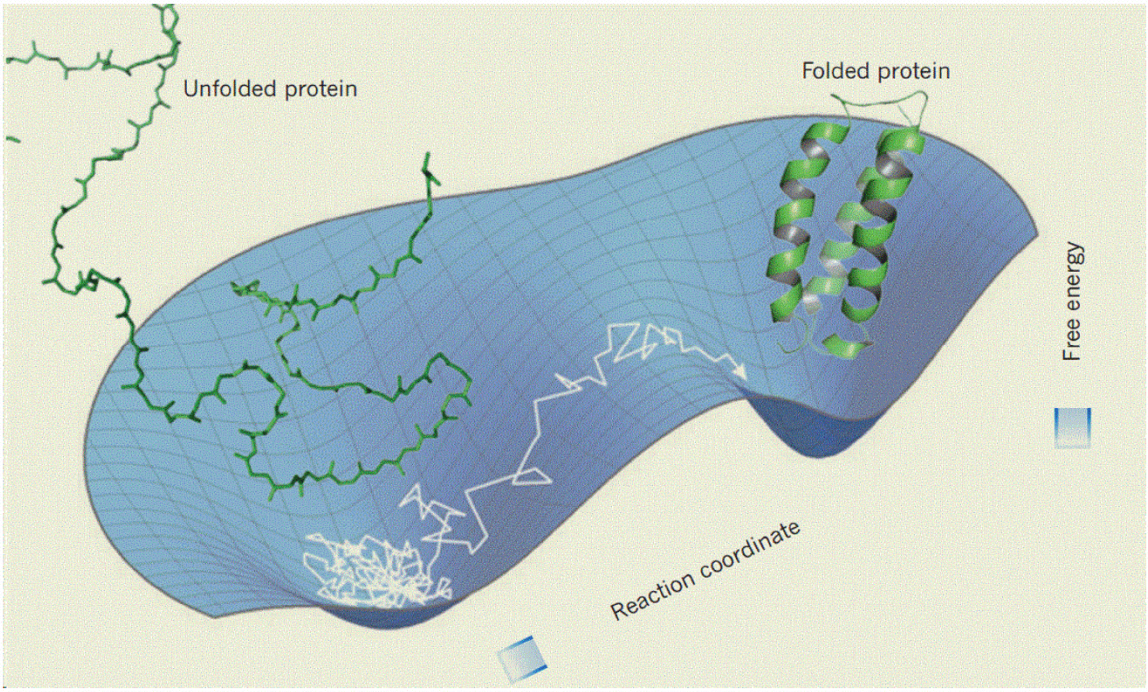


Figure 1.5. A 3D energy landscape depicting a protein acting as a Brownian particle diffusing over the free energy barrier (white arrow) separating the denatured and native assemblies. Kramers theory describing a one-dimensional cut of this landscape descripting the curvatures of the energy landscape in the native, denatured and transition state, taking into account the friction between the particle and the energy landscape. Reprinted from reference [79].

In order to achieve this, the variables that are taken into account must describe the energy landscape in much greater detail than just the height of the barrier that separates the states. This entails a way to estimate the curvatures of the energy landscape, that inform on how well a polypeptide can diffuse between the various microstates along the unfolding/folding trajectory. The model also accounts for friction between the diffusing particle and the free energy surface. In the simplest case this friction is entirely due to solvent viscosity, dampening the trajectories over the barrier. However, in the case of a protein molecule, internal protein-protein interactions might also contribute to the friction within the system [78]. Such a detailed description of an energy landscape for an unfolding/folding process is experimentally challenging, even when just considering a one-dimensional free energy landscape. However, the model has been successfully used to describe barrier crossing of an α-helical synthetic protein using single molecule fluorescence spectroscopy [78, 79]. Thus, it can be said that modeling of kinetic processes can be a great challenge, especially for complex systems such as protein folding and unfolding.

Kinetically stable proteins

In general, when proteins are referred to as kinetically stable, it usually encompasses all proteins that unfold irreversibly, meaning that the unfolded state is either trapped in a very deep free energy "valley", and consequently unable to refold, or has been chemically altered or precipitated and thus essentially removed from the original equilibrium reaction [69]. This can be the case for thermodynamically stable proteins that may unfold in an irreversible manner under certain circumstances, often causing pathological conditions [69, 80]. On the other side there are proteins that are truly kinetically stable, meaning that the native active structure is not necessarily favored thermodynamically, but the stability of the molecule is solely reliant on the free energy barrier separating the states. This type of free energy landscape was first shown to exists in the case of α-lytic protease from the bacterium *Lysobacter enzymogenes* [81]. The kinetic aspect is highly important for the overall stability of many different proteins, and seems to have evolved for some systems in order to deal with harsh environmental conditions such as crowded intracellular compartments or secreted into harsh environments in case of extracellular proteins [69]. The latter seemingly is especially prevalent and is best evidenced by the convergent evolution of the intramolecular chaperones of different proteases, most notably from the subtilisin clan, chymotrypsin clan and the carboxypeptidase clan, among others [82-89]. Being produced with an intramolecular chaperone (IMC), the unprocessed protein is thermodynamically stable and folds to its native state in a facile manner due to a low kinetic barrier. The maturation process of auto-cleavage produces a free catalytic domain that now is disconnected from the original folding-equilibrium and locked into a kinetically stable state (Fig. 1.6) [84]. Kinetically stable proteins tend to exhibit simple first order unfolding kinetics, although that tendency is also often observed in the cases of simpler monomeric systems [3, 90-99]. The simplest descriptor for a system that unfolds irreversibly, would be:

$$N \overset{k}{\rightarrow} D \tag{17}$$

where N and D denote the native and denatured states and k stands for the first order rate constant of unfolding. This leaves some proteins that solely rely on the kinetic barrier for stability out of the scope of the transition state theory in the conventional sense, such as α-lytic protease and *S. griseus* protease B (SGPB) and many subtilisin-like proteases that rely on IMC to fold (Fig. 1.6) [1, 100-102]. However, the two-step model can be interpreted as a version of the Lumry-Eyring model, where a quasi-equilibrium between the native state and

the $TS^{‡}$ is not established at observed unfolding temperatures [91, 103]. Thus, instead of describing the N to $TS^{‡}$ with an equilibrium constant the term has been replaced with the rate constants:

$$N \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} TS^{‡} \overset{k_2}{\rightarrow} D \tag{18}$$

where k_1 is the rate of $TS^{‡}$ formation from the native state, k_{-1} is the rate of which the $TS^{‡}$ returns to the native state and k_2 is the rate of $TS^{‡}$ irreversible step to the fully denatured state. This model would appear as a two-state system if one forward rate constant completely controls the apparent rate of unfolding under given experimental conditions. The scenario described here by a high kinetic barrier entails a low k_1 rate constant and a very fast k_2 rate constant, essentially being in a "pseudo" steady-state [104, 105]. The rate of deactivation of the $TS^{‡}$ to the native state (k_{-1}) being much slower than k_2 at temperatures where k_1 is readily measurable.

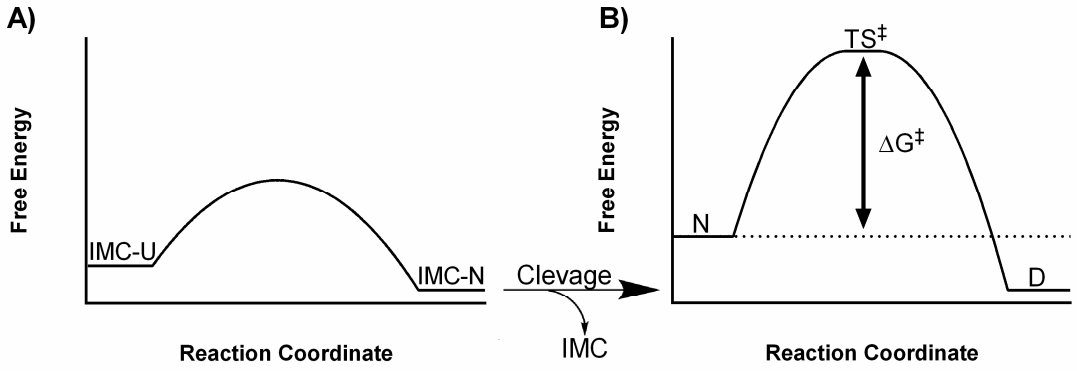


Figure 1.6. A) A theoretical energy landscape for the folding of a protein utilizing an intramolecular chaperone (IMC) from an unfolded state (IMC-U) to a folded state (IMC-N) followed by a cleavage of the IMC. B) The new theoretical energy landscape of the newly formed active protein, where the native (N) and denatured (D) are separated by a large free energy activation barrier ($\Delta G^{‡}$) and TS* is the transition state

The measured rate of unfolding will then change as a function of temperature according to the activation energy of that transition as described by the Arrhenius equation. This Arrhenius behavior can only be estimated to occur at strongly denaturing temperatures. Over that temperature range, where the assumption is made that the apparent rate of unfolding represents the rate of $TS^{‡}$ formation, the free energy barrier could be calculated by the Eyring equation (Eq. 15). Furthermore, by fitting the Eyring equation to the Arrhenius plot over the limited temperature range that the observed transition encompasses, and remains linear, the activation enthalpy and entropy can be derived. This is done by replacing the terms of

activation energy (E_a) and the pre-exponential factor A from eq. 13 with the enthalpic term and the entropic term from eq. 14, setting the transmission factor to unity, giving the following relationships for a unimolecular reaction:

$$E_a = \Delta H^{\ddagger} + RT \tag{19}$$

$$A = \left(\frac{k_B T}{h}\right) e^{\left(1+\left(\frac{\Delta S^{\ddagger}}{R}\right)\right)} \tag{20}$$

The relationship of the height of the activation free energy barrier ($\Delta G^{\ddagger}$) and the activation enthalpy and activation entropy being:

$$\Delta G^{\ddagger} = \Delta H^{\ddagger} - T\Delta S^{\ddagger} \tag{21}$$

The validity of state functions acquired this way is questionable and could only at the best of times be considered as apparent free energy parameters over the narrow temperature range where the unfolding transition is readily observed. This is because the assumptions that are being made exclude evaluation of how the thermodynamic parameters change as a function of temperature, as there is no information on the activation heat capacity change ($\Delta C^{\ddagger}$) for the $TS^{\ddagger}$ formation.

One method to estimate $\Delta C^{\ddagger}$ and the other thermodynamic properties of the $TS^{\ddagger}$ was demonstrated for the α-subunit of a tryptophan synthase, a thermodynamically stable protein, by measuring unfolding rates at selected temperatures using various strongly unfolding concentrations of a denaturant [106]. This method works on the premise that the apparent rate of unfolding only reflects on the rate of $TS^{\ddagger}$ formation [101, 107]. By using a denaturant, unfolding rates at low temperatures become experimentally available, and these can be extrapolated to unfolding rates in water using the appropriate model. In the case of the α-subunit of tryptophan synthase the data could be fitted by a denaturant-binding model [49]. Thus, faster unfolding rates due to a positive activation heat capacity value could be observed at low temperatures. This allows for fitting of the apparent unfolding rates to an extended version of the Eyring equation where values for $\Delta C^{\ddagger}$, $\Delta H^{\ddagger}$, $\Delta S^{\ddagger}$, and thus $\Delta G^{\ddagger}$, can be calculated:

$$\ln\left(\frac{k_{app}}{T}\right) = \frac{\Delta S^{\ddagger}_{T_0} - \Delta C_p^{\ddagger}}{R} + \ln\left(\frac{h}{k_B}\right) + \frac{\Delta C_p^{\ddagger} - {\Delta H^{\ddagger}_{T_0}}/{T_0}}{R}\left(\frac{T_0}{T}\right) - \frac{\Delta C_p^{\ddagger}}{R}\ln\left(\frac{T_0}{T}\right) \tag{22}$$

where k_{app} is the apparent rate of $TS^{\ddagger}$ formation, $\Delta S^{\ddagger}_{T0}$ and $\Delta H^{\ddagger}_{T0}$ are the known activation entropy and activation enthalpies at the reference temperature T_0, T is the absolute temperature and $\Delta C_p^{\ddagger}$ is the activation heat capacity, considered to be constant over the temperature range. Being a detailed gauge on the nature of the unfolding barrier these values should only be considered apparent thermodynamic properties of the transition state. On top of that, the front factor of the Eyring equation ($\kappa(k_B * T/h)$) is potentially not very applicable to protein folding and unfolding/folding in a water solution [108, 109], as the water medium affects movements due to viscosity as Kramers accounted for [77], and the high complexity of various interactions between the water solvent and the myriad of chemically diverse surfaces proteins can have. Thus, there is a potential for erroneous determinations of the absolute values of the entropic term and the Gibbs free energy barrier itself. However, using the Eyring front factor is at least reasonable in relative estimations between proteins and protein variants. The caveat is, however, that the front factor is taken to be the same for all water-soluble proteins, which is a statement that is certainly not above scrutiny.

Despite of these limitations and approximations used to estimate kinetic barriers, these analyses have provided interesting insights into kinetic stability of proteins. Cases in point are studies on the kinetically stable proteases, α-lytic protease and SGPB, which have provided interesting insights into the mechanism of their kinetic stability [101]. The results indicated abnormally high values for the activation heat capacity change of unfolding ($\Delta C_p^{\ddagger}$) and a very high activation free energy barrier with its maximum around the optimum growth temperature of the organism of origin [101]. The abnormally high activation heat capacities ($\Delta C_p^{\ddagger}$) indicate large disruptions in the intramolecular network within the protease and the surrounding solvent. This may be linked to the concept of the molten-globule state [110], or a somewhat frustrated and dynamic state with a multitude of unfulfilled contacts. Such a process in an aqueous solution, would make the polypeptide structure become permeable to water molecules as internal connections break down, exposing new surfaces.

This adds to the picture the cost of breaking up the preferential solvation sphere of the native state and solvating the transition state, namely, the effects of solvent barriers on stability. Indeed, solvation/de-solvation barriers seem to impact the energetics of the free energy barrier to a major degree, possibly being a rate limiting step that is responsible for the major part of the activation energy of folding and unfolding, explaining in part the remarkably high values for activation energies acquired from experimental data of many proteins [111-114]. As a

direct consequence, the size of the protein molecule would affect the solvation barrier in addition to the extra surface area accessible to solvent in the $TS^{\ddagger}$ upon unfolding [111]. Activation energies or the enthalpic contribution to the free energy barrier are, however, only half the story. It can be visualized that the entropic part is also quite significant, as displacement of water molecules into crevices due to increased surface area would likely be an entropically unfavorable process due to the restrictions placed on available conformations of water molecules at the protein-water interface. There would be an entropic gain from more allowed protein conformations as a result of looser packing because of loss of internal interactions (Fig. 1.7). These two different variables counteract each other until a temperature is reached where the conformational entropy gain of the peptide chain overcomes the entropic penalty of the water solvation, as the hydrogen bonding network of water becomes more and more transient [59]. Indeed, this is observed for the α-lytic protease and SGPB. At low temperatures, the free energy barrier is completely governed by the activation entropy and the increase in conformational entropy does not compensate until at relatively high temperatures [101]. Due to rather high values for $\Delta C_p^{\ddagger}$ in the case of those kinetically stable protein, and the steep dependence of $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ on temperature, the activation enthalpy favors the $TS^{\ddagger}$ at low temperatures, presumably due to protein-solvent interactions which would also be highly entropically unfavorable. Then, at higher temperatures, the free energy barrier becomes dominated by the activation enthalpy, as at those high temperatures the intramolecular forces are strengthened as dictated by a high positive activation heat capacity ($\Delta C_p^{\ddagger}$) [101]. This could be thought of as being a result of the low prevalence of protein-solvent interactions at high temperatures favoring protein-protein interactions, as can be surmised from the iceberg model of water and how it surrounds the protein at different temperatures [115, 116].

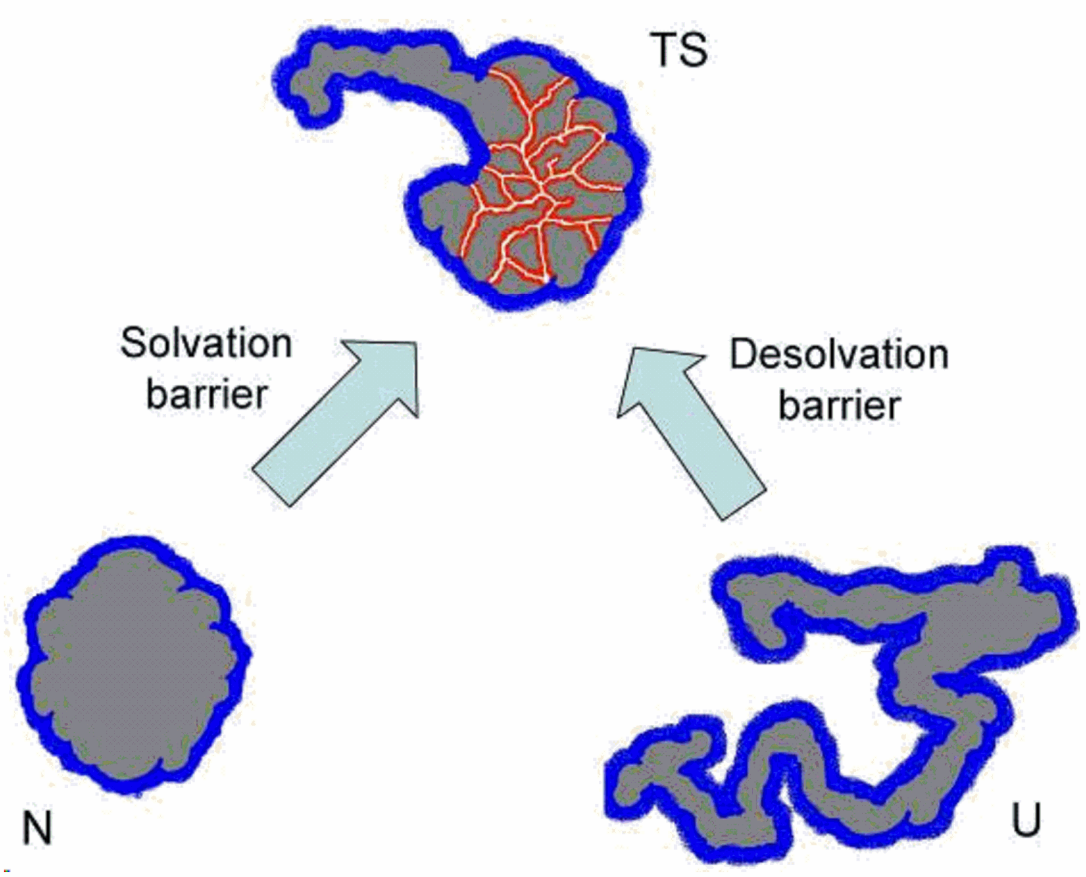


Figure 1.7. Illustration of the solvation shell (blue outlines) for the native state (N), transition state (TS) and the unfolded state (U). The red lines representing lost internal contacts in the TS, not yet solvated. Reprinted from reference [111].

Being such a multifaceted process, the solvation barrier is bound to be affected by many variables and has indeed been found to differ greatly within the same protein family. This has been seen in the family of triose-phosphate isomerases [117], implying sensitivity to other variables than just structure and size. The energy contributors that control the height of the kinetic barrier and the unfolding pathway could, thus, be dictated by small changes in the amino acid sequence at crucial points within the structure [118]. In line with this are observations from molecular dynamics simulations showing that unfolding cooperativity was much higher for the kinetically stable α-lytic protease than for its structural homolog, bovine trypsin, that unfolded in a gradual manner [102]. These simulations indicated that the kinetically stable α-lytic protease has a much narrower unfolding pathway creating a bottleneck of available conformations in the TS‡, i.e. a very steep saddle point in a 3D free energy landscape. Trypsin on the other hand seems to a have a much broader conformational space at the TS‡, allowing for better diffusion back and forth across it (Fig. 1.8). This means

that a well-defined TS‡ i.e. with few allowed microstates at the peak, does not allow for much diffusion back and forth over the TS‡.

The conformational space of the unfolded peptide chain is enormous compared to the conformational space at the TS‡. The TS‡ would thus be inaccessible from the unfolded assembly, largely due to the difference in conformational entropy between the states.

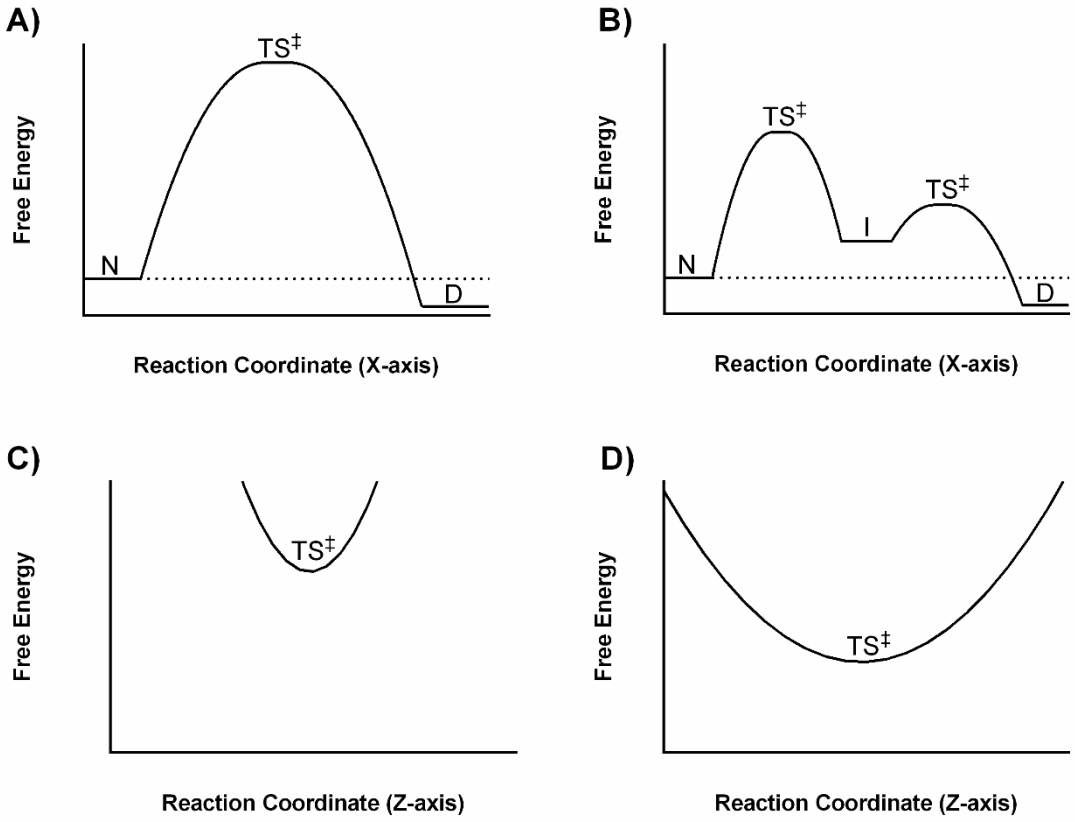


Figure 1.8. A) A theoretical energy landscape of an unfolding progress for a protein that unfolds in a cooperative manner, resulting in a high free energy barrier, along the native (N) to denatured (D) axis (X-axis). B) A theoretical energy landscape of an unfolding progress for the same protein that does not unfold in a cooperative manner due to an intermediate formation (I) , resulting in lower individual free energy barriers, along the native (N) to denatured (D) axis (X-axis). C) A theoretical view along the Z-axis of figure A) demonstrating a narrow conformational space at the transition state saddle point. D) A theoretical view along the Z-axis of figure B) demonstrating a broad conformational space at the transition state saddle point.

Bringing these aspects together, the basic part of the kinetic barrier seems be rooted in the cooperativity of unfolding. The highest free enthalpy change would certainly be achieved by the instantaneous collapse of the native state, thus breaking the highest number of bonds at once. In addition, this would result in the near simultaneous solvation of a maximal area of the newly exposed surface providing a high entropic solvation penalty at lower temperatures, contributing to the free energy barrier. In the case where there is low cooperativity, the unfolding would be gradual, allowing for many more conformations and possibly different

routes to a TS‡ containing many more conformations. In addition, the formation of metastable unfolding intermediates along the unfolding pathway as a result of loss of cooperativity could yield transitions states with lower individual peaks. The analogy being that a very cooperative unfolding process would resemble an un-catalyzed reaction and a non-cooperative process could resemble a catalyzed reaction (Fig. 1.8), if the systems are presumed to have the same energetics otherwise.

The description above would fit the observation of an extracellular kinetically stable protease with restricted conformational dynamics, that unfolds in an extremely cooperative manner [100, 119]. In the case of proteases, suppressing any native conformations, or unfolding intermediates, that could expose a potential proteolytic cleavage site, would also be beneficial for the overall fitness of the protein. Thus, a cooperative unfolding pathway and kinetic stability would be under even more evolutionary pressure [4, 120]. So, for a kinetically stable protein that must adapt to high temperatures, a highly connected system of internal protein-protein interactions, allows for the absorption of the highest amount of energy while maintaining activity before complete unfolding occurs. So, the ability of the interaction network to distribute the energy about its system and release it as molecular motions without the loss of activity has been correlated with kinetic stability and called rather aptly the thermal flexibility of the protein [121].

Overall, kinetic stability seems to be Nature´s answer to the most hostile environments and even protection against detrimental events such as protein aggregation and thus many diseases caused by protein aggregation [122]. So, learning from the highly kinetically stable molecules is certainly warranted, as learning from the extremes can often help to eluate the small nuances of other systems.

1.3 Temperature adaptation strategies

The molecular interactions and energetics that determine protein stability are highly diverse. The structures that can be devised from the twenty proteogenic amino acids are truly limitless, although only a very small fraction results in polypeptides that fold spontaneously into one stable three-dimensional from. In addition, the various post translational modifications that some structures are subjected to, give rise to structurally similar molecules

with varying stabilities both kinetic and thermodynamic. Thus, it is of no surprise that a myriad of different strategies are utilized by nature in order to obtain the desired functionalities and stabilities. Here, the trends seen in nature will be discussed, by comparing the differences in structures and activities observed from enzymes operating at the two temperature extremes of living organisms.

1.3.1 Thermostable proteins

As touched upon earlier, at high temperature, the main challenge of proteins is maintaining an active state for a biological relevant time. Adapting to high temperature by increased stability is obtainable through thermodynamic routes and/or kinetic routes. The thermodynamic approach would be to stabilize the native state i.e. lowering the free energy of the native state, or by destabilizing the denatured, i.e. by increasing its free energy, or by a combination of the two approaches. Examples of how varied the free energy stability curves for a protein evolving from mesophilic to thermophilic stability characteristics might look like by such manipulations is depicted in Fig. 1.9 [123]. Those free energy curves are all the product of the delicate balance between the thermodynamic factors and the enthalpy/entropy tradeoffs that are expected in a protein-water system [29]. The kinetic route would then be increasing the energy barriers separating the states and/or restricting available routes traversing the energy barrier. These different strategies can be traced to the attributes of the primary, secondary, tertiary and the quaternary structure of proteins and their interactions with their environment. The overall stability is eventually dependent on both thermodynamic and kinetics contributions, giving either ordered or disordered proteins.

Several trends have provided insights into the molecular determinants of thermostability that have come to light when comparing the genomes of organisms that have adapted to high temperatures versus those adapted to lower temperatures and differences between structural homologues [124]. However, there is no single mechanism that can be said to be responsible for ensuring folding into a stable structure, as different protein families adopt different strategies to maintain their active form at high temperatures. The structural factors that are however most frequently observed include, a more tightly packed hydrophobic protein core, a thicker hydration layer, a higher abundance of water-protein H-bonds potentials, more polar and charged surface residues, critically located disulfide bridges, increased number of salt bridges, higher number of π-charge interactions, more abundant π-π interactions and

more rigid structures that often have shorter loops and thus higher ratios of secondary structure elements [123, 125, 126]. Each of these interactions have enthalpic and an entropic contributions making their effectiveness dependent on temperature, thus helping in identifying their role and their effectiveness in each case [127].

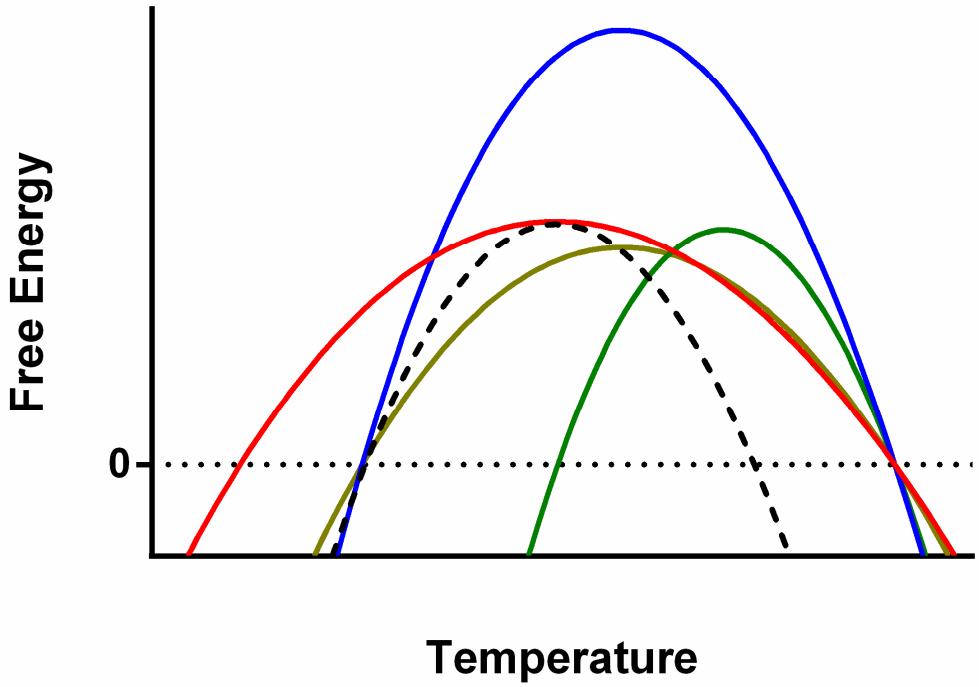


Figure 1.9. Comparison of free energy stability diagrams of theoretical protein from a mesophilic organism (dashed line) and four different theoretical proteins from a thermophilic organism (red, blue, green and gold lines) that have reached a high melting point via different strategies but all with the same melting point. The dotted line representing the zero point, below which the equilibrium favors the unfolded state. Inspired by reference [123].

The hydrophobic effect, a highly complex phenomenon, is the strongest for biological systems between the temperatures of approximately 30°C and 80°C, although it differs on more parameters than just temperature, such as the surface area and the volume of the hydrophobic solute [128, 129]. Thus, the burial of hydrophobic residues is the most beneficial for proteins from mesophilic and thermophilic organisms as a free-energy contributor for folding. This trend of having a more efficient packing of their hydrophobic core residues and fewer water accessible cavities have been very prominent structural features of thermophilic proteins when compared to cold active proteins [25]. Comparison of thermostable proteins to their counterparts from hyperthermophiles, the latter also tend to have tightly packed protein cores. However, the internal contacts between residues in hyperthermostable proteins are

more likely to form bulkier hydrophobic clusters than their mesophilic and thermophilic counterparts, indicating a growing role of internal protein contacts at higher temperatures [125, 130].

Thermostable proteins are frequently observed to have surfaces rich in polar residues and charged residues not partaking in ion-ion interactions, resulting in an increased number of potential solvent-protein H-bonds. MD-simulations have indicated that those residues allow for the formation of a solvent shell at high temperatures [131]. This could be mediated by solvation of lone charges, forming a water anchor of sorts, maybe partly explaining why arginine is so frequent in thermostable proteins [132], as the guanidino group of the arginine side chain can provide a maximum potential for formation of H-bonds. This would provide thermostable proteins with considerable solvation barriers at higher temperatures and prevent water molecules from penetrating the structure. These surface charges would also ensure that the native protein could function at higher temperatures by retaining water molecules that are essential for protein-protein and protein-ligand interactions [125, 133].

Disulfide bridges serve a stabilization role as they can reduce the entropy of the unfolded state by interlocking distant parts in primary sequence close in 3D space and helping to maintain correct orientations between parts of the protein as molecular movements increase as a function of higher temperatures. The location itself being key as a disulfide bridge between rigid areas would be expected to have less of an impact than between flexible parts in the structure [125, 134]. The sulfur-sulfur covalent bond is, however, susceptible to decomposition at higher temperatures and even more so in alkaline environments. This fact has brought into question how effective they are for proteins that have to retain their structure at temperatures around and over 100°C [124]. However, their location in the native and unfolded states, as well as all conformations along the unfolding pathway, will determine how susceptible these disulfide bridges are to cleavage. In certain cases, disulfide bridges have been found to infer stability to hyperthermostable proteins. An example is a hyperthermostable ferredoxin variant without a native disulfide bridge which had a melting point of 113°C, as compared to 121°C for the disulfide containing wild type when measured at neutral pH [135].

Hyperthermostable proteins that have to function at temperatures well above the boiling point of water also require more strategies to maintain their structure [126]. Ion-ion interactions

or salt bridges being a prime example of this. Salt bridges are highly dependent on many environmental factors such as pH, location and orientation in the protein, salt concentration and temperature. Salt ions can screen surface charges that are dependent on the pH of the aqueous solvent as it has to be in the appropriate range so that the acidic side chains of aspartic and glutamic acid residues and those of the basic groups of arginine, lysine and histidine, to maintain their charges [136]. Water plays another big role, being a very polar solvent, such that charges are readily hydrated by water. This entails, however, that the desolvation penalty of salt bridge formation is rather high, especially at low temperatures, because of the high dielectric constant of water. The dielectric constant of water gets lower as a function of temperature between 0°C and 100°C [137]. Thus, according to Coulomb's law, salt bridges are stronger at higher temperatures, due to lesser screening of water and are also dependent on the proximity of the interacting charges:

$$U_{1-2} = k\frac{q_1 q_2}{\varepsilon r_{1-2}} \tag{23}$$

where $U_{1\text{-}2}$ is there electrostatic potential between the charges q_1 and q_2 that are separated by the distance $r_{1\text{-}2}$ and k is the conversion factor to the appropriate units and ε is the dielectric constant.

The desolvation penalty of the charges also decreases as temperature rises, likely due to the change in the dielectric constant of water with increasing temperature and possibly due to partial solvation at higher temperatures [125, 138-140]. This property may have some interesting implications for protein stability, as the salt bridges could lead to lowering of the free energy change at low temperatures, but increase it at higher temperatures. This could be achieved by broadening of the free energy stability curve, that could entail among other things, a lower ΔC_p of unfolding (Fig. 1.9. the cold curve). Observations of lower ΔC_p of unfolding due to the introduction of salt-bridges have been made in the case of the thermostable ribosomal protein L30e from *Thermococcus celer* [141].

Hyperthermostable proteins are observed to have large surface networks of salt bridges. Interaction networks being one of the basis of a cooperative structure. A popular protein to point this out in, is the glutamate dehydrogenase found in the hyperthermophile *Pyrococcus furiosus*, that has a salt bridge network of 18 charged residues [142]. The strongest evidence for the importance of the salt bridge networks to gain the thermotolerance for maintaining

an active structure at temperatures rising above 100°C is the prevalence of these networks in proteins from hyperthermophilic organisms that are not found as frequently in counterparts from thermophilic, mesophilic and psychrophilic organisms [143].

Charge-Charge interactions and single charges are not the only charge interactions that are observed in higher frequency in proteins from thermophiles and hyperthermophiles, as cation-π interactions are also found more frequently in these proteins [144]. Their effectiveness in increasing thermal stability has been found to be highly dependent on temperature, in general being stabilizing at high temperatures, but at low temperatures these interactions can be destabilizing [36]. This behavior is possibly related to desolvation of the charge, resulting in a shifted free energy curve towards higher temperatures, or lower heat capacity leading to a broader free energy curve. Aromatic interactions are not exclusively found as part of charge-π interactions but also found in aromatic clusters on rigid surface patches that are found in greater number in thermophilic proteins [145].

It seems that interaction clusters of various kinds are found in more abundance in proteins as a function of increased temperature of adaptation. These interaction networks can be vast, connecting distant parts of the protein structure together (Fig. 1.10) [125, 146]. A high degree of interconnectedness within a structure is presumably the basis of a cooperative structure. Indicating that interaction networks are important for high free energy barriers between the folded and unfolded states.

As a result of these differences in the structural elements of thermostable proteins, they exhibit structures that are more rigid or stiff as compared to their cold adapted homologs when examined at the same temperature. However, the molecular motions that are needed to facilitate activity are presumably similar in both thermostable and cold active protein structures. Thus, molecular motions at temperatures that are found in the natural environments of these proteins should be similar, as stated by the corresponding state hypothesis [147, 148].

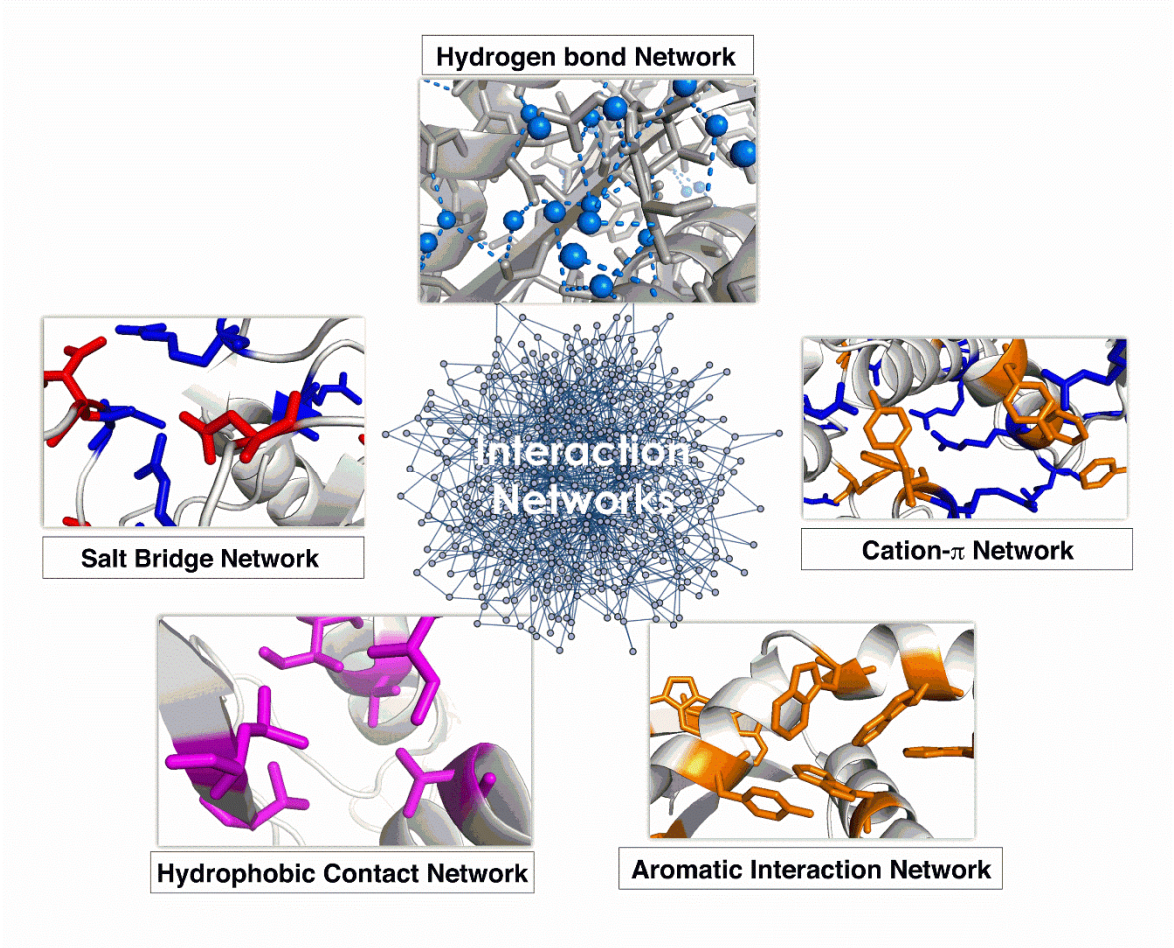


Figure 1.10. Examples of interactions that can form networks within the protein structure, contributing to the overall stability and cooperativity. Reprinted from reference [125].

The proline residue

A particular interest in the work underlying this thesis is the role of the proline residue [4] with regards to flexibility and rigidity, focusing on their effect on protein stability when located in loops. The proline residue differentiates from all the other natural amino acids found in proteins in that it possesses a secondary amine i.e. the α-carbon is linked to the amino group via its side chain forming a pyrrolidine ring. This unique structural property is the basis of the conformational restricting properties of the residue [149]. The structure of the residue also promotes a relatively slow cis-trans isomerization of the peptide bond, with a rate constant that has been found to be in ms^{-1} range at ambient temperatures, favoring the trans isomer in most oligopeptides and being a key part of the rate-determining step of protein folding for many systems (Fig. 1.11) [150, 151].

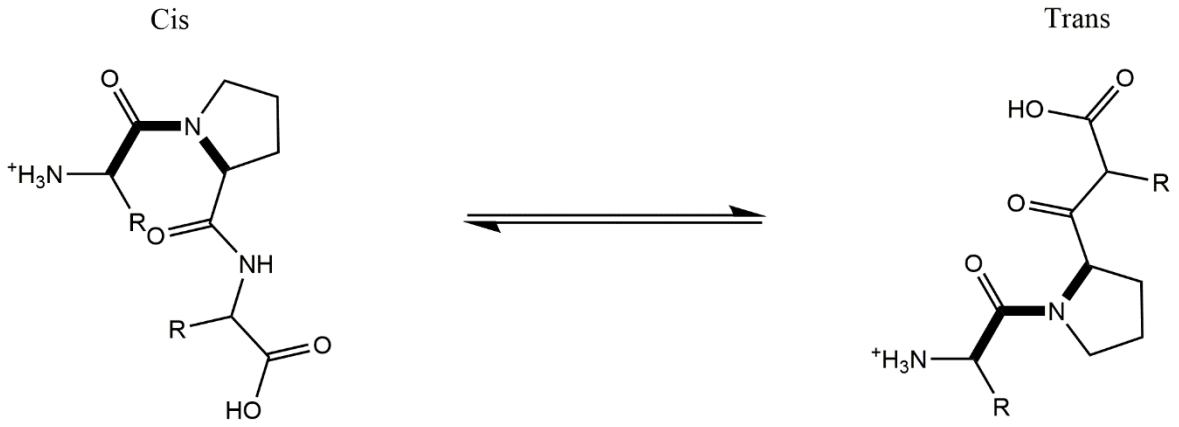


Figure 1.11. The cis and trans isomers of the X-Pro peptide bond of the tripeptide X-Pro-X. The bold bonds are for clarity to emphasize the orientation of the α-carbons on either side of the peptide bond.

Proline residues do seem to be more prevalent in thermostable proteins than in their cold active counterparts and proteomic analysis of cryophiles have suggested that proline residues occur less frequently in their proteomes [132, 152-155]. One mechanistic way the proline residue can infer thermostability is by reducing the entropy of the denatured state as the proline residue is uniquely suited to restrict available conformations of the unfolded peptide [156]. However, due to the secondary amide of the proline residue, the amide cannot partake in H-bond formation thus, location of the residue would be of utmost importance as unfulfilled H-bonds within the protein core and be detrimental for the overall stability of α-helices and β-sheets [157, 158]. Indeed, proline residues seem to be more prevalent on the N-terminus of α-helices and in loops of proteins, providing some structural rigidity and dictating direction of the main-chain [132, 159, 160]. Proline residues in loops can also increase kinetic stability [4, 161], possibly via anchoring important structural parts together, thus allowing for more heat to be applied to the system while maintaining correct interactions, and thereby increasing the thermal flexibility of the protein [4, 121].

1.3.2 Cold active proteins

Low temperatures bring a whole new set of evolutionary challenges to organisms. Focusing on enzymes, the main problem is that their activity must pair with the overall reduction in rates of chemical reactions. As a rule of thumb, it is often said that for every 10°C increase in temperature, the reaction rates double. So, in order to maintain biological relevant reactions rates at 5°C compared to 85°C the increased catalytic activity required would be enormous. It has thus been of much interest to decipher what structural aspects contribute to increasing the activity of an enzyme. This topic has been debated for decades, as linking

certain structural elements and/or their movements to faster reaction rates is no easy task. However, the increased catalytic rates at low temperatures have most often been attributed to different dynamics of the cold active structures [162-164]. Observations have been made that cold active proteins have more flexible and dynamic structures, and are more unstable than their homologs adapted to higher temperatures. This view has been supported by reports of larger mobile loops, larger distorted protein parts and even a higher occurrence of methionine, a residue that has high conformational entropy, as general characteristics of cold adapted proteins [25, 165-169]. Cold adapted proteins also tend to contain water accessible cavities in their structures as well as more hydrophobic residues on the accessible surfaces [125, 170, 171]. These attributes have been linked to the destabilization of the solvent shell and weakening of internal contacts in order to facilitate movements resulting in increased flexibility. The ability to catalyze reactions at low temperatures is often directly attributed to this increased flexibility. The simplest picture that emerges is the weakening of protein-protein and protein-solvent interactions would allow for more flexibility that would directly cause an increased number of native conformations with lower energy barriers separating them. This would tend to create a "lubricated" structure which would facilitate the transversion of structural states at lower temperatures, thus an easier transversion into conformations necessary for catalysis. The native assembly has lower enthalpy as a result of having fewer electrostatic interactions and a structure that has more water accessible cavities reducing intra-protein contacts. This reduction of enthalpic interactions leads to a native assembly of highly flexible structures manifested in a larger conformational space of the native assembly (Fig. 1.12) [23].

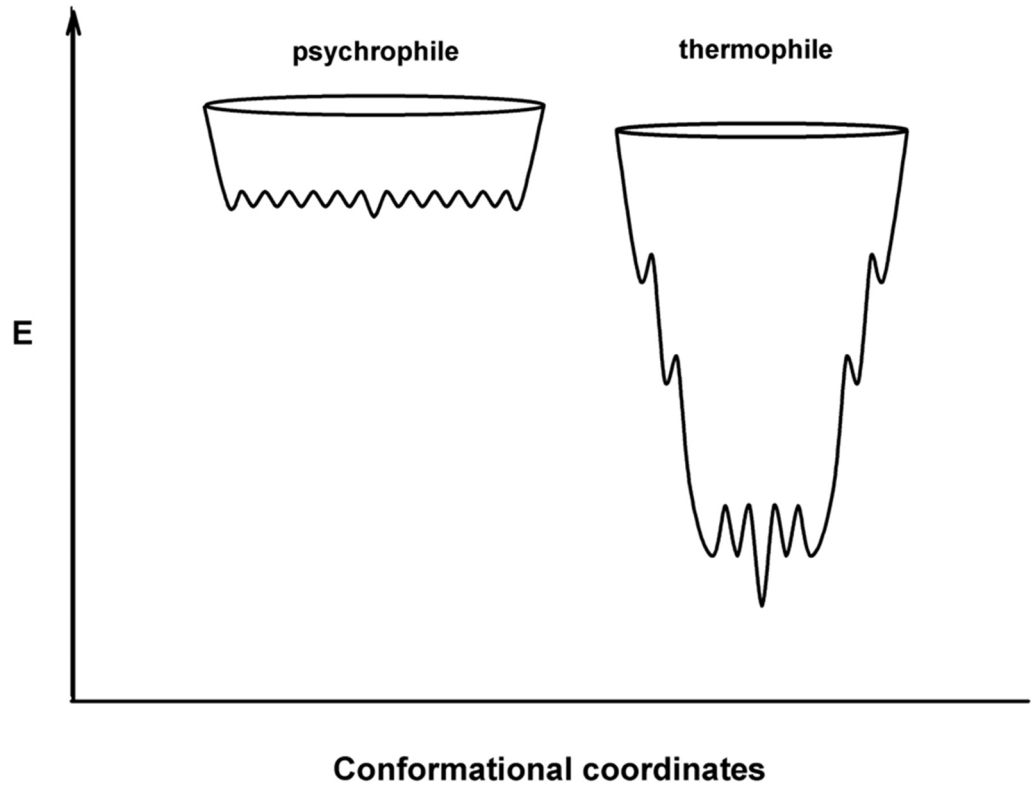


Figure 1.12. Hypothetical one-dimensional energy landscapes of a protein form a psychrophile and a thermophile, demonstrating the differences in conformational diversity as a function of E (free energy). Reprinted from reference [23].

The transversion of such broad conformational space into the presumably restricted conformational space needed for catalysis would, thus, be very entropically unfavorable. That entropic penalty would dampen the effects that fewer/weaker enthalpic interactions could have on reaction rates, this being a prime example of enthalpy-entropy compensation [23, 172].

The observed increased flexibility does not, however, seem to present itself uniformly in cold active structures. Increased activity is often linked to local flexibility rather than global flexibility as certain parts in the structure, that are directly involved in catalysis, require high flexibility and others might still have an evolutionary pressure to stay rigid to maintain a certain degree of stability [173-176]. Thus, conformational flexibility at crucial parts is a key component for activity at lower temperatures and the obvious question would be how and where increased flexibility does facilitate faster reactions rates. To address this question, a simple enzyme catalyzed reaction is described as:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} ES \underset{k_{-2}}{\overset{k_2}{\rightleftarrows}} TS^{\ddagger} \overset{k_3}{\rightarrow} E + P \qquad (24)$$

is considered, where E is the enzyme, S is the substrate, ES is a loosely bound enzyme-substrate complex, TS‡ is the ES transition state complex and P is the product (Fig. 1.13). Each step is controlled by a forward and backwards rate constant. For simplification, the last step is considered irreversible and different energy landscapes due to different microstates are disregarded [177].

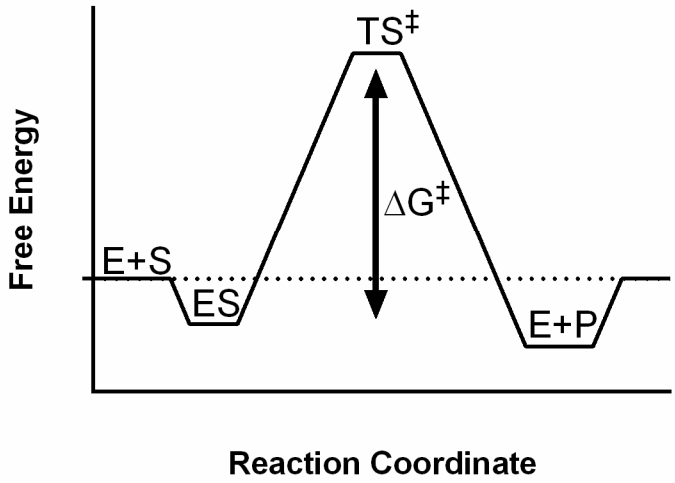


Figure 1.13. Reaction diagram of a simple reaction pathway, depicting the creation of the loosely bound enzyme-substrate complex (ES), its activation to the transition state (TS‡) and product release.

A classical view of enzyme catalysis is where the ES complex is lower in free energy than the free enzyme. This complex can dissociate back, or continue towards the product, the rate of association and dissociation of that complex being described by the K_M constant in classical Michaels-Menten kinetics. Then, the binding of the complex is maximized to the transition state that can involve the formation and breaking up of a myriad of weak and covalent interactions. All these events rely on correct conformations of both the enzyme and the substrate which is all accompanied by solvation and de-solvation of various elements. This tight binding, guides the substrate into a highly reactive conformation and/or creates a highly reactive intermediate chemical species leading to lower the reaction enthalpy [178, 179]. The height of this energy barrier is then directly reflected in the turnover number of the enzyme k_{cat}, the maximum number of chemical conversions of the substrate an enzyme molecule can carry out over a defined unit of time.

Cold active enzymes are often found to have higher k_{cat} values and higher K_M values than their thermostable homologues, meaning that substrate binding affinity is sacrificed while increasing turnover [24, 25, 175, 180-182]. A reasonable interpretation of those observations is that high flexibility causes destabilization of the ES complex as interactions between the enzyme

and the substrate are fewer and possibility weaker, resulting in a lower activation free energy barrier from the ES complex to the transition state (Fig. 1.13) [172]. This idea of a flexible active site has also been supported by additional observations, such as those of a cold adapted elastase from North Atlantic salmon and a cold adapted alcohol dehydrogenase from the Antarctic psychrophile *Moraxella* sp. TAE123, which exhibited loss of substrate specificity [183, 184]. This caused both enzymes to be able to catalyze bulky substrates, presumably due to a more flexible or more open active site than related enzymes.

However, this effect of low temperatures raising K_M values is not universal. There are examples of cold active enzymes that do not share the high K_M - high k_{cat} trends, but rather exhibit high k_{cat} values and low K_M values, meaning that they have high affinities and low transition barriers [1, 185-187]. This demonstrates that cold activity is obtainable through more strategies than just weakening of the ES-complex. The reason for this discrepancy is most likely due to an additional evolutionary pressure, low substrate concentrations in the natural environment being a likely candidate. In line with this view are observations that indicate that this trend is more likely to occur in secreted marine enzymes [25]. This property could be ascribed to beneficial electrostatic interactions and/or a rigid active site, allowing for the formation of more enthalpic interactions [25, 188]. The idea of a rigid active site with an otherwise flexible structure is supported by MD-simulations on citrate synthases and trypsin homologues from organisms adapted to different temperatures. Those simulations did not show change in flexibilities in the active sites of these cold adapted enzymes. Rather the surface area surrounding the active site was more flexible, while the protein core was still relatively rigid, describing those cold active systems as having soft surfaces compared to their thermostable homologues [189]. Thus, it cannot be said that there is any consensus on where increased flexibility is needed in order to facilitate biological relevant rates at low temperatures. Thus, apparently the topic of local flexibilities must be viewed on a case to case basis.

Monitoring protein flexibilities

Big strides have been made during the last decades in methods to monitor protein motions and locating flexible parts within a protein structure. Protein motions occur on a very wide timescale, from very local motion of bond vibrations on the femtosecond scale to large collective motions of large protein systems that can take milliseconds or longer. Smaller motions, such as side-chain rotations or loop fluctuations, are, however, often in the pico-,

nano- or microsecond ranges [190]. Estimations of flexible areas within protein structures can be made directly from protein structures determined by X-ray or neutron scattering crystallography. This is done by utilizing B-factors (Debye-Waller factor) that describes the attenuation of X-ray or neutron scattering due to thermal motions within the crystal [191]. Thus, B-factors reflect on parts of the protein that are prone to thermal motions. Deuterium exchange monitored with NMR or mass spectrometry can inform on larger motions in solution or unfolding events with timescales of milliseconds or longer [192]. Another method to monitor flexibility in solution is fluorescence quenching that can give information about the accessibility of a quencher to a fluorophore. Depending on the context, this method can be utilized to assess flexibility [193]. These methods give us a general idea of the flexible areas within protein structures. To ascertain the kinetics of these motions, time-resolved techniques need to be utilized. Time-resolved fluorescence spectroscopy, utilizing pulsed light sources to monitor fluorescence decay, can give site specific information. This is often done while monitoring fluorescence anisotropy and/or Förster resonance energy transfer (FRET), when FRET is utilized to measure the distances between groups in moving protein parts [194]. Single molecule FRET has proven to be a powerful tool to unravel detailed molecular mechanisms, as has been demonstrated for F_0F_1-ATP synthase [195]. With increasing power of NMR instruments and improved analytical techniques, more detailed information of specific protein dynamics can be determined from relaxation rates [196]. Time-resolved X-ray spectroscopy has also been utilized to characterize protein motions on the microsecond timescale, shown to be important for catalysis of cyclophilin A [197]. By using temperature jumps, two distinct relaxation processes were resolved, a fast process related to surface loop motions and a slower process related to motions within the protein core [197]. MD-simulations are also an invaluable tool in unraveling the molecular dynamics and specific interactions within the protein contributing to their activities. With increasing computational power and more refined models and algorithms, longer and more detailed simulations can be performed. Together, *in vitro* and *in silico* methods can provide a comprehensive toolset to decipher the molecular interactions and fluctuations necessary for their functions [198].

Macromolecular rate theory (MMRT)

So far, the temperatures of maximum activity (T_{opt}) have not been discussed, but it is a universal behavior of enzymes that their catalytic rates increase up to a point and start falling thereafter. This behavior is most easily attributed to simultaneous unfolding during measurements. However, as discussed previously, the native assembly, the ES-complex and particularly the transition state complex, are distinct thermodynamic states. The activation entropy and activation enthalpy of catalysis are also affected by temperatures as dictated by their activation heat capacities ($\Delta C_p^{\ddagger}$), controlling the temperature dependence of enzyme catalyzed rates. The temperature dependence of enzymatic reaction rates has been addressed by the so-called macromolecular rate theory (MMRT) [199]. Central to the MMRT are observations from crystal structures and MD simulations that $\Delta C_p^{\ddagger}$ for the transition state formation is a negative value (Fig. 1.14) [200].

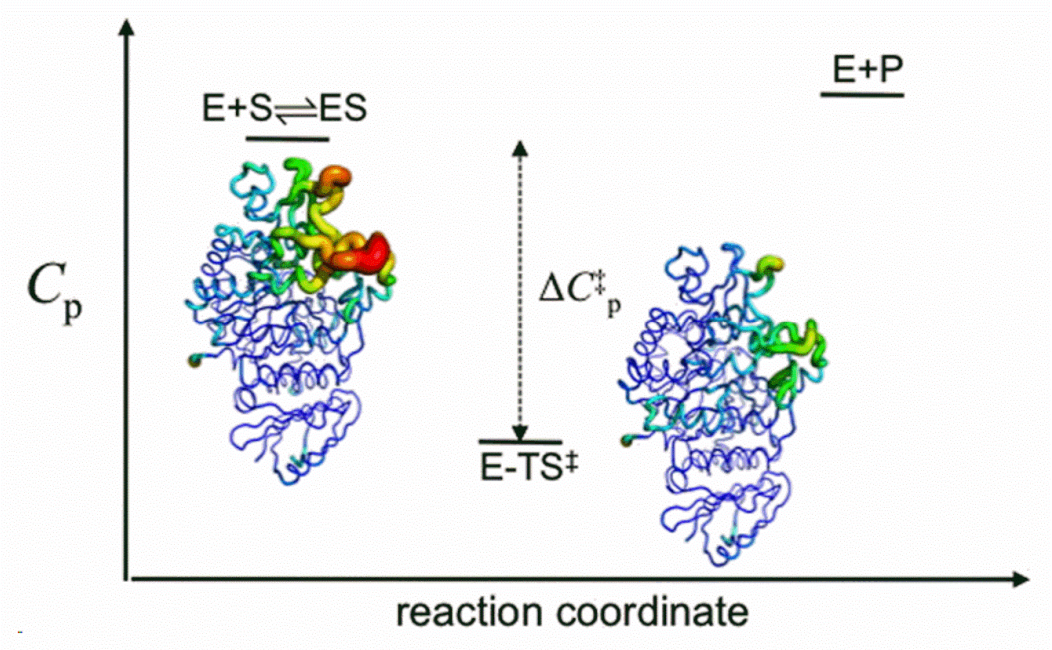


Figure 1.14. Theoretical heat capacity diagram of an enzyme catalyzed reaction, demonstrating the reduction of available conformations during the activated transitions state resulting in a lower heat capacity of the transition state. Reprinted from reference [199].

A negative value for $\Delta C_p^{\ddagger}$ leads to a semi bell shaped reaction rate profile (Fig. 1.15). Since the negative value for $\Delta C_p^{\ddagger}$ can be quite large, the activation enthalpy and entropy of the reactions change steeply with temperature. Thus, assigning a major part of the free energy barrier in a catalyzed reaction to an enthalpic or entropic term does not provide deep insights about the reaction, as a system with a negative $\Delta C_p^{\ddagger}$ has free energy barrier that is predominantly enthalpic in nature at low temperatures and entropic at high temperatures. T_{opt} is where $\Delta H^{\ddagger}$ approaches zero.

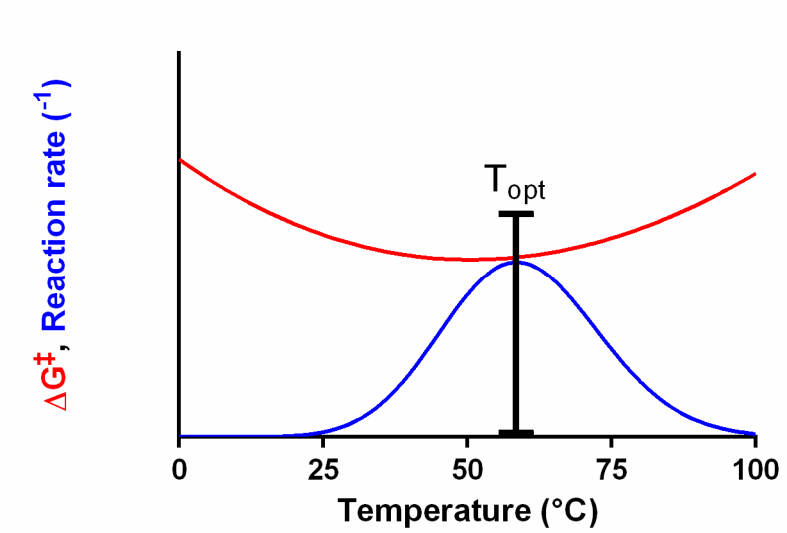


Figure 1.15. The effect of temperature on the activation free energy (red line), where the activation heat capacity change from a loosely bound ES-complex to the transition state is negative value and the resulting reaction rates (blue line).

MMRT thus fit the idea of an enzyme that binds loosely to substrate and then maximizes binding to the transition state via funneling energy to the active site through translational, rotational and vibrational energies. The source of this internal free energy will be the sum of every protein-water, protein-ligand, external protein-protein and internal protein-protein interactions. This could in part explain why larger catalytic domains often exhibit better rate enhancements when comparing catalyzed rates versus uncatalyzed reaction rates [199]. In addition, this also indicates that multimeric enzymes might also be beneficial to reactions rates, as larger systems can funnel more energy. This means funneling of vibrational, rotational and translational modes to the active site reduces available interactions in the structure causing the system to less efficiently distribute energy, thus lowering the heat capacity of the "rigid" transition state. Thus, a flexible cold active enzyme could generally be expected to have a more negative $\Delta C_p^{\ddagger}$ if the role of the flexible structure is to funnel more energy into the active site to compensate for low thermal energy in the system. In its most simple form, this view would be consistent with the activity/stability trade-off, if this increased flexibility is due to the lowering of enthalpic interactions. However, there are certainly some evolutionary routes that can lead to $\Delta C_p^{\ddagger}$ of transition state formation that approach zero, as seen in adenylate kinases that have evolved from hot to colder environments and then to hot environments again, that exhibit higher activities at lower temperatures than those that have been restricted to hot environments [26]. This would reflect a near zero $\Delta C_p^{\ddagger}$ of the transition state formation [26].

This idea of a non-zero $\Delta C_p^{\ddagger}$ in enzyme catalysis is, however, debated. A research using MD-simulations of reaction rates and conformational populations during catalysis, utilizing a model system of a psychrophilic and a mesophilic α-amylase, resulted in a dead-end steady-state model (substrate inhibition model) to explain the observed T_{opt} [201]. The cold active α-amylase has T_{opt} for activity at 25°C, that is well below the unfolding of the enzyme. So, activity loss cannot be explained by concurrent unfolding of the cold active α-amylase. The reaction rates could be fitted with very good agreement with both MMRT and the dead-end model to explain this loss of activity above 25°C. However, in support of the dead-end model was that populations of substrate-inhibited complexes did appear in higher amounts at higher temperatures during simulations. This "dead-end" side pathway was used to purely explain away this so-called anomalous temperature optimum [201].

Deciphering what interactions and molecular movements govern the temperature dependence of the catalytic reaction rates still leaves a lot of questions open. Considering the myriad of different chemical reactions different enzymes catalyze, the answer needs to be based on the individual enzyme-substrate system under study, as changes in substrate and other environmental conditions can affect the apparent rate determining step, hence enzyme catalysis not being generally as simple as stated in eq. 24. Thus, the answer in each case must be a combination of determinants such as changes in $\Delta C_p^{\ddagger}$ between relevant states in each case and the different temperature dependencies of rates between relevant states, and to complicate this even further, the temperature dependence of $\Delta C_p^{\ddagger}$ for the various reactions.

1.4 Proteases

This chapter will serve as a brief introduction into the world of proteases, outlining their properties and role in nature. This chapter will also serve to introduce the research model that involved comparison of the properties of the thermophilic subtilase from *Thermus aquaticus* (AQUI) and the cold active subtilase-like protease from a marine *Vibrio* sp (VPR).

1.4.1 Classification of proteases

Proteases belong to the class of hydrolases (EC 3) [202] and serve the role of protein catabolism in addition to being responsible for the regulation of multiple biological processes. The

hydrolysis of peptide bonds is one of the critical reactions carried out in every organism and must have been an important function for the metabolism of early life, providing a source of energy and building blocks for proteogenic processes [203, 204]. Point in case, proteases are found in all kingdoms of life and many viruses use them to sculpt the proteomes of their targets often in order to hijack their functions [205].

Proteases are classified as either endo-proteases or exo-proteases. Exo-proteases either cleave their substrate at the amino- or carboxy-termini, whereas endo-proteases cleave peptide bonds between residues somewhere within the peptide substrate. [206]. Proteases do not just act as aggressive tools of peptide degradation, as their functions are highly diverse, from rather non-specific enzymes like proteinase K, to highly specific proteases that have specificity towards cleaving only one type of bond in a specific protein, e.g. angiotensin-converting enzyme [203, 207]. The importance of this class of enzymes is reflected in their abundance in various organisms, typically accounting for 2 – 4 % of encoded genes and, for example, it is estimated that around 2 % of the human genome codes for proteases [207-209]. As a result, several diseases are caused by dysregulation of proteases and they have been of interest to the pharmaceutical industry as targets, diagnostic biomarkers, or even as therapies [210, 211]. In addition, proteases are important in other industries, chiefly in the food and detergent industries [212, 213].

The MEROPS database [208] classifies proteases by catalytic type. To date, there are nine different catalytic types listed in the database, those are: Aspartic (A), Cysteine (C), Glutamic (G), Metallo (M) Asparagine (N), Mixed (P), Serine (S), Threonine (T) and Unknown (U). These types of proteases are then classified into families based on amino acid sequence identity of the catalytic unit denoted by a number, or into clans based on evidence of common ancestry assigned by another letter [214]. As of this writing, there are 273 families listed, and 64 clans. Of all sequences in the database, serine and metallo proteases are the most abundant, accounting for 37 % and 36 %, respectively [208].

1.4.2 Serine proteases

Serine proteases are identified by the active serine residue responsible for the nucleophilic attack on the carbonyl carbon of the peptide bond with the formation of a covalent intermediate that eventually is hydrolyzed from the enzyme. As of this writing, there are 13 different clans of serine proteases listed in the MEROPS database that can be divided into

38 families that belong purely to serine proteases and 3 clans belonging to the mixed category, containing 16 families of serine proteases (Table 1-1) [208]. In the field of biochemistry, serine proteases have the historical value of being among the first enzymes that were studied in great detail [215]. The interest in serine proteases is an interdisciplinary one due to their abundance in genomes of all living organisms and their roles broad functions, from digestion to cell differentiation [216-218].

The widespread distribution of serine proteases demonstrates the effectiveness of this type of proteases in nature. This is also reflected in the fact that at least four of the clans SB, SC, SH and PA have evolved to have the same active site, the so-called catalytic triad, consisting of serine, histidine and aspartic acid residues. In these four distantly related clans, the residues take up the same geometry of these catalytic residues (Fig. 1.16) and have an identical reaction mechanism (Fig. 1.17), indicating that this motif has evolved on four different occasions [219, 220]. In addition, some members of the SK clan have the same catalytic triad but in a novel configuration [220]. The other clans all seem to utilize some sorts of catalytic triads, dyads, or an environment that enables a chemical moiety to act as a general base to accept the proton form the active Ser residue [220, 221].

Table 1-1. Classification of the serine proteases, their clans and families along with a model enzyme for each family. Adapted from the MEROPS database (As of April 2020).

CLAN	FAMILY	MODEL PROTEASE	STRUCTURE
PA	S1	Chymotrypsin A	Yes
	S3	Togavirin	Yes
	S6	IgA1-Specific Serine Peptidase	Yes
	S7	Flavivirin	Yes
	S29	Hepacivirin	Yes
	S30	Potyvirus P1 Peptidase	-
	S31	Pestivirus NS3 Polyprotein Peptidase	Yes
	S32	Equine Arteritis Virus Serine Peptidase	Yes
	S39	Sobemovirus Peptidase	Yes
	S46	Dipeptidyl Peptidase 7	Yes
	S55	SpoIVB Peptidase	-
	S64	Ssy5 Peptidase	-
	S65	Picornain-like Serine Peptidase	-
	S75	White Bream Virus Serine Peptidase	-
PB	S45	Penicillin G Acylase Precursor	Yes
PC	S51	Dipeptidase E	Yes
SB	S8	Subtilisin Carlsberg	Yes
	S53	Sedolisin	Yes
SC	S9	Prolyl Oligo-peptidase	Yes
	S10	Carboxypeptidase Y	Yes
	S15	Xaa-Pro Dipeptidyl-peptidase	Yes
	S28	Lysosomal Pro-Xaa Carboxy-peptidase	Yes
	S33	Prolyl Amino peptidase	Yes
	S37	PS-10 Peptidase	-
	S82	Autocrine Proliferation Repressor Protein A	-
SE	S11	D-Ala-D-Ala Carboxy-peptidase A	Yes
	S12	D-Ala-D-Ala Carboxy-peptidase B	Yes
	S13	D-Ala-D-Ala Peptidase C	Yes
SF	S24	Repressor LexA	Yes
	S26	Signal Peptidase I	Yes
SH	S21	Cytomegalovirus Assemblin	Yes
	S73	gpO Peptidase	-
	S77	Prohead Peptidase gp21	Yes
	S78	Prohead Peptidase	-
	S80	Prohead Peptidase gp175	-
SJ	S16	Lon-A Peptidase	Yes
	S50	Infectious Pancreatic Necrosis Birnavirus Vp4 Peptidase	Yes
	S69	Tellina Virus 1 VP4 Peptidase	Yes
SK	S14	Peptidase Clp	Yes
	S41	C-terminal Processing Peptidase-1	Yes
	S49	Signal Peptide Peptidase A	Yes
SO	S74	*E. coli* phage K1F Endosialidase CIMCD Self-cleaving Protein	Yes
SP	S59	Nucleoporin 145	Yes
SR	S60	Lactoferrin	Yes
SS	S66	Murein Tetrapeptidase LD-Carboxy-peptidase	Yes
ST	S54	Rhomboid-1	Yes
UNASSIGNED	S48	HetR Putative Peptidase	Yes
	S62	Influenza A PA Peptidase	Yes
	S68	PIDD Auto-processing Protein Unit 1	-
	S71	MUC1 Self-cleaving Mucin	Yes
	S72	Dystroglycan	Yes
	S79	CARD8 Self-cleaving Protein	Yes
	S81	Destabilase	Yes
	S85	Small protease	Yes

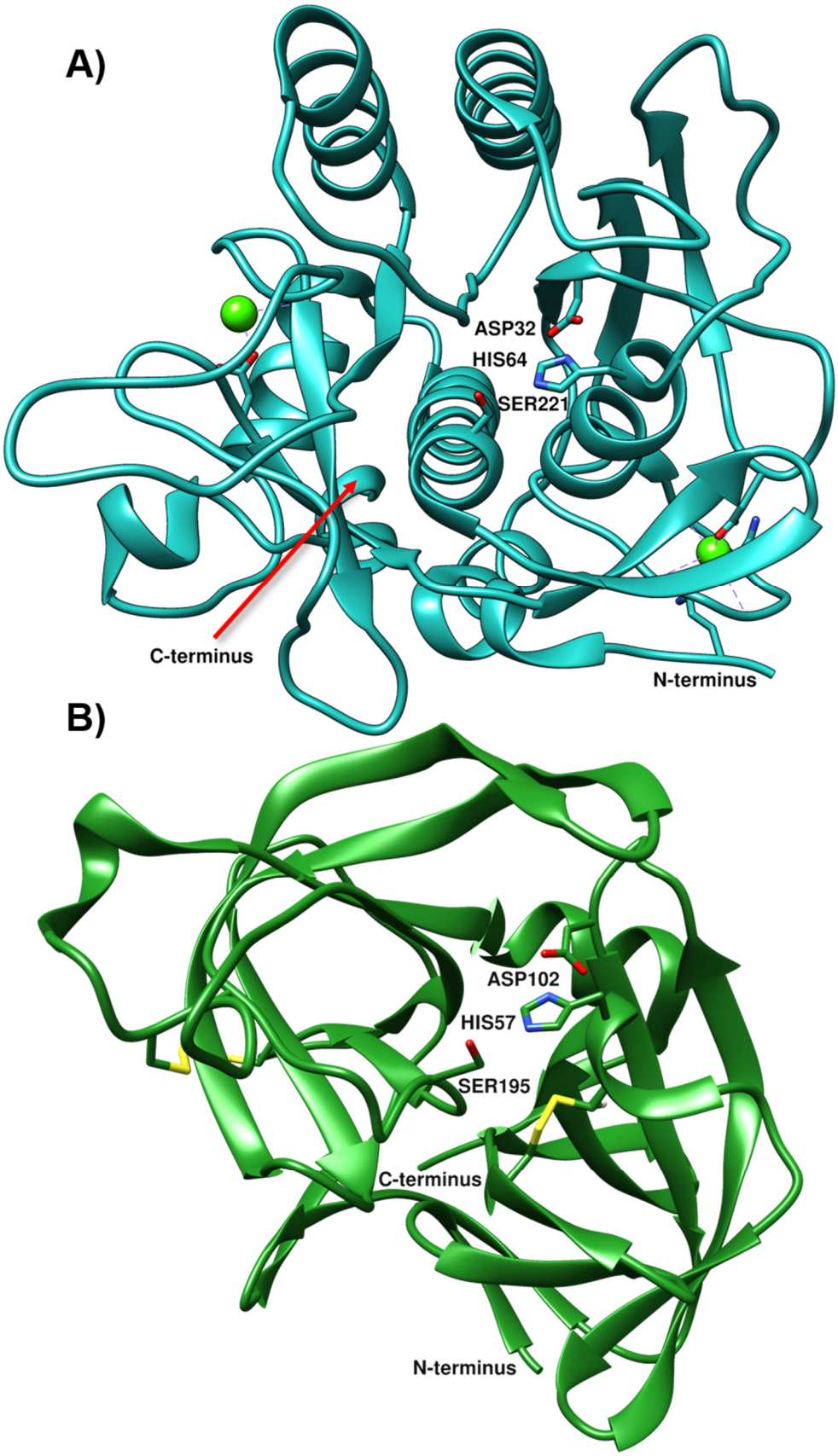


Figure 1.16. A) The crystal structure of subtilisin BPN´ (PDB: 1ST2) marking the catalytic triad. B) The crystal structure of α-lytic protease, a chymotrypsin-like protease from *Lysobacter enzymogenes* (PDB: 1SSX) marking the catalytic triad and the N- and C-terminals.

The reaction mechanism of serine proteases has been studied for decades and its deciphering shed light on the fine tuning of the molecular movements and interactions involved. The identification of this well-known triad was made on chymotrypsin, the model enzyme of the S1 family. In the 1950s stoichiometric inhibition of chymotrypsin by diisopropyl-fluorophosphate (DFP) led researchers towards identifying that a reactive serine residue was essential for the activity of chymotrypsin [222-225]. In the 1960s the histidine residue was discovered and identified by the usage of the tosyl phenylalanine chloromethyl ketone (TPCK) that reacted in a stochiometric manner with a histidine residue that was later identified by labeling with radioactive TPCK [226, 227]. And lastly, the role of the aspartic acid residue as a part of the electrostatic chain that creates the nucleophilic serine residue was brought forth in the late 1960s (Fig. 1.17) [228]. The involvement of all these residues have been shown by NMR, indicating weak Asp-His interactions that would be more in line with the view that the Asp residue has a role in orientating the side chain of the His residue correctly towards the active serine [229]. However, the strength of the Asp-His interaction among the catalytic residues might differ between the clans of serine proteases [230]. The proposed mechanism is shown in Fig. 1.17 as it is generally accepted. Starting with substrate binding, an unstable tetrahedral intermediate is formed due to the nucleophilic attack of the active serine to the carbonyl carbon, that is stabilized by hydrogen bonds from the oxyanion hole. The collapse of the tetrahedral intermediate releases the first product, the C-end part of the peptide, leaving the rather stable intermediate, the acyl-enzyme. The acyl-enzyme then reacts with an activated water molecule, forming the second tetrahedral intermediate which collapses, releasing a free enzyme and the second product [231]. The specificity of serine proteases has throughout the years provided a clear example of the structure/function relationships underlying catalysis, as for a reaction to take place the catalytic residues and the peptide substrate must be brought in close proximity and in the right orientation for a successful reaction to occur. The specificity is achieved by binding pockets on the surface of the enzyme close to the catalytic residues. The nomenclature used to describe the positions around the scissile peptide bond are from the N-terminus: P_n-P_2-P_1-P_1'-P_2'-P_n' for the sidechains of the substrate, the scissile bond being between P1 and P1′. Corresponding binding pockets or “subsites” are then marked with a S instead of a P (Fig. 1.17) [232]. For example, trypsin has a negative charged residue in the S1 pocket and cleaves at the C-site of P1 positioned arginine and lysine [233]. Elastase on the other hand has two aliphatic residues in the same pocket that reduce the size of the pocket resulting in cleavage at the C-site of

small aliphatic residues [234]. The subtilisin proteases cleave at aliphatic and aromatic residues, thus have multiple cleavage sites on almost every natural polypeptide, similar to that of chymotrypsin [216, 235, 236].

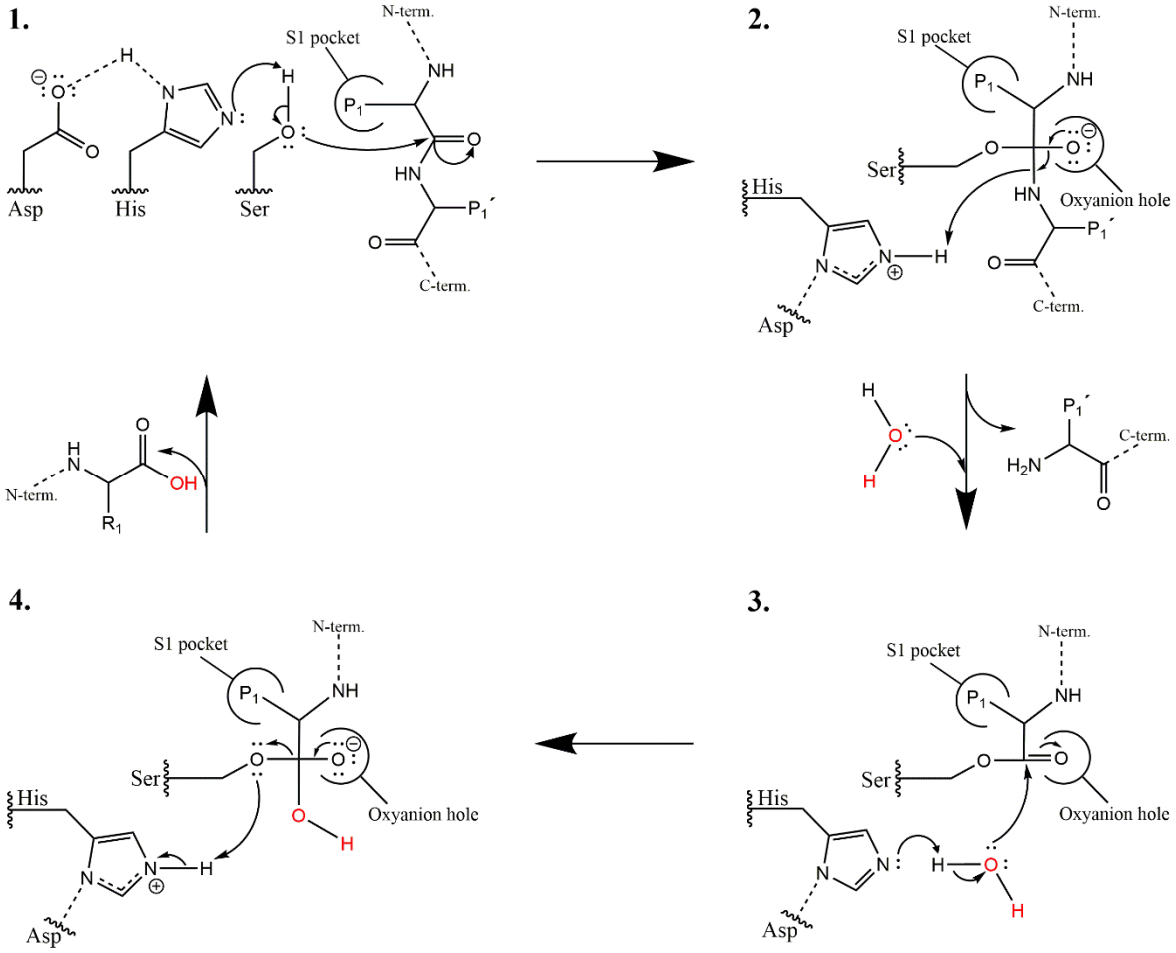


Figure 1.17. A reaction scheme depicting the catalytic hydrolysis of a peptide bond as carried out by serine proteases that utilize the Ser-His-Asp triad. Part 1 depicts the general base catalysis of the Ser residue by the His residue and the subsequent nucleophilic attack on a carbonyl carbon in the substrate by the alkoxide ion of the Ser residue. Part 2 shows the first tetrahedral intermediate stabilized via interactions to the oxyanion hole and the subsequent break down of the intermediate followed by the release of the first product. Part 3 shows the acyl-enzyme and the deprotonation of water by His and the nucleophilic attack of the resulting hydroxide ion to the acyl-enzyme. Part 4 shows the second tetrahedral intermediate and its collapse followed by regeneration of the enzyme and the second product release.

1.4.3 Subtilisin-like serine proteases

The SB clan of serine proteases contains two families, the S8 family of subtilisins and the S53 family of sedolisins [208]. The S8 family can be further divided into five different subfamilies, the subtilisins, the thermitases, the proteinase K-family, the lantibiotic peptidases and the kexin subfamilies [237]. The substrate specificity of these proteases is often

quite broad, but with a hydrophobic S1 binding pocket, thus cleaving at the C-end of hydrophobic and aromatic residues. There are known exceptions, however, e.g. enzymes that have acidic residues in the S1 binding pocket. This exception includes members of the kexin subfamily that have even more substrate specificity, cleaving only after dibasic or multibasic residues [237-239].

Members of the S8 family are found in every kingdom of life, being highly prevalent in plant and bacterial genomes, but are also involved in various protein processing in both the animal and fungi kingdoms [220, 240]. The namesake and representative of the family is subtilisin BPN´ or "bacterial protease Novo" from *Bacillus subtilis* and was one of the first protein crystal structures to be solved (Fig. 1.16) [220, 241]. Nine enzymes from the family are found in the mammalian secretory pathway working as pro-protein convertases, two of which function in the cholesterol and/or fatty acid metabolism [242]. One of those two is the proprotein convertase subtilisin-kexin type 9 (PCSK9) which has been found to regulate the levels of low-density lipoprotein receptor in the liver and thus contributing to cholesterol homeostasis [243]. In prokaryotes, subtilisin-like serine proteases are usually secreted to provide the organisms with nutrition and to assist in host invasion [220, 239]. IvaP is an example. This is a subtilisin-like serine protease secreted by *Vibrio cholerae*, and is responsible for the potentially fatal diarrheal disease cholera [244]. IvaP´s role in cholera seems to be to regulate a wide array of events, notably downregulating several host proteins. One of them is intelectin that has been postulated to be a part of the immune response to bacterial infections as it binds to bacterial specific carbohydrates [244, 245]. In addition, they also serve a role in the maturation of various polypeptides such as bacteriocins and adhesins [246-248]. The ability to function in those different processes and under various conditions is likely facilitated by adjacent protein domains and/or domain insertions that allow for specific activity and/or location in the hydrolytic machinery that regulates cell defenses, predatory mechanisms and other relevant cellular functions mediated by proteases [248].

Structural organization

Subtilisin-like serine proteases are produced in general as large pre-pro-proteins (Fig. 1.18) [249]. The organization of this pre-pro-protein is generally found to be as follows; at the N-terminus the polypeptide begins with a the signal peptide, a cornerstone for protein localization, that controls the secretion and sorting of pre-pro-proteins to the correct environment where correct folding can be initiated [82, 250-252].

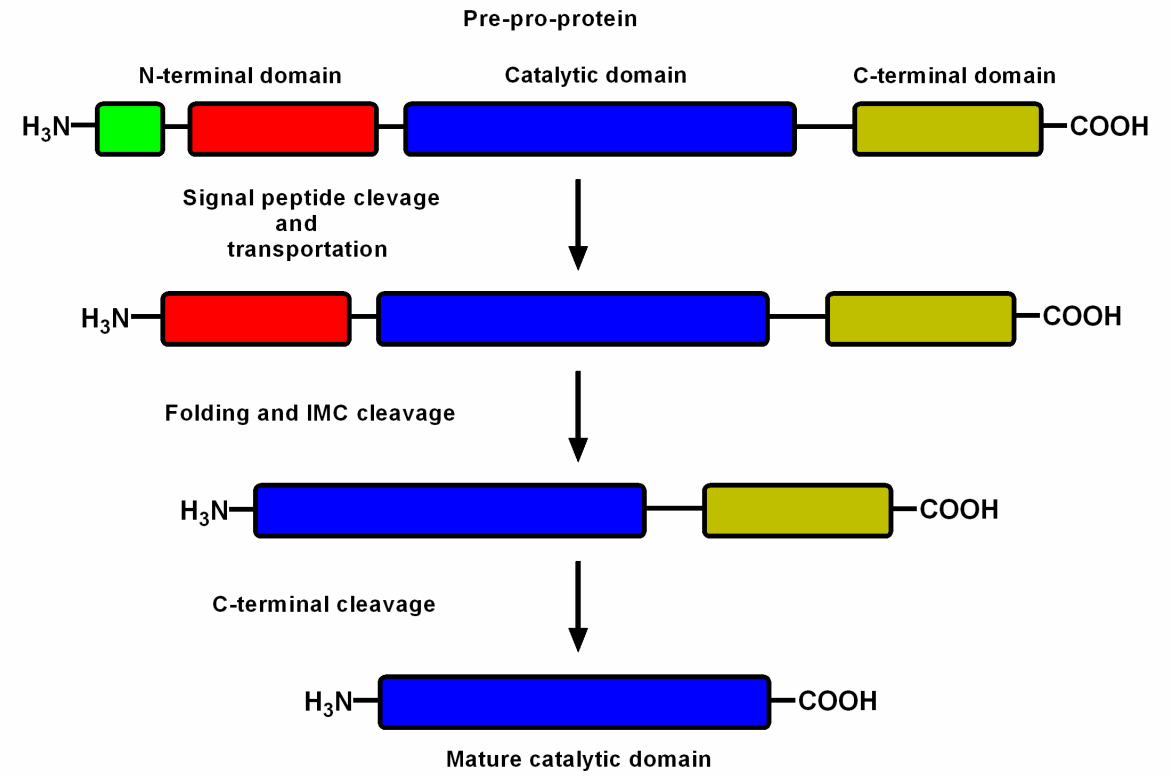


Figure 1.18. The maturation of a subtilisin pre-pro-protein. First there is transport to the correct location, followed by signal peptide (green) cleavage. Folding of the protein comes next and is followed by subsequent cleavage and degradation of the N-terminal IMC (red). Lastly C-terminal (gold) cleavage that will yield a free catalytic domain (blue).

Following the signal peptide in the primary sequence is the intramolecular chaperone (IMC), often around 75-130 residue long, responsible for fast and correct folding. The catalytic domain, in addition to being an inhibitor of its catalytic activity [85, 253-257]. The IMC likely is an intrinsically disordered protein (IDP) that assists in correct folding. The IMC is known to adopt an α-β-structure in the presence of its folded catalytic domain and inhibit the catalytic activity (Fig. 1. 19) [258]. The auto-cleavage of the IMC and subsequent release is the rate-determining step in the whole process and is heavily dependent on environmental conditions, thus controlling when the protease is activated. The first liberated catalytic domains then initiate an activation cascade of other folded IMC-pro-proteins and hydrolyze any remaining free IMCs [258, 259]. It must be noted, however, that the IMC does not need to be covalently linked to the rest of the pro-protein, as free IMC domains can induce and/or increase the rate of refolding of a lone unfolded catalytic domain and form a dimer IMC-protease complex, that will then cleave the IMC and liberate the fully active protease [83, 260-262].

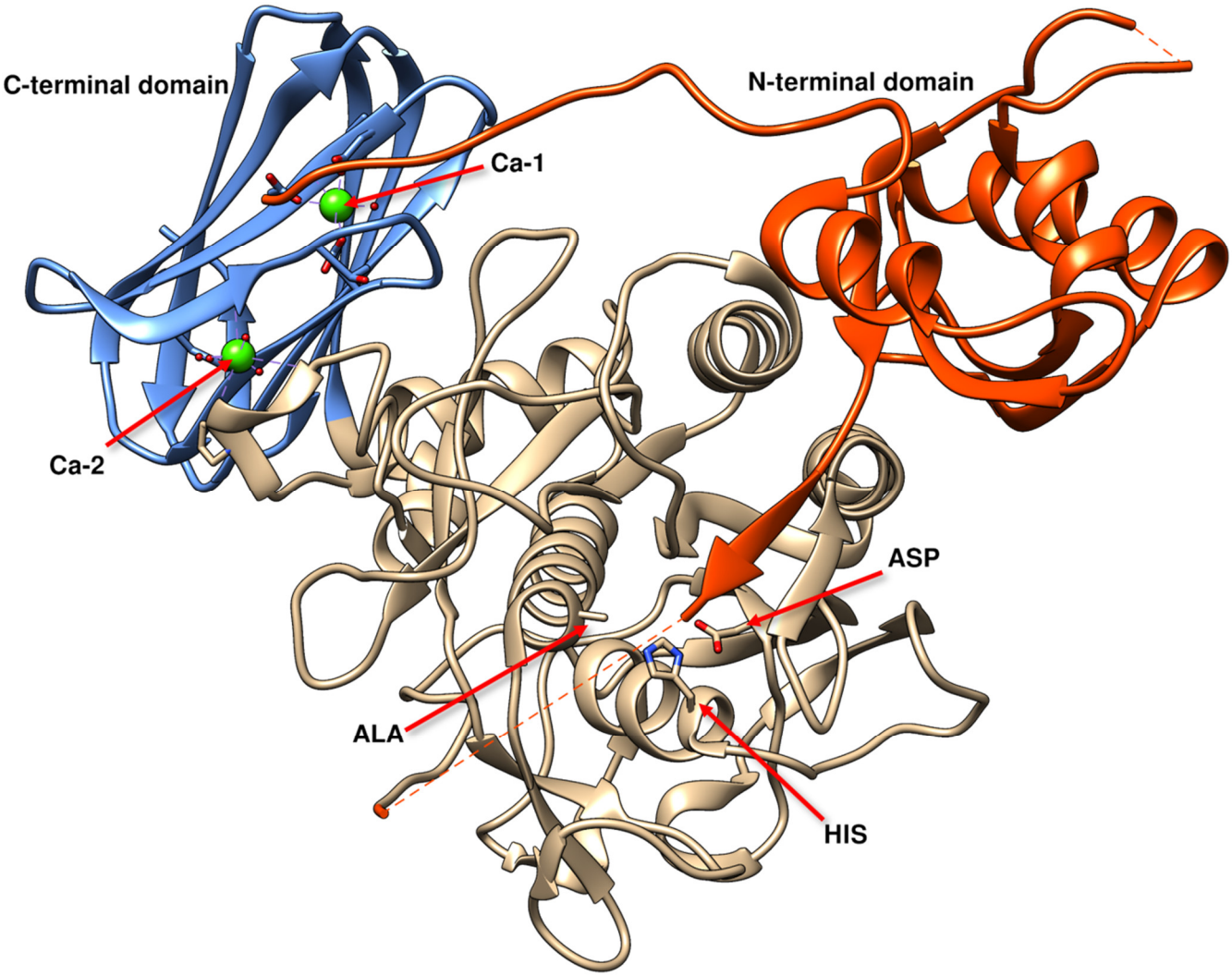


Figure 1.19. The crystal structure of subtilisin TK-SP (ProN-Tk-SP) (PDB: 3AFG) containing both the N-terminal domain (orange) and the calcium ion (green spheres) containing C-terminal domain (blue). The catalytic domain (light brown) showing the mutated active site, where the active serine has been mutated to alanine (S359A).

The active catalytic domain, often around 270-290 residue long, is characterized by a α/β-structure, where seven parallel β-sheets, aligned in the order 2314567, are located between α-helices, forming a three-layer αβα Rossmann-like sandwich. In the primary structure the organization of the catalytic triad is Asp-His-Ser. The nucleophilic Ser and the activating His residues are located at the ends of two adjacent α-helices (helix-E and helix-B, respectively) with the Asp residue on the primary β-sheet (Fig. 1. 20 and 1.22). Coming together in the tertiary structure, they form the correct geometry of these residues to form the catalytic triad [249].

Following the catalytic domain is a C-terminal domain, often around 100-150 residues long. The role of this domain is somewhat unclear, as it is often autocleaved and discarded after

the IMC has been processed leaving just the mature catalytic domain in prokaryotes, but the domain is retained in eukaryotes. The exact role of this C-terminal domain is thus related to the organism of origin and the role of the specific subtilisin-like protease in the proteolytic machinery. As an example of this is the C-terminal domain of aqualysin I (AQUI) from the thermophilic *Thermus aquaticus*. When the protease is expressed in *Escherichia coli* with the C-terminal domain intact, the protease was mostly found to be bound in the outer membrane of the bacterium, whereas in a C-terminal domain deficient mutants the localization of AQUI was mostly in the cytoplasmic space [263, 264]. This indicated a role of the C-terminal domain in the translocation of AQUI through membranes and possibly membrane tethering. Further mutations on the C-terminal domain of AQUI indicated that the possible mode of action of the C-terminal domain, facilitating transmembrane export, was by inhibiting the rapid folding of the catalytic domain by the IMC [265]. The C-terminal domain of AQUI shares a partial sequence homology to the C-terminal domain of the subtilisin-like serine protease KP-43 from *Bacillus* sp. KSM-KP43 [266]. The crystal structure of KP-43 contains the C-terminal domain that consists of an eight-stranded β-barrel, resembling a jelly roll β-barrel domain, binding two calcium atoms. As a C-terminal truncated form of KP-43 could not be expressed, it suggested that it might be involved in folding or the stability of the active protease domain [266] In the case of the hyperthermostable subtilisin-like serine protease Tk-SP from the archaeon *Thermococcus kodakaraensis*, the β-barrel might play a role in its hyperthermostability (Fig. 1.19) [267, 268]. The β-barrel interacts with the catalytic domain through salt-bridges, in addition to partaking in calcium binding at the domain-domain interface, where chelating conditions using EDTA result in dramatic loss in thermostabilty [249, 267-269]. However, these C-terminal domains are usually cleaved off and discarded in prokaryotic subtilisins, at least under laboratory conditions. So, their functionality for the mature protease seems rather diverse. In eukaryotes, jelly roll β-barrel C-terminal domains are also found in the genes of subtilisin-like serine proprotein convertases of the kexin subfamily. However, those C-terminal domains do not share sequence homology with the prokaryote subtilisin C-terminal domains, possibly indicating a very distant relationship, if any at all. These C-terminal domains, called P-domains, appear to be important for the proteolytic activity and serve a stabilizing role as they are maintained in the mature protease [249, 269]. In addition, these P-domains contain a variable C-termini that play a role in cell surface anchoring and therefore appear to be important for correct localization of the active protease [269, 270].

Calcium binding

Calcium binding is found almost ubiquitously in all natural subtilisin-like serine proteases, except for plant subtilisins, There are, however, indications of bacterial subtilisins that seem to function well in the presence of EDTA such as the IvaP from *Vibrio cholerae* [245, 271]. This indicates that either there are bacterial subtilisins that do not rely on bound calcium ions, or have extremely tightly bound calcium ions. However, as a rule of thumb, the role of calcium in these structures is mainly to maintain the correct native fold, i.e. protecting against autolysis, increasing thermostability, folding, and possibly in some cases calcium ions are involved in substrate binding, presumably through stabilization of binding pockets [3, 267, 272-276]. The number of calcium binding sites and their location within the structure does however vary greatly, both within and between subfamilies of subtilisins [249]. There is also no direct correlation between the number of calcium binding sites and thermostability of the structure. An example is the crystal structure of the thermostable AQUI, which contains two bounds ions, whereas the cold active structural homolog VPR from *Vibrio* sp. PA-44 contains three, including the two found in AQUI [277, 278]. At the low end for number of calcium binding sites is the thermostable thermomycolase from the fungus *Malbranchea pulchella* that contains just one calcium binding site [279], to the high end of seven bound calcium ions as found in TK-subtilisin from *Thermococcus kodakaraensis* [280]. Thus, TK-subtilisin and TK-SP, both being hyperthermostable, have such varying calcium ion content, seven versus the two as found in TK-SP, and those two are part of the C-terminal β-domain, albeit one being on the domain-domain interface (Fig. 1.19). This high number of calcium ions found in TK-subtilisin stems in part from a four calcium ion cluster that is located in a loop insertion, that may be a reason for the calcium mediated refolding of the catalytic domain without the IMC [280]. These calcium binding loop insertions are not unique, however, to the TK-subtilisin, as the SUB1 subtilisin from the protozoan parasite of the *Plasmodium* genus that causes malaria contains three calcium ions, two of which are located in a loop insertion close to the N-terminus [281]. In the case of SUB1, calcium ions are essential for folding, but may also have an activating property via stabilization around the S1 binding pocket, allowing for the formation of a disulfide bridge which is essential for activity [281]. In addition, this calcium loop in SUB1 is not the same as is found in TK- subtilisin, meaning that a myriad of different calcium binding sites are found in the structures of subtilisin-like serine proteases often with rather specific roles. One calcium site has been shown to be highly important for the kinetic

stability of many of the bacterial subtilisins, as has been well demonstrated for subtilisin BPN´ that has a calcium binding site near the N-terminus (Fig. 1.20) [282].

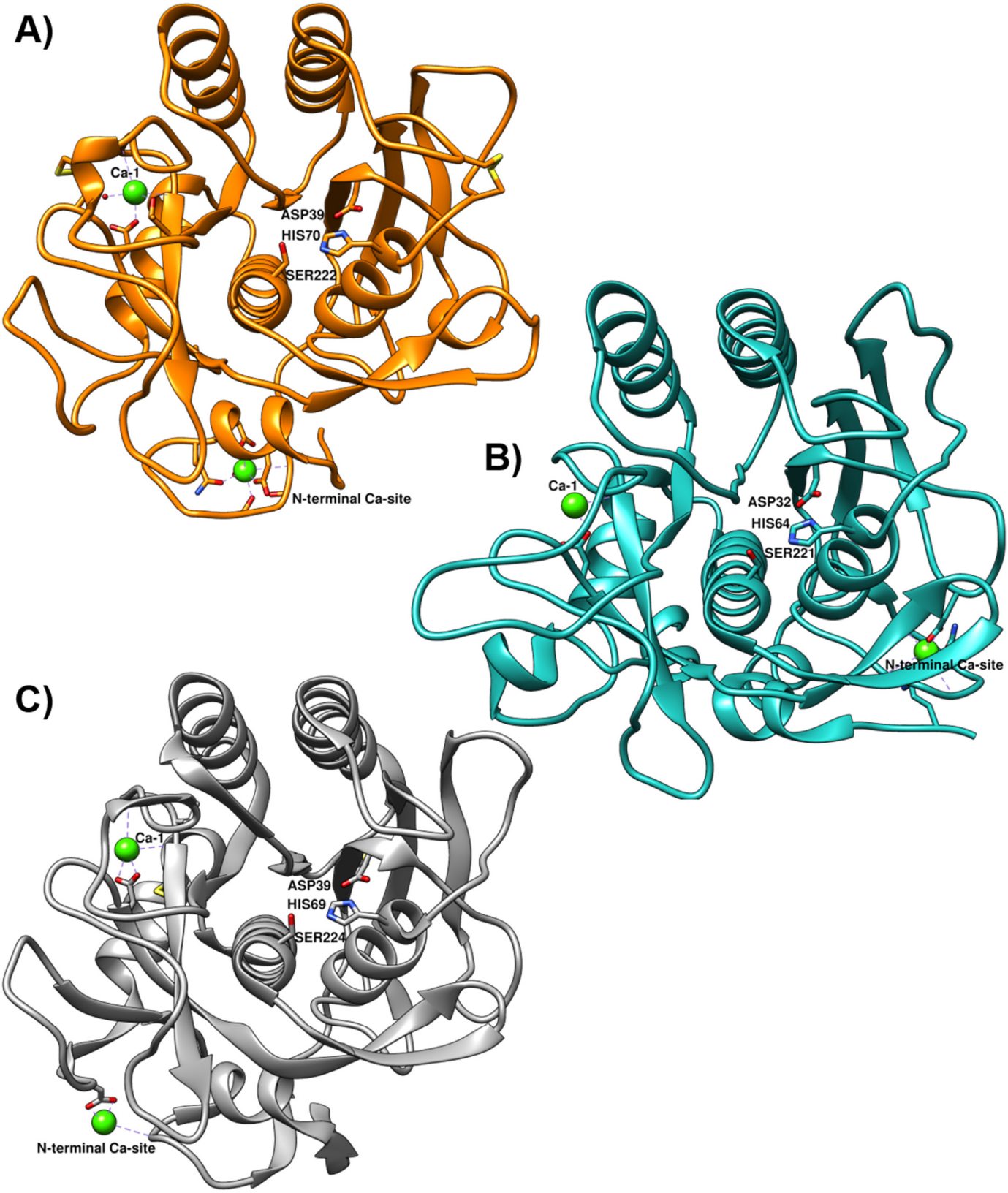


Figure 1.20. The crystal structures of A) AQUI (orange) (PDB: 4DZT), B) subtilisin BPN´ (blue) (PDB: 1ST2) and C) PRK (grey) (PDB: 2PRK) marking the calcium ion (green spheres) binding sites.

This contains a calcium binding site near the N-terminus in the structure of the mature active subtilisin BPN′, cannot be formed until after the IMC-catalytic domain peptide bond has been hydrolyzed, allowing for the formation of the N-terminus. Only at that time, the N-terminal calcium binding site can be formed, which binds calcium tightly, in the nM range. When this calcium binding site was deleted, an active protease was formed, but which had become much less stable against thermal denaturation, but which could refold independently [282]. The calcium independent subtilisin BPN′ variant did fold an order of magnitude faster than the wild type in the presence of the IMC. These results, therefore, indicated that bacterial proteases may have adopted strategies that compromise folding in order to gain kinetic stability.

Plants are apparently the only organisms to have evolved calcium free subtilisins. In the case of the tomato subtilisin STB3, the role of calcium seems to have been solved by the insertion of a PA-domain (protease associated) leading to homodimerization, loop extensions and by utilizing the positive charge at the side chain of a lysine residue where a calcium ion is usually found (Fig. 1.21) [271, 283].

As mentioned earlier, bacterial subtilisins have been subjected to mutagenesis to delete calcium binding sites. But as a result, the stability of the calcium depleted variants was greatly diminished. A notable work on designing a calcium independent subtilisin BPN′ has been carried out, producing a calcium-independent subtilisin variant with restored thermal stability and that could fold without the IMC [284-286]. Stable calcium free subtilisins are sought after for industrial application, where proteolytic activity is required under chelating conditions.

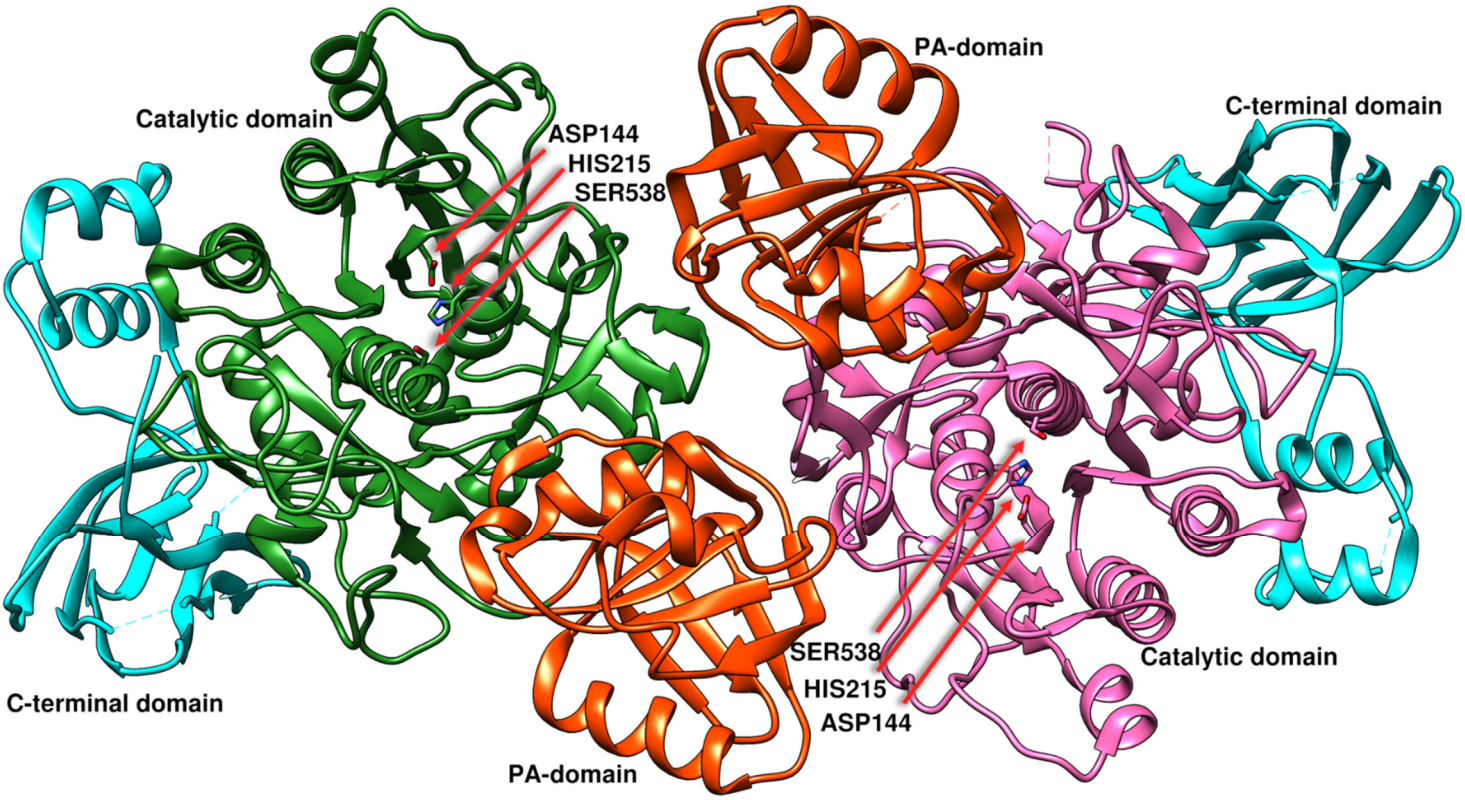


Figure 1.21. The crystal structure of the tomato subtilisin STB3 (PDB: 3I74). Shown are the catalytic residues of both monomers, Asp144, His215 and Ser539. The catalytic domains are colored green and pink to distinguish between each monomer comprising the dimer. The PA-domains at the dimer interface are colored orange and the C-terminal domain is colored light blue.

Industrial and research value

Enzymes have widespread applications as industrial tools ranging from food applications and as additives in detergents to pharmaceutical and medical diagnosis. The estimated value of the enzyme industry stood at 1.5 billion dollars at the turn of the 21st century. Of that amount, proteases provided around 60-65% of the total value and alkaline proteases, there among subtilisins, 25% of the total market share [287]. Subtilisins have also been of great value as research models, notably assisting in the development and pioneering of directed evolution methods [288-292], that was awarded the Nobel prize in chemistry in 2018 to Francis H. Arnold for her work in directed evolution.

1.4.4 The research model: VPR and AQUI

The research model utilized in this doctoral thesis work consists of two extremophilic subtilisin-like serine proteases of the proteinase K subfamily, a cold active protease and a thermostable protease [1]. The proteinase K subfamily is named after proteinase K (PRK) form

the fungi *Engyodontium album* and mostly consist of secreted alkaline endopeptidases found in fungi and bacteria, sharing a relatively high sequence identity of 37 % [237, 293]. The cold active subtilisin-like serine protease VPR was isolated from the Gram-negative psychrotrophic bacterium *Vibrio* sp. PA-44 [1]. The gene that encodes for VPR consists of 1593 base pairs that yield a 530 residue long pre-pro-protein [294]. The pre-pro-protein consists of a 139-residue long N-terminal sequence, containing the signal peptide and the IMC, a 291-residue long catalytic domain and a 100-residue long C-terminal sequence [294, 295].

The thermostable subtilisin-like serine protease AQUI was isolated from the thermophilic bacterium *Thermus aquaticus* YT-1 of the *Deinococcus-Thermus* phylum [2]. Despite testing Gram-positive, bacteria of the *Deinococcus-Thermus* phylum processes complex cell envelop, including an outer lipid layer, thus being more akin to Gram-negative bacteria [1, 2, 296, 297]. AQUI is encoded by 1539 base pairs yielding 513-residue long pre-pro-protein, 127-residue long N-terminal sequence, 276-residue long catalytic domain and a 110-residue long C-terminal sequence [295, 298].

VPR and AQUI share a high structural homology, in addition to having a sequence identity of around 60 % [294]. When comparing the crystal structures of the catalytic domains of VPR and AQUI, a notable difference is seen in the C-terminus, as VPR contains a 15 residue longer C-terminal tail that contains one disulfide-bridge between residues 277 and 281, using VPR numbering for clarification (Fig. 1. 22). Both VPR and AQUI share two other disulfide-bridges, between residues 67 and 99 and 163-194. At the N-terminus, AQUI contains two extra residues, showing that the autocatalytic site of the N-pro-peptide cleavage is shifted by two positions into the catalytic domain of VPR in the maturation process. The crystal structure of the mature catalytic domain exhibits the classical subtilisin αβα-sandwich structure, with the residues of the catalytic triad located on β-sheet 1 (Asp37), α-helix-B (His70) and α-helix-E (Ser222). The S1 to S4 binding pockets are then formed by the loops preceding α-helix-C, α-helix-D and the loop between β-sheet 5 and 6, that also takes part in coordinating the calcium ion in the Ca-1 site (Fig. 1.22) [277].

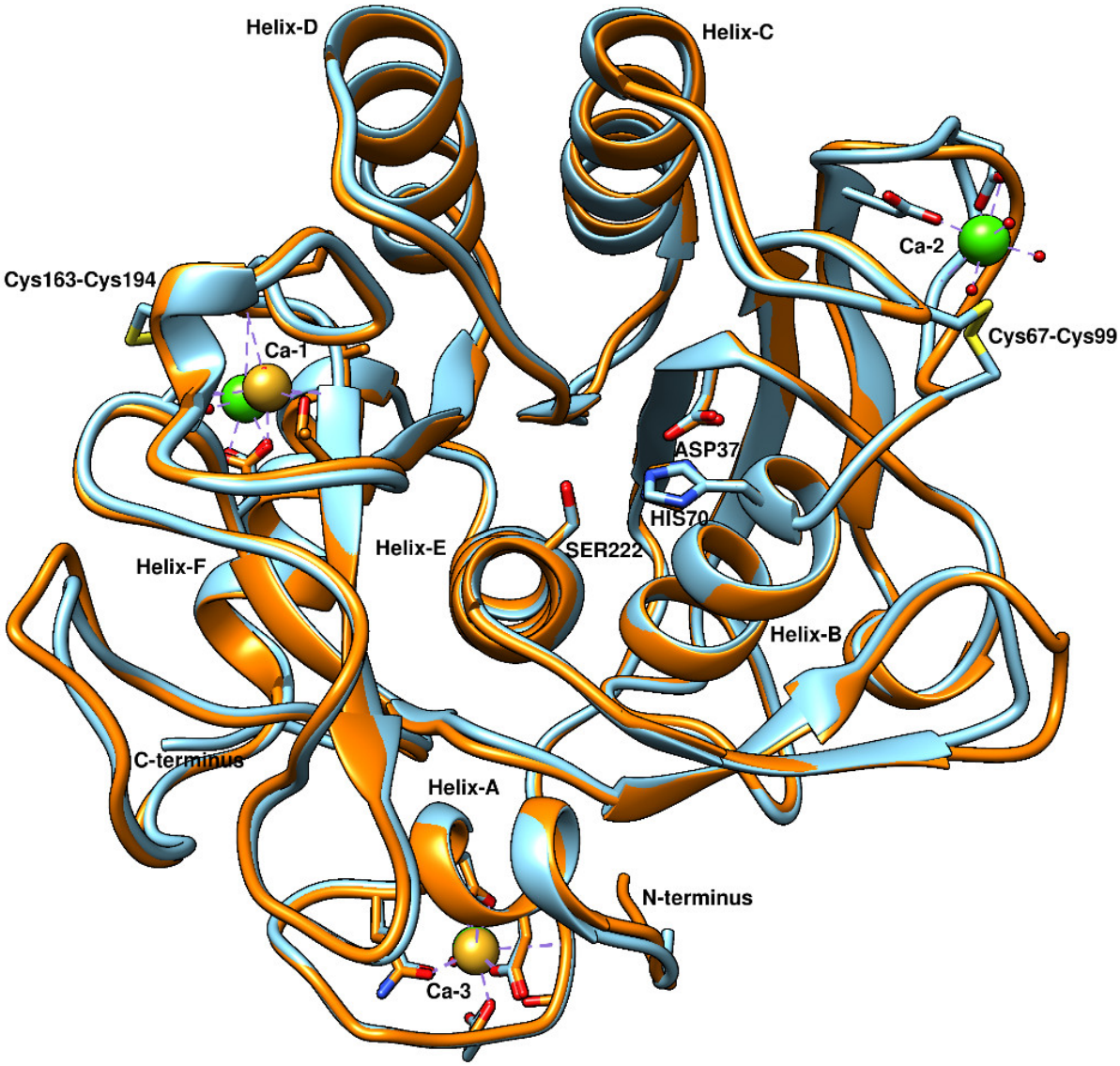


Figure 1.22. The crystal structures of VPR (light blue) (PDB: 1SH7) and AQUI (orange) (PDB: 4DZT). Marked are the active site residues, the calcium binding sites, disulfide bridges, the N- and C-terminals. The helices are marked in an order from A to F. Calcium ions in the structure of VPR are shown as green spheres and the calcium ions from AQUI as golden spheres.

Additionally, AQUI has a two-residue loop deletion corresponding to residues 61 and 62 in VPR. This loop harbors a calcium binding site in VPR (Ca-2), but not in AQUI. Both enzymes share the remaining two calcium binding sites. Those sites are the conserved calcium 1 binding site (Ca-1), shared with PRK, and the N-terminal calcium 3 binding site (Ca-3), first identified in the crystal structure of VPR [277]. The Ca-3 site in VPR and AQUI binds a highly coordinated calcium ion, located within a α-helical-loop motif that envelops the calcium ion (Fig. 1.22). This N-terminal motif interacts directly with C-terminal regions of the protein and takes part in forming the hydrophobic core of the protein. In PRK, the stabilization of the N-terminus seems also to be mediated by a N-terminal calcium binding

site. However, in the PRK structure the calcium ion seems more to be solvent exposed and bridges the N- and C-terminal parts of the protein, whereas the N-terminus itself takes up a antiparallel β-sheet motif, with the α-helical-loop motif separating the two β-strands forming the sheet (Fig 1.21). The amino acid sequences of the N-terminal regions of VPR and AQUI share a few conserved residues with PRK, notably a tryptophan residue located at the first α-helix, coming in close proximity with residues at the C-terminals of two β-sheets that take part in forming the Ca-1 site, including the conserved Asp196 a main coordinator of that calcium ion. The numbering of the Ca-1 site comes from its presence in the structure of PRK. The calcium binding at that site is much stronger as compared to the much more weakly bound calcium ion bridging the N- and C-terminals [299]. This is likely not the case for VPR and AQUI, as another cold active subtilisin-like serine protease, SPRK from a psychrotrophilic *Serratia* species, also has the Ca-3 site, while the Ca-1 and Ca-2 sites are depleted of calcium in that crystal structure. SPRK, however, contains all the necessary structural components for these calcium binding motifs as found in VPR and the relevant calcium binding residues [300]. This strongly suggests that the Ca-1 and Ca-2 sites are weaker binders of calcium than the Ca-3 site in those structures. In addition, the Ca-2 site which is not found in AQUI, has been shown not to affect the dynamics of the structure of VPR in MD simulations, when comparing simulations of the enzyme with a calcium ion either present, or absent, in the Ca-2 site [301]. This may be linked to the close proximity to the Cys67-Cys99 disulfide bridge that essentially rigidifies that part of the protein and the Ca-2 site.

Previous research into the determinants of temperature adaptation

Possessing a system like VPR and AQUI can give a great insight into the structural factors that determine the ability of these enzymes to function on their respective ends of the temperature spectrum [1]. Over the temperature range where the activity of VPR is readily measured, the enzyme exhibits much higher k_{cat} values and lower K_M values as compared to AQUI, resulting in catalytic efficiencies that are far higher than for AQUI. Additionally, AQUI is far more stable both against heat (Fig. 1.23) and denaturants [1, 302].One of the larger differences observed between the processed catalytic domains of VPR and AQUI, is the extended, fifteen residue long, disulfide-bridge containing, C-terminus of the former. In order to create an even more compatible model system for our comparative structural studies, a C-terminal truncated variant of VPR, VPR_C277$_{stop}$ ($VPR_{\Delta C}$) was constructed [303]. The $VPR_{\Delta C}$ variant did not show large changes in properties but had slightly lower catalytic

efficiency and an increase in melting temperature of about 1.5°C, up from 63.6°C, when measured by circular dichroism (CD). These and other results showed that shortening of the C-terminus of VPR had minor effects on the properties of the enzyme, at least under laboratory conditions using a synthetic substrate.

On the basis of structural comparisons of the structures of VPR and AQUI, a multitude of mutations have been made on both VPR and AQUI, to test hypotheses on structural aspects of their respective temperature adaptation. These involve Ser/Ala exchanges at selected sites in VPR that had moderately stabilizing effects. Such specific Ser to Ala exchanges were among those most frequently observed between the cold and heat adapted structures of enzymes in general [304]. Single and double proline substitutions were incorporated into the loops of VPR based on location of prolines in loops of AQUI, that had variable effects, as proline mutations at the N-terminus were much more stabilizing than at other selected sites [304, 305]. A more systematic and extended study on the effects of such incorporations of prolines into loops of VPR are the subject of a part of this thesis [4].

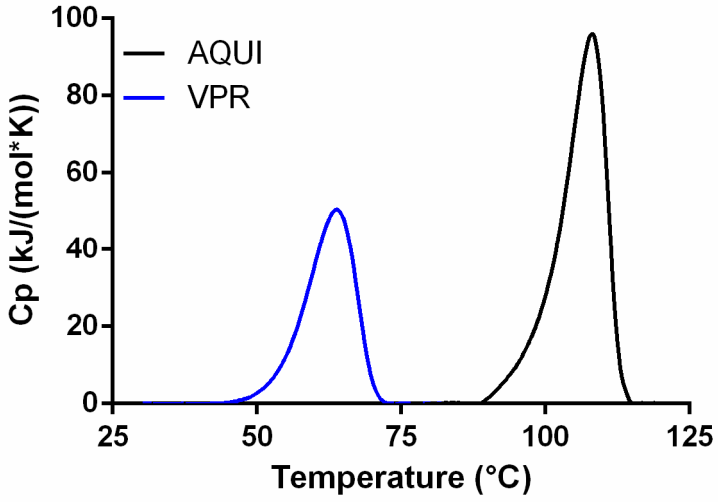


Figure 1.23. Comparison of normalized thermograms of $VPR_{\Delta C}$ (blue line) and AQUI (black line) at pH 8.6 and 15 mM $CaCl_2$, with a temperature gradient of 1°C/min recorded separately on DSC in our laboratory.

Some chosen residues of VPR have also been incorporated into the structure of AQUI at corresponding sites. Four variants were created containing a single Pro-Xxx mutation, all of which displayed diminished stabilities, especially for proline exchanges at the N-terminal. All the variants had similar catalytic efficiencies [161]. Those experiments were however conducted at calcium chelating conditions, due to the usage of a phosphate buffer [161, 306]. Interestingly, one of those variants (AQUI_P5N), displayed a DSC thermogram with two

peaks, a similar observation as is found for many of the proline variants measured in the work of this thesis. This will be discussed in the results and discussion chapter.

AQUI has the potential of forming thirteen salt-bridges that are not found in VPR, six of whom are single salt-bridges and the rest being in small clusters of three or four residues [307]. A study was undertaken to delete those putative salt-bridges from the structure of AQUI by site directed mutagenesis. Eight single variants were produced to delete individual salt-bridges that had been predicted form MD-simulations to be important in the structure of AQUI. Results showed relatively small decreases in stability of the enzyme, except for one variant. That variant was AQUI_D17N, where this single mutation resulted in a lowering of the melting point by 4.1°C. The effect on thermal inactivation of the active conformation was even more pronounced with an 8.8°C decrease in $T_{50\%}$ (the temperature which half of the activity is lost in half an hour) [307, 308]. The side chain of Asp17 has the possibility of interacting with Arg12 and Arg259. The Asp17-Arg259 interaction would most likely be more prevalent, as Arg12 is expected to interact strongly with Asp183. From MD-simulations of these salt-bridges in AQUI, it turns out that many of them are dynamic, some just persisting in around 2 % of simulations frames [307]. Thus, the stabilization role of surface exposed putative salt-bridges is not fully clear, as interpretation of the effects of such single charge deletions is highly dependent on conditions and might need to be viewed in a protein interaction network context. Of note, however, is that the Asp17-Arg259 salt-bridge, being important for thermal stability, essentially links together the same structural parts as the calcium binding site 2 does in PRK, as the main coordinators of that calcium ion are the side chain of Asp260 and the peptide bond oxygen of Thr16. That calcium ion binding site of PRK and the salt-bridge of AQUI are thus linking the C- and N-terminal parts of the proteins, indicating that interactions between those parts are important. Linking those regions together could thus be important for the kinetic stability of the structures, possibly by increasing the unfolding cooperativity.

In VPR, there is a lysine residue (Lys257) at the corresponding location to Arg259 in AQUI. Furthermore, at the site corresponding to Asp17, an asparagine residue is present (Asn15). Thus, salt-bridge formation is not possible at this site. The introduction of the N15D mutation into the structure of VPR, however, was stabilizing, increasing the melting point by about 3°C, without affecting the catalytic efficiency, at least when using $VPR_{\Delta C}$ as a template [303].

Further studies have also shown that increasing connectivity/contacts between the N- and C-terminal parts impacts the stability of these structures. One of the interconnected clusters found in AQUI is located between helix D and the loop that makes up a part of the Ca-1 site. In AQUI, this cluster consists of salt-bridges between Asp138-Arg169 and Lys142-Glu172, with Lys142 being able to interact with Asp138, both being on helix D.

In the structure of VPR, only the salt-bridge between Asp138-Arg169 is found. Thus, the variants $VPR_{\Delta C}$_Q142K, $VPR_{\Delta C}$_S172E and $VPR_{\Delta C}$_S172E/Q142K were produced [309]. None of these variants had stabilizing effects on the structure, instead they were rather destabilizing. This was evident in the case of $VPR_{\Delta C}$_S142E/Q142K where the exchange of residues was found to lower the melting point of the variant by 4°C compared to $VPR_{\Delta C}$. The $VPR_{\Delta C}$_S142E/Q142K modification had no effects on the activity. However, the single variants did show an increase in the catalytic efficiency, especially in case of $VPR_{\Delta C}$_Q142K. This was related to Lys142 interacting with Asp138, thus interrupting the Asp138-Arg169 salt-bridge and further increasing the mobility important for catalysis (Fig. 1.24) [309]. MD-simulations of VPR and AQUI have been carried out in order to compare the flexibilities of the structures. These studies have shown that the structure of VPR is overall more flexible, with a general noisier RMSF of the peptide backbone, as compared with AQUI. Observations of a higher degree of flexibility of VPR have also been supported by fluorescence quenching experiments, showing increased permeability of acrylamide into the structure [310]. These results are in line with the idea of the higher flexibility of enzymes that are adapted to colder environments. In general, flexibility of these structures has a similar overall profile, with notable regions of high flexibility separated by more rigid areas. Of special note are three areas; around β-sheet 2 and into the loop that contains the Ca-2 binding site in VPR, the loop connecting β-sheet 3 to helix-C and the loop between β-sheet 4 and helix-D. The loops preceding helix-C and helix-D are a part of the substrate binding pockets in those structures. In addition, some increased mobility was seen in the loop preceding β-sheet 7 and the loop just preceding the C-terminus (Fig. 1.25) [309, 310].

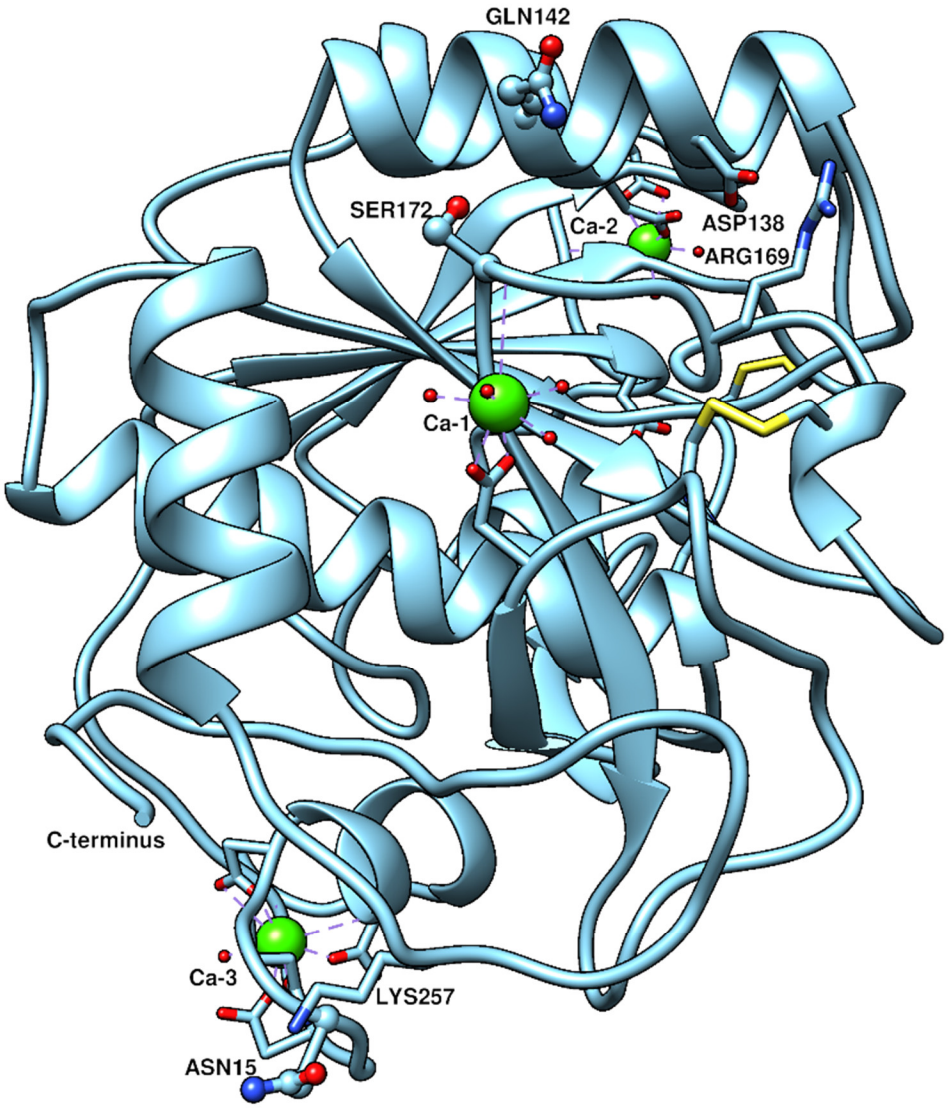


Figure 1.24. The crystal structure of AQUI (PDB: 4DZT). The highlighted areas are the parts that both VPR and AQUI show increased flexibility during MD-simulations, excluding the highly mobile terminal regions.

The description that has been presented and summarized above indicates the state of the research model at the beginning of the work presented in this dissertation. The aim was to examine further what effects additive proline exchanges would have on stability, activity and flexibility. Flexibility estimations would be conducted using active site labeling, steady-state and time-resolved fluorescence quenching, in addition to MD-simulations, with the aim of comparing those results to AQUI. Assisting in resolving the fluorescence spectra, Trp-to-Phe variants were to be produced, in order to resolve the contribution of individual

tryptophan residues to the fluorescence spectra of VPR, in addition to monitoring their effects on stability, activity and flexibility. Results obtained in the early parts of the study led to more thorough investigation into the structural basis of kinetic stability and unfolding cooperativity of VPR and its variants, utilizing differential scanning calorimetry.

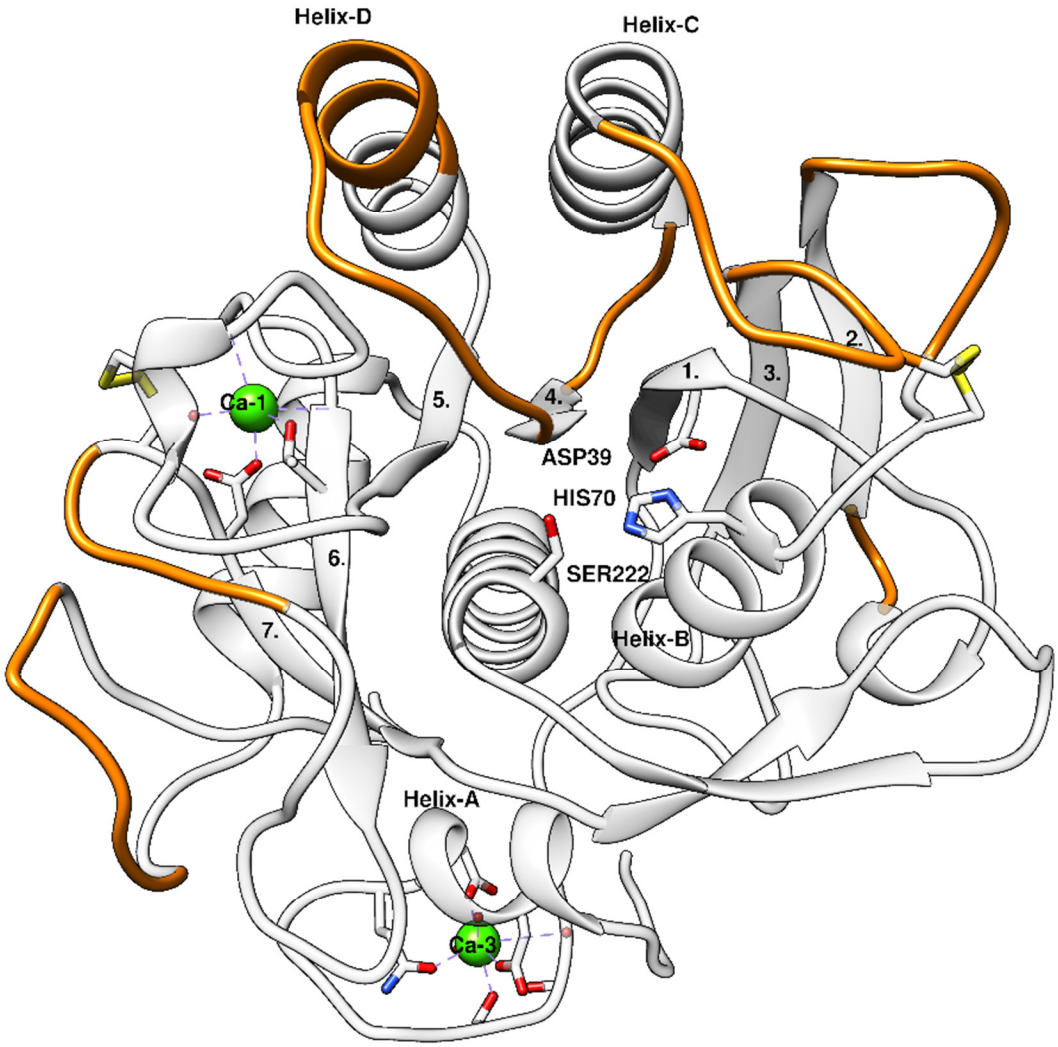


Figure 1.25. The crystal structure of VPR (PDB: 1SH7), highlighting the residues (balls and sticks) that have been mutated to investigate the role of salt-bridges in the structure of VPR.

2 Methodologies

This chapter will outline methodologies utilized in the current work, published and unpublished.

2.1 Molecular modelling and data fitting

Molecular graphics, analysis of crystal structures and sequence homology analysis of crystal structures was carried out with UCSF Chimera [311]. Non-linear and linear fitting of data sets was carried out using either GraphPad Prism 6 for Windows or the analysis software KaleidaGraph 3.6 for Windows.

2.2 Production and purification

The expression system utilized in earlier work involving VPR had the drawback of usually only yielding less than 500 U per 1 L of expression culture. When initial expression tests of tryptophan exchange variants yielded extremely low levels of active protein, the decision was taken to try out other expression systems to improve yields. As described in the paper "Improved expression, purification and characterization of VPR, a cold active subtilisin-like serine proteinase and the effects of calcium on expression and stability" [3] we opted to try out the T7 based expression system using the pET-11a-d vector in the titratable T7 polymerase/T7lysosyme system of the *E. coli* strain Lemo21 [312].

2.2.1 Cloning

Cloning was performed via overlap extension PCR cloning [313] using Phusion® High-Fidelity DNA polymerase. The genes coding for the C-terminal truncated form of $VPR_{\Delta C}$ [277] and AQUI YT-1 [263, 308, 314] were both attempted to be cloned into the pET-11a vector. The PCR reaction mixture used to construct the megaprimers was as follows: 3.0 µM of sense and anti-sense primers, 1.5 nM of the vector containing the target gene, 200 µM dNTP, 4 U of Phusion polymerase and using the GC buffer from New England Biolabs, in a total volume of 20 µL. The following program was used: the mixture was pre-heated at 98°C for 30

seconds, then thermo-cycles ran 25-times in the following order: 98°C for 10 seconds, 55°C for 20 seconds, 72°C for 50 seconds and repeat. The final elongation step was performed at 72°C for 10 minutes followed by cooling to 4°C. After amplification, the PCR product was gel purified from a 1% agar gel and the appropriate band was isolated using a gel extraction kit. The purified megaprimer was then used to clone the gene into the pET-11-a-d vector. The PCR reaction mixture contained the purified megaprimer in a 250-fold excess over the pET-11a vector under the same conditions as in the first PCR reaction. The PCR program was as follows: denaturation at 99°C for 2 minutes followed by 20 cycles of 95°C for 50 seconds, 60°C for 50 seconds and 68°C for 12 minutes. The final elongation step was carried out at 68°C for 12 minutes and then the reaction was quenched by cooling to 4°C. DpnI was then added to digest any methylated DNA in the samples and incubated at room temperature for 1.5 hours, followed by transformation into XL10-Gold cells and then plated on LB-discs that were grown overnight at 37°C. Single colonies were then picked and grown overnight in LB-Miller broth containing ampicillin at 37°C and rotated at 230 rpm. Plasmid purification was carried out using the Monarch plasmid miniprep kit from New England Biolabs, following their instructions. Cloning was then verified by Sanger sequencing at Genewiz.

2.2.2 Strains, plasmids, media and transformation

All plasmid production was carried out in the *E. coli* strain XL10-Gold having the genotype: *TetrD(mcrA)183 D(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte [F′ proAB lacIqZDM15 Tn10 (Tetr) Amy Camr]*. The expression of VPR was performed in strain Lemo21 (DE3) having the genotype: *fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS/pLemo(CamR) λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5 pLemo = pACYC184-PrhaBAD-lysY*. Transformation of both strains was carried out by diluting PCR samples or purified plasmids (50 – 100 ng/µL) 50-fold into Eppendorf tubes containing competent cells and kept on ice for 10 to 30 minutes. Heat shock was carried out at 42°C for 60 seconds and placed on ice to chill to ambient temperatures fast. Samples were then diluted 5 to 20-fold into LB-Miller broth and grown for 1 hour at 37°C. Cells were then plated and grown at 37°C overnight. Cultures and agar plates for plasmid production were made from LB-Miller broth containing 0.1mg/mL ampicillin. Lemo21 cells for expression were grown on LB-Miller plates containing 0.1 mg/mL ampicillin, 0.03 mg/mL

chloramphenicol and 0.1% (w/v) L-rhamnose. Liquid media for protein expression was 2xYT broth containing 0.1 mg/mL ampicillin and 0.03 mg/mL chloramphenicol.

The plasmid pJOE3075 containing the gene coding for AQUI YT-1 was transformed into the *E. coli* strain BL21 with the genotype fhuA2 [lon] ompT gal [dcm] ΔhsdS, following the same protocol as for VPR.

2.2.3 Expression and optimization

VPR expression and optimization

Starter cultures of transformed Lemo21 cells in 2xYT broth were cultivated at 37°C and 230 rpm overnight. Starter cultures were then diluted 50x into 2xYT expression cultures containing varying concentrations of L-rhamnose. Cultivation was performed in Innova44 incubators at 37°C and 230 rpm. Cells were grown until A_{600} had reached 0.4 - 0.8 A.U. At that point the culture was made 400 µM in IPTG and varying amounts of a sterile 4 M $CaCl_2$ were added. At that point the temperature was lowered to 18°C and the flasks shaken at 230 rpm for 20 hours. Cells were harvested by centrifugation at 4500xg at 10°C for 15 min. Samples of cultures were then withdrawn and diluted 10x for absorbance measurements (A_{600}) and supernatant activity measurements to estimate cell density and the amount of enzyme leaking into the growth medium. Harvested cell pellets were stored at -25°C until use. In the case of $VPR_{\Delta C}$, and especially for the less stable variants of VPR, the addition of $CaCl_2$ to a final concentration of 100 mM has been observed to be highly beneficial [3]. Thus, all expression cultivations of VPR and its variants in this improved expression system have contained 100 mM $CaCl_2$ and 76 µM L-rhamnose, yielding upwards of one hundred times more units of $VPR_{\Delta C}$ and made the expression of unstable variants such as some of the tryptophan variants possible. Improved expression of AQUI in this system has not been achieved in this system at this time.

AQUI expression

Starter cultures of transformed BL21 cells in 2xYT broth containing 0.1 mg/mL were grown overnight at 37°C. Cultures were then diluted approximately 50x into 2xYT containing 0.1 mg/mL ampicillin and grown at 37°C until A_{600} reached 0.4 - 0.8 A.U. Expression was then induced by adding rhamnose to a final concentration of 0.2 % weight/volume and the expression cultivation made 10 mM $CaCl_2$. Cultures were grown at 37°C for approximately 6 hours, cells then harvested and stored as described for VPR.

2.2.4 Site directed mutagenesis

During this project, site directed mutagenesis has been carried out on VPR and AQUI and the list of primers and mutations is to be found in their relevant publications. Throughout this project, two methods of site directed mutagenesis have been utilized; Quikchange Site-directed Mutagenesis Kit from Stratagene and Q5 Site-directed Mutagenesis Kit form New England Biolabs. Presently, the Q5 system is almost exclusively used as it produces much higher amounts of modified plasmids as it does not rely on mutagenic primer pairs but uses non-overlapping primers that allows for exponential amplification instead of linear.

2.2.5 Purification

Purification of $VPR_{\Delta C}$

Harvested cell pellets were dissolved in 1/20 of the original culture volume in a buffer containing 25 mM Tris, 10 mM $CaCl_2$ and set to pH 8.0 at room temperature. To that sample, dissolved DNase and lysozyme were added to a final concentration of 1 µg/mL and 1 mg/mL, respectively. The sample was then gently shaken for 2 hours, or until the pellet was completely resuspended. This was followed by a flash freeze and thaw cycle using liquid nitrogen and thawed at room temperature with gentle shaking. This freeze/thaw cycle was repeated three times with the last thawing being carried out at 4°C overnight. The sample was then centrifuged at 20,000 xg at 4°C for 45 minutes. The supernatant was kept and made 80% saturated by ammonium sulfate, followed by another centrifuge step at 20,000 xg at 4°C for 45 minutes. The supernatant was discarded, and the precipitate dissolved in the same Tris buffer as before.

All chromatography steps were performed on a BioLogic LP workstation at 4°C and all buffer solutions used in the purification protocol contained 25 mM Tris, 10 mM $CaCl_2$ and were adjusted to pH 8.0 at room temperature beforehand. The first chromatography step consisted of an affinity column (N-carbobenzoxy-D-phenylalanyl-triethylenetetramine-Sepharose (z-D-Phe-TETA) column)), previously equilibrated with the Tris-buffer. After loading the sample onto the column, it was washed with 1 M NaCl until absorbance at 280 nm had stabilized. Prior to elution the column was washed with the buffer without NaCl. Elution was then performed with a 2 M guanidinium chloride solution and the eluted fractions diluted into 3 M ammonium sulfate in a 5 to 4 ratio. That sample was then immediately loaded onto a hydrophobic interaction column (Phenyl Sepharose), previously

equilibrated with 1 M ammonium sulfate. The ammonium sulfate concentration was then lowered to 0 M, in a gradient or stepwise manner, taking care to wash off contaminants that start eluting at or below 0.4 M ammonium sulfate. $VPR_{\Delta C}$ was then eluted by 50% ethylene glycol in Tris-buffer. The final step in the purification procedure consisted of an anion exchange column (Q-Sepharose) which was meant to concentrate the sample and to ensure that no RNA contamination was present [3]. The column was equilibrated with the Tris-buffer before the sample was loaded. The elution was then performed with the same buffer followed by elution using a linear NaCl gradient from 0 M to 0.5 M. Active portions were pooled together and made 20% in ethylene glycol before being flash frozen in liquid nitrogen. Samples were then stored at -25°C.

Purification was then confirmed by using Blue Silver stained [315] stacking SDS-PAGE (4%/12%). Specific activity of all samples was estimated by measuring activity against 0.5 mM Suc-AAPF-NH-Np and protein concentration was determined by using the Coomassie G-250 based assay as described by Zaman and Vervilghen [316]. $VPR_{\Delta C}$ samples purified to homogeneity migrated as a single band with an estimated molecular mass of 34 – 35 kDa as judged by SDS-PAGE and with estimated specific activity of around 800 – 900 U/mg.

Purification of AQUI

The purification of AQUI was carried out on a BioLogic LP workstation at 4°C, the purification protocol described in detail in [308]. The protocol consists of heat treatment at 70°C for an hour, ammonium sulfate precipitation, hydrophobic interaction column (Phenyl Sepharose) and cation exchange column (CM Sepharose). Purification to homogeneity confirmed by SDS-PAGE as described for VPR and samples stored in the same manner.

2.3 Activity assays

The standard conditions under which the kinetic parameters of VPR and its mutated variants were characterized, were as follows. Enzyme solutions were dialyzed against the assay buffer, that contained 100 mM Tris,10 mM $CaCl_2$ and set to pH 8.6 at 25°C, overnight at 4°C. Activity against the substrate Suc-AAPF-NH-Np dissolved in the assay buffer was monitored at 410 nm over 15 seconds at 25°C. Seven different substrate concentrations were used ranging between 0.1 mM and 1.0 mM. For each set of experiments, each substrate

concentration was measured in triplicates. Reported values are averages of at least three sets of measurements.

For protein concentration measurements, absorbance at 280 nm were utilized, using the molar attenuation coefficient 34170 $M^{-1}cm^{-1}$ for $VPR_{\Delta C}$ and 28670 $M^{-1}cm^{-1}$ in the case of the tryptophan variants [317]. For accurate substrate concentrations, assay solutions containing enzyme and substrate were incubated overnight at room temperature before diluting 10-fold and measuring absorbance at 410 nm, using the molar attenuation coefficient 8480 $M^{-1}cm^{-1}$ for calculations [318]. Results were then fitted to the Michaelis-Menten equation and parameters calculated. For estimation of kinetic parameters under non-standard conditions, activity assays were carried out on a Shimadzu UV-2700 spectrophotometer connected to a Shimadzu TCC-100 temperature controller and measuring against four different substrate concentrations, all done in triplicates.

2.4 Stability assays

2.4.1 Inactivation experiments

The standard method for inactivation experiments were as follows. Samples were prepared by dialysis overnight at 4°C against a buffer containing 25 mM Tris, 15 mM $CaCl_2$, 100 mM NaCl, 1 mM EDTA and tuned to pH 8.95 at 25°C. Samples were incubated at several selected temperatures and activity monitored against 0.5 mM Suc-AAPF-NH-Np withdrawing aliquots at timed intervals. The rates of inactivation were then plotted up in the form of an Arrhenius plot, where the activation energy of inactivation ($E_{act\ (inactivation)}$) is calculated along with the $T_{50\%}$ value. $T_{50\%}$ is defined as the temperature at which half of the activity is lost after 30 minutes. The half-life ($t_{1/2}$) at a chosen temperature and varying conditions was also measured and calculated for further comparisons. The samples were generally diluted 50-100-fold into the appropriate incubation buffer just prior to measurements and aliquots were withdrawn at intervals and assayed for remaining activity

2.4.2 Melting curves by circular dichroism (CD) (T_m)

To determine the stability of the secondary structure circular dichroism (CD) was utilized. Sample preparations were as follows. Samples were inhibited by PMSF by making the enzyme solution 2.5 mM in phenylmethylsulphonyl fluoride (PMSF) and incubating that

solution for 30 minutes at room temperature (the more unstable variants were incubated at 4°C for periods exceeding 30 minutes). Following that incubation period, the sample was dialyzed overnight at 4°C against the standard stability assay buffer that consists of 25 mM glycine, 15 mM $CaCl_2$ concentrations, 100 mM NaCl and tuned to pH 8.6 at 25°C. The secondary structure unfolding of VPR and variants was monitored on a Jasco J-1100 spectropolarimeter at 222 nm using a 1 mm cuvette. Protein concentration in these experiments was kept between 0.1-0.4 mg/mL and recorded between temperatures that made sure that proper baselines for the native and denatured state were established. For T_m determination, melting curves were normalized and then fitted to a sigmoidal curve from which the T_m value was calculated. Reported T_m values are all based on at least three separate experiments recorded at the same temperature gradient.

2.4.3 Protein stability determined by differential scanning calorimetry (DSC) ($T_{m\ (app)}$)

Differential scanning calorimetry (DSC) scans were carried out using a MicroCal VP-DSC. Sample preparation for DSC experiments was the same as for CD measurements. Protein concentrations were between 0.4–1.2 mg/mL and prior to loading, samples were degassed for 15 - 30 minutes. Background determination was done by recording the thermal history of the appropriate buffer overnight. Samples were then loaded in the cooling stage between 25°C and 10°C. Experimental setup did vary but the standard conditions were to scan from 15°C and until the proteins were fully unfolded followed by cooling and a reheat run recording the heat capacity of the denatured state over the temperature range. The standard assay conditions were 25 mM glycine, 15 mM $CaCl_2$ and 100 mM NaCl at pH 8.6 scanning with a rate of 1.0°C/min. To confirm unfolding models, scans were conducted using 0.5°C/min and 1.5°C/min, varying pH and calcium concentrations in addition to recording partial unfolding DSC thermograms, where heating and cooling rates were kept constant.

Initial data analysis consisted of buffer background subtraction, performed using the Origin software and unmodified data sets of native and denatured runs exported for further analysis. In addition, data sets of all native state thermograms were converted into plots of excess heat capacity versus temperature, by manually fitting the linear segments close to the peak and creating a baseline using the Origin software. The apparent melting point ($T_{m\ (app)}$) was calculated by fitting a cubic function to the peaks of the thermograms, solving the first derivative for the local maximum.

The unfolding process of $VPR_{\Delta C}$ is under kinetic control [4] and conforms to the two-state irreversible unfolding model:

$$N \overset{k_1}{\rightarrow} D$$

where N stands for the native state, D for the denatured one and k_1 is the first order rate constant of unfolding. The rate of unfolding (k_1) will then follow the Arrhenius equation and is calculated as:

$$k_1 = \frac{vC_p}{Q_t - Q}$$

where v is the speed of the temperature gradient C_p is the excess heat capacity at a given temperature, Q_t is the total heat evolved and Q is heat evolved at a given temperature [91]. The activation energy of unfolding (E_{act}) is then calculated from the slope of the Arrhenius plot using the calculated unfolding rates above 5% heat evolved and under 95% heat evolved to avoid the relative high uncertainty at the end and beginning of the of the transitions [91]. From the activation energy of unfolding the activation enthalpy is calculated:

$$\Delta H^{\ddagger} = E_{act} - RT$$

where R is the universal gas constant and T is the absolute temperature. The activation Gibbs free energy ($\Delta G^{\ddagger}$) is then calculated as:

$$\Delta G^{\ddagger} = -RT \ln\left(\frac{k_1 \times h}{k_b \times T}\right)$$

where R is the universal gas constant, T is the absolute temperature, k_1 is the rate of unfolding at a given temperature, h is the Planck constant and k_b is the Boltzmann constant. The activation entropy ($\Delta S^{\ddagger}$) can then be calculated as:

$$\Delta S^{\ddagger} = \frac{\Delta H^{\ddagger} - \Delta G^{\ddagger}}{T}$$

Deviations of the two-state model were observed in several variants and for the wild type at low pH values. The unfolding model used to describe deviating variants was a three-state irreversible model, were both transitions are under kinetic control:

$$N \overset{k_1}{\rightarrow} I \overset{k_2}{\rightarrow} D$$

where N stands for the native state, I represents the intermediate state, D the denatured state, k_1 stands for the first order rate constant of the first transition and k_2 for the first order rate constant of the second transition. We utilize the online tool CalFitter [319, 320] to analyze these more complex unfolding processes. The model CalFitter uses being a modification of the Arrhenius equation:

$$k = \exp\left(-\frac{E_{act}}{R}\left(\frac{1}{T} - \frac{1}{T_{act}}\right)\right)$$

where k is the rate of unfolding, E_{act} is the activation energy of unfolding, R is the gas constant, T is the absolute temperature and T_{act} is an expression of the preexponential factor A, that has been transformed into exponent with the single new parameter T_{act} for more robust parameter estimation. Activation energy parameters can then be calculated as described for the two-state model of each transition. Concurrent with DSC unfolding experiments, CD secondary structure melting measurements were carried out under the same conditions. CalFitter was thus used to simultaneously fit DSC and CD data for more accurate fit and correlations between unfolding events and secondary structure content of those events. Under our standard assay conditions many measured variants displayed unstable heat capacity traces post to the unfolding transition, indicative of aggregation. As of now, CalFitter cannot model aggregation events. Thus, comparisons between variants is based on fitting of deconvoluted DSC thermograms and normalized CD melting profiles. In this case $\Delta C_p^{\ddagger}$ is fixed as 0 kJ/mol as is the slope of the pre-baseline set to 0. Although baseline subtraction leads to a loss of information and possibly poorer precision in parameter estimation, our observation is that the accuracy holds and was improved when working with data sets that showed signs of slow aggregation appearing as downward sloping of post heat capacities after the unfolding transition.

2.5 Structure monitoring

2.5.1 Circular dichroism (CD)

Concurrent to the DSC and CD unfolding experiments, CD wavelength scans were also recorded on a Jasco J-1100 from 250 nm to 200 nm at 25°C using a 1 mm cuvette.

2.5.2 Fluorescence spectroscopy (FS)

Steady state fluorescence was recorded for each variant at 15°C, 25°C and 35°C using a Fluoromax-4 spectrofluorometer equipped with a circulating water bath for temperature control. All samples were inhibited with PMSF at a final concentration of 2.5 mM, followed by dialysis against 50 mM Tris, 10 mM $CaCl_2$ and pH 8.0 overnight at 4°C. Prior to fluorescence experiments absorbance spectra were recorded from 400 nm down to 220 nm and absorbance adjusted to 0.03 - 0.05 A.U. at 295 nm in a 0.4 cm quartz cuvette used for fluorescence experiments. In addition to recording native fluorescence of all variants, steady state fluorescence of the denatured state was also recorded for $VPR_{\Delta C}$, $VPR_{\Delta C}$_N3P/I5P, $VPR_{\Delta C}$_N3P/I5P/N238P/T265P and $VPR_{\Delta C}$_W6F, where samples were heated to 90°C for 15 minutes and fluorescence measured at 25°C. All samples were excited at 295 nm using 3 nm entrance slit width and fluorescence monitored between 310 nm and 450 nm using a 5-8 nm exit slit width for native samples and 2-3 nm for denatured samples. Relative fluorescence was then calculated as:

$$F_{n.} = \frac{\left(\frac{CPS}{[P] * exi.^2}\right)}{F_{\mathrm{VPR\Delta C}}}$$

where $F_{n.}$ is the normalized fluorescence intensity, CPS the recorded fluorescence intensity, [P] the protein concentration, exi. the exit slit width used and $F_{VPR\Delta C}$ the concentration and exit slit width normalized fluorescence for native $VPR_{\Delta C}$. The peak of each fluorescence spectrum was then fitted to a cubic function, solving the first derivative for the local maximum (λ_{max}). The area under each curve (AUC) was calculated for all variants via the trapezoidal rule and the relative emission efficacy calculated by dividing the results with the average fluorescence intensity for native $VPR_{\Delta C}$.

Acrylamide quenching experiments were also conducted on all variants, using a 2.5 M stock of molecular biology grade acrylamide. Sample preparation and experimental conditions

were as described for samples for steady state experiments, with the addition of recording steady state fluorescence at varying concentrations of acrylamide. After each aliquot addition of acrylamide, the solution was thoroughly mixed and followed by one min resting time for temperature equilibration. The effectiveness of quenching was calculated by fitting the data to the Stern-Volmer equation:

$$\frac{\mathrm{F^o}}{F} = 1 + K_{SV}[Q]$$

where F^0 and F are the fluorescence intensities in the absence and presence of quencher between 310 nm and 410 nm, [Q] is concentration of quencher and K_{sv} is the Stern-Volmer constant calculated by linear regression. Corrections of fluorescence intensities were performed on the data to account for dilutions due to additions of acrylamide.

2.5.3 Molecular dynamic simulations (MD)

MD simulations were carried out using the Gromacs software package (www.gromacs.org) on a parallel architecture computer utilizing the CHARMM22* force field [321] for simulations. To provide starting structures for the simulations, the X-ray crystal structure of VPR (PDB ID: 1SH7) [277] was modified using UCSF Chimera [311]. In-silico modifications of VPR consisted of C-terminal deletion by deleting Gly276 and succeeding residues. In-silico mutagenesis was performed by choosing a rotamer for the new residue that closely resembled the native rotamer of the native residue. Structures were soaked in TIP3P dodecahedral water with a minimum of 1 nm between the protein and the edge of the water box, all succeeding MD simulations were then conducted under the periodic boundary condition. Preparation of models was as follows: (i) steepest energy minimization of 10,000 steps; (ii) addition of two sodium ions to counter the system charge; (iii) steepest descent energy minimization of 10,000 steps; (iv) solvent equilibration for 100 ps at 300 K while restraining the protein and bound calcium atoms by a harmonic potential with a force constant of 1000 kJ $mol^{-1}nm^{-2}$; (v) pressure equilibration of the system without any positional restraints was conducted under NPT ensemble at 1 bar, 300 K for 4 ns using a Parrinello-Rahman barostat; (vi) temperature equilibration of the system to 300 K utilizing Berendsen thermostat under NVT ensemble; (vii) a 2 ns thermalization in NVT conditions at 300 K under isotropic pressure at 1 bar using the velocity-rescale thermostat [322]. The productive MD simulations were carried out for 500 ns, however for one of the variants simulated,

$VPR_{\Delta C}$_W6F, a 1000 ns run was needed to equilibrate, under the NVT ensemble using the velocity-rescale thermostat [322] with a coupling constant 0.2 ps at 300 K. For heavy atom bonds the LINCS algorithm was used setting both LINCS-order and LINCS-iter as 4 [323]. Van der Waals and Coulomb interactions were cutoff at 0.8 nm. Long-range electrostatic interactions were calculated using the Particle-Mesh Ewald (PME) summation scheme [324]. In depth analysis of MD simulations is not finished at this time.

2.6 Active site labeling

Active site labeling was carried out on $VPR_{\Delta C}$ and AQUI using 5-dimethylaminonaphtalene-1-sulfonyl fluoride or dansyl fluoride (DNSF). Dansyl fluoride is much less reactive than dansyl chloride, known for its reactivity against amines, phenols, thiols, imidazoles and alcohols. Dansyl fluoride however has rather specific reactivity against active alcohols such as is in the case of the catalytic serine residue of serine proteases [325]. DNSF is thus, a suitable label for the proteases studied here (Fig. 2.1).

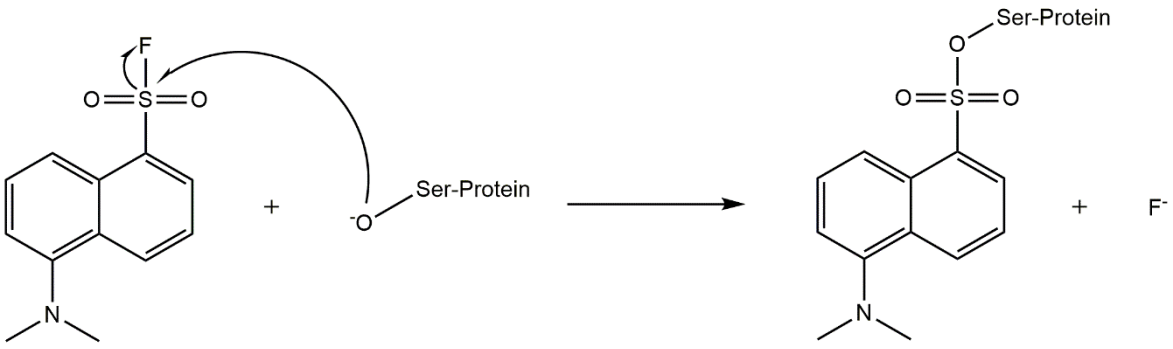


Figure 2.1. A reaction scheme for the reaction between the active site Ser residue of a serine protease with dansyl fluoride.

Optimization of the reaction condition was carried out by testing various buffers systems applicable between pH 7 and 9 and usable with calcium. Tests were performed by adding various concentrations of dansyl fluoride dissolved in 1,4-dioxane to a final volume/volume percentage of 5 % to the enzyme/buffer solutions. It was observed that imidazole and the Good´s buffers [326] tested resulted in poor labeling with apparent second order rates of inactivation, indicative of some side reactivity against the buffer. Sufficient labeling without side reactivity was achieved by dialyzing samples against a 25 mM sodium cacodylate, 10 mM $CaCl_2$ and adjusted to pH 7.0 at 25°C, overnight at 4°C. Following mixing of the enzyme solution to a stock solution of 0.005 mM dansyl fluoride the mixture was shaken gently in

the dark, at 25°C. The labeling efficiency was estimated by monitoring activity against 0.5 mM Suc-AAPF-NH-Np dissolved in the assay buffer. The reaction was quenched with the addition PMSF to a final concentration of 2.5 mM when estimated labeling efficiency exceeded 99%. Samples were then filtered to remove excess precipitated dansyl fluoride, followed by either extensive sequential dialysis, or buffer exchange by diafiltration using centrifugal filter units (Amicon) monitoring absorbance at 340 nm in the flow through until a steady value was achieved. To completely ensure that all free labels had been washed off the sample was subjected to a final dialysis against the Tris buffer used for steady state fluorescence overnight at 4°C.

The success of labeling was confirmed by absorbance scans from 450 nm to 220 nm, showing a new absorbance peak around 330 nm - 340 nm, as expected for a DNS protein derivative. To compare the absorbance spectra of labeled proteases and non-labeled, the approach was taken to calculate the molar attenuation coefficient spectra for the labeled proteins at 280 nm by adding the known molar attenuation coefficient of 1507 $M^{-1}cm^{-1}$ at 280 nm for dansyl-glycine in dioxane [327, 328] to the calculated molar attenuation coefficients for VPR and AQUI at 280 nm. This yielded the molar attenuating coefficients 35677 $M^{-1}cm^{-1}$ for VPR-DNS and 36137 $M^{-1}cm^{-1}$ for AQUI-DNS. Using this method, the maximum molar attenuation coefficient for VPR-DNS on the dansyl part of the spectrum was calculated as 3600 $M^{-1}cm^{-1}$ at 334 nm and for AQUI-DNS on the dansyl part of the spectrum as 3900 $M^{-1}cm^{-1}$ at 342 nm. These values are in good agreement with reported values for 1/1 labeling ratio of dansylated chymotrypsin having molar attenuation coefficient of 3360 $M^{-1}cm^{-1}$ at the peak of the dansyl absorbance peak [329]. In addition, CD wavelength scan of VPR-DNS in the Tris-buffer compares well with VPR, suggesting no noticeable changes in the secondary structure of the dansylated protease derivatives just prior to fluorescence measurements.

Steady state fluorescence was recorded of native and denatured samples as described in chapter 2.5.2, with the exception that the spectra were recorded between 580 nm and 310 nm when exciting at 295 nm. Steady state fluorescence was also recorded by excitation of

the dansyl moiety at 370 nm and recording was between 725 nm and 385 nm. For measurements of denatured AQUI-DNS, samples were autoclaved at 121°C for 15 minutes.

Fluorescence lifetimes of the dansyl labeled proteases were recorded on a Fluoromax-4 spectrofluorometer equipped with a time-correlated single-photon counter (TCSPC), using NanoLED-370 as a light source, having a peak emission at 369 nm and pulse durations of 1.2 ns. Emission was monitored at 500 nm and 580 nm for denatured samples and 580 nm for native samples. Signal collection times were kept under one minute by varying slit widths while collecting 10,000 counts. The instrumental response time was estimated by using the Tris-buffer as prompt, sometimes with added colloidal silica. Lifetimes were recorded from 5°C to 35°C for VPR-DNS and from 5°C to 75°C for AQUI-DNS, with at least 10°C intervals. Lifetime calculations were carried out using the decay analysis software (DAS6) included with the instrument from Horiba Scientific. Lifetimes of both VPR-DNS and AQUI-DNS exhibited complex decays, consisting of at least three lifetimes, two short and one long. Lifetimes of denatured samples of both VPR-DNS and AQUI-DNS exhibited one long lifetime and fitting the native data with three lifetimes fixing one of them as the lifetime of the denatured state resulted in good fits that were easily repeatable between batches. The lifetimes of the denatured forms of both VPR-DNS and AQUI-DNS were identical and had a linear relationship with temperature and may, therefore, indicate similar assemblies for both denatured proteases. The linear relationship could be used to calculate expected denatured lifetime to help fitting native data.

3 Results and discussions

This chapter will outline the projects that have been undertaken in this doctoral work and the results of published and unpublished work will be discussed.

3.1 Aims of the study

The general aim of the study was to gain a deeper insight into the temperature adaptation of the kinetically stable extremophilic subtilisin-like serine proteases VPR and AQUI. An important reason for this choice was the fact that these enzymes are structural homologs and are adapted to extremely different temperatures, that is reflected in their different catalytic activity and stability.

Of main interest was to elucidate the role of molecular flexibility in the temperature adaptation of these enzymes, especially regarding the role of prolines in loops and their effect on loop rigidity. This was to be done by mutational studies using the structure of AQUI as a template for the selection of mutation sites in VPR. AQUI contains four proline residues in loops that are not found in VPR. Thus, single and multiple proline variants were to be produced of VPR, culminating in the creation of a quadruple VPR proline variant (Fig. 3.1). Then, characterization of activity and stability of these constructs would be conducted. This included estimation of changes in structural flexibility as a result of these mutations. Flexibility changes were to be measured by steady-state and time-resolved fluorescence spectroscopy in addition to probing the effect of acrylamide quenching on fluorescence emissions and probing active site dynamics with labeling of the active site serine residue with a dansyl group. Supporting these experimental procedures, MD-simulations were to be carried out to gain deeper insight into the molecular dynamics of those variants.

In order to resolve native steady-state and time-resolved fluorescence spectra of the native VPR structure, tryptophan variants were to be produced where tryptophan residues were eliminated from the structure by exchanging all four tryptophan residues found in the structure of VPR (Fig. 3.1) to phenylalanine and resolving the properties of each variant and possibly multiple phenylalanyl variants.

The effects of these mutations were also to be studied via MD simulations using the information gained from them to help construct a clear picture of the effects of the different mutants produced.

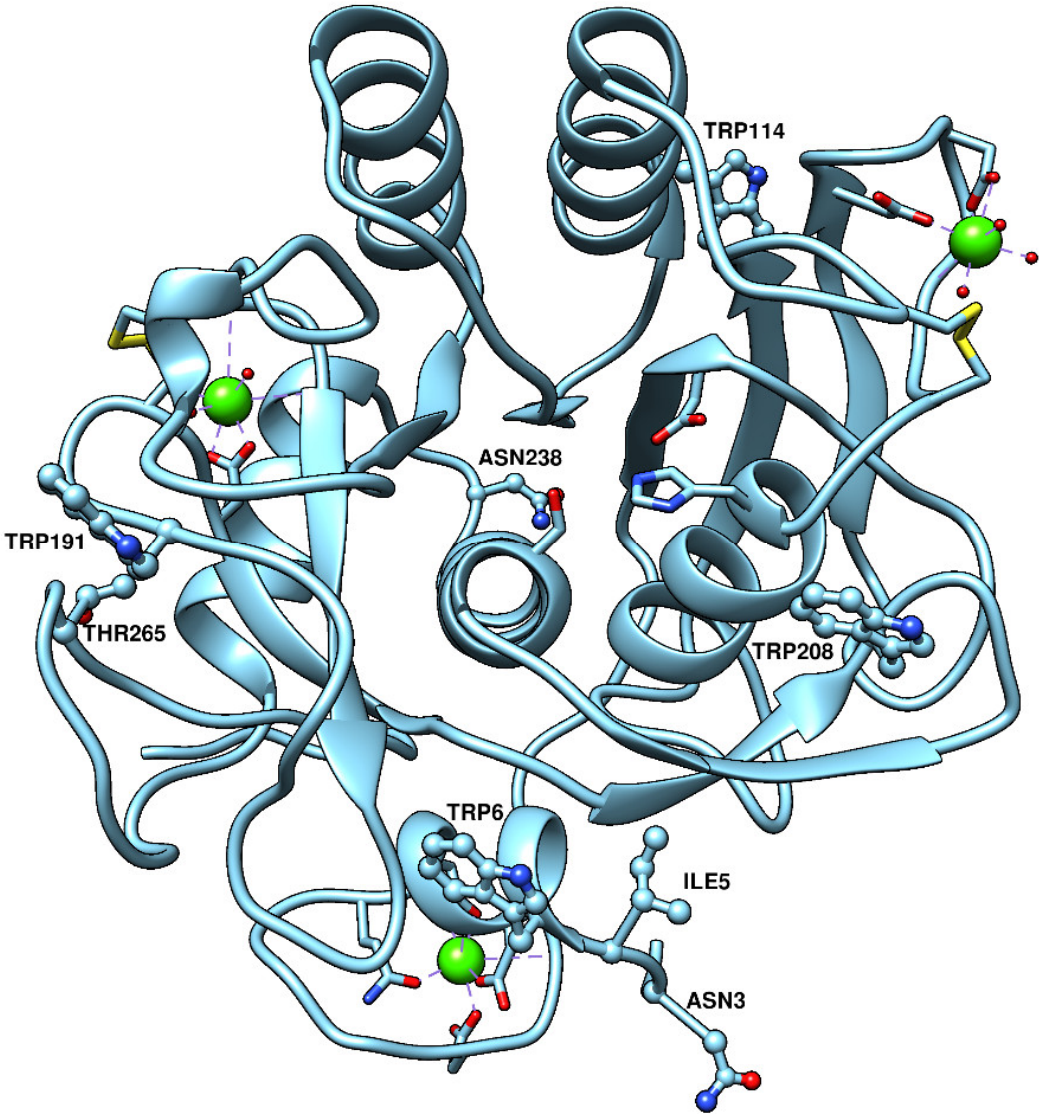


Figure 3.1. The crystal structure of VPR (PDB: 1SH7). Highlighted as ball and sticks are the Trp residues that were exchanged to Phe and Asn3, Ile5, Asn238 and Thr265 that were exchanged to Pro in this project.

Due to setbacks in production and characterization of some of these proposed variants, the project did steer into optimization of the production and purification systems. This work yielded much better system, for overexpressing VPR and its mutants, with a purification protocol that yielded a pure protease in high yields. The larger amount of protein produced opened the door for the utilization of differential scanning calorimetry that provided a wealth of information on the properties of the system. In addition to expressing useable yields of

variants, it made it feasible to obtain variants that were barely producible in the older system, an example being the tryptophan variants. In the following sections, the focus will largely be on the effects these VPR mutants had on the unfolding process and the role of calcium in the structure. This demonstrated the role of the calcium binding N-terminus as a major contributor to the stability of the protein structure and how it influences the cooperativity of the structure.

In the following sections, each of these projects will be discussed more closely, starting with the first paper, the reporting on the new and improved expression system and the role the different calcium binding sites have on the structure of the protease and how they can affect the unfolding cooperativity.

In the second section, the effects of the proline exchanges on the stability and activity will be discussed, focusing on the apparent loss of unfolding cooperativity as a result of these modifications. Furthermore, it will be discussed how that information can possibility be utilized to design an even more stable variant of VPR than the final product of that project, $VPR_{\Delta C}$_N3P/I5P/N238P/T265P, which did show a greatly improved resistance against thermal denaturation.

The third segment will discuss the current status of characterization of the Trp-to-Phe variants, with focus on the $VPR_{\Delta C}$_W6F mutant on which a manuscript is in preparation. $VPR_{\Delta C}$_W6F demonstrated some drastic detrimental effects on stability and unfolding cooperativity, seemingly causing a structural collapse at distant parts of the protein.

In the fourth and final section of this chapter, the current status of dansylation experiments of VPR and AQUI will be discussed, giving insights into dynamics of their active sites.

3.2 Improving the expression of VPR and the role of calcium in its stability

At the beginning of this doctoral work, all published work regarding VPR had either been carried out on the protease purified from cultures of the psychrotrophic *Vibrio* species (strain PA-44), or on the purified recombinant protease from *E. coli*, using a pBAD vector and Top10 cells [1, 277, 294, 302, 303, 305, 309, 310]. The plan was to continue using that system. However,

when the first proline and tryptophan variants were being produced it became apparent that sufficient amounts of some of the variants could not be obtained for the planned experiments.

Thus, the decision was taken to revamp the expression system. We opted to clone the gene of $VPR_{\Delta C}$ into a pET 11-a vector that utilizes the T7 expression system, known for efficient overexpression of cloned genes [330]. The chosen *E. coli* expression strain was Lemo21 [312], which contains the pLemo vector that has a T7 lysozyme under a rhamnose titratable promoter. As T7 lysozyme is a natural inhibitor of the T7 RNA polymerase [331], this system provides extra control over the rate of expression. During that work, it was observed that a mutation had occurred in the signal peptide of VPR, where an isoleucine had been replaced by a threonine in position ten of the signal peptide. Thus, prior to cloning the mutation T10I was incorporated and that led to around two- to four-fold increase in activity units recovered from cells. Likely, this was due to more efficient export into the periplasmic space where folding is not under as reducing conditions as in the cytosolic space [332].

The cloning was successful and work on optimizing conditions was started. During that work a mistake led to the discovery that high concentrations of calcium increased the yields of $VPR_{\Delta C}$ considerably. This was especially clear in the case of some tryptophan variants that were being worked on at the time. In some of those cases, $CaCl_2$ concentrations up to 100 mM did improve the yields of the active protease considerably. These observations prompted the interest in the effects of those high calcium concentrations on the stability of $VPR_{\Delta C}$. For $VPR_{\Delta C}$, the optimum conditions for expression were found to be in 2xYT broth, 76 µM rhamnose and calcium concentrations above 25 mM. Thus, for uniformity in the expression of all VPR variants, the conditions for the cells after inducing the expression contained 100 mM $CaCl_2$, 76 µM rhamnose, as they were shaken at 18°C for 18 – 20 hours, before harvesting. This optimized system yielded around 100 times more enzyme than the older system, in addition to opening the door to expressing variants that were non-expressible in the older system.

The first chromatography step in the purification protocol has been the affinity column N-carbobenzoxy-D-phenylalanyl-triethylenetetramine (TETA). The z-D-Phe ligand of the TETA column binds into the active site of the protease very tightly and to elute the column, 2 M guanidinium hydrochloride is required. As the column is thoroughly washed prior to elution, the peak observed in previous purifications was presumed to be just the protease.

However, when the protease expressed in the new system was eluted, it was very clear that the elution peak did consist of two separate peaks. One being sharp and eluting immediately, but the latter being very broad and eluting slowly. It also became clear, that activity coincided with the broad peak (See supplementary figure 2. of paper I). This indicated that previously observed sharp peaks in older purifications might be due to some unknown impurities.

These impurities appeared to be RNA fragments as observed from spectroscopic data of the contaminants (Fig. 3.2).

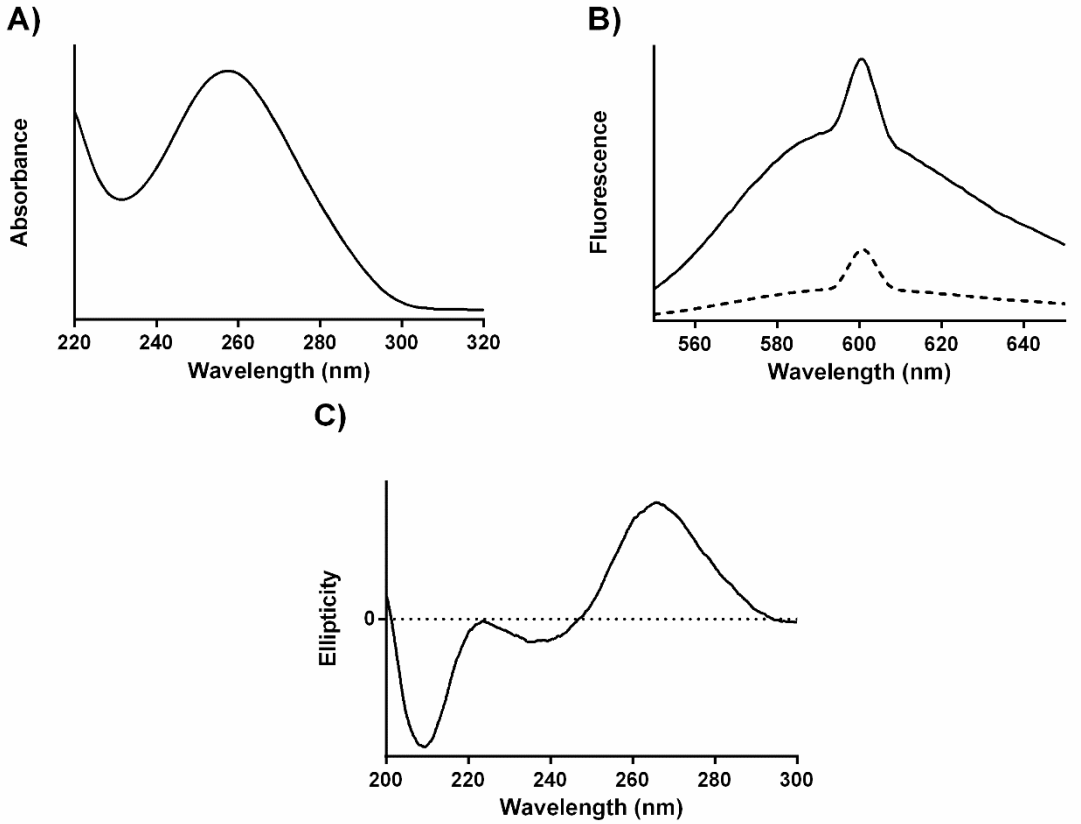


Figure 3.2. The A) absorbance spectrum of the contaminant isolated from the phenyl-sepharose column. B) Fluorescence emission spectra of contaminants binding ethidium bromide before (solid line) and after (dashed line) RNase A digestion, excited at 300 nm. C) Circular dichroism wavelength scans of the contaminants.

An agarose gel electrophoresis of the impurities indicated that these impurities were apparently two distinct RNAs in size range of around 100 nucleotides (data not shown). This realization and the utilization of a new expression system demanded a reestimation of the kinetic and thermostability parameters, as RNA may interfere with protein concentration estimation and have some interactions with the protein itself, which may affect stability. Reestimation of the enzymatic properties of the new recombinant $VPR_{\Delta C}$ showed that earlier

kinetic data had underestimated the turnover number greatly, or by about of 3.3 times lower (Table 3.1).

Table 3-1. Measured stability and kinetic parameters of recombinant $VPR_{\Delta C}$ from the new expression system utilizing the Lemo21 for expression and the stability and kinetic parameters of recombinant $VPR_{\Delta C}$ from the older expression system of Top10.

	T_m (°C)	$T_{50\%}$ (°C)	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M ($s^{-1}mM^{-1}$)
$VPR_{\Delta C}$ (Lemo21)	61.9 ± 0.4	53.8 ± 0.4	225.7 ± 12.0	0.178 ± 0.016	1238 ± 149
$VPR_{\Delta C}$ (Top10)	65.1 ± 0.2	56.4 ± 0.2	68.2 ± 17.0	0.184 ± 0.017	371 ± 26

No significant change was observed for the binding affinity (K_M) of the new recombinant enzyme for the substrate, as would be expected for contaminants not interacting with the enzyme. The effects found on stability were more puzzling, as the new recombinant $VPR_{\Delta C}$ was consistently measured with a T_m and a $T_{50\%}$ that was about 3°C lower than what had been reported for the older recombinant $VPR_{\Delta C}$. These results are still rather ambiguous as no explicit explanation has been found. If changes to the protease due to the change of the expression system are ruled out, these results suggest some sort of an interactions between the protease and the RNA in the solution. Initial testing of inactivation rates for $VPR_{\Delta C}$ in the presence and absence of added RNA did not show any significant effects on $T_{50\%}$. However, this has not been pursued fully, in terms of incubation times and different concentration of RNA in the experiments. So, interactions between these RNA fragments and the protease cannot be ruled out as of now.

As mentioned earlier, the fortuitous observation was made that an increase in the concentrations of $CaCl_2$ up to 100 mM in the growth medium led to increased yields of the active protease. This suggested that added calcium interacts with the protease in the periplasmic space and/or components inside the periplasmic space of the host cell. The focus was set on elucidating how high levels of calcium stabilize $VPR_{\Delta C}$. So, the effects of calcium concentration on melting points (T_m) of the PMSF inhibited enzyme, as well as on the rate of inactivation ($T_{50\%}$) of active samples, were measured. From these measurements, it was observed that concentrations of calcium in the low mM range did increase the melting point from 54.3°C, at 1 mM up to 62.2°C at 15 mM. From thermal inactivation experiments, however, the resistance against activity loss was still increasing at 250 mM. Inspection of the hyperbolic curve fitted to the half-lives at a constant temperature against calcium

concentration indicated that it was made of two separate curves (non-linear Scatchard plot: not shown) (Fig. 3.3).

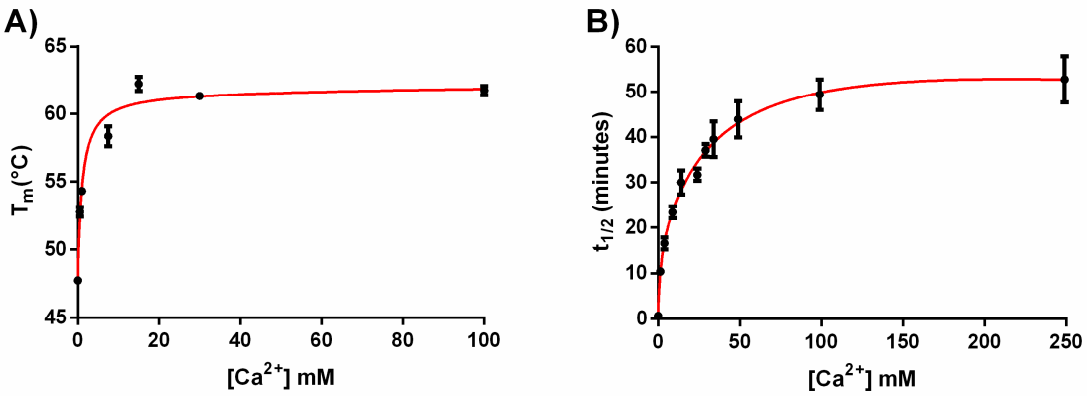


Figure 3.3. A) The effects of calcium ions on the melting point of PMS-VPR$_{\Delta C}$ as measured by CD. B) The effects of calcium inactivation rates of VPR$_{\Delta C}$. Data is shown with error bars that represent the standard deviations of measured values. Red lines represent the best fits of data, a single hyperbolic fit for figure A and a double hyperbolic fit for figure B.

The apparent binding constants (K_{app}) from these binding curves could be resolved into a single value of ~1 mM from the melting point data, and two apparent binding constants from the inactivation data with apparent binding constants (K_{app}) of about ~1 mM and ~22 mM. This strongly suggested that two distinct calcium binding sites that bind calcium in the mM range exist in the structure of VPR (Fig. 3.3). In addition, the stronger of these sites was clearly stabilizing the structure whereas the weaker one only protected against inactivation, presumably through protection against proteolysis. In our paper [3], by comparisons of crystal structures of VPR, AQUI and SPRK, and info from MD-simulations, it was deduced that the stronger site was the Ca-1 site, whereas the weaker one is the Ca-2 site. Supporting this suggestion are results from a study where the corresponding loop in subtilisin BPN′ was found to be the one of the first sites of proteolysis in that structure. By introducing a calcium binding site into that loop, the resistance of BPN′ against proteolysis was increased in the presence of calcium [275].

Examination of the effects of different calcium concentrations on the thermograms recorded on DSC showed that at higher calcium concentrations the unfolding transition becomes steeper and higher, indicating that the bound calcium in the Ca-1 site leads to higher unfolding cooperativity of the structure. However, at calcium concentrations above 15 mM, the activation Gibbs free energy barrier stopped increasing. This was due to a higher entropic penalty at higher calcium concentrations, presumably due to the fixation of the Ca-2 loop

(Fig. 3.3 and Fig. 3.4), restricting available conformations of the structure of the native assembly that are not restricted in the transition state.

3.2.1 Information on the unfolding process of VPR$_{\Delta C}$ from DSC

This work has opened the door to utilizing DSC and other experimental techniques requiring protein in high concentrations. In fact, information obtained on the unfolding process as recorded by DSC has become the cornerstone of this doctoral research. VPR$_{\Delta C}$ behaves, under slightly alkaline buffer conditions where the protease is active, like a typical two-state irreversible system (Fig. 3.5). As shown in Fig. 3.4. B. the thermogram of VPR$_{\Delta C}$ is free of aggregating events, making it a relatively good system to work with on DSC. This is best seen in the similarities of the heat capacity traces after the unfolding event, that are almost identical in the rerun of the sample, demonstrating that the unfolded assemblies in both runs are essentially identical, with minimal amounts of aggregated or degraded protein.

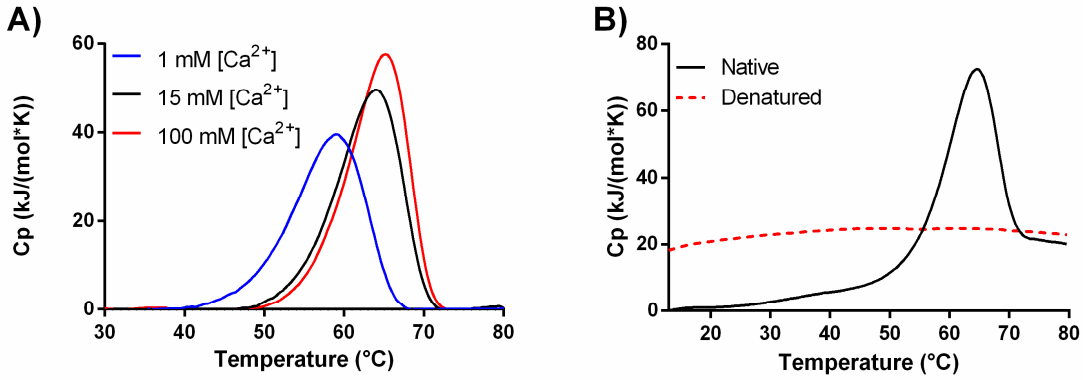


Figure 3.4. A) The effect of different calcium concentrations on the unfolding transition of VPR$_{\Delta C}$ as recorded on DSC, at pH 8.6, with a temperature gradient of 1°C/min. B) Non-normalized thermograms of native VPR$_{\Delta C}$ at 15 mM calcium and pH 8.6 (black line) and the second run of the sample (red dotted line) demonstrating irreversibility at the time-scale of the experiment.

From the above it is clear, that the ΔC_p change between the native and denatured states can be deduced from the difference of the heat capacity traces of the native and unfolded states pre-unfolding transitions, yielding a ΔC_p of around ~16 kJ/mol to ~ 18 kJ/mol. This is a rather high value but might be expected for a kinetically stable system. Note, that this does

not give any information on the $\Delta C_p^{\ddagger}$ of unfolding, but one would presume that value to be some degree lower than the difference between native and denatured assemblies.

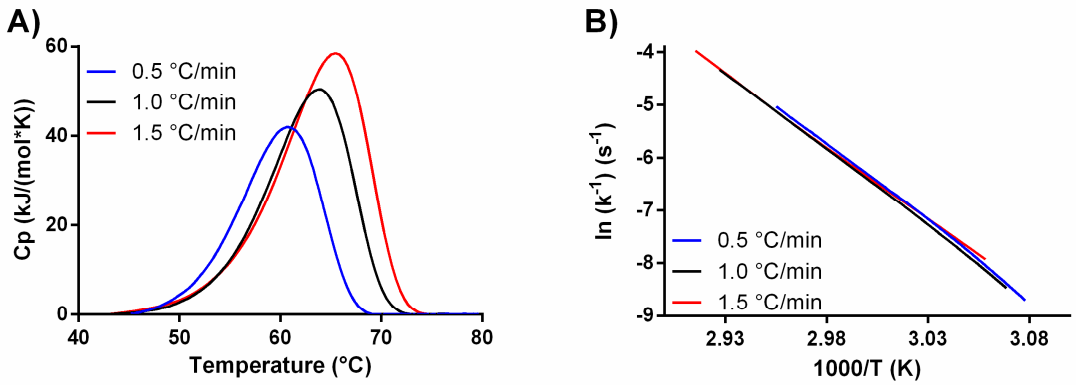


Figure 3.5. A) The effect of different scan rates on the unfolding transitions of VPR$_{\Delta C}$ as recorded by DSC at 15 mM calcium and pH 8.6. B) Arrhenius graphs calculated from unfolding thermograms, yielding a value for the activation energy of unfolding ~ 235 kJ/mol for all transitions.

VPR$_{\Delta C}$ consists of 276 amino acids and is, therefore, a relatively small protein. It can unfold in a more complex manner than described by the simple two-state irreversible model. This is seen for instance when the unfolding process is monitored at pH 5.0. At that pH, the thermogram showed a clear second unfolding event. This event became even clearer at low calcium concentrations (Fig. 3.6. A), as at pH 5.0 with 1 mM calcium a clear second peak became apparent, which is only slightly observable in the thermogram of VPR$_{\Delta C}$ at 15 mM calcium and pH 5.0 (Fig. 3.6. B).

From these results, it is interesting to see that the structure of VPR$_{\Delta C}$ which appears to be a single domain with an active site in the middle, is seemingly composed of two distinct structural parts. This is indicated by an apparent loss of unfolding cooperativity within the structure at pH 5.0 at low concentrations of calcium. The loss of unfolding cooperativity may be related to the possible protonation of acidic residues, such as Asp, many of which appear to be coordinators of calcium in the structure of VPR. It may be hypothesized that this effect is due to compromised binding of calcium within the structure.

Noting that the Ca-1 and Ca-3 sites appear to be the most relevant for the stability of the structure, either or both of these sites might have developed a reduced affinity for calcium at low pH. As discussed earlier in this thesis, interactions that enforce contacts between the C- and N-terminal parts of VPR tend to be important for thermostability. Thus, destabilizing either site might cause disconnection between the C- and N-terminal regions of the protein,

resulting in two peaks appearing in their thermograms. This appearance of a second peak is an exciting observation, as identifying these structural components could shed light on important contacts for thermostability within the protein structure.

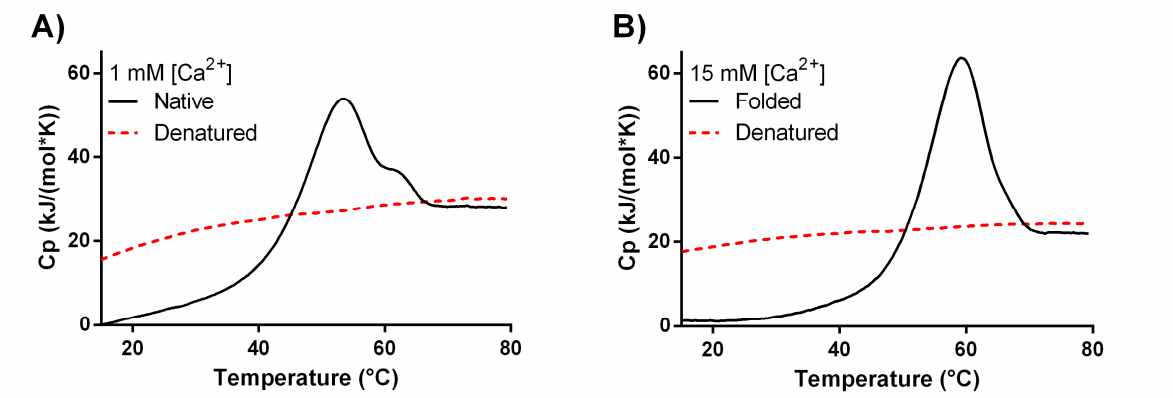


Figure 3.6. A) Non-normalized thermograms of native VPR$_{\Delta C}$ at 1 mM calcium and pH 5.0 (black line) and the second run of the sample (red dotted line) demonstrating irreversibility. B) Non-normalized thermograms of native VPR$_{\Delta C}$ at 15 mM calcium and pH 5.0 (black line) and the second run of the sample (red dotted line).

3.3 The role of proline residues in loops

In total, eight different VPR proline variants were constructed, comprising the single variants VPR$_{\Delta C}$_N3P, VPR$_{\Delta C}$_I5P, VPR$_{\Delta C}$_N238P and VPR$_{\Delta C}$_T265P, the double variant VPR$_{\Delta C}$_N3P/I5P, the triple variants VPR$_{\Delta C}$_N3P/I5P/N238P and VPR$_{\Delta C}$_N3P/I5P/T265P, and the quadruple variant VPR$_{\Delta C}$_N3P/I5P/N238P/T265P. As these were produced in the new expression and purification system, all the single variants and the double variant were recharacterized with respect to Michaelis Menten kinetic parameters (Table 3-2) and thermal stability (Table 3-3). Insertion of proline residues into the loops of VPR was found to have small effect on the catalytic efficiency of the mutants, contrary to what had been observed previously [304, 305]. There was a trend in the direction of higher K_m values and higher k_{cat} values with increased number of prolines inserted, but that resulted in similar catalytic efficiencies for different variants. The only exception was VPR$_{\Delta C}$_T265P, although with no explicit explanation. This trend in the catalytic properties, being most prominent in the VPR$_{\Delta C}$_N3P/I5P/N238P/T265P variant, is interesting as it goes against the idea of a rigid active site having lower K_m values. It implies that the accumulative effects of proline exchanges cause a destabilization of the ES complex, not something that would be expected if these mutations cause rigidification around the active site.

Table 3-2. The kinetic parameters of all proline variants measured proline variants at 25°C and pH 8.6.

Variant	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M ($s^{-1}mM^{-1}$)
$VPR_{\Delta C}$	225.7 ± 12.0	0.177 ± 0.016	1238 ± 149
$VPR_{\Delta C}$_N3P	235.4 ± 21.8	0.173 ± 0.013	1364 ± 60
$VPR_{\Delta C}$_I5P	201.6 ± 8.2	0.187 ± 0.010	1077 ± 37
$VPR_{\Delta C}$_N238P	224.6 ± 16.6	0.189 ± 0.026	1196 ± 84
$VPR_{\Delta C}$_T265P	166.5 ± 11.6	0.152 ± 0.019	1101 ± 104
$VPR_{\Delta C}$_N3P/I5P	231.8 ± 10.5	0.187 ± 0.009	1243 ± 77
$VPR_{\Delta C}$_N3P/I5P/N238P	229.5 ± 18.3	0.199 ± 0.024	1158 ± 72
$VPR_{\Delta C}$_N3P/I5P/T265P	221.5 ± 7.8	0.219 ± 0.014	1017 ± 74
$VPR_{\Delta C}$_N3P/I5P/N238P/T265P	259.3 ± 27.4	0.212 ± 0.014	1222 ± 95

Table 3-3. Stability parameters of measured proline variants, showing the $T_{50\%}$ values and $E_{act\ (inactivation)}$ calculated from inactivation Arrhenius graphs. The melting point as measured by CD and fitted to a single sigmoidal curve and the melting point determined from DSC defined as the highest point of each graph.

Variant	$T_{50\%}$ (°C)	$E_{act\ (inactivation)}$ (kJ/mol)	$T_{m\ (CD)}$ (°C)	$T_{m\ (DSC)}$ (°C)
$VPR_{\Delta C}$	53.8 ± 0.4	218 ± 9	61.9 ± 0.4	63.9 ± 0.3
$VPR_{\Delta C}$_N3P	56.6 ± 0.3	203 ± 12	64.8 ± 0.1	66.8 ± 0.3
$VPR_{\Delta C}$_I5P	56.1 ± 0.2	199 ± 14	65.1 ± 0.2	65.7 ± 0.5
$VPR_{\Delta C}$_N238P	52.3 ± 0.2	209 ± 17	60.7 ± 0.1	63.6 ± 0.2
$VPR_{\Delta C}$_T265P	54.3 ± 0.2	206 ± 4	61.6 ± 0.2	64.5 ± 0.2
$VPR_{\Delta C}$_N3P/I5P	60.3 ± 0.4	208 ± 8	67.8 ± 0.3	72.0 ± 0.8
$VPR_{\Delta C}$_N3P/I5P/N238P	60.9 ± 0.5	207 ± 27	68.8 ± 0.2	72.0 ± 0.4
$VPR_{\Delta C}$_N3P/I5P/T265P	62.2 ± 0.6	194 ± 27	69.2 ± 0.2	73.6 ± 0.3
$VPR_{\Delta C}$_N3P/I5P/N238P/T265P	61.6 ± 0.6	182 ± 29	72.1 ± 0.3	77.2 ± 0.2

The effects of the proline mutations on stability demonstrated that the N-terminal prolines were the most effective in increasing the stability of VPR, whereas the N238P and T265P variants were not effective, unless they were added on the $VPR_{\Delta C}$_N3P/I5P variant. This

indicates that the stability of the loops and neighboring structural regions that N238P and T265P were proposed to rigidify are dependent on the stability of the N-terminus. The overall conclusion of these results is, therefore, that proline residues in loops increase the kinetic stability of these proteins. The likely mechanism is that the increased rigidity caused by the mutations increases the thermal flexibility of the structure. Thus, by maintaining the correct orientations of other interactions at higher temperatures, the proline residues act as some sort of anchor points at these sites in the structure. However, this increased stability was accompanied by loss in unfolding cooperativity, for most of the proline variants.

3.3.1 Loss of unfolding cooperativity in proline variants

Loss of the single step unfolding process was observed in all proline variants, except for $VPR_{\Delta C}$_N238P and $VPR_{\Delta C}$_T265P. The single N-terminal proline variants both showed a clear disconnect between structural parts within the protein during unfolding, as shown for $VPR_{\Delta C}$_I5P in Fig.3.7.

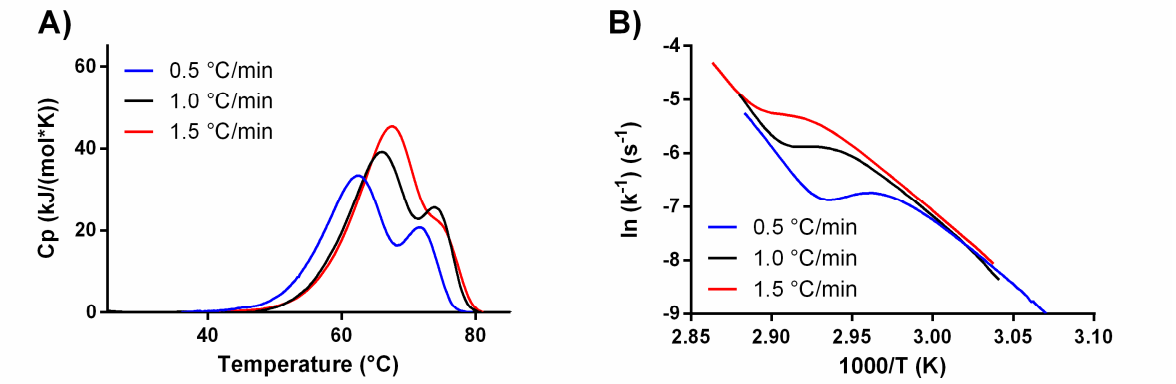


Figure 3.7. A) The effect of different scan rates (0.5°C/min: blue, 1.0°C/min: black and 1.5°C/min: red) on the unfolding transitions of $VPR_{\Delta C}$_I5P recorded by DSC at 15 mM calcium and pH 8.6. B) Arrhenius graphs calculated from the unfolding thermograms from figure A).

This suggests that the stability of the N-terminal region is crucial for maintaining the rest of the structure in the native conformation. Consequently, it is possible to overstabilize the N-terminal region resulting in a two-step unfolding process. $VPR_{\Delta C}$_I5P is a prime example of such an effect (Fig. 3.7). In the DSC thermograms of $VPR_{\Delta C}$_I5P, two very distinct peaks appeared and according to Arrhenius graphs calculated form the unfolding rates at different scan-speeds, both transitions are under kinetic control (Fig. 3.7 B). Thus, there is a relatively stable intermediate in the unfolding process of $VPR_{\Delta C}$_I5P (Fig. 3.8).

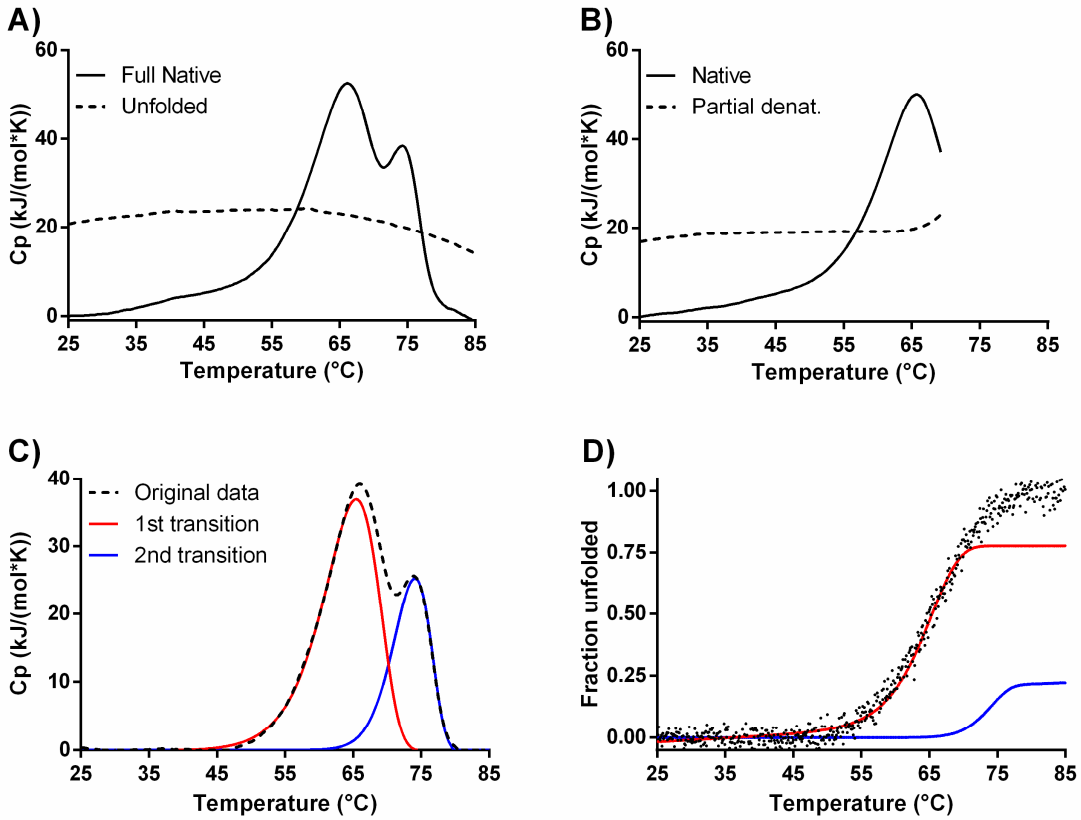


Figure 3.8. A) The unfolding process of VPR$_{\Delta C}$_I5P (solid line) as monitored by DSC at pH 8.6 and 15 mM, recorded with a 1°C/min temperature gradient and a rescan of the same sample (dotted line). B) The partial unfolding of VPR$_{\Delta C}$_I5P (solid line) and the sequential reheating of the sample demonstrating the irreversibility of the first event. C) Normalized DSC thermogram of the unfolding of VPR$_{\Delta C}$_I5P (dotted line). D) Normalized CD melting curves of VPR$_{\Delta C}$_I5P (black dots). The red and blue line represent the best fits of the unfolding process by CalFitter.

The nature of this intermediate in the case of VPR$_{\Delta C}$_I5P could partly be uncovered by the comparison and fitting of DSC and CD unfolding data (Fig. 3.8). From that comparison, it could be deduced that the first transition consists of around 70 % to 80 % of the secondary structure and two-thirds of the calorimetric enthalpy. This paints the picture of an intermediate that is quite energetic with a low secondary structure content, possibly due to a calcium binding site still being present. Considering the location of the mutation, and how the unfolding process of VPR$_{\Delta C}$ is affected by calcium binding (see chapter 3.2), we have come up with the hypothesis that the intermediate consists of the N-terminal region containing the Ca-3, site in addition to some other secondary structures. These may well be helix-E, containing the active serine residue, an idea which is supported by unfolding MD-simulations published by another group [333]. However, proper isolation and structural characterization of this intermediate has not been carried out yet.

One very relevant observation from these DSC experiments of variants containing N-terminal proline mutations, is that there is increased kinetic stability but an apparent loss of unfolding cooperativity. This loss of unfolding cooperativity may explain lower activation energies of inactivation observed for proline variants (Table 3-3). A prime example is the final product with a quadruple proline replacement $VPR_{\Delta C}$_N3P/I5P/N238P/T265P, that did not show improved resistance against thermal inactivation ($T_{50\%}$) compared to the triple variants despite showing increased resistance against thermal denaturation (T_m). This could be interpreted as follows; the loss of molecular contacts along important protein-protein interfaces at high temperatures allow parts of the structure to become more flexible and prone to auto-proteolysis. However, by strengthening contacts at this presently unknown interface in $VPR_{\Delta C}$_N3P/I5P/N238P/T265P, thermostability could be increased even further. That variant would unfold cooperatively with a melting point of the secondary structure that is closer to the observed highest peak of the DSC thermogram, that might result in greatly enhanced resistance against inactivation. That project that would need the crystal structure of $VPR_{\Delta C}$_N3P/I5P/N238P/T265P along with unfolding MD-simulations, in order to pinpoint the interface that needs stabilization. By comparing that area of VPR with the structure of AQUI, a mutational strategy could be implemented to stabilize the interface that divides the VPR structure into separate unfolding units.

3.3.2 Structural rigidity of proline variants

Structural flexibility of proline variants was probed by using acrylamide fluorescence quenching. The results indicated restrictions of movements/accessibility in mutants containing the N-terminal mutations, which would be in line with a restricted access to Trp6. The N238P and T265P mutations however, apparently caused an increase in accessibility, resulting in a similar quenching profile for $VPR_{\Delta C}$_N3P/I5P/N238P/T265P and $VPR_{\Delta C}$ at 25°C (Fig. 3 in paper II and supplementary Table 3 of paper II) (Fig. 3.9). However, as will be discussed later, the fluorescence properties of $VPR_{\Delta C}$ are rather convoluted. This stems from the fact, that the enzyme contains four tryptophan residues that are highly intrinsically quenched in the native structure (Fig. 3.11). This means, that minor changes in the average rotameric position of a single tryptophan residue might cause large changes in the fluorescence properties of the protein due to relief of the suppression that prevented the emission of light. Thus, increased acrylamide quenching might not necessarily indicate that

the structure has become more flexible but rather that a tryptophan residue does not reside in an intrinsically quenched position as much as in other variants.

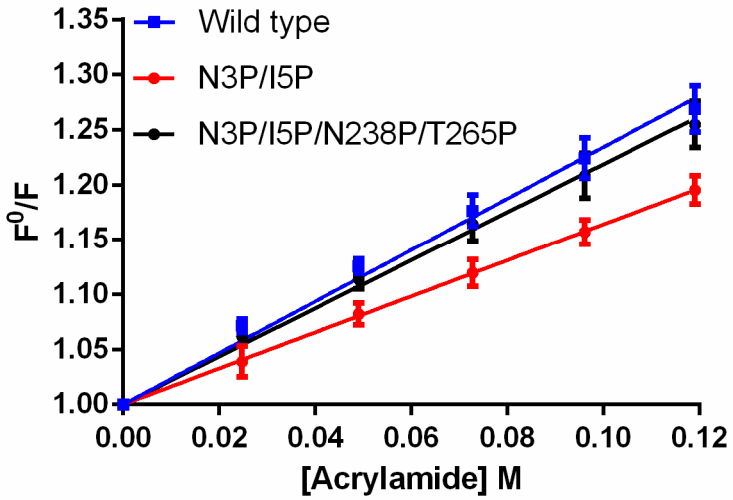


Figure 3.9. Stern-Volmer graphs calculated from fluorescence quenching data of $VPR_{\Delta C}$, $VPR_{\Delta C}$_N3P/I5P and $VPR_{\Delta C}$_N3P/I5P/N238P/T265P, between 310 nm – 410 nm.

A noteworthy observation that may be related to increased global rigidity of the protein structure of proline variants, especially in the case of the quadruple variant is derived from their DSC thermograms. The fact that $VPR_{\Delta C}$ and $VPR_{\Delta C}$_N3P/I5P/N238P/T265P both produce well behaved DSC thermograms allowed for direct comparisons of the heat capacities of the native and denatured assemblies over a broad temperature range (Fig. 3.10). What is apparent from these two runs, is that the difference between the heat capacities of the native and denatured states is much higher in the case of $VPR_{\Delta C}$_N3P/I5P/N238P/T265P compared with wild type. This means, that either the denatured assembly of the proline variant has a higher heat capacity than the denatured assembly of $VPR_{\Delta C}$, or that the heat capacity of native state of the proline variant is lower than that of $VPR_{\Delta C}$.

The unfolded assemblies of the two have higher heat capacities, i.e. are able to distribute heat better, than their highly structured native states. In case of the proline variants, the conformational restrictions caused by the inserted residues could reduce the ability of the peptide chain to distribute heat throughout the water-protein system, hence it would lead to lower heat capacities of the native and denatured assemblies. Disregarding differences in the unfolded assembly between the variants for simplification, it would suggest that the native state of $VPR_{\Delta C}$_N3P/I5P/N238P/T265P has a lower heat capacity than the native state of $VPR_{\Delta C}$, possibly even lower if the assumption about the denatured assembly is correct. This

would fit the picture of a structure that is more rigid, with restricted molecular movements and that cannot distribute the heat as well in the system.

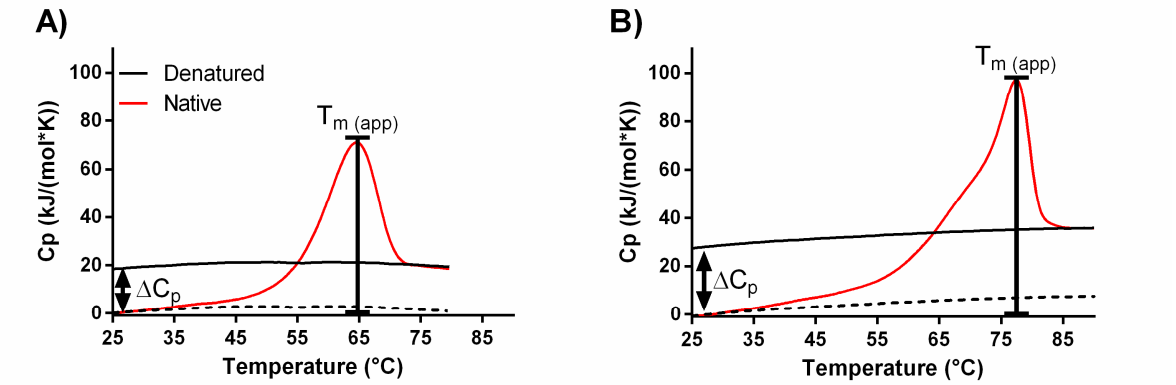


Figure 3.10. Non-normalized thermograms of native A) VPR$_{\Delta C}$ and B) VPR$_{\Delta C}$_N3P/I5P/N238P/T265P (red lines) and a second run of same samples (black line) recording the heat capacities of the denatured assembly. Black dotted lines are for highlighting the differences in heat capacity of the native state as a function of temperature compared to how the unfolded assembly changes with temperature.

Adding to this is the non-zero heat capacity change observed prior to the unfolding transition of VPR and its variants, possibly indicating increased mobility within the structure as a function of temperature (Fig. 3.10). This pre-transition heat absorption appears to be a more prominent feature of VPR$_{\Delta C}$_N3P/I5P/N238P/T265P than the wild type VPR. This would be suggestive of a structure that is able to absorb more thermal energy leading to increased flexibility before unfolding and is in line with the idea of increased thermal flexibility. However, to confirm speculations such as these, other experimental techniques are needed, such as NMR, fluorescence anisotropy, neutron scattering, deuterium exchange or MD simulations.

3.4 Mapping the fluorescence attributes of VPR

In the structure of VPR$_{\Delta C}$, there are four tryptophan residues, W6, W114, W191 and W208. All of these, except for W191, are found in AQUI, which contains a tyrosine residue at this location. This residue has been exchanged for tryptophan which showed minimal effects on thermal stability, but did indicate some loss of catalytic activity [310]. In the new expression system using high calcium concentrations, all Trp-to-Phe variants of VPR were expressible in sufficient yields. Initial characterization with respect to kinetics and stability were carried out on all tryptophan variants. All Trp single variants did exhibit some loss of stability, especially the W6F and W191F variants. The results for the W191F variant were quite

surprising, considering that this tryptophan residue is not found in AQUI, instead AQUI contains a tyrosine in that position. The W6F variant was of special interest to this project being located at the N-terminal region, proximal to sites 3 and 5 as discussed previously. The activity parameters did follow a trend of lower k_{cat} values and lower K_M values, except for the W208F variant where the K_M value was measured higher (Table 3-4). This was not investigated in more detail, but we also produced Trp-to-Tyr variants in the older expression system, but in very low yields. Initial data with that material suggested that the W208Y variant had a K_M value around two times higher than $VPR_{\Delta C}$. This result relates this residue to substrate binding, and thus provides an angle that might be beneficial to study, i.e. a site which can be monitored to learn more about how these structures maintain high catalytic efficiencies.

Table 3-4. The stability parameters T_m and $T_{50\%}$ and the Michaelis-Menten parameters of all Trp-Phe variants.

Variant	$T_{m\ (CD)}$ (°C)	$T_{50\%}$ (°C)	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M ($s^{-1}mM^{-1}$)
$VPR_{\Delta C}$	61.9 ± 0.4	53.8 ± 0.4	225.7 ± 12.0	0.178 ± 0.016	1238 ± 149
$VPR_{\Delta C}$_W6F	49.0 ± 0.6	40.7 ± 2.0	150.9 ± 31.9	0.145 ± 0.028	1098 ± 167
$VPR_{\Delta C}$_W114F	60.8 ± 0.2	52.5 ± 0.1	179.9 ± 21.9	0.151 ± 0.026	1249 ± 119
$VPR_{\Delta C}$_W191F	47.2 ± 0.1	37.5 ± 0.1	182.2 ± 17.9	0.116 ± 0.015	1583 ± 138
$VPR_{\Delta C}$_W208F	59.9 ± 0.2	52.3 ± 0.2	201.0 ± 32.4	0.231 ± 0.044	874 ± 57

3.4.1 Fluorescence properties of Trp variants

As the main objective of this part of the study was to map the contribution of each tryptophan residue to fluorescence of $VPR_{\Delta C}$, the steady state fluorescence of the native and denatured assemblies was measured in addition to acrylamide fluorescence quenching. However, it turned out that the native fluorescence of $VPR_{\Delta C}$ is highly intrinsically quenched, resulting in a native structure that is about ten-times less fluorescent than its denatured state (Fig. 3.11). This entails that tryptophan fluorescence is highly sensitive to unfolding, or any structural changes in the microenvironment of the tryptophan residues.

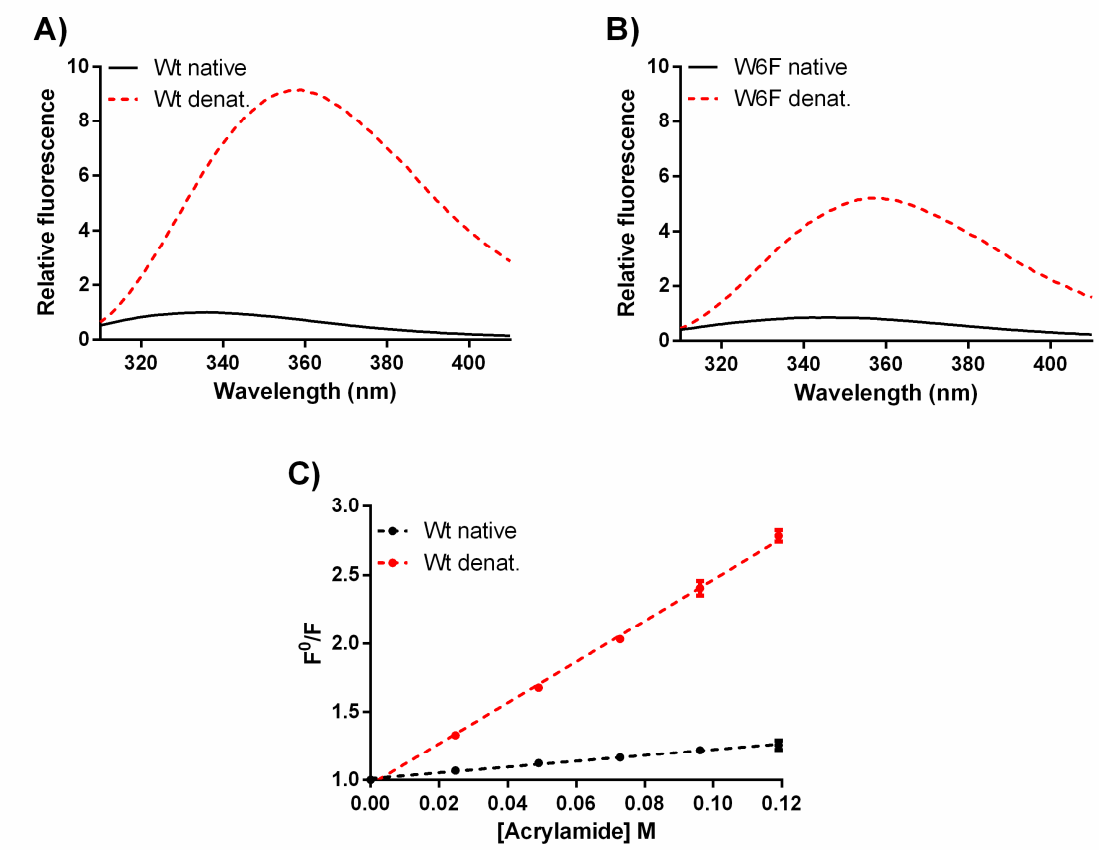


Figure 3.11. Comparisons of steady state fluorescence of the native (black line) and denatured (dotted red line) states of A) $VPR_{\Delta C}$ and B) $VPR_{\Delta C}$_W6F. C) Stern-Volmer graphs calculated from fluorescence quenching by acrylamide of the native (black dotted line) and the denatured state (red dotted line) of $VPR_{\Delta C}$.

As seen in Fig. 3.12, all the variants that lost a Trp for Phe cause some blue shift in emission that would suggest a more polar environment and usually some lowering of emitted light. However, these spectra do not allow for any resolution of the contribution of the fluorescence signal of the residues. Acrylamide quenching suggested a more permeable structure, the difference being most prominent for the two most unstable tryptophan variants W6F and W1191F (Fig. 3.13).

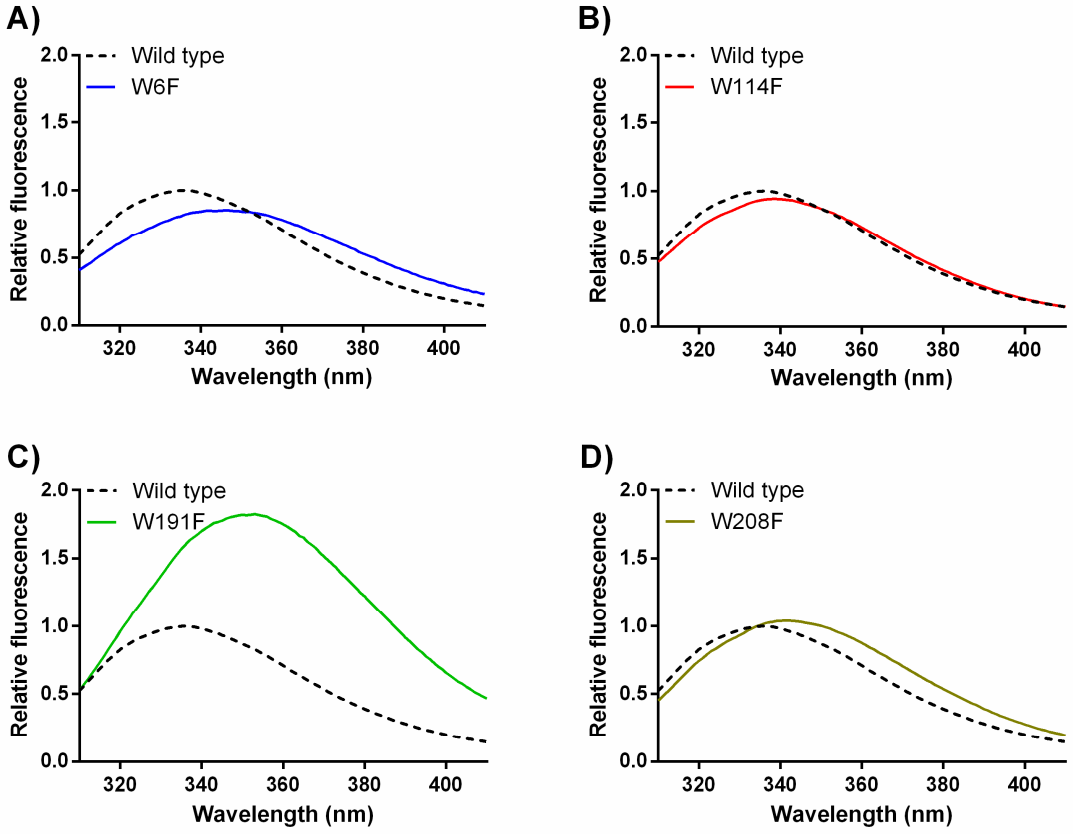


Figure 3.12. Steady state fluorescence emission spectra of $VPR_{\Delta C}$_Trp-Phe variants at 25°C at pH 8.0 of the native states of A) W6F (blue line), B) W114F (red line), C) W191F (green line) and D) W208F (golden line) compared to $VPR_{\Delta C}$ (dotted black line).

It was decided that since mapping the fluorescence properties of the protein was unattainable, the unstable variants would be investigated further. Most of the focus in the next section will be on $VPR_{\Delta C}$_W6F due to its location close to the N-terminus of the molecule, a part of the protein that has been shown to have a high impact on the stability of the protein structure.

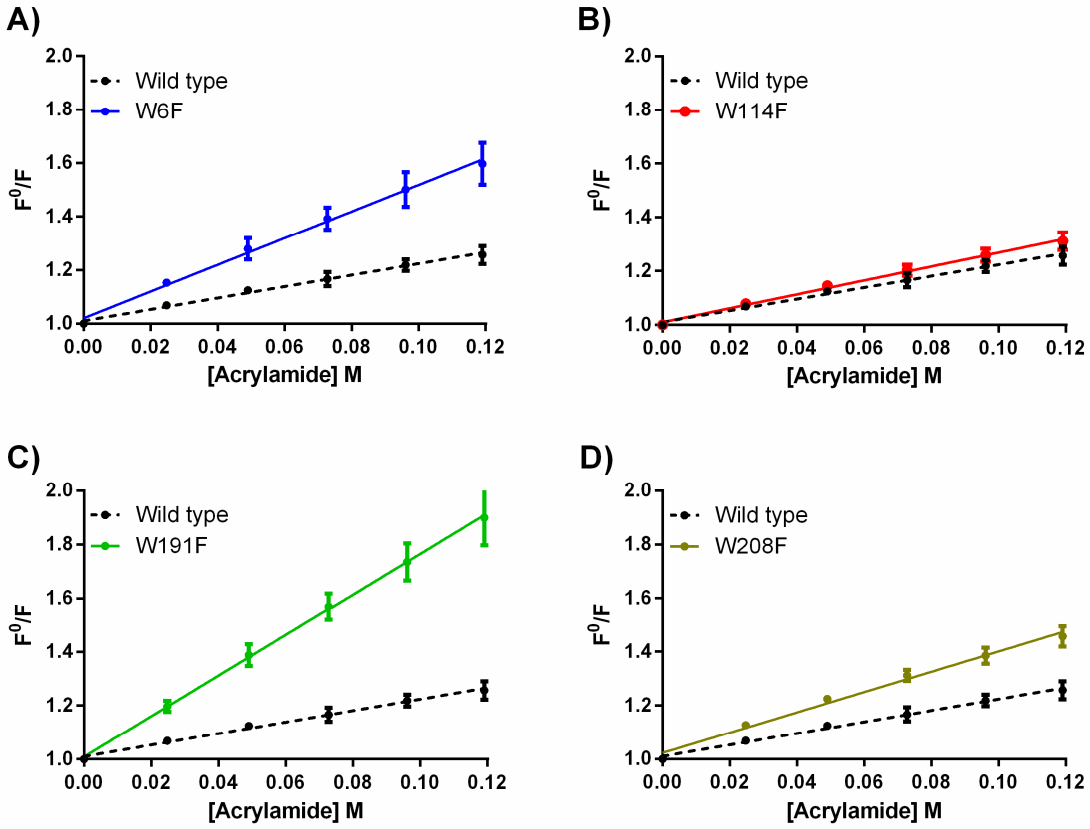


Figure 3.13. Stern-Volmer graphs for Trp variants of VPR. Acrylamide fluorescence quenching determined at 25°C at pH 8.0 for the native state of A) W6F (blue), B) W114F (red), C) W191F (green) and D) W208F (gold) compared to VPR$_{\Delta C}$ (black dotted line).

3.4.2 The VPR$_{\Delta C}$_W6F variant

VPR$_{\Delta C}$_W6F was studied and characterized in further detail due to its high impact on the structure and the stability of the protein. The mutation caused a loss of unfolding cooperativity and stability, but the variant exhibited similar looking thermograms as the N-terminal proline mutations did (see chapter 3.3) (Fig. 3.14 and 3.15). It is noteworthy, that the DSC and CD melting data showed that approximately 80 % of the secondary structure content was lost in the first transition (Fig. 3.15), almost the same as was found for the I5P variant (Fig. 3.8). This further supports the idea that the observed transitions consist of similar protein unfolding units.

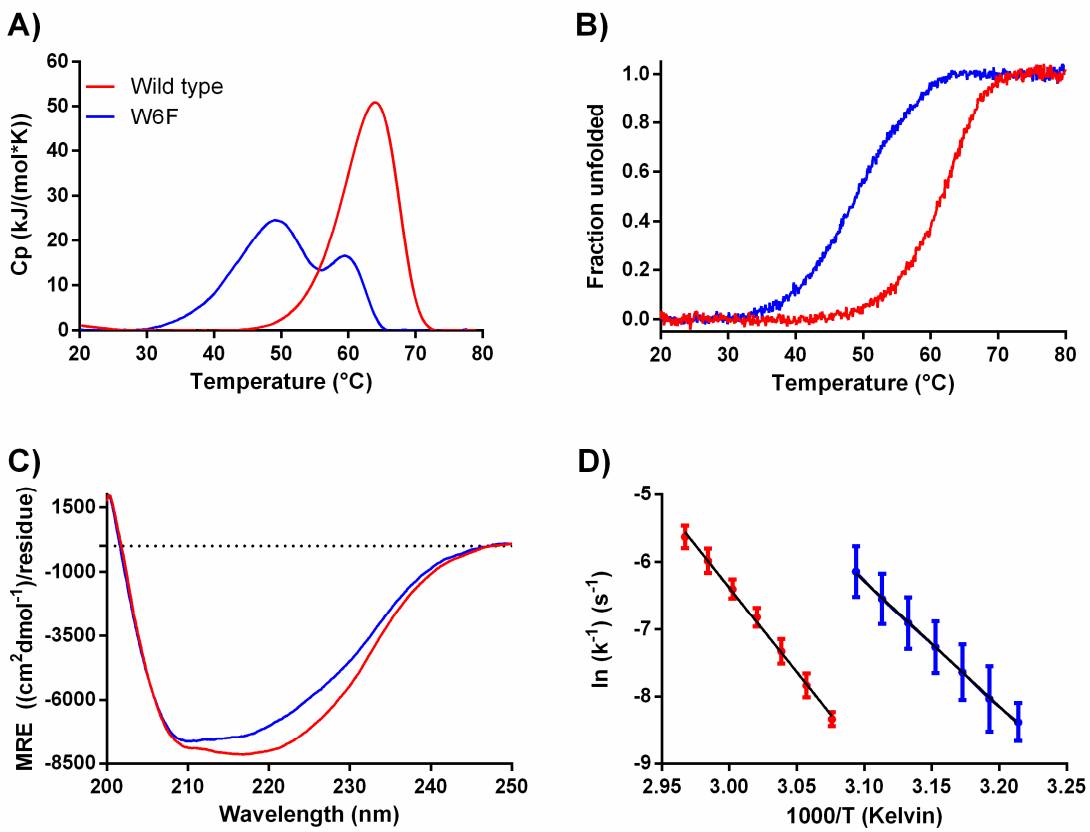


Figure 3.14. A) A comparison of normalized DSC thermograms of VPR$_{\Delta C}$ (red) and VPR$_{\Delta C}$_W6F (blue). B) Normalized CD melting curves of VPR$_{\Delta C}$ (red) and VPR$_{\Delta C}$_W6F (blue). C) Mean residual ellipticity of VPR$_{\Delta C}$ (red) and VPR$_{\Delta C}$_W6F (blue) measured by CD. D) Arrhenius graphs calculated from the rate of inactivation for VPR$_{\Delta C}$ (red) and VPR$_{\Delta C}$_W6F (blue).

In addition to being very destabilizing, the W6F modification is the first substitution that seems to have some real impact on the secondary structure of the protein. According to CD-wavelength scans, there was apparently reduced ellipticity in the case of the variant compared with the wild type, particularly between 220 nm and 230 nm (Fig 3.14. C). This part of the spectrum chiefly reports on α-helical content within the structure. This loss in secondary structure content apparently decreases intra-molecular contacts resulting in destabilization of the protein. Increased loop content could also promote increased number of proteolytic sites and how accessible they are. From this data is it clear that Trp6 is highly important for the facilitation of contacts between important parts of the enzyme.

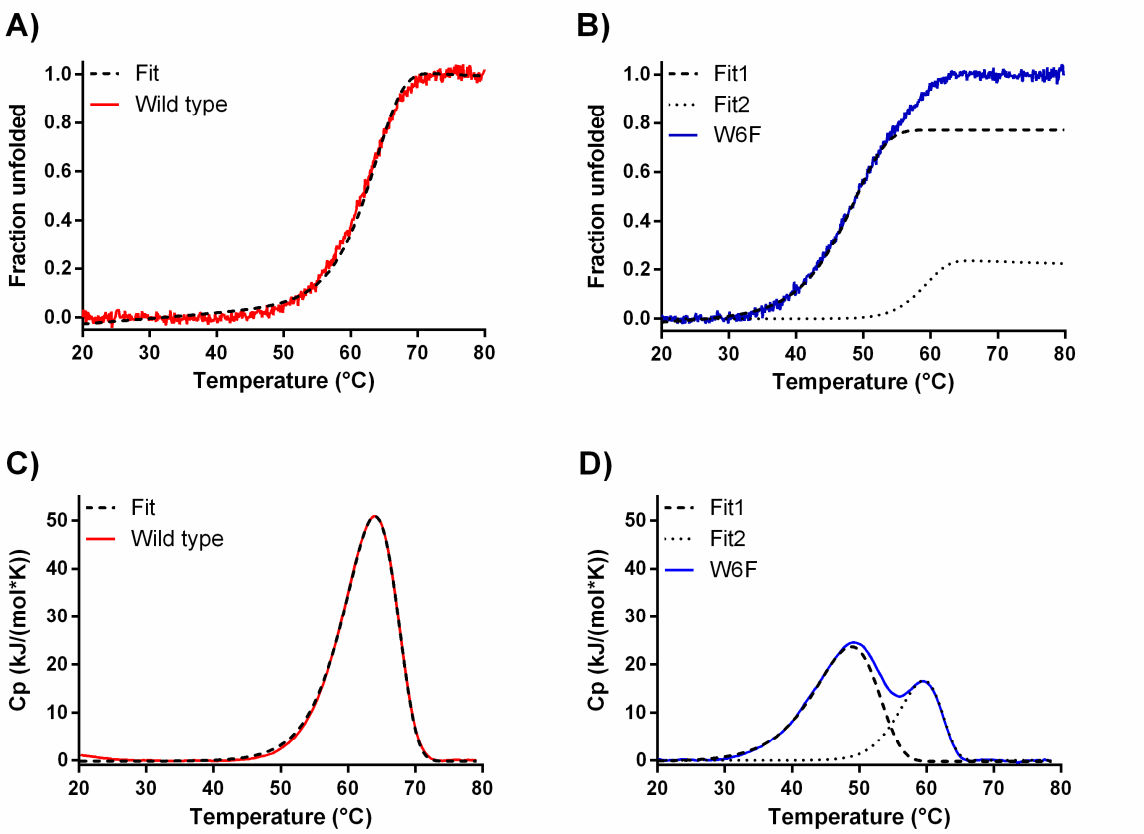


Figure 3.15. Comparisons of CalFitter global fitting of DSC and CD data for $VPR_{\Delta C}$ (red) and $VPR_{\Delta C}$_W6F (blue). A) CD data and fit of $VPR_{\Delta C}$ and C) the DSC data and fit of $VPR_{\Delta C}$. B) CD data and fits of the two apparent transitions observed and D) the DSC data and fit of $VPR_{\Delta C}$_W6F. The thick dotted line represents the fit of the first transitions and the small dotted line represents the second transition.

MD simulations

Initial inspection of MD simulation runs, paints a picture that support the statement made above. A general comparison of MD simulations of $VPR_{\Delta C}$ and $VPR_{\Delta C}$_W6F suggest a difference in behavior of the two forms of the enzyme. In $VPR_{\Delta C}$, the Trp6 residue spends most of its time tucked into the crevice between α-helix-A, β-sheet-6 and the loop between β-sheet-7 and the antiparallel β-sheet preceding α-helix-E (Fig. 3.16). However, in the MD-simulation of $VPR_{\Delta C}$_W6F, the aromatic side chain of the Phe6 residue does not reside in that pocket and is pointed towards the solvent during most of the simulation. This causes a cascade of structural changes within the structure, leaving the N-terminus and helix-A relatively unchanged except for residues 1 to 5, that are much less mobile than in the $VPR_{\Delta C}$ simulation (Fig. 3.16).

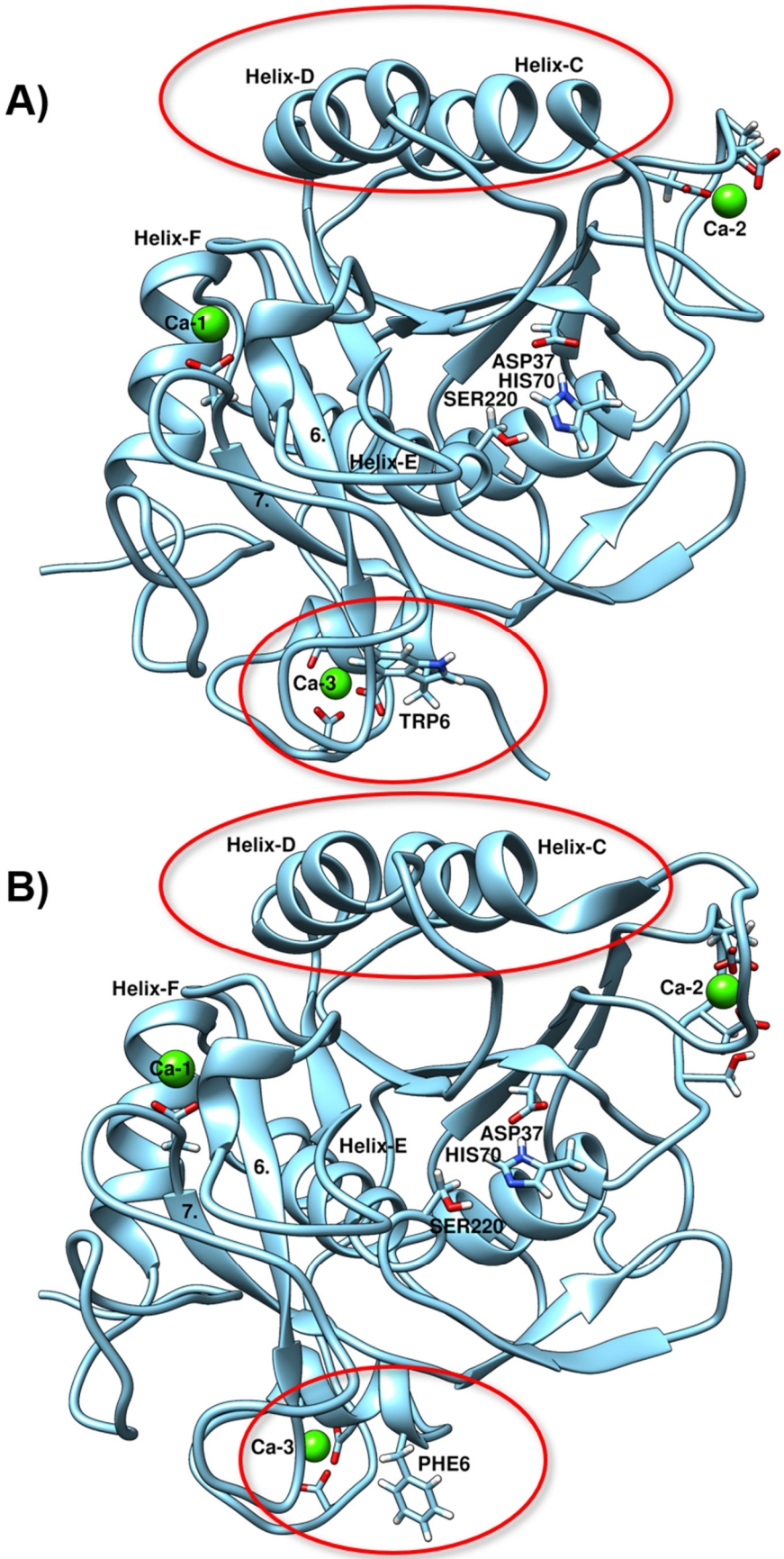


Figure 3.16. Snapshots from MD-simulations of A) VPR$_{\Delta C}$ frame 807/1251 and B) VPR$_{\Delta C}$_W6F frame 1398/2502. Structures are in silico mutants created from the structure of VPR (1SH7). Red circles are to highlight areas of interest.

The simulations show that few of the α-helices that have unwound (Fig. 3.16). Most prominent changes can be associated with helix-C and helix-D that have shortened by a whole turn. Similarly, the last turn of helix-E, containing the active Ser220, becomes a little more stretched. Thus, there is evidence of structural changes in the MD-simulations that are supported by the CD-wavelength scans (Fig 3.14. C). However, at this time further analysis of the MD-simulations to pinpoint the changes in protein interactions that cause these structural changes have not been carried out.

3.5 Active site dansylation of VPR and AQUI

In order to test the hypothesis on whether the active site of $VPR_{\Delta C}$ is more mobile at lower temperatures than the active site of AQUI, a convenient label was needed to monitor active site dynamics. For that purpose, the fluorescent label dansyl fluoride was utilized. Dansyl fluoride reacts specifically with active alcohols and suits the purpose well of labeling the active site serine residue of $VPR_{\Delta C}$ and AQUI. Thus, by using steady-state and time-resolved fluorescence measurements, the differences in the environment and dynamics of the active site region of the two extremophilic enzymes would be examined. Dansyl fluoride was found to react with the active serine residue of $VPR_{\Delta C}$ and AQUI in a pseudo-first order fashion when carried out in cacodylate buffer containing 10 mM $CaCl_2$, at 25°C and pH 7.0 (Fig. 3.17). Other buffers applicable at around neutral pH, such as Tris, MOPS and imidazole, displayed second order reaction rates (Fig 3.17. B). These results indicated some side reactions, possibly between the buffer ion and the dansyl label that do not occur in the cacodylate buffer. Labeling efficiency was estimated to be around 99 % according to activity measurements (Fig. 3.17C). In addition, the calculated molar attenuation coefficient for the DNS-label at its λ_{max}, around 340 nm, indicated a protein/label ratio of 1:1 (see Materials and Methods) (Fig. 3. 18). It is noteworthy, that the labeling reaction progressed at almost twice the rate in the case of AQUI versus $VPR_{\Delta C}$ (Fig. 3.17). The reason for this observation is not known but could indicate that the environment around the active serine in $VPR_{\Delta C}$ creates some steric hindrance, or some repulsive interactions. Labeling of the enzymes, followed by exhaustive washing to get rid of free labels, resulted in fully labeled protein molecules, having no observable effects on the structure of the protein as monitored by CD (Fig. 3.18).

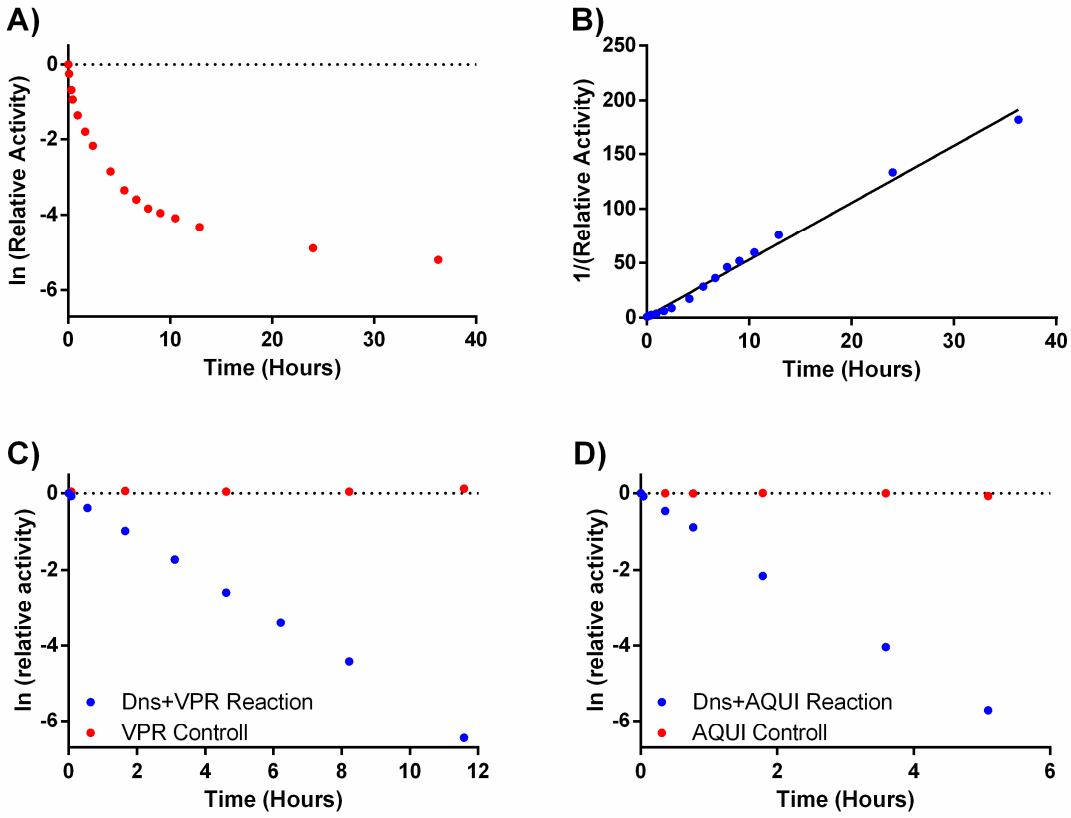


Figure 3.17. A) Rate of VPR inactivation in the reaction with dansyl fluoride (DNSF) in MOPS buffer at pH 7.0 and 25°C, showing deviation from first order kinetics. B) Same data as in A) plotted 1/relative activity showing linearity and thus indicative of 2nd order kinetics. C) Example of the rate of inactivation of VPR (blue) in the reaction with DNSF in cacodylate buffer at pH 7.0 at 25°C and the control (red) in the same buffer containing 5 % 1.4-dioxane. D) Example of the rate of inactivation of AQUI (blue) in its reaction with DNSF in cacodylate buffer at pH 7.0 at 25°C and the control (red) in the same buffer containing 5 % 1.4-dioxane.

Steady-state fluorescence measurements of native and denatured states of AQUI-DNS and $VPR_{\Delta C}$-DNS, after exciting at 295 nm and 370 nm, were recorded. Excitation at 295 nm should be rather specific for tryptophan fluorescence. Our results indicated that there was some Förster resonance energy transfer (FRET) between tryptophan residues and the dansyl-label, as tryptophan fluorescence of $VPR_{\Delta C}$-DNS was about half of that emitted by unlabeled $VPR_{\Delta C}$ (Fig. 3.19). The label having a more complex emission, seemingly consisting of two separate peaks (Fig. 3.19. A). Similar trends were seen in the emission spectra of the denatured state, demonstrated by the fact that the denatured state of $VPR_{\Delta C}$-DNS shows reduced tryptophan emission (Fig. 3.19. B and Fig. 3.11. A). However, the signal from the DNS-label is a single peak for the denatured states emitting much more light compared to

the native state, with a λ_{max} at 500 nm (Fig. 3.19 B). This assessment of apparent FRET could be somewhat overestimated, however, as the label itself might be absorbing some light at 295 nm, therefore, causing an overestimation of reduced Trp fluorescence.

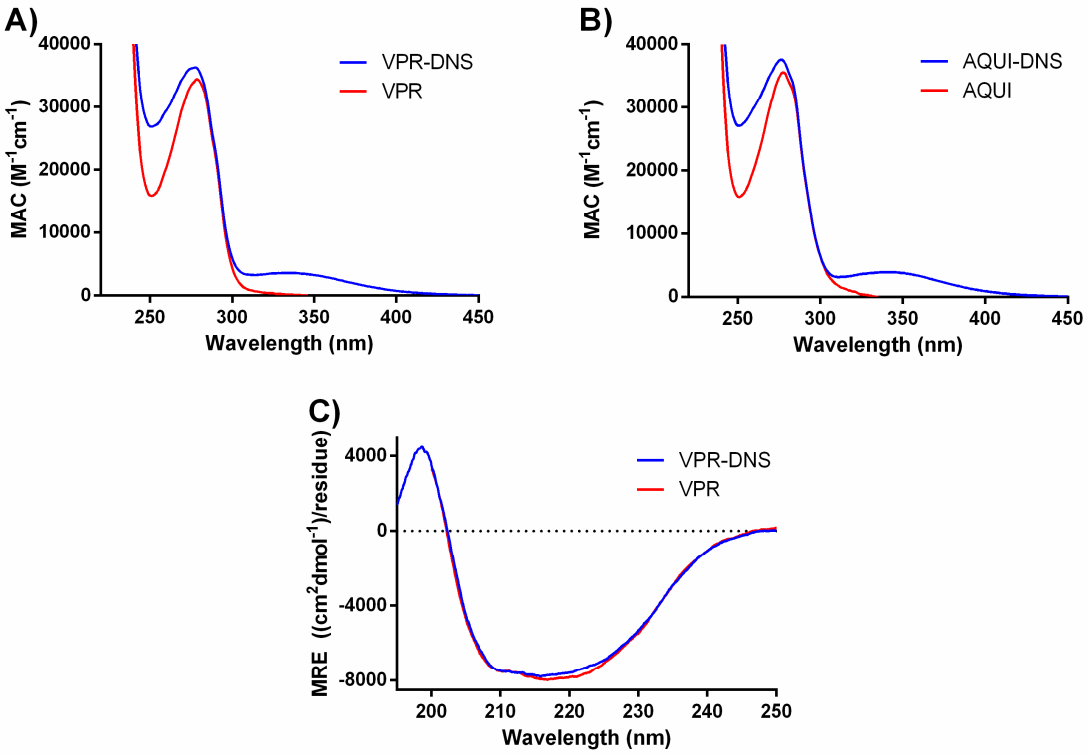


Figure 3.18. Comparison of the absorption spectra of label free VPR (red) and dansylated VPR (blue), expressed as molar attenuation coefficients (MAC). B) Comparison of the absorption spectra of label free AQUI (red) and dansylated AQUI (blue), expressed as molar attenuation coefficients (MAC) C) Comparison of the CD wavelength scans of dansylated VPR (blue) and non-labeled VPR (red), expressed as mean residual ellipticity.

The emission spectra obtained when exciting the DNS-label specifically at 370 nm showed similar trends. The denatured states of AQUI-DNS and $VPR_{\Delta C}$-DNS had essentially the same emission spectra, meaning that the environment of the label in both unfolded assemblies was very similar. The native emission signal showed that the label in the case of $VPR_{\Delta C}$-DNS was more quenched than in the case of AQUI. The λ_{max} (the emission peak), however, was around 585 nm in both cases (Figs. 3.19 C and D), indicating similar overall electrostatic environments in the active sites of both AQUI-DNS and $VPR_{\Delta C}$-DNS. However, the more quenched fluorescence emission of $VPR_{\Delta C}$-DNS indicated that the label in that case interacted more with a chemical moiety that is able to intrinsically quench the fluorescence of the dansyl label.

The λ_{max} values for the native states were much higher than for the denatured assembly meaning that the environment of the label in the native state was much more polar than in the denatured state and thus more polar than a water environment [194]. In addition, acrylamide quenching of label emission resulted in no apparent effect on the native states. However, the denatured signal was readily quenched. Taken together, these results suggest that the label is located in a highly polar environment within the protein, where the label emission is intrinsically quenched, and very inaccessible to extrinsic quenchers such as acrylamide.

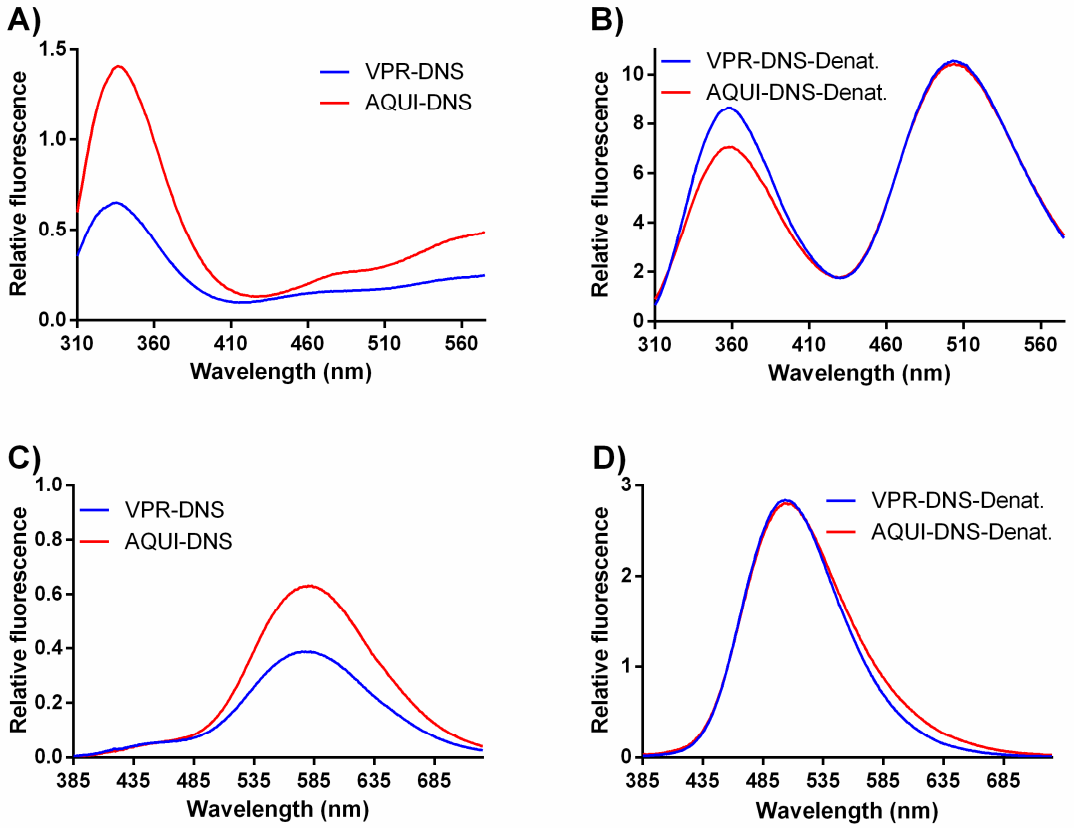


Figure 3.19. Emission spectra of $VPR_{\Delta C}$-DNS and AQUI-DNS at 25°C and pH 8.0. A) The emission spectra of native structures when excited at 295 nm. B) The emission spectra of denatured assemblies when excited at 295 nm. C) The emission spectra of native structures when excited at 370 nm. D) The emission spectra of denatured assemblies when excited at 370 nm.

In addition to measuring steady state fluorescence, we measured time-resolved fluorescence decay at varying temperatures in order to determine if there were notable changes in active site dynamics between the two enzymes at varying temperatures (Table 3-5).

The recorded fluorescence decay of native AQUI-DNS and $VPR_{\Delta C}$-DNS exhibited complex decay patterns (Fig. 3.20. A), whereas the decay of the unfolded assembly for AQUI-DNS and $VPR_{\Delta C}$-DNS display a single exponent decay (Fig. 3.20. B). The decays of the unfolded

assemblies were essentially the same for AQUI-DNS and VPR$_{\Delta C}$-DNS, decreasing linearly with temperatures (Fig. 3.21). These lifetimes could then be fitted, and the fit used to predict the lifetimes of the unfolded assemblies over the temperature range and used in fitting denatured decay in the native decay data sets (see Materials and Methods for details).

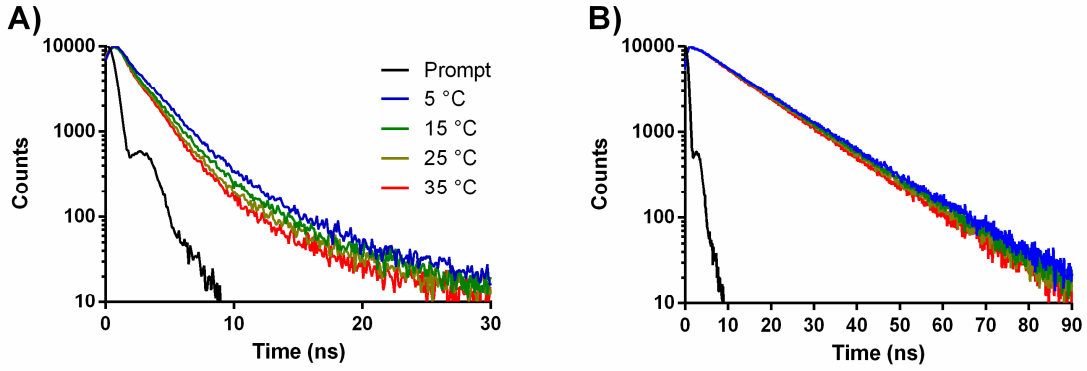


Figure 3.20. Examples of lifetime measurements of A) native VPR$_{\Delta C}$-DNS and B) denatured VPR$_{\Delta C}$-DNS at varying temperatures. Showing the prompt (black), VPR$_{\Delta C}$-DNS lifetime at 5°C (blue), 15°C (green), 25°C (gold) and 35°C (red).

As both AQUI-DNS and VPR$_{\Delta C}$-DNS are kinetically stable proteases that unfold irreversibly, the assumption that some unfolded molecules are found in the samples is reasonable. Furthermore, a small fraction of unfolded molecules would give a high background due to the long lifetime of the dansyl-label in the unfolded structures. Utilizing this, the native decay could be reliably fitted by three different lifetimes, one of them being the unfolded lifetime, accounting for less than 1 % of the signal in most experiments.

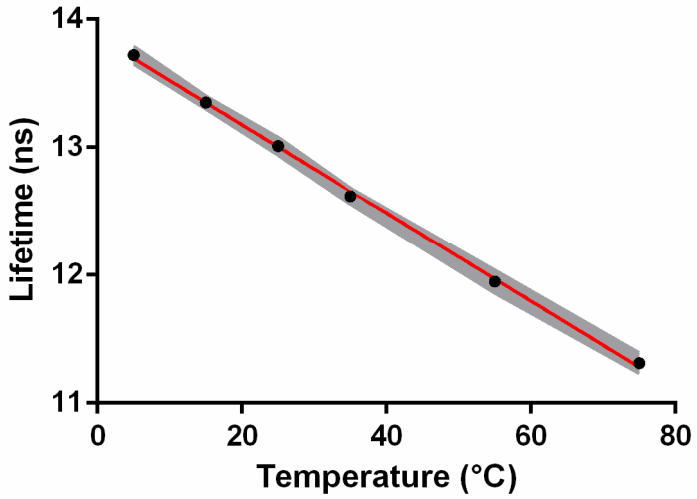


Figure 3.21. The temperature dependence of lifetimes of denatured VPR$_{\Delta C}$-DNS and AQUI-DNS excited at at 370 nm. Black dots represent average values, the red line is the best line fit of the data and the gray area represents the standard deviation of the mean.

According to those measurements, both AQUI-DNS and $VPR_{\Delta C}$-DNS have two native lifetimes, a shorter one, about 1 ns and a longer, one being between 2 ns and 3 ns (Table 3-5). This means that the DNS-label has at least two different microenvironments in the native structure of AQUI-DNS and $VPR_{\Delta C}$-DNS. The timescales of these two separate lifetimes of the two enzymes might suggest that these microenvironments are similar in both cases. However, these environments are not evenly populated over the measured temperatures, the short lifetime being favored in both enzymes at higher temperatures. In the case of AQUI-DNS, this trend seems to be more extreme, with initial estimates at 5°C for indicating those states being almost equally populated. In the case of $VPR_{\Delta C}$-DNS on the other hand, the distribution was about 80/20 between the states $VPR_{\Delta C}$-DNS (Fig. 3.22).

Table 3-5. The two fitted lifetimes of native $VPR_{\Delta C}$-DNS and AQUI-DNS (t_1 and t_2) and their fraction of the total signal (f_1 and f_2).

		$VPR_{\Delta C}$-DNS		
Temperature	t_1 (ns)	t_2 (ns)	f_1	f_2
5°C	1.26 ± 0.03	3.12 ± 0.13	0.79 ± 0.04	0.21 ± 0.04
15°C	1.12 ± 0.06	2.81 ± 0.25	0.82 ± 0.01	0.18 ± 0.01
25°C	1.05 ± 0.07	2.73 ± 0.21	0.84 ± 0.06	0.16 ± 0.06
35°C	1.07 ± 0.05	2.66 ± 0.23	0.89 ± 0.01	0.11 ± 0.01
		AQUI-DNS		
5°C	0.67 ± N.A.	2.61 ± N.A.	0.47 ± N.A.	0.52 ± N.A.
15°C	1.45 ± 0.35	2.66 ± 0.14	0.71 ± 0.09	0.29 ± 0.09
25°C	1.29 ± 0.22	2.58 ± 0.19	0.80 ± 0.10	0.20 ± 0.09
35°C	1.22 ± 0.04	2.62 ± 0.21	0.90 ± 0.03	0.10 ± 0.03
55°C	1.02 ± 0.02	2.65 ± 0.43	0.95 ± 0.01	0.05 ± 0.01
75°C	0.91 ± 0.05	2.25 ± 0.17	0.94 ± 0.02	0.06 ± 0.01

These results suggest that there may be differences in the active sites of AQUI and $VPR_{\Delta C}$ that can be detected by using the dansyl-label as a reporter group. Identifying what these two microenvironments represent in the actual structures of the enzymes is a bit harder and solving that problem might be problematic. However, by measuring under more varied conditions such as other pH levels in addition to labeling variants with enhanced or diminished catalytic efficiencies might shed a light on the meaning of these two lifetimes.

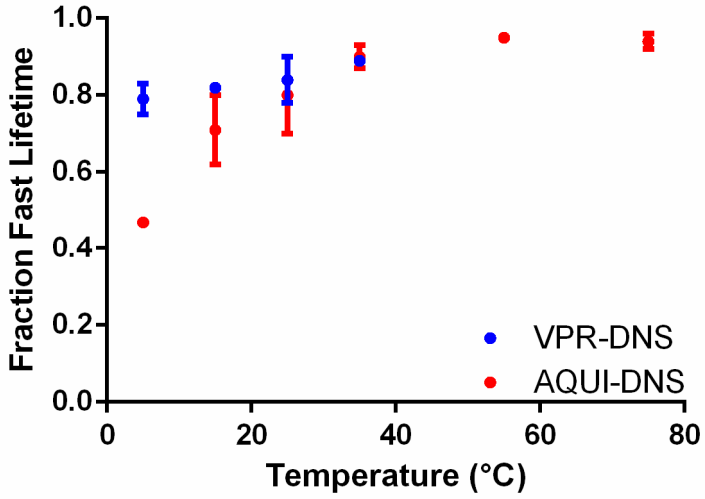


Figure 3.22. The effects of temperature on the dominant shorter lifetime (t_1) in the native states of AQUI-DNS and $VPR_{\Delta C}$-DNS.

4 Conclusions

This work highlights the important characteristics and properties of the interaction network that is vital for the kinetic stability of VPR. A prime example is how the W6F mutation affects the whole protein, seemingly causing a vast collapse of the internal non-covalent interactions all around the protein structure, leading to destabilization of the protein fold. The same can be concluded from the proline exchange experiments, where observations indicated long-range synergic effects within the structure as a result of multiple proline replacements, contributing to the stability of the structure.

Summarizing the results from the mutational studies reported in this thesis and the experiments revealing how calcium ions effect the unfolding process, there is evidence to indicate the presence of a weak point in the non-covalent bond interaction network. This network connecting two separate unfolding units. Loss of contact between those unfolding units or greatly increased stability in one of them, will cause a separation of unfolding events. A hypothesis was put forward as to which parts of the protein harbor the two structural moieties that the stability effects are centered around. Mutational studies at the N-terminal region indicated that one of the calcium ion binding sites (Ca-3) is responsible for strengthening the more stable part of the protein. Calcium ion and proline exchange experiments indicated that the region around a different calcium binding site (Ca-1) is responsible for contributing to the stability of the less stable region of the protein. Thus, strengthening the interactions within and between those two regions of the protein could be a good strategy in increasing the thermal stability of VPR.

As shown in this thesis, the research model of comparing the similar structures of VPR and AQUI from the standpoint of their very different operational temperatures still has insights to give regarding kinetic stability and what molecular interactions are important for kinetic stability. In addition, the firsts steps in monitoring active site dynamics have been made. That work might help to shed some light on more detailed view on movements within the active site of these enzymes, in order to further explain the differences in their catalytic efficiencies.

References

1. Kristjansson, M.M., Magnusson, O.T., Gudmundsson, H.M., Alfredsson, G.A. & Matsuzawa, H. Properties of a subtilisin-like proteinase from a psychrotrophic *Vibrio* species comparison with proteinase K and aqualysin I. *Eur J Biochem* **260**, 752-760 (1999).
2. Matsuzawa, H., Hamaoki, M. & Ohta, T. Production of thermophilic extracellular proteases (aqualysin-I and aqualysin-II) by *Thermus aquaticus* YT-1, an extreme thermophile. *Agr Biol Chem* **47**, 25-28 (1983).
3. Oskarsson, K.R. & Kristjansson, M.M. Improved expression, purification and characterization of VPR, a cold active subtilisin-like serine proteinase and the effects of calcium on expression and stability. *Biochim Biophys Acta Proteins Proteom* **1867**, 152-162 (2019).
4. Oskarsson, K.R., Saevarsson, A.F. & Kristjansson, M.M. Thermostabilization of VPR, a kinetically stable cold adapted subtilase, via multiple proline substitutions into surface loops. *Sci Rep* **10** (2020).
5. Wacey, D., Kilburn, M.R., Saunders, M., Cliff, J. & Brasier, M.D. Microfossils of sulphur-metabolizing cells in 3.4-billion-year-old rocks of Western Australia. *Nat Geosci* **4**, 698-702 (2011).
6. Ohtomo, Y., Kakegawa, T., Ishida, A., Nagase, T. & Rosing, M.T. Evidence for biogenic graphite in early Archaean Isua metasedimentary rocks. *Nat Geosci* **7**, 25-28 (2014).
7. Bell, E.A., Boehnke, P., Harrison, T.M. & Mao, W.L. Potentially biogenic carbon preserved in a 4.1 billion-year-old zircon. *Proc Natl Acad Sci USA* **112**, 14518-14521 (2015).
8. Manhes, G., Allegre, C.J., Dupre, B. & Hamelin, B. Lead isotope study of basic-ultrabasic layered complexes - Speculations about the age of the earth and primitive mantle characteristics. *Earth Planet Sc Lett* **47**, 370-382 (1980).
9. Darwin, C. On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life. (London : John Murray, 1859, 1859).
10. Theobald, D.L. A formal test of the theory of universal common ancestry. *Nature* **465**, 219-219 (2010).
11. Weiss, M.C. et al. The physiology and habitat of the last universal common ancestor. *Nat Microbiol* **1** (2016).
12. Gogarten, J.P. & Deamer, D. Is LUCA a thermophilic progenote? *Nat Microbiol* **1** (2016).
13. Gaucher, E.A., Govindarajan, S. & Ganesh, O.K. Palaeotemperature trend for Precambrian life inferred from resurrected proteins. *Nature* **451**, 704-707 (2008).
14. Robert, F. & Chaussidon, M. A palaeotemperature curve for the Precambrian oceans based on silicon isotopes in cherts. *Nature* **443**, 969-972 (2006).
15. Merino, N. et al. Living at the extremes: Extremophiles and the limits of life in a planetary context. *Front Microbiol* **10** (2019).
16. Geiges, O. Microbial processes in frozen food. *Adv Space Res* **18**, 109-118 (1996).
17. Takai, K. et al. Cell proliferation at 122 degrees C and isotopically heavy CH4 production by a hyperthermophilic methanogen under high-pressure cultivation. *Proc Natl Acad Sci USA* **105**, 10949-10954 (2008).

18. Clarke, A. The thermal limits to life on Earth. *Int J Astrobiol* **13**, 141-154 (2014).
19. Frosler, J., Panitz, C., Wingender, J., Flemming, H.C. & Rettberg, P. Survival of *Deinococcus geothermalis* in biofilms under desiccation and simulated space and martian conditions. *Astrobiology* **17**, 431-447 (2017).
20. Price, P.B. & Sowers, T. Temperature dependence of metabolic rates for microbial growth, maintenance, and survival. *Proc Natl Acad Sci USA* **101**, 4631-4636 (2004).
21. Clarke, A. et al. A low temperature limit for life on earth. *Plos One* **8** (2013).
22. Cowan, D.A. The upper temperature for life - where do we draw the line? *Trends Microbiol* **12**, 58-60 (2004).
23. D'Amico, S., Marx, J.C., Gerday, C. & Feller, G. Activity-stability relationships in extremophilic enzymes. *J Biol Chem* **278**, 7891-7896 (2003).
24. Gudjonsdottir, K. & Asgeirsson, B. Effects of replacing active site residues in a cold-active alkaline phosphatase with those found in its mesophilic counterpart from *Escherichia coli*. *FEBS J* **275**, 117-127 (2008).
25. Siddiqui, K.S. & Cavicchioli, R. Cold-adapted enzymes. *Annu Rev Biochem* **75**, 403-433 (2006).
26. Nguyen, V. et al. Evolutionary drivers of thermoadaptation in enzyme catalysis. *Science* **355**, 289-294 (2017).
27. Miyazaki, K., Wintrode, P.L., Grayling, R.A., Rubingh, D.N. & Arnold, F.H. Directed evolution study of temperature adaptation in a psychrophilic enzyme. *J Mol Biol* **297**, 1015-1026 (2000).
28. Wintrode, P.L., Miyazaki, K. & Arnold, F.H. Cold adaptation of a mesophilic subtilisin-like protease by laboratory evolution. *J Biol Chem* **275**, 31635-31640 (2000).
29. Bigman, L.S. & Levy, Y. Proteins: molecules defined by their trade-offs. *Curr Opin Struct Biol* **60**, 50-56 (2019).
30. Robertson, A.D. & Murphy, K.P. Protein structure and the energetics of protein stability. *Chem Rev* **97**, 1251-1268 (1997).
31. Lumry, R. & Biltonen, R. Validity of the "two-state" hypothesis for conformational transitions of proteins. *Biopolymers* **4**, 917-944 (1966).
32. Anfinsen, C.B. Principles that govern the folding of protein chains. *Science* **181**, 223-230 (1973).
33. Pace, C.N. et al. Contribution of hydrogen bonds to protein stability. *Protein Sci* **23**, 652-661 (2014).
34. Pace, C.N. et al. Contribution of hydrophobic interactions to protein stability. *J Mol Biol* **408**, 514-528 (2011).
35. Perutz, M.F. Electrostatic effects in proteins. *Science* **201**, 1187-1191 (1978).
36. Prajapati, R.S., Sirajuddin, M., Durani, V., Sreeramulu, S. & Varadarajan, R. Contribution of cation-Contribution of cation-π interactions to protein stability. *Biochemistry* **45**, 15000-15010 (2006).
37. Burley, S.K. & Petsko, G.A. Amino-aromatic interactions in proteins. *FEBS Lett* **203**, 139-143 (1986).
38. Breberina, L.M., Milcic, M.K., Nikolic, M.R. & Stojanovic, S.D. Contribution of anion-π interactions to the stability of Sm/LSm proteins. *J Biol Inorg Chem* **20**, 475-485 (2015).
39. Majewski, M., Ruiz-Carmona, S. & Barril, X. An investigation of structural stability in protein-ligand complexes reveals the balance between order and disorder. *Commun Chem* **2** (2019).
40. Dill, K.A. Dominant forces in protein folding. *Biochemistry* **29**, 7133-7155 (1990).

41. Bye, J.W., Baxter, N.J., Hounslow, A.M., Falconer, R.J. & Williamson, M.P. Molecular mechanism for the hofmeister effect derived from NMR and DSC measurements on barnase. *ACS Omega* **1**, 669-679 (2016).
42. Lucas, X., Bauza, A., Frontera, A. & Quinonero, D. A thorough anion-π interaction study in biomolecules: on the importance of cooperativity effects. *Chem Sci* **7**, 1038-1050 (2016).
43. Kuriyan, J., Konforti, B. & Wemmer, D. The molecules of life : Physical and chemical principles. (Garland Science ,Taylor & Francis Group, New York; 2013).
44. Dill, K.A. & Alonso, D.O.V. Conformational entropy and protein stability, Vol. 39. (Springer, Berlin, Heidelberg, Berlin, Heidelberg; 1988).
45. Kinoshita, M. Importance of translational entropy of water in biological self-assembly processes like protein folding. *Int J Mol Sci* **10**, 1064-1080 (2009).
46. Chandler, D. Interfaces and the driving force of hydrophobic assembly. *Nature* **437**, 640-647 (2005).
47. Gerstein, M. & Chothia, C. Packing at the protein-water interface. *Proc Natl Acad Sci USA* **93**, 10167-10172 (1996).
48. Gibbs, J.W. A method of geometrical representation of the thermodynamic properties of substances by means of surfaces. (New Haven, 1873).
49. Tanford, C. Protein denaturation. C. Theoretical models for the mechanism of denaturation. *Adv Protein Chem* **24**, 1-95 (1970).
50. Greene, R.F., Jr. & Pace, C.N. Urea and guanidine hydrochloride denaturation of ribonuclease, lysozyme, alpha-chymotrypsin, and beta-lactoglobulin. *J Biol Chem* **249**, 5388-5393 (1974).
51. Cooper, A. Protein heat capacity: An anomaly that maybe never was. *J Phys Chem Lett* **1**, 3298-3304 (2010).
52. Cooper, A., Johnson, C.M., Lakey, J.H. & Nollmann, M. Heat does not come in different colours: entropy-enthalpy compensation, free energy windows, quantum confinement, pressure perturbation calorimetry, solvation and the multiple causes of heat capacity effects in biomolecular interactions. *Biophys Chem* **93**, 215-230 (2001).
53. Becktel, W.J. & Schellman, J.A. Protein stability curves. *Biopolymers* **26**, 1859-1877 (1987).
54. Hopkins, F.G. Denaturation of proteins by urea and related substances. *Nature* **126**, 383-384 (1930).
55. Privalov, P.L. Cold denaturation of proteins. *Crit Rev Biochem Mol Biol* **25**, 281-305 (1990).
56. Pastore, A. et al. Unbiased cold denaturation: low- and high-temperature unfolding of yeast frataxin under physiological conditions. *J Am Chem Soc* **129**, 5374-5375 (2007).
57. Sanfelice, D., Morandi, E., Pastore, A., Niccolai, N. & Temussi, P.A. Cold denaturation unveiled: Molecular mechanism of the asymmetric unfolding of yeast Frataxin. *ChemPhysChem* **16**, 3599-3602 (2015).
58. Brandts, J.F. & Hunt, L. The thermodynamics of protein denaturation. 3. The denaturation of ribonuclease in water and in aqueous urea and aqueous ethanol mixtures. *J Am Chem Soc* **89**, 4826-4838 (1967).
59. Cooper, A. Heat capacity of hydrogen-bonded networks: an alternative view of protein folding thermodynamics. *Biophys Chem* **85**, 25-39 (2000).
60. National Research, C. et al. International critical tables of numerical data, physics, chemistry and technology. (Pub. for the National research council by the McGraw-Hill book company, inc., New York; 1926).

61. Cooper, A., Cameron, D., Jakus, J. & Pettigrew, G.W. Pressure perturbation calorimetry, heat capacity and the role of water in protein stability and interactions. *Biochem Soc Trans* **35**, 1547-1550 (2007).
62. Dunitz, J.D. Win some, lose some - Enthalpy-entropy compensation in weak intermolecular interactions. *Chem Biol* **2**, 709-712 (1995).
63. Ebbinghaus, S. et al. An extended dynamical hydration shell around proteins. *Proc Natl Acad Sci USA* **104**, 20749-20752 (2007).
64. Makhatadze, G.I. & Privalov, P.L. Energetics of protein structure. *Adv Protein Chem* **47**, 307-425 (1995).
65. Camilloni, C. et al. Towards a structural biology of the hydrophobic effect in protein folding. *Sci Rep* **6** (2016).
66. Aznauryan, M., Nettels, D., Holla, A., Hofmann, H. & Schuler, B. Single-molecule spectroscopy of cold denaturation and the temperature-induced collapse of unfolded proteins. *J Am Chem Soc* **135**, 14040-14043 (2013).
67. Wuttke, R. et al. Temperature-dependent solvation modulates the dimensions of disordered proteins. *Proc Natl Acad Sci USA* **111**, 5213-5218 (2014).
68. Dragan, A.I., Read, C.M. & Crane-Robinson, C. Enthalpy-entropy compensation: the role of solvation. *Eur Biophys J Biophy* **46**, 301-308 (2017).
69. Sanchez-Ruiz, J.M. Protein kinetic stability. *Biophys Chem* **148**, 1-15 (2010).
70. Laidler, K.J. & King, M.C. The development of transition-state theory. *J Phys Chem-Us* **87**, 2657-2664 (1983).
71. Svante, A. Über die dissociationswärme und den einfluss der temperatur auf den dissociationsgrad der elektrolyte. *Zeitschrift für Physikalische Chemie* **4U**, 96-116 (1889).
72. Svante, A. Über die reaktionsgeschwindigkeit bei der inversion von rohrzucker durch säuren. *Zeitschrift für Physikalische Chemie* **4U**, 226-248 (1889).
73. Truhlar, D.G., Garrett, B.C. & Klippenstein, S.J. Current status of transition-state theory. *J Phys Chem* **100**, 12771-12800 (1996).
74. Eyring, H. The activated complex in chemical reactions. *J Chem Phys* **3**, 107-115 (1935).
75. Vestergaard, C.L., Mikkelsen, M.B., Reisner, W., Kristensen, A. & Flyvbjerg, H. Transition state theory demonstrated at the micron scale with out-of-equilibrium transport in a confined environment. *Nat Commun* **7** (2016).
76. Lumry, R. & Eyring, H. Conformation changes of proteins. *J Phys Chem* **58**, 110-120 (1954).
77. Kramers, H.A. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **7**, 284-304 (1940).
78. Chung, H.S. & Eaton, W.A. Single-molecule fluorescence probes dynamics of barrier crossing. *Nature* **502**, 685-688 (2013).
79. Schuler, B. & Clarke, J. Biophysics: Rough passage across a barrier. *Nature* **502**, 632-633 (2013).
80. Gershenson, A., Gierasch, L.M., Pastore, A. & Radford, S.E. Energy landscapes of functional proteins are inherently risky. *Nat Chem Biol* **10**, 884-891 (2014).
81. Sohl, J.L., Jaswal, S.S. & Agard, D.A. Unfolded conformations of alpha-lytic protease are more stable than its native state. *Nature* **395**, 817-819 (1998).
82. Ikemura, H., Takagi, H. & Inouye, M. Requirement of pro-sequence for the production of active subtilisin-E in *Escherichia Coli*. *J Biol Chem* **262**, 7859-7864 (1987).

83. Shinde, U. & Inouye, M. Intramolecular chaperones and protein-folding. *Trends Biochem Sci* **18**, 442-446 (1993).
84. Shinde, U. & Inouye, M. Intramolecular chaperones: polypeptide extensions that modulate protein folding. *Semin Cell Dev Biol* **11**, 35-44 (2000).
85. Shinde, I.P., Fu, X., Marie-Claire, C., Yabuta, Y. & Inouye, M. Intramolecular chaperone mediated folding of subtilisin. *Abstr Pap Am Chem S* **219**, 16871-16878 (2000).
86. Epstein, D.M. & Wensink, P.C. The alpha-lytic protease gene of *Lysobacter Enzymogenes* - The nucleotide-sequence predicts a large prepro-peptide with homology to pro-peptides of other Chymotrypsin-like enzymes. *J Biol Chem* **263**, 16586-16590 (1988).
87. Baardsnes, J. et al. *Streptomyces griseus* protease B: secretion correlates with the length of the propeptide. *J Bacteriol* **180**, 3241-3244 (1998).
88. Winther, J.R. & Sorensen, P. Propeptide of carboxypeptidase-Y provides a chaperone-like function as well as inhibition of the enzymatic-activity. *Proc Natl Acad Sci USA* **88**, 9330-9334 (1991).
89. Eder, J. & Fersht, A.R. Pro-sequence-assisted protein folding. *Mol Microbiol* **16**, 609-614 (1995).
90. Duy, C. & Fitter, J. Thermostability of irreversible unfolding alpha-amylases analyzed by unfolding kinetics. *J Biol Chem* **280**, 37360-37365 (2005).
91. Sanchez-Ruiz, J.M., Lopez-Lacomba, J.L., Cortijo, M. & Mateo, P.L. Differential scanning calorimetry of the irreversible thermal denaturation of thermolysin. *Biochemistry* **27**, 1648-1652 (1988).
92. Weijers, M., Barneveld, P.A., Cohen Stuart, M.A. & Visschers, R.W. Heat-induced denaturation and aggregation of ovalbumin at neutral pH described by irreversible first-order kinetics. *Protein Sci* **12**, 2693-2703 (2003).
93. Violet, M. & Meunier, J.C. Kinetic-study of the irreversible thermal-denaturation of *Bacillus Licheniformis* alpha-amylase. *Biochem J* **263**, 665-670 (1989).
94. Siddiqui, K.S. et al. A chemically modified alpha-amylase with a molten-globule state has entropically driven enhanced thermal stability. *Protein Eng Des Sel* **23**, 769-780 (2010).
95. Torpenholt, S. et al. Effect of mutations on the thermostability of *Aspergillus aculeatus* beta-1,4-galactanase. *Comput Struct Biotec* **13**, 256-264 (2015).
96. Jackson, S.E. & Fersht, A.R. Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry* **30**, 10428-10435 (1991).
97. Jackson, S.E. How do small single-domain proteins fold? *Fold Des* **3**, R81-R91 (1998).
98. Yi, Q. & Baker, D. Direct evidence for a two-state protein unfolding transition from hydrogen-deuterium exchange, mass spectrometry, and NMR. *Protein Sci* **5**, 1060-1066 (1996).
99. Johnson, F.H. The kinetic basis of molecular biology. (Wiley, New York; 1954).
100. Truhlar, S.M.E., Cunningham, E.L. & Agard, D.A. The folding landscape of *Streptomyces griseus* protease B reveals the energetic costs and benefits associated with evolving kinetic stability. *Protein Sci* **13**, 381-390 (2004).
101. Jaswal, S.S., Truhlar, S.M.E., Dill, K.A. & Agard, D.A. Comprehensive analysis of protein folding activation thermodynamics reveals a universal behavior violated by kinetically stable proteases. *J Mol Biol* **347**, 355-366 (2005).

102. Salimi, N.L., Ho, B. & Agard, D.A. Unfolding simulations reveal the mechanism of extreme unfolding cooperativity in the kinetically stable alpha-lytic protease. *Plos Comput Biol* **6** (2010).
103. Sanchez-Ruiz, J.M. Theoretical analysis of Lumry-Eyring models in differential scanning calorimetry. *Biophys J* **61**, 921-935 (1992).
104. Perez-Benito, J.F. Some Considerations on the Fundamentals of Chemical Kinetics: Steady State, Quasi-Equilibrium, and Transition State Theory. *J Chem Educ* **94**, 1238-1246 (2017).
105. del Pino, I.M.P., Ibarra-Molero, B. & Sanchez-Ruiz, J.M. Lower kinetic limit to protein thermal stability: A proposal regarding protein stability in vivo and its relation with misfolding diseases. *Proteins* **40**, 58-70 (2000).
106. Chen, X.W. & Matthews, C.R. Thermodynamic properties of the transition state for the rate-limiting step in the folding of the alpha-subunit of tryptophan synthase. *Biochemistry* **33**, 6356-6362 (1994).
107. Matthews, C.R. Effect of point mutations on the folding of globular proteins. *Methods Enzymol* **154**, 498-511 (1987).
108. Akmal, A. & Munoz, V. The nature of the free energy barriers to two-state folding. *Proteins* **57**, 142-152 (2004).
109. Parker, M.J., Lorch, M., Sessions, R.B. & Clarke, A.R. Thermodynamic properties of transient intermediates and transition states in the folding of two contrasting protein structures. *Biochemistry* **37**, 2538-2545 (1998).
110. Ohgushi, M. & Wada, A. 'Molten-globule state': a compact form of globular proteins with mobile side-chains. *FEBS Lett* **164**, 21-24 (1983).
111. Rodriguez-Larrea, D., Ibarra-Molero, B. & Sanchez-Ruiz, J.M. Energetic and structural consequences of desolvation/solvation barriers to protein folding/unfolding assessed from experimental unfolding rates. *Biophys J* **91**, L48-50 (2006).
112. Rank, J.A. & Baker, D. A desolvation barrier to hydrophobic cluster formation may contribute to the rate-limiting step in protein folding. *Protein Sci* **6**, 347-354 (1997).
113. Rodriguez-Larrea, D., Minning, S., Borchert, T.V. & Sanchez-Ruiz, J.M. Role of solvation barriers in protein kinetic stability. *J Mol Biol* **360**, 715-724 (2006).
114. Liu, Z. & Chan, H.S. Solvation and desolvation effects in protein folding: native flexibility, kinetic cooperativity and enthalpic barriers under isostability conditions. *Phys Biol* **2**, S75-85 (2005).
115. Frank, H.S. & Evans, M.W. Free volume and entropy in condensed systems III. Entropy in binary liquid mixtures; Partial molal entropy in dilute solutions; Structure and thermodynamics in aqueous electrolytes. *J Chem Phys* **13**, 507 (1945).
116. Southall, N.T., Dill, K.A. & Haymet, A.D.J. A view of the hydrophobic effect. *J Phys Chem B* **106**, 521-533 (2002).
117. Costas, M. et al. Between-species variation in the kinetic stability of TIM proteins linked to solvation-barrier free energies. *J Mol Biol* **385**, 924-937 (2009).
118. Subbian, E., Yabuta, Y. & Shinde, U. Positive selection dictates the choice between kinetic and thermodynamic protein folding and stability in subtilases. *Biochemistry* **43**, 14348-14360 (2004).
119. Truhlar, S.M.E. & Agard, D.A. The folding landscape of an alpha-lytic protease variant reveals the role of a conserved beta-hairpin in the development of kinetic stability. *Proteins* **61**, 105-114 (2005).
120. Cunningham, E.L., Jaswal, S.S., Sohl, J.L. & Agard, D.A. Kinetic stability as a mechanism for protease longevity. *Proc Natl Acad Sci USA* **96**, 11008-11014 (1999).

121. Quezada, A.G. et al. Interplay between protein thermal flexibility and kinetic stability. *Structure* **25**, 167-179 (2017).
122. Colon, W. et al. Biological roles of protein kinetic stability. *Biochemistry* **56**, 6179-6186 (2017).
123. Luke, K.A., Higgins, C.L. & Wittung-Stafshedel, P. Thermodynamic stability and folding of proteins from hyperthermophilic organisms. *FEBS J* **274**, 4023-4033 (2007).
124. Vieille, C. & Zeikus, G.J. Hyperthermophilic enzymes: Sources, uses, and molecular mechanisms for thermostability. *Microbiol Mol Biol R* **65**, 1-43 (2001).
125. Pucci, F. & Rooman, M. Physical and molecular bases of protein thermal stability and cold adaptation. *Curr Opin Struc Biol* **42**, 117-128 (2017).
126. Szilagyi, A. & Zavodszky, P. Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. *Structure* **8**, 493-504 (2000).
127. Folch, B., Dehouck, Y. & Rooman, M. Thermo- and mesostabilizing protein interactions identified by temperature-dependent statistical potentials. *Biophys J* **98**, 667-677 (2010).
128. van Dijk, E., Hoogeveen, A. & Abeln, S. The hydrophobic temperature dependence of amino acids directly calculated from protein structures. *Plos Comput Biol* **11** (2015).
129. Sun, Q. The physical origin of hydrophobic effects. *Chem Phys Lett* **672**, 21-25 (2017).
130. Sammond, D.W. et al. Comparing residue clusters from thermophilic and mesophilic enzymes reveals adaptive mechanisms. *Plos One* **11** (2016).
131. Sterpone, F., Bertonati, C., Briganti, G. & Melchionna, S. Water around thermophilic proteins: the role of charged and apolar atoms. *J Phys-Condens Mat* **22** (2010).
132. Kumar, S., Tsai, C.J. & Nussinov, R. Factors enhancing protein thermostability. *Protein Eng* **13**, 179-191 (2000).
133. Barillari, C., Taylor, J., Viner, R. & Essex, J.W. Classification of water molecules in protein binding sites. *J Am Chem Soc* **129**, 2577-2587 (2007).
134. Niu, C.T., Zhu, L.J., Xu, X. & Li, Q. Rational design of disulfide bonds increases thermostability of a mesophilic 1,3-1,4-beta-glucanase from *Bacillus terquilensis*. *Plos One* **11** (2016).
135. Meyer, J. et al. A hyperthermophilic plant-type [2Fe-2S] ferredoxin from *Aquifex aeolicus* is stabilized by a disulfide bond. *Biochemistry* **41**, 3096-3108 (2002).
136. Donald, J.E., Kulp, D.W. & DeGrado, W.F. Salt bridges: Geometrically specific, designable interactions. *Proteins* **79**, 898-915 (2011).
137. Malmberg, C.G. & Maryott, A.A. Dielectric constant of water from 0° to 100° C. *J Res Natl Inst Stand Technol* **56**, 8 (1956).
138. Bosshard, H.R., Marti, D.N. & Jelesarov, I. Protein stabilization by salt bridges: concepts, experimental approaches and clarification of some misunderstandings. *J Mol Recognit* **17**, 1-16 (2004).
139. Salari, R. & Chong, L.T. Effects of high temperature on desolvation costs of salt bridges across protein binding interfaces: Similarities and differences between implicit and explicit solvent models. *J Phys Chem B* **116**, 2561-2567 (2012).
140. Elcock, A.H. The stability of salt bridges at high temperatures: Implications for hyperthermophilic proteins. *J Mol Biol* **284**, 489-502 (1998).

141. Chan, C.H., Yu, T.H. & Wong, K.B. Stabilizing salt-bridge enhances protein thermostability by reducing the heat capacity change of unfolding. *Plos One* **6** (2011).
142. Yip, K.S.P. et al. The structure of *Pyrococcus Furiosus* glutamate-dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures. *Structure* **3**, 1147-1158 (1995).
143. Karshikoff, A. & Ladenstein, R. Ion pairs and the thermotolerance of proteins from hyperthermophiles: a 'traffic rule' for hot roads. *Trends Biochem Sci* **26**, 550-556 (2001).
144. Gromiha, M.M., Thomas, S. & Santhosh, C. Role of cation-pi interactions to the stability of thermophilic proteins. *Prep Biochem Biotech* **32**, 355-362 (2002).
145. Kannan, N. & Vishveshwara, S. Aromatic clusters: a determinant of thermal stability of thermophilic proteins. *Protein Eng* **13**, 753-761 (2000).
146. Sengupta, D. & Kundu, S. Role of long- and short-range hydrophobic, hydrophilic and charged residues contact network in protein's structural organization. *BMC Bioinform* **13** (2012).
147. Jaenicke, R. Protein stability and molecular adaptation to extreme conditions. *Eur J Biochem* **202**, 715-728 (1991).
148. Jaenicke, R. & Bohm, G. The stability of proteins in extreme environments. *Curr Opin Struct Biol* **8**, 738-748 (1998).
149. Macarthur, M.W. & Thornton, J.M. Influence of proline residues on protein conformation. *J Mol Biol* **218**, 397-412 (1991).
150. Wedemeyer, W.J., Welker, E. & Scheraga, H.A. Proline cis-trans isomerization and protein folding. *Biochemistry* **41**, 14637-14644 (2002).
151. Grathwohl, C. & Wuthrich, K. NMR-Studies of the rates of proline cis-trans isomerization in oligopeptides. *Biopolymers* **20**, 2623-2633 (1981).
152. Haney, P., Konisky, J., Koretke, K.K., LutheySchulten, Z. & Wolynes, P.G. Structural basis for thermostability and identification of potential active site residues for adenylate kinases from the archaeal genus *Methanococcus*. *Proteins* **28**, 117-130 (1997).
153. Watanabe, K., Hata, Y., Kizaki, H., Katsube, Y. & Suzuki, Y. The refined crystal structure of Bacillus cereus oligo-1,6-glucosidase at 2.0 Ångstrom resolution: Structural characterization of proline-substitution sites for protein thermostabilization. *J Mol Biol* **269**, 142-153 (1997).
154. Bogin, O. et al. Enhanced thermal stability of *Clostridium beijerinckii* alcohol dehydrogenase after strategic substitution of amino acid residues with prolines from the homologous thermophilic *Thermoanaerobacter brockii* alcohol dehydrogenase. *Protein Sci* **7**, 1156-1163 (1998).
155. Raymond-Bouchard, I. et al. Conserved genomic and amino acid traits of cold adaptation in subzero-growing Arctic permafrost bacteria. *Fems Microbiol Ecol* **94** (2018).
156. Matthews, B.W., Nicholson, H. & Becktel, W.J. Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. *Proc Natl Acad Sci USA* **84**, 6663-6667 (1987).
157. Stickle, D.F., Presta, L.G., Dill, K.A. & Rose, G.D. Hydrogen bonding in globular proteins. *J Mol Biol* **226**, 1143-1159 (1992).
158. McDonald, I.K. & Thornton, J.M. Satisfying hydrogen bonding potential in proteins. *J Mol Biol* **238**, 777-793 (1994).

159. Takano, K. et al. Proline effect on the thermostability and slow unfolding of a hyperthermophilic protein. *J Biochem* **145**, 79-85 (2009).
160. Balasco, N., Esposito, L., De Simone, A. & Vitagliano, L. Role of loops connecting secondary structure elements in the stabilization of proteins isolated from thermophilic organisms. *Protein Sci* **22**, 1016-1023 (2013).
161. Sakaguchi, M. et al. Role of proline residues in conferring thermostability on aqualysin I. *J Biochem* **141**, 213-220 (2007).
162. Nagel, Z.D. & Klinman, J.P. A 21(st) century revisionist's view at a turning point in enzymology. *Nat Chem Biol* **5**, 543-550 (2009).
163. Kamerlin, S.C.L. & Warshel, A. At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis? *Proteins* **78**, 1339-1375 (2010).
164. Glowacki, D.R., Harvey, J.N. & Mulholland, A.J. Taking Ockham's razor to enzyme dynamics and catalysis. *Nat Chem* **4**, 169-176 (2012).
165. Thomas, T. & Cavicchioli, R. Archaeal cold-adapted proteins: structural and evolutionary analysis of the elongation factor 2 proteins from psychrophilic, mesophilic and thermophilic methanogens. *FEBS Lett* **439**, 281-286 (1998).
166. Russell, R.J.M., Gerike, U., Danson, M.J., Hough, D.W. & Taylor, G.L. Structural adaptations of the cold-active citrate synthase from an Antarctic bacterium. *Structure* **6**, 351-361 (1998).
167. Wallon, G. et al. Sequence and homology model of 3-isopropylmalate dehydrogenase from the psychrotrophic bacterium *Vibrio* sp I5 suggest reasons for thermal instability. *Protein Eng* **10**, 665-672 (1997).
168. Gerike, U., Danson, M.J., Russell, N.J. & Hough, D.W. Sequencing and expression of the gene encoding a cold-active citrate synthase from an Antarctic bacterium, strain DS2-3R. *Eur J Biochem* **248**, 49-57 (1997).
169. D'Amico, S., Gerday, C. & Feller, G. Structural determinants of cold adaptation and stability in a large protein. *J Biol Chem* **276**, 25791-25796 (2001).
170. DasSarma, S., Capes, M.D., Karan, R. & DasSarma, P. Amino acid substitutions in cold-adapted proteins from *Halorubrum lacusprofundi*, an extremely halophilic microbe from Antarctica. *Plos One* **8** (2013).
171. Paredes, D.I., Watters, K., Pitman, D.J., Bystroff, C. & Dordick, J.S. Comparative void-volume analysis of psychrophilic and mesophilic enzymes: Structural bioinformatics of psychrophilic enzymes reveals sources of core flexibility. *BMC Struct Biol* **11** (2011).
172. D'Amico, S. et al. Molecular basis of cold adaptation. *Philos T R Soc B* **357**, 917-924 (2002).
173. Lonhienne, T., Gerday, C. & Feller, G. Psychrophilic enzymes: revisiting the thermodynamic parameters of activation may explain local flexibility. *Biochim Biophys Acta Prot Struct Mol Enzymol* **1543**, 1-10 (2000).
174. D'Amico, S., Collins, T., Marx, J.C., Feller, G. & Gerday, C. Psychrophilic microorganisms: challenges for life. *EMBO Rep* **7**, 385-389 (2006).
175. Johns, G.C. & Somero, G.N. Evolutionary convergence in adaptation of proteins to temperature: A_4-Lactate dehydrogenases of pacific damselfishes (Chromis spp.). *Mol Biol Evol* **21**, 314-320 (2004).
176. Collins, T., Meuwis, M.A., Gerday, C. & Feller, G. Activity, stability and flexibility in Glycosidases adapted to extreme thermal environments. *J Mol Biol* **328**, 419-428 (2003).
177. Xie, X.S. & Trautman, J.K. Optical studies of single molecules at room temperature. *Annu Rev Phys Chem* **49**, 441-480 (1998).

178. Wolfenden, R. & Snider, M.J. The depth of chemical time and the power of enzymes as catalysts. *Acc Chem Res* **34**, 938-945 (2001).
179. Jencks, W.P. Binding energy, specificity, and enzymic catalysis: the circe effect. *Adv Enzymol Relat Areas Mol Biol* **43**, 219-410 (1975).
180. Somero, G.N. Adaptation of enzymes to temperature: searching for basic "strategies". *Comp Biochem Phys B* **139**, 321-333 (2004).
181. Garsoux, G., Lamotte, J., Gerday, C. & Feller, G. Kinetic and structural optimization to catalysis at low temperatures in a psychrophilic cellulase from the Antarctic bacterium *Pseudoalteromonas haloplanktis*. *Biochem J* **384**, 247-253 (2004).
182. D'Amico, S., Sohier, J.S. & Feller, G. Kinetics and energetics of ligand binding determined by microcalorimetry: Insights into active site mobility in a psychrophilic alpha-amylase. *J Mol Biol* **358**, 1296-1304 (2006).
183. Berglund, G.I., Smalas, A.O., Outzen, H. & Willassen, N.P. Purification and characterization of pancreatic elastase from North Atlantic salmon (*Salmo salar*). *Mol Mar Biol Biotech* **7**, 105-114 (1998).
184. Tsigos, I., Velonia, K., Smonou, I. & Bouriotis, V. Purification and characterization of an alcohol dehydrogenase from the Antarctic psychrophile *Moraxella* sp. TAE123. *Eur J Biochem* **254**, 356-362 (1998).
185. Kim, S.Y. et al. Structural basis for cold adaptation - Sequence, biochemical properties, and crystal structure of malate dehydrogenase from a psychrophile *Aquaspirillium arcticum*. *J Biol Chem* **274**, 11761-11767 (1999).
186. Hoyoux, A. et al. Cold-adapted beta-galactosidase from the Antarctic psychrophile *Pseudoalteromonas haloplanktis*. *Appl Environ Microb* **67**, 1529-1535 (2001).
187. Lonhienne, T. et al. Modular structure, local flexibility and cold-activity of a novel chitobiase from a psychrophilic Antarctic bacterium. *J Mol Biol* **310**, 291-297 (2001).
188. Georlette, D. et al. in Physics and Chemistry Basis of Biotechnology. Cold-Adapted Enzymes. (eds. M. De Cuyper & J.W.M. Bulte) 177-196 (Springer, Dordrecht, 2001).
189. Aqvist, J., Kazemi, M., Isaksen, G.V. & Brandsdal, B.O. Entropy and enzyme catalysis. *Accounts Chem Res* **50**, 199-207 (2017).
190. Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450**, 964-972 (2007).
191. Sun, Z.T., Liu, Q., Qu, G., Feng, Y. & Reetz, M.T. Utility of B-factors in protein science: Interpreting rigidity, flexibility, and internal motion and engineering thermostability. *Chem Rev* **119**, 1626-1665 (2019).
192. Englander, S.W. Hydrogen exchange and mass spectrometry: A historical perspective. *J Am Soc Mass Spectr* **17**, 1481-1489 (2006).
193. Matyus, L., Szollosi, J. & Jenei, A. Steady-state fluorescence quenching applications for studying protein structure and dynamics. *J Photoch Photobio B* **83**, 223-236 (2006).
194. Lakowicz, J.R. Principles of fluorescence spectroscopy, Edn. 3. (Springer Science, 233 Spring Street, New York, NY 10013, USA; 1983).
195. Diez, M. et al. Proton-powered subunit rotation in single membrane-bound F0F1-ATP synthase. *Nat Struct Mol Biol* **11**, 135-141 (2004).
196. Ishima, R. & Torchia, D.A. Protein dynamics from NMR. *Nat Struct Biol* **7**, 740-743 (2000).
197. Thompson, M.C. et al. Temperature-jump solution X-ray scattering reveals distinct motions in a dynamic enzyme. *Nat Chem* **11**, 1058-1066 (2019).

198. Orellana, L. Large-scale conformational changes and protein function: Breaking the in silico barrier. *Front Mol Biosci* **6** (2019).
199. Arcus, V.L. et al. On the temperature dependence of enzyme-catalyzed rates. *Biochemistry* **55**, 1681-1688 (2016).
200. Hobbs, J.K. et al. Change in heat capacity for enzyme catalysis determines temperature dependence of enzyme catalyzed rates. *Acs Chem Biol* **8**, 2388-2393 (2013).
201. Socan, J., Purg, M. & Aqvist, J. Computer simulations explain the anomalous temperature optimum in a cold-adapted enzyme. *Nat Commun* **11**, 2644 (2020).
202. International Union of Biochemistry and Molecular Biology. Nomenclature Committee. & Webb, E.C. Enzyme nomenclature 1992 : recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the nomenclature and classification of enzymes. (Published for the International Union of Biochemistry and Molecular Biology by Academic Press, San Diego; 1992).
203. Lopez-Otin, C. & Bond, J.S. Proteases: multifunctional enzymes in life and disease. *J Biol Chem* **283**, 30433-30437 (2008).
204. Neurath, H. Protein-digesting enzymes. *Sci Am* **211**, 68-79 (1964).
205. Babe, L.M. & Craik, C.S. Viral proteases: evolution of diverse structural motifs to optimize function. *Cell* **91**, 427-430 (1997).
206. Rao, M.B., Tanksale, A.M., Ghatge, M.S. & Deshpande, V.V. Molecular and biotechnological aspects of microbial proteases. *Microbiol Mol Biol R* **62**, 597-635 (1998).
207. Marnett, A.B. & Craik, C.S. Papa's got a brand new tag: advances in identification of proteases and their substrates. *Trends Biotechnol* **23**, 59-64 (2005).
208. Rawlings, N.D. et al. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res* **46**, D624-D632 (2018).
209. Puente, X.S., Sanchez, L.M., Overall, C.M. & Lopez-Otin, C. Human and mouse proteases: A comparative genomic approach. *Nat Rev Genet* **4**, 544-558 (2003).
210. Craik, C.S., Page, M.J. & Madison, E.L. Proteases as therapeutics. *Biochem J* **435**, 1-16 (2011).
211. Turk, B. Targeting proteases: successes, failures and future prospects. *Nat Rev Drug Discov* **5**, 785-799 (2006).
212. Tavano, O.L. Protein hydrolysis using proteases: An important tool for food biotechnology. *J Mol Catal B-Enzym* **90**, 1-11 (2013).
213. Maurer, K.H. Detergent proteases. *Curr Opin Biotechnol* **15**, 330-334 (2004).
214. Rawlings, N.D. & Barrett, A.J. Evolutionary families of peptidases. *Biochem J* **290 (Pt 1)**, 205-218 (1993).
215. Neurath, H. Proteolytic enzymes, past and present. *Fed Proc* **44**, 2907-2913 (1985).
216. Perona, J.J. & Craik, C.S. Structural basis of substrate specificity in the serine proteases. *Protein Sci* **4**, 337-360 (1995).
217. Neurath, H. Evolution of proteolytic enzymes. *Science* **224**, 350-357 (1984).
218. Chasan, R. & Anderson, K.V. The role of easter, an apparent serine protease, in organizing the dorsal-ventral pattern of the *Drosophila* embryo. *Cell* **56**, 391-400 (1989).
219. Di Cera, E. Serine proteases. *IUBMB Life* **61**, 510-515 (2009).
220. Page, M.J. & Di Cera, E. Serine peptidases: classification, structure and function. *Cell Mol Life Sci* **65**, 1220-1236 (2008).

221. Bompard-Gilles, C. et al. Crystal structure of a D-aminopeptidase from *Ochrobactrum anthropi*, a new member of the 'penicillin-recognizing enzyme' family. *Structure* **8**, 971-980 (2000).
222. Balls, A.K. & Jansen, E.F. Stoichiometric inhibition of chymotrypsin. *Adv Enzymol Relat Subj Biochem* **13**, 321-343 (1952).
223. Schaffer, N.K., May, S.C., Jr. & Summerson, W.H. Serine phosphoric acid from diisopropylphosphoryl chymotrypsin. *J Biol Chem* **202**, 67-76 (1953).
224. Cohen, J.A., Jansz, H.S. & Oosterbaan, R.A. The chemical structure of the reactive group of esterases. *Biochim Biophys Acta* **20**, 402-403 (1956).
225. Dixon, G.H. & Neurath, H. The reaction of DFP with trypsin. *Biochim Biophys Acta* **20**, 572-574 (1956).
226. Schoellmann, G. & Shaw, E. Direct evidence for the presence of histidine in the active center of chymotrypsin. *Biochemistry* **2**, 252-255 (1963).
227. Ong, E.B., Shaw, E. & Schoellmann, G. The identification of the histidine residue at the active center of chymotrypsin. *J Biol Chem* **240**, 694-698 (1965).
228. Blow, D.M., Birktoft, J.J. & Hartley, B.S. Role of a buried acid group in the mechanism of action of chymotrypsin. *Nature* **221**, 337-340 (1969).
229. Everill, P., Sudmeier, J.L. & Bachovchin, W.W. Direct NMR observation and pKa determination of the Asp102 side chain in a serine protease. *J Am Chem Soc* **134**, 2348-2354 (2012).
230. Day, R.M. et al. Tautomerism, acid-base equilibria, and H-bonding of the six histidines in subtilisin BPN' by NMR. *Protein Sci* **12**, 794-810 (2003).
231. Hedstrom, L. Serine protease mechanism and specificity. *Chem Rev* **102**, 4501-4524 (2002).
232. Schechter, I. & Berger, A. On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun* **27**, 157-162 (1967).
233. Stroud, R.M. A family of protein-cutting proteins. *Sci Am* **231**, 74-88 (1974).
234. Naughton, M.A. & Sanger, F. Purification and specificity of pancreatic elastase. *Biochem J* **78**, 156-163 (1961).
235. Morihara, K., Oka, T. & Tsuzuki, H. Comparison of alpha-chymotrypsin and subtilisin BPN': size and specificity of the active site. *Biochem Biophys Res Commun* **35**, 210-214 (1969).
236. Wright, C.S. Comparison of the active site stereochemistry and substrate conformation in chymotrypsin and subtilisin BPN'. *J Mol Biol* **67**, 151-163 (1972).
237. Siezen, R.J. & Leunissen, J.A. Subtilases: The superfamily of subtilisin-like serine proteases. *Protein Sci* **6**, 501-523 (1997).
238. Barr, P.J. Mammalian subtilisins - The long-sought dibasic processing endoproteases. *Cell* **66**, 1-3 (1991).
239. Vandeven, W.J.M., Roebroek, A.J.M. & Vanduijnhoven, H.L.P. Structure and function of eukaryotic proprotein processing enzymes of the subtilisin family of serine proteases. *Crit Rev Oncogenesis* **4**, 115-136 (1993).
240. Rockwell, N.C. & Thorner, J.W. The kindest cuts of all: crystal structures of Kex2 and furin reveal secrets of precursor processing. *Trends Biochem Sci* **29**, 80-87 (2004).
241. Wright, C.S., Alden, R.A. & Kraut, J. Structure of subtilisin BPN' at 2.5 Ångstrom resolution. *Nature* **221**, 235-242 (1969).
242. Seidah, N.G. et al. The activation and physiological functions of the proprotein convertases. *Int J Biochem Cell Biol* **40**, 1111-1125 (2008).

243. Abifadel, M. et al. Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nat Genet* **34**, 154-156 (2003).
244. Hatzios, S.K. et al. Chemoproteomic profiling of host and pathogen enzymes active in cholera. *Nat Chem Biol* **12**, 268-+ (2016).
245. Howell, M., Dumitrescu, D.G., Blankenship, L.R., Herkert, D. & Hatzios, S.K. Functional characterization of a subtilisin-like serine protease from *Vibrio cholerae*. *J Biol Chem* **294**, 9888-9900 (2019).
246. Coutte, L., Antoine, R., Drobecq, H., Locht, C. & Jacob-Dubuisson, F. Subtilisin-like autotransporter serves as maturation protease in a bacterial secretion pathway. *EMBO J* **20**, 5040-5048 (2001).
247. Siezen, R.J., Kuipers, O.P. & deVos, W.M. Comparison of lantibiotic gene clusters and encoded proteins. *Anton Leeuw Int J G* **69**, 171-184 (1996).
248. Tripathi, L.P. & Sowdhamini, R. Genome-wide survey of prokaryotic serine proteases: Analysis of distribution and domain architectures of five serine protease families in prokaryotes. *BMC Genet* **9** (2008).
249. Kristjansson, M.M. in Thermostable Proteins. Structural Stability and Design. Thermostable subtilases (subtilisin-like serine proteinases). (eds. S. Sen & L. Nilsson) 67-104 (Taylor & Francis/CRC Press., Boca Raton, Florida.; 2012).
250. Blobel, G. & Dobberstein, B. Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma. *J Cell Biol* **67**, 835-851 (1975).
251. Payne, S.H. et al. Unexpected diversity of signal peptides in prokaryotes. *mBio* **3** (2012).
252. Wong, S.L. & Doi, R.H. Determination of the signal peptidase cleavage site in the preprosubtilisin of *Bacillus Subtilis*. *J Biol Chem* **261**, 176-181 (1986).
253. Fu, X., Inouye, M. & Shinde, U. Folding pathway mediated by an intramolecular chaperone - The inhibitory and chaperone functions of the subtilisin propeptide are not obligatorily linked. *J Biol Chem* **275**, 16871-16878 (2000).
254. Lee, Y.C., Miyata, Y., Terada, I., Ohta, T. & Matsuzawa, H. Involvement of NH_2-terminal pro-sequence in the production of active aqualysin-I (a thermophilic serine protease) in *Escherichia coli*. *Agr Biol Chem* **55**, 3027-3032 (1991).
255. Bryan, P. et al. Catalyis of a protein-folding reaction - Mechanistic implications of the 2.0 Ångstrom structure of the subtilisin-prodomain complex. *Biochemistry* **34**, 10310-10318 (1995).
256. Ohta, Y. et al. Pro-peptide as an intermolecular chaperone - renaturation of denatured subtilisin-E with a synthetic pro-peptide. *Mol Microbiol* **5**, 1507-1510 (1991).
257. Bryan, P.N. Prodomains and protein folding catalysis. *Chem Rev* **102**, 4805-4815 (2002).
258. Subbian, E., Yabuta, Y. & Shinde, U.P. Folding pathway mediated by an intramolecular chaperone: Intrinsically unstructured propeptide modulates stochastic activation of subtilisin. *J Mol Biol* **347**, 367-383 (2005).
259. Shinde, U.P., Liu, J.J. & Inouye, M. Protein memory through altered folding mediated by intramolecular chaperones. *Nature* **389**, 520-522 (1997).
260. Bryan, P. et al. Catalysis of a protein folding reaction: mechanistic implications of the 2.0 Å structure of the subtilisin-prodomain complex. *Biochemistry* **34**, 10310-10318 (1995).
261. Zhu, X.L., Ohta, Y., Jordan, F. & Inouye, M. Pro-sequence of subtilisin can guide the refolding of denatured subtilisin in an intermolecular process. *Nature* **339**, 483-484 (1989).

262. Marie-Claire, C., Yabuta, Y., Suefuji, K., Matsuzawa, H. & Shinde, U. Folding pathway mediated by an intramolecular chaperone: The structural and functional characterization of the Aqualysin I propeptide. *J Mol Biol* **305**, 151-165 (2001).
263. Terada, I., Kwon, S.T., Miyata, Y., Matsuzawa, H. & Ohta, T. Unique precursor structure of an extracellular protease, aqualysin I, with NH_2- and COOH-terminal pro-sequences and its processing in *Escherichia coli*. *J Biol Chem* **265**, 6576-6581 (1990).
264. Kim, D.W. & Matsuzawa, H. Requirement for the COOH-terminal pro-sequence in the translocation of aqualysin I across the cytoplasmic membrane in *Escherichia coli*. *Biochem Bioph Res Commun* **277**, 216-220 (2000).
265. Kim, J.Y. et al. Role of the C-terminal pro-domain of aqualysin I in the maturation and translocation of the precursor across the cytoplasmic membrane in *Escherichia coli*. *Biotechnol Lett* **24**, 433-439 (2002).
266. Nonaka, T. et al. The crystal structure of an oxidatively stable subtilisin-like alkaline serine protease, KP-43, with a C-terminal beta-barrel domain. *J Biol Chem* **279**, 47344-47351 (2004).
267. Foophow, T. et al. Crystal structure of a subtilisin homologue, Tk-SP, from *Thermococcus kodakaraensis*: Requirement of a C-terminal beta-jelly roll domain for hyperstability. *J Mol Biol* **400**, 865-877 (2010).
268. Foophow, T., Tanaka, S., Koga, Y., Takano, K. & Kanaya, S. Subtilisin-like serine protease from hyperthermophilic archaeon *Thermococcus kodakaraensis* with N- and C-terminal propeptides. *Protein Eng Des Sel* **23**, 347-355 (2010).
269. Shinde, U. & Thomas, G. Insights from bacterial subtilases into the mechanisms of intramolecular chaperone-mediated activation of furin. *Methods Mol Biol* **768**, 59-106 (2011).
270. Nour, N. et al. The cysteine-rich domain of the secreted proprotein convertases PC5A and PACE4 functions as a cell surface anchor and interacts with tissue inhibitors of metalloproteinases. *Mol Biol Cell* **16**, 5215-5226 (2005).
271. Ottmann, C. et al. Structural basis for Ca^{2+}-independence and activation by homodimerization of tomato subtilase 3. *Proc Natl Acad Sci USA* **106**, 17223-17228 (2009).
272. Bajorath, J., Hinrichs, W. & Saenger, W. The enzymatic-activity of proteinase-K Is controlled by calcium. *Eur J Biochem* **176**, 441-447 (1988).
273. Lin, S.J., Yoshimura, E., Sakai, H., Wakagi, T. & Matsuzawa, H. 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 (1999).
274. Smith, C.A., Toogood, H.S., Baker, H.M., Daniel, R.M. & Baker, E.N. Calcium-mediated thermostability in the subtilisin superfamily: The crystal structure of *Bacillus* Ak.1 protease at 1.8 Ångstrom resolution. *J Mol Biol* **294**, 1027-1040 (1999).
275. Braxton, S. & Wells, J.A. Incorporation of a stabilizing Ca^{2+}-binding loop into subtilisin BPN'. *Biochemistry* **31**, 7796-7801 (1992).
276. Tanaka, S.I. et al. Crystal structure of unautoprocessed precursor of subtilisin from a hyperthermophilic Archaeon - Evidence for Ca^{2+} induced folding. *J Biol Chem* **282**, 8246-8255 (2007).
277. Arnorsdottir, J., Kristjansson, M.M. & Ficner, R. Crystal structure of a subtilisin-like serine proteinase from a psychrotrophic *Vibrio* species reveals structural aspects of cold adaptation. *FEBS J* **272**, 832-845 (2005).

278. Barnett, B.L., Green, P.R., Strickland, L.C., Oliver, J.D., Rydel, T., Sullivan, J.F. Aqualysin I: the crystal structure of a serine protease from an extreme thermophile, *Thermus aquaticus* YT-1. (http://www.rcsb.org/structure/4DZT, 2012).
279. Voordouw, G. & Roche, R.S. The role of bound calcium ions in thermostable, proteolytic enzymes. I. Studies on thermomycolase, the thermostable protease fron the fungus *Malbranchea pulchella*. *Biochemistry* **14**, 4659-4666 (1975).
280. Tanaka, S.I., Matsumura, H., Koga, Y., Takano, K. & Kanaya, S. Four new crystal structures of Tk-subtilisin in unautoprocessed, autoprocessed and mature forms: Insight into structural changes during maturation. *J Mol Biol* **372**, 1055-1069 (2007).
281. Withers-Martinez, C. et al. The malaria parasite egress protease SUB1 is a calcium-dependent redox switch subtilisin. *Nat Commun* **5** (2014).
282. Bryan, P. et al. Energetics of folding subtilisin BPN'. *Biochemistry* **31**, 4937-4945 (1992).
283. Vartapetian, A.B., Tuzhikov, A.I., Chichkova, N.V., Taliansky, M. & Wolpert, T.J. A plant alternative to animal caspases: subtilisin-like proteases. *Cell Death Differ* **18**, 1289-1297 (2011).
284. Gallagher, T., Bryan, P. & Gilliland, G.L. Calcium-independent subtilisin by design. *Proteins* **16**, 205-213 (1993).
285. Strausberg, S.L. et al. Directed Evolution of a Subtilisin with Calcium-Independent Stability. *Bio Technol* **13**, 669-673 (1995).
286. Almog, O. et al. Crystal structure of calcium-independent subtilisin BPN ' with restored thermal stability folded without the prodomain. *Proteins* **31**, 21-32 (1998).
287. D. Kumar, S., N. Thakur, R. Verma and T. C. Bhalla Microbial proteases and application as laundry detergent additive. *Res J Microbiol* **3**, 661-672 (2008).
288. Zhao, H.M. & Arnold, F.H. Directed evolution converts subtilisin E into a functional equivalent of thermitase. *Protein Eng* **12**, 47-53 (1999).
289. Miyazaki, K. & Arnold, F.H. Exploring nonnatural evolutionary pathways by saturation mutagenesis: Rapid improvement of protein function. *J Mol Evol* **49**, 716-720 (1999).
290. Arnold, F.H., Giver, L., Gershenson, A., Zhao, H.M. & Miyazaki, K. Directed evolution of mesophilic enzymes into their thermophilic counterparts. *Ann NY Acad Sci* **870**, 400-403 (1999).
291. Shao, Z.X., Zhao, H.M., Giver, L. & Arnold, F.H. Random-priming in vitro recombination: an effective tool for directed evolution. *Nucleic Acids Res* **26**, 681-683 (1998).
292. Zhao, H.M., Giver, L., Shao, Z.X., Affholter, J.A. & Arnold, F.H. Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat Biotechnol* **16**, 258-261 (1998).
293. Betzel, C., Pal, G.P. & Saenger, W. Synchrotron X-ray data collection and restrained least-squares refinement of the crystal structure of proteinase K at 1.5 Å resolution. *Acta Crystallogr B* **44 (Pt 2)**, 163-172 (1988).
294. Arnorsdottir, J. et al. Characterization of a cloned subtilisin-like serine proteinase from a psychrotrophic *Vibrio* species. *Eur J Biochem* **269**, 5536-5546 (2002).
295. Bateman, A. et al. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* **47**, D506-D515 (2019).
296. Sutcliffe, I.C. A phylum level perspective on bacterial cell envelope architecture. *Trends Microbiol* **18**, 464-470 (2010).
297. Battista, J.R. Against all odds: The survival strategies of *Deinococcus radiodurans*. *Annu Rev Microbiol* **51**, 203-224 (1997).

298. Matsuzawa, H. et al. Purification and characterization of aqualysin-I (a thermophilic alkaline serine protease) produced by *Thermus aquaticus* YT-1. *Eur J Biochem* **171**, 441-447 (1988).
299. Betzel, C., Pal, G.P. & Saenger, W. 3-Dimensional structure of proteinase-K at 0.15-nm resolution. *Eur J Biochem* **178**, 155-171 (1988).
300. Helland, R., Larsen, A.N., Smalas, A.O. & Willassen, N.P. The 1.8 Ångstrom crystal structure of a proteinase K-like enzyme from a psychrotroph *Serratia* species. *FEBS J* **273**, 61-71 (2006).
301. Tiberti, M. & Papaleo, E. Dynamic properties of extremophilic subtilisin-like serine-proteases. *J Struct Biol* **174**, 69-83 (2011).
302. Kristjansson, M.M. & Manusson, O.T. Effect of lyotropic salts on the stability of a subtilisin-like proteinase from a psychrotrophic *Vibrio* species, proteinase K and aqualysin I. *Protein Peptide Lett* **8**, 249-255 (2001).
303. Sigurdardottir, A.G. et al. Characteristics of mutants designed to incorporate a new ion pair into the structure of a cold adapted subtilisin-like serine proteinase. *Biochim Biophys Acta* **1794**, 512-518 (2009).
304. Arnorsdottir, J., Helgadottir, S., Thorbjarnardottir, S.H., Eggertsson, G. & Kristjansson, M.M. Effect of selected Ser/Ala and Xaa/Pro mutations on the stability and catalytic properties of a cold adapted subtilisin-like serine proteinase. *Biochim Biophys Acta Proteins Proteom* **1774**, 749-755 (2007).
305. Arnorsdottir, J., Sigtryggsdottir, A.R., Thorbjarnardottir, S.H. & Kristjansson, M.M. Effect of proline substitutions on stability and kinetic properties of a cold adapted subtilase. *J Biochem* **145**, 325-329 (2009).
306. Chow, L.C. Solubility of calcium phosphates. *Monogr Oral Sci* **18**, 94-111 (2001).
307. Jonsdottir, L.B. et al. The role of salt bridges on the temperature adaptation of aqualysin I, a thermostable subtilisin-like proteinase. *Biochim Biophys Acta Proteins Proteom* **1844**, 2174-2181 (2014).
308. Arnorsdottir, J., Magnusdottir, M., Fridjonsson, O.H. & Kristjansson, M.M. The effect of deleting a putative salt bridge on the properties of the thermostable subtilisin-like proteinase, aqualysin I. *Protein Peptide Lett* **18**, 545-551 (2011).
309. Oskarsson, K.R. et al. A single mutation Gln142Lys doubles the catalytic activity of VPR, a cold adapted subtilisin-like serine proteinase. *Biochim Biophys Acta* **1864**, 1436-1443 (2016).
310. Sigtryggsdottir, A.R., Papaleo, E., Thorbjarnardottir, S.H. & Kristjansson, M.M. Flexibility of cold- and heat-adapted subtilisin-like serine proteinases evaluated with fluorescence quenching and molecular dynamics. *Biochim Biophys Acta Proteins Proteom* **1844**, 705-712 (2014).
311. Pettersen, E.F. et al. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612 (2004).
312. Wagner, S. et al. Tuning *Escherichia coli* for membrane protein overexpression. *Proc Natl Acad Sci USA* **105**, 14371-14376 (2008).
313. Bryksin, A.V. & Matsumura, I. Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. *Biotechniques* **48**, 463-465 (2010).
314. Kwon, S.T., Terada, I., Matsuzawa, H. & Ohta, T. Nucleotide sequence of the gene for aqualysin I (a thermophilic alkaline serine protease) of *Thermus aquaticus* YT-1 and characteristics of the deduced primary structure of the enzyme. *Eur J Biochem* **173**, 491-497 (1988).
315. Candiano, G. et al. Blue silver: a very sensitive colloidal Coomassie G-250 staining for proteome analysis. *Electrophoresis* **25**, 1327-1333 (2004).

316. Zaman, Z. & Verwilghen, R.L. Quantitation of proteins solubilized in sodium dodecyl sulfate mercaptoethanol Tris electrophoresis buffer. *Anal Biochem* **100**, 64-69 (1979).
317. Pace, C.N., Vajdos, F., Fee, L., Grimsley, G. & Gray, T. How to measure and predict the molar absorption coefficient of a protein. *Protein Sci* **4**, 2411-2423 (1995).
318. Delmar, E.G., Largman, C., Brodrick, J.W. & Geokas, M.C. Sensitive new substrate for chymotrypsin. *Anal Biochem* **99**, 316-320 (1979).
319. Mazurenko, S. et al. CalFitter: a web server for analysis of protein thermal denaturation data. *Nucleic Acids Res* **46**, W344-W349 (2018).
320. Mazurenko, S. et al. Exploration of protein unfolding by modelling calorimetry data from reheating. *Sci Rep* **7** (2017).
321. Brooks, B.R. et al. CHARMM: The biomolecular simulation program. *J Comput Chem* **30**, 1545-1614 (2009).
322. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J Chem Phys* **126** (2007).
323. Hess, B., Bekker, H., Berendsen, H.J.C. & Fraaije, J.G.E.M. LINCS: A linear constraint solver for molecular simulations. *J Comput Chem* **18**, 1463-1472 (1997).
324. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J Chem Phys* **98**, 10089-10092 (1993).
325. Vaz, W.L. & Schoellmann, G. Specific fluorescent derivatives of macromolecules. Reaction of dansyl fluoride with serine proteinases. *Biochim Biophys Acta* **439**, 194-205 (1976).
326. Good, N.E. et al. Hydrogen ion buffers for biological research. *Biochemistry* **5**, 467-477 (1966).
327. Taniguchi, M., Du, H. & Lindsey, J.S. PhotochemCAD 3: Diverse modules for photophysical calculations with multiple spectral databases. *Photochem Photobiol* **94**, 277-289 (2018).
328. Taniguchi, M. & Lindsey, J.S. Database of absorption and fluorescence spectra of >300 common compounds for use in PhotochemCAD. *Photochem Photobiol* **94**, 290-327 (2018).
329. Hartley, B.S. & Massey, V. The active centre of chymotrypsin. I. Labelling with a fluorescent dye. *Biochim Biophys Acta* **21**, 58-70 (1956).
330. Studier, F.W. & Moffatt, B.A. Use of bacteriophage-T7 RNA-polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* **189**, 113-130 (1986).
331. Studier, F.W. Use of bacteriophage-T7 lysozyme to improve an inducible T7 expression system. *J Mol Biol* **219**, 37-44 (1991).
332. Bessette, P.H., Aslund, F., Beckwith, J. & Georgiou, G. Efficient folding of proteins with multiple disulfide bonds in the *Escherichia coli* cytoplasm. *Proc Natl Acad Sci USA* **96**, 13703-13708 (1999).
333. Du, X. et al. Comparative thermal unfolding study of psychrophilic and mesophilic subtilisin-like serine proteases by molecular dynamics simulations. *J Biomol Struct Dyn* **35**, 1500-1517 (2017).

Paper I

BBA - Proteins and Proteomics 1867 (2019) 152–162

Contents lists available at ScienceDirect

BBA - Proteins and Proteomics

journal homepage: www.elsevier.com/locate/bbapap

Improved expression, purification and characterization of VPR, a cold active subtilisin-like serine proteinase and the effects of calcium on expression and stability

Kristinn R. Óskarsson, Magnús M. Kristjánsson*

Department of Biochemistry, Science Institute, University of Iceland, Reykjavík, Iceland

ARTICLE INFO

Keywords:
Subtilase
Serine proteinase
T7 expression
Calcium dependence
Temperature adaptation
Differential scanning calorimetry

ABSTRACT

Cloning into a pET 11a vector, followed by high-level expression of the cold adapted subtilase, VPR, utilizing the rhamnose titratable T7 system of Lemo21, resulted in a dramatic increase of soluble protein compared to the older system used. Expression optimization clearly shows the importance of calcium in the medium after induction, both for stability of the proteinase and cell health. Characterization of the purified enzyme obtained in a redesigned purification protocol which removed apparent RNA contaminants, resulted in a significantly higher value for k_{cat} than previously reported. The new recombinant protein exhibited slightly lower stability against thermal denaturation and thermal inactivation. Our results also indicate that two of the calcium binding sites have apparent binding constants in the mM range. Binding of calcium to the weaker of those two sites only affects resistance of the enzyme against irreversible thermal inactivation. Differential scanning calorimetry revealed a non-two-state denaturation process, with indication of presence of intermediates caused by unfolding of calcium binding motifs.

1. Introduction

VPR is a subtilisin-like serine proteinase of the proteinase K family, produced by the psychrophilic gram-negative bacterium *Vibrio* sp. PA-44 [1,2]. The gene encoding the proteinase consists of 1593 base pairs that yields the 530-residue precursor protein [3]. The precursor protein consists of 3 domains, a 139 residue N-terminal intramolecular chaperone, a 291-residue catalytic domain and a 100 residue C-terminal domain [3]. During folding and maturation of the protein the N-terminal domain is cleaved off in an intramolecular autocatalytic reaction and so is the C-terminal domain, leaving a 29.7 kDa proteinase [3]. In the present study we used a C-terminal truncated form of the enzyme, $VPR_{\Delta C}$, which we produced previously, for the purpose of mimicking the structure of the thermophilic homologue, aqualysin I, for comparative studies of temperature adaptation [4]. $VPR_{\Delta C}$, and several of its mutated variants have been studied in relation to temperature adaptation, activity/stability trade-offs and molecular flexibility [4–6]. In these studies, the properties of the psychrophilic $VPR_{\Delta C}$ have been compared to those of the structural homologue aqualysin I (AQUI) from the thermophile *Thermus aquaticus* [5–8]. Despite being closely structurally related, these enzymes differ greatly with respect to thermal stability and kinetic properties, reflecting their different temperature adaptation [5–8]. $VPR_{\Delta C}$ contains three calcium binding sites which have been shown to be highly important for structural stability of the enzyme [1,7,9]. In addition, it has two disulfide bridges to stabilize the protein structure [9]. These former studies of $VPR_{\Delta C}$ and its variants have been somewhat hampered by the relatively low level of enzyme production in the expression system we have used. We therefore aimed for applying a more efficient expression system for production of $VPR_{\Delta C}$ and variants for detailed structural and functional studies. In this paper we present an efficient procedure to produce large quantities of highly purified $VPR_{\Delta C}$. To that end we utilized the T7 system of *E. coli* Lemo21 cells [10] and demonstrated the importance of high concentration of calcium ions during protein expression and its interplay with rhamnose concentrations in the tunable T7 system of Lemo21. In the aftermath of this increased protein production, previously undetected impurities were observed. Therefore, a revised protein purification protocol was developed. For this reason, we report new values for measured kinetic constants and a comprehensive examination of other physical properties of the new recombinant proteinase such as the effects of calcium on structural stability and inactivation.

* Corresponding author.
E-mail address: mmk@hi.is (M.M. Kristjánsson).

https://doi.org/10.1016/j.bbapap.2018.11.010
Received 5 July 2018; Received in revised form 22 November 2018; Accepted 26 November 2018
Available online 28 November 2018

2. Materials and methods

2.1. E. coli strains, plasmids and media

For plasmid production, the *E. coli* strain XL10-Gold acquired from Agilent Technologies having the genotype: *TetrD(mcrA)183 D(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte [F´ proAB lacIqZDM15 Tn10 (Tetr) Amy Camr]* was used. The expression strain used was Lemo21(DE3) from New England BioLabs, genotype: *fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS/pLemo(CamR) λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5 pLemo = pACYC184-PrhaBAD-lysY*. Plasmids used in this study where pBAD containing the truncated form of VPR ($VPR_{\Delta C}$) [3,4] and pET-11a-d used in high level expression of the proteinase. Ingredients for media and plates were yeast extract, tryptone and agar from Bacto. For plasmid cultures and agar plates LB-Miller broth containing 0.1 mg/mL ampicillin (Sigma) was used. Lemo21 cells for expression were grown on LB-Miller plates containing 0.1 mg/mL ampicillin (Sigma), 0.03 mg/mL chloramphenicol (AppliChem) and 0.1% (*w/v*) L-rhamnose (Sigma). Liquid media for protein expression was 2xYT broth containing 0.1 mg/mL ampicillin, 0.03 mg/mL chloramphenicol, 400 μM isopropyl β-D-1-thiogalactopyranoside (IPTG) (AppliChem) and varying concentrations of L-rhamnose and Ca^{2+}.

2.2. Cloning

Cloning of the subtilase was performed via overlap extension PCR cloning [11] using Phusion® High-Fidelity DNA polymerase from NEB. The primers used for the amplification of the $VPR_{\Delta C}$ gene from the pBAD vector were the sense primer 5´CCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGTTAAAGAAAGTATTAAGTTGTTG´3 and the anti-sense primer 5´CAAGGGGTTATGCTAGTTATTGCTCAGCGGTTAAAAGTTTGCTTGGAGCGTC´3. These primers were acquired from TAG Copenhagen. For amplification of the VPR megaprimer a PCR reaction mixture containing 3.0 μM of sense and anti-sense primers, 1.5 nM of the pBAD vector containing the $VPR_{\Delta C}$ gene, 200 μM dNTP, 4 U of Phusion polymerase and using the GC buffer, in a total volume of 20 μL. The following program was used: mixture was pre-heated at 98 °C for 30 s then thermo-cycles ran 25-times in the following order: 98 °C for 10 s, 55 °C for 20 s, 72 °C for 50 s and repeat. The final elongation step was performed at 72 °C for 10 min followed by cooling to 4 °C. After amplification, the PCR product was gel purified from a 1% agar gel and the appropriate band was isolated using gel extraction kit from NEB. The purified megaprimer was then used to clone the gene into a pET-11-a-d vector. The PCR reaction mixture contained the purified megaprimer in a 250-fold excess over the pET-11a vector. Otherwise the mixture was as in the previous PCR reaction mix. The PCR program was as following: denaturation at 99 °C for 2 min followed by 20 cycles of 95 °C for 50 s, 60 °C for 50 s and 68 °C for 12 min. The final elongation step was carried out at 68 °C for 12 min and then the reaction was quenched by cooling to 4 °C. The remaining methylated DNA of the template was then digested by adding *Dpn*I (Thermo Scientific) to a final concentration of 1 U/μL and incubated at room temperature for 1.5 h, followed by transformation into XL10-Gold cells and then plated on LB-discs that were grown overnight at 37 °C. Single colonies were then picked and grown overnight in LB-Miller broth containing ampicillin at 37 °C and rotated at 230 rpm. Plasmid purification was carried out using the Monarch plasmid miniprep kit from NEB, following their instructions. To verify that the cloning was a success, samples were sent to Genewiz® for Sanger sequencing.

2.3. Transformation

Transformation of XL10-Gold was carried out by diluting PCR samples or purified plasmids 50-fold into Eppendorf tubes containing competent cells and kept on ice for 10 to 30 min. Heat shock was carried out at 42 °C for 60 s and again put on ice until being cooled down. Samples were then diluted 5 to 20-fold into LB-Miller broth and grown for 1 h at 37 °C. Cells were then plated onto LB-plates containing ampicillin and grown overnight at 37 °C. Transformation of Lemo21 was carried out by adding 50 ng – 100 ng of purified plasmid to 50 μL of competent cells and then following the same protocol as described above. Lemo21 cells were plated onto LB-plates containing ampicillin, chloramphenicol, rhamnose and then grown overnight at 37 °C.

2.4. Expression and optimization

For expression, single colonies of transformed Lemo21 cells were incubated in starter cultures of 2xYT broth and cultivated at 37 °C and 230 rpm overnight. Starter cultures were then diluted 50× into 2xYT expression cultures containing varying concentrations of L-rhamnose. Cultivation was performed in an Innova®44 incubator (New Brunswick) at 37 °C and 230 rpm in Erlenmeyer flasks with a volume of five times the media volume. Cells were grown until A_{600} had reached 0.4–0.8 A.U. At that point the culture was made 400 μM in IPTG and varying amounts of a sterile 4 M $CaCl_2$ solution was added. Cultivation was then carried out in an Innova®44R incubator at 18 °C and shaken at 230 rpm for 20 h. Cells were harvested by centrifugation at 4500 ×*g* at 10 °C for 15 min. Samples of growth culture were withdrawn for A_{600} and supernatant activity measurements. Harvested cell pellets were stored at −25 °C until used.

2.5. Protein purification

Cell lysis was performed by dissolving a pellet in 1/20 of its culture volume in 25 mM Tris (Sigma), 10 mM $CaCl_2$ (Sigma) at pH 8.0. Then dissolved DNase (Sigma) and lysozyme (Sigma) were added to a final concentration of 1 μg/mL and 1 mg/mL, respectively. The sample was gently shaken for 2 h and flash frozen in liquid nitrogen and thawed at room temperature with gentle shaking. These freeze/thaw cycles were repeated 3 times with the last thawing being carried out at 4 °C overnight. The sample was then centrifuged at 20,000 ×*g* at 4 °C for 45 min in a Beckman Coulter Avanti® J-26XP centrifuge. At that point activity measurements were performed to estimate expression yields. The rest of the supernatant was kept and made 80% saturated in $(NH_4)_2SO_4$ (Sigma) and centrifuged at 20,000 ×*g* at 4 °C for 45 min. The supernatant was discarded, and the precipitate dissolved in 1/20–1/10 of the original culture volume in 25 mM Tris, 10 mM $CaCl_2$ at pH 8.0. All column purification steps were performed using a BioLogic LP workstation from BioRad. The sample was loaded on to a N-carbobenzoxy-D-phenylalanyl-triethylenetetramine-Sepharose (z-D-Phe-TETA) column, previously equilibrated with 25 mM Tris, 10 mM $CaCl_2$ and pH 8.0 buffer. A wash was then preformed with 1 M NaCl Tris-buffer followed with another wash with buffer without salt. Elution was performed with 2 M guanidinium chloride in the Tris-buffer and the eluted fractions were diluted into 3 M $(NH_4)_2SO_4$ in the Tris-buffer in a 5:4 ratio. That sample was then loaded immediately onto a phenyl Sepharose column, previously equilibrated with 1 M $(NH_4)_2SO_4$ in Tris-buffer at pH 8.0. The $(NH_4)_2SO_4$ concentration was then lowered stepwise to 0 M, usually in two steps: to 0.4 M, and the down to 0 M. The elution buffer consisted of 50% ethylene glycol in the same Tris-buffer. The final step in the purification procedure was a run on a Q-sepharose column which was meant to concentrate the sample and ensure that no RNA contamination was present. The column was equilibrated with 25 mM Tris, 10 mM $CaCl_2$, pH 8.0, before the sample was applied. Wash was performed with the same buffer followed by elution using a linear NaCl gradient from 0 M to 0.5 M. Active portions were pooled together and made 20% ethylene glycol before being flash frozen in liquid nitrogen. All samples were then stored at −25 °C.

2.6. SDS-PAGE and estimation of protein purification

SDS-PAGE was performed on home cast 4%/12% gels. Mixing 40% stock solution of acrylamide/bis-acrylamide (37/1) (Sigma), dH_2O, sodium dodecyl sulfate (Thermo Scientific), tetraacetylethylenediamine (TEMED) (Merck) and ammonium persulfate (Sigma) [12] using Mini-PROTEAN tetra Cell from BioRad. Samples for SDS-PAGE electrophoreses were all inhibited with PMSF (Sigma) to a final concentration of 2.5 mM prior to mixing with 4× LDS sample buffer (Invitrogen) and denatured at 80 °C for 10 min. After electrophoresis, gels were fixed in 2.5% H_3PO_4 and 50% ethanol. Gels were rinsed in water two times before staining using Blue Silver stain [13]. The resulting gel was then photographed on a BioRad GelDoc™ EZ Imager. Specific activity of all samples was estimated by measuring activity against 0.5 mM Suc-AAPF-NH-Np (Bachem) and protein concentration was determined by using the Coomassie G-250 based assay as described by Zaman and Vervilghen [14]. UV wavelength scans on contaminants were performed on a Shimadzu UV-2700 spectrophotometer and ethidium bromide (Sigma) fluorescence measurement were conducted using a Horiba FluoroMax-4 spectrofluorometer, with excitation wavelength of 300 nm and fluorescence was monitored between 550 and 650 nm.

2.7. Effects of temperature and pH on activity

Activity assays were carried out using a Shimadzu UV-2700 spectrophotometer connected to a Shimadzu TCC-100 temperature controller. Activity was monitored at 410 nm over 15 s by using Suc-AAPF-NH-Np as a substrate. To determine the effects of pH on activity, six pH values were chosen and measured in triplicates. These were 3.8, 5.0, 6.2, 7.4, 8.6 and 9.8. Buffers used were glycine (Sigma) for pH 9.8. Tris for 8.6 and 7.4, MES (Sigma) for pH 6.0 and acetate (Sigma) for 5.0 and 3.8. The substrate concentration of 1 mM was used in those measurements. Apparent pKa was fitted and calculated using a sigmoidal curve fit using GraphPad Prism (GraphPad Software©). To estimate the activity dependence on temperature, assays were carried out at seven different temperatures in triplicates using 1 mM Suc-AAPF-NH-Np in 100 mM Tris,10 mM $CaCl_2$, pH 8.6 at the temperature being measured. To further classify the effects accompanied by the changes in pH and temperature, kinetic properties of the protease were determined by Michaelis-Menten assay at pH 7.4, 8.6 and 9.8, at 25 °C and at the temperatures 5 °C, 25 °C and 45 °C, at pH 8.6. Four different substrate concentrations between 0.1 and 1.0 mM were used for all experiments except for measurements done at pH 8.6 and 25 °C were seven substrate concentrations were used. All measurements were done in triplicates. Protein concentration was estimated by absorbance measurements at 280 nm using the molar attenuation coefficient calculated by the web-based program ProtParam (https://web.expasy.org/protparam/) (34,170 $M^{-1}\,cm^{-1}$) [15]. For substrate concentration estimation, assay solutions containing enzyme and substrate were incubated overnight at room temperature before diluting 10-fold and measuring absorbance at 410 nm, using 8480 $M^{-1}\,cm^{-1}$ for calculations [16]. All measurements were at minimum done in triplicates. The calculations of kinetic constants were performed by fitting the data to the Michaelis-Menten equation using the analysis software KaleidaGraph (Synergy Software©).

2.8. Stability measurements

$T_{50\%}$ was determined by following the thermal inactivation at several selected temperatures. From rates of inactivation we determined $T_{50\%}$, defined as the temperature at which half of the activity was lost after 30 min. Samples were prepared by dialysis against a buffer containing 25 mM Tris, 15 mM $CaCl_2$, 100 mM NaCl, 1 mM EDTA (Sigma) and pH 8.95 at 25 °C. Half-life ($t_{1/2}$) measurements on the other hand were carried out at $T_{50\%}$ under the same conditions as described, except at varying Ca^{2+} concentrations from 0 to 250 mM. Samples were diluted 50–100-fold into the incubation buffer just prior to measurements and aliquots were withdrawn and assayed for remaining activity. Data analysis was performed using GraphPad Prism fitting data with a double hyperbolic binding curve. To determine $T_{m,}$ we used circular dichroism (CD) spectroscopy and differential scanning calorimetry (DSC). Samples were prepared by inhibition with PMSF followed by dialysis overnight at 4 °C against 25 mM glycine (Sigma), with varying $CaCl_2$ concentrations, 100 mM NaCl and pH 8.6. Prior to T_m measurements a CD spectrum was recorded for each sample on a Jasco J-1100 from 250 nm down to 200 nm at 25 °C using a 1 mm cuvette. Melting curves were recorded at 222 nm with a heating of 1 °C/min from 25 °C to 85 °C. Protein concentration in these experiments was set to 0.1–0.3 mg/mL. Data analysis was performed by normalizing data following a fitting to a sigmoidal curve using Kaleidagraph. Melting points at varying calcium concentrations were fitted to a hyperbolic curve using GraphPad Prism. Concurrent to these measurements differential scanning calorimetry (DSC) scans were carried out using a MicroCal VP-DSC. Experiments were run at 1 °C/min temperature gradient from 15 °C to 95 °C. Protein concentrations were between 0.8 and 1.2 mg/mL. Data analysis for DSC was performed using the Origin® software. After buffer subtraction and concentration normalization, the data was converted into plots of excess heat capacity versus temperature. Differences in heat capacity of the folded and unfolded states were eliminated by manually fitting linear segments close to the peak and baseline created using the progress baseline option. Baseline was then subtracted and the deconvoluted data set fitted to non-two-state model [17]. Further analysis of data the sets included calculations of apparent melting points ($T_{m(DSC)}$) by fitting a cubic function to the peaks of the thermograms and solving the first derivative for the local maximum. Also, the rate of unfolding ($k_{(unfold)}$) was calculated:

$$k_{(unfold)} = \frac{vC_p}{Q_t - Q}$$

where v is the speed of the temperature gradient, C_p is the excess heat capacity at a given temperature, Q_t is the total heat evolved and Q is heat evolved at a given temperature [18,19]. The activation energy of unfolding (E_a) was calculated using the Arrhenius equation using calculated unfolding rates above 5% heat evolved and under 95% to avoid the relative high uncertainty in the end and beginning of the data. Half-times of unfolding $t_{1/2(DSC)}$ were then calculated from $k_{(unfold)}$ along with the activation free energy ($\Delta G^{\ddagger}$):

$$t_{1/2(DSC)} = \frac{\ln(2)}{k_{(unfold)}}$$

$$\Delta G^{\ddagger} = -RT\ln\left(\frac{k_{(unfold)} \times h}{k_b \times T}\right)$$

where R is the universal gas constant, T is the absolute temperature, $k_{(unfold)}$ is the rate of unfolding at a given temperature, h is the Planck constant and k_b is the Boltzmann constant. The activation enthalpy ($\Delta H^{\ddagger}$) was calculated by:

$$\Delta H^{\ddagger} = E_a - RT$$

and the activation entropy ($\Delta S^{\ddagger}$) was calculated as:

$$\Delta S^{\ddagger} = \frac{\Delta H^{\ddagger} - \Delta G^{\ddagger}}{T}$$

The thermodynamic activation parameters $\Delta G^{\ddagger}$, $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ were calculated at temperatures corresponding to the apparent melting points ($T_{m(DSC)}$).

3. Results

3.1. Cloning and expression optimization

Successful amplification of the VPR megaprimer was confirmed by

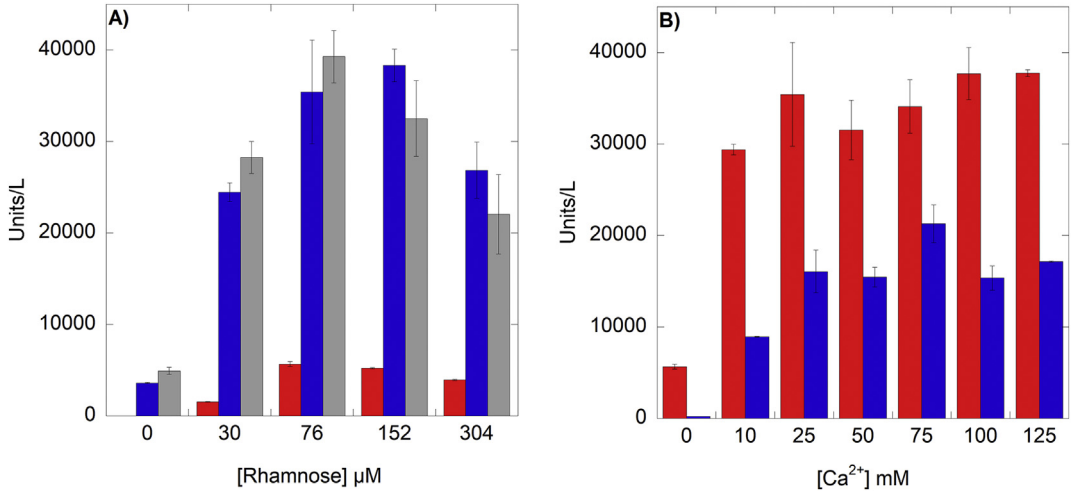


Fig. 1. A. Effects of rhamnose concentration on VPR expression yields without added calcium (red bars), in the presence of 25 mM $CaCl_2$ (blue bars) and 100 mM $CaCl_2$ (grey bars), recovered from harvested cells. **B.** Effects of $CaCl_2$ concentration on VPR expression at 76 μM rhamnose. Red bars represent activity units recovered from harvested cells. Blue bars represent activity units that leaked into the culture supernatant during cultivation. Error bars represent the standard deviation of the mean.

Table 1
Purification table for $VPR_{\Delta C}$ expressed in Lemo21. Activity per mL after column steps is calculated from the average activity from collected active fraction.

Step.	Volume (mL)	Concentration (mg/mL)	Activity (U/mL)	Units (U)	Total protein (mg)	Specific activity (U/mg)	Yield (%)	Purification (fold)
Soluble fraction.	50	5.19	580	28,715	257	112	100	1.0
80% $(NH_4)_2SO_4$ precipitate.	108	1.41	242	25,999	152	171	91	1.5
Z-D-Phe-TETA-Sepharose. * (Affinity column)	590	0.20 *(0.13) **	33	19,677	118 *(79) **	166 (250)	69	1.5 (2.2)
Phenyl-Sepharose. (Hydrophobic interaction column)	203	0.17	109	22,105	34	654	77	5.9
Q-Sepharose. (Anion exchange column)	50	0.46	407	20,165	23	876	70	7.8

* Protein concentration is likely be overestimated in these samples due to interference by ammonium sulfate and RNA present in the assay mixture in the Zaman and Vervilghen procedure [14].

** In brackets are corrected values for protein concentration, where a blank was used simulating the composition of the assay mixture with respect to guanidinium chloride and ammonium sulfate in Tris buffer. A possible interference by RNA in the sample has not been accounted for in these values, however.

agarose electrophoresis of the product and showed a band that was around 1600 bp (Supplementary. Fig. 1) which was in good agreement with the actual size of 1669 bp. After gel purification, overlap extension PCR cloning, transformation and plasmid purification, four of the six clones that were sequenced gave active clones. Optimization of conditions during overexpression of VPR was carried out by varying rhamnose concentration at a constant calcium concentration, and also where calcium concentrations were varied, but at selected rhamnose concentrations. In the absence of rhamnose toxic effects of the proteinase were clearly observed, as cultures grew slower and ended up at lower cell density (Supplementary. Table 1). Even though the tunable Lemo21 system has proven to be a powerful tool in the expression of difficult proteins [14] by using the titratable induction of T7 lysozyme via rhamnose addition, this was not enough for the expression of $VPR_{\Delta C}$. The presence of Ca^{2+} after induction by IPTG was just as crucial for overexpression of the proteinase. Experiments where rhamnose concentrations were varied at fixed Ca^{2+} concentrations showed that a final concentration of 76 μM rhamnose resulted in the highest yield of soluble protein at almost all Ca^{2+} concentrations (Fig. 1A) (Supplementary. Table 1). In the absence of rhamnose the difference in expression levels in the absence and presence of 100 mM Ca^{2+} was around 1000-fold. This is almost the same change as observed in the absence of both Ca^{2+} and rhamnose, compared to that in the presence of 76 μM rhamnose, but without Ca^{2+} (Fig. 1B) (Supplementary. Table 1). Evaluation of how Ca^{2+} affects expression levels shows that even small amounts of Ca^{2+} increase the yield of soluble protein drastically. At higher Ca^{2+} concentrations yields do only increase slightly, however. By using 100 mM Ca^{2+} and 76 μM rhamnose in 1 L cultures could yield up to 40,000 U as compared to 4 U without added calcium and rhamnose, and 5500 U at this rhamnose concentration, but in the absence of Ca^{2+}. The effect of rhamnose on cell growth was quite clear as in its absence A_{600} was ~ 2–4, but increased steadily to ~ 8–10 at higher concentrations, levelling off at around 152 μM rhamnose. The effect of calcium on cell density was much smaller, but a trend was observed indicating that higher calcium concentrations were beneficial for cell growth (Supplementary. Table 1). This indicates that the calcium present in the growth medium is not only aiding in correct folding and stabilization of VPR but is also increasing the resistance of Lemo21 against the toxic effects caused by VPR.

3.2. Purification

The purification procedure consists of a $(NH_4)_2SO_4$ salting out step, followed by three column chromatography steps, an affinity column using *Z*-D-Phe-TETA-Sepharose, followed by steps using phenyl-Sepharose and Q-Sepharose columns. During elution from the affinity column by 2 M guanidinium chloride a sharp peak appeared with low activity that was then followed by a broad peak with increased activity (Supplementary. Fig. 2A). It turned out that the sharp front of the peak contained contaminants. These contaminants could be separated from

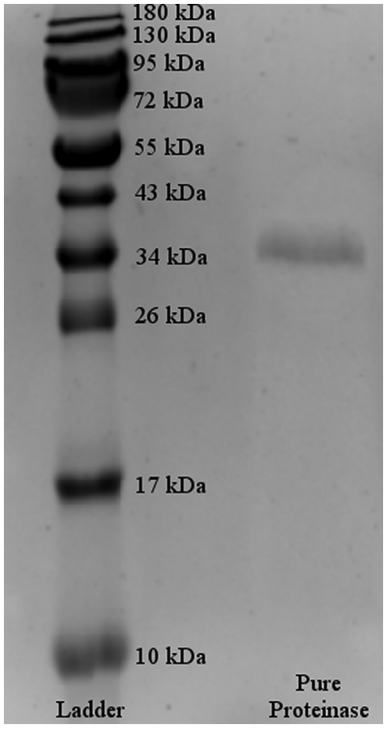


Fig. 2. SDS-PAGE of the purified $VPR_{\Delta C}$ expressed in Lemo21. Prestained Protein Ladder with indicated molecular masses in kilodaltons are indicated for size determination.

the proteinase on the phenyl-Sepharose column by lowering $(NH_4)_2SO_4$ from 1 M to 0.35–0.45 mM where the contaminant peak elutes from the column (Supplementary. Fig. 2B). The active peak fractions, which were eluted from the column with 50% ethylene glycol, were pooled and applied to a Q-sepharose column, which was eluted with a linear salt gradient which resulted in a single sharp peak (Supplementary. Fig. 2C), which migrated as a single band around 34–35 kDa on SDS-PAGE (Fig. 2). Yields of activity units using this purification protocol has been estimated to be from 60 to 75%. Specific activity of samples has been measured at 874 ± 130 U/mg at room temperature (Table 1). The contaminant that coeluted with the proteinase from the affinity column was examined further. A wavelength scan (Fig. 3A.) revealed a spectrum that is identical to that of pure nucleotides. Adding ethidium bromide to the solution made it highly fluorescent indicating that it contained either DNA or RNA strands. By adding RNase A and incubating for 20 min (Fig. 3B.) fluorescence was lost however, showing that the contaminant is most likely a single stranded RNA [20]. Attempts to identify the contaminants further indicated that the impurities peak collected from the phenyl-Sepharose column contained two rather small RNA fragments, but that were at least larger than 14 kDa as they did not pass through the cellulose membrane used for dialysis of samples.

3.3. Characterization of enzymatic properties

The purified recombinant $VPR_{\Delta C}$ was characterized with respect to effects of temperature and pH on its enzymatic properties. The temperature dependence of activity was measured and showed that in the presence of 10 mM Ca^{2+}, when monitoring activity only for 15 s, the activity steadily increased up to 65 °C (Fig. 4A). The measured pH profile is consistent with what to expect from a subtilisin-like serine protease [21] with an ionizable histidine residue in the active site having an apparent pKa of approximately 7 (Fig. 4B). As measurements were done at 1 mM Suc-AAPF-NH-Np these changes represent mostly changes in k_{cat}. Michaelis-Menten kinetic parameters measured on the recombinant $VPR_{\Delta C}$ produced and purified in this new procedure was found to have considerably higher values of k_{cat}, whereas the K_m values were not significantly changed as compared to older values [4]. This change in activity is most likely due to the removal of RNA from samples as A_{280} measurements have been used to estimate protein concentration, leading to overestimation of protein if RNA is present. With increasing temperature, the turnover number increases to around 550 s^{-1} at 45 °C from 41 s^{-1} at 5 °C but is accompanied by a concomitant twofold increase in K_m (Table 2). These trends are in line with previously reported results for wild type VPR and other similar enzymes [1,22]. The catalytic efficiency (kcat/K_m) also increased with pH from 7.4 to 9.4 when measured at 25 °C, both as a result of increased k_{cat} and lowered K_m (Table 3). This effect of pH may at least partly be attributed to deprotonation of the active site histidine.

3.4. The stability of $VPR_{\Delta C}$ and the effects of Ca^{2+} on structural stability and inactivation

Measurements of thermal inactivation rates showed that $VPR_{\Delta C}$ has a $T_{50\%}$ value of 53.8 ± 0.4 °C obtained from Arrhenius plots (Supplementary. Fig. 3). This value is in line with results reported by Kristjansson et al. [1] for the wild type, but is around 2.5 °C lower than

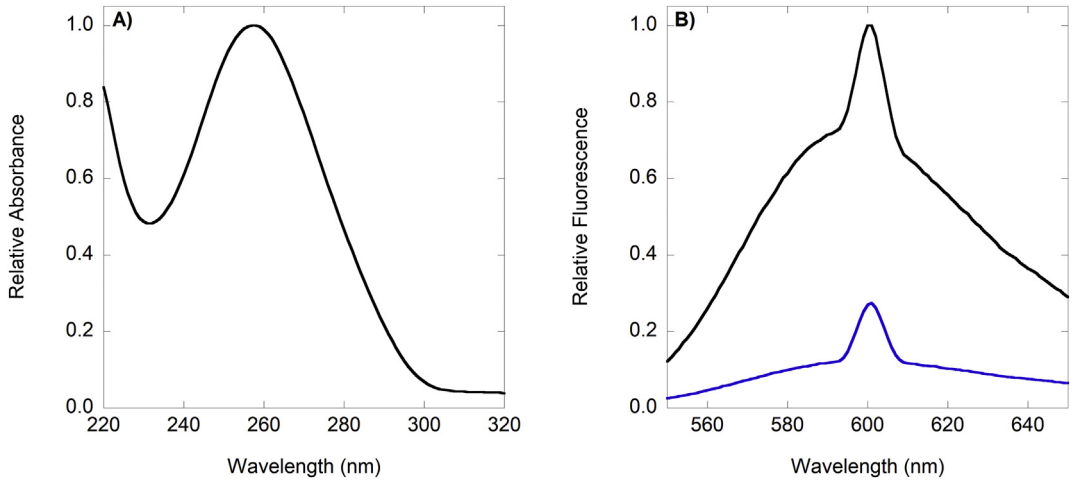


Fig. 3. A. Absorbance spectrum of contaminants recovered from phenyl-sepharose column. **B.** Fluorescence emission spectrum of ethidium bromide containing contaminants, before RNase A digestion (black line) and after RNase A digestion (blue line). Sharp peaks observed are second order diffractions of the scattered 300 nm excitation light.

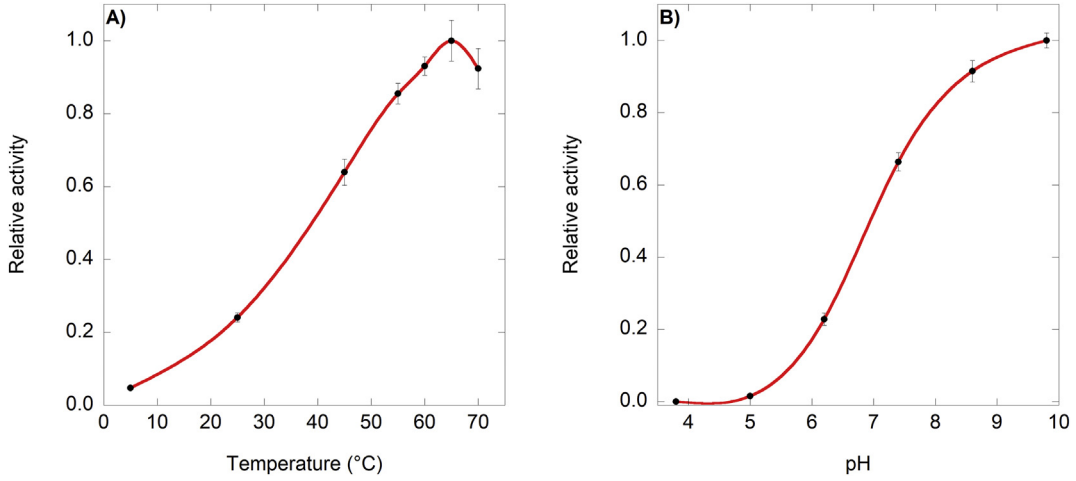


Fig. 4. A. The effect of temperature on the activity of VPR measured in Tris-buffer containing 10 mM $CaCl_2$. Buffer was adjusted to pH 8.6 at each measured temperature. The average of maximum activity measured was set to 1. **B.** The effect of pH on the activity of VPR in the presence of 10 mM $CaCl_2$. The average of maximum activity measured was set to 1. Error bars represent one standard deviation of the mean.

Table 2
Kinetic parameters for the activity of $VPR_{\Delta C}$ against Suc-AAPF-NH-Np at varying temperatures.

Conditions. Temperature (°C) / pH	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (s^{-1} mM^{-1})
5 °C / 8.6	40.8 ± 2.1	0.115 ± 0.002	354 ± 24
25 °C / 8.6	225.7 ± 12.0	0.178 ± 0.016	1239 ± 149
45 °C / 8.6	552.3 ± 32.6	0.227 ± 0.038	2471 ± 347

Table 3
Kinetic parameters for the activity of $VPR_{\Delta C}$ against Suc-AAPF-NH-Np at varying pH values.

Conditions. Temperature (°C) / pH	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (s^{-1} mM^{-1})
25 °C / 7.4	193.1 ± 4.1	0.198 ± 0.004	984 ± 9
25 °C / 8.6	225.7 ± 12.0	0.178 ± 0.016	1239 ± 149
25 °C / 9.4	248.7 ± 14.3	0.133 ± 0.018	1948 ± 240

what has been reported for the truncated $VPR_{\Delta C}$ variant [4]. The reason for this discrepancy is not known, but experiments where the proteinase was incubated with various RNA-contaminant concentrations did not show any change in the rate of thermal inactivation. $VPR_{\Delta C}$ has three calcium binding sites [9] and binding of calcium ions has been shown to be highly important for the correct fold, thermal stability and protection against thermal inactivation of subtilases [2,7,23–25]. To estimate the contribution of Ca^{2+} binding for stabilization of the enzyme against thermal inactivation, we measured $t_{1/2}$, the half-life of inactivation, at different Ca^{2+} concentrations at $T_{50\%}$ (53.8 °C). Results of these experiments revealed that the resistance of the enzyme against thermal inactivation was still increasing at 250 mM Ca^{2+}, the highest calcium concentration used in these experiments (Fig. 5B). Values for melting temperatures (T_m) obtained from melting curves measured by CD were not increased however, at Ca^{2+} concentrations higher than 15 mM (Fig. 5A and 6B). Thermal stability of $VPR_{\Delta C}$ was also determined by DSC in the presence of 1 mM, 15 mM and 100 mM $CaCl_2$ (Fig. 7A). Prior to these experiments CD wavelength scans were recorded of the samples and revealed no observable changes in the secondary structure of the enzyme over that calcium concentration range (Fig. 6A). Plots of excess heat capacity exhibited changes over the calcium range, where at higher calcium concentration the total calorimetric heat increased.

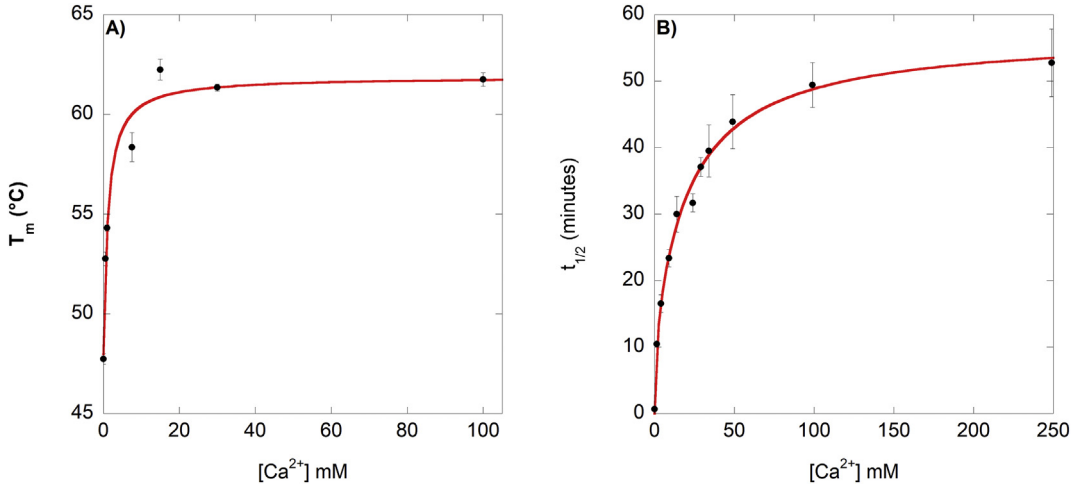


Fig. 5. A. T_m at different $CaCl_2$ concentrations. Black dots represent average experimental values whereas the red line represents the hyperbolic fit. **B).** $t_{1/2}$ at different $CaCl_2$ concentration measured at $T_{50\%}$ (53.8 °C). Black dots represent average experimental values whereas the red line represents the double hyperbolic fit. Error bars represent one standard deviation of the mean.

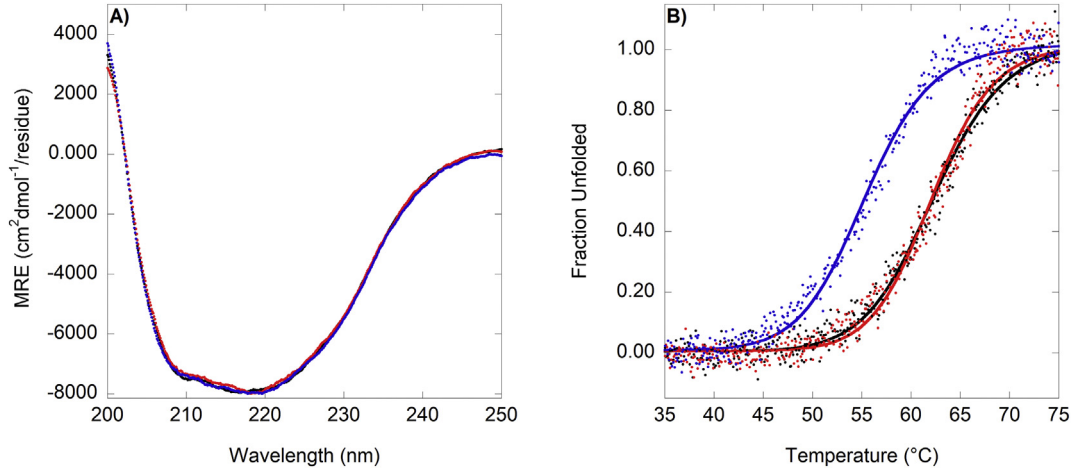


Fig. 6. A. CD wavelength scan of $VPR_{\Delta C}$ with varying calcium concentrations at 25 °C; 1 mM (blue dots), 15 mM (black dots) and 100 mM (red dots) $CaCl_2$. **B.** Normalized melting curves of $VPR_{\Delta C}$ at varying calcium concentrations recorded by CD measurements. The buffer used was 25 mM glycine, 100 mM NaCl and pH 8.6 and with varying $CaCl_2$ concentrations;1 mM (blue), 15 mM (black) and 100 mM (red) $CaCl_2$. Dots represent normalized data points were solid lines represent sigmoidal fits.

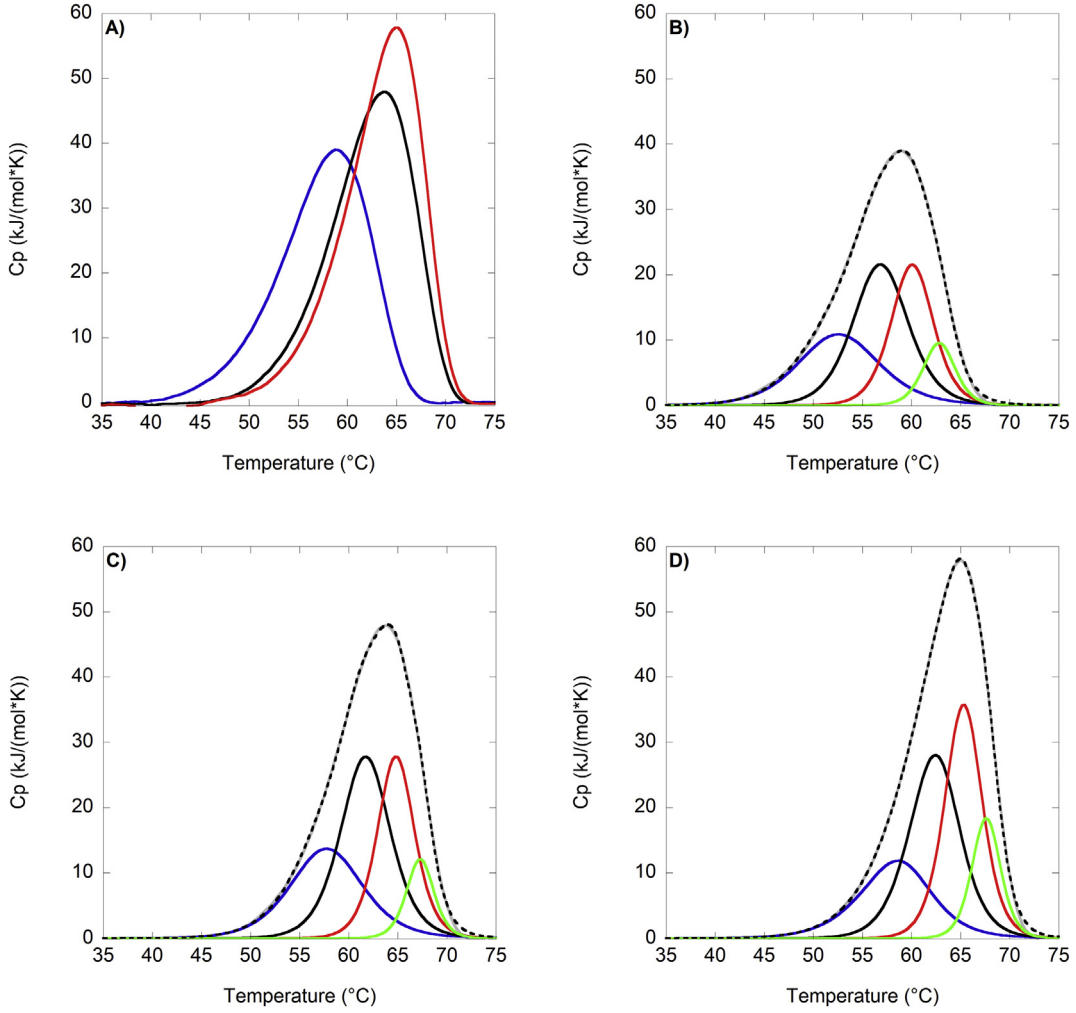


Fig. 7. A. Differential scanning calorimetric melting transitions for $VPR_{\Delta C}$. Measurements were carried out in 25 mM glycine, 100 mM NaCl and pH 8.6 and at 1 mM $CaCl_2$ (blue), 15 mM $CaCl_2$ (black) and 100 mM $CaCl_2$ (red). Scan rate was at 1 °C/min. **B.** Deconvolution of excess heat during the unfolding of $VPR_{\Delta C}$ at 1 mM $CaCl_2$. Black dotted line represents the best fit of the total excess heat, blue solid line represents the excess heat of the first transition, black solid line represents the excess heat of the second transition, red solid line represents the excess heat of the third transition, green solid line represents the excess heat of the fourth transition and the grey solid line represents the total excess heat measured. **C).** Deconvolution of excess heat during the unfolding of $VPR_{\Delta C}$ at 15 mM $CaCl_2$. **D).** Deconvolution of excess heat during the unfolding of $VPR_{\Delta C}$ at 100 mM $CaCl_2$.

Table 4
Melting temperatures as measured by CD and the four T_m recorded on DSC along with the calorimetric heats (ΔH) for each transition and the total calorimetric heat. Standard deviation of the mean of experimental data is shown for every value except for total heat of denaturation where the normal sum distribution is shown.

$[Ca^{2+}]$ mM	T_m (°C) (CD)	T_{m1} (°C)	ΔH_1 (kJ/mol)	T_{m2} (°C)	ΔH_2 (kJ/mol)	T_{m3} (°C)	ΔH_3 (kJ/mol)	T_{m4} (°C)	ΔH_4 (kJ/mol)	Total ΔH (kJ/mol)
1 mM	54.3 ± 0.2	52.5 ± 0.9	118 ± 22	56.8 ± 0.7	160 ± 7	60.1 ± 0.6	123 ± 20	62.9 ± 0.5	41 ± 12	442 ± 16
15 mM	62.2 ± 0.5	57.8 ± 0.6	133 ± 10	61.7 ± 0.4	187 ± 29	64.6 ± 0.5	140 ± 15	67.3 ± 0.2	45 ± 7	504 ± 17
100 mM	61.8 ± 0.3	58.1 ± 1.4	113 ± 11	62.3 ± 1.1	190 ± 1	65.3 ± 0.7	180 ± 19	67.7 ± 0.6	70 ± 15	553 ± 13

Table 5
The thermodynamic activation parameters $\Delta G^{\ddagger}$, $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ calculated from DSC thermograms at their respective apparent melting points ($T_{m(DSC)}$) along with the half time at that temperature.

$[Ca^{2+}]$ mM	$T_{m(DSC)}$ (°C)	$t_{1/2(DSC)}$ (min^{-1})	$\Delta G^{\ddagger}$ (kJ/mol)	$\Delta H^{\ddagger}$ (kJ/mol)	$\Delta S^{\ddagger}$ (J/(mol*K))
1 mM	58.9 ± 0.2	3.1 ± 0.1	97.0 ± 0.1	211 ± 3	343 ± 8
15 mM	63.7 ± 0.3	2.5 ± 0.1	98.1 ± 0.2	238 ± 12	413 ± 35
100 mM	65.0 ± 0.3	2.4 ± 0.1	98.2 ± 0.1	260 ± 9	478 ± 25

However, no changes were observable between unfolding traces during CD melting between 15 mM and 100 mM Ca^{2+} (Fig. 6B). Fitting the DSC melting curves to a non-two-state model gave the best the fits. We observed that a model with four transitions best described the thermal denaturation process of $VPR_{\Delta C}$, as it adequately fits the curvature of the data and the observed effects of calcium (Fig. 7B-D). The largest heat of denaturation was associated with the second transition observed, that also has a T_m corresponding to that measured by CD (Table 4.). The calorimetric heat of denaturation of the second transition also only shows difference between 1 mM and 15 mM calcium, following the same trend as the melting temperatures determined by CD, indicating effects of calcium binding occurring in that concentration range on the α/β core structure of the protein. The calorimetric heat of denaturation of the first transition is mostly unaffected by calcium concentration but has a T_m around 4 °C lower than the melting point measured by CD. Heat of denaturation of transitions three and four that occur 3 °C and 6 °C above the CD melting point are very responsive to calcium in the solution over all calcium concentrations measured (Fig. 7B-D). Effects of calcium binding on activation energies of unfolding (Table 5.) tell a similar tale, as there is no change in activation free energy ($\Delta G^{\ddagger}$) between 15 mM and 100 mM calcium but an increase of 1.1 kJ/mol from 1 mM to 15 mM. Over the calcium range measured activation enthalpy ($\Delta H^{\ddagger}$) increases, however this is followed by an increase in activation entropy ($\Delta S^{\ddagger}$), that after 15 mM calcium starts to compensate for the effect of the increasing activation enthalpy ($\Delta H^{\ddagger}$). Binding of calcium ions to specific binding sites clearly is the single most important factor contributing to the stability of $VPR_{\Delta C}$ as is clearly demonstrated by these results. This is further underlined by the fact that our attempts to measure the stability of calcium depleted $VPR_{\Delta C}$, by either DSC or CD, were unsuccessful. Samples were almost completely denatured during dialysis against a calcium free T_m buffer at 4 °C which contained 1 mM EDTA prior to CD and DSC experiments. The melting temperature was however measurable on CD with no added calcium and no EDTA (47.7 ± 0.3 °C). We contribute this destabilization in the presence of EDTA to chelation of a tightly bound calcium ion by EDTA, rather than destabilizing effects caused by EDTA on the structure, as $t_{1/2}$ experiments with no added calcium but in the presence of 1 mM, 10 mM and 100 mM EDTA at 25 °C did not show any change in the rate of inactivation.

4. Discussion

The successful cloning and expression of $VPR_{\Delta C}$ in Lemo21 cells has opened the door for more protein production, making it feasible to carry out experiments demanding higher protein concentrations as expression levels are up to 100 times higher than in the older system. The expression levels were found to be highly dependent on optimized concentrations of calcium and rhamnose in the growth medium, prompting the question of the role of high concentration of calcium in the culture medium. This system is now being used in high-level production of various mutants of VPR, some of which were non-expressible in the older system. Some of the more unstable variants have even been found to be more dependent on high levels of calcium in the media.

Characterization of the enzymatic properties of the new recombinant VPR shows that after implementing a reformed purification protocol, kinetic analysis yielded significantly higher k_{cat} values for the enzyme against its synthetic substrate, than previously had been reported. This is attributed to RNA contaminants which coeluted with the proteinase from the affinity column and would interfere with A_{280} measurements when present in samples. The activity of the purified proteinase is very responsive to temperature changes between 5 °C and 45 °C, showing a 14-fold increase in turnover number, but with a concomitant 2-fold increase in K_m. Changes accompanied by changes in pH between 7.4 and 9.8 show a one third increase in k_{cat} and similar decrease in K_m, these changes being attributed to changes in protonation of the active His70 residue, in addition to neighboring ionizable residues possibly affecting substrate binding to some extent.

Stability against thermal unfolding and inactivation however, has been measured consistently 2–3 °C lower than previously reported [4,26] with no explicit explanation. The observed difference between the effect of calcium concentration on rates of thermal inactivation ($t_{1/2}$) and global thermal stability (T_m) as determined by CD, strongly suggests that two low affinity calcium binding sites may be present in the enzyme, with calcium binding affinities in the mM range. This is best demonstrated from the comparison of $t_{1/2}$ and T_m against Ca^{2+} concentration. Plotting a hyperbolic curve to melting temperatures against calcium concentration resolves in an apparent K_D of ~ 1 mM (Fig. 5A), whereas plotting $t_{1/2}$ as a function of calcium concentration and fitting the data to a double hyperbolic curve (Fig. 5B) can be resolved to two apparent binding constants of ~ 1 mM and ~ 22 mM, respectively. Thus, binding site with an apparent K_D of ~1 mM is crucial for the thermal stability of the secondary structure and a second binding site with an apparent K_D of ~ 22 mM is mostly important in relation to resistance against thermal inactivation, rather than the stability of the α/β core structure of VPR. This explains why such high calcium concentrations are beneficial in the culture media during expression in addition to some beneficial effects on cell density observed at higher calcium levels. The last calcium binding site then likely has a K_D in the sub-mM range as indicated by the loss of stability of the enzyme in the presence of EDTA but without calcium, as compared to conditions where neither calcium nor EDTA were added.

Changes to activation energies of unfolding determined by DSC at

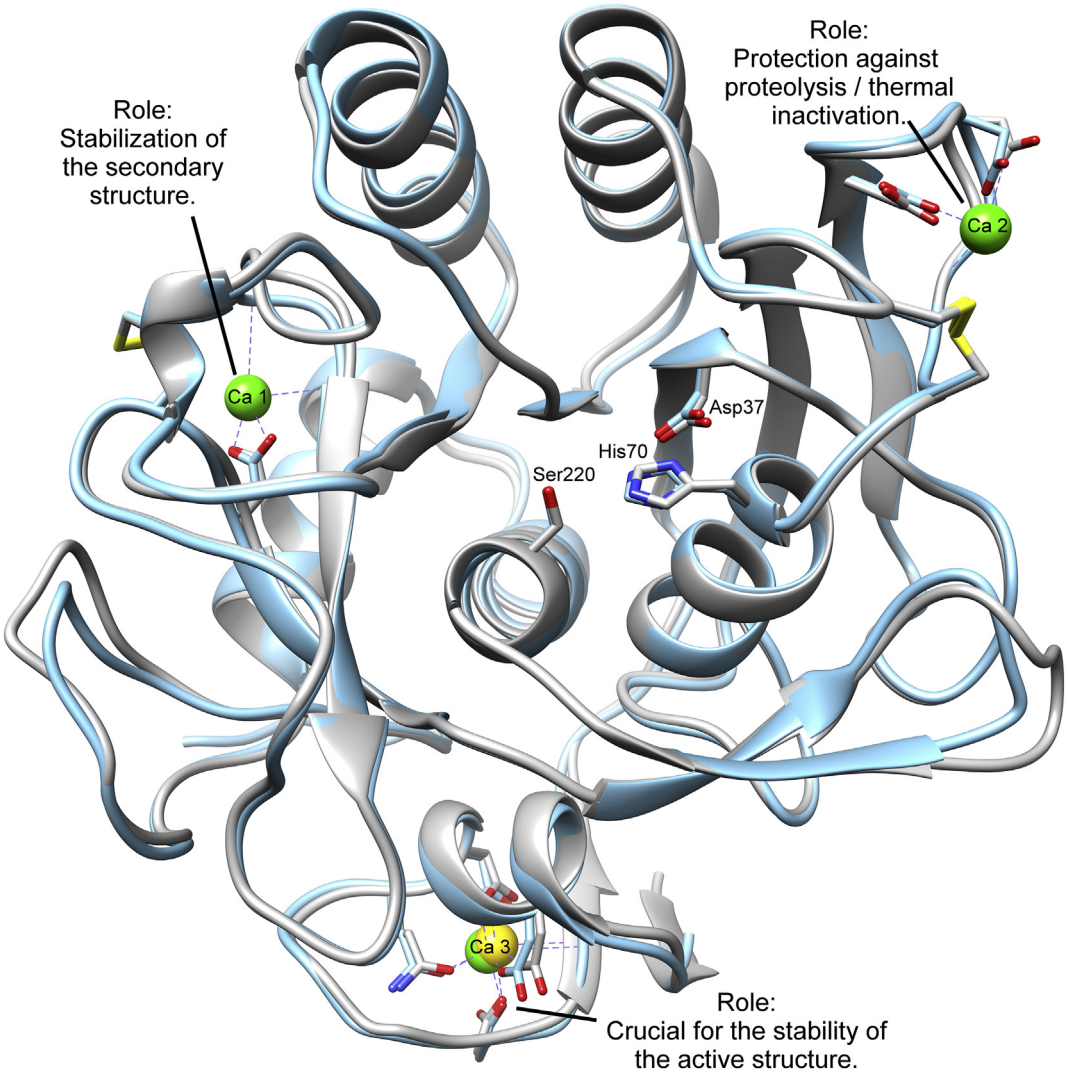


Fig. 8. Superimposed three-dimensional structures of VPR (PDB ID: 1SH7) and SPRK (PDB ID: 2B6N). VPR is coloured light blue with green calcium atoms, whereas SPRK is coloured grey with a golden calcium atom.

different calcium concentrations show clear effects of calcium binding, as up to 100 mM calcium the activation enthalpy ($\Delta H^{\ddagger}$) is increasing indicative of new bonds forming or strengthening preexisting interactions (Table 5). The accompanied increase in activation entropy ($\Delta S^{\ddagger}$) could then be attributed to a more ordered native state due to the fixation of the weaker calcium binding motifs. As the activation free energy ($\Delta G^{\ddagger}$) does not increase after 15 mM calcium due to entropy/enthalpy compensation would then indicate that binding to the weakest calcium binding site is entropically unfavorable for the stability of $VPR_{\Delta C}$. Melting profiles of the protein as determined by DSC revealed a rather complex denaturation process. Based on a best fit to a non-two-state model of the unfolding curves, a model with four transitions best described the thermal denaturation process of $VPR_{\Delta C}$ (Fig. 7B-C). Comparison of the T_m values of the four-transition model obtained from the DSC data, showed that the second transition yielded T_m values corresponding to those measured by CD (Table 4). This suggests that the second transition, which also has the highest heat of denaturation, may reflect on the unfolding of the major part of the secondary structure. When denaturation was measured at different calcium concentrations, it was observed that the heat of the first transition was little affected by calcium concentration, but the last two (T_{m3} and T_{m4}) were highly responsive to calcium concentration. This may suggest that these latter transitions, which become more prominent as the calcium concentration is increased, may reflect on unfolding events at one or more of the calcium binding sites or neighboring parts of the protein stabilized by these binding sites. These observations could then lend some credence to a hypothetical Lumry Eyring style model and written out as:

$$N \rightleftharpoons P \leftrightarrow D_1 \leftrightarrow D_2 \rightarrow U$$

were the double headed arrows are representing either a reversible or an irreversible step. The native state (N) first undergoes partial unfolding. The resulting transition state (P) is then a starting point for the denaturing process ultimately leading to a near complete loss of the secondary structure (D_1). This state is however still retaining some calcium binding capabilities. These calcium binding motifs are then last parts to denature (D_2), ending in the completely unfolded state of $VPR_{\Delta C}$ (U).

The crystal structure of SPRK, a proteinase K-like subtilase from a psychrotrophic *Serratia* species, was reported to have only one calcium binding site occupied [22]. Structural similarity, with respect to geometries and amino acid composition, of the calcium sites between VPR and SPRK suggests however, that calcium binding site 2 (Ca-2), as well as the Ca-1 site should also be present in SPRK [9]. The presence of a Ca^{2+} ion at the site in the crystal structure of SPRK corresponding to the Ca-3 site in VPR, strongly suggests that the Ca-3 site is the strongest binding site in the protein (Fig. 8). The crystal structure shows that the calcium ion at this site is coordinated by six different protein atoms [9]. AQUI, a thermostable homologue of VPR contains two calcium binding

sites, corresponding to Ca-1 and Ca-3 in VPR [27]. The Ca-3 binding site shares a high degree of similarity in geometry and protein ligands to that of VPR and so does the Ca-1 site although some differences in orientation of the calcium ion are observed in the binding site [9,27]. AQUI has also been shown to have a weak calcium binding site that is important for thermostabilization of the protein structure [28]. It is highly likely that these weak binding calcium sites, that are important for the stability of the secondary structure, are the same in both VPR and AQUI, and based on the arguments discussed above, it would suggest that this site is the Ca-1 site. Proteinase K also has the Ca-1 binding site, which has been shown to have a higher affinity for calcium [25] than a second binding site in the protein. The second binding site is located close to the N-terminus, bridging that part of the protein to the C-terminus [29]. This second site is not present in VPR and other closely related subtilases. Outside of the proteinase K family Ca-1 is often considered to be a weak binding site such as in the case of subtilisin BPN´ [24]. In addition, structures belonging to a thermostable subtilase from *Bacillus* species Ak.1 and thermitase from *Thermoactinomyces vulgaris* the Ca-1 site was occupied with a sodium ion, while a calcium ion was bound in the Ca-2 binding site [30,31]. So vastly different affinities for calcium have been observed for similar sites in the superfamily of subtilases.

In a molecular dynamics simulations study where VPR with the Ca-2 site occupied by calcium was compared to that of the enzyme with the site without calcium bound, showed no effect of calcium binding on the flexibility of that region of the protein. This was attributed to the presence of the disulfide bridge Cys67-Cys99 rigidifying that site and minimizing the effects of calcium binding there [32]. This could indicate that calcium binding to Ca-2 has minimal effects on conformational stability and could therefore be a site only affecting thermal inactivation. This loop has been shown to contain the initial autolytic site in the thermal inactivation of subtilisin BPN´ [33]. In fact, by eliminating this initial autolysis site and incorporating a calcium binding site corresponding to the Ca-2 site loop in thermitase, it was possible to improve the autolytic stability of the enzyme significantly in the presence of calcium. In the absence of calcium this mutant harboring the new calcium binding site was less stable to thermal inactivation [33]. By analogy, calcium binding to the Ca-2 site in VPR may increase the resistance of the enzyme to autolytic inactivation without contributing significantly to the global stability of the folded protein.

Compiling this information, we hypothesize that the Ca-2 site is the low affinity calcium binding site, with an apparent K_D of ~ 22 mM, and mainly affects the rate of thermal inactivation, leaving that part of the protein vulnerable to proteolytic degradation in the absence of bound calcium. From this reasoning the Ca-1 is the second weak binding site with an apparent K_D of ~ 1 mM, important for thermal stability of the secondary structure and the Ca-3 site, is the strongest binding site, formed during maturation of the subtilase and which would lock the folded proteinase, in its kinetically stable and active form (Fig. 8). These observations will pave the way for further investigations into thermostability and temperature adaptation of VPR and AQUI.

Acknowledgments

This work was supported by The Icelandic Research Fund [grant number 162977-051].

Abbreviations

PMSF	Phenylmethylsulfonyl fluoride
Suc-AAPF-NH-Np	Succinyl-AlaAlaProPhe-p-nitroanilide
DSC	Differential scanning calorimetry
VPR	A subtilisin-like serine proteinase from a psychrotrophic *Vibrio* species
SPRK	A subtilisin-like serine proteinase form a psychrotrophic *Serratia* species
AQUI	Aqualysin I; PRK, Proteinase K

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbapap.2018.11.010.

References

[1] M.M. Kristjansson, O.T. Magnusson, H.M. Gudmundsson, G.A. Alfredsson, H. Matsuzawa, Properties of a subtilisin-like proteinase from a psychrotrophic Vibrio species comparison with proteinase K and aqualysin I, Eur. J. Biochem. 260 (1999) 752–760.
[2] R.J. Siezen, J.A. Leunissen, Subtilases: the superfamily of subtilisin-like serine proteases, Protein Sci. 6 (1997) 501–523.
[3] J. Arnorsdottir, R.B. Smaradottir, O.T. Magnusson, S.H. Thorbjarnardottir, G. Eggertsson, M.M. Kristjansson, Characterization of a cloned subtilisin-like serine proteinase from a psychrotrophic Vibrio species, Eur. J. Biochem. 269 (2002) 5536–5546.
[4] A.G. Sigurdardottir, J. Arnorsdottir, S.H. Thorbjarnardottir, G. Eggertsson, K. Suhre, M.M. Kristjansson, Characteristics of mutants designed to incorporate a new ion pair into the structure of a cold adapted subtilisin-like serine proteinase, Biochim. Biophys. Acta 1794 (2009) 512–518.
[5] K.R. Oskarsson, M. Nygaard, B.O. Ellertsson, S.H. Thorbjarnardottir, E. Papaleo, M.M. Kristjansson, A single mutation Gln142Lys doubles the catalytic activity of VPR, a cold adapted subtilisin-like serine proteinase, Biochim. Biophys. Acta 1864 (2016) 1436–1443.
[6] A.R. Sigtryggsdottir, E. Papaleo, S.H. Thorbjarnardottir, M.M. Kristjansson, Flexibility of cold- and heat-adapted subtilisin-like serine proteinases evaluated with fluorescence quenching and molecular dynamics, Biochim. Biophys. Acta 1844 (2014) 705–712.
[7] M.M. Kristjansson, Thermostable subtilases (Subtilisin-Like Serine Proteinases), in: S. Sen, L. Nilsson (Eds.), Thermostable Proteins. Struct. Stab. Des, Taylor & Francis group, 2012, pp. 67–104.
[8] L.B. Jonsdottir, B.O. Ellertsson, G. Invernizzi, M. Magnusdottir, S.H. Thorbjarnardottir, E. Papaleo, M.M. Kristjansson, The role of salt bridges on the temperature adaptation of aqualysin I, a thermostable subtilisin-like proteinase, Biochim. Biophys. Acta 1844 (2014) 2174–2181.
[9] J. Arnorsdottir, M.M. Kristjansson, R. Ficner, Crystal structure of a subtilisin-like serine proteinase from a psychrotrophic Vibrio species reveals structural aspects of cold adaptation, FEBS J. 272 (2005) 832–845.
[10] S. Wagner, M.M. Klepsch, S. Schlegel, A. Appel, R. Draheim, M. Tarry, M. Hogbom, K.J. van Wijk, D.J. Slotboom, J.O. Persson, J.W. de Gier, Tuning Escherichia coli for membrane protein overexpression, Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 14371–14376.
[11] A.V. Bryksin, I. Matsumura, Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids, BioTechniques 48 (2010) 463–465.
[12] D.W. Cleveland, S.G. Fischer, M.W. Kirschner, U.K. Laemmli, Peptide mapping by limited proteolysis in sodium dodecyl sulfate and analysis by gel electrophoresis, J. Biol. Chem. 252 (1977) 1102–1106.
[13] G. Candiano, M. Bruschi, L. Musante, L. Santucci, G.M. Ghiggeri, B. Carnemolla, P. Orecchia, L. Zardi, P.G. Righetti, Blue silver: a very sensitive colloidal Coomassie G-250 staining for proteome analysis, Electrophoresis 25 (2004) 1327–1333.
[14] Z. Zaman, R.L. Verwilghen, Quantitation of Proteins Solubilized in Sodium Dodecyl Sulfate Mercaptoethanol Tris Electrophoresis Buffer, Anal. Biochem. 100 (1979) 64–69.
[15] C.N. Pace, F. Vajdos, L. Fee, G. Grimsley, T. Gray, How to measure and predict the molar absorption coefficient of a protein, Protein Sci. 4 (1995) 2411–2423.
[16] E.G. Delmar, C. Largman, J.W. Brodrick, M.C. Geokas, Sensitive new substrate for chymotrypsin, Anal. Biochem. 99 (1979) 316–320.
[17] A. Matouschek, J.M. Matthews, C.M. Johnson, A.R. Fersht, Extrapolation to water of kinetic and equilibrium data for the unfolding of barnase in urea solutions, Protein Eng. 7 (1994) 1089–1095.
[18] J.M. Sanchez-Ruiz, J.L. Lopez-Lacomba, M. Cortijo, P.L. Mateo, Differential scanning calorimetry of the irreversible thermal denaturation of thermolysin, Biochemistry 27 (1988) 1648–1652.
[19] K.S. Siddiqui, A. Poljak, D. De Francisci, G. Guerriero, O. Pilak, D. Burg, M.J. Raftery, D.M. Parkin, J. Trewhella, R. Cavicchioli, A chemically modified alpha-amylase with a molten-globule state has entropically driven enhanced thermal stability, Protein Eng. Des. Sel. 23 (2010) 769–780.
[20] C.M. Cuchillo, M.V. Nogues, R.T. Raines, Bovine Pancreatic Ribonuclease: Fifty Years of the first enzymatic reaction mechanism, Biochemistry 50 (2011) 7835–7841.
[21] M.B. Rao, A.M. Tanksale, M.S. Ghatge, V.V. Deshpande, Molecular and biotechnological aspects of microbial proteases, Microbiol, Mol. Biol. R. 62 (1998) 597.
[22] R. Helland, A.N. Larsen, A.O. Smalas, N.P. Willassen, The 1.8 angstrom crystal structure of a proteinase K-like enzyme from a psychrotroph Serratia species, FEBS J. 273 (2006) 61–71.
[23] S.I. Tanaka, K. Saito, H. Chon, H. Matsumura, Y. Koga, K. Takano, S. Kanaya, Crystal structure of unautoprocessed precursor of subtilisin from a

hyperthermophilic Archaeon - evidence for Ca2+ −induced folding, J. Biol. Chem. 282 (2007) 8246–8255.
[24] M.W. Pantoliano, M. Whitlow, J.F. Wood, M.L. Rollence, B.C. Finzel, G.L. Gilliland, T.L. Poulos, P.N. Bryan, The Engineering of binding-affinity at metal-ion binding-sites for the stabilization of proteins - Subtilisin as a test case, Biochemistry 27 (1988) 8311–8317.
[25] J. Bajorath, W. Hinrichs, W. Saenger, The enzymatic-activity of proteinase-K is controlled by calcium, FEBS J. 176 (1988) 441–447.
[26] J. Arnorsdottir, A.R. Sigtryggsdottir, S.H. Thorbjarnardottir, M.M. Kristjansson, Effect of proline substitutions on stability and kinetic properties of a cold adapted subtilase, J. Biochem. 145 (2009) 325–329.
[27] P.R. Green, J.D. Oliver, L.C. Strickland, D.R. Toerner, H. Matsuzawa, T. Ohta, Purification, crystallization and preliminary X-ray investigation of aqualysin I, a heat-stable serine protease, Acta Crystallogr. D. Biol. Crystallogr. 49 (1993) 349–352.
[28] S.J. Lin, E. Yoshimura, H. Sakai, T. Wakagi, H. Matsuzawa, 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 (1999) 132–138.
[29] C. Betzel, G.P. Pal, W. Saenger, Synchrotron X-ray data collection and restrained least-squares refinement of the crystal structure of proteinase K at 1.5 A resolution, Acta Crystallogr. B 44 (Pt 2) (1988) 163–172.
[30] C.A. Smith, H.S. Toogood, H.M. Baker, R.M. Daniel, E.N. Baker, Calcium-mediated thermostability in the subtilisin superfamily: the crystal structure of Bacillus Ak.1 protease at 1.8 angstrom resolution, J. Mol. Biol. 294 (1999) 1027–1040.
[31] A.V. Teplyakov, I.P. Kuranova, E.H. Harutyunyan, B.K. Vainshtein, C. Frommel, W.E. Hohne, K.S. Wilson, Crystal-structure of thermitase at 1.4 Å resolution, J. Mol. Biol. 214 (1990) 261–279.
[32] M. Tiberti, E. Papaleo, Dynamic properties of extremophilic subtilisin-like serine-proteases, J. Struct. Biol. 174 (2011) 69–83.
[33] S. Braxton, J.A. Wells, Incorporation of a stabilizing Ca(2+)-binding loop into subtilisin BPN, Biochemistry 31 (1992) 7796–7801.

Improved Expression, Purification and Characterization of VPR, a Cold Active Subtilisin-like Serine Proteinase and the Effects of Calcium on Expression and Stability.

Supplementary

Kristinn R. Óskarsson & Magnús M. Kristjánsson*

Department of Biochemistry, Science Institute, University of Iceland, Reykjavík, Iceland. Correspondence and requests for materials should be addressed to M.M.K. (email: mmk@hi.is)

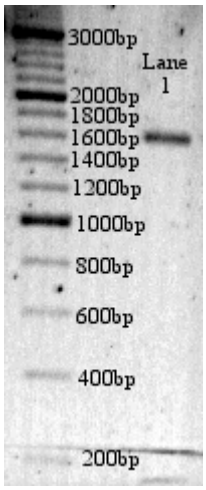
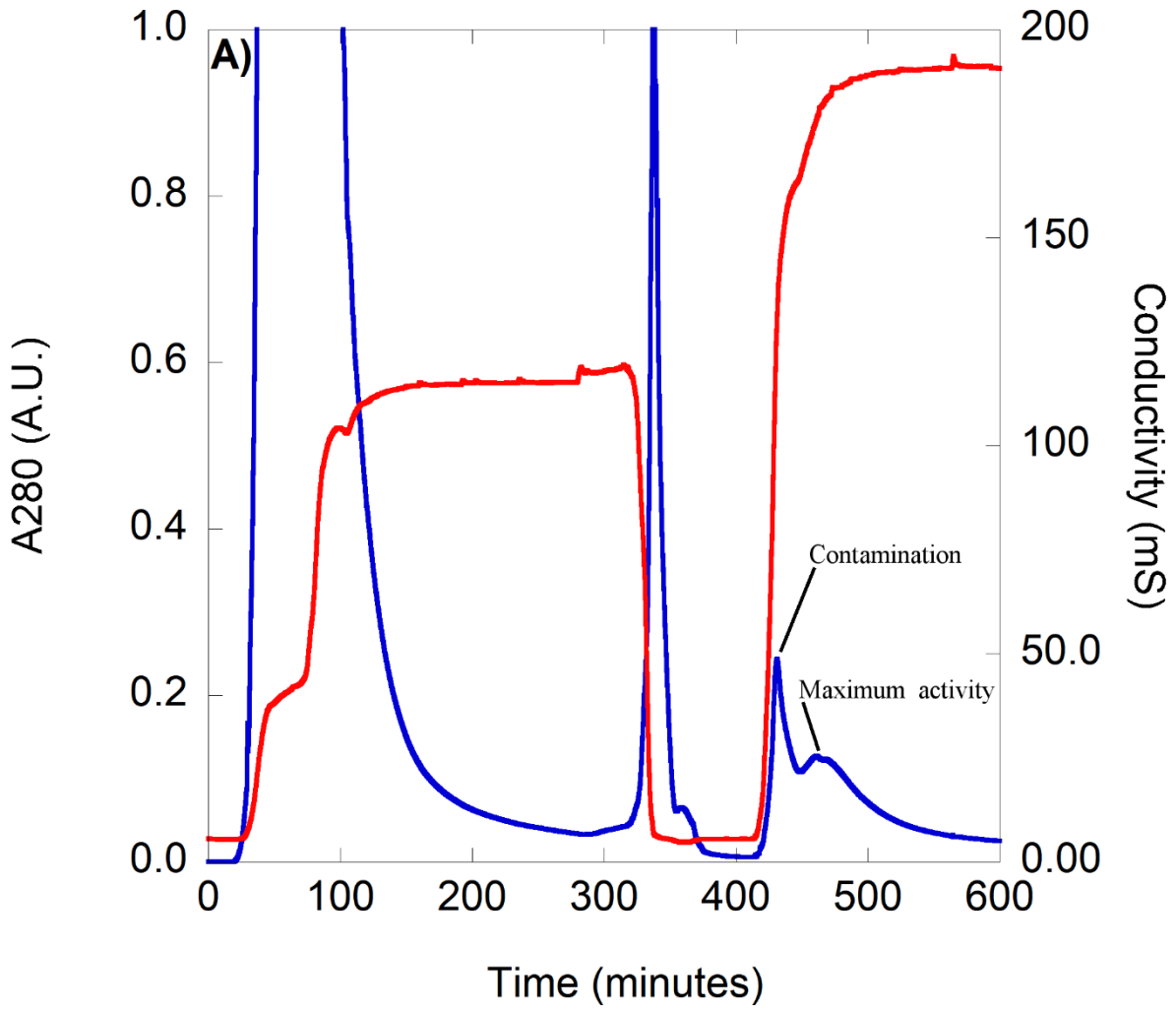
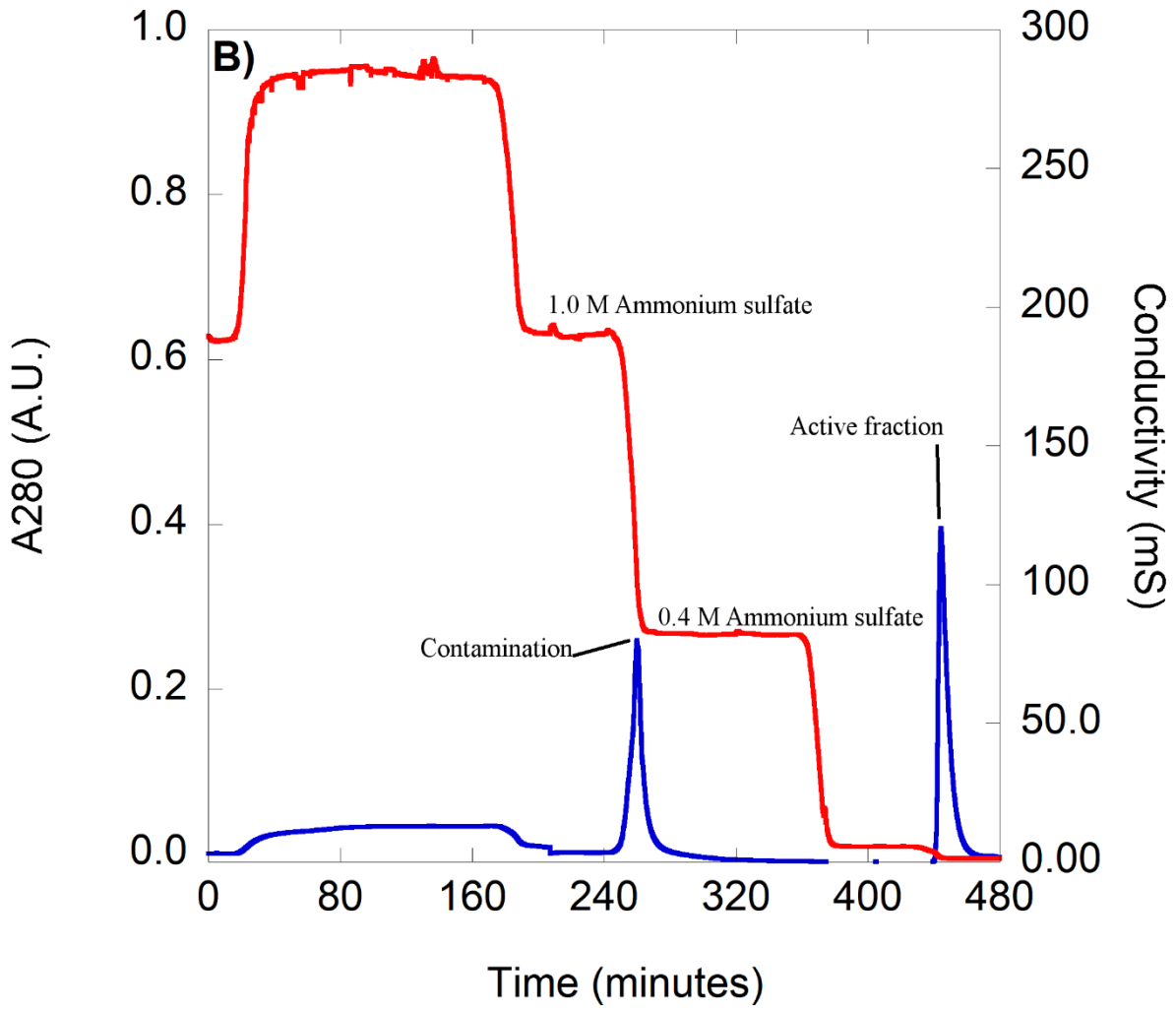


Figure 1. Supplementary. 1% agarose gel electrophoresis of the megaprimer used for gel extraction. Ladder used was O´RangeRuler 200bp DNA ladder from ThermoFisher Scientific. Lane 1 contains the megaprimer of the approximate size 1600bp.

Table 1. Supplementary. Overview of results from expression optimization experiments. All cultures were 20 mL and 10 mL were harvested for cell lysis. Values are based on averages from 1 to 7 measurements and standard deviation is shown were applicable.

0 mM Calcium																	
[Rhamnose]	0 µM			30 µM			76 µM			152 µM			304 µM			609 µM		
	U/L cells	**U/L Supernatant**	**A600**	***U/L cells***	***U/L Supernatant***	***A600***	**U/L cells**	**U/L Supernatant**	**A600**	***U/L cells***	***U/L Supernatant***	***A600***	**U/L cells**	**U/L Supernatant**	**A600**	***U/L cells***	***U/L Supernatant***	***A600***
Average	4	49	2.50	*1547*	*237*	*4.42*	5679	259	6.72	*5226*	*2074*	*7.58*	3948	53	7.76	*839*	*53*	*7.50*
S.D.	1	N.A.	N.A.	*15*	*N.A.*	*N.A.*	274	N.A.	N.A.	*69*	*N.A.*	*N.A.*	73	N.A.	N.A.	*126*	*N.A.*	*0.71*
10 mM Calcium																		
[Rhamnose]	0 µM			30 µM			76 µM			152 µM			304 µM			609 µM		
	U/L cells	**U/L Supernatant**	**A600**	***U/L cells***	***U/L Supernatant***	***A600***	**U/L cells**	**U/L Supernatant**	**A600**	***U/L cells***	***U/L Supernatant***	***A600***	**U/L cells**	**U/L Supernatant**	**A600**	***U/L cells***	***U/L Supernatant***	***A600***
Average	10352	13050	7.00	*N.A.*	*N.A.*	*N.A.*	29388	8935	8.59	*20976*	*5733*	*9.00*	13368	396	8.00	*5872*	*396*	*8.00*
S.D.	N.A.	N.A.	N.A.	*N.A.*	*N.A.*	*N.A.*	578	170	0.07	*N.A.*	*N.A.*	*N.A.*	N.A.	N.A.	N.A.	*N.A.*	*N.A.*	*N.A.*
25 mM Calcium																		
[Rhamnose]	0 µM			30 µM			76 µM			152 µM			304 µM			609 µM		
	U/L cells	**U/L Supernatant**	**A600**	***U/L cells***	***U/L Supernatant***	***A600***	**U/L cells**	**U/L Supernatant**	**A600**	***U/L cells***	***U/L Supernatant***	***A600***	**U/L cells**	**U/L Supernatant**	**A600**	***U/L cells***	***U/L Supernatant***	***A600***
Average	3610	7953	7.14	*26242*	*20717*	*7.76*	35428	8882	8.07	*38322*	*2422*	*8.49*	26862	340	8.64	*10417*	*340*	*8.25*
S.D.	47	N.A.	N.A.	*3686*	*3470*	*0.35*	5666	196	0.31	*1777*	*N.A.*	*N.A.*	3082	N.A.	N.A.	*3742*	*N.A.*	*0.36*
50 mM Calcium																		
[Rhamnose]	0 µM			30 µM			76 µM			152 µM			304 µM			609 µM		
	U/L cells	**U/L Supernatant**	**A600**	***U/L cells***	***U/L Supernatant***	***A600***	**U/L cells**	**U/L Supernatant**	**A600**	***U/L cells***	***U/L Supernatant***	***A600***	**U/L cells**	**U/L Supernatant**	**A600**	***U/L cells***	***U/L Supernatant***	***A600***
Average	2613	4482	3.12	*18539*	*17832*	*6.30*	31539	11081	6.50	*33163*	*3935*	*7.24*	17870	871	7.61	*9201*	*871*	*6.89*
S.D.	127	N.A.	N.A.	*8339*	*1288*	*0.99*	3246	2241	0.24	*4524*	*N.A.*	*N.A.*	1351	N.A.	N.A.	*3479*	*N.A.*	*0.16*
75 mM Calcium																		
[Rhamnose]	0 µM			30 µM			76 µM			152 µM			304 µM			609 µM		
	U/L cells	**U/L Supernatant**	**A600**	***U/L cells***	***U/L Supernatant***	***A600***	**U/L cells**	**U/L Supernatant**	**A600**	***U/L cells***	***U/L Supernatant***	***A600***	**U/L cells**	**U/L Supernatant**	**A600**	***U/L cells***	***U/L Supernatant***	***A600***
Average	3964	8032	4.10	*19112*	*23297*	*5.50*	34110	17892	6.79	*31566*	*3899*	*8.04*	29612	2119	8.78	*27435*	*2119*	*8.69*
S.D.	889	3223	0.26	*5825*	*4095*	*0.70*	3327	6315	0.61	*7615*	*3348*	*0.82*	4414	N.A.	N.A.	*2709*	*N.A.*	*N.A.*
100 mM Calcium																		
[Rhamnose]	0 µM			30 µM			76 µM			152 µM			304 µM			609 µM		
	U/L cells	**U/L Supernatant**	**A600**	***U/L cells***	***U/L Supernatant***	***A600***	**U/L cells**	**U/L Supernatant**	**A600**	***U/L cells***	***U/L Supernatant***	***A600***	**U/L cells**	**U/L Supernatant**	**A600**	***U/L cells***	***U/L Supernatant***	***A600***
Average	4919	7507	3.95	*28122*	*14679*	*6.09*	37697	15173	6.84	*29100*	*3322*	*8.59*	22026	1251	8.99	*14908*	*1251*	*9.49*
S.D.	388	N.A.	N.A.	*1247*	*5119*	*1.04*	4937	5593	0.36	*7632*	*2576*	*0.62*	4339	N.A.	N.A.	*4545*	*N.A.*	*N.A.*
125 mM Calcium																		
[Rhamnose]	0 µM			30 µM			76 µM			152 µM			304 µM			609 µM		
	U/L cells	**U/L Supernatant**	**A600**	***U/L cells***	***U/L Supernatant***	***A600***	**U/L cells**	**U/L Supernatant**	**A600**	***U/L cells***	***U/L Supernatant***	***A600***	**U/L cells**	**U/L Supernatant**	**A600**	***U/L cells***	***U/L Supernatant***	***A600***
Average	N.A.	N.A.	N.A.	*N.A.*	*N.A.*	*N.A.*	37754	17158	6.73	*16377*	*1688*	*7.78*	N.A.	N.A.	N.A.	*N.A.*	*N.A.*	*N.A.*
S.D.	N.A.	N.A.	N.A.	*N.A.*	*N.A.*	*N.A.*	368	188	0.04	*N.A.*	*N.A.*	*N.A.*	N.A.	N.A.	N.A.	*N.A.*	*N.A.*	*N.A.*


A)
1.0
0.8
0.6
0.4
0.2
0.0
A280 (A.U.)
200
150
100
50.0
0.00
Conductivity (mS)
Contamination
Maximum activity
0
100
200
300
400
500
600
Time (minutes)


B)
1.0 M Ammonium sulfate
0.4 M Ammonium sulfate
Contamination
Active fraction
A280 (A.U.)
Conductivity (mS)
Time (minutes)
1.0
0.8
0.6
0.4
0.2
0.0
300
250
200
150
100
50.0
0.00
0
80
160
240
320
400
480

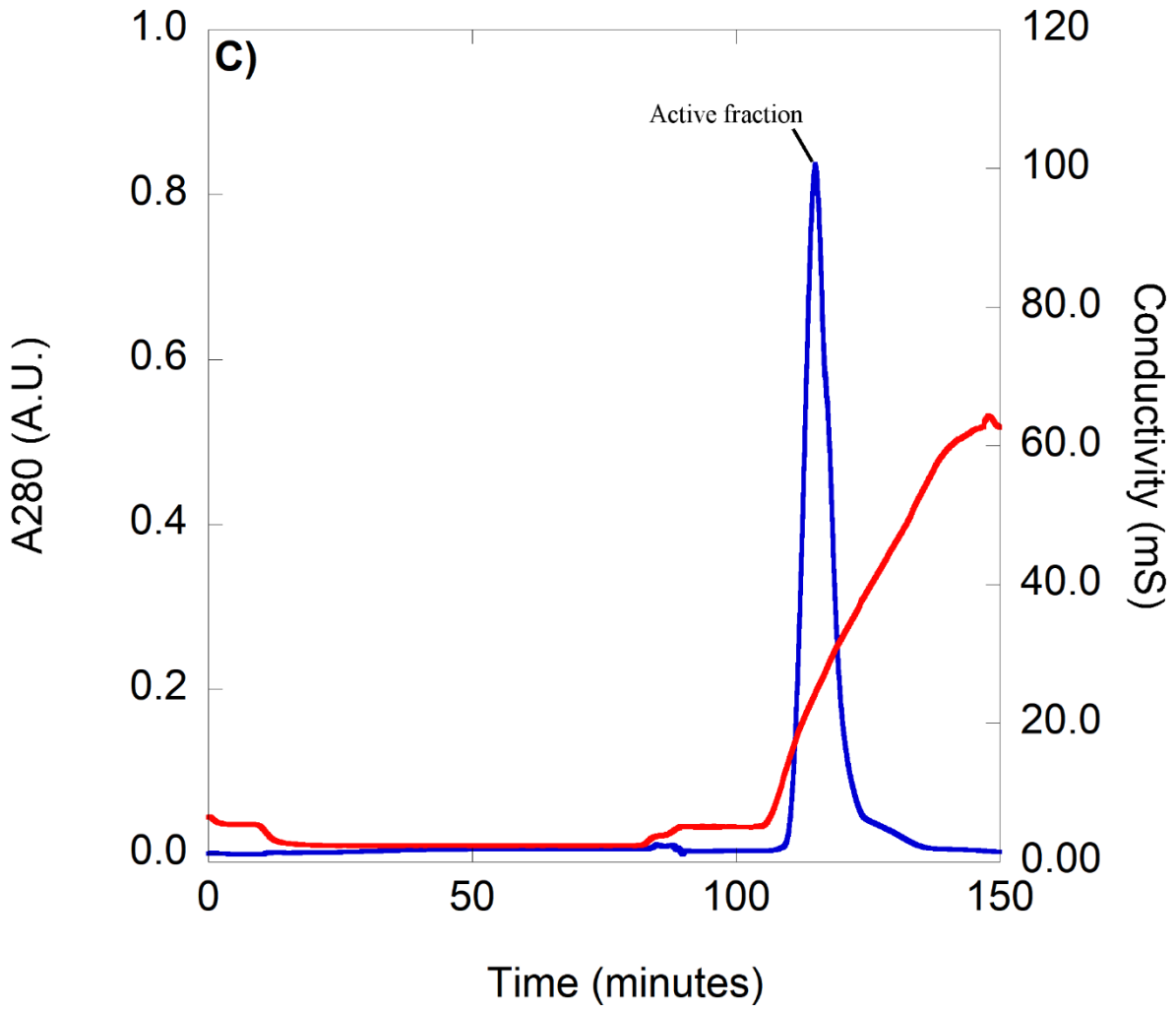


Figure 2. Supplementary. Purification steps of $VPR_{\Delta C}$. **A.** Affinity column, equilibrated with a Tris calcium buffer. Washed with Tris calcium buffer containing 1 M NaCl followed by NaCl free buffer and then eluted with 2 M guanidinium chloride in Tris calcium buffer. **B.** Phenyl-sepharose column, equilibrated with a Tris calcium buffer containing 1 M ammonium sulphate. Stepwise lowered to 0 M ammonium sulphate and eluted with 50% ethylene glycol. **C.** Q-sepharose, equilibrated with Tris calcium buffer and washed with the same buffer. Eluted by a NaCl gradient fram 0.0 M to 0.5 M NaCl. Blue lines represent absorbance at 280 nm and red lines represent conductivity in mS (Simens). Flow rate was set to 2 mL/min except for loading onto phenyl-sepharose (up to 8 mL/min) and elution from Q-sepharose (1 mL/min).

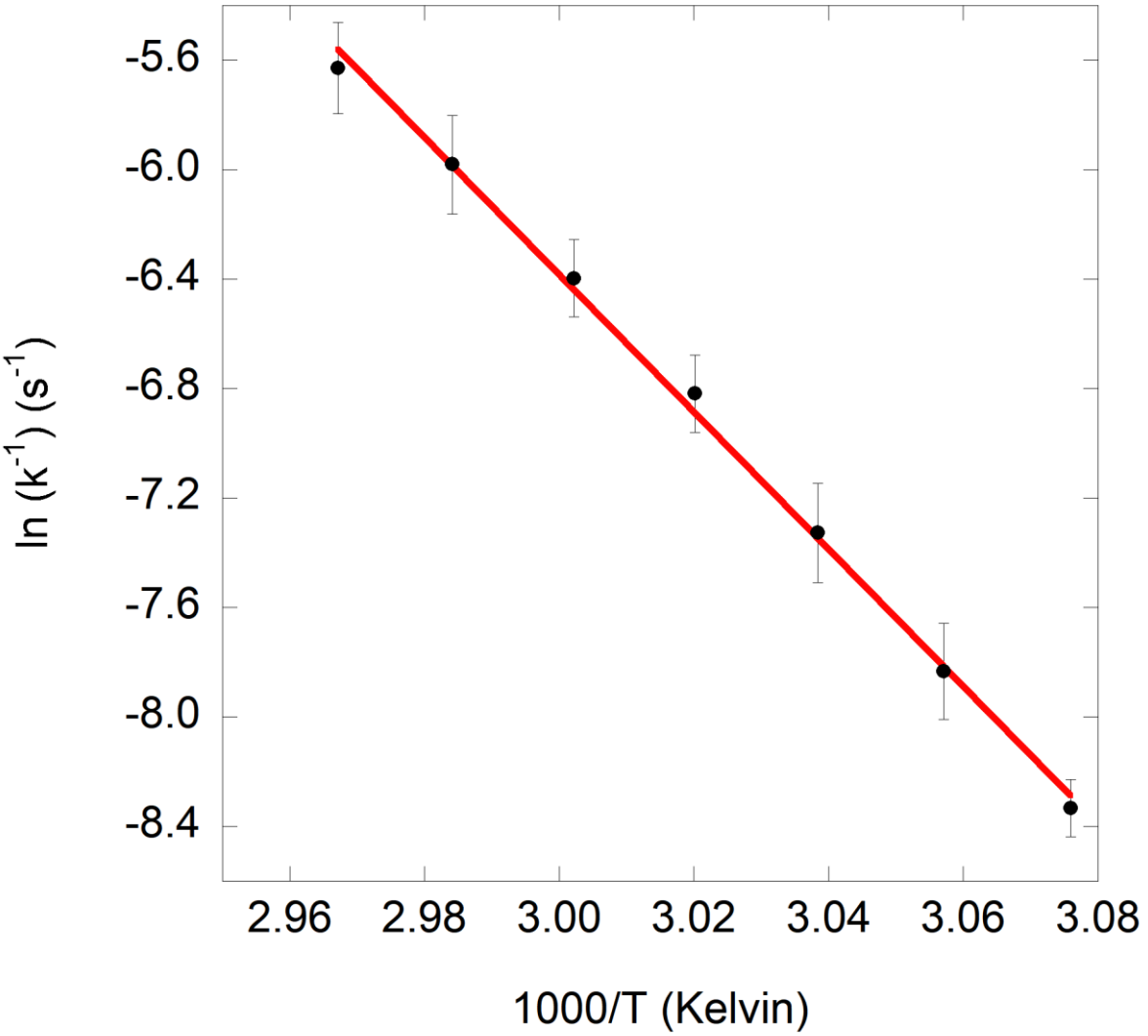


Figure 3. Supplementary. Arrhenius plot for rates of thermal inactivation of $VPR_{\Delta C}$. Black dots represent the average values from experimental data. The red line is the linear best fit of the data. Error bars represent one standard deviation of the mean.

Table 2. Supplementary. Half-lives of inactivation ($t_{1/2}$) and unfolding ($t_{1/2(DSC)}$) of $VPR_{\Delta C}$. Half-lives of inactivation were calculated using the best fit of half-lives of inactivation against absolute calcium concentration (accounting for EDTA concentration) (Figure 5. B).). Half-lives of unfolding were calculated from DSC at temperatures corresponding to calculated melting points from DSC data fitting.

$[Ca^{2+}]$ mM	$t_{1/2}$ (min^{-1}) @53.8°C	$t_{1/2(DSC)}$ (min^{-1}) @53.8°C	$t_{1/2(DSC)}$ @T_{m1} (min^{-1}) / T_{m1} (°C)	$t_{1/2(DSC)}$ @ T_{m2} (min^{-1}) / T_{m2} (°C)	$t_{1/2(DSC)}$ @ T_{m3} (min^{-1}) / T_{m3} (°C)	$t_{1/2(DSC)}$ @ T_{m4} (min^{-1}) / T_{m4} (°C)
1 mM	9 ± 1	10 ± 1	14 ± 3 / 52.5 ± 0.9	4.9 ± 0.6 / 56.8 ± 0.7	2.4 ± 0.2 / 60.1 ± 0.6	1.3 ± 0.1 / 62.9 ± 0.5
15 mM	29 ± 3	38 ± 6	13 ± 1 / 57.8 ± 0.6	4.6 ± 0.2 / 61.7 ± 0.4	2.1 ± 0.2 / 64.8 ± 0.3	1.1 ± 0.1 / 67.3 ± 0.2
100 mM	49 ± 3	65 ± 20	17 ± 6 / 58.1 ± 1.4	5.2 ± 1.0 / 62.3 ± 1.1	2.2 ± 0.2 / 65.3 ± 0.7	1.1 ± 0.1 / 67.7 ± 0.6

Paper II

SCIENTIFIC REPORTS
nature research

OPEN

Thermostabilization of VPR, a kinetically stable cold adapted subtilase, via multiple proline substitutions into surface loops

K. R. Óskarsson, A. F. Sævarsson & M. M. Kristjánsson*

Protein stability is a widely studied topic, there are still aspects however that need addressing. In this paper we examined the effects of multiple proline substitutions into loop regions of the kinetically stable proteinase K-like serine protease VPR, using the thermostable structural homologue AQUI as a template. Four locations for proline substitutions were chosen to imitate the structure of AQUI. Variants were produced and characterized using differential scanning calorimetry (DSC), circular dichroism (CD), steady state fluorescence, acrylamide fluorescence quenching and thermal inactivation experiments. The final product $VPR_{\Delta C}$_N3P/I5P/N238P/T265P was greatly stabilized which was achieved without any noticeable detrimental effects to the catalytic efficiency of the enzyme. This stabilization seems to be derived from the conformation restrictive properties of the proline residue in its ability to act as an anchor point and strengthen pre-existing interactions within the protein and allowing for these interactions to prevail when thermal energy is applied to the system. In addition, the results underline the importance of the synergy between distant local protein motions needed to result in stabilizing effects and thus giving an insight into the nature of the stability of VPR, its unfolding landscape and how proline residues can infer kinetic stability onto protein structures.

Stabilization of proteins against various environmental factors are of interest in many fields of industry and science, as application range and storage limit their usability in various processes. One of these factors, temperature, plays a pivotal role in this regard. In nature, temperature is one of the main evolutionary drivers of enzymes due to its direct effect on kinetic energies involved in biochemical reactions necessary to maintain life. With early life likely to have existed at high temperatures[1–3], many branches of life have had to adapt to colder environments over time and thus overcoming slower reaction rates, with no pressure on selecting for thermostable proteins. Many examples of highly active unstable enzymes and thermostable enzymes with low activities at ambient temperatures exist[4–7]. Observations like these prompted the activity/stability trade-off hypothesis. It states that to achieve high stability, molecular motions needed for rapid catalysis at lower temperatures are sacrificed. A recent study has however indicated that adenylate kinases from organisms that throughout their evolutionary history have evolved to cooler temperatures and again toward higher temperatures still retain relatively higher activities at lower temperatures, as in the case of *B. stearothermophilus*[8]. In contrast, thermostable enzymes that have never adapted to lower temperatures showed a much steeper dependence on temperature in order to maintain catalytic rates, as in the case of the enzyme from *C. subterraneus* and *A. aeolicus*[8]. This indicates that the activity/stability trade-off is more of an evolutionary artifact due to different evolutionary pressures, rather than an absolute relationship. In addition, lessons learnt from directed evolution on subtilisins also indicated that the activity/stability trade-off did not share a strict relationship, as higher stability without compromising activity at low temperatures could be achieved[9] and higher activity at lower temperatures was possible without the loss of stability[10]. Working on that premise, thermostabilization of enzymes from cold adapted organisms would be a feasible method in designing stable enzymes that are highly active at a broad range of temperatures. One of the ways to achieve that goal would be by engineering of cold adapted enzymes via site directed mutagenesis using their thermostable structural homologs as templates. Thus, the aim of this study was to enhance the stability of a kinetically stable, cold adapted subtilisin-like serine protease, VPR[5] and gain more insight into the molecular basis of kinetic stability of proteins. To this end a truncated version of VPR, $VPR_{\Delta C}$[11], was subjected to single point mutations

Department of Biochemistry, Science Institute, University of Iceland, Reykjavík, Iceland. *email: mmk@hi.is

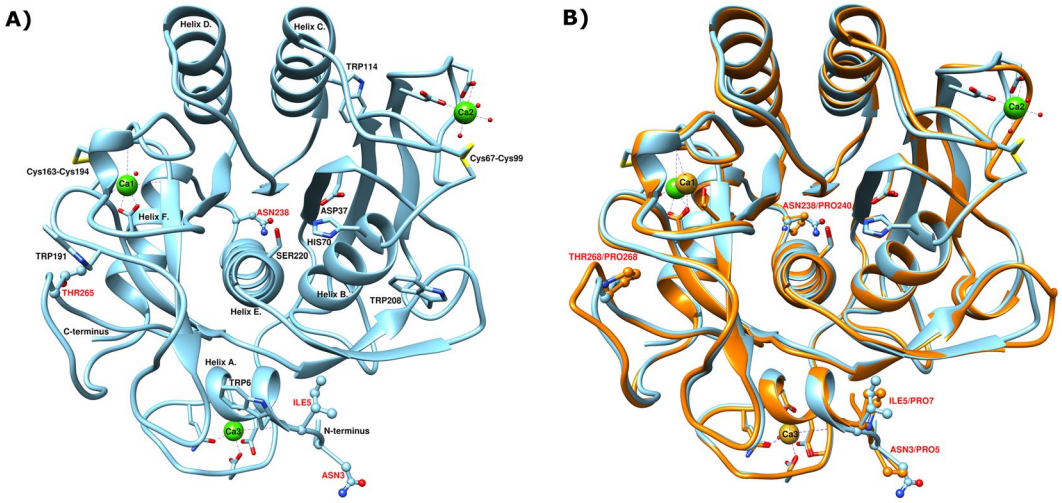


Figure 1. (**A**) The three-dimensional structure of VPR (PDB ID: 1SH7). Residues shown as sticks and balls and marked with red labels are the native residues mutated to prolines in this study. In addition, the catalytic triad Ser220, His70 and Asp37 are also shown along with all the Trp residues in the structure as well as calcium ion coordinators. Calcium ions are shown as green spheres. (**B**) Superimposed three-dimensional structures of VPR (light blue) (PDB ID: 1SH7) and AQUI (Orange) (PDB ID: 4DZT). The native prolines of AQUI are shown as sticks and balls along with the native VPR residues mutated to prolines. Calcium ions are shown as green spheres for VPR and golden spheres for AQUI. Atomic specifiers for side chains are as follows: carbon atoms are coloured same as the secondary structure; nitrogen atoms are coloured blue; oxygen atoms are coloured red and sulphur atoms coloured yellow.

incorporating the desired proline residues. The positions of proline residues were decided by using structural information from the thermostable structural homologue AQUI[5,12]. AQUI has four proline residues in loops not found in VPR, two of which are located near to the N-terminus, a third is located near a short loop between helices E and F in position 238 (VPR numbering) and the fourth is located on a loop following helix F (position 265) (Fig. 1)[13]. Although many aspects of thermostability have been identified[14] there is an observation of increased occurrence of proline residues in thermostable proteins[14–17]. This trend seems to be rather prevalent, genomic analysis of five cryophile genomes revealed a trend towards lower proline content in their proteomes[18]. Proline is unique among the natural amino acid residues in protein structures in containing a secondary amine group. This structural fact is the basis of the unique properties of the residue that restrict allowed conformations of the peptide backbone[19]. The effect of these restrictions is of interest with regards to kinetic stability. Kinetically stable proteins unfold irreversibly thus rely on high free energy barriers between the native and denatured states to maintain their activity[20]. Kinetically stable proteases also have evolved to have rigid native states that unfold in a highly cooperative manner[21,22]. Thus, the restriction of movements caused by proline substitutions could enforce pre-existing interaction within the protein structure[23,24]. In this study eight different proline variants were produced, purified and their properties measured. The variants produced were the single proline variants $VPR_{\Delta C}$_N3P, $VPR_{\Delta C}$_I5P, $VPR_{\Delta C}$_N238P, $VPR_{\Delta C}$_T265P, the double proline variant $VPR_{\Delta C}$_N3P/I5P, the triple proline variants $VPR_{\Delta C}$_N3P/I5P/N238P and $VPR_{\Delta C}$_N3P/I5P/T265P and lastly the quadruple proline variant $VPR_{\Delta C}$_N3P/I5P/N238P/T265P. The effects of these mutations on the properties of the enzyme were studied by circular dichroism (CD), differential scanning calorimetry (DSC), steady state fluorescence, acrylamide fluorescence quenching and Michaelis-Menten kinetics. The aim of enhancing the stability of $VPR_{\Delta C}$ was successful, as the final product, the quadruple proline variant was significantly stabilized without losing catalytic efficiency. In addition, the measured effects of proline exchange of the different variants did shed some light on the mode of action by which proline residues confer stability to the structure of VPR. The observed effects of prolines can be interpreted as restriction of movements leading to strengthening of pre-existing interactions by anchoring certain points within the structure that may lead to more allowed movements within the structure at higher temperatures without unfolding taking place. The effects of some proline substitutions showed clear signs of high local stabilization and as a result unfolding intermediates were observed as cooperativity of unfolding is lost to some degree. However, incorporating proline residues at distant parts of the protein displays synergic effects causing overall higher stability of the protein structure.

Results

All proline variants were successfully overexpressed in the *E. coli* strain Lemo21 and purified to homogeneity following the reformed production and purification protocol[25]. Thus, all single and the N3P/I5P proline variants[26] have been recharacterized with respect to activity, thermal inactivation ($T_{50\%}$) and the melting of the secondary structure ($T_{m\,(CD)}$).

Variant	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_M ($s^{-1}mM^{-1}$)
$VPR_{\Delta C}$	225.7 ± 12.0	0.177 ± 0.016	1238 ± 149
$VPR_{\Delta C}$/N3P	235.4 ± 21.8	0.173 ± 0.013	1364 ± 60
$VPR_{\Delta C}$/I5P	201.6 ± 8.2	0.187 ± 0.010	1077 ± 37
$VPR_{\Delta C}$/N238P	224.6 ± 16.6	0.189 ± 0.026	1196 ± 84
$VPR_{\Delta C}$/T265P	166.5 ± 11.6	0.152 ± 0.019	1101 ± 104
$VPR_{\Delta C}$/N3P/I5P	231.8 ± 10.5	0.187 ± 0.009	1243 ± 77
$VPR_{\Delta C}$/N3P/I5P/N238P	229.5 ± 18.3	0.199 ± 0.024	1158 ± 72
$VPR_{\Delta C}$/N3P/I5P/T265P	221.5 ± 7.8	0.219 ± 0.014	1017 ± 74
$VPR_{\Delta C}$/N3P/I5P/N238P/T265P	259.3 ± 27.4	0.212 ± 0.014	1222 ± 95

Table 1. Kinetic parameters of $VPR_{\Delta C}$ and proline variants. Values are expressed as the averages and the standard deviations of the mean.

Variant	Stern-Volmer (M^{-1})	λ_{max} (nm)	Relative intensity
$VPR_{\Delta C}$	2.24 ± 0.12	335 ± 1	1.00 ± 0.03
$VPR_{\Delta C}$/N3P	2.32 ± 0.18	337 ± 2	1.37 ± 0.02
$VPR_{\Delta C}$/I5P	2.05 ± 0.13	337 ± 2	1.16 ± 0.05
$VPR_{\Delta C}$/N238P	2.63 ± 0.25	339 ± 2	1.16 ± 0.07
$VPR_{\Delta C}$/T265P	2.41 ± 0.14	338 ± 1	1.15 ± 0.08
$VPR_{\Delta C}$/N3P/I5P	1.64 ± 0.07	334 ± 1	1.00 ± 0.03
$VPR_{\Delta C}$/N3P/I5P/N238P	2.18 ± 0.05	339 ± 1	1.18 ± 0.08
$VPR_{\Delta C}$/N3P/I5P/T265P	2.17 ± 0.05	338 ± 1	1.24 ± 0.03
$VPR_{\Delta C}$/N3P/I5P/N238P/T265P	2.12 ± 0.19	336 ± 2	1.19 ± 0.06

Table 2. Relative intensity of fluorescence calculated as area under the curve (AUC), the maxima of curves (λ_{max}) and the Stern-Volmer constant at 25 °C and pH 8.0 of $VPR_{\Delta C}$ and the proline variants. Values are expressed as the averages and the standard deviations of the mean.

Kinetics. Kinetic parameters determined by Michaelis-Menten assays at 25 °C and pH 8.6 indicated only small changes in turnover numbers and affinity for the substrate for the different variants (Table 1). The only exception was the $VPR_{\Delta C}$_T265P variant, where k_{cat} and K_m were consistently measured a little lower than for $VPR_{\Delta C}$ but resulted however in a similar value in terms of catalytic efficiency. The double and both triple proline variants showed little changes in turnover numbers, however this was accompanied by a trend towards slightly higher K_m values. The final product of this study, the quadruple proline variant, also still retained its catalytic efficiency as compared to $VPR_{\Delta C}$ but had a slightly higher turnover number and K_m value.

Fluorescence steady state emission and acrylamide quenching. VPR contains four Trp residues (Trp6, Trp114, Trp191 and Trp208) (Fig. 1). In the native state of VPR these Trp residues are highly intrinsically quenched. This is well demonstrated by examination of the native and denatured steady state fluorescence spectra of $VPR_{\Delta C}$ (Supplementary Fig. 1 and Supplementary Table 1) where the denatured state was found to be ten times as fluorescent as the native state at 25 °C. This makes the protein sensitive to changes of the microenvironments around these Trp residues. All single proline variants showed a trend of higher λ_{max} values (Table 2) indicating higher polarity around one or more of the Trp residues. These changes were notably higher for the $VPR_{\Delta C}$_N238P and $VPR_{\Delta C}$_T265P variants and these variants were also more quenchable by acrylamide than the wild type and had around 16% higher fluorescence, strongly suggesting changes in the environment of at least one Trp residue. The $VPR_{\Delta C}$_N3P variant had 37% higher fluorescence than the wild type, but had an unchanged Stern-Volmer constant, but with a minor increase in λ_{max}. $VPR_{\Delta C}$_I5P had a lower Stern-Volmer constant with 16% higher fluorescence and a minor increase in λ_{max}. Possibly this indicates some changes in the environment of Trp6 due to its proximity to these mutation sites. The fluorescence properties of the double proline variant $VPR_{\Delta C}$_N3P/I5P were different from the single variants. $VPR_{\Delta C}$_N3P/I5P had the same relative amplitude as $VPR_{\Delta C}$ but with a slight blue shift in the spectrum indicating a more buried Trp residue. This was further supported by the quenching data as $VPR_{\Delta C}$_N3P/I5P had a considerably lower Stern-Volmer constant than $VPR_{\Delta C}$, indicating reduced flexibility of the N-terminal and/or different dynamics of the N-terminus. These effects of the N3P/I5P mutation seem to be undone by the addition of N238P and T265P, as both relative intensity and the Stern-Volmer constants were higher for the triple proline variants. Those observations however, are likely to be due to changes in accessibility of Trp residues other than Trp6 as these effects are similar as seen by these mutations on the wild type. The quadruple proline variant exhibited similar properties as the triple variants but with a small blue shift in its spectrum (Figs. 2 and 3). The effects of temperature on fluorescence properties were also investigated. Temperatures measured were 15 °C, 25 °C and 35 °C. All variants were stable under those conditions during measurements as seen in the λ_{max} values and the gradual lowering in relative fluorescence intensities (Supplementary Tables 1 and

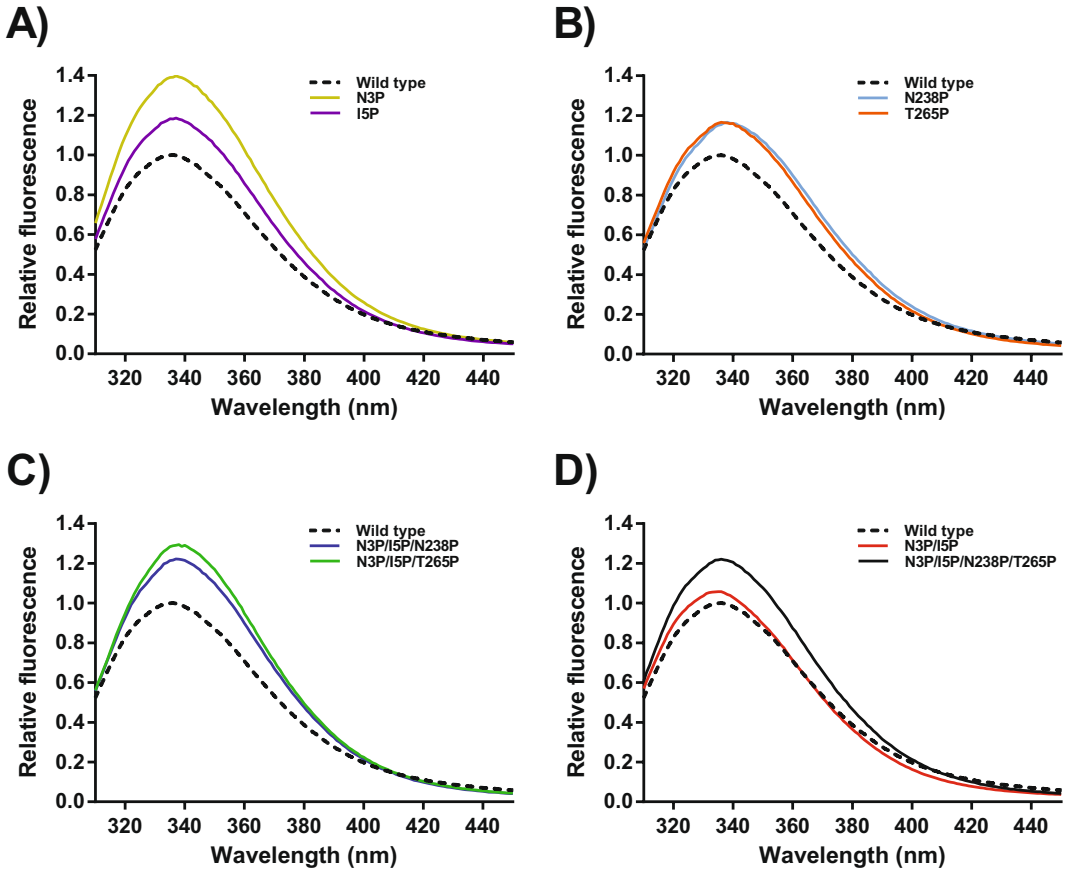


Figure 2. Fluorescence emission of proline variants after excitation at 295 nm at pH 8.0. Intensities of emissions have been normalized against $VPR_{\Delta C}$ (dotted black line). (**A**) Emission of $VPR_{\Delta C}$_N3P (Gold) and $VPR_{\Delta C}$_I5P (purple). (**B**) Emission of $VPR_{\Delta C}$_N238P (light blue) and $VPR_{\Delta C}$_T265P (orange). (**C**) Emission of $VPR_{\Delta C}$_N3P/I5P/N238P (blue) and $VPR_{\Delta C}$_N3P/I5P/T265P (green). (**D**) Emission of $VPR_{\Delta C}$_N3P/I5P (red) and $VPR_{\Delta C}$_N3P/I5P/N238P/T265P (black).

2). At these temperatures accessibility to fluorophores did not seem so be affected to any extent as Stern-Volmer constants showed just a marginal trend of higher Stern-Volmer constants, often under one standard deviation (Supplementary Table 3). This is consistent with the notion that kinetically stable proteinases are highly rigid structures to reduce auto-proteolysis in their native state[22,27]. However, there is some information to be obtained from these results recorded between 15 °C and 35 °C. Cooperative effects caused by the combination of N238P and T265P on top of the $VPR_{\Delta C}$_N3P/I5P variant seem to cause the native structure of the final product to become less responsive to acrylamide quenching as a function of temperature, which might be indicative of a more rigid and temperature tolerant structure because of changed dynamics within the protein due to synergy between N238P and T265P.

Stability. The impact of proline exchange on the stability of $VPR_{\Delta C}$ showed strong evidence for local stabilizing effects. This is best seen in the DSC thermograms with the emergence of an unfolding intermediate. None of the proline variants produced seem to have any notable effects on the secondary structure of the native state as seen in their CD wavelength spectra, the only notable changes being ascribed to concentration estimations i.e. the calculated depth of the spectra (Supplementary Fig. 2). However, structural changes were observed in the microenvironments of Trp fluorophores (Table 2) resulting from these mutations as shown in the steady-state spectra and acrylamide quenching data sets of proline variants (Figs. 2 and 3).

The N-terminal variants $VPR_{\Delta C}$_N3P and $VPR_{\Delta C}$_I5P. Of the single proline variants, the N-terminal mutations had the most impact on the measured melting points as determined by CD and the rates of thermal inactivation (Table 3). $VPR_{\Delta C}$_N3P and $VPR_{\Delta C}$_I5P had higher melting points of the secondary structure ($T_{m\ (CD)}$) by 2.9 °C and 3.2 °C, respectively (Fig. 4) and $T_{50\%}$ by 2.8 °C and 2.3 °C (Fig. 5). Accompanied with this increased stability was a considerable change observed in the recorded DSC thermograms of these variants. As for the truncated wild type, the curvature of the thermogram resembles a classic thermogram of a kinetically stable protein that unfolds in a rather cooperative manner following a two-state unfolding model (Fig. 6 and Supplementary Fig. 4)[28]. However, the thermograms of $VPR_{\Delta C}$_N3P and $VPR_{\Delta C}$_I5P exhibit a second transition peak, present with a maximum at considerably higher temperature than $T_{m\ (CD)}$, or around 71.5 °C for

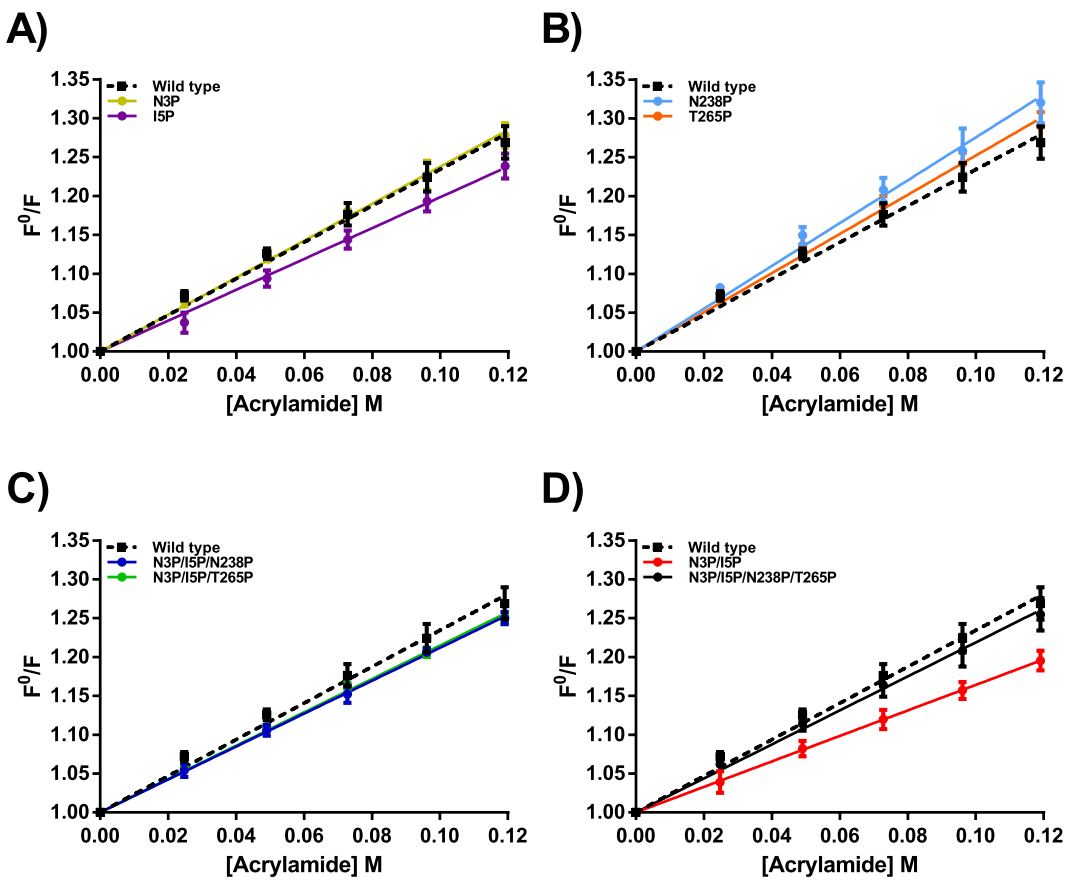


Figure 3. Stern-Volmer graphs calculated form fluorescence quenching of proline variants between 310–410 nm at pH 8.0. $VPR_{\Delta C}$ (black boxes with a dotted black line). (**A**) Quenching of $VPR_{\Delta C}$_N3P (Gold) and $VPR_{\Delta C}$_I5P (purple). (**B**) Quenching of $VPR_{\Delta C}$_N238P (light blue) and $VPR_{\Delta C}$_T265P (orange). (**C**) Quenching of $VPR_{\Delta C}$_N3P/I5P/N238P (blue) and $VPR_{\Delta C}$_N3P/I5P/T265P (green). (**D**) Quenching of $VPR_{\Delta C}$_N3P/I5P (red) and $VPR_{\Delta C}$_N3P/I5P/N238P/T265P (black).

Variant	$T_{m\,(CD)}$ (°C)	$T_{m\,(DSC)}$ (°C)	ΔH_{cal} (kJ/mol)	$T_{50\%}$ (°C)	$E_{act\,(inactivation)}$ (kJ/mol)	$t_{1/2\,(60\,°C)}$ (min)
$VPR_{\Delta C}$	61.9 ± 0.4	63.9 ± 0.3	528 ± 35	53.8 ± 0.4	218 ± 9	7 ± 1
$VPR_{\Delta C}$_N3P	64.8 ± 0.1	66.8 ± 0.3	533 ± 33	56.6 ± 0.3	203 ± 12	14 ± 1
$VPR_{\Delta C}$_I5P	65.1 ± 0.2	65.7 ± 0.5	570 ± 11	56.1 ± 0.2	199 ± 14	13 ± 1
$VPR_{\Delta C}$_N238P	60.7 ± 0.1	63.6 ± 0.2	556 ± 8	52.3 ± 0.2	209 ± 17	5 ± 1
$VPR_{\Delta C}$_T265P	61.6 ± 0.2	64.5 ± 0.2	451 ± 51	54.3 ± 0.2	206 ± 4	8 ± 1
$VPR_{\Delta C}$_N3P/I5P	67.8 ± 0.3	72.0 ± 0.8	646 ± 38	60.3 ± 0.4	208 ± 8	33 ± 3
$VPR_{\Delta C}$_N3P/I5P/N238P	68.8 ± 0.2	72.0 ± 0.4	679 ± 37	60.9 ± 0.5	207 ± 27	38 ± 5
$VPR_{\Delta C}$_N3P/I5P/T265P	69.2 ± 0.2	73.6 ± 0.3	697 ± 49	62.2 ± 0.6	194 ± 27	48 ± 7
$VPR_{\Delta C}$_N3P/I5P/N238P/T265P	72.1 ± 0.3	77.2 ± 0.2	683 ± 16	61.6 ± 0.6	182 ± 29	39 ± 3

Table 3. Thermostability parameters of $VPR_{\Delta C}$ and its proline variants. Parameters shown are $T_{m\,(CD)}$ the melting point of PMSF inhibited enzymes as measured by CD, $T_{m\,(DSC)}$ the apparent melting point of PMSF inhibited enzymes defined as the highest peak of DSC thermograms, ΔH_{cal} the excess calorimetric heat released during unfolding, $T_{50\%}$ the temperature where half of the activity has been lost over 30 min, $E_{act\,(inactivation)}$ calculated form the slope of Arrhenius graphs used to calculate $T_{50\%}$ and $t_{1/2}$ at 60 °C calculated form the same Arrhenius graphs. Values are expressed as the averages and the standard deviations of the mean.

$VPR_{\Delta C}$_N3P (Fig. 6 and Supplementary Fig. 5) and 74.1 °C for $VPR_{\Delta C}$_I5P (Fig. 6 and Supplementary Fig. 6). This shows that the major part of the three-dimensional structure had dissipated at or around 70– 80% according to CalFitter global fitting of CD melting curves that were recorded at 222 nm in tandem with DSC thermograms (Supplementary Figs. 4 and 5). This might indicate that the local stability of the N-terminus had been increased to such an extent that the cooperativity of the unfolding process was disrupted, thus leading to this apparent intermediate. In the case of $VPR_{\Delta C}$_N3P the activation energy (E_{act}) of unfolding for the first transition was

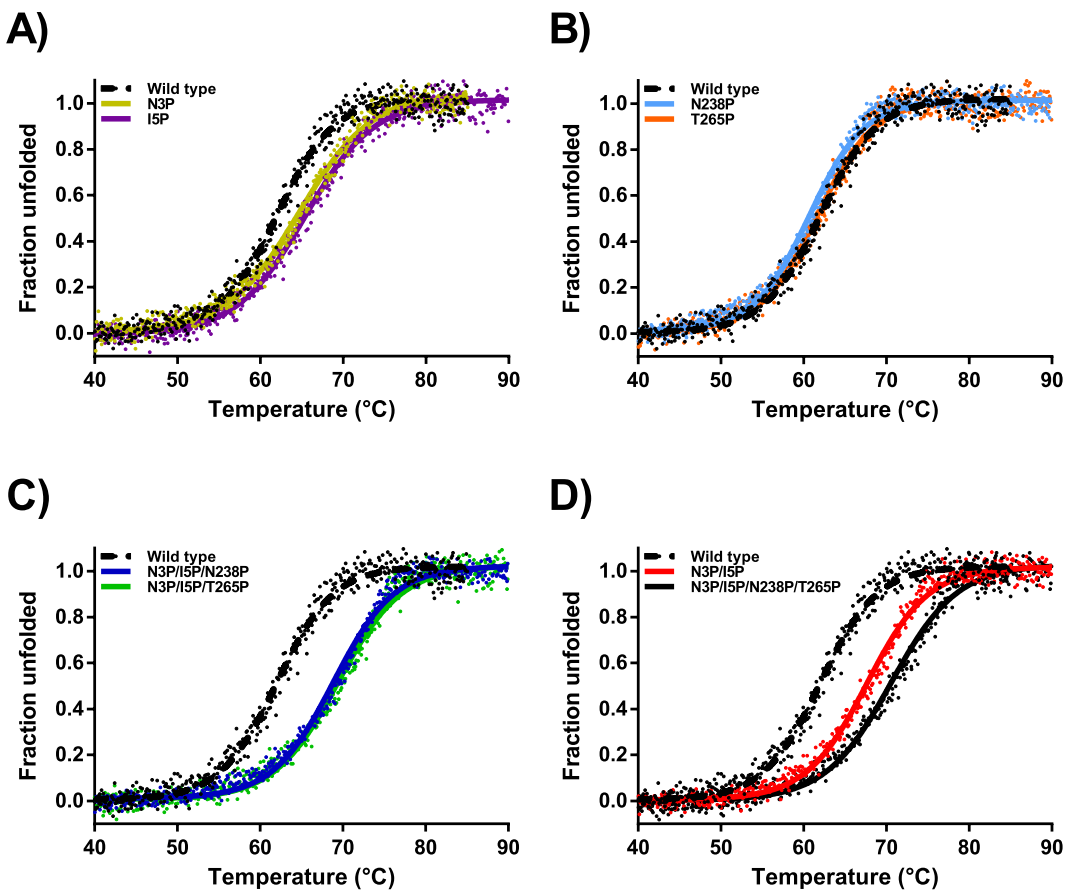


Figure 4. Normalized melting curves of proline variants in glycine buffer containing 15 mM $CaCl_2$ and 100 mM NaCl. $VPR_{\Delta C}$ (black dotted line). (**A**) Melting of $VPR_{\Delta C}$_N3P (Gold) and $VPR_{\Delta C}$_I5P (purple). (**B**) Melting of $VPR_{\Delta C}$_N238P (light blue) and $VPR_{\Delta C}$_T265P (orange). (**C**) Melting of $VPR_{\Delta C}$_N3P/I5P/N238P (blue) and $VPR_{\Delta C}$_N3P/I5P/T265P (green). (**D**) Melting of $VPR_{\Delta C}$_N3P/I5P (red) and $VPR_{\Delta C}$_N3P/I5P/N238P/T265P (black).

increased when compared to the wild type (Table 4). However, when comparing $VPR_{\Delta C}$_I5P, where the loss of cooperativity in the unfolding process was even more pronounced, to that of the wild type, the E_{act} values for the first transition did not change but was followed by the second transition having a much higher E_{act}. This infers that the first transition state is not as entropically favoured and that a good part of the enthalpic interactions are concentrated within the regions of the protein stabilized by the N-terminus and within the N-terminus itself, as E_{act} reflects on the activation enthalpy of unfolding. The N-terminus, where both these mutations are located also harbours the calcium 3 binding site, which has been suggested to be highly important for the overall stability of the enzyme (Fig. 7)[25]. Combined with the high activation energy of the second transition, this could indicate that the intermediate still retains the calcium-binding site 3. Even though $VPR_{\Delta C}$ unfolds in a rather cooperative manner, an elucidation of the chronological order of events during the thermal unfolding of wild type VPR using MD has been reported[29]. There, helix D close to the Ca1 binding site appeared to be the initiation point of thermal unfolding. However, partial unfolding of helix A started soon thereafter and completely dissipated shortly after, but the Ca3 loop seemed to retain itself much longer. In the light of those restricting movements around the N-terminal site of helix A (Fig. 7) as a result of proline substitutions likely slows down the unfolding of that region substantially, promoting the appearance of an unfolding intermediate. According to these unfolding simulations helix E is one of the most stable parts of the protein and accounts for approximately 20% of the helical content of VPR. Thus, speculations that the unfolding intermediate might be consisting of the N-terminal calcium binding loop and helix E fits nicely.

The $VPR_{\Delta C}$_N238P and $VPR_{\Delta C}$_T265P variants. The $VPR_{\Delta C}$_N238P variant was the only mutation that caused destabilizing effects, having a ~1.2 °C lower $T_{m\,(CD)}$ and 1.5 °C lower $T_{50\%}$. According to the Arrhenius graphs calculated from the DSC thermograms, $VPR_{\Delta C}$_N238P (Supplementary Figs. 3 and 7) unfolded cooperatively like the wild type. Even though lowered stability was observed using CD and inactivation experiments, the same observations were not as evident in the DSC thermograms. The apparent melting point ($T_{m\,(DSC)}$) corresponded very well with that of the wild type, but the variant exhibited higher E_{act} and calorimetric heat of unfolding (ΔH_{cal}) (Tables 3 and 4). This suggests that the mutation is indeed enforcing some interactions within the structure, but possibly not as entropically favourable as in the wild type. The $VPR_{\Delta C}$_T265P variant did cause the smallest changes in the observed stability having $T_{m\,(DSC)}$, $T_{m\,(CD)}$ and $T_{50\%}$ values around a half a degree

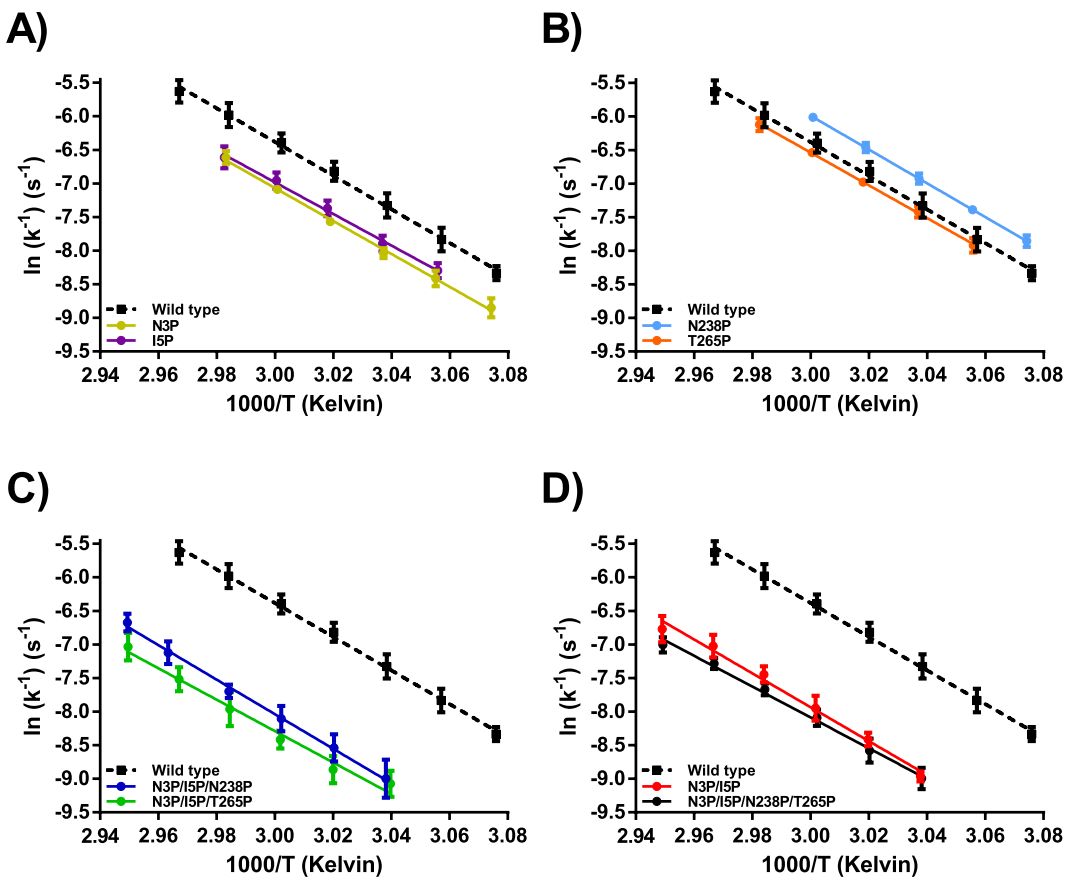


Figure 5. Arrhenius plots calculated form the thermal inactivation of proline variants in Tris buffer containing 15 mM $CaCl_2$. $VPR_{\Delta C}$ (black boxes with a dotted black line). (**A**) Thermal inactivation of $VPR_{\Delta C}$_N3P (Gold) and $VPR_{\Delta C}$_I5P (purple). (**B**) Thermal inactivation of $VPR_{\Delta C}$_N238P (light blue) and $VPR_{\Delta C}$_T265P (orange). (**C**) Thermal inactivation of $VPR_{\Delta C}$_N3P/I5P/N238P (blue) and $VPR_{\Delta C}$_N3P/I5P/T265P (green). (**D**) Thermal inactivation of $VPR_{\Delta C}$_N3P/I5P (red) and $VPR_{\Delta C}$_N3P/I5P/N238P/T265P (black).

higher than that of the wild type. The variant however, had one of the least stable pre and post heat capacities recorded on DSC before and after the unfolding transition, complicating the baseline estimation and data analysis. However, the Arrhenius graph constructed from DSC data (Supplementary Figs. 3 and 8) did indicate some loss of cooperativity during unfolding, but the data could not be reliably fitted to a more complex model than a simple two-state irreversible model, indicating that divergence from linearity observed may have arisen from poor baseline generation.

The $VPR_{\Delta C}$_N3P/I5P variant. The double N-terminal proline variant $VPR_{\Delta C}$_N3P/I5P showed clear additive properties at first glance as it had an increased melting point ($T_{m\ (CD)}$) by 5.9 °C and $T_{50\%}$ by 5.1 °C. However, the DSC thermogram did not exhibit the distinct second peak as was observed for the single proline N-terminal variants. The apparent melting point ($T_{m\ (DSC)}$) was around 4.2 °C higher than the $T_{m\ (CD)}$, which is 2.0 °C higher than the difference obtained between $T_{m\ (DSC)}$ and $T_{m\ (CD)}$ for the wild type. Also, the $T_{m\ (DSC)}$ of $VPR_{\Delta C}$_N3P/I5P more closely coincided with the estimated second peaks from the single N-terminal variants. Moreover, the Arrhenius graph calculated from thermograms showed a slight divergence from linearity and was also most reliably fitted to a three-state model (Supplementary Figs. 3 and 9). This indicates that the unfolding is much more cooperative when the N-terminal proline substitutions are combined in the double variant. This combination of the two most stabilizing proline mutations therefore causes extra stabilization throughout the protein structure, including parts of the protein that did not directly benefit from the single proline mutations. This implies that the single N-terminal proline mutations better stabilize the N-terminus, but synergic effects caused by the combination of N3P and I5P lead to global stabilization of the structure. This effect may be partly explained by a two amino acid residue shift in the auto-cleavage site of the N-terminal during maturation of the protease as was reported on previously[26]. The two-residue extension at the N-terminus potentially adds new hydrogen bonds between the N-terminus and the loop following helix A and anchors these interactions via the movement restricting proline residues. In the crystal structure, residues Tyr22 to Phe26 in that loop can form hydrogen bonds via the facilitation of water molecules to residues Ser271 to Ala273 on the C-terminus of the protein, with mainchain-mainchain distances between residues Asn25 and Leu272 being as low as 7.4 Å (Fig. 7). In addition, the hydrophobic interface of helix A on the N-terminus takes part in the formation of a hydrophobic core along with residues on helix E, helix F, residues Leu268 and Leu269 of the C-terminal region that also form a part of this

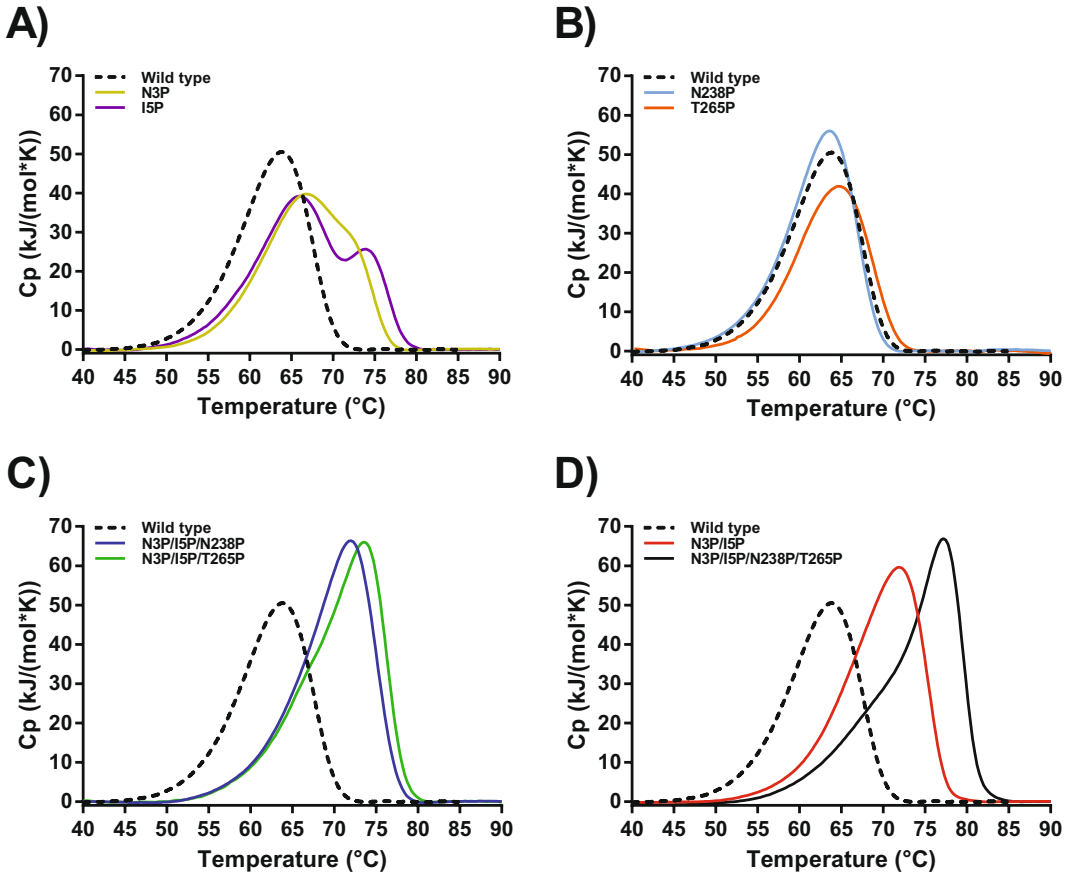


Figure 6. Deconvoluted differential scanning thermograms showing the excess heat during the unfolding process of the proline variants in a glycine buffer containing 15 mM $CaCl_2$ and100 mM NaCl. $VPR_{\Delta C}$ (dotted black line). (**A**) Unfolding of $VPR_{\Delta C}$_N3P (Gold) and $VPR_{\Delta C}$_I5P (purple). (**B**) Unfolding of $VPR_{\Delta C}$_N238P (light blue) and $VPR_{\Delta C}$_T265P (orange). (**C**) Unfolding of $VPR_{\Delta C}$_N3P/I5P/N238P (blue) and $VPR_{\Delta C}$_N3P/I5P/T265P (green). (**D**) Unfolding of $VPR_{\Delta C}$_N3P/I5P (red) and $VPR_{\Delta C}$_N3P/I5P/N238P/T265P (black).

Variant	E_{act}^{1} (kJ/mol)	E_{act}^{2} (kJ/mol)	$\Delta H_{cal\text{-}fit}^{1}$ (kJ/mol)	$\Delta H_{cal\text{-}fit}^{2}$ (kJ/mol)
$VPR_{\Delta C}$	235 ± 2	N.A.	542 ± 5	N.A.
$VPR_{\Delta C}$_N3P	251 ± 5	285 ± 20	318 ± 20	227 ± 19
$VPR_{\Delta C}$_I5P	235 ± 3	356 ± 13	399 ± 7	180 ± 6
$VPR_{\Delta C}$_N238P	248 ± 2	N.A.	564 ± 4	N.A.
$VPR_{\Delta C}$_T265P	229 ± 2	N.A.	471 ± 5	N.A.
$VPR_{\Delta C}$_N3P/I5P	261 ± 9	283 ± 6	176 ± 20	477 ± 19
$VPR_{\Delta C}$_N3P/I5P/N238P	270 ± 11	279 ± 3	120 ± 13	567 ± 13
$VPR_{\Delta C}$_N3P/I5P/T265P	259 ± 10	326 ± 8	224 ± 20	481 ± 18
$VPR_{\Delta C}$_N3P/I5P/N238P/T265P	215 ± 8	383 ± 7	275 ± 17	425 ± 15

Table 4. Parameters resulting from fitting deconvoluted DSC thermograms (Fig. 6), averaged from at least three separate runs using CalFitter 1.2. Values shown are the activation energy (E_{act}) of unfolding transitions of PMSF inhibited VPR variants and $\Delta H_{cal\text{-}fit}$ the calorimetric enthalpy of the fits. Numbers in superscript refer to the chronological order of transitions from the native to unfolded state. All values are represented with their 95% confidence interval.

hydrophobic core. Thus, reduced movements of the N-terminal region which are observed in the fluorescence quenching experiments of this variant (Table 2), may facilitate the increased cooperativity of unfolding compared to the single N-terminal variants, by stabilizing interactions between these parts of the protein at higher temperatures. This is achieved by possible extra H-bonds or enforcement of pre-existing ones reflected in the higher

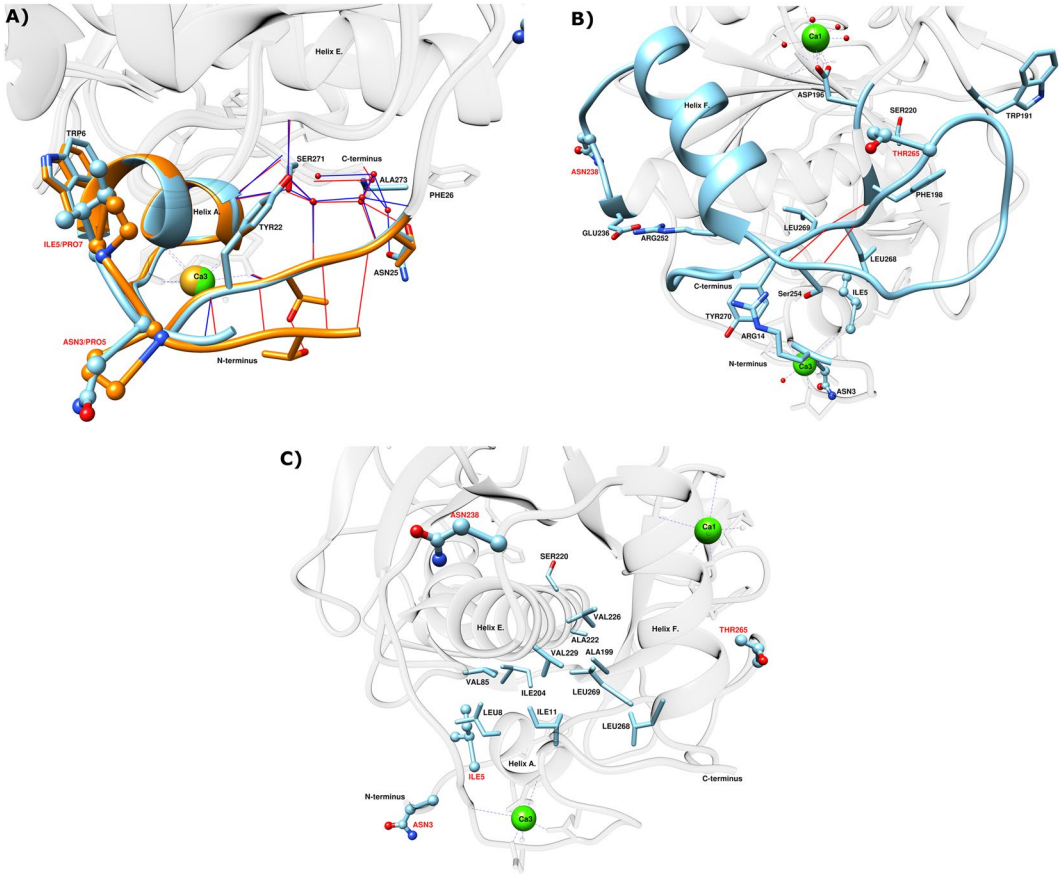


Figure 7. (**A**) Superimposed closeup of the N-terminals of VPR (light blue) (PDB ID: 1SH7) and AQUI (orange) (PDB ID: 4DZT) comparing the H-bond potential of both enzymes (red lines H-bonds in AQUI and blue lines H-bonds in VPR). (**B**) Closeup of the part of the protein in the closest vicinity of the N238P and T265P mutations in the structure of VPR. (**C**) Closeup highlighting the residues making up a part of the hydrophobic core and N-terminal interface to the main part of the protein in the structure of VPR. Calcium ions are shown as green spheres for VPR and golden spheres for AQUI. Atomic specifiers for side chains are as follows: carbon atoms are coloured same as the secondary structure; nitrogen atoms are coloured blue; oxygen atoms are coloured red and sulphur atoms coloured yellow.

E_{act} values that would indicate that the stabilization is rather enthalpic in nature and explaining the higher ΔH_{cal} observed for $VPR_{\Delta C}$_N3P/I5P and variants containing that mutational combination.

The $VPR_{\Delta C}$_N3P/I5P/N238P variant. Due to the considerable stabilization of $VPR_{\Delta C}$_N3P/I5P it was of interest to see how the addition of the mutations N238P and T265P would impact the stability. $VPR_{\Delta C}$_N3P/I5P/N238P increased the $T_{m\ (CD)}$ by 1.0 °C compared to N3P/I5P but had negligible beneficial effects on $T_{50\%}$ (Table 3) This variant had the same apparent melting point $T_{m\ (DSC)}$ as $VPR_{\Delta C}$_N3P/I5P indicating some higher degree of cooperativity in the unfolding as is reflected in the Arrhenius graph constructed form DSC thermograms (Supplementary Figs. 3 and 10). However, fitting of data using CalFitter a three-state unfolding process was needed to adequately fit it. The results gave E_{act} values (Table 4) slightly higher than for $VPR_{\Delta C}$_N3P/I5P for the first transition, which was consistent with the effects observed from the N238P mutation on the wild type. The fact that the second transition was almost unaffected suggests that the N238P mutation does not affect the stability of the N-terminal region or parts of the protein directly impacted by the N-terminal mutations. The increased stability of $VPR_{\Delta C}$_N3P/I5P/N238P is thus localized to parts of the protein corresponding to the first transition. Thus, placing the initiation point of thermal unfolding around that mutation. MD-simulation at different temperatures have suggested that the initiation point of unfolding is helix D[29]. Helix D is proximal to the Ca1 site, a site that is important for the stability of the structure[25,30,31]. Thus, by restricting movements and enforce interactions to that part of the protein might stabilize it at higher temperatures. Why this mutation had destabilizing effects when added to the wild type but stabilizes the structure when $VPR_{\Delta C}$_N3P/I5P is the template, is an interesting observation. The site of the N238P mutation is on a loop at the C-end of helix E, a helix that may well be one of the more stable parts of the structure[29]. The mutation does cause loss of H-bond potentials that could be the reason for the detrimental effects observed on the wild type template. However, having stabilized the N-terminal region to a higher degree compared to the rest of the protein the loss of H-bond potential does not show detrimental effects. This may mean that the part of the protein affected by the N238P mutation i.e. helix E, helix F and possibly

helix D are reliant on the N-terminal and parts directly stabilized by the proline N-terminal mutations. Thus, the discrepancies in the effects of N238P on $VPR_{\Delta C}$ and $VPR_{\Delta C}$_N3P/I5P can be explained if there is a chronological order to events in the unfolding of VPR and its variants. In the case of $VPR_{\Delta C}$ the area affected by N238P unfolds early and immediately thereafter the rest of the protein unfolds. In the case of $VPR_{\Delta C}$_N238P the proline might anchor that part of the protein and rigidify it, but the parts that this area is being anchored to do not provide the matching interactions needed for increased stability, hence the structure unfolds as in the case of the wild type, possibly with more cooperativity. When the N-terminus has been stabilized by the N3P/I5P mutations the anchoring points within the structure crucial for the stability of the N238P region do stay intact at higher temperatures, thus increasing the stability of the variant and increasing the apparent cooperativity of unfolding compared to $VPR_{\Delta C}$_N3P/I5P (Fig. 6). This shows that the terminal regions and the major bulk of the protein are finely tuned in the wild type to unfold in a very cooperative manner, strengthening the idea that cooperativity in the unfolding of kinetically stable proteins are under some evolutionary pressures. It also suggests that the chronological order of the unfolding events is an aspect worth looking into regarding the effectiveness of proline substitution.

The $VPR_{\Delta C}$_N3P/I5P/T265P variant. $VPR_{\Delta C}$_N3P/I5P/T265P increased the $T_{m\,(CD)}$ by 1.4 °C, the $T_{50\%}$ by 1.9 °C and $T_{m\,(DSC)}$ by 1.6 °C when compared to $VPR_{\Delta C}$_N3P/I5P (Table 3). The thermogram for this variant shows clear signs of an intermediate state and was thus fitted to a three-state model (Supplementary Figs. 3 and 11). The T265P mutation had little to no effect on E_{act} of the first transition indicating that the mutation causes the first transition state to be less entropically favored as the melting point is indeed higher, a similar observation as was made in the case of $VPR_{\Delta C}$_T265P. However, in this case it was very clear that a second transition was present which yielded a considerably higher E_{act} values for $VPR_{\Delta C}$_N3P/I5P/T265P than $VPR_{\Delta C}$_N3P/I5P and an increased apparent melting point. This may suggest considerable synergic interaction between the area affected by T265P and the region of the enzyme affected by N3P/I5P. This rings true as the loop where T265P is located interacts with the Ca3 site, most notably Arg14 and has multiple H-bonding potentials with Arg252, Ser254 and Asp 274 along with the addition of cation-π interaction with Tyr270, possibly yielding the higher activation energy of the second peak. Enforcing these interactions would in turn help maintaining the hydrophobic core of the protein due to the proximity to Leu268 and Leu269 to the mutation site (Fig. 7). In addition, Thr265 lies on an adjacent loop to Asp196 a main coordinator of the calcium ion at the Ca1 site with α-carbon distances in the crystal structure of just 7.0 Å (Fig. 7). Anchoring this part of the protein which harbours many interactions between terminals of the protein, in addition of being in close proximity to the Ca1 site, has the potential of facilitating the correct orientations of these interactions at higher temperatures thus explaining the increased stability of both transitions.

The $VPR_{\Delta C}$_N3P/I5P/N238P/T265P variant. The final product, $VPR_{\Delta C}$_N3P/I5P/N238P/T265P was also the most stable variant with a melting point of 72.1 °C ($T_{m\,(CD)}$), 10.2 °C higher than the wild type and 4.3 °C higher than $VPR_{\Delta C}$_N3P/I5P. $T_{50\%}$ values were however least affected, being slightly lower than that of $VPR_{\Delta C}$_N3P/I5P/T265P, but yielded an increase of 7.8 °C as compared to the wild type. The largest increase in stability was recorded by DSC with an apparent melting of 77.2 °C (Table 3), which was 13.3 °C higher than the wild type and 5.2 °C higher than $VPR_{\Delta C}$_N3P/I5P. However, this was accompanied by a very clear unfolding intermediate, with both transitions being more stable than seen in any other variant especially the second transition (Supplementary Fig. 12). A notable change that was observed was the low E_{act} fitted to the first transition and the very high E_{act} of the second transition. This is a similar observation as for $VPR_{\Delta C}$_I5P, where the overlap of unfolding events was low and the estimated E_{act} of the first transition was also low. This supports the idea that the first part to unfold is a major part of the α/β structure, as its unfolding would lead to the exposure of hydrophobic residues which would be entropically more unfavourable than the subsequent unfolding of the terminal region, along with at least calcium binding site 3. This latter process would be entropically favourable due to the release of the bound calcium ion and as a consequence would transform a rigid calcium binding loop into a flexible loop thus increasing the entropy of the system[32,33]. The synergic effects in this variant show that both transitions are highly stabilized, but the second transition is the more affected. From the $VPR_{\Delta C}$_N3P/I5P/T265P variant it is clear, that the T265P mutation has synergic effects with the terminal region of the protein. The quadruple proline variant adds implication of synergic effects between N238P and T265P. These two mutation sites are located on loops on the either side of helix F and neighbouring residue of the N238P site is Glu236 on helix E that likely forms a salt-bridge to Arg252 located on helix F. Ser254 located on the loop at the C-end of helix F can form a H-bond to Arg14 as mentioned earlier and is also involved in two mainchain-mainchain H-bonds that can be formed between Ser254 and Leu268, a partner in the hydrophobic core of the protein. In addition, Leu268 forms two mainchain-mainchain H-bonds to Phe198, possibly providing extra stabilization to the Ca1 binding site by restricting the neighbouring Asp196, a main coordinator in that calcium binding site (Fig. 7). Restricting movements at this site might be crucial for maintaining local interactions at higher temperatures, increasing the thermostability of the enzyme resulting from the combined effects of N238P and T265P.

Discussion

This work focused on $VPR_{\Delta C}$, a cold active subtilisin-like serine protease from the proteinase K family[5]. $VPR_{\Delta C}$ is expressed as a preproenzyme, containing a N-terminal intramolecular chaperone that is cleaved off during maturation leaving a 28 kDa active protease with an α/β-fold[11,13,34]. This maturation leaves a kinetically stable enzyme in a process that may be similar to what has been described for α-lytic protease[11,13,25,34,35]. During this maturation the calcium binding site 3 (Ca3) (Fig. 1) is likely formed. $VPR_{\Delta C}$ contains three calcium binding sites and calcium binding is one of the most important structural factors contributing to the stability of the enzyme.

The most important calcium binding sites for the stability are believed to be calcium binding site 1 (Ca1) (Fig. 1), mainly coordinated by the conserved Asp196 residue and Ca3 located in the N-terminal region[25,30]. The third calcium binding site is the low affinity Ca2 site thought to mainly serve as a defence against exogenous proteolysis[25]. The aim of the study was to explore the effects of insertion of proline residues into loops on the stability and activity of VPR and to construct a more stable VPR variant. The aim of creating a more thermostable variant of VPR by proline substitutions into loops was successful and this stabilization was achieved without losing catalytic efficiency. The use of the thermostable structural homolog AQUI as a template for selecting proline mutations was therefore a successful strategy. Research has been carried out on AQUI[36] where the same proline residues were exchanged to the corresponding residues in VPR, also indicated that the N-terminal proline residues were the most important with regards to thermostability. In that study the DSC thermograms were recorded at pH 7.4 in a filtered phosphate buffer containing 1 mM calcium, that would leave a very low concentration of calcium in the samples after dialysis and filtration[37]. There the mutant AQUI_P7I (corresponding to I5P) exhibited a thermogram with an apparent melting point almost 20 °C lower than the wild type. However, the thermogram of AQUI_P5N (corresponding N3P) a small peak was observed in front of the main transition, which had a similar unfolding initial temperature as AQUI_P7I. In the light of the results obtained for the VPR proline variants in the present study, these rather drastic but varied effects observed in the DSC thermograms of these N-terminal AQUI variants, further support the hypothesis that the Ca3 site is one of more important sites for the stability of these protein structures. The extremely low calcium concentrations in those AQUI experiments may however further exacerbate the destabilizing effects of the N-terminal mutations, as the two corresponding calcium binding sites in AQUI (Ca1 and Ca3) may be partly or fully depleted of calcium due to phosphate coprecipitation. In addition, P240N (corresponding to N238P) and P268T (corresponding to T265P) did also cause destabilization although not to the same extent as in the N-terminal variants. Thus, all these proline residues do serve a stabilizing role in the structure of AQUI.

This study also sheds some light on the way proline residues affect protein kinetic stability. Proline exchanges have been observed to have either a beneficial or detrimental effect on protein stability, even detrimental when mutation sites were selected on basis of structural comparisons to more stable structural homologues[38–41]. What causes these discrepancies in observations made on proline substitutions is not clear and a subject of debate regarding the importance of proline residues in protein structures. In the present study none of the variants showed any significant changes in their secondary structures, according to the far UV CD wavelength scans (Supplementary Fig. 2). An aspect of proline exchanges to consider is the loss of H-bonds due to the cyclic nature of the side chain of the residue which cannot therefore act as a H-bond donor. The N-terminal proline substitutions are not expected to cause any loss of H-bonds based on the crystal structure. The N238P and T265P substitutions however, do cause a loss of hydrogen bonding potential. In the case of N238P the potential of the N238 side-chain to Gln235 main-chain H-bond is lost, possibly, to a degree explaining the loss of stability resulting from this mutation when the wild type was used as a template. In the case of the T265P mutation the most likely H-bond potential to be lost is from the side-chain of Thr265 and the side-chain of Lys267, which effect is expected to be rather benign as it is solvent exposed, thus reducing its expected lifetime in the structure. The only measurable evidence for structural changes in the native state of proline variants were obtained from steady-state fluorescence spectroscopy and acrylamide quenching experiments (Figs. 2 and 3). Quenching experiments showed restricted accessibility to Trp residues accompanied with $VPR_{\Delta C}$_I5P, and even more in the case of the $VPR_{\Delta C}$_N3P/I5P variant. In those experiments Trp6 could be acting as a reporter on dynamics of the N-terminal region. The N-terminal mutations are thus likely to be causing restrictions of movements within the N-terminal region. These observations agree with the idea that the disproportional local stabilization within the N-terminus of the protein is caused by the restricted movements resulting from the N-terminal proline residues. This increased local stability leads to the emergence of an unfolding intermediate that could be observed in DSC thermograms of all proline variants containing N-terminal mutations, implicating the N-terminus as a part of the unfolding intermediate. The highly energetic second transition also fits well with observations on the effects of calcium on the calorimetric enthalpy of unfolding for VPR, as it has been shown that the calorimetric enthalpy of denaturation increases with increased calcium concentration in the buffers[25]. Measured melting points by CD do not increase above 10 mM calcium, but the rates of inactivation, the calorimetric enthalpy and the apparent melting point measured by DSC still increase up to 100 mM due to binding to the low affinity Ca2 site (Fig. 1) and the moderate affinity Ca1 site[25]. This shows that in the unfolding of VPR the calcium binding sites contribute greatly to the calorimetric heat evolved during unfolding and that changes to these sites would likely be well observable in DSC thermograms. Stabilization of the overall structure is also achieved however, thus the local stabilization of the N-terminal part reverberates throughout the structure. The N-terminal region interacts with several parts of the protein. Helix A that holds two of the Ca3 coordinating residues also is part of the hydrophobic interface between the N-terminus and the main body of the protein through Ile11 and Leu8 (Fig. 7). The Ca3 binding loop also contains Arg10 that can form a H-bond network with several residues at the C-terminal part of the protein, in addition to cation-π interaction with Tyr270. For $VPR_{\Delta C}$_N3P/I5P it is apparent that the stability of the second transition of the DSC thermograms is not affected to a large extent but most of the stabilization is reflected in the first transition. $VPR_{\Delta C}$_N3P/I5P shifts the autocatalytic site by two residues[26] and this provides more H-bond potential between these extra residues and the loop following helix A in a location where the loop is proximal to the C-terminal parts (Fig. 7). Thus, the N-terminal prolines might be seen acting as anchors maintaining these interactions between distant parts of the protein at higher temperatures. In addition, the stabilization of the protein structure caused by the $VPR_{\Delta C}$_N3P/I5P creates an environment where N238P and T265P can contribute more to stabilization. Fluorescence quenching experiments of the single N238P and T265P variants indicate changes in the microenvironments of a Trp residue or residues. As seen in lower intrinsic quenching, higher sensitivity to acrylamide quenching and a red shift in λ_{max} indicating a more polar environment. These same observations are very apparent when these mutations are added onto the $VPR_{\Delta C}$_N3P/I5P template (Table 2).

How T265P and N238P effect $VPR_{\Delta C}$ and $VPR_{\Delta C}$_N3P/I5P show that the microenvironment of a Trp residue, likely other than Trp6, is clearly influenced. Considering proximity in the structure, Trp191 is a strong candidate for being that Trp residue, a residue that might probe for change in movements around the Ca1 site and its main coordinator Asp196. Synergic effects between the N238P and T265P mutations indicate reduced movements as a function of temperature on the affected fluorophore as the change in Stern-Volmer constants between 15 °C and 35 °C are lower for $VPR_{\Delta C}$_N3P/I5P/N238P/T265P compared to the temperature effects observed for $VPR_{\Delta C}$_N3P/I5P/N238P and $VPR_{\Delta C}$_N3P/I5P/T265P. These synergic effects observed in fluorescence quenching might indicate that N238P and T265P are indeed affecting the same fluorophore.

In this study we demonstrated that insertion of proline residues into loops contributed significantly to the stability of $VPR_{\Delta C}$ variants. The mode of action is likely by restricting movements at critical points in the structure that enforces pre-existing interactions by anchoring certain parts of the protein in correct positions at higher temperatures. The restrictive nature of proline residues could thus decrease the flexibility of the structure at low temperatures but allow for more movement at higher temperatures without losing the structural integrity of the protein by retaining the interactions as more thermal energy is applied to the system. The conclusion of these observations would be that the role of proline residues in loops in the kinetic stability of proteins is to allow for more thermal flexibility of the structure[42]. Proline substitution is an effective way to stabilize kinetically stable proteins, however as shown their position within the protein structure is of utmost importance. Surface loops are good targets due to the structural nature of the residue that may leave some H-bonds unfulfilled within the protein core, possibly causing destabilization of the structure[43,44]. As proline residues seem to enforce important interactions, including interactions between distant parts of the primary structure, it would mean that the effects of prolines in the structure could be rendered useless if a crucial counterpart to the interaction being strengthened is destabilized to a certain degree or not present. Explaining the non-additive nature of combining proline mutations in this study and some of the discrepancies observed in proline mutagenic studies so far. Although the stability of the final product is greatly increased compared to the wild type, an unfolding intermediate was observed in the unfolding process. This may explain the higher degree of stabilization observed in CD and DSC as compared to $T_{50\%}$ values and the trend of lower activation energies of inactivation ($E_{act\ (inactivation)}$) (Table 4) of the more stable proline variants as the intermediate state may be a good target for exogenous proteolysis[27]. Under our standard experimental conditions this intermediate is not observable in the unfolding process of the wild type. However, when unfolding takes place at pH 5 and 1 mM $CaCl_2$, conditions known to be destabilizing for the enzyme, an intermediate could be observed (Supplementary Fig. 13). This intermediate bears many similarities to what was observed in case of the $VPR_{\Delta C}$_I5P unfolding process (Supplementary Figs. 6, 13, 17 and 18). Under these destabilising conditions it may be that calcium binding is compromised. However, if the calcium ion concentration is increased to 15 mM the first transition is stabilized and the intermediate is not as readily observed (Supplementary Fig. 13). These observations fit our ideas concerning the unfolding pathway and the role of the Ca1 binding site[25], as a compromised Ca1 binding site could lead to destabilisation of a possible unfolding initiation point at helix D[29]. In addition, low pH values could destabilise further interactions important for the stability of the protein structure but less so for the intermediate state. This information thus indicates that there is a metastable intermediate along the unfolding pathway that is poorly structured and which by tweaking conditions, or by increasing the local stability of the N-terminal region can become kinetically trapped along the unfolding pathway causing the apparent loss of cooperativity in the unfolding process. This also suggests that the quadruple variant has the potential to be stabilized even further by tweaking the stability of the first transition and increasing the cooperativity of unfolding. As a result, $VPR_{\Delta C}$_N3P/I5P/N238P/T265P is a prime candidate for further work exploring enzyme kinetic stability, thermostability, protein engineering and temperature adaptation.

Materials and Methods

Site directed mutagenesis. All mutations were done on the gene of $VPR_{\Delta C}$ (a C-terminal truncated form of wild type VPR)[13], contained in a pET-11a-d vector[25]. Proline variants were obtained with site-directed mutagenesis using Q5 site-directed mutagenesis kit from New England Biolabs (NEB) following their protocol. Mutagenic primers used to produce the variants were designed using the web tool NEBaseChanger (NEBaseChanger.neb.com), except for the I5P variant which was made using the Quickchange Site directed Mutagenesis Kit from Stratagene, following their protocol (Table 5). Mutagenic PCR products were all transformed into XL10-Gold from Agilent Technologies, genotype: *TetrD(mcrA)183 D(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte [F´ proAB lacIqZDM15 Tn10 (Tetr) Amy Camr]*. Plasmid purification was done with the Monarch plasmid miniprep kit from NEB, following their instructions. All mutations were verified by Sanger sequencing performed by Genewiz.

Expression and purification. All proline variants were expressed in the *E. coli* strain Lemo21 (NEB) from a pET-11a-d vector, utilizing the T7 polymerase/T7 lysozyme system[45]. Liquid media used for the expression of all variants was 2xYT broth containing 0.1 mg/mL ampicillin (Sigma), 0.03 mg/mL chloramphenicol (Sigma) and 76 µM rhamnose (Sigma). Cultures were grown to a density of A_{600} ~ 0.4–0.8 A.U. and expression was initialized by adding isopropyl β-D-1-thiogalactopyranoside (IPTG) (AppliChem) to a final concentration of 400 µM followed by the addition of sterile 4 M $CaCl_2$ (Sigma) to give a final concentration of 100 mM and grown at 18 °C and 230 rpm for 20–24 hours. All proline variants were purified to homogeneity as described in[25].

Activity assays. All activity assays were performed in 100 mM Tris, 10 mM $CaCl_2$ at pH 8.6 using Suc-AAPF-NH-Np as a substrate. Kinetic parameters of proline variants were characterized by Michaelis-Menten assay monitoring activity at 25 °C against Suc-AAPF-NH-Np at seven different substrate concentrations, up to 1.00 mM, and monitoring ΔA_{410} over 15 seconds. Enzyme samples were dialyzed against the assay buffer

Primer	Sequence
N3P fw.	CGTTCAAAGCCCGGCGATTTGGG
N3P rv.	GCTTCATTTGAAACAACAG
I5P fw.	GTTCAAAGCAACGCGCCGTGGGGGCTAGACCG
I5P rv.	CGGTCTAGCCCCCACGGCGCGTTGCTTTGAAC
N3P/I5P fw.	GCCGTGGGGGCTAGACCGAATA
N3P/I5P rv.	GCCGGGCTTTGAACGGCTTCATTTG
N238P fw.	ACAAGAAAACCCGGGCTTAACTCCGCTTC
N238P rv.	AAGTACAAGGCTGCAACG
T265P fw.	AAGAGGCACCCCGAATAAACTGC
T265P rv.	GTGTCAGAAACCTTATTCTC

Table 5. Sequences of mutagenic primers used.

overnight at 4 °C and concentration was estimated by A_{280} measurements using the calculated molar attenuation coefficient 34,170 $M^{-1}cm^{-1}$[46]. Exact substrate concentrations were determined at 410 nm, using the molar attenuation coefficient 8,480 $M^{-1}cm^{-1}$[47]. Data points were then fitted to the Michaelis-Menten equation using the analysis software KaleidaGraph (Synergy Software).

Fluorescence. Steady state fluorescence was recorded for each variant at 15 °C, 25 °C and 35 °C on a Fluoromax-4 spectrofluorometer (Horiba Scientific) equipped with a circulating water bath for temperature control. All samples were inhibited by PMSF to a final concentration of 2.5 mM followed by dialysis against 50 mM Tris, 10 mM $CaCl_2$ and pH 8.0 overnight at 4 °C. Prior to fluorescence experiments absorbance spectra were recorded from 400 nm down to 220 nm and absorbance tuned to 0.03–0.05 A.U. at 295 nm in a 0.4 cm quartz cuvette (Spectrocell) used for fluorescence experiments. In addition to recording native fluorescence of all variants, steady state fluorescence of the denatured state was also recorded for $VPR_{\Delta C}$, $VPR_{\Delta C}$_N3P/I5P and $VPR_{\Delta C}$_N3P/I5P/N238P/T265P, where samples were heated to 90 °C for 15 minutes and fluorescence measured at 25 °C. All samples were excited at 295 nm using 3 nm entrance slit width and fluorescence monitored between 310 nm and 450 nm using a 5–8 nm exit slit width for native samples and 2–3 nm for denatured samples. Relative fluorescence was then calculated as:

$$F_{n.} = \frac{\left(\frac{CPS}{[P] * exi.^2}\right)}{F_{\mathrm{VPR}\Delta\mathrm{C}}}$$

where $F_{n.}$ is the normalized fluorescence intensity, CPS the recorded fluorescence intensity, [P] the protein concentration, exi. the exit slit width used and $F_{VPR\Delta C}$ the concentration and exit slit width normalized fluorescence for native $VPR_{\Delta C}$. The peak of each fluorescence spectra was then fitted to a cubic function, solving the first derivative for the local maximum (λ_{max}). AUC (area under curve) was calculated for all variants via the trapezoidal rule and the relative emission efficacy calculated by dividing the results with the average fluorescence intensity for native $VPR_{\Delta C}$. In addition, acrylamide quenching was conducted on all variants, using a 2.5 M stock of molecular biology grade acrylamide (Sigma). Sample preparation and experimental conditions were as described above. Each aliquot of acrylamide added to samples was followed by thorough mixing and one min resting time for temperature equilibration. The effectiveness of quenching was calculated by fitting the data with the Stern-Volmer equation:

$$\frac{F^o}{F} = 1 + K_{SV}[Q]$$

where F^0 and F are the fluorescence intensities in the absence and presence of quencher between 310 nm and 410 nm, [Q] is concentration of quencher and K_{sv} is the Stern-Volmer constant calculated via linear regression. Corrections of fluorescence intensities were performed on the data to account for dilutions due to additions of acrylamide.

Thermal stability. Prior to thermal inactivation experiments, samples were dialyzed against a 25 mM Tris buffer containing 15 mM $CaCl_2$, 100 mM NaCl, 1 mM EDTA and at pH 8.95 (Sigma) overnight at 4 °C. Samples were then heated to selected temperatures and aliquots withdrawn at timed intervals for assaying remaining activity using 0.5 mM Suc-AAPF-NH-Np. The observed first order rate constants were then used to construct Arrhenius-plots that were analysed by linear regression using KaleidaGraph, from which the $T_{50\%}$ (the temperature where half of the activity was lost after thirty minutes) and $E_{act\ (inactivation)}$ (corresponding to the slope of the Arrhenius graph) values were then calculated.

Unfolding of the secondary structure was monitored by circular dichroism (CD). Prior to measurements samples were inhibited by PMSF at a final concentration of 2.5 mM followed by dialysis against a 25 mM glycine buffer containing 100 mM NaCl and 15 mM $CaCl_2$ at pH 8.6 overnight at 4 °C. Melting curves of protein samples (0.1–0.4 mg/mL) were recorded at 222 nm with a heating rate of 1 °C/min from 25 °C to 90 °C on a Jasco J-1100 spectropolarimeter. Data analysis and $T_{m\ (CD)}$ determination was performed as described in[25]. Concurrent CD

wavelength scans were also recorded on a Jasco J-1100 from 250 nm down to 200 nm at 25 °C using a 1 mm cuvette.

Differential scanning calorimetry (DSC) was used to record thermograms of the unfolding process using a MicroCal VP-DSC. Prior to measurements, samples were inhibited by PMSF at a final concentration of 2.5 mM followed by dialysis against a 25 mM glycine buffer containing 100 mM NaCl and 15 mM $CaCl_2$ and pH 8.6 overnight at 4 °C. Prior to loading, protein samples (0.4–1.2 mg/mL) and buffers were degassed for 15–30 min at 10 °C. Thermograms were then recorded from 15 °C to 95 °C with a temperature gradient of 1 °C/min. Initial data analysis was performed by Origin software where buffer subtraction and concentration normalization was carried out. Due to a slow downward sloping post heat capacities recorded at high temperatures for some variants (Supplementary Fig. 14), the Origin software was used to normalize data sets by baseline generation to convert data sets into plots of excess heat capacity versus temperature. Initial data analysis consisted of calculating the AUC (area under curve) of each excess heat capacity plot via the trapezoidal rule yielding the calorimetric enthalpy (ΔH_{cal}), including recalculations for $VPR_{\Delta C}$ data sets from[25]. The apparent melting points ($T_{m\,(DSC)}$) were found by fitting a cubic function to the highest peak of the thermograms and solving the first derivative for the local maximum. The rate of unfolding ($k_{(unfold)}$) was calculated as:

$$k_{(unfold)} = \frac{vC_p}{Q_t - Q}$$

where v is the speed of the temperature gradient, C_p is the excess heat capacity at a given temperature, Q_t is the total heat evolved and Q is heat evolved at a given temperature[4,25,28,48]. The unfolding rates were then used to plot Arrhenius graphs and used as assistance in further analysis. Due to the complexity of some thermograms CalFitter v1.2 (https://loschmidt.chemi.muni.cz/calfitter/)[49] was utilized to separate unfolding events. The model CalFitter used for fitting irreversible transitions is a modification of the Arrhenius equation:

$$k = \exp\left(-\frac{E_{act}}{R}\left(\frac{1}{T} - \frac{1}{T_{act}}\right)\right)$$

where k is the rate of unfolding, E_{act} is the activation energy of unfolding, R is the gas constant, T is the absolute temperature and T_{act} is an expression of the preexponential factor A, that has been transformed into exponent with the single new parameter T_{act} for more robust parameter estimation. The average of normalized DSC scans of variants and a normalized CD melting curves were simultaneously subjected to global fitting using CalFitter. As CalFitter cannot account for the downward slopes as observed in some thermograms the normalized excess heat thermograms were fitted instead. For fitting of normalized curves ΔC_p of each transition was fixed at 0 along with the slope (see supplementary for reflections on data analysis). The unfolding model selected for the wild type and variants that did not exhibit complex unfolding was a two-state irreversible model confirmed by DSC scan-rate experiments (Supplementary Fig. 15)[28]. Variants that did exhibit more complex unfolding a three-state model with both transitions being irreversible, was used. The model was chosen by running scan-rate experiments on $VPR_{\Delta C}$_I5P (the variant with the most prominent second unfolding transition) (Supplementary Figs. 16, 17 and 18) revealing that both transitions exhibited scan-rate independent activation energies with more accumulation of the intermediate at slower scan-rates. In addition, reheating experiments[50] on $VPR_{\Delta C}$_I5P up to 70 °C and $VPR_{\Delta C}$_N3P/I5P/N238P/T265P up to 72 °C showed no signs of refolding (Supplementary Figs. 19 and 20) in case of the first transition. In addition, protein stability dependence on protein concentration which was measured by recording melting points on CD at 0.1 mg/mL, 0.2 mg/mL and 0.4 mg/mL for both $VPR_{\Delta C}$ and $VPR_{\Delta C}$_I5P (Supplementary Fig. 21), showed no concentration dependence for either variant, excluding the possibility of oligomerization causing the observed second transitions.

Molecular modelling and graphical data representation. Molecular graphics and H-bond analysis was performed with UCSF Chimera[51] using the crystal structure of AQUI (4DZT) and VPR (1SH7)[13]. Data sets were plotted using GraphPad Prism 6 for Windows, GraphPad software.

Received: 30 August 2019; Accepted: 27 December 2019;
Published online: 23 January 2020

References

1. Gaucher, E. A., Govindarajan, S. & Ganesh, O. K. Palaeotemperature trend for Precambrian life inferred from resurrected proteins. *Nature* **451**, 704–707, https://doi.org/10.1038/nature06510 (2008).
2. Akanuma, S. *et al.* Experimental evidence for the thermophilicity of ancestral life. *Proc. Natl. Acad. Sci. USA* **110**, 11067–11072, https://doi.org/10.1073/pnas.1308215110 (2013).
3. Knauth, L. P. & Lowe, D. R. High Archean climatic temperature inferred from oxygen isotope geochemistry of cherts in the 3.5 Ga Swaziland Supergroup, South Africa. *Geol. Soc. Am. Bull.* **115**, 566–580, 10.1130/0016-7606(2003)115<0566:Hactif>2.0.Co;2 (2003).
4. D'Amico, S., Marx, J. C., Gerday, C. & Feller, G. Activity-stability relationships in extremophilic enzymes. *J. Biol. Chem.* **278**, 7891–7896, https://doi.org/10.1074/jbc.M212508200 (2003).
5. Kristjansson, M. M., Magnusson, O. T., Gudmundsson, H. M., Alfredsson, G. A. & Matsuzawa, H. Properties of a subtilisin-like proteinase from a psychrotrophic Vibrio species comparison with proteinase K and aqualysin I. *Eur. J. Biochem.* **260**, 752–760, https://doi.org/10.1046/j.1432-1327.1999.00205.x (1999).
6. Gudjonsdottir, K. & Asgeirsson, B. Effects of replacing active site residues in a cold-active alkaline phosphatase with those found in its mesophilic counterpart from Escherichia coli. *FEBS J* **275**, 117–127, https://doi.org/10.1111/j.1742-4658.2007.06182.x (2008).
7. Siddiqui, K. S. & Cavicchioli, R. Cold-adapted enzymes. *Annu. Rev. Biochem.* **75**, 403–433, https://doi.org/10.1146/annurev.biochem.75.103004.142723 (2006).

8. Nguyen, V. *et al*. Evolutionary drivers of thermoadaptation in enzyme catalysis. *Science* **355**, 289–293, https://doi.org/10.1126/science.aah3717 (2017).
9. Miyazaki, K., Wintrode, P. L., Grayling, R. A., Rubingh, D. N. & Arnold, F. H. Directed evolution study of temperature adaptation in a psychrophilic enzyme. *J. Mol. Biol.* **297**, 1015–1026, https://doi.org/10.1006/jmbi.2000.3612 (2000).
10. Wintrode, P. L., Miyazaki, K. & Arnold, F. H. Cold adaptation of a mesophilic subtilisin-like protease by laboratory evolution. *J. Biol. Chem.* **275**, 31635–31640, https://doi.org/10.1074/jbc.M004503200 (2000).
11. Sigurdardottir, A. G. *et al*. Characteristics of mutants designed to incorporate a new ion pair into the structure of a cold adapted subtilisin-like serine proteinase. *Biochim. Biophys. Acta.* **1794**, 512–518, https://doi.org/10.1016/j.bbapap.2008.11.018 (2009).
12. Matsuzawa, H. *et al*. Purification and Characterization of Aqualysin-I (a Thermophilic Alkaline Serine Protease) Produced by Thermus-Aquaticus Yt-1. *Eur. J. Biochem.* **171**, 441–447, https://doi.org/10.1111/j.1432-1033.1988.tb13809.x (1988).
13. Arnorsdottir, J., Kristjansson, M. M. & Ficner, R. Crystal structure of a subtilisin-like serine proteinase from a psychrotrophic Vibrio species reveals structural aspects of cold adaptation. *FEBS J.* **272**, 832–845, https://doi.org/10.1111/j.1742-4658.2005.04523.x (2005).
14. Kumar, S., Tsai, C. J. & Nussinov, R. Factors enhancing protein thermostability. *Protein Eng.* **13**, 179–191, https://doi.org/10.1093/protein/13.3.179 (2000).
15. Haney, P., Konisky, J., Koretke, K. K., Luthey-Schulten, Z. & Wolynes, P. G. Structural basis for thermostability and identification of potential active site residues for adenylate kinases from the archaeal genus Methanococcus. *Proteins* **28**, 117–130, 10.1002/(Sici)1097-0134(199705)28:1<117::Aid-Prot12>3.0.Co;2-M (1997).
16. Watanabe, K., Hata, Y., Kizaki, H., Katsube, Y. & Suzuki, Y. The refined crystal structure of Bacillus cereus oligo-1,6-glucosidase at 2.0 angstrom resolution: Structural characterization of proline-substitution sites for protein thermostabilization. *J. Mol. Biol.* **269**, 142–153, https://doi.org/10.1006/jmbi.1997.1018 (1997).
17. Bogin, O. *et al*. Enhanced thermal stability of Clostridium beijerinckii alcohol dehydrogenase after strategic substitution of amino acid residues with prolines from the homologous thermophilic Thermoanaerobacter brockii alcohol dehydrogenase. *Protein Sci.* **7**, 1156–1163, https://doi.org/10.1002/pro.5560070509 (1998).
18. Raymond-Bouchard, I. *et al*. Conserved genomic and amino acid traits of cold adaptation in subzero-growing Arctic permafrost bacteria. *Fems. Microbiol. Ecol.* **94**, https://doi.org/10.1093/femsec/fiy023 (2018).
19. Macarthur, M. W. & Thornton, J. M. Influence of Proline Residues on Protein Conformation. *J. Mol. Biol.* **218**, 397–412, https://doi.org/10.1016/0022-2836(91)90721-H (1991).
20. Sanchez-Ruiz, J. M. Protein kinetic stability. *Biophys. Chem.* **148**, 1–15, https://doi.org/10.1016/j.bpc.2010.02.004 (2010).
21. Truhlar, S. M. E., Cunningham, E. L. & Agard, D. A. The folding landscape of Streptomyces griseus protease B reveals the energetic costs and benefits associated with evolving kinetic stability. *Protein Sci.* **13**, 381–390, https://doi.org/10.1110/ps.03336804 (2004).
22. Jaswal, S. S., Truhlar, S. M. E., Dill, K. A. & Agard, D. A. Comprehensive analysis of protein folding activation thermodynamics reveals a universal behavior violated by kinetically stable proteases. *J. Mol. Biol.* **347**, 355–366, https://doi.org/10.1016/j.jmb.2005.01.032 (2005).
23. Xie, Y. *et al*. Enhanced Enzyme Kinetic Stability by Increasing Rigidity within the Active Site. *J. Biol. Chem.* **289**, 7994–8006, https://doi.org/10.1074/jbc.M113.536045 (2014).
24. Manning, M. & Colon, W. Structural basis of protein kinetic stability: resistance to sodium dodecyl sulfate suggests a central role for rigidity and a bias toward beta-sheet structure. *Biochemistry* **43**, 11248–11254, https://doi.org/10.1021/bi0491898 (2004).
25. Oskarsson, K. R. & Kristjansson, M. M. Improved expression, purification and characterization of VPR, a cold active subtilisin-like serine proteinase and the effects of calcium on expression and stability. *Biochim. Biophys. Acta. Proteins. Proteom.* **1867**, 152–162, https://doi.org/10.1016/j.bbapap.2018.11.010 (2019).
26. Arnorsdottir, J., Sigtryggsdottir, A. R., Thorbjarnardottir, S. H. & Kristjansson, M. M. Effect of proline substitutions on stability and kinetic properties of a cold adapted subtilase. *J. Biochem.* **145**, 325–329, https://doi.org/10.1093/jb/mvn168 (2009).
27. Fontana, A., deLaureto, P. P., DeFilippis, V., Scaramella, E. & Zambonin, M. Probing the partly folded states of proteins by limited proteolysis. *Fold. Des.* **2**, R17–R26, https://doi.org/10.1016/S1359-0278(97)00010-2 (1997).
28. Sanchez-Ruiz, J. M., Lopez-Lacomba, J. L., Cortijo, M. & Mateo, P. L. Differential scanning calorimetry of the irreversible thermal denaturation of thermolysin. *Biochemistry* **27**, 1648–1652, https://doi.org/10.1021/bi00405a039 (1988).
29. Du, X. *et al*. Comparative thermal unfolding study of psychrophilic and mesophilic subtilisin-like serine proteases by molecular dynamics simulations. *J. Biomol. Struct. Dyn.* **35**, 1500–1517, https://doi.org/10.1080/07391102.2016.1188155 (2017).
30. Tiberti, M. & Papaleo, E. Dynamic properties of extremophilic subtilisin-like serine-proteases. *J. Struct. Biol.* **174**, 69–83, https://doi.org/10.1016/j.jsb.2011.01.006 (2011).
31. Helland, R., Larsen, A. N., Smalas, A. O. & Willassen, N. P. The 1.8 angstrom crystal structure of a proteinase K-like enzyme from a psychrotroph Serratia species. *FEBS J.* **273**, 61–71, https://doi.org/10.1111/j.1742-4658.2005.05040.x (2006).
32. King, J. T., Arthur, E. J., Brooks, C. L. & Kubarych, K. J. Site-Specific Hydration Dynamics of Globular Proteins and the Role of Constrained Water in Solvent Exchange with Amphiphilic Cosolvents. *J. Phys. Chem. B.* **116**, 5604–5611, https://doi.org/10.1021/jp300835k (2012).
33. Dragan, A. I., Read, C. M. & Crane-Robinson, C. Enthalpy-entropy compensation: the role of solvation. *Eur. Biophys. J. Biophy.* **46**, 301–308, https://doi.org/10.1007/s00249-016-1182-6 (2017).
34. Arnorsdottir, J. *et al*. Characterization of a cloned subtilisin-like serine proteinase from a psychrotrophic Vibrio species. *Eur. J. Biochem.* **269**, 5536–5546, https://doi.org/10.1046/j.1432-1033.2002.03259.x (2002).
35. Sohl, J. L., Jaswal, S. S. & Agard, D. A. Unfolded conformations of alpha-lytic protease are more stable than its native state. *Nature* **395**, 817–819, https://doi.org/10.1038/27470 (1998).
36. Sakaguchi, M. *et al*. Role of proline residues in conferring thermostability on aqualysin I. *J. Biochem.* **141**, 213–220, https://doi.org/10.1093/jb/mvm025 (2007).
37. Chow, L. C. Solubility of calcium phosphates. *Monogr. Oral. Sci.* **18**, 94–111 (2001).
38. Kini, R. M. & Evans, H. J. A hypothetical structural role for proline residues in the flanking segments of protein-protein interaction sites. *Biochem. Biophys. Res. Commun.* **212**, 1115–1124, https://doi.org/10.1006/bbrc.1995.2084 (1995).
39. Matthews, B. W., Nicholson, H. & Becktel, W. J. Enhanced Protein Thermostability from Site-Directed Mutations That Decrease the Entropy of Unfolding. *P. Natl. Acad. Sci. USA* **84**, 6663–6667, https://doi.org/10.1073/pnas.84.19.6663 (1987).
40. Vieille, C. & Zeikus, G. J. Hyperthermophilic enzymes: Sources, uses, and molecular mechanisms for thermostability. *Microbiol. Mol. Biol. R.* **65**, 1–43, https://doi.org/10.1128/MMBR.65.1.1-43.2001 (2001).
41. Sriprapundh, D., Vieille, C. & Zeikus, J. G. Molecular determinants of xylose isomerase thermal stability and activity: analysis of thermozymes by site-directed mutagenesis. *Protein Eng.* **13**, 259–265 (2000).
42. Quezada, A. G. *et al*. Interplay between Protein Thermal Flexibility and Kinetic Stability. *Structure* **25**, 167–179, https://doi.org/10.1016/j.str.2016.11.018 (2017).
43. Stickle, D. F., Presta, L. G., Dill, K. A. & Rose, G. D. Hydrogen bonding in globular proteins. *J. Mol. Biol.* **226**, 1143–1159 (1992).
44. McDonald, I. K. & Thornton, J. M. Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**, 777–793, https://doi.org/10.1006/jmbi.1994.1334 (1994).
45. Wagner, S. *et al*. Tuning Escherichia coli for membrane protein overexpression. *Proc. Natl. Acad. Sci. USA* **105**, 14371–14376, https://doi.org/10.1073/pnas.0804090105 (2008).
46. Pace, C. N., Vajdos, F., Fee, L., Grimsley, G. & Gray, T. How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* **4**, 2411–2423, https://doi.org/10.1002/pro.5560041120 (1995).

47. Delmar, E. G., Largman, C., Brodrick, J. W. & Geokas, M. C. Sensitive New Substrate for Chymotrypsin. *Anal. Biochem.* **99**, 316–320, https://doi.org/10.1016/S0003-2697(79)80013-5 (1979).
48. Siddiqui, K. S. *et al.* A chemically modified alpha-amylase with a molten-globule state has entropically driven enhanced thermal stability. *Protein. Eng. Des. Sel.* **23**, 769–780, https://doi.org/10.1093/protein/gzq051 (2010).
49. Mazurenko, S. *et al.* CalFitter: a web server for analysis of protein thermal denaturation data. *Nucleic Acids Res.* **46**, W344–W349, https://doi.org/10.1093/nar/gky358 (2018).
50. Mazurenko, S. *et al.* Exploration of Protein Unfolding by Modelling Calorimetry Data from Reheating. *Sci. Rep.* **7**, https://doi.org/10.1038/s41598-017-16360-y (2017).
51. Pettersen, E. F. *et al.* UCSF Chimera–a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605–1612, https://doi.org/10.1002/jcc.20084 (2004).

Acknowledgements

This work was supported by The Icelandic Research Fund [Grant Number 162977-051].

Author contributions

M.M.K. conceived and supervised the study. K.R.Ó. and M.M.K. wrote the manuscript and all authors proof read the manuscript. K.R.Ó. and A.F.S. carried out all experiments, purifications and the construction of mutants.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-57873-3.

Correspondence and requests for materials should be addressed to M.M.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Thermostabilization of VPR, a kinetically stable cold adapted subtilase, via multiple proline substitutions into surface loops.

Supplementary.

K. R. Óskarsson, A. F. Sævarsson & M. M. Kristjánsson*

Department of Biochemistry, Science Institute, University of Iceland, Reykjavík, Iceland. Correspondence and requests for materials should be addressed to M.M.K. (email: mmk@hi.is)

Reflections on data analysis.

In this paper we utilized DSC and CD to analyse the unfolding processes of $VPR_{\Delta C}$ and its proline variants. The unfolding process of $VPR_{\Delta C}$ follows a classical two state irreversible pathway as seen from its DSC thermogram (Fig. 6) and resulting Arrhenius graph (Supplementary Fig. 3). However, six out of nine enzyme constructs measured in this work did show evidence of divergences from the cooperative unfolding observed for the wild type. In this paper we utilized CalFitter 1.2 [1] in order to analyse all data sets in the same working environment. However due to unstable pre and post heat capacities of some variants, most notably $VPR_{\Delta C}$_T265P, all data sets were converted into excess heat thermograms. Thus, the change in heat capacity due to unfolding could not be considered. Thus, we opted to report on E_{act} values instead of activation Gibbs free energies, activation enthalpy or activation entropy. Although baseline subtraction leads to a loss of information and possibly precision in parameter estimation, our observation was that the accuracy holds and was improved when working with data sets that showed signs of slow aggregation appearing as downward sloping post heat capacities after the thermogram peak. In addition, proteases such as αLP and SGPB that are produced with intramolecular chaperones and are entirely dependent on them for folding, as is the case for VPR, have shown to have very large $\Delta C_p^{\ddagger}$ and $\Delta G^{\ddagger}$ values [2]. Together this results in highly temperature dependent $\Delta G^{\ddagger}$ curves, making free energy comparisons hard between variants without having a value for $\Delta C_p^{\ddagger}$. Thus, we feel that reporting E_{act} values we get more reliable results for comparison between proline variants as baseline subtraction was always performed in the same manner. For the wild type and variants showing cooperative unfolding the simplest model for irreversible unfolding holds true:

$$N \overset{k_1}{\rightarrow} D$$

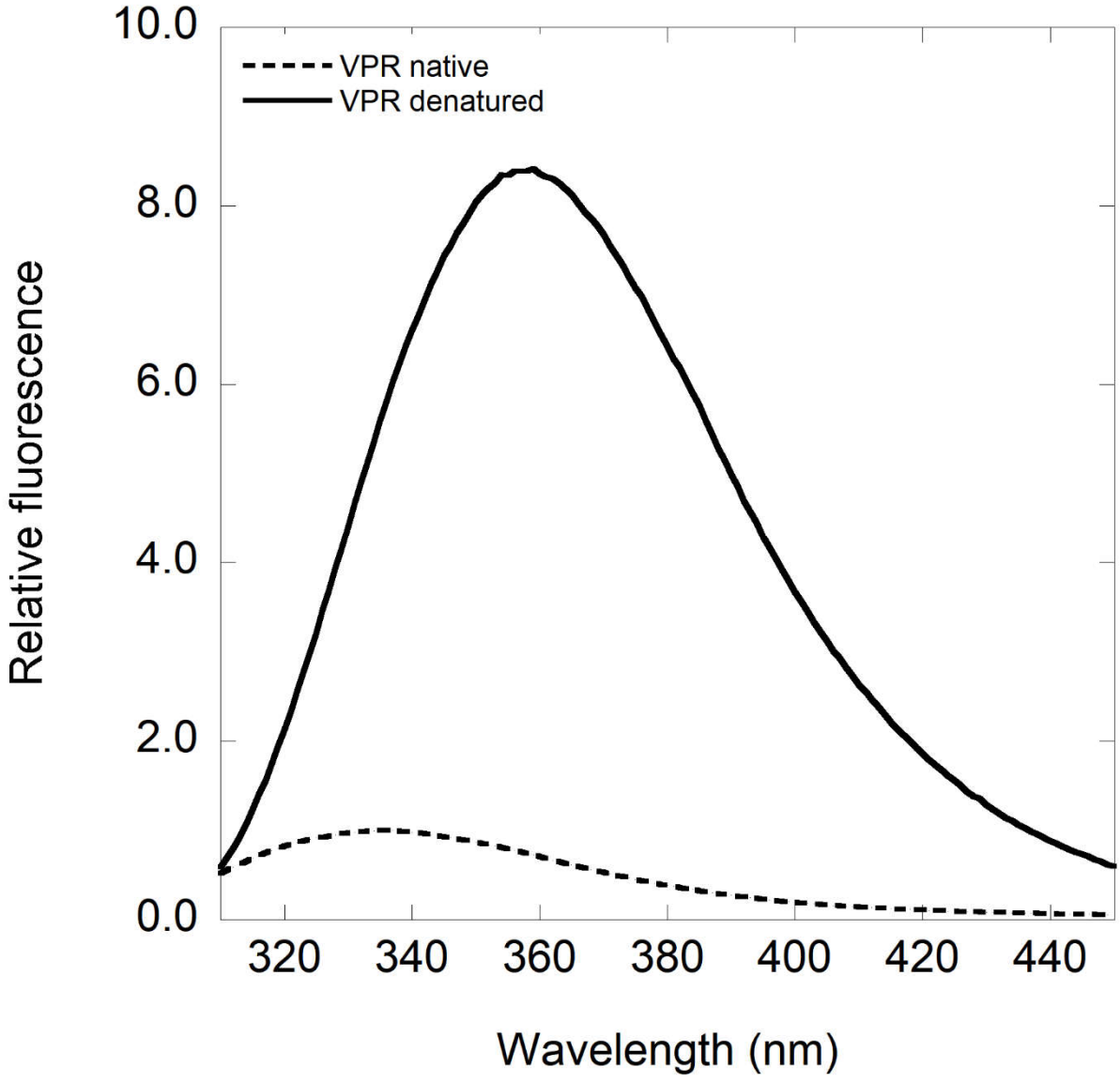
where N stands for the native state, D for the denatured one and k_1 is the rate of unfolding at a given temperature. Variants that exhibited divergences from the irreversible two-state model

did all fit best to a three-state irreversible model rather than a Lumry-Eyring unfolding. The model that best fitted the data had both steps being irreversible:

$$N \overset{k_1}{\rightarrow} I \overset{k_2}{\rightarrow} D$$

where I stands for the intermediate and k_2 stands for the unfolding rate of the intermediate state at a given temperature and used to calculate the activation energy of the second transition. Partial unfolding experiments also indicated that the first transition was irreversible or at least slow enough that it did not come into effect during experiments (Supplementary Fig. 19 & 20).

Supplementary Figures and Tables.



Supplementary Fig. 1. Fluorescence emission of $VPR_{\Delta C}$ at 25°C and pH 8.0 before heat treatment (dotted line) and after heat treatment (solid line). Heat treatment was carried out by heating at 85 °C for 15 min and then cooled down before measurements.

Supplementary Table 1. Relative fluorescence intensities of proline variants calculated as the AUC of fluorescence curves at different temperatures and of the denatured state at 25 °C. Along with the standard deviation of the mean.

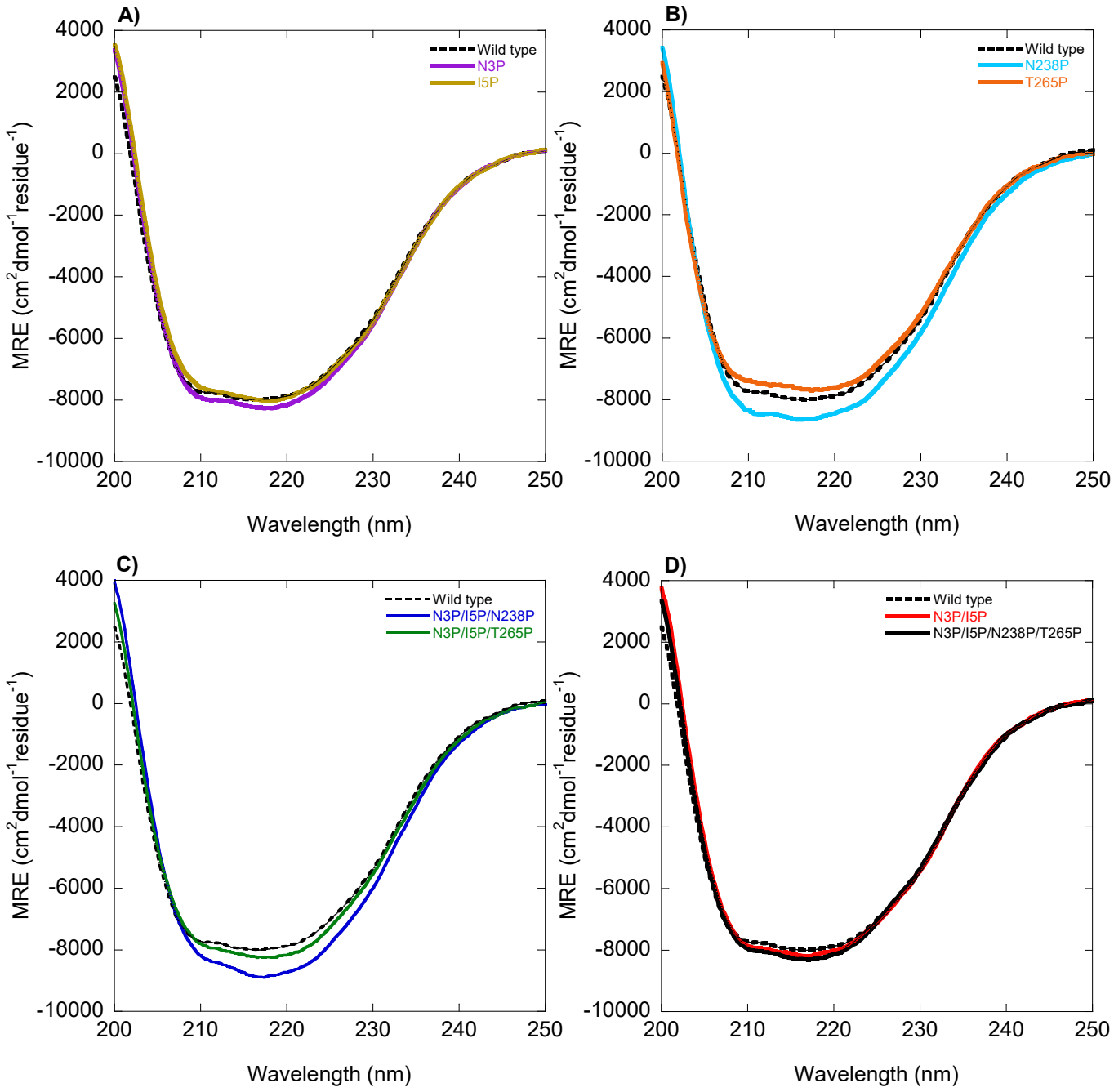
Variant	15 °C (Relative intensity)	25 °C (Relative intensity)	35 °C (Relative intensity)	Denatured 25 °C (Relative intensity)
VPR$_{\Delta C}$	1.12 ± 0.02	1.00 ± 0.03	0.95 ± 0.03	10.02 ± 0.04
VPR$_{\Delta C}$/N3P	1.46 ± 0.08	1.37 ± 0.02	1.28 ± 0.06	Not measured
VPR$_{\Delta C}$/I5P	1.28 ± 0.04	1.16 ± 0.05	1.12 ± 0.07	Not measured
VPR$_{\Delta C}$/N238P	1.26 ± 0.07	1.16 ± 0.07	1.09 ± 0.06	Not measured
VPR$_{\Delta C}$/T265P	1.25 ± 0.08	1.15 ± 0.08	1.07 ± 0.09	Not measured
VPR$_{\Delta C}$/N3P/I5P	1.10 ± 0.04	1.00 ± 0.03	0.94 ± 0.01	11.01 ± 0.27
VPR$_{\Delta C}$/N3P/I5P/N238P	1.30 ± 0.10	1.18 ± 0.08	1.13 ± 0.10	Not measured
VPR$_{\Delta C}$/N3P/I5P/T265P	1.34 ± 0.01	1.24 ± 0.03	1.17 ± 0.01	Not measured
VPR$_{\Delta C}$/ N3P/I5P/N238P/T265P	1.31 ± 0.08	1.19 ± 0.06	1.10 ± 0.05	10.73

Supplementary Table 2. λ_{max} values for proline variants calculated as the maxima of fluorescence curves at different temperatures and for the denatured state at 25 °C. Along with the standard deviation of the mean.

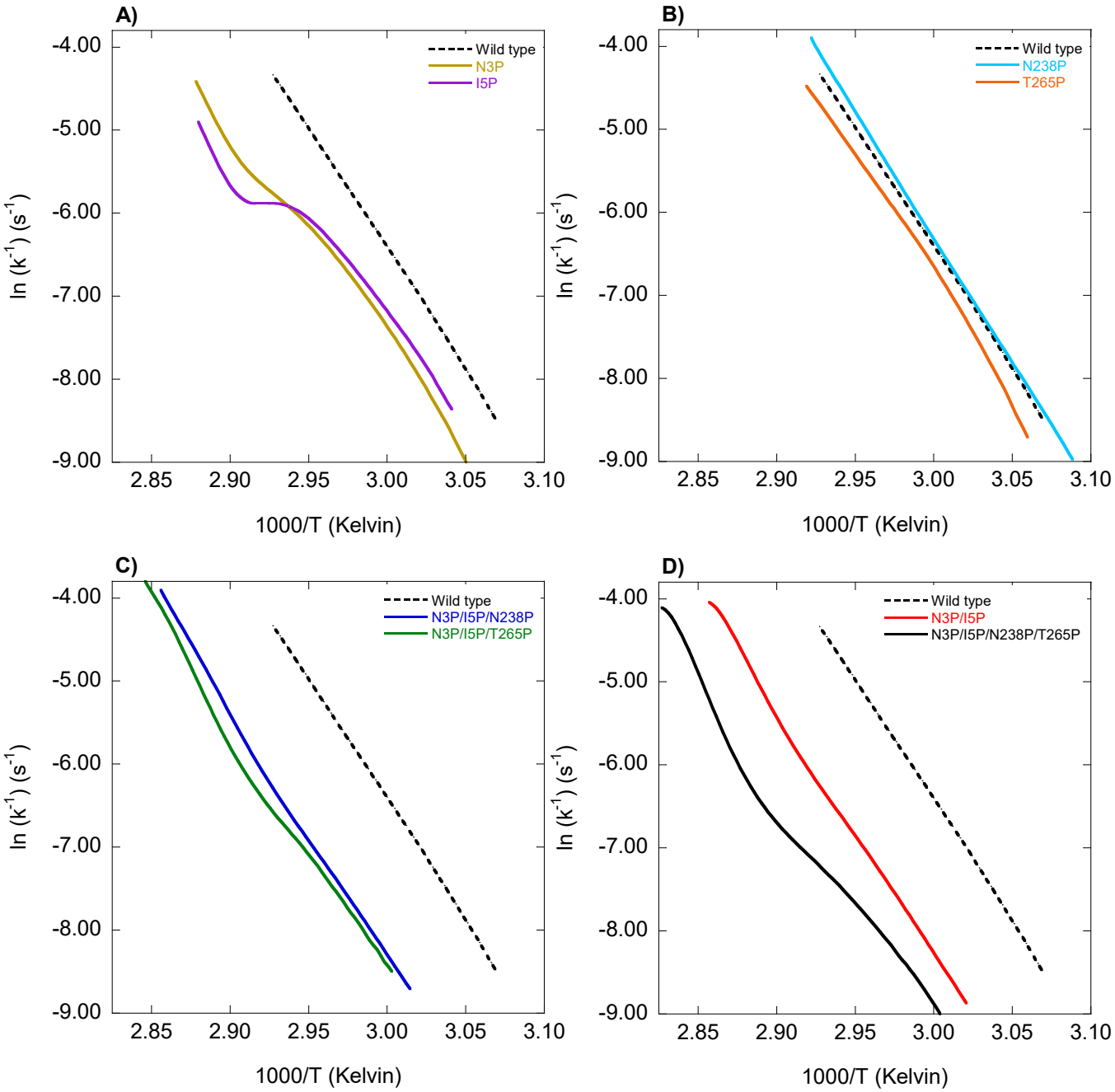
Variant	15 °C (nm)	25 °C (nm)	35 °C (nm)	Denatured 25 °C (nm)
VPR$_{\Delta C}$	335 ± 1	335 ± 1	336 ± 1	357 ± 1
VPR$_{\Delta C}$/N3P	337 ± 2	337 ± 2	337 ± 2	Not measured
VPR$_{\Delta C}$/I5P	337 ± 2	337 ± 2	337 ± 2	Not measured
VPR$_{\Delta C}$/N238P	339 ± 1	339 ± 2	339 ± 2	Not measured
VPR$_{\Delta C}$/T265P	337 ± 1	338 ± 1	338 ± 1	Not measured
VPR$_{\Delta C}$/N3P/I5P	334 ± 1	334 ± 1	334 ± 1	358 ± 1
VPR$_{\Delta C}$/N3P/I5P/N238P	339 ± 1	339 ± 1	339 ± 1	Not measured
VPR$_{\Delta C}$/N3P/I5P/T265P	338 ± 1	338 ± 1	338 ± 1	Not measured
VPR$_{\Delta C}$/ N3P/I5P/N238P/T265P	336 ± 2	336 ± 2	336 ± 2	358

Supplementary Table 3. Stern-Volmer constants calculated from quenching data between 310 nm – 410 nm at different temperatures and for the denatured state at 25 °C. Along with the standard deviation of the mean.

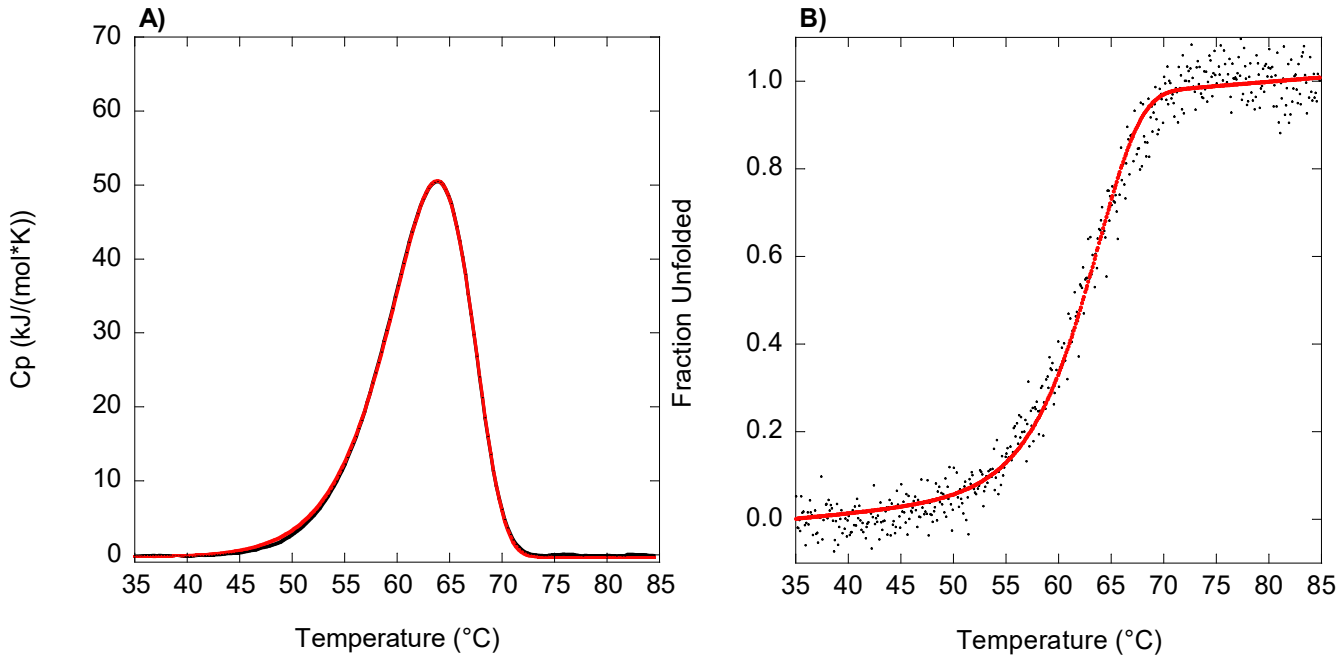
Variant	15 °C (M^{-1})	25 °C (M^{-1})	35 °C (M^{-1})	Denatured 25 °C (M^{-1})
VPR$_{\Delta C}$	2.13 ± 0.14	2.24 ± 0.12	2.25 ± 0.26	15.00 ± 0.30
VPR$_{\Delta C}$/N3P	2.00 ± 0.12	2.32 ± 0.18	2.66 ± 0.19	Not measured
VPR$_{\Delta C}$/I5P	1.96 ± 0.04	2.05 ± 0.13	2.30 ± 0.22	Not measured
VPR$_{\Delta C}$/N238P	2.49 ± 0.17	2.63 ± 0.25	2.75 ± 0.23	Not measured
VPR$_{\Delta C}$/T265P	2.17 ± 0.08	2.41 ± 0.14	2.45 ± 0.25	Not measured
VPR$_{\Delta C}$/N3P/I5P	1.55 ± 0.11	1.64 ± 0.07	1.77 ± 0.02	15.69 ± 0.24
VPR$_{\Delta C}$/N3P/I5P/N238P	2.05 ± 0.06	2.18 ± 0.05	2.41 ± 0.23	Not measured
VPR$_{\Delta C}$/N3P/I5P/T265P	1.94 ± 0.06	2.17 ± 0.05	2.48 ± 0.07	Not measured
VPR$_{\Delta C}$/ N3P/I5P/N238P/T265P	2.05 ± 0.23	2.12 ± 0.19	2.19 ± 0.22	15.67



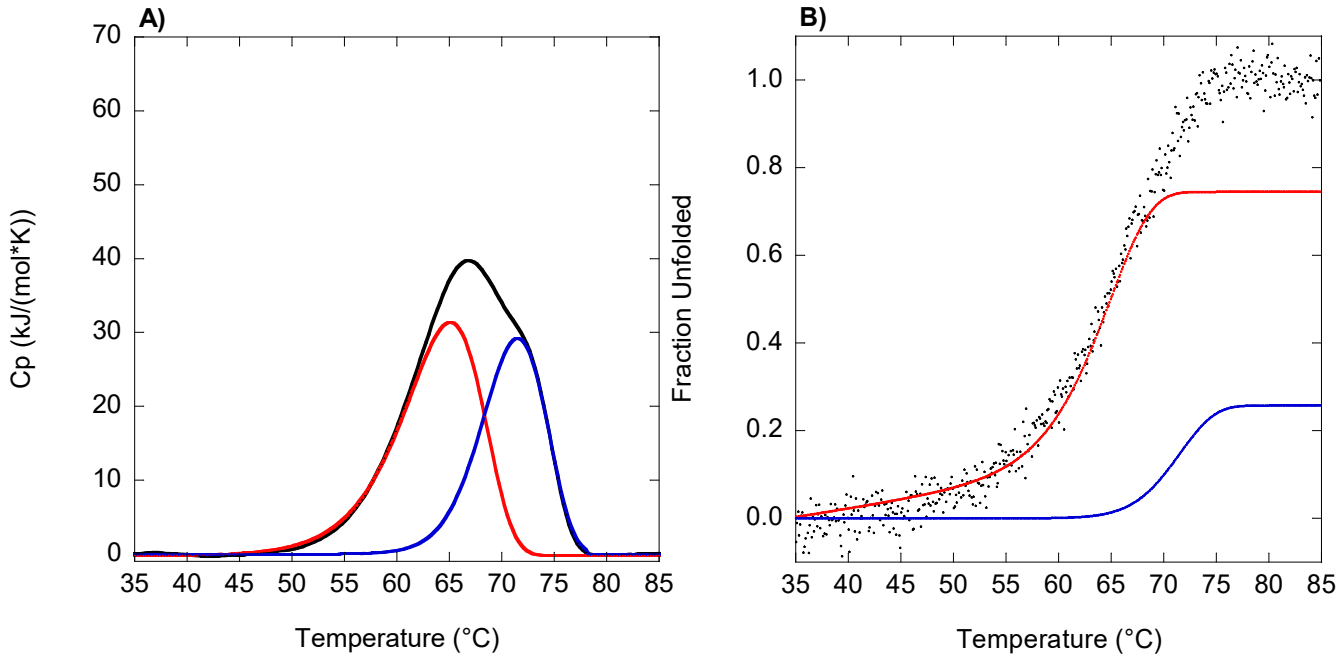
Supplementary Fig. 2. CD wavelength scans of proline variants in glycine buffer containing 15 mM $CaCl_2$ and 100 mM NaCl at 25 °C. $VPR_{\Delta C}$ (black dotted line). **A.** CD wavelength scans of $VPR_{\Delta C}$_N3P (Gold) and $VPR_{\Delta C}$_I5P (purple). **B.** CD wavelength scans of $VPR_{\Delta C}$_N238P (light blue) and $VPR_{\Delta C}$_T265P (orange). **C.** CD wavelength scans of $VPR_{\Delta C}$_N3P/I5P/N238P (blue) and $VPR_{\Delta C}$_N3P/I5P/T265P (green). **D.** CD wavelength scans of $VPR_{\Delta C}$_N3P/I5P (red) and $VPR_{\Delta C}$_N3P/I5P/N238P/T265P (black).



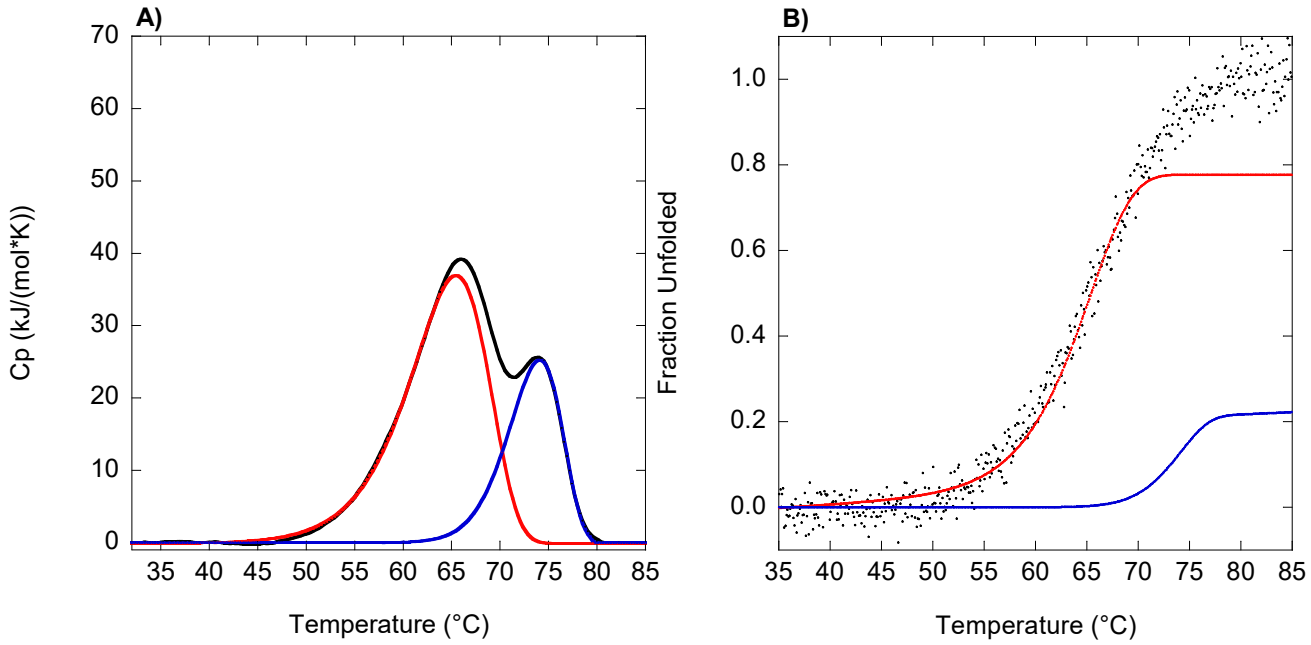
Supplementary Fig. 3. Arrhenius graphs calculated from unfolding rates calculated from DSC thermographs for $VPR_{\Delta C}$ (black dotted line) **A.** $VPR_{\Delta C}$_N3P (Gold) and $VPR_{\Delta C}$_I5P (purple). **B.** $VPR_{\Delta C}$_N238P (light blue) and $VPR_{\Delta C}$_T265P (orange). **C.** $VPR_{\Delta C}$_N3P/I5P/N238P (blue) and $VPR_{\Delta C}$_N3P/I5P/T265P (green). **D.** $VPR_{\Delta C}$_N3P/I5P (red) and $VPR_{\Delta C}$_N3P/I5P/N238P/T265P (black).



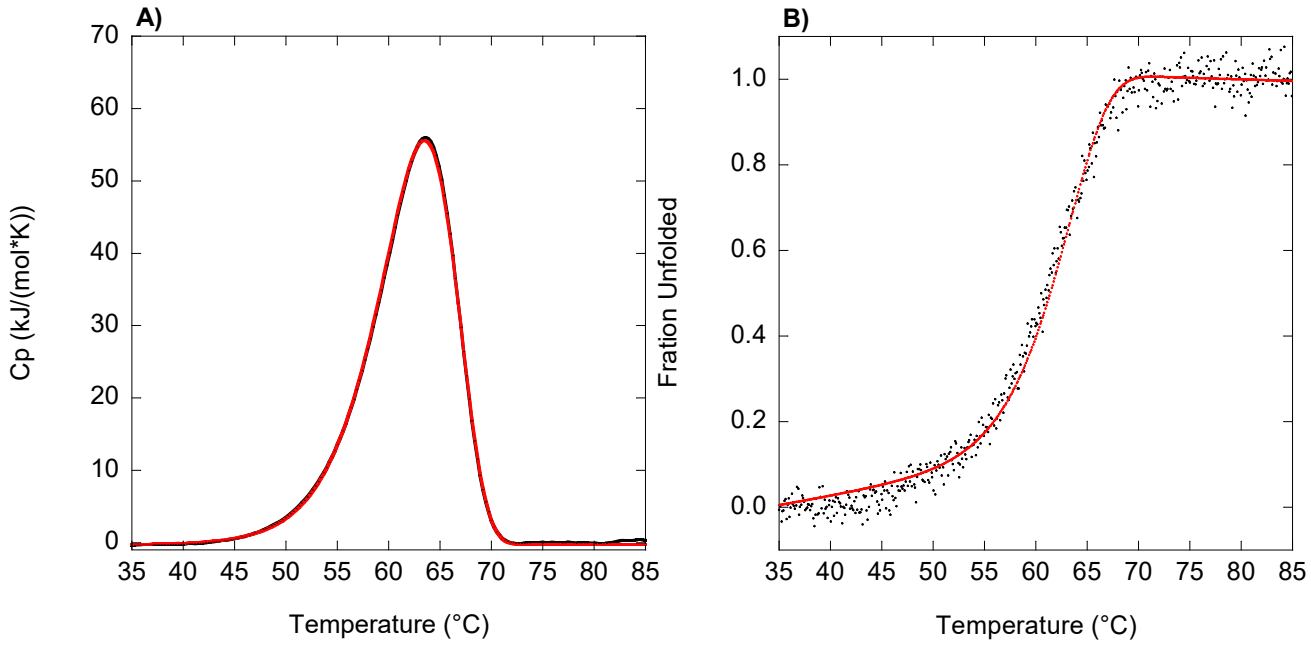
Supplementary Fig. 4. CalFitter 1.2 global fits (red line) of $VPR_{\Delta C}$ normalized DSC thermogram **(A)** (black line) and CD melting profile **(B)** (black dots). Assay conditions: 25 mM glycine, 15 mM $CaCl_2$ and 100 mM NaCl at pH 8.6. Global fit parameters were: E_{act} = 235 ± 2 kJ/mol, T_{act} = 87.4 ± 0.3 °C and $\Delta H_{cal\text{-}fit}$ = 542 ± 5 kJ/mol.



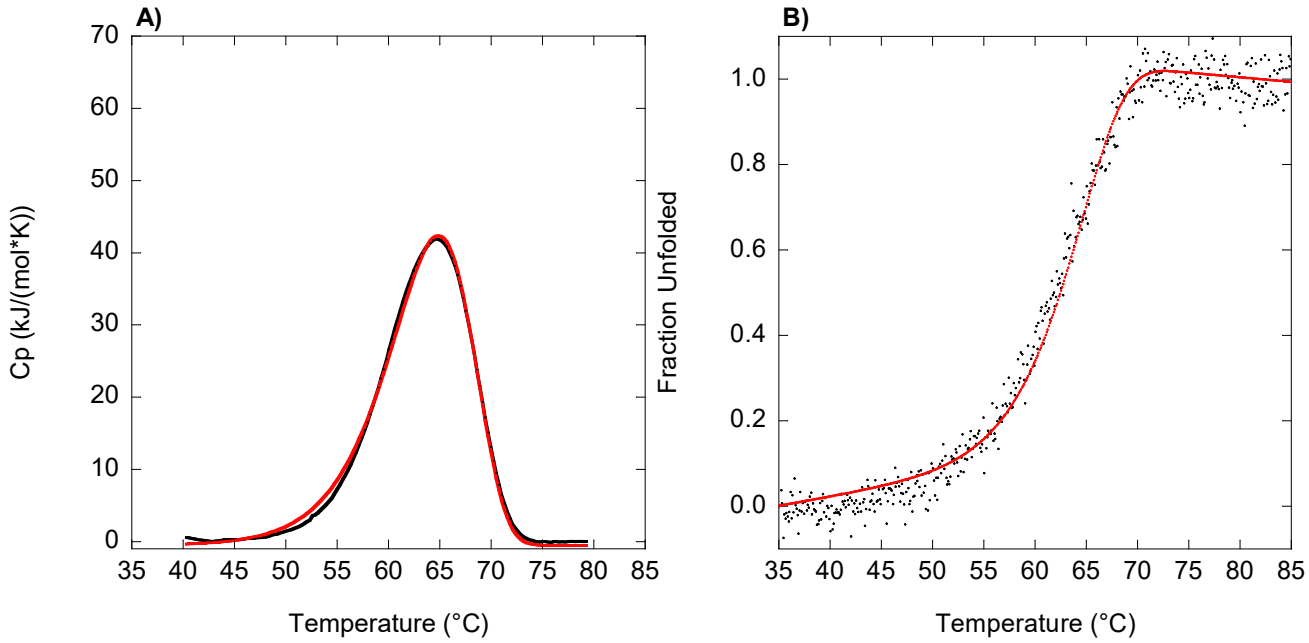
Supplementary Fig. 5. CalFitter 1.2 global fits (red line for transition 1 and blue line for transition 2) of $VPR_{\Delta C}$_N3P normalized DSC thermogram **(A)** (black line) and CD melting profile **(B)** (black dots). Assay conditions: 25 mM glycine, 15 mM $CaCl_2$ and 100 mM NaCl at pH 8.6. Global fit parameters were: E_{act}^1 = 251 ± 5 kJ/mol, E_{act}^2 = 285 ± 20 kJ/mol, T_{act}^1 = 87.0 ± 0.8 °C, T_{act}^2 = 90.8 ± 1.3 °C, $\Delta H_{cal\text{-}fit}^1$ = 318 ± 20 kJ/mol and $\Delta H_{cal\text{-}fit}^2$ = 227 ± 19 kJ/mol.



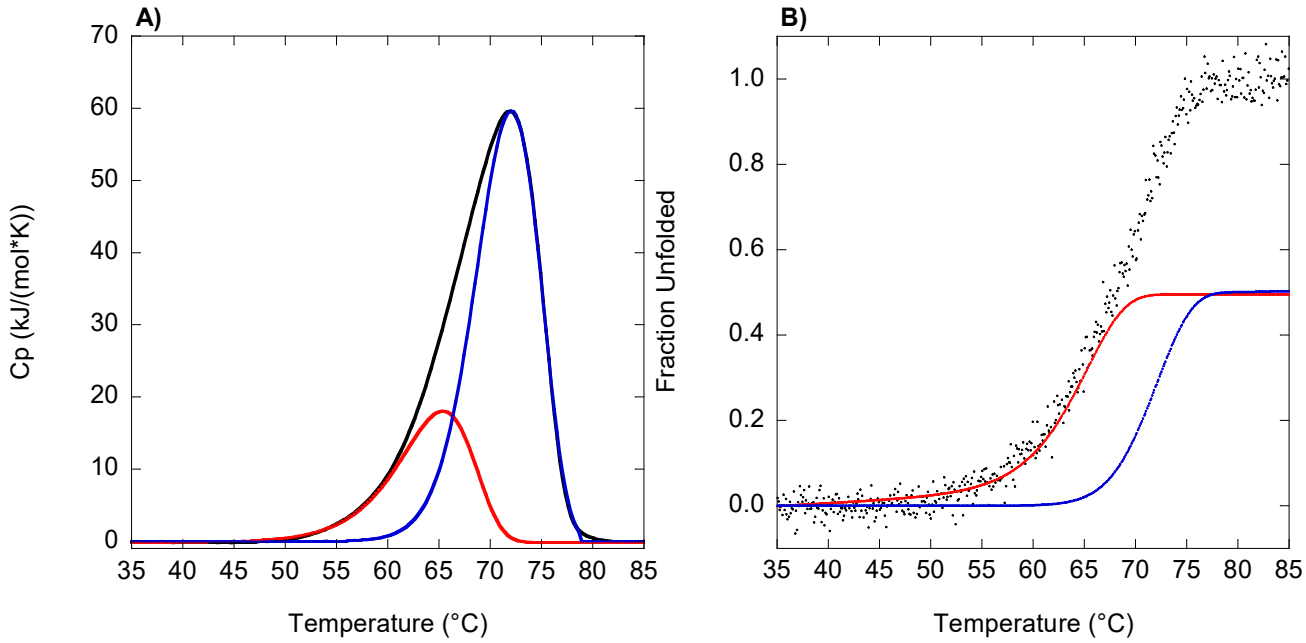
Supplementary Fig. 6. CalFitter 1.2 global fits (red line for transition 1 and blue line for transition 2) of $VPR_{\Delta C}$_I5P normalized DSC thermogram **(A)** (black line) and CD melting profile **(B)** (black dots). Assay conditions: 25 mM glycine, 15 mM $CaCl_2$ and 100 mM NaCl at pH 8.6. Global fit parameters were: E_{act}^1 = 235 ± 3 kJ/mol, E_{act}^2 = 356 ± 13 kJ/mol, T_{act}^1 = 89.3 ± 0.4 °C, T_{act}^2 = 89.2 ± 0.6 °C, $\Delta H_{cal\text{-}fit}^1$ = 399 ± 7 kJ/mol and $\Delta H_{cal\text{-}fit}^2$ = 180 ± 6 kJ/mol.



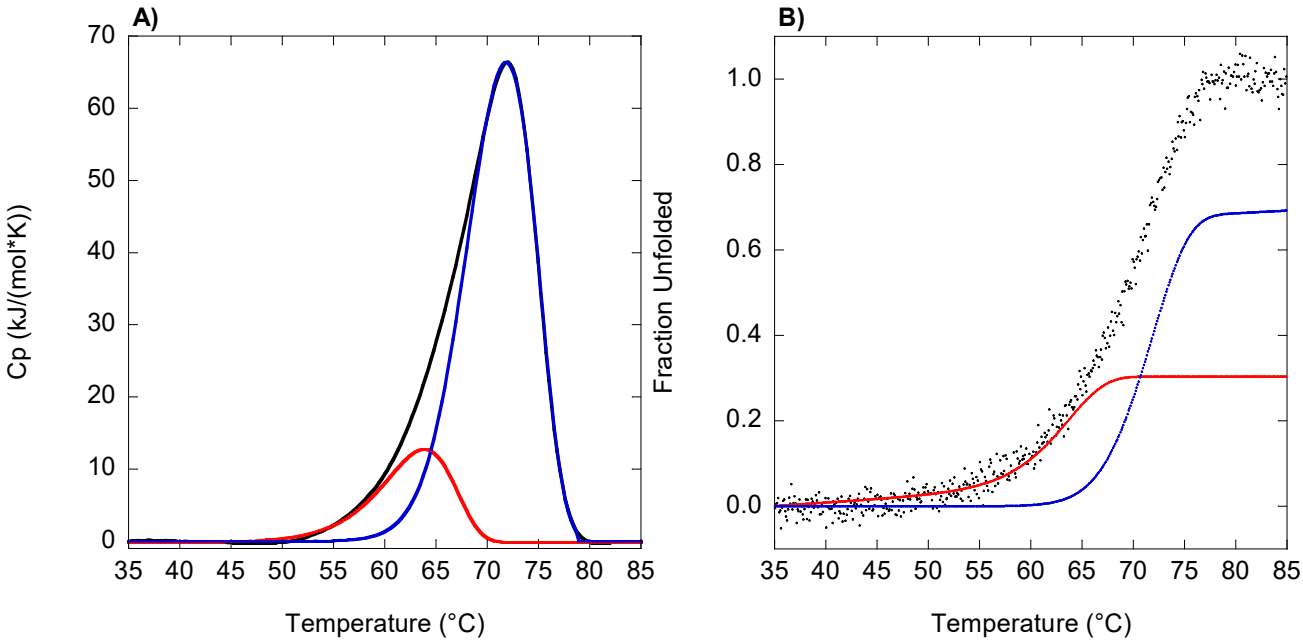
Supplementary Fig. 7. CalFitter 1.2 global fits (red line) of $VPR_{\Delta C}$_N238P normalized DSC thermogram **(A)** (black line) and CD melting profile **(B)** (black dots). Assay conditions: 25 mM glycine, 15 mM $CaCl_2$ and 100 mM NaCl at pH 8.6. Global fit parameters were: E_{act} = 248 ± 2 kJ/mol, T_{act} = 85.5 ± 0.2 °C and $\Delta H_{cal\text{-}fit}$ = 564 ± 4 kJ/mol.



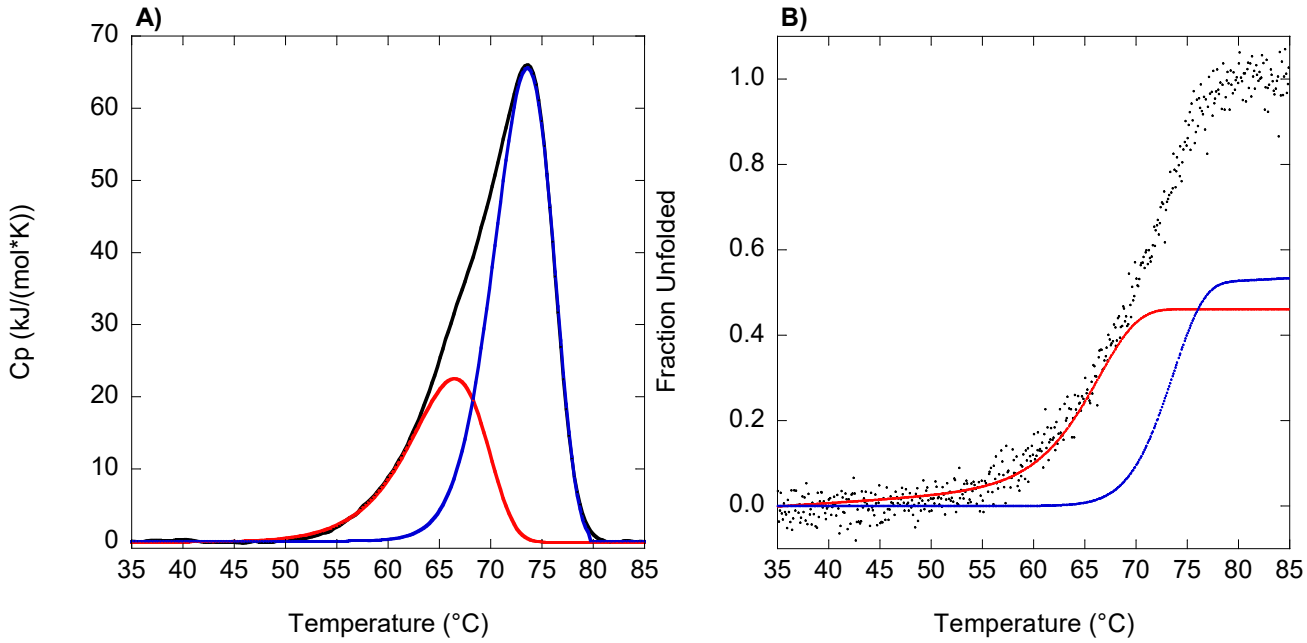
Supplementary Fig. 8. CalFitter 1.2 global fits (red line) of $VPR_{\Delta C}$_T265P normalized DSC thermogram **(A)** (black line) and CD melting profile **(B)** (black dots). Assay conditions: 25 mM glycine, 15 mM $CaCl_2$ and 100 mM NaCl at pH 8.6. Global fit parameters were: E_{act} = 229 ± 2 kJ/mol, T_{act} = 89.4 ± 0.3 °C and $\Delta H_{cal\text{-}fit}$ = 471 ± 5 kJ/mol.



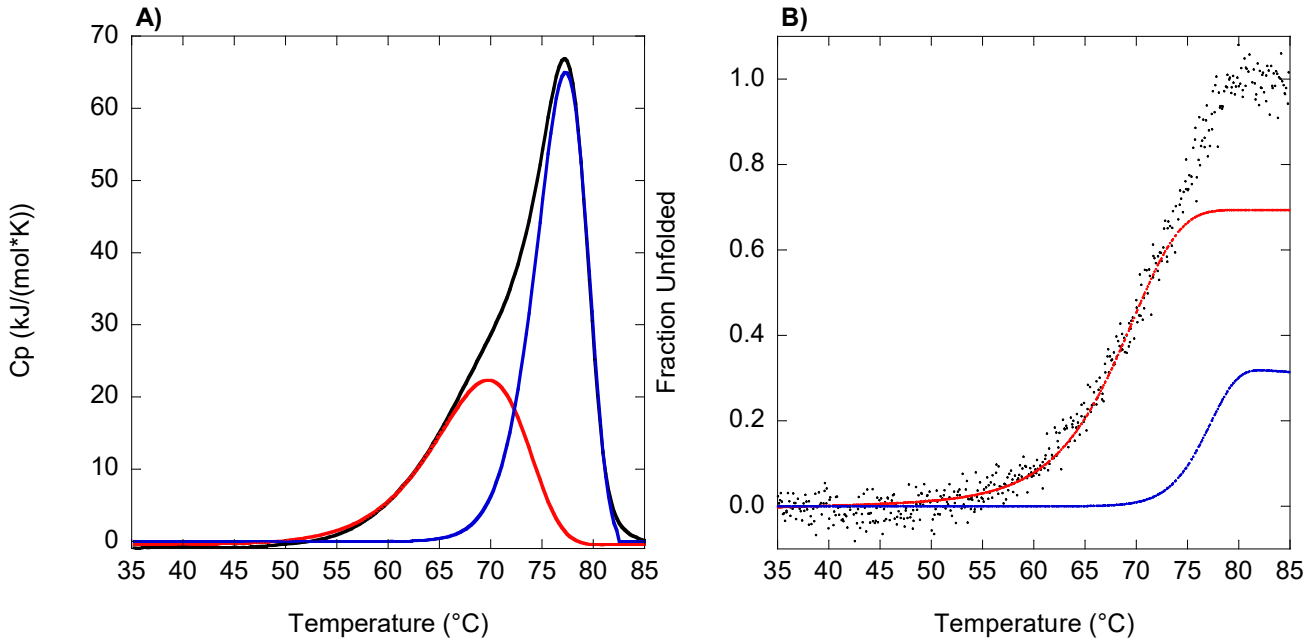
Supplementary Fig. 9. CalFitter 1.2 global fits (red line for transition 1 and blue line for transition 2) of $VPR_{\Delta C}$_N3P/I5P normalized DSC thermogram **(A)** (black line) and CD melting profile **(B)** (black dots). Assay conditions: 25 mM glycine, 15 mM $CaCl_2$ and 100 mM NaCl at pH 8.6. Global fit parameters were: E_{act}^1 = 261 ± 9 kJ/mol, E_{act}^2 = 283 ± 6 kJ/mol, T_{act}^1 = 86.2 ± 1.3 °C, T_{act}^2 = 91.8 ± 0.4 °C, $\Delta H_{cal\text{-}fit}^1$ = 176 ± 20 kJ/mol and $\Delta H_{cal\text{-}fit}^2$ = 477 ± 19 kJ/mol.



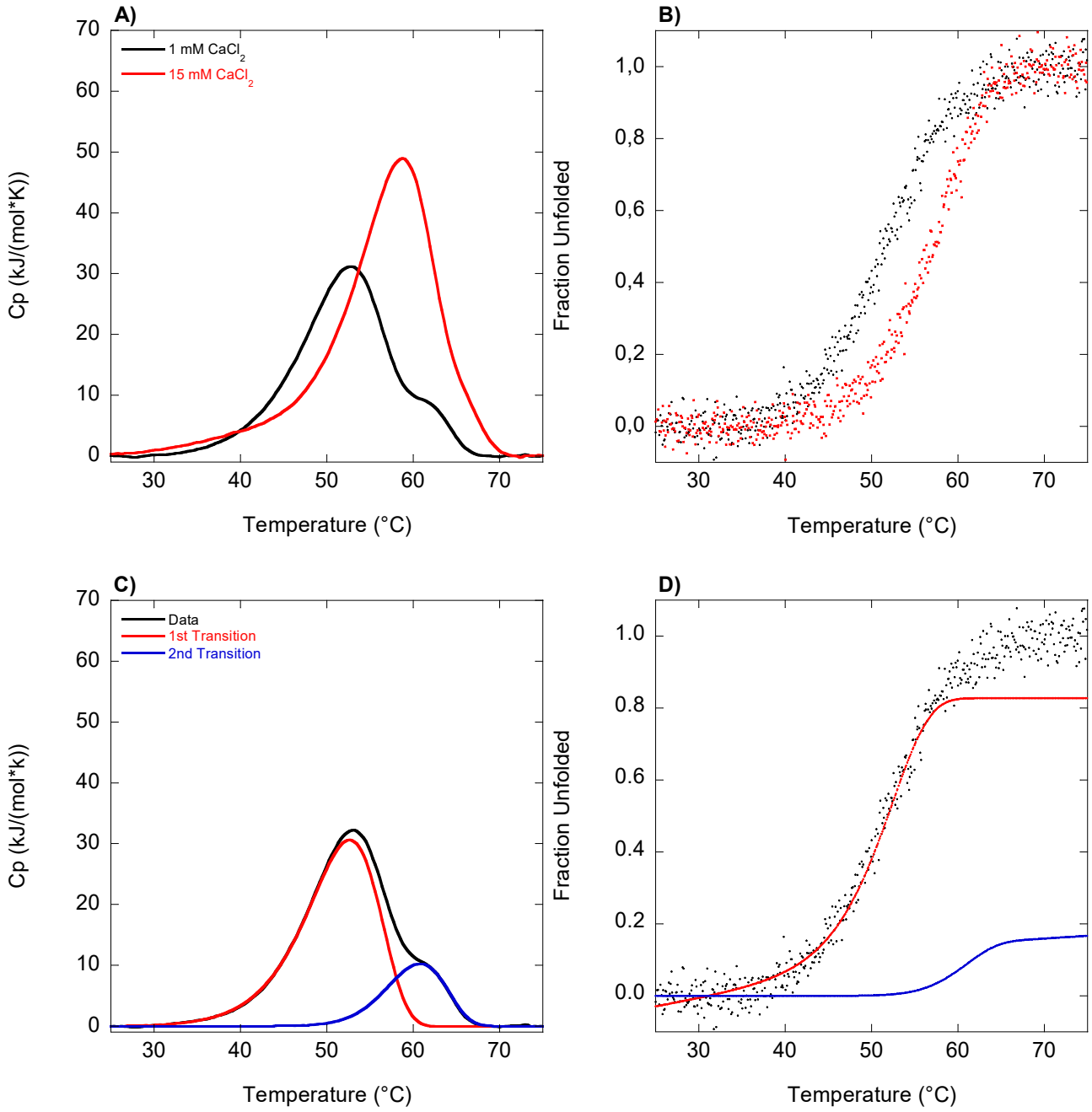
Supplementary Fig. 10. CalFitter 1.2 global fits (red line for transition 1 and blue line for transition 2) of $VPR_{\Delta C}$_N3P/I5P/N238P normalized DSC thermogram **(A)** (black line) and CD melting profile **(B)** (black dots). Assay conditions: 25 mM glycine, 15 mM $CaCl_2$ and 100 mM NaCl at pH 8.6. Global fit parameters were: E_{act}^1 = 270 ± 11 kJ/mol, E_{act}^2 = 279 ± 3 kJ/mol, T_{act}^1 = 83.6 ± 1.4 °C, T_{act}^2 = 92.0 ± 0.3 °C, $\Delta H_{cal\text{-}fit}^1$ = 120 ± 13 kJ/mol and $\Delta H_{cal\text{-}fit}^2$ = 567 ± 13 kJ/mol.



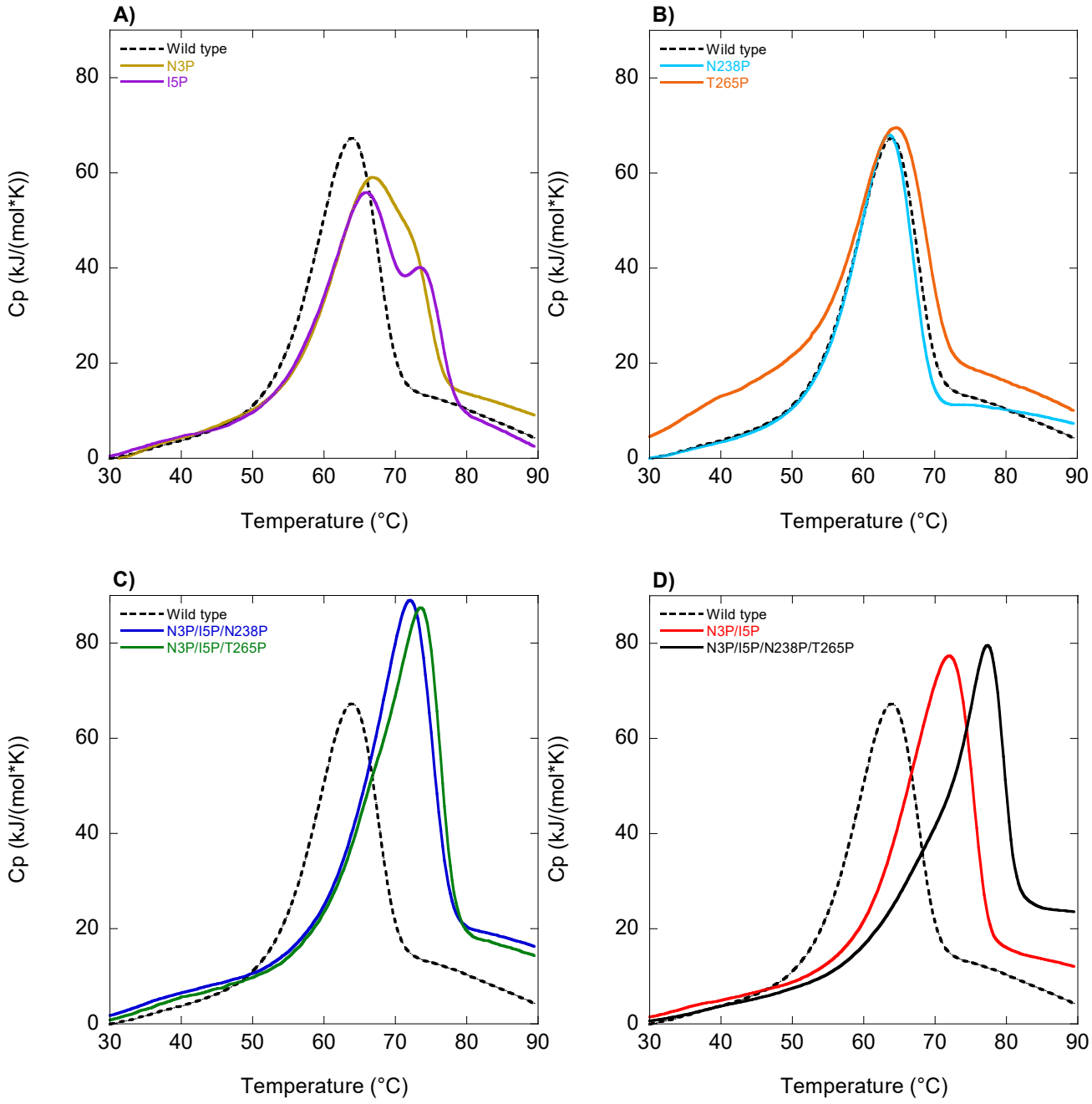
Supplementary Fig. 11. CalFitter 1.2 global fits (red line for transition 1 and blue line for transition 2) of $VPR_{\Delta C}$_N3P/I5P/T265P normalized DSC thermogram **(A)** (black line) and CD melting profile **(B)** (black dots). Assay conditions: 25 mM glycine, 15 mM $CaCl_2$ and 100 mM NaCl at pH 8.6. Global fit parameters were: E_{act}^1 = 259 ± 10 kJ/mol, E_{act}^2 = 326 ± 8 kJ/mol, T_{act}^1 = 87.7 ± 1.4 °C, T_{act}^2 = 90.3 ± 0.4 °C, $\Delta H_{cal\text{-}fit}^1$ = 224 ± 20 kJ/mol and $\Delta H_{cal\text{-}fit}^2$ = 481 ± 18 kJ/mol.



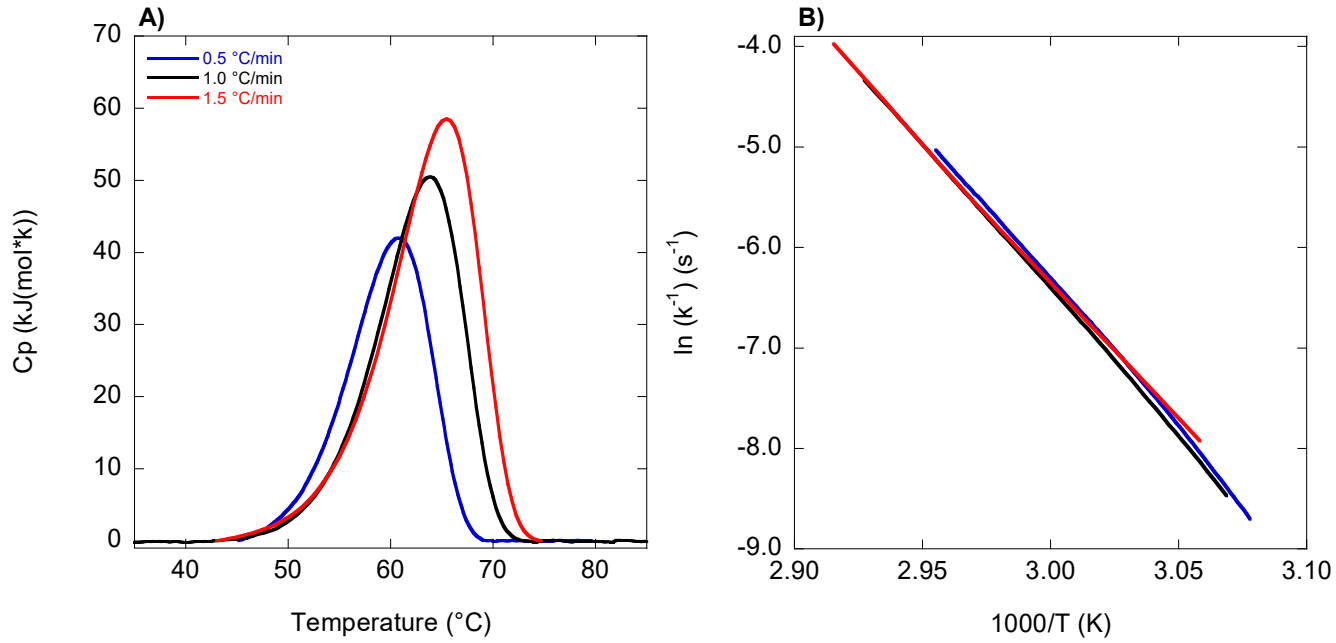
Supplementary Fig. 12. CalFitter 1.2 global fits (red line for transition 1 and blue line for transition 2) of $VPR_{\Delta C}$_N3P/I5P/N238P/T265P normalized DSC thermogram **(A)** (black line) and CD melting profile **(B)** (black dots). Assay conditions: 25 mM glycine, 15 mM $CaCl_2$ and 100 mM NaCl at pH 8.6. Global fit parameters were: E_{act}^1 = 215 ± 8 kJ/mol, E_{act}^2 = 383 ± 7 kJ/mol, T_{act}^1 = 97.3 ± 1.6 °C, T_{act}^2 = 91.3 ± 0.3 °C, $\Delta H_{cal\text{-}fit}^1$ = 275 ± 17 kJ/mol and $\Delta H_{cal\text{-}fit}^2$ = 425 ± 15 kJ/mol.



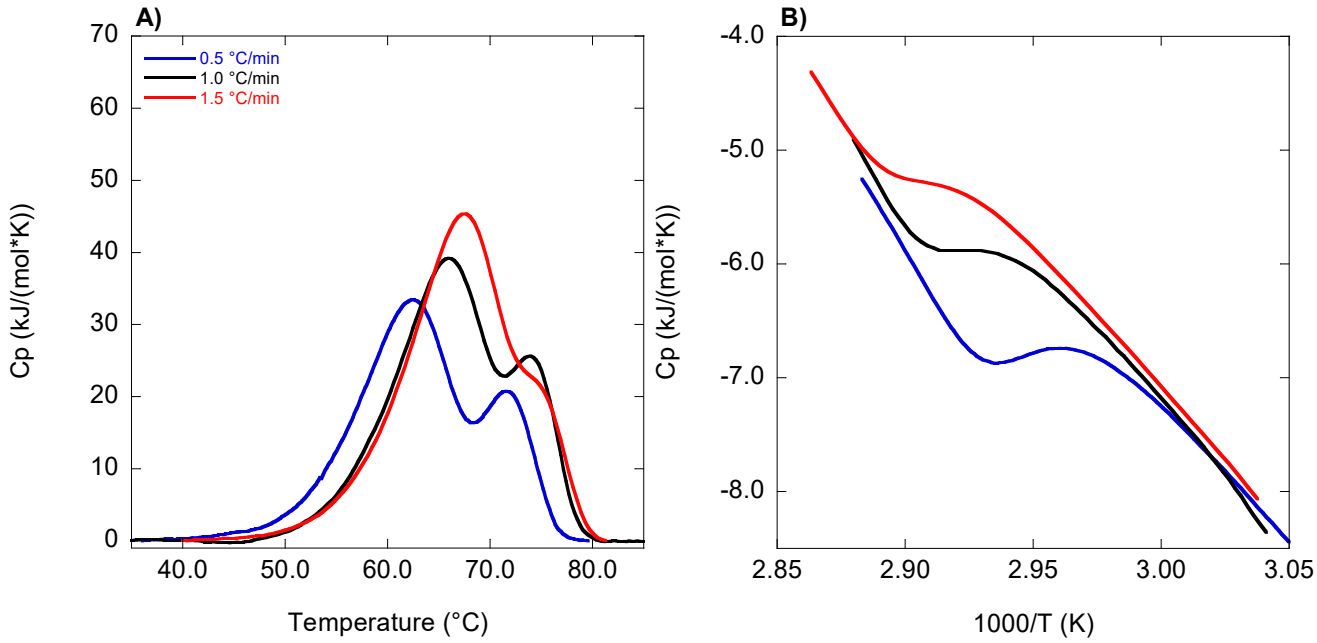
Supplementary Fig. 13. DCS thermogram **(A)** and CD melting profile **(B)** of $VPR_{\Delta C}$ at two different calcium ion concentrations, 1 mM (black line) and 15 mM (red line) at pH 5.0 in a 25 mM acetate buffer and 100 mM NaCl. CalFitter 1.2 global fits (red line for transition 1 and blue line for transition 2) of $VPR_{\Delta C}$ at 1 mM $CaCl_2$ and pH 5.0 of normalized DSC thermogram (black line) **(C)** and normalized CD melting profile (black dots) **(D)** Global fit parameters were: E_{act}^1 = 210 ± 3 kJ/mol, E_{act}^2 = 252 ± 25 kJ/mol, T_{act}^1 = 77.7 ± 0.5 °C, T_{act}^2 = 81.9 ± 2.2 °C, $\Delta H_{cal\text{-}fit}^1$ = 342 ± 9 kJ/mol and $\Delta H_{cal\text{-}fit}^2$ = 89 ± 8 kJ/mol.



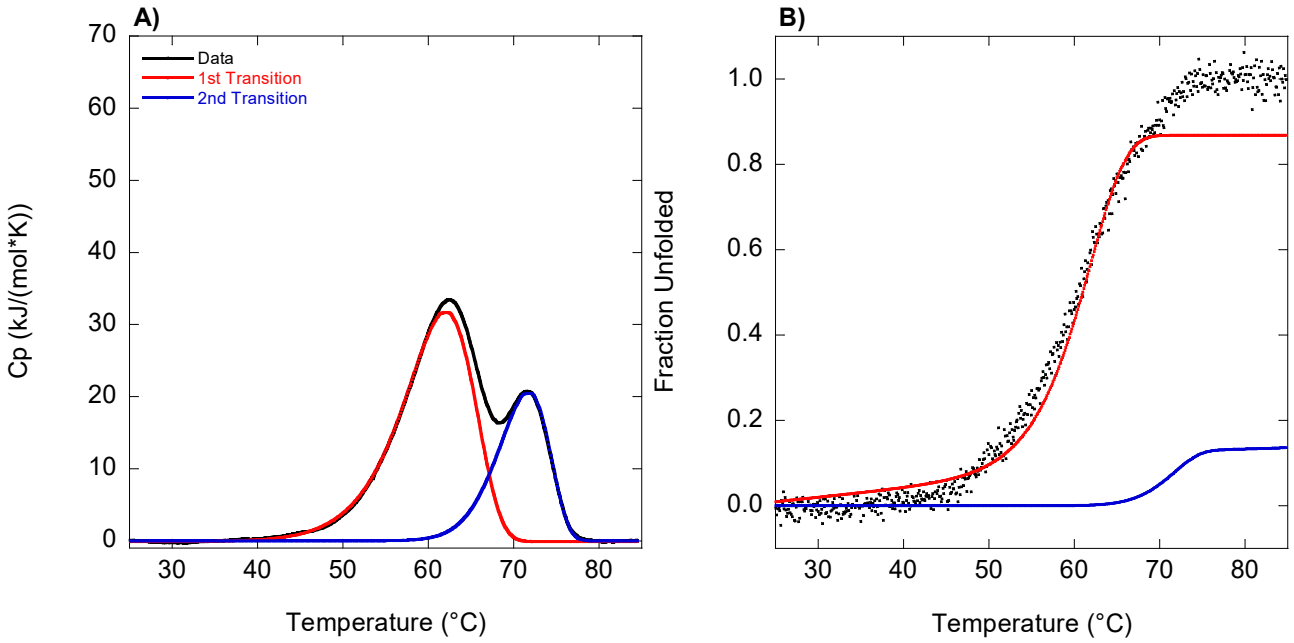
Supplementary Fig. 14. Buffer subtracted differential scanning thermograms showing the excess heat during the unfolding process of the proline variants in a 25 mM glycine buffer containing 15 mM $CaCl_2$ and 100 mM NaCl. $VPR_{\Delta C}$ (dotted black line). **(A)** Unfolding of $VPR_{\Delta C}$_N3P (Gold) and $VPR_{\Delta C}$_I5P (purple). **(B)** Unfolding of $VPR_{\Delta C}$_N238P (light blue) and $VPR_{\Delta C}$_T265P (orange). **(C)** Unfolding of $VPR_{\Delta C}$_N3P/I5P/N238P (blue) and $VPR_{\Delta C}$_N3P/I5P/T265P (green). **(D)** Unfolding of $VPR_{\Delta C}$_N3P/I5P (red) and $VPR_{\Delta C}$_N3P/I5P/N238P/T265P (black).



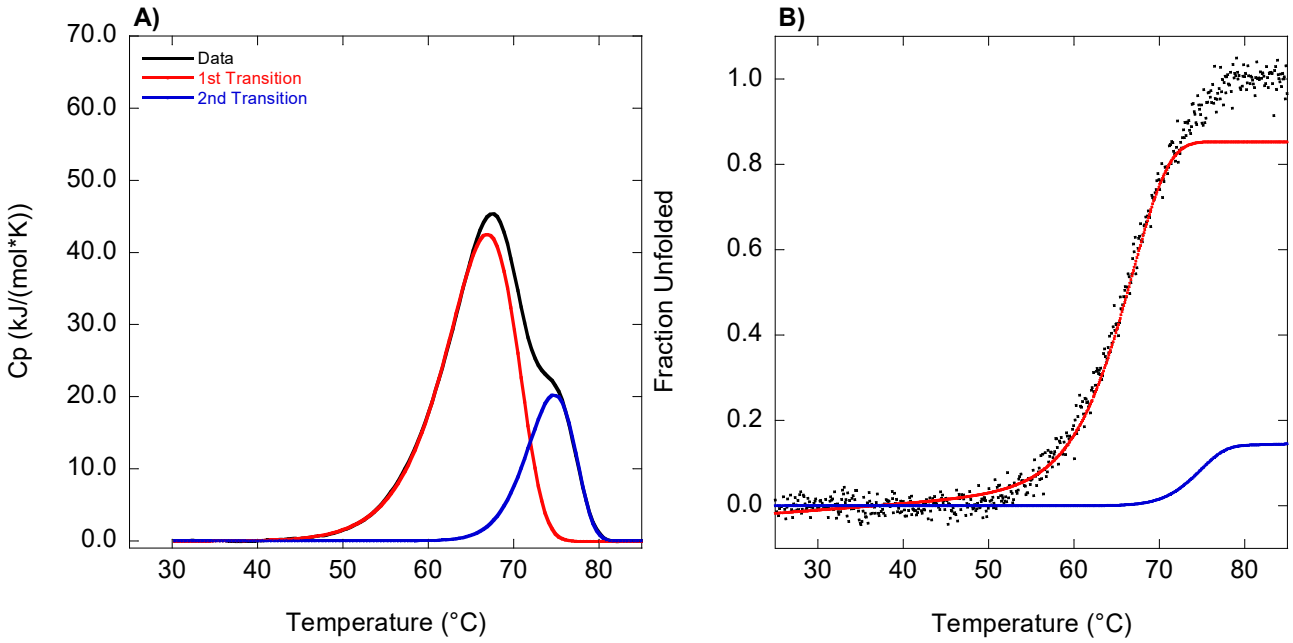
Supplementary Fig. 15. (A) Thermograms of $VPR_{\Delta C}$ in a 25 mM glycine buffer containing 15 mM $CaCl_2$ and100 mM NaCl at three different scan rates, 0.5 °C/min (blue line), 1.0 °C/min (black line) and 1.5 °C/min (red line). **(B)** Arrhenius graph showing the rate of unfolding at three different scan rates calculated from $VPR_{\Delta C}$ thermograms 25 mM glycine buffer containing 15 mM $CaCl_2$ and100 mM NaCl, 0.5 °C/min (blue line), 1.0 °C/min (black line) and 1.5 °C/min (red line).



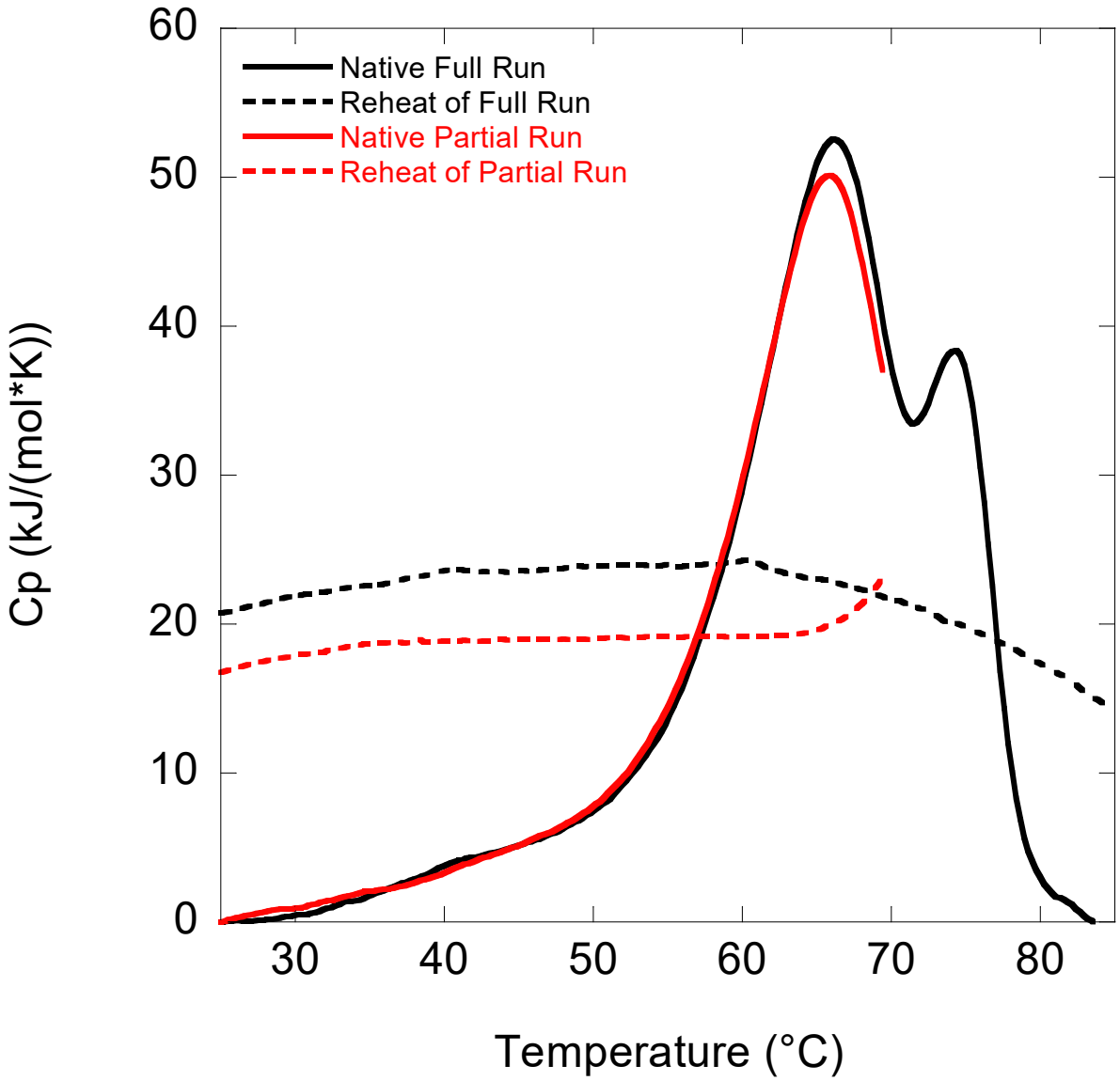
Supplementary Fig. 16. (A) Thermograms of $VPR_{\Delta C}$_I5P in a 25 mM glycine buffer containing 15 mM $CaCl_2$ and100 mM NaCl at three different scan rates, 0.5 °C/min (blue line), 1.0 °C/min (black line) and 1.5 °C/min (red line). **(B)** Arrhenius graph showing the rate of unfolding at three different scan rates calculated from $VPR_{\Delta C}$_I5P thermograms 25 mM glycine buffer containing 15 mM $CaCl_2$ and100 mM NaCl, 0.5 °C/min (blue line), 1.0 °C/min (black line) and 1.5 °C/min (red line).



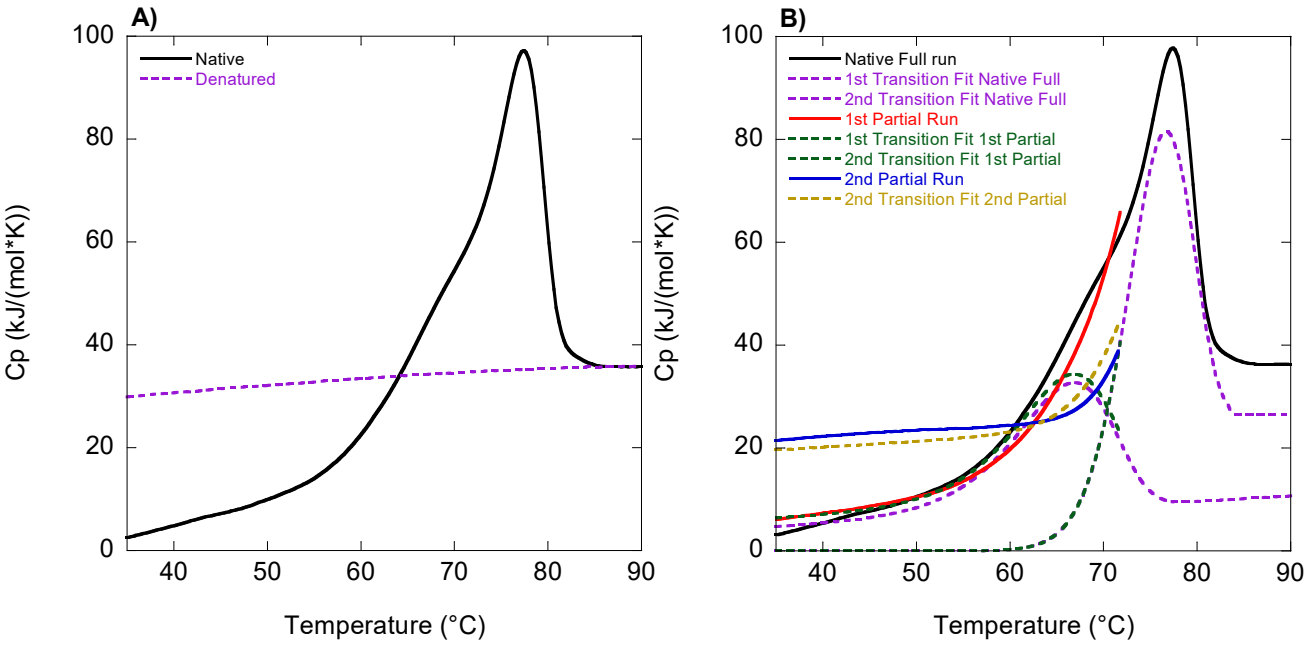
Supplementary Fig. 17. CalFitter 1.2 global fits (red line for transition 1 and blue line for transition 2) of $VPR_{\Delta C}$_I5P normalized DSC thermogram **(A)** (black line) and CD melting profile **(B)** (black dots). Assay conditions: 0.5 °C/min, 25 mM glycine, 15 mM $CaCl_2$ and 100 mM NaCl at pH 8.6. Global fit parameters were: E_{act}^1 = 227 ± 2 kJ/mol, E_{act}^2 = 330 ± 7 kJ/mol, T_{act}^1 = 89.6 ± 0.3 °C, T_{act}^2 = 90.3 ± 0.5 °C, $\Delta H_{cal\text{-}fit}^1$ = 348 ± 3 kJ/mol and $\Delta H_{cal\text{-}fit}^2$ = 158 ± 3 kJ/mol.



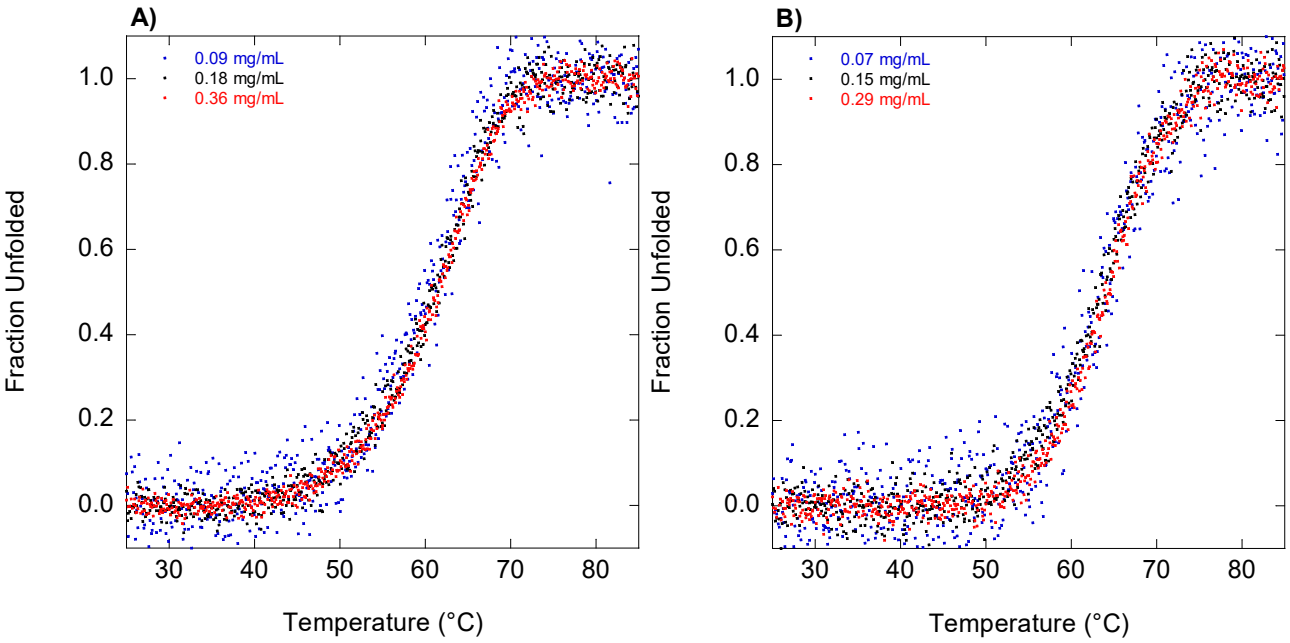
Supplementary Fig. 18. CalFitter 1.2 global fits (red line for transition 1 and blue line for transition 2) of $VPR_{\Delta C}$_I5P normalized DSC thermogram **(A)** (black line) and CD melting profile **(B)** (black dots). Assay conditions: 1.5 °C/min, 25 mM glycine, 15 mM $CaCl_2$ and 100 mM NaCl at pH 8.6. Global fit parameters were: $E_{act}^1 = 229 \pm 2$ kJ/mol, $E_{act}^2 = 338 \pm 16$ kJ/mol, $T_{act}^1 = 89.9 \pm 0.4$ °C, $T_{act}^2 = 89.5 \pm 0.7$ °C, $\Delta H_{cal\text{-}fit}^1 = 475 \pm 8$ kJ/mol and $\Delta H_{cal\text{-}fit}^2 = 147 \pm 7$ kJ/mol.



Supplementary Fig. 19. Complete (black line) and partial (red line) unfolding of $VPR_{\Delta C}$_I5P followed by a reheating of complete unfolding (black dotted line) and reheating of partial unfolding (red dotted line). Assay conditions: 1.0 °C/min, 25 mM glycine, 15 mM $CaCl_2$ and 100 mM NaCl at pH 8.6.



Supplementary Fig. 20. A) Complete unfolding of $VPR_{\Delta C}$_N3P/I5P/N238P/T265P (black line) followed by reheating of the resulting denatured assembly (purple dotted line) confirming the irreversibility of the full unfolding process. **B)** Partial unfolding of $VPR_{\Delta C}$_N3P/I5P/N238P/T265P fitted using CalFitter 1.2. CalFitter analysis was carried out using the reheat feature in the program. Input data consisted of three thermograms: total denaturation of the protein (black solid line and fitted curves purple dotted lines), partial unfolding of the protein by heating a native sample to 72 °C (red solid line and fitted curves green dotted lines) and reheating of that partially unfolded sample to 72 °C (blue solid line and the fitted curve golden dotted line). Prior to fitting via CalFitter samples were subjected to thermocycle baseline subtraction. The best fitting model was a three-state model with two irreversible transitions. Assay conditions were 25 mM glycine, 15 mM $CaCl_2$ and 100 mM NaCl at pH 8.6.



Supplementary Fig. 21. Protein concentration effects on melting points measured on CD of $VPR_{\Delta C}$ **(A)** and $VPR_{\Delta C}$_I5P **(B)**. The three protein concentrations tested were approximately 0.1 mg/mL (blue dots), 0.2 mg/mL (black dots) and 0.4 mg/mL (red dots). Assay conditions were 25 mM glycine, 15 mM $CaCl_2$ and 100 mM NaCl at pH 8.6.

References.

1. Mazurenko, S. *et al.* CalFitter: a web server for analysis of protein thermal denaturation data. *Nucleic Acids Res* **46**, W344-W349, doi:10.1093/nar/gky358 (2018).
2. Jaswal, S. S., Truhlar, S. M. E., Dill, K. A. & Agard, D. A. Comprehensive analysis of protein folding activation thermodynamics reveals a universal behavior violated by kinetically stable proteases. *J. Mol. Biol.* **347**, 355-366, doi:10.1016/j.jmb.2005.01.032 (2005).

Paper III

N-terminal mutation of a conserved Tryptophan residue causes structural changes and decreases the kinetic stability of VPR, a cold adapted subtilase.

Kristinn R. Óskarsson[1], Matteo Lambrughi[2], Elena Papaleo[2] & Magnús M. Kristjánsson[1]*

1. Department of Biochemistry, Science Institute, University of Iceland, Reykjavík, Iceland.
2. Computational Biology Laboratory, Unit of Statistics, Bioinformatics and Registry, Danish Cancer Society Research Center, Strandboulevarden 49, 2100 Copenhagen, Denmark

* Corresponding author: Magnús M. Kristjánsson, E-mail: mmk@hi.is

Abstract

Kinetic stabilization seems to be especially prevalent among proteases from different clans, to combat harsh environments. These enzymes are produced as pro-proteins with intramolecular chaperones, facilitating correct folding followed by autoproteolysis, leaving the active folded protease domain. The folded domains are often highly structured proteins with the the active state locked in a kinetic trap. In this paper we demonstrate the effects of substituting a conserved N-terminal Trp residue to Phe in the cold adapted subtilase VPR, belonging to the proteinase K subfamily. Results show that this Trp residue plays a pivotal role in the kinetic stability of VPR. Replacing the N-terminal Trp residue to Phe caused a drastic reduction in resistance against thermal denaturation, proteolysis and diminished unfolding cooperativity. This is a result of the collapse of certain internal protein interactions, facilitated by the Trp residue. The Trp-Phe mutation leading to a less structured protein as indicated by circular dichroism, fluorescence spectroscopy and MD-simulations. These results demonstrate that the kinetic stability of VPR might be in large part facilitated by few key residues that maintain the “lock” that keeps the structure in its kinetic trap.

1. Introduction

Protein stability can be described in general terms by having two major facets, a thermodynamic one and a kinetic one. The thermodynamic part describes the equilibrium between the native state and the unfolded states along with all intermediates on that pathway. The kinetic part describes the rate of which these states can sample, thus describing the energy barriers separating the states [1]. These two facets do however not contribute evenly to the prevalence of the native active form of proteins. Thus, proteins can be roughly divided into kinetically stable and thermodynamically stable structures. Thermodynamically stable proteins are proteins that fold to their native state under "physiological" conditions without any major kinetic barriers. Kinetically stable proteins on the other hand can undergo an irreversible transition trapping them in a nonfunctional state. The purest examples of structures that are under kinetic control, might be many extracellular proteinases [1, 2]. These proteins are expressed with intramolecular chaperones (IMC) that are essential for folding and when folding has finished the protease cleaves the IMC off and degrades it, leaving the active structure [1, 3-5]. These IMCs are often quite sizeable, reaching 77 residues in the case of subtilisin E and 174 residues in the case of α-lytic protease where the IMC is almost as large as the active enzyme (198 residues) [3, 6, 7]. This amount of resources spent on the production of these IMCs is a marker of the importance of these kinetically stable structures for their ability to carry out their biological function, which is the hydrolysis of peptide bonds in the harsh extracellular environment providing their bacterial originators with nutrients. As a result the native structure of these kinetically stable proteases tend to be extremely compact and rigid structures that unfold in a highly cooperative manner with large $\Delta G^{\ddagger}$ and $\Delta C_p^{\ddagger}$ compared to structural homologs that are not under kinetic control [8-10]. In addition, those native states are often just marginally, or not at all thermodynamically stable compared to their unfolded states and are only able to carry out their biological function due to the enormous energy barrier separating the two states [1, 4, 11]. In this paper we will report on changes in the kinetic stability of $VPR_{\Delta C}$ [12] accompanied by a single Trp-Phe mutation located on the N-terminal region of the enzyme. The effects were probed utilizing molecular dynamics (MD), differential scanning colorimetry (DSC), circular dichroism spectroscopy (CD), steady state fluorescence, fluorescence quenching, thermal inactivation assays ($T_{50\%}$) and Michalis-Menten activity assays. $VPR_{\Delta C}$ is a cold adapted subtilisin-like serine proteinase of the proteinase K family whose gene was isolated from the Gram-negative bacterium *Vibrio* sp. PA-44 [13, 14]. The C-terminal truncated form of VPR, $VPR_{\Delta C}$ is produced as a pre-pro protein containing a 139 residue IMC and a 276

residue catalytic domain [15, 16]. The mature active form of $VPR_{\Delta C}$ contains three calcium ions that have been shown to be extremely dependent on calcium ion concentration for its stability both as protection against proteolytic degradation as well as the stability of its overall structure [15]. One of these calcium ion binding sites (Ca3) is in a loop on the N-terminal part of the protein, that likely is formed after the cleavage of the loop connecting the IMC domain and the catalytic domain, locking the structure in its kinetic trap. The N-terminal region has been shown to be highly important part of the protein structure and mutation studies, where Pro residues were introduced in order to rigidify and stabilize it, have been shown to be successful [17]. However, located near to that calcium binding site is a highly conserved Trp residue that is found in a broad array of related subtilases. In the structure of $VPR_{\Delta C}$_Trp6 is tucked in a pocket located between the N-terminal calcium binding loop and the major bulk of the enzyme, thus having contact to many distant parts of the protein. It was thus of interest to probe its effects on the stability of $VPR_{\Delta C}$ as its conservation, location and orientation might indicate an important role in the kinetic stability of VPR. The resulting variant $VPR_{\Delta C}$_W6F showed evidence of drastic changes both in activity, stability and ease of expression, where it was found to be much more dependent on high calcium ion concentrations in the expression cultivation. The variant showed drastic loss in stability, with a 12.6 °C lower melting point and 13.1 °C lower $T_{50\%}$ along with the emergence of an unfolding intermediate that was observable in DSC thermograms. Fluorescence and CD wavelength scans indicated strongly that major changes to the structure of the enzyme occurred as a result of this mutation, with indication of a reduced α-helical content in the structure. Those observations were supported by MD simulations that showed unraveling of distant helical structures due to the breakdown of an H-bond network and contacts that were reliant on the presence of that Trp residue. The mutation seemingly does not disrupt the N-terminus itself in such a drastic way. MD simulations show a Phe rotamer that is much more solvent exposed pointing out of the pocket that Trp residue resides in in the wild type. The effects the mutation had on the activity parameters suggested stronger substrate binding, but slower turnover numbers using Suc-AAPF-NH-Np as a substrate at 25 °C. However, using Arrhenius graphs to calculate the thermodynamic parameters of activation for catalysis indicated that the W6F variant had lower activation enthalpy and a more negative value for the activation entropy. These results that could indicate a more flexible/softer protein surface, suggesting that the W6F variant might be more active at very low temperatures [18].

2. Materials and Methods

2.1 Site directed mutagenesis, expression and purification

The W6F mutation was introduced into the gene of $VPR_{\Delta C}$, contained in a pET-11-a-d vector [15], via Quick-change utilizing Phusion® High-Fidelity DNA polymerase from NEB, following their PCR protocols. Primers used to introduce the mutation were acquired from TAG Copenhagen, having the sequence 5'-CAAAGCAACGCGATTT**TT**GGGCTAGACCG-3' for the forward primer and 5'-CGGTCTAGCCC**AA**AAATCGCGTTGCTTTG-3' for the reverse. Following the PCR protocol, DpnI (Thermo Scientific) was added to a final concentration of 1 U/mL and incubated at 37 °C overnight. For plasmid amplification, samples were transformed into XL10-Gold cells (Agilent Technologies) following the producer´s protocol. Plasmid purification was carried out using the Monarch plasmid miniprep kit (NEB), following their instructions. Samples were then sent to Genewiz® for verification via Sanger sequencing. Expression of the mutant was carried out in Lemo21 cells (NEB) [19], transformation and cultivation protocol were the same as described in [15], except for the addition of 100 mM $CaCl_2$ during expression, which was needed for successful expression of the variant. Purification was as described in [15].

2.2 Activity assays

All activity assays were performed in 100 mM Tris, 10 mM $CaCl_2$ at pH 8.6 using Suc-AAPF-NH-Np as a substrate. Kinetic parameters for the wild type and the W6F variant were characterized by Michaelis-Menten assay monitoring activity at varying temperatures against Suc-AAPF-NH-Np at different substrate concentrations, up to 1.00 mM, and monitoring ΔA_{410} over 15 seconds. Enzyme samples were dialyzed against the assay buffer overnight at 4 °C and concentration was estimated by A_{280} measurements using the calculated molar attenuation coefficient 34,170 $M^{-1}cm^{-1}$ for the wild type and 28,670 $M^{-1}cm^{-1}$ for the W6F variant [20]. Exact substrate concentrations were determined at 410 nm, using the molar attenuation coefficient 8,480 $M^{-1}cm^{-1}$ [21]. Data points were then fitted to the Michaelis-Menten equation using the analysis software KaleidaGraph (Synergy Software).

2.3 Fluorescence

Steady state fluorescence was recorded at 25°C on a Fluoromax-4 spectrofluorometer (Horiba Scientific) equipped with a circulating water bath for temperature control. All samples were inhibited by PMSF to a final concentration of 2.5 mM followed by dialysis against 50 mM Tris,

10 mM $CaCl_2$ and pH 8.0 overnight at 4°C. Prior to fluorescence experiments absorbance spectra were recorded from 400 nm down to 220 nm and absorbance adjusted to 0.03 - 0.05 A.U. at 295 nm in a 0.4 cm quartz cuvette (Spectrocell) used for fluorescence experiments. In addition to recording native fluorescence, steady state fluorescence of the denatured state was also recorded, where samples were heated to 90°C for 15 minutes and fluorescence measured at 25°C. All samples were excited at 295 nm using 3 nm entrance slit width and fluorescence monitored between 310 nm and 450 nm using a 5-8 nm exit slit width for native samples and 2-3 nm for denatured samples. Relative fluorescence was then calculated as:

$$F_{n.} = \frac{\left(\frac{CPS}{[P] * exi.^2}\right)}{F_{\mathrm{VPR\Delta C}}}$$

where $F_{n.}$ is the normalized fluorescence intensity, CPS the recorded fluorescence intensity, [P] the protein concentration, exi. the exit slit width used and $F_{VPR\Delta C}$ the concentration and exit slit width normalized fluorescence for native $VPR_{\Delta C}$. The peak of each fluorescence spectra was then fitted to a cubic function, solving the first derivative for the local maximum (λ_{max}). AUC (area under curve) was calculated for all variants via the trapezoidal rule and the relative emission efficacy calculated by dividing the results with the average fluorescence intensity for native $VPR_{\Delta C}$. In addition, acrylamide quenching was conducted on all variants, using a 2.5 M stock of molecular biology grade acrylamide (Sigma). Sample preparation and experimental conditions were as described above. Each aliquot of acrylamide added to samples was followed by thorough mixing and one min resting time for temperature equilibration. The effectiveness of quenching was calculated by fitting the data with the Stern-Volmer equation:

$$\frac{\mathrm{F^o}}{F} = 1 + K_{SV}[Q]$$

where F^0 and F are the fluorescence intensities in the absence and presence of quencher between 310 nm and 410 nm, [Q] is concentration of quencher and K_{sv} is the Stern-Volmer constant calculated via linear regression. Corrections of fluorescence intensities were performed on the data to account for dilutions due to additions of acrylamide.

2.4 Thermal stability

Prior to thermal inactivation experiments, samples were dialyzed against a 25 mM Tris buffer containing 15 mM $CaCl_2$, 100 mM NaCl, 1 mM EDTA and at pH 8.95 overnight at 4 °C. Samples were then heated to selected temperatures and aliquots withdrawn at timed intervals

for assaying remaining activity using 0.5 mM Suc-AAPF-NH-Np. The observed first order rate constants were then used to construct Arrhenius-plots that were analysed by linear regression using KaleidaGraph, from which the $T_{50\%}$ (the temperature where half of the activity was lost after thirty minutes) and $E_{act\ (inactivation)}$ (corresponding to the slope of the Arrhenius graph) values were then calculated.

Unfolding of the secondary structure was monitored by circular dichroism (CD). Prior to measurements samples were inhibited by PMSF at a final concentration of 2.5 mM followed by dialysis against a 25 mM glycine buffer containing 100 mM NaCl and 15 mM $CaCl_2$ at pH 8.6 overnight at 4 °C. Melting curves of protein samples (0.1 – 0.4 mg/mL) were recorded at 222 nm with a heating rate of 1°C/min from 25 °C to 90°C on a Jasco J-1100 CD spectrometer. Data analysis and $T_{m\ (CD)}$ determination was performed as described in [17]. Concurrent CD wavelength scans were also recorded on a Jasco J-1100 from 250 nm down to 200 nm at 25°C using a 1 mm cuvette.

Differential scanning calorimetry (DSC) was used to record thermograms of the unfolding process using a MicroCal VP-DSC. Prior to measurements, samples were inhibited by PMSF at a final concentration of 2.5 mM followed by dialysis against a 25 mM glycine buffer, containing 100 mM NaCl and 15 mM $CaCl_2$ and pH 8.6, overnight at 4 °C. Prior to loading, protein samples (0.3 – 1.2 mg/mL) and buffers were degassed for 15 - 30 min at 10 °C. Thermograms were then recorded from 5 °C until fully unfolded with a temperature gradient of 1 °C/min. Buffer subtraction and concentration normalization was carried out using the Origin software. Reheat runs of unfolded samples were carried out confirming irreversibility [22]. Prior to data analysis DSC data sets were normalized by generating baselines to convert data sets into plots of excess heat capacity versus temperature using the Origin software. The apparent melting points ($T_{m\ (DSC)}$) were found by fitting the peaks with a cubic function and solving the first derivative for the local maximum. Data fitting was carried out using the online tool CalFitter 1.3 (https://loschmidt.chemi.muni.cz/calfitter/) [23]. To fit the wild type DSC data a two-state irreversible unfolding model was used:

$$N \overset{k_1}{\rightarrow} D$$

where N stands for the native state, D for the denatured one and k_1 for the first order rate constant of unfolding that follows the Arrhenius equation [24]. In order to fit the data for the W6F variant a three-state model was used:

$$N \overset{k_1}{\rightarrow} I \overset{k_2}{\rightarrow} D$$

where N stands for the native state, I stands for an intermediate state, D for the denatured state, k_1 stands for the first order rate constant of the first transition and k2 for the first order rate constant of the second transition as has been done for other variants of VPR exhibiting a second transition [17]. Prior to fitting, the average of at least three separate excess heat DSC thermograms and the average of at least three separate normalized CD melting profiles were subjected to global fitting on the CalFitter server. As the data sets were deconvoluted and normalized $\Delta C_p^{\ddagger}$ was set as a fixed value of 0 kJ/mol. The model which CalFitter uses for fitting irreversible transitions is a modification of the Arrhenius equation:

$$k = \exp\left(-\frac{E_{act}}{R}\left(\frac{1}{T} - \frac{1}{T_{act}}\right)\right)$$

where k is the rate of unfolding, E_{act} is the activation energy of unfolding, R is the gas constant, T is the absolute temperature and T_{act} is an expression of the preexponential factor A, that has been transformed into an exponent with the single new parameter T_{act}, for more robust parameter estimation. From the resulting fits the activation Gibbs free energy ($\Delta G^{\ddagger}$) can be calculated:

$$\Delta G^{\ddagger} = -RT \ln\left(\frac{k_{(T)} \times h}{k_b \times T}\right)$$

where R is the universal gas constant, T is the absolute temperature, $k_{(T)}$ is the first order rate constant of unfolding at a given temperature, h is the Planck constant and k_b is the Boltzmann constant. The activation enthalpy ($\Delta H^{\ddagger}$) was calculated by:

$$\Delta H^{\ddagger} = E_a - RT$$

and the activation entropy ($T\Delta S^{\ddagger}$) was calculated as:

$$T\Delta S^{\ddagger} = \Delta H^{\ddagger} - \Delta G^{\ddagger}$$

The thermodynamic activation parameters $\Delta G^{\ddagger}$, $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ were calculated at temperatures corresponding to the apparent melting points ($T_{m\,(DSC)}$).

2.5 Molecular dynamics (MD) simulations

MD simulations were carried out for both $VPR_{\Delta C}$ and $VPR_{\Delta C}$_W6F using the Gromacs X.X.X software package (www.gromacs.org) on a parallel architecture utilizing the CHARMM22*

force field [25] for simulations. In order to provide starting structures for the simulations the X-ray crystal structure of VPR (PDB ID: 1SH7) [26] was modified using UCSF Chimera [27]. In-silico modifications of $VPR_{\Delta C}$ consisted of C-terminal deletion, deleting Gly276 and succeeding residues. In the case of $VPR_{\Delta C}$_W6F the newly made $VPR_{\Delta C}$ structure was subjected to in-silico mutagenesis of Trp6 choosing a rotamer for the new Phe6 residue that closely resembled the native Trp rotamer. Structures were soaked in TIP3P dodecahedral water with a minimum of 1 nm between the protein and the edge of the water box, all succeeding MD simulations were then conducted under the periodic boundary condition. Preparation of models was as follows: (i) steepest energy minimization of 10,000 steps; (ii) addition of two sodium ions to counter the system charge; (iii) steepest descent energy minimization of 10,000 steps; (iv) solvent equilibration for 100 ps at 300 K while restraining the protein and bound calcium atoms by a harmonic potential with a force constant of 1000 $kJ(mol^{-1}nm^{-2})$; (v) pressure equilibration of the system without any positional restraints was conducted under NPT ensemble at 1 bar, 300 K for 4 ns using a Parrinello-Rahman barostat; (vi) temperature equilibration of the system to 300 K utilizing Berendsen thermostat under NVT ensemble; (vii) a 2 ns thermalization in NVT conditions at 300 K under isotropic pressure at 1 bar using the velocity-rescale thermostat [28].

The productive MD simulation was carried out for 500 ns for $VPR_{\Delta C}$ and 1000 ns for $VPR_{\Delta C}$_W6F under the NVT ensemble using the velocity-rescale thermostat [28] using a coupling constant 0.2 ps at 300 K. For heavy atom bonds the LINCS algorithm was used setting both LINCS-order and LINCS-iter as 4 [29]. Waals and Coulomb interactions were cutoff at 0.8 nm. Long-range electrostatic interactions were calculated using the Particle-Mesh Ewald (PME) summation scheme [30].

3. Results

The tryptophan exchange variant $VPR_{\Delta C}$_W6F was successfully expressed and purified to homogeneity. However, the variant was found to be much more sensitive to calcium concentrations in the expression cultivate compared to the wild type. Additions of calcium up to 100 mM concurrently with inducing the expression were needed for reliant optimal production of the variant [15].

3.1 Kinetics

The determined Michaelis Menten kinetic parameters showed both lower turnover number and K_m for the variant as compared to the wild type at 25 °C at pH 8.6, that resulted in somewhat lower catalytic efficiency (Table 1).

Table 1. *Kinetic parameters of $VPR_{\Delta C}$ and $VPR_{\Delta C}$_W6F. Values are shown as averages and the standard deviation of the mean of at least three separate experiments.*

	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M ($s^{-1}mM^{-1}$)
$VPR_{\Delta C}$	225.7 ± 12.0	0.178 ± 0.016	1238 ± 149
$VPR_{\Delta C}$/W6F	150.9 ± 31.9	0.145 ± 0.028	1098 ± 167

Monitoring the effects of temperature on the turnover number show changes in activity/temperature relationships. Plotting measured k_{cat} values in Arrhenius graphs (Fig. 1) to calculate the thermodynamic parameters of activation for catalysis show that the W6F variant has lower activation enthalpy and a more negative values for activation entropy (Table 2).

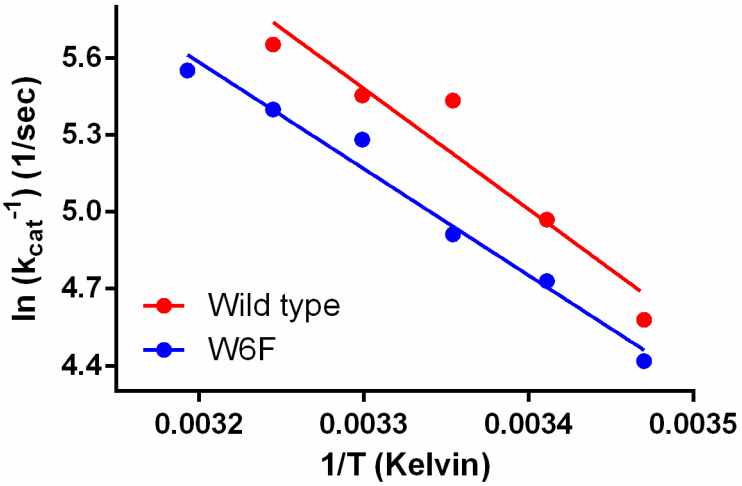


Figure 1. *Arrhenius plot calculated form the rate of catalysis of Suc-AAPF-NH-Np at pH 8.6 and temperatures ranging from 15 - 45 °C. Red dots represent measured values for VPR and the red line represent the best linear fit of the data. Blue dots represent measured values for the W6F variant, and the blue line represents the best linear fit.*

Table 2. *The thermodynamic activation parameters for the catalysis of Suc-AAPF-NH-Np at 25 °C. Parameters shown are the activation Gibbs free energy of catalysis ($\Delta G^{\ddagger}$), activation enthalpy of catalysis ($\Delta H^{\ddagger}$) and the activation entropy of catalysis ($T\Delta S^{\ddagger}$).*

	$\Delta G^{\ddagger}$ (kJ/mol)	**$\Delta H^{\ddagger}$ (kJ/mol)**	**$T\Delta S^{\ddagger}$ (kJ/mol)**
$VPR_{\Delta C}$	59.6	36.7	-22.9
$VPR_{\Delta C}$/W6F	60.6	32.2	-28.4

3.2 Structural effects

To monitor changes in the structure of the variant compared to the wild type, fluorescence spectroscopy and CD wavelength scans did provide information strongly suggesting a less structured protein (Fig. 2). The steady-state fluorescence spectrum of the variant suggests that the microenvironments around some or all of the remaining Trp residues (Trp114, Trp191 and Trp 208) are much more exposed to a more polar environment as seen in the shift of around 11 nm of the maxima of the emission curve. In addition, the remaining fluorophores do emit more in comparison to the wild type fluorophores as the area under the curve stays the same with one less Trp residue (Table 3), indicating less intrinsic quenching, reflecting changes in dynamics and/or orientation of the residues. In addition, the microenvironments of the remaining Trp residues have become more permeable to acrylamide, as evidenced by a considerably higher Stern-Volmer constant, adding more evidence for a less structured protein (Fig. 3C) (Table 3).

Table 3. *Calculated Stern-Volmer constants from acrylamide quenching experiments, the maxima of emission curves (λ_{max}) and the relative fluorescence calculated as area under curve (AUC) recorded at pH 8.0 at 25 °C. Values are represented as averages and standard deviation of the mean of at least three separate experiments.*

	Stern-Volmer (M^{-1})		**λ_{max} (nm)**		**AUC (relative)**	
	Native	Denatured	Native	Denatured	Native	Denatured
$VPR_{\Delta C}$	2.23 ± 0.31	14.71 ± 0.66	335.4 ± 0.6	357.4 ± 0.2	1.00 ± 0.03	10.02 ± 0.03
$VPR_{\Delta C}$/W6F	4.98 ± 0.66	13.50 ± 0.58	345.7 ± 3.8	356.6 ± 0.8	1.03 ± 0.11	5.71 ± 1.21

CD wavelength scans also indicate changes in the secondary structure. The upward shift in the spectrum of $VPR_{\Delta C}$_W6F indicates that the protein is less structured than the wild type. In addition, that shift is very noticeable around 222 nm, suggesting that a major part of the structure lost may be α-helical [31].

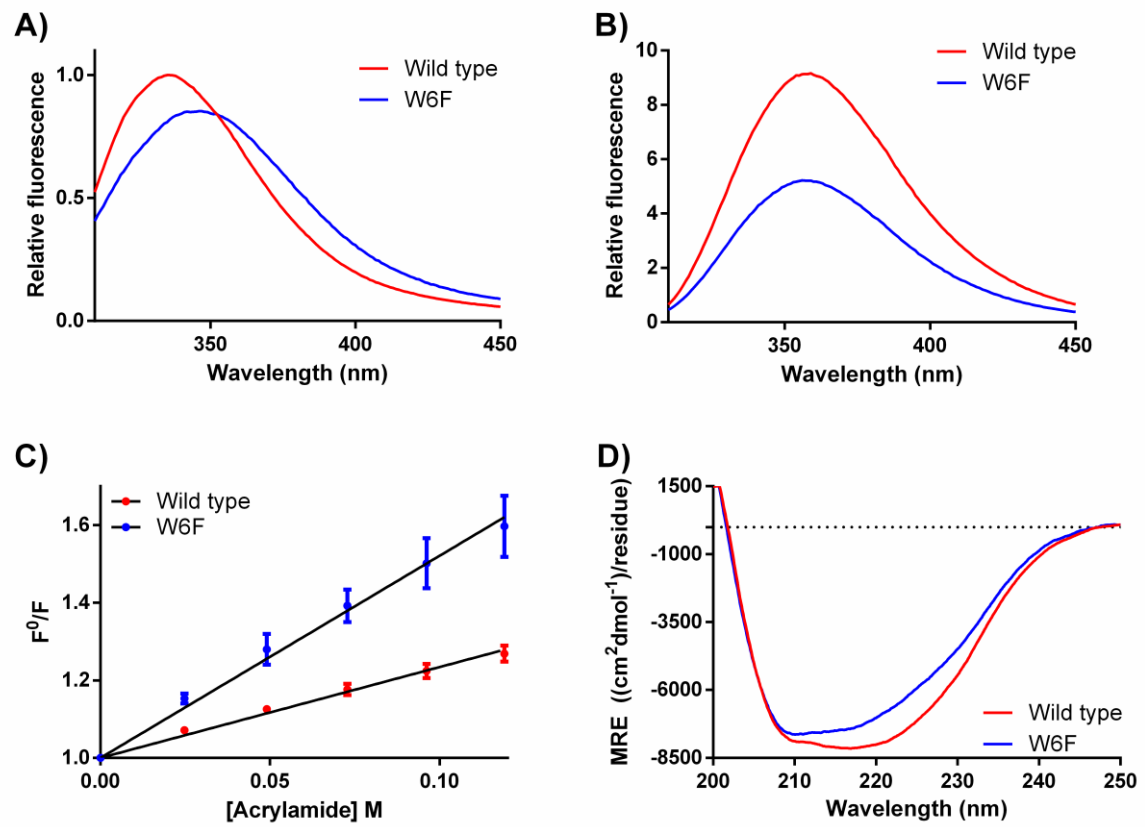


***Figure 2**. **A)** Steady state fluorescence spectrum of the native states of $VPR_{\Delta C}$ and $VPR_{\Delta C}$_W6F at pH 8.0 and 25 °C. **B)** Steady state fluorescence spectrum of the denatured states of $VPR_{\Delta C}$ and $VPR_{\Delta C}$_W6F at pH 8.0 and 25 °C. **C)** Stern-Volmer graph showing the quenching effects of acrylamide on $VPR_{\Delta C}$ and $VPR_{\Delta C}$_W6F at pH 8.0 and 25 °C. Error bars represent standard deviation of the mean of measurements. **D)** CD wavelength scans of $VPR_{\Delta C}$ and $VPR_{\Delta C}$_W6F at pH 8.6 and 25 °C. All measurements were done in triplicates at least.*

Those observations were supported by MD simulations that showed unraveling of helices D and C (Crystal structure), that are located in close proximity to calcium binding sites 1 and 2, due to the breakdown of an H-bond network and contacts that were reliant on the presence of that Trp residue. The mutation seemingly does not disrupt the N-terminus itself in such a drastic way. As MD simulations show a Phe rotamer that is much more solvent exposed pointing out of the pocket that Trp residue resides in in the wild type. This may lead to some changes in the stability of this region. The N-terminal which is highly mobile in wild type simulations, becomes much more static in the $VPR_{\Delta C}$_W6F simulation however, possibly indicating that the N-terminus is mostly dependent on the calcium binding site for its stability whereas the Trp residue is a relay of contacts stabilizing the rest of the structure.

3.3 Stability

The stability of the W6F variant was greatly diminished, with $T_{50\%}$ and T_m values dropping by 13.1 °C and 12.6 °C, respectively (Table 4) (Fig. 3). It is therefore clear that the mutation greatly affects the kinetic stability of the variant and the resistance against proteolytic degradation (Fig. 3. A). In addition to this, the W6F variant does show loss of cooperativity during unfolding, as the DSC thermogram clearly shows a second peak (Fig. 3. C). Thus, an intermediate state forms during the unfolding of the W6F variant. This possibly explains the lower activation energies measured in inactivation experiments (Table 4) as a less structured intermediate would be more prone to autoproteolytic degradation. DSC thermograms also indicate that the heat capacity of the native state of the W6F variant might be lower than that of the wild type. This is hard to estimate however, due to the apparent increased aggregation of the W6F variant. This can be observed as downward sloping of heat capacity traces post unfolding transition, as well as from the difference in heat capacities of the first and second run of the W6F variant (Fig. 3. C and D).

Table 4. *Thermostability parameters of VPRΔC and its W6F variant. Parameters shown are* $T_{50\%}$, *the temperature where half of the activity has been lost over 30 min calculated from Arrhenius graphs,* E_{act} *(inactivation) calculated from the slope of Arrhenius graphs,* T_m *(CD) the melting point of the PMSF inhibited enzymes as measured by CD, Tm (DSC) the apparent melting point of PMSF inhibited enzymes calculated as the highest peak of DSC thermograms. Numbering refers to the order of peaks. Values are represented as averages and standard deviations of the mean of at least four separate experiments.*

	$T_{50\%}$ (°C)	$E_{act\ (inactivation)}$ (kJ/mol)	$T_{m\ (CD)}$ (°C)	$T_{m\ (DSC)}^{1}$ (°C)	$T_{m\ (DSC)}^{2}$ (°C)
$VPR_{\Delta C}$	53.8 ± 0.4	218 ± 9	61.6 ± 0.5	63.9 ± 0.4	N.A.
$VPR_{\Delta C}$/W6F	40.7 ± 2.0	154 ± 14	49.0 ± 0.6	49.3 ± 0.2	59.4 ± 0.4

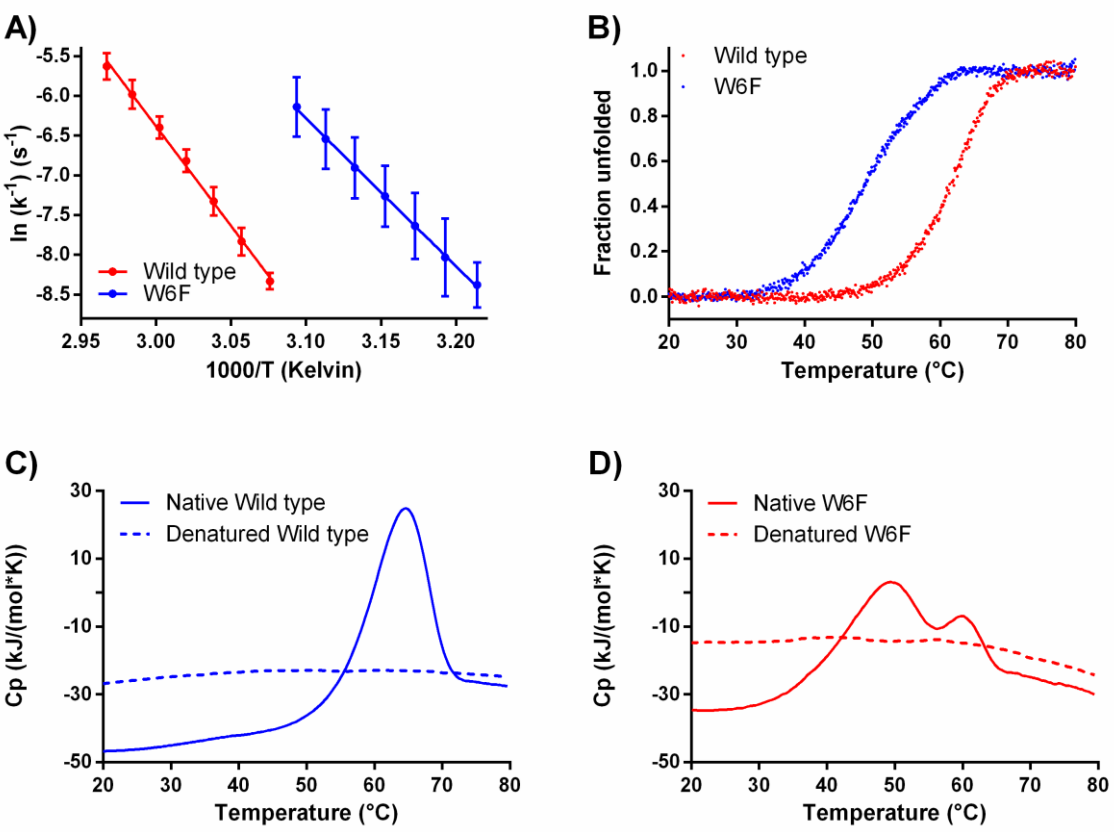


Figure 3. A) *Arrhenius plots calculated from thermal inactivation of the wild type (red) and the W6F variant (blue) in Tris buffer containing 15 mM $CaCl_2$, error bars represent the standard deviation of the mean.* ***B)*** *Normalized secondary structure melting profiles of the wild type (red) and the W6F variant (blue) recorded on CD at 222 nm in a glycine buffer at pH 8.6 containing 15 mM $CaCl_2$.* ***C)*** *Example of a native (solid line) and reheat scan (dotted line) DSC scan of the wild type.* ***D)*** *Example of a native (solid line) and reheat scan (dotted line) DSC scan of the W6F variant.*

Averages of deconvoluted DSC thermograms and averaged CD melting profiles were subjected to global fitting using CalFitter 1.3 (Fig. 4) (Table 5). The wild type was fitted to a two-state irreversible model, whereas the W6F variant was fitted to a three-state kinetically stable model. The fitting shows that the first transition of the W6F variant involves the major part of the secondary structure with the second transition being around 20% of the CD signal. The second transition however has very high values for both the activation enthalpy and activation entropy compared to the first transition. This information suggests that the second transition is that of an intermediate that has lost most of its regular secondary structure, but which still maintains strong and ordered molecular interactions. In addition to the location of the mutation this strongly suggests that the second transition involves the N-terminal part of the protein and the

calcium 3 binding site located there (Fig. 5). The loss of cooperativity due to this single mutation might thus be due to disconnection between the N-terminal part and the rest of the protein molecule. In the crystal structure of VPR, the side chain of Trp6 resides in a pocket in close proximity to the loop between the main coordinators of the calcium 1 binding site, Pro171, Gly173 and Asp196, a site proposed to be highly important for the stability of the secondary structure [15, 17]. In addition, Trp6 has the potential to form sidechain-mainchain H-bond to Ser202, a residue relatively close to Asp196 (Fig. 5) that might be important for the stability of sheets 6 and 7 (Figs. 5).

Table 5. *Results of CalFitter 1.3 using the global fit function fitting together deconvoluted DSC data and normalized CD data. Parameters shown are the activation Gibbs free energy ($\Delta G^{\ddagger}$), activation enthalpy ($\Delta H^{\ddagger}$) and the activation entropy ($\Delta S^{\ddagger}$). Values are calculated at temperatures corresponding to their respective $T_{m\ (DSC)}$. Numerals indicate observed order of transitions. Data is represented with their 95% confidence interval.*

	$\Delta G^{\ddagger}_{peak}{}^{1}$ (kJ/mol)	$\Delta G^{\ddagger}_{peak}{}^{2}$ (kJ/mol)	$\Delta H^{\ddagger}_{peak}{}^{1}$ (kJ/mol)	$\Delta H^{\ddagger}_{peak}{}^{2}$ (kJ/mol)	$T\Delta S^{\ddagger}_{peak}{}^{1}$ (kJ/mol)	$T\Delta S^{\ddagger}_{peak}{}^{2}$ (kJ/mol)
VPR$_{\Delta C}$	99.3 ± 0.1	N.A.	235 ± 1	N.A.	135 ± 1	N.A.
VPR$_{\Delta C}$/W6F	94.1 ± 0.1	96.4 ± 0.1	177 ± 1	285 ± 2	83 ± 1	189 ± 2

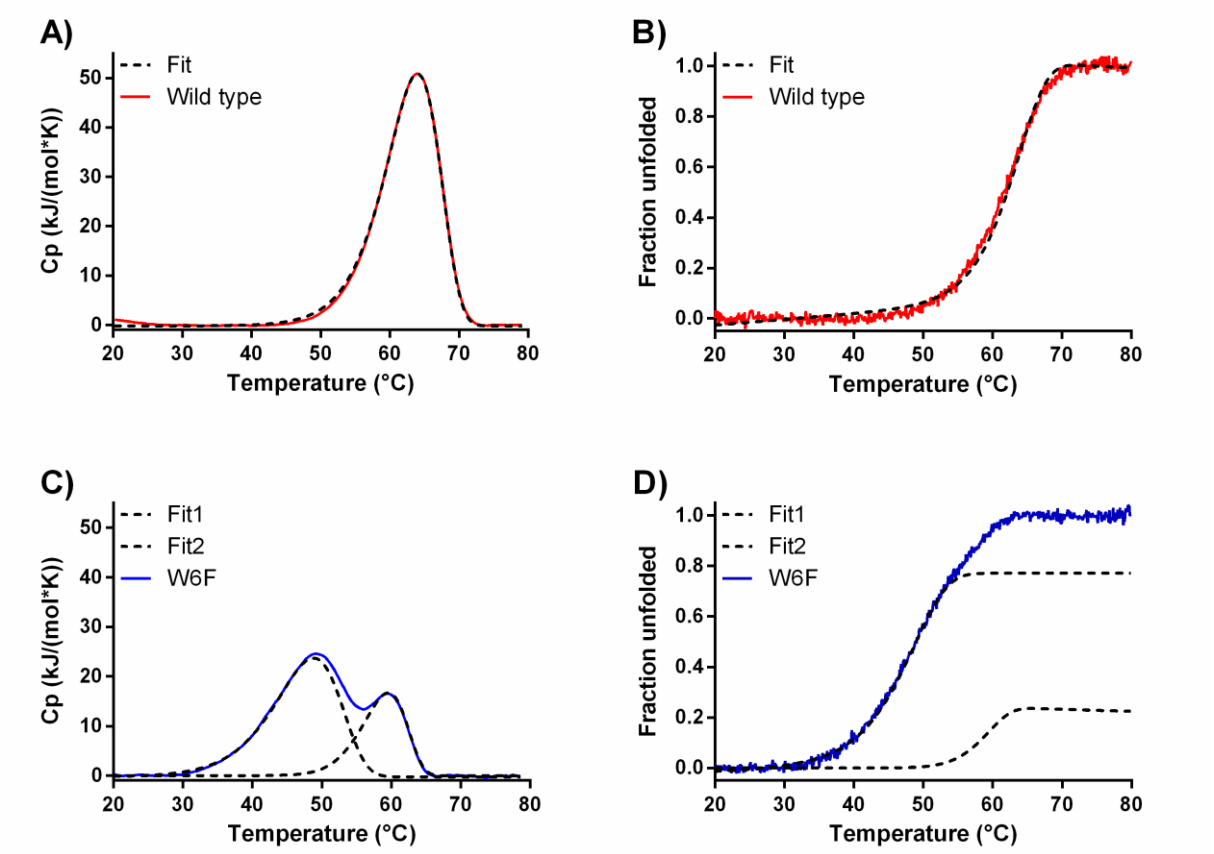


Figure 4. *Results of CalFitter 1.3 global fits of DSC and CD data of $VPR_{\Delta C}$ and $VPR_{\Delta C}$W6F.* ***A)*** *Fit (black dotted line) of averaged and deconvoluted DSC thermograms of $VPR_{\Delta C}$ (red line).* ***B)*** *Fit (black dotted line) of averaged and normalized CD melting profiles of $VPR_{\Delta C}$ (red line).* ***C)*** *Fits (black dotted lines) of averaged and deconvoluted DSC thermograms of $VPR_{\Delta C}$W6F (blue line).* ***D)*** *Fits (black dotted lines) of averaged and normalized CD melting profiles of $VPR_{\Delta C}$W6F (blue line).*

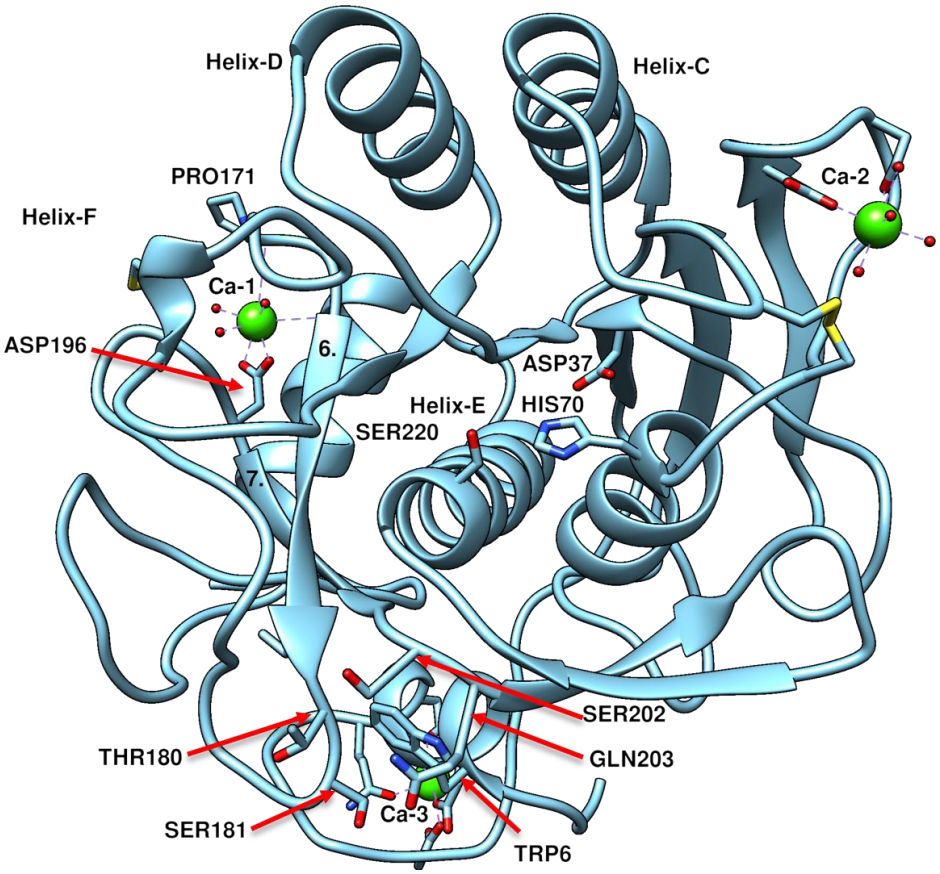


Figure 5. *The crystal structure of VPR$_{\Delta C}$. Marking the calcium ion binding sites, the active site (Ser220, His70 and Asp37), relevant α-helixes (C, D, E and F), β-sheets (6 and 7), the coordinators of the Ca-1 ion Asp196 and Pro171, neighboring residues of Trp6, Thr180, Ser181, Ser202 and Gln203 and the crystalline rotameric state of theTrp6 residue.*

4. Discussion

The work presented here demonstrates in further detail the importance of the N-terminal part of VPR and its role in maintaining the kinetic stability of the structure. In addition, this work highlights how finely tuned the unfolding process of the wild type is under our standard assay conditions. Unfolding cooperativity has been pointed out to be extremely important for kinetically stable proteinases to carry out their biological roles [8]. Intermediate unfolded states are targets for proteolytic cleavage and are aggregation prone [1]. Using mutagenic studies and investigation of the effects of calcium on the unfolding process, the N-terminus of the protein has become a site of interest [15, 17, 32]. The N-terminal region harbors the highly coordinated calcium binding site 3 (Ca3) that is likely formed during its maturation process, locking the enzyme in its kinetic trap and has been shown to be very influential for the stability of the protein structure [17]. Thus, we pointed our focus on the highly conserved Trp6 residue, located

on the small helix A with its sidechain tucked between the N-terminus and other parts of the protein. The decision was taken to exchange the residue for a smaller aromatic group (Phe) to investigate the effects of how fewer contacts in that area might affect the protein. It became quite clear early on that the stability of the resulting variant was highly compromised, as expression of the variant provided very low yields of protein. Sufficient expression was not achieved until relatively high concentrations of calcium were added to the expression culture medium along with IPTG for induction. A final concentration of 100 mM $CaCl_2$ was observed to give constant acceptable yields. These high calcium concentrations have only been correlated with resistance against proteolytic cleavage [15]. Interestingly the structure of the Trp variant seems to be affected and evidence of dissolution of parts of the secondary structure became apparent in CD wavelength scans. In addition, changes were observed in the steady state fluorescence spectrum of the variant where a large red shift indicated that the remaining three Trp residues in the structure were in a more polar environment. Furthermore, this was accompanied by higher Stern-Volmer constants indicating better access to fluorophores by acrylamide [33]. MD simulations provided extra insight into the effects on the structure and show far reaching effects on the structure. Accompanying the 12.6 °C lower melting point and 13.1 °C lower $T_{50\%}$, was a clear indication that an intermediate state was formed during unfolding. From observation on how the single mutation W6F completely disrupted the cooperativity of the unfolding process, it is clear that important interactions must have been broken. In MD simulations the rotameric states of Trp6 in the wild type and Phe6 in the variant differ remarkably. Trp6 resides almost always in its pocket maintaining the H-bond to Ser202 and proximity to Gln203 and the loop area around Thr180 and Ser181. However, Phe6 in the mutant spends most of the simulation pointing out toward the solvent. Thus, the aforementioned interactions are completely lost. Residues 180, 181, 202 and 203 are located on the C-terminals of β-sheets 6 and 7, that both contain coordinators of the calcium ion in the Ca1 site (Fig. 5). Changed dynamics there might be a part of the reason why helices D and C partly unravel during the simulation. This idea would fit with the hypothesis that the Ca1 site is very important for maintaining the secondary structure of VPR [15]. Thus, the role of Trp6 might be to infer stability from the N-terminus to the main structure of VPR, possibly in part through interactions with the Ca1 site. Moreover, these structural changes cause changes to the activity of the W6F variant. Showing lower activation enthalpy change during catalysis and higher activation entropy change. These results would suggest a more flexible/softer surface, of the protein, suggesting that the W6F variant might be more active at very low temperatures.

Acknowledgments

This work was supported by The Icelandic Research Fund [grant number 162977-051].

References

[1] J.M. Sanchez-Ruiz, Protein kinetic stability, Biophysical Chemistry, 148 (2010) 1-15.
[2] D. Baker, A.K. Shiau, D.A. Agard, The role of pro regions in protein folding, Curr Opin Cell Biol, 5 (1993) 966-970.
[3] U. Shinde, M. Inouye, Intramolecular Chaperones and Protein-Folding, Trends Biochem Sci, 18 (1993) 442-446.
[4] J.L. Sohl, S.S. Jaswal, D.A. Agard, Unfolded conformations of alpha-lytic protease are more stable than its native state, Nature, 395 (1998) 817-819.
[5] U. Shinde, M. Inouye, Intramolecular chaperones: polypeptide extensions that modulate protein folding, Semin Cell Dev Biol, 11 (2000) 35-44.
[6] H. Ikemura, H. Takagi, M. Inouye, Requirement of Pro-Sequence for the Production of Active Subtilisin-E in Escherichia-Coli, Journal of Biological Chemistry, 262 (1987) 7859-7864.
[7] D.M. Epstein, P.C. Wensink, The Alpha-Lytic Protease Gene of Lysobacter-Enzymogenes - the Nucleotide-Sequence Predicts a Large Prepro-Peptide with Homology to Pro-Peptides of Other Chymotrypsin-Like Enzymes, Journal of Biological Chemistry, 263 (1988) 16586-16590.
[8] N.L. Salimi, B. Ho, D.A. Agard, Unfolding Simulations Reveal the Mechanism of Extreme Unfolding Cooperativity in the Kinetically Stable alpha-Lytic Protease, Plos Comput Biol, 6 (2010).
[9] E.L. Cunningham, S.S. Jaswal, J.L. Sohl, D.A. Agard, Kinetic stability as a mechanism for protease longevity, P Natl Acad Sci USA, 96 (1999) 11008-11014.
[10] S.S. Jaswal, S.M.E. Truhlar, K.A. Dill, D.A. Agard, Comprehensive analysis of protein folding activation thermodynamics reveals a universal behavior violated by kinetically stable proteases, J Mol Biol, 347 (2005) 355-366.
[11] I.P. Shinde, X. Fu, C. Marie-Claire, Y. Yabuta, M. Inouye, Intramolecular chaperone mediated folding of subtilisin., Abstr Pap Am Chem S, 219 (2000) U154-U154.
[12] A.G. Sigurdardottir, J. Arnorsdottir, S.H. Thorbjarnardottir, G. Eggertsson, K. Suhre, M.M. Kristjansson, Characteristics of mutants designed to incorporate a new ion pair into the structure of a cold adapted subtilisin-like serine proteinase, Biochim Biophys Acta, 1794 (2009) 512-518.
[13] M.M. Kristjansson, O.T. Magnusson, H.M. Gudmundsson, G.A. Alfredsson, H. Matsuzawa, Properties of a subtilisin-like proteinase from a psychrotrophic Vibrio species comparison with proteinase K and aqualysin I, Eur J Biochem, 260 (1999) 752-760.
[14] R.J. Siezen, J.A. Leunissen, Subtilases: the superfamily of subtilisin-like serine proteases, Protein Sci, 6 (1997) 501-523.
[15] K.R. Oskarsson, M.M. Kristjansson, Improved expression, purification and characterization of VPR, a cold active subtilisin-like serine proteinase and the effects of calcium on expression and stability, Biochim Biophys Acta Proteins Proteom, 1867 (2019) 152-162.
[16] J. Arnorsdottir, R.B. Smaradottir, O.T. Magnusson, S.H. Thorbjarnardottir, G. Eggertsson, M.M. Kristjansson, Characterization of a cloned subtilisin-like serine proteinase from a psychrotrophic Vibrio species, Eur J Biochem, 269 (2002) 5536-5546.

[17] K.R. Oskarsson, A.F. Saevarsson, M.M. Kristjansson, Thermostabilization of VPR, a kinetically stable cold adapted subtilase, via multiple proline substitutions into surface loops, Sci Rep, 10 (2020) 1045.
[18] J. Aqvist, M. Kazemi, G.V. Isaksen, B.O. Brandsdal, Entropy and Enzyme Catalysis, Accounts Chem Res, 50 (2017) 199-207.
[19] S. Wagner, M.M. Klepsch, S. Schlegel, A. Appel, R. Draheim, M. Tarry, M. Hogbom, K.J. van Wijk, D.J. Slotboom, J.O. Persson, J.W. de Gier, Tuning Escherichia coli for membrane protein overexpression, Proc Natl Acad Sci U S A, 105 (2008) 14371-14376.
[20] C.N. Pace, F. Vajdos, L. Fee, G. Grimsley, T. Gray, How to measure and predict the molar absorption coefficient of a protein, Protein Sci, 4 (1995) 2411-2423.
[21] E.G. Delmar, C. Largman, J.W. Brodrick, M.C. Geokas, Sensitive New Substrate for Chymotrypsin, Anal Biochem, 99 (1979) 316-320.
[22] S. Mazurenko, A. Kunka, K. Beerens, C.M. Johnson, J. Damborsky, Z. Prokop, Exploration of Protein Unfolding by Modelling Calorimetry Data from Reheating, Sci Rep-Uk, 7 (2017).
[23] S. Mazurenko, J. Stourac, A. Kunka, S. Nedeljkovic, D. Bednar, Z. Prokop, J. Damborsky, CalFitter: a web server for analysis of protein thermal denaturation data, Nucleic Acids Res, 46 (2018) W344-W349.
[24] J.M. Sanchez-Ruiz, J.L. Lopez-Lacomba, M. Cortijo, P.L. Mateo, Differential scanning calorimetry of the irreversible thermal denaturation of thermolysin, Biochemistry, 27 (1988) 1648-1652.
[25] B.R. Brooks, C.L. Brooks, A.D. Mackerell, L. Nilsson, R.J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A.R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R.W. Pastor, C.B. Post, J.Z. Pu, M. Schaefer, B. Tidor, R.M. Venable, H.L. Woodcock, X. Wu, W. Yang, D.M. York, M. Karplus, CHARMM: The Biomolecular Simulation Program, Journal of Computational Chemistry, 30 (2009) 1545-1614.
[26] J. Arnorsdottir, M.M. Kristjansson, R. Ficner, Crystal structure of a subtilisin-like serine proteinase from a psychrotrophic Vibrio species reveals structural aspects of cold adaptation, Febs J, 272 (2005) 832-845.
[27] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera--a visualization system for exploratory research and analysis, J Comput Chem, 25 (2004) 1605-1612.
[28] G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling, J Chem Phys, 126 (2007).
[29] B. Hess, H. Bekker, H.J.C. Berendsen, J.G.E.M. Fraaije, LINCS: A linear constraint solver for molecular simulations, Journal of Computational Chemistry, 18 (1997) 1463-1472.
[30] T. Darden, D. York, L. Pedersen, Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems, J Chem Phys, 98 (1993) 10089-10092.
[31] N.J. Greenfield, Using circular dichroism spectra to estimate protein secondary structure, Nat Protoc, 1 (2006) 2876-2890.
[32] J. Arnorsdottir, A.R. Sigtryggsdottir, S.H. Thorbjarnardottir, M.M. Kristjansson, Effect of proline substitutions on stability and kinetic properties of a cold adapted subtilase, J Biochem, 145 (2009) 325-329.
[33] J.R. Lakowicz, Principles of Fluorescence Spectroscopy, 3 ed., Springer Science, 1983.