

Crystallographic studies on a cold adapted subtilase and proteins involved in mRNA processing

Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultäten
der Georg-August-Universität zu Göttingen

vorgelegt von
Jóhanna Arnórsdóttir
aus Reykjavík

Göttingen 2005

D7

Referent: Prof. Dr. Ralf Ficner

Korreferent: Prof. Dr. Oliver Einsle

Tag der mündlichen Prüfung:

Abstract

The crystal structure of a subtilisin-like serine proteinase from the psychrotrophic marine bacterium, *Vibrio* sp. PA-44, was solved by means of molecular replacement and refined at 1.84 Å resolution. This is the first structure of a cold adapted subtilase to be determined and its elucidation facilitates examination of the molecular principles underlying temperature adaptation of enzymes. The cold adapted *Vibrio*-proteinase was compared to known three-dimensional structures of homologous enzymes of meso- and thermophilic origin, proteinase K and thermitase, to which it has high structural resemblance. The main structural features standing out as plausible determinants of the different temperature adaptation of the enzymes involve the character of their exposed and buried surface areas. The hydrophobic effect is found to play a significant role for the structural stability of the meso- and thermophile enzymes, whereas the cold enzyme exposes more of its apolar surface. In addition, the cold adapted *Vibrio*-proteinase is distinguished from the more stable enzymes by its strong anionic character arising from the high occurrence of uncompensated negatively charged residues on its surface. Interestingly, both the cold and thermophile proteinases differ from the mesophile enzyme in having more extensive hydrogen- and ion pair interactions in their structures, supporting suggestions of the dual role of electrostatic interactions in adaptation of enzymes to both high and low temperatures. The *Vibrio*-proteinase has three calcium ions associated with its structure, one of which is in a calcium-binding site not described in other subtilases.

A 61 kDa protein component of the human spliceosome is indispensable for the assembly of the [U4/U6·U5] triple snRNP. The gene encoding the 61 kDa spliceosomal protein was overexpressed in *E. coli* and purified for crystallisation trials. No crystals were obtained. Evidence of incomplete folding and heterogeneity due to unspecific binding of *E. coli* nucleic acids emerged in the purification process.

A 42 kDa protein component of the *Trypanosoma brucei* RNA editosome has endo- and exonuclease activity and is suggested to mediate protein-protein and/or protein-RNA interactions. Attempts were made to crystallise the 42 kDa protein with and without non-hydrolysable double stranded nucleic acid ligands but without success.

Zusammenfassung

Die Kristallstruktur einer Subtilisin-ähnlichen Serin-Proteinase des psychrotrophen marinen Bakteriums, *Vibrio* sp. PA-44, wurde durch molekularen Ersatz bei 1,84 Å Auflösung aufgeklärt. Damit konnte zum ersten Mal die Struktur einer kälteadaptiven Subtilase aufgeklärt werden. Dies macht es möglich, die molekularen Prinzipien, die der Temperaturadaption von Enzymen zu Grunde liegen, genauer zu untersuchen. Die kälteadaptive *Vibrio*-Proteinase wurde mit bekannten dreidimensionalen Strukturen von homologen Enzymen meso- und thermophilen Ursprungs, Proteinase K und Thermitase, verglichen, zu denen eine hohe strukturelle Ähnlichkeit besteht. Zu den strukturellen Hauptmerkmalen, die als mögliche Determinanten der unterschiedlichen Temperaturadaption der verglichenen Enzyme in Frage kommen, gehört der Charakter ihrer exponierten und verborgenen Oberfläche. So scheint der hydrophobe Effekt wichtig zu sein für die strukturelle Stabilität der meso- und thermophilen Enzyme, während die kälteadaptiven Enzyme einen größeren Teil ihrer apolaren Oberfläche exponieren. Hinzu kommt, dass sich die kälteadaptive *Vibrio*-Proteinase von den stabileren Enzymen durch einen stark anionischen Charakter unterscheidet, der durch das hohe Auftreten von negativ geladenen Resten auf ihrer Oberfläche hervorgerufen wird. Interessanterweise unterscheiden sich die kälteadaptive sowie die thermophile Proteinase von dem mesophilen Enzym durch ausgeprägtere Wasserstoff- und Ioneninteraktionen in ihrer Struktur, was auf eine duale Rolle der elektrostatischen Interaktionen in der Adaption von Enzymen zu hohen sowie tiefen Temperaturen andeutet. Die *Vibrio*-Proteinase hat drei Kalziumione an ihrer Oberfläche gebunden, davon eine an einer Kalzium bindenden Stelle, die bei anderen Subtilasen nicht bekannt ist.

Ein 61 kDa Protein des humanen Spleissosoms ist unverzichtbar für den Aufbau des [U4/U6·U5] tri-snRNPs. Das 61 kDa Protein wurde in *E. coli* hergestellt und für Kristallisationsansätze gereinigt. Kristalle von dem 61 kDa Protein konnten nicht gewonnen werden.

Ein 42 kDa Protein des *Trypanosoma brucei* RNA Editosoms hat Endo- und Exonuklease Aktivität und soll Protein-Protein und/oder Protein-RNA Interaktionen vermitteln. Versuche, dieses 42 kDa Protein zu kristallisieren, waren sowohl mit als auch ohne Nukleinsäureliganden nicht erfolgreich.

INTRODUCTION.....	3
Structural principles of cold adaptation	4
Extremophiles	4
Extreme temperature challenges	4
Thermodynamics aspects of stability	6
Marginal stability	8
Serine proteinases	10
Subtilases	12
Proteinase K and Thermitase.....	14
The cold adapted <i>Vibrio</i> -proteinase	16
Posttranscriptional RNA processing	19
Splicing	20
The [U4/U6·U5] triple snRNP	22
Protein p61	23
RNA editing.....	25
Mechanism of RNA editing	28
Editosome	30
TbMP42 and its relatives	32
METHODS AND MATERIALS	34
Materials.....	34
Chemicals.....	34
Equipment	34
Buffers.....	34
Enzymes and inhibitors.....	34
Kits.....	35
Bacterial strains.....	35
Vectors used for production of recombinant proteins in <i>E. coli</i>	36
Medium and antibiotics.....	36
Oligonucleotides	37
Molecular biology techniques	38
Polymerase Chain Reaction	38
Ethanol precipitation	38
Plasmide preparation.....	38
Agarose-gel electrophoresis	39
DNA restriction digest	39
Gel extraction from agarose gels.....	39
Ligation.....	39
Competent cell preparation	40
Transformation.....	40
Identification of positive clones	41
SDS-PAGE.....	41
Western Blotting	42
Crystallisation methods.....	43
Crystal growth.....	43
Seeding.....	44
Protein preparation, crystallisation and structure determination of the <i>Vibrio</i>-proteinase.....	45
Expression and purification of VPR.....	45
Activity assay.....	46
Crystallisation and data collection	46
Structure solution and refinement	47
Structure analysis	47

Protein preparation for crystallisation trials of p61	48
Expression and purification of p61	48
Protein preparation for crystallisation trials TbMP42.....	49
Cloning, expression and purification of TbMP42	49
RESULTS	54
The cold adapted <i>Vibrio</i>-proteinase.....	54
Preparation of the <i>Vibrio</i> -proteinase for crystallisation trials	54
Crystallisation of the <i>Vibrio</i> -proteinase	56
Structure determination of the <i>Vibrio</i> -proteinase	57
Crystal structure of the <i>Vibrio</i> -proteinase	58
Overall structure comparison of the cold adapted <i>Vibrio</i> -proteinase and related enzymes from meso- and thermophiles	59
Charged residues and ion pairs.....	62
Hydrogen bonds	65
Calcium binding sites.....	65
Disulfide bridges	67
Surface properties and packing	69
Protein preparation of p61, a component of the [U4/U6·U5] triple snRNP.....	73
Expression and purification of p61	73
Coexpression with 15.5 kDa protein	75
Pull down with the 15.5 kDa protein.....	76
RNase stability test.....	77
Crystallisation trials of p61	77
Protein preparation and crystallisation attempts on TbMP42, a component of the <i>Trypanosoma brucei</i> editosome	78
TbMP42 sequence	78
TbMP42 constructs	80
Purification of TbMP42-His ₆	81
Crystallisation trials of TbMP42-His ₆	82
Purification of TbMP42-GST.....	83
Purification of TbMP42-MBP.....	84
Crystallisation trials of TbMP42-MBP	87
TbMP42_C-terminal (250-393)-MBP.....	89
DISCUSSION	90
The cold adapted <i>Vibrio</i>-proteinase.....	90
Structure comparisons of the <i>Vibrio</i> -proteinase, proteinase K and thermitase.....	90
Previous mutagenic experiments reviewed	91
Concluding remarks	97
The 61 kDa protein component of the [U4/U6·U5] triple snRNP	98
The editosomal protein TbMP42	99
ACKNOWLEDGEMENTS	101
REFERENCES.....	102
APPENDICES	112
ABBREVIATIONS	119
CURRICULUM VITAE.....	121

Introduction

In order to improve the understanding of the structural principles of temperature adaptation a subtilisin-like serine proteinase from the psychrotrophic marine bacterium, *Vibrio* sp. PA-44, was crystallised and its three-dimensional structure determined. The *Vibrio*-proteinase belongs to the family of proteinase K and shares a high sequence identity of 60 – 87 % with several meso- and thermophilic members of the family (Arnórsdóttir *et al.*, 2002). Furthermore, it shares 41 % sequence identity and 57 % similarity with proteinase K, the best characterised representative of this protein family, of which the three-dimensional structure has been determined to atomic resolution (Betzel *et al.*, 2001). The *Vibrio*-proteinase showed clear cold adaptive traits in comparison with its homologs of meso- and thermophilic origin (Kristjánsson *et al.*, 1999). A thorough sequence- and computer model comparisons performed on the *Vibrio*-proteinase and the most related enzymes of meso- and thermophilic origin revealed some differences, possibly relevant to temperature adaptation (Arnórsdóttir *et al.*, 2002). The results gave rise to ongoing mutagenic research at the University of Iceland, where single and combined amino acid substitutions aimed at increasing the stability of the *Vibrio*-proteinase are being tested. The elucidation of the *Vibrio*-proteinase structure now, the first structure of a cold adapted subtilase to be determined, enables a more focused examination on plausible determinants of different temperature adaptation among subtilases.

The cold adapted *Vibrio*-proteinase was crystallised in the presence of bound inhibitor, phenyl-methyl-sulfonate, and the structure was refined at 1.84 Å resolution. In order to identify parameters that might be important with respect to cold adaptation, structural features in the *Vibrio*-proteinase and the two most related enzymes of known three-dimensional structure, proteinase K from the mesophilic fungi *Tritirachium album* Limber and thermitase from the thermophilic eubacterium *Thermoactinomyces vulgaris*, were analysed and compared.

In addition, two proteins involved in posttranscriptional RNA processing, a 61 kDa protein component of the human spliceosome and a 42 kDa protein component of the *Trypanosoma brucei* editosome, were purified and attempts were made to crystallise the proteins.

Structural principles of cold adaptation

Extremophiles

Microorganisms inhabit the most diverse environments on earth. Extremophiles are microorganisms that have adapted to environmental conditions regarded by humans to fall out of the range of being normal with respect to factors like temperature, pressure, salinity or pH. Extremophiles have had to develop strategies to deal with environmental stress, mainly by molecular adaptation of their cell inventory. Of major importance in adapting to extreme environmental conditions is the optimisation of protein function and stability. Enzymes from extremophiles are essentially alike their mesophilic counterparts, sharing the same overall fold and catalysing identical reactions by the same mechanisms, while having adopted different traits regarding kinetic- and structural properties. Therefore they provide excellent tools to examine the molecular basis of different protein properties as well as the relation between structure and function of enzymes. Regarding temperature, organisms have been isolated from places with temperatures as high as 113 °C (Blochl *et al.*, 1997) and biological activity has been detected in microbial samples at as low as -20 °C (Deming, 2002).

Extreme temperature challenges

Thermo- and hyperthermophiles are by definition organisms that grow at temperatures in the range of 60-80 °C and above 80 °C respectively (Vieille *et al.*, 1996). These organisms face the challenge of maintaining their macromolecules functional under the environmental stress imposed by extreme thermal motion. As a response they have evolved enzymes characterised by a high degree of stability against heat and other denaturants. The increased stability of enzymes from thermo- and hyperthermophiles is considered to reflect structural rigidity, which in turn would account for their observed poor catalytic efficiency at low temperatures. The properties of thermophilic enzymes have awoken great interest as they have opened new potentials in biotechnology and diverse industrial processes (Vieille & Zeikus, 2001; Haki & Rakshit, 2003; van den Burg, 2003). In addition, the production of thermophilic recombinant enzymes is facilitated by their relatively straightforward overexpression and purification, which makes them feasible candidates for various biochemical experiments as well as for crystal structure determination. Those factors have enhanced research on thermostability, which has been extensively studied in past years, mainly by

comparisons of structural properties of thermo- and mesophilic enzymes, as well as with mutagenic experiments (for review see Vieille & Zeikus, 2001). In contrast to enzymes from thermophiles, cold adapted enzymes are relatively poorly examined, in particular considering their extensive distribution and occurrence in our biosphere. The organisms occupying the permanently cold areas dominating the majority of earth's surface, collectively called psychrophiles, have to rely on enzymes that can compensate for the low reaction rate at their physiological temperatures. The properties that characterise and distinguish cold adapted enzymes from enzymes of higher temperature origin are their increased turnover rate (k_{cat}) and inherent higher catalytic efficiency (k_{cat}/K_m) at low temperatures (Gerday *et al.*, 1997). It is assumed that optimisation of the catalytic parameters in cold adapted enzymes is accomplished by developing increased structural flexibility, allowing the conformational changes required for catalysis at low temperatures (Hochachka & Somero, 1984).

In recent years, a few crystal structures of cold adapted enzymes have been determined (Smalås *et al.*, 1994; Aghajari *et al.*, 1998; Alvarez *et al.*, 1998; Russel *et al.*, 1998; Kim *et al.*, 1999; Aghajari *et al.*, 2003; Leiros *et al.*, 2003; Van Petegem *et al.*, 2003; Bae & Phillips, 2004). These structures have served as basis in comparative studies on structural aspects of cold adaptation. Additionally, information from site directed mutagenesis experiments, homology modelling and directed evolution have been used in the effort of shedding light on the molecular principles underlying adaptation of enzymes to low temperatures (Kano *et al.*, 1997; Narinx *et al.*, 1997; Taguchi *et al.*, 1998; Miyazaki *et al.*, 2000; Sheridan *et al.*, 2000; Wintrode *et al.*, 2000; D'Amico *et al.*, 2001; Tindbaek *et al.*, 2004).

In general, regardless of whether research is directed at thermo- or psychrophilic adaptation, the results primarily show that each protein family adopts its own strategies for coping at extreme temperatures. Although no general rules have been found to apply in temperature adaptation of enzymes, some structural tendencies have emerged. The most frequently reported features related to temperature adaptation, going from higher growth temperature to lower, are a reduced number of non-covalent intra- and intermolecular interactions, a less compact packing of the hydrophobic core, an increased apolar surface area, decreased metal ion affinity, longer surface loops and a reduced number of prolines in loops (Gerday *et al.*, 1997; Aghajari *et al.*, 1998;

Jaenicke & Böhm, 1998; Smalås *et al.*, 2000; Vieille & Zeikus, 2001; Kristjánsson & Ásgeirsson, 2002; Feller & Gerday, 2003). In general, a correlation is seen in naturally occurring enzymes between catalytic efficiency at low temperatures and susceptibility towards heat and other denaturants (Feller, 2003). Still, with methods of directed evolution, mutants have been obtained with changes in one of the properties, stability or catalytic efficiency, irrespective of the other indicating that these properties are not essentially intertwined (Miyazaki *et al.*, 2000; Wintrode *et al.*, 2000). The observed instability of cold adapted enzymes is regarded not as a selected for property but rather as a consequence of the reduction of stabilising features arising from the requirement of increased flexibility to keep up catalytic efficiency at low temperatures (Fields, 2001). Structural flexibility of cold adapted enzymes is as yet a poorly defined term of which little direct experimental evidence is available on. Attempts to assess and compare structural flexibility of a psychrophilic α -amylase and more thermostable homologues by dynamic fluorescence quenching supported the idea of an inverse correlation between protein stability and structural flexibility (D'Amico *et al.*, 2003). Comparisons of hydrogen-deuterium exchange rates as an estimate of flexibility in enzymes of different temperature origin (Svingor *et al.*, 2001) have supported the idea of “corresponding states” (Jaenicke, 2000), which assumes that at their physiological temperatures, enzymes possess a comparable flexibility and a structural stability adequate to maintain their active conformation.

Thermodynamics aspects of stability

Protein stability is a term that refers to the resistance of an active conformation towards denaturation. Protein stability depends on the difference in free energy (ΔG) between the native (N) and the unfolded (U) state. The stability (ΔG) of a protein that denatures in a reversible two state manner is expressed as a function of temperature according to the modified Gibbs–Helmholtz equation:

$$\Delta G_{N \rightarrow U} = \Delta H_{N \rightarrow U} - T\Delta S_{N \rightarrow U} + \Delta C_p(T - T_m - T \ln T / T_m) \quad \text{Equation 1}$$

where ΔH and ΔS are changes of enthalpy and entropy, respectively, upon unfolding (N \rightarrow U) at a reference temperature (here melting temperature, T_m). The heat capacity difference (ΔC_p) is taken as a constant and is larger than zero. The Gibbs-Helmholtz equation describes a parabolic curve, termed the protein stability curve (Figure 1). The

stability curve predicts that proteins have maximum stabilities at a temperature (T_{\max}), when $\Delta S = 0$. The stability decreases on both sides of the maximum and intersects the temperature axis twice ($\Delta G = 0$), at temperatures of cold and heat denaturation. The stability curve reflects important properties of the protein; its curvature reflects heat capacity change ($\delta^2\Delta G/\delta T^2 = -\Delta C_p/T$) and its slope shows entropy change ($\delta\Delta G/\delta T = -\Delta S$) of unfolding at a given temperature.

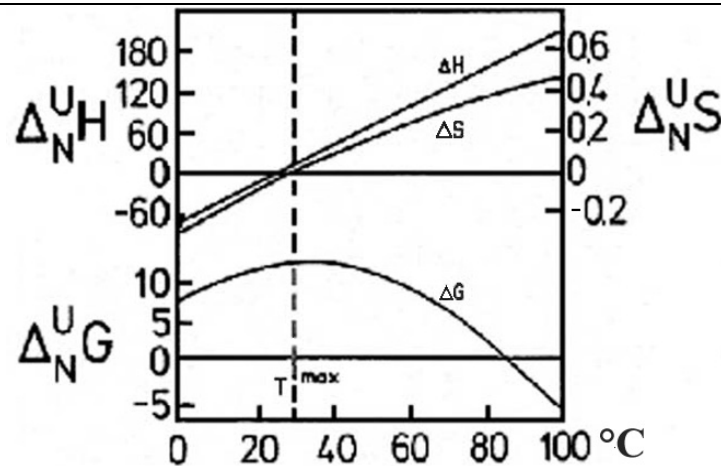


Figure 1. Temperature dependent enthalpy-, entropy and free energy changes upon unfolding of myoglobin.

Unfolding at low temperatures can be either exo- or endothermic. Enthalpy of transfer of polar groups from the protein interior into water is positive at low temperatures and negative at higher temperatures (Makhatadze & Privalov, 1993). This is due to the polar groups creating their own ordered water, which generates a negative enthalpy change due to the increased molecular interactions. Counterbalanced against this is the positive enthalpy arising from the energy requiring disruption of polar interactions in the native state and of the pre-existing water structure. The positive enthalpy contribution due to transfer of polar groups to the solvent is more significant at lower temperatures since water naturally has more structure at lower temperatures than at higher. In contrast, the enthalpy of transfer of non-polar groups from the protein interior into water is negative at lower temperatures because non-polar groups enhance pre-existing order. At higher temperatures, this order is lost and the creation of clathrate structures requires enthalpic input. Thus, there is an overall positive enthalpy of unfolding at higher temperatures. In other words, whether unfolding at low temperatures is endo- or exothermic, depends on which weighs more, the favourable interactions formed between polar groups and the

solvent and non-polar groups enhancing the ordered solvent structure or the unfavourable breaking of the pre-existing solvent structure and disruption of interactions between polar groups. In any way, the enthalpy change of unfolding becomes increasingly positive with higher temperature.

Entropy changes of unfolding can be divided into three categories (Edgcomb & Murphy, 2000):

- i. configurational, due to rotations about single bonds (ΔS_{conf}),
- ii. due to changes in the degree of order of the solvent molecules (ΔS_{solv}),
- iii. due to mixing (ΔS_{mix})

Changes in heat capacity are mainly the result of different solvent exposure of non-polar groups. Upon unfolding, hydrophobic groups, which in the native state are buried in the protein core, are transferred to the solvent causing a decrease in entropy (ΔS_{solv}) due to increased order of water structure. This is reflected in a large positive change in heat capacity (Privalov & Makhatadze, 1993). At temperatures where unfolding is endothermic the driving force of unfolding is increased configurational entropy (ΔS_{conf}).

Figure 1 shows well the temperature dependence of the effects of enthalpy and entropy changes on protein stability. In this example, unfolding at low temperature is exothermic and the folded protein structure is supported by the decrease in entropy upon unfolding. On approaching the temperature of maximal stability the entropy factor becomes less significant and the structure stability relies on the positive enthalpy change of unfolding. At temperatures above maximum stability the enthalpy factor contributes to stability whereas the entropy change has now an opposite sign as compared to low temperatures and acts as a destabilising factor.

Marginal stability

All conceivable forces and interactions between atoms of the peptide chain and solvent molecules contribute to the folding of a protein. All these small forces are temperature dependent and have different effects on enthalpy and entropy changes that stabilise or destabilise the protein structure. Proteins, due to their delicate balance of stabilising and

destabilising interactions, are only marginally stable if physiological conditions are considered as the standard state. Their free energies of stabilisation are minute compared to the total molecular energy. Molecular adaptation to extremes of physical conditions at the protein level only requires small alterations of intramolecular interactions. This is reflected in small differences of stability, between mesophilic proteins and related proteins from thermo- and hyperthermophiles, not exceeding 100 kJ/mol, which is merely equivalent to a few non-covalent interactions (Jaenicke & Böhm, 1998). According to experimental evidence, thermophilic proteins are not essentially more stable, i.e. have a higher maximum value of free energy of stability, than mesophilic counterparts (Beadle *et al.*, 1999). A theoretical model describing three types of deviations of the protein stability curve has been proposed to account for the increased stability of thermophilic proteins compared to their relatives adapted to lower temperatures (Figure 2). First, the stability curve can be translated vertically towards a higher stability over the whole temperature range. Second, it can be translated horizontally resulting in higher melting point and higher stabilities at temperatures above maximum stability. Third, the profile of thermophilic stability curve can be flattened reflecting a similar stability around the maximum stability but increased stability to both sides yielding a broader temperature range of stability (Jaenicke & Böhm, 1998; Beadle *et al.*, 1999).

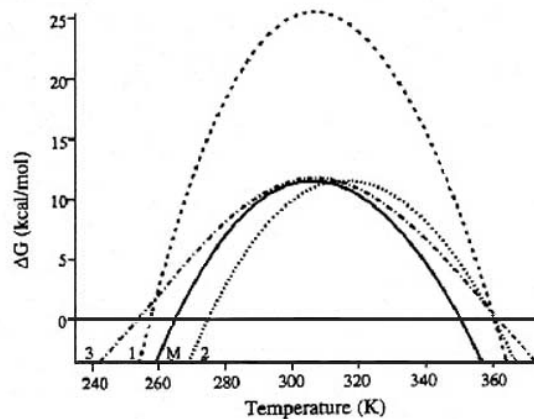


Figure 2. Three plausible mechanisms for the deviation of stability curve of thermostable proteins relative to curves of mesophilic proteins (solid curve M). Curve 1 (dashed): the thermophilic protein is stabilised across the temperature range and has a greater maximum stability. Curve 2 (dotted) is shifted towards a higher temperature but otherwise the relationship ΔG vs. T is similar to the mesophilic case. Curve 3 (dash-dot) is flattened relative to the mesophile's curve, reflecting less temperature dependent stability and a shift towards higher and lower temperature of heat and cold denaturation, respectively (Beadle *et al.*, 1999).

Serine proteinases

Proteinases (by definition: endopeptidases) catalyse the hydrolysis of peptide bonds in peptides and proteins. Proteinases are divided into four classes named after the characteristic group in the catalytic centre: serine-, cysteine, aspartic- and metallo proteinases. Serine proteinases are the most extensively studied group of proteinases. Their catalytic triad consists of aspartic acid, histidine and a serine residue. The cleavage of the peptide bond is a two step reaction (Figure 3).

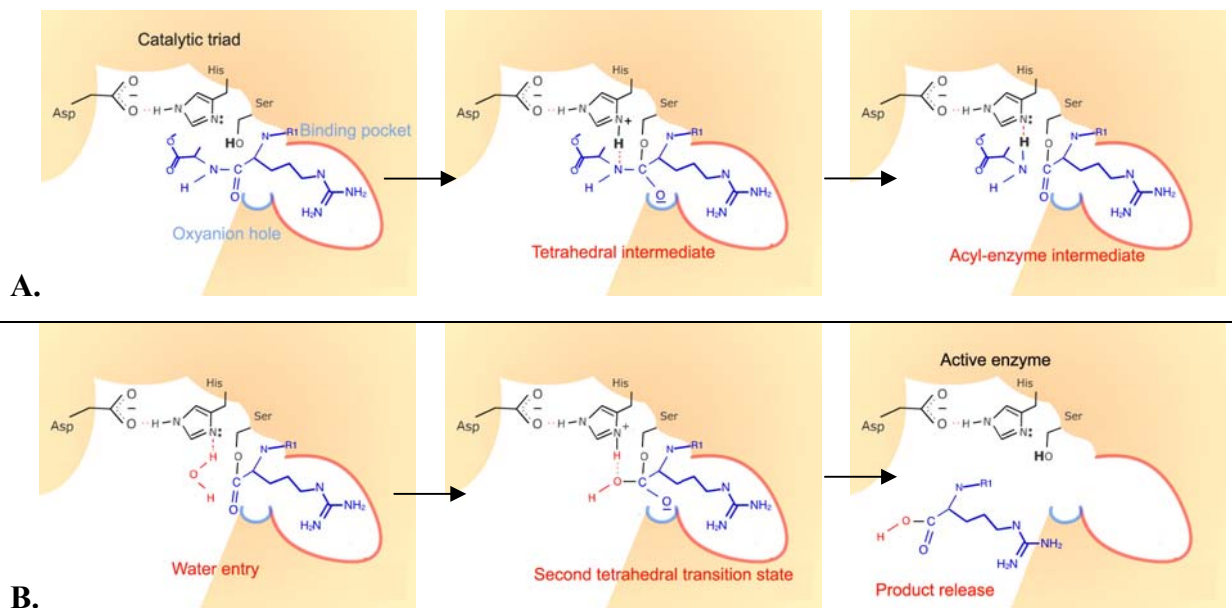


Figure 3. Schematic drawing of the reaction mechanism of a serine proteinase catalysed hydrolysis of a peptide bond. A. Formation of acyl-enzyme intermediate. Histidine of the catalytic triad abstracts a proton from the serine hydroxyl group, of which oxygen makes a nucleophilic attack on the peptide carbonyl, resulting in the formation of a tetrahedral intermediate. The acyl-enzyme intermediate is generated upon transfer of a proton from the histidine to the leaving group. B. Deacylation. This second step is practically a reverse first step with a water molecule in the role of the amino-group of the substrate. Histidine abstracts a proton from a water molecule and activates it to make a nucleophilic attack on the carbon of the acyl-enzyme intermediate. Again a tetrahedral intermediate is formed. Histidine delivers the proton to the serine residue and the product is released. The figure was adapted from <http://telstar.ote.cmu.edu/biology/animation/SerineProtease/biochem.html>.

In the first step an acyl enzyme intermediate is formed between the substrate and the essential serine. Formation of this covalent intermediate proceeds through a negatively charged tetrahedral transition state intermediate and then the peptide bond is cleaved. During the second step or deacylation, the acyl-enzyme intermediate is hydrolyzed by a water molecule to release the peptide and to restore the serine-hydroxyl of the enzyme. The deacylation, which also involves the formation of a tetrahedral transition state intermediate, proceeds through the reverse reaction pathway of acylation. A water molecule is the attacking nucleophile instead of the serine residue. The histidine residue

provides a general base and accepts the hydroxyl group of the reactive serine residue. In order to form a stable tetrahedral intermediate, the enzyme needs to form a hydrogen bond with a negatively charged oxygen atom at the cleavage site in the oxyanion hole. The serine proteinases exhibit different substrate specificities, which are dependent on the types of residues in the various enzyme subsites interacting with the substrate residues. The nomenclature to describe the interaction of a substrate with a proteinase was introduced by Schechter & Berger (1967) and is now widely used in the literature. By convention, the subsites on the proteinase are called S (for subsites) and the substrate amino acid residues are called P (for peptide). The amino acid residues of the N-terminal side of the scissile bond are numbered P3, P2, P1 and those of the C-terminal side are numbered P1', P2', P3', etc. (Figure 6). The P1 or P1' residues are those residues located next to the scissile bond. The subsites on the proteinase that complement the substrate binding residues are numbered S3, S2, S1, S1', S2', S3', etc.

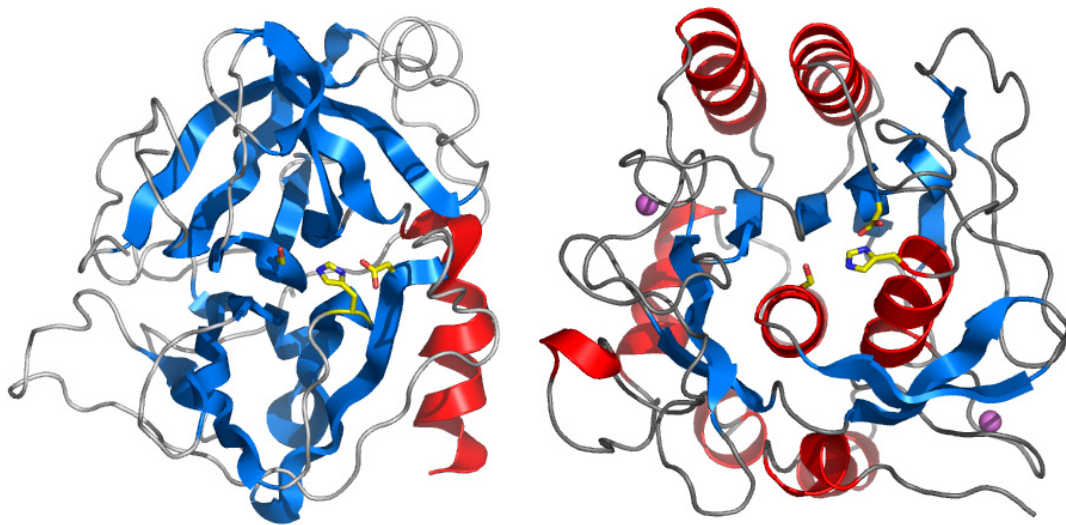


Figure 4. The two superfamilies of serine proteinases, chymotrypsin- (left) and subtilisin like (right) proteinases have different protein scaffolds, although they have similar active site geometry and catalyse cleavage of the peptide bond via the same mechanism. The chymotrypsin like enzymes are characterised by the β/β motif, while the subtilases have an α/β motif. The catalytic triad residues are coloured yellow and the calcium ions associated with the subtilisin are shown in purple. The enzymes representing the two clans are bovine γ -chymotrypsin (PDB ID: 2GMT) and subtilisin from *Bacillus lentus* (PDB ID: 1ST3).

Serine proteinases are divided into two main superfamilies: the chymotrypsin like and the subtilisin like serine proteinases. The chymotrypsin family includes well known mammalian enzymes such as chymotrypsin, trypsin and elastase. The subtilisin like serine proteinases, also called subtilases, were for a long time considered to be solely bacterial enzymes but have now been found to occur in archaea, plants, fungi and

animals. The two clans, chymotrypsins and subtilases, are distinguished by highly similar active site geometry and the catalysis proceeds via the same mechanism, however in radically different β/β (chymotrypsin) and α/β (subtilisin) protein scaffolds (Figure 4).

Subtilases

The majority of the subtilases are extracellular enzymes synthesised as a precursor with a pre- and pro-sequence extension of the N-terminus of the mature protein (Bryan *et al.*, 1995). The pre-sequence acts as a signal peptide, driving translocation through the cell membrane, whilst the pro-sequence acts both as an intramolecular chaperone that guides the correct folding of the mature protein and as a protease self-inhibitor (Yabuta *et al.*, 2003). The pro-sequence is usually cleaved from the mature protein by autoproteolysis to produce active mature protease. In addition, some members of the family have C-terminal extensions that display a variety of properties, such as sequence repeats, cysteine rich regions or transmembrane segments. Based on sequence homology of the catalytic domain, subtilases are currently divided into 6 families (Figure 5) (Siezen & Leunissen, 1997):

- i. Subtilisin: Apart from the true subtilisins, which stem from *Bacillus* species, this family includes subgroups of high-alkaline- and intracellular proteinases from microorganisms. Several crystal structures of subtilisins are known.
- ii. Thermitase: Enzymes found only in microorganisms including thermo- and halophiles. The crystal structure of the representative enzyme for this group, thermitase, is known.
- iii. Proteinase K: A large family of secreted proteinases from fungi, yeasts and gram-negative bacteria. Proteinase K was until now the only enzyme of known three-dimensional structure in this family.
- iv. Lantibiotic peptidases: A small family of gram positive bacterial enzymes highly specific for cleaving leader peptides of lantibiotics, a group of antimicrobial peptides.
- v. Kexin: A large group of enzymes involved in activation of peptide hormones, growth factors, viral proteins and other peptides. Most known kexins stem from eukaryotes.
- vi. Pyrolysins: Heterogeneous group of enzymes of varied origin and low sequence conservation

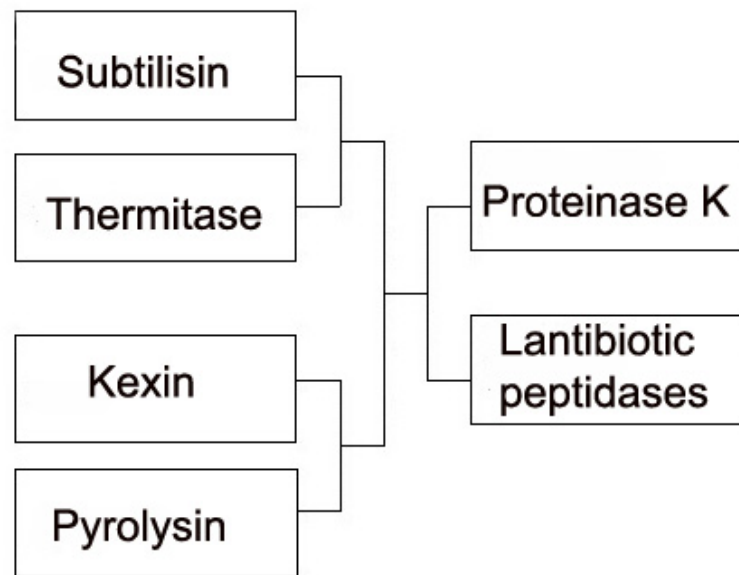


Figure 5. A general layout of the relationship between the six families of subtilisin like serine proteinases.

The substrate binding region of subtilases resembles a surface channel that can accommodate at least six amino acid residues (P4 – P2') of a substrate or an inhibitor. Both backbone and sidechain interactions contribute to substrate binding. The substrate residues P4-P1 are bound by hydrogen bonds to the backbone of two beta-strands and form the central strand of a triple stranded antiparallel beta sheet (Figure 6). Interactions of the side chains of the P4 – P1 and S4 – S1 of the substrate and enzyme, respectively, largely determine substrate specificity. However, the type of S4 and especially S1 appears to dominate in substrate preference at least in subtilisins (Siezen & Leunissen, 1997). In general subtilases have a relatively broad specificity with preference for big hydrophobic groups in the P1 site (Perona & Craik, 1995).

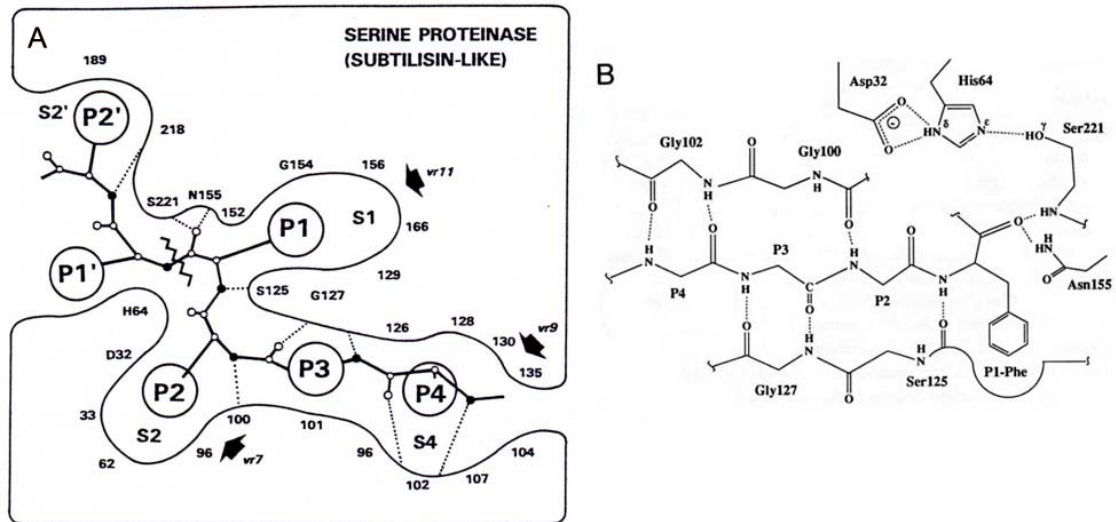


Figure 6. Substrate binding in subtilases. A. Schematic drawing of the interactions between substrate and enzyme subsites. The numbers of residues relate to Subtilisin BPN'. Nomenclature P4 – P2' and S4 – S2' is according to Schechter and Berger (1967). Hydrogen bonds are shown as dotted lines and the scissile bond is shown by a jagged line. (Siezen & Leunissen, 1997) B. Formation of a triple beta-sheet upon substrate binding. (Perona & Craik, 1995).

Calcium binding is a universal feature of all known subtilases. Calcium binding has been shown to be fundamental for structural stability of subtilases enhancing resistance towards denaturing agents and reducing rate of autolysis. Furthermore, stabilizing mutations in subtilases are usually calcium-dependent in their stabilizing effect, irrespective of their proximity to the calcium binding sites (Bryan, 2000). Four calcium binding sites were identified in the first crystal structures of subtilases, subtilisin, proteinase K and thermitase (Siezen & Leunissen, 1997). These were termed Ca1-Ca4, defined as strong, medium and weak calcium binding sites and found, according to sequence alignments, to be conserved in different subtilase families. Thus, a strong calcium binding site, Ca1, and a weak one, Ca3, were considered to be present in most subtilisins, thermitases and kexins, whereas a medium strength one, Ca2, was less common. Ca4 was only found in proteinase K family members. Since the first calcium binding sites were identified, many more have been reported and up to five calcium ions have been found associated with a single subtilisin molecule, none of which correlates with the conserved calcium binding sites described previously (Almog *et al.*, 2003).

Proteinase K and Thermitase

Proteinase K from the fungus *Tritirachium album* Limber was the first enzyme with proteolytic activity to be isolated from that genus and got its name from its ability to digest native keratin (Ebeling *et al.*, 1974; Betzel *et al.*, 1988). It was soon clear that

proteinase K had great potentials for basic and applied research due to its broad peptide cleavage activity and unusual stability over a wide range of temperatures and pH values and even at low concentrations of SDS and urea. Proteinase K is now widely used for facilitating nucleic acid isolation by degradation of contaminating proteins in cell lysates and for inactivation of enzymes such as DNase and RNase without recourse to a denaturation process. Proteinase K was extensively studied not least due to its prospects for application and its structure was determined by X-ray crystallography in 1984 (Paehler *et al.*, 1984). Interestingly, the three-dimensional structure of proteinase K was known prior to its primary structure. Thus, the classification of proteinase K as subtilisin like serine proteinase was originally based on its three-dimensional structure and was later confirmed by sequence analysis (Jany *et al.*, 1986). The structure of a subtilisin like enzyme from a thermophile, thermitase from *Thermoactinomyces vulgaris* was first determined in 1989 (Gros *et al.*, 1989). Proteinase K and thermitase were assigned to a subgroup characterised by a free cysteine at equivalent position in the sequence and located near the active histidine. Other known subtilases at that time were cysteine free. According to current classification proteinase K and thermitase each represent one of the six subtilase families.

There are two calcium ions associated with the native structure of proteinase K. First, there is a tightly bound calcium ion in a well defined calcium binding site close to but not directly equivalent to a conserved calcium binding sites in members of the subtilisin family. The second more mobile calcium ion bridges two loops close to the amino- and carboxy termini and is not conserved. Proteinase K is very dependent on calcium binding for structural stability. Removal of calcium ions from proteinase K did not affect its proteolytic activity but reduced its stability drastically (Muller *et al.*, 1994), Kristjánsson, personal communication). A study on the three-dimensional structure of a calcium free proteinase K revealed subtle but concerted changes relative to the native structure, mostly affecting loops and turns while helices and strands are structurally invariant (Muller *et al.*, 1994). As an overall effect of calcium depletion, the proteinase K structure was reported to become more open, reflected by an increase in solvent accessibility.

Three calcium binding sites have been identified in thermitase. The first is equivalent to the major subtilisin calcium binding site, Ca1 (see above). The second calcium binding

site is bound to an extended surface loop. The residues coordinating the second calcium binding site are conserved in several enzymes including members of the proteinase K family, such as the cold adapted *Vibrio*-proteinase and its most related meso- and thermophilic homologs. The third calcium ion was found in a site only 2.5 Å from a monovalent cation binding site (Gros *et al.*, 1991). At 100 mM CaCl₂ concentration the site is occupied by calcium but at lower CaCl₂ concentrations the monovalent binding site is occupied by potassium or sodium. The differences in calcium binding, regarding number of sites and especially increased number of protein ligands in thermitase, whereas water more often serves as a ligand in proteinase K, was suggested to be a likely source of the enhanced stability of thermitase (Betzl *et al.*, 1988; Teplyakov *et al.*, 1990).

The cold adapted *Vibrio*-proteinase

A psychrotrophic *Vibrio* species, termed *Vibrio* strain PA-44, isolated from a cold marine source secretes a proteinase, which belongs to the proteinase K family of subtilases (Kristjánsson *et al.*, 1999). The bacterium has an optimum growth temperature at 19 °C and production of the *Vibrio*-proteinase (VPR) coincides with the growth temperature (Alfredsson *et al.*, 1995). The VPR-gene encodes a 530 amino acid pre-pro-enzyme consisting of an N-terminal pre-prosequence, a proteinase or catalytic domain, and a C-terminal extension. The N-terminal pre-prosequence comprises 139 amino acids and probably functions as a molecular chaperone for correct folding, but is subsequently cleaved off by autolysis to give the active protein (Lee *et al.*, 1991; Eder *et al.*, 1993). The C-terminal extension is also present in related proteins and is suggested to facilitate secretion (Terada *et al.*, 1990). Recombinant VPR produced in *E. coli* is, like the wild type enzyme from the *Vibrio* strain PA44, isolated in the form of a 40 kDa protein without the 139 residues N-prosequence. Under relatively mild conditions, the enzyme undergoes further autolysis, cleaving off the C-terminal extension leaving a 30 kDa mature proteinase (Figure 7).

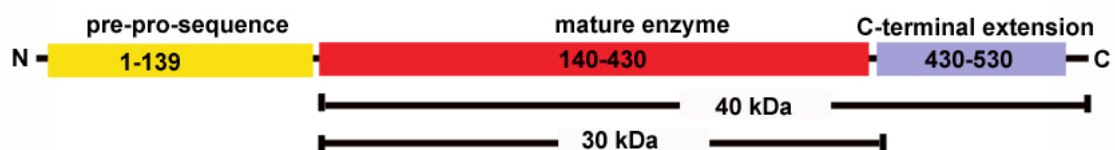


Figure 7. VPR is produced as a pre-pro enzyme.

VPR was purified from cultures of *Vibrio* and selected enzymatic properties were compared with those of related enzymes of thermo- and mesophilic origin, aqualysin I from *Thermus aquaticus* and proteinase K from *Tritirachium album* Limber (Kristjánsson *et al.*, 1999). The study revealed a clear trend of protein stability and catalytic efficiency related to the temperature adaptation of the host organism. Thus, the catalytic activity of the cold adapted VPR was higher for amidase activity over a temperature range of 15 - 55 °C. The stability of VPR both towards heat and denaturants was significantly lower than that of both proteinase K and aqualysin I. Furthermore, disulfide cleavage, either by reduction with dithiothreitol or by sulfitolysis, led to a loss in VPR activity. Under the same conditions, aqualysin I was also partially inactivated by dithiothreitol, but the activity of proteinase K was unaffected. The disulfides of proteinase K and aqualysin I were reactive to sulfitolysis only under strong denaturing conditions, while all disulfides of VPR reacted in absence of a denaturant. The same trend was observed for inactivation of the enzymes by H₂O₂-oxidation. H₂O₂ oxidises the methionine adjacent to the active site serine residue in the subtilisin-, thermitase- and proteinase K family, which leads to inactivation of the enzyme. The difference in reactivity in these reactions is probably due to different accessibility of the reactive cysteine or methionine and might reflect different degree of flexibility in the proteins.

The mature VPR has 86 % and 70 % sequence identity to proteinases from the mesophilic bacteria *Vibrio alginolyticus* and *Vibrio cholerae*, respectively. The VPR sequence is 60 % identical to two thermophilic proteinases, aqualysin I from *Thermus aquaticus* and proteinase from *Thermus* sp. Rt41a. VPR and proteinase K have 41 % sequence identity. A thorough sequence- and computer model comparisons between VPR and the most related enzymes of meso- and thermophilic origin revealed some trends possibly relevant to temperature adaptation regarding the amino-acid composition and amino acid substitutions among these enzymes (Arnórsdóttir *et al.*, 2002). All the enzymes have high Ser content, the cold adapted VPR the highest. The thermophilic enzymes contain a higher Arg/Lys ratio and a lower frequency of Asn and Gln residues than the other enzymes. In addition, the thermophilic enzymes have a relatively higher number of hydrophobic amino acids, as well as a higher aliphatic index (Arnórsdóttir *et al.*, 2002). A noticeable trend was seen in the number of Ala, increasing on going from the cold adapted *Vibrio*-proteinase to the thermophilic enzymes. A closer look on individual amino acid substitutions revealed a striking number of the

replacement of Ala in the more thermostable enzymes, in particular in aqualysin, for Ser in the cold adapted *Vibrio*-proteinase. This Ser-Ala exchange has been observed as highly correlated to thermostability in other comparative studies (Argos *et al.*, 1979; Peek *et al.*, 1992; Haney *et al.*, 1999). Another feature related to thermostability was the more frequent occurrence of proline residues in surface loops of the thermostable enzymes. Proline residues are assumed to stabilise protein structures by stabilising interactions between structural elements, e.g. hydrogen bonds between adjacent beta strands (Vieille & Zeikus, 2001). Prolines in loops also reduce the force of entropy driven unfolding by reducing the conformational degree of freedom relative to other residues more in the unfolded state than in the folded protein. The results of the comparison gave rise to ongoing mutagenic research where single and combined amino acid substitutions aimed at increasing the stability of VPR are tested. The structure determination of VPR facilitates rational design of mutants and interpretation of mutagenic experiments.

Posttranscriptional RNA processing

The central dogma, which forms the backbone of molecular biology, states that the flow of genetic information is from DNA to RNA to protein. Three major processes are responsible for the inheritance and conversion of genetic information from one form to another: replication, transcription and translation. With the discovery of reverse transcriptase in retroviruses (Baltimore, 1970), the central dogma was extended to also account for the reverse flow of information, from RNA to DNA. The findings of posttranscriptional processing of RNA transcripts: splicing of eukaryotic mRNA (Berget *et al.*, 1977; Chow *et al.*, 1977) and alterations of RNA by the so-called RNA-editing (Benne *et al.*, 1986) further expanded the central dogma. In order to improve our understanding of the machineries that drive these complex processes, attempts were made to crystallise critical components of the spliceosome and the RNA editosome. A human 61 kDa protein involved in the assembly of spliceosomal subunits and a 42 kDa protein component of the *Trypanosoma Brucei* editosome, endowed with both endo- and exonuclease activity, were produced in *E. coli* and purified for crystallisation trials.

Splicing

Eukaryotic genes contain segments of noncoding sequences, called introns, which break up the amino acid coding sequences, the exons. The introns are removed from the precursor mRNA, and the exons are spliced together to produce the translatable mRNA. The splicing mechanism is catalysed via a highly dynamic ribonucleoprotein complex, the spliceosome. The spliceosome is formed from the association of five uridine-rich small nuclear RibonucleoProtein particles, the U1, U2, U4, U5 and U6 snRNPs, and an undefined number of other splicing factors with the pre mRNA (Will & Lührmann, 1997; Jurica & Moore, 2003).

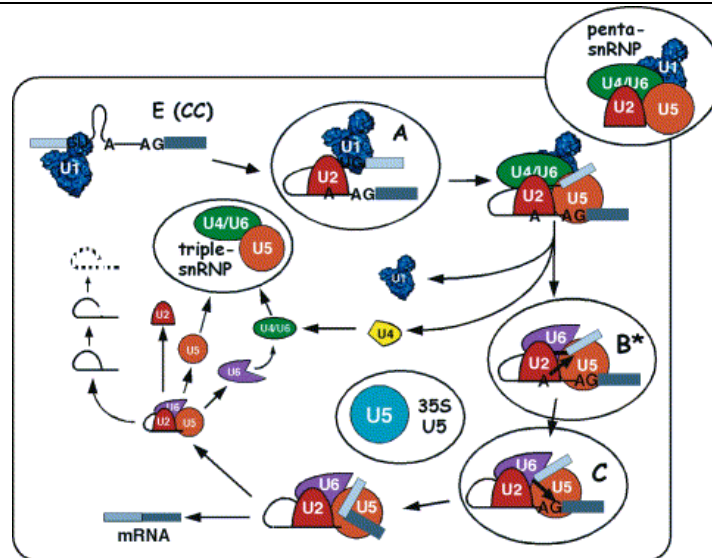


Figure 8. The spliceosomal assembly cycle derived from *in vitro* splicing experiments, where individual complexes were subject to proteomics analysis and mass spectrometry (Jurica & Moore, 2003).

Consensus models of the splicing cycle describe spliceosomal assembly as an ordered process with a series of short-lived intermediate complexes (reviewed in Staley & Guthrie, 1998; Hastings & Krainer, 2001; Jurica & Moore, 2003). The spliceosomal assembly cycle is depicted in (Figure 8). The initial step in the assembly of the spliceosome is the formation of complex E by recognition and interaction of U1 snRNP to the 5' splice site of pre-mRNA. The formation of complex E also involves some non-snRNP splicing factors and the U2 auxiliary factor (U2AF), which binds to the pyrimidin tract upstream of the 3' splice site. Upon binding of U2 snRNP to the branch point, complex A is formed. Subsequently, the U4, U5 and U6 snRNPs, already associated in a complex, the [U4/U6-U5] triple snRNP, is integrated to form the spliceosome or complex B. Immediately after the integration of the [U4/U6-U5] triple snRNP into the spliceosome, U6 and U4 RNA dissociate, the U4 snRNP is released

from the complex and new base pairing interactions are formed between U2 and U6 snRNA. Simultaneously the U1 snRNP leaves the 5' splice site, which is then recognised by a conserved sequence of the U6 snRNA. The resulting highly remodelled structure is the catalytically active spliceosome, complex B* (Makarov *et al.*, 2002). The splicing catalysis is a two step transesterification reaction. In the first step the 2'-hydroxyl of the branch point adenosine attacks the 5'-splice site producing 5'-exon and a branched intron lariat-3'-exon intermediates (Figure 9) and complex C is formed. In the second step the 3' hydroxyl of the 5'-splice site attacks the 3' splice site to give the spliced mRNA and an excised lariat intron.

In recent years some evidence has emerged for an earlier association of some snRNPs with pre-mRNA than previously thought (see for examples and discussion: Das *et al.*, 2000; Maroney *et al.*, 2000; Hastings & Krainer, 2001; Will & Lührmann, 2001; Nilsen, 2002; Stevens *et al.*, 2002; Jurica & Moore, 2003). Thus, the U2 snRNP was found to be present in a functionally active complex E (Das *et al.*, 2000) and the tri-snRNP was shown to interact with the 5' splice site prior to assembly of complex A (Maroney *et al.*, 2000). Furthermore, the isolation of a penta-snRNP complex has been reported (Stevens *et al.*, 2002). This might indicate that the splicing cycle takes place on a preassembled spliceosome and that the assembly intermediates in the model described above might just reflect key structural changes that are stabilised as the splicing machinery proceeds through splice site recognition and catalysis (Nilsen, 2002). Thus, although these observations require some revision of our ideas about the splicing cycle they do not violate the conclusions drawn from previous data, holding that interactions between spliceosomal components change in an ordered fashion throughout the splicing cycle.

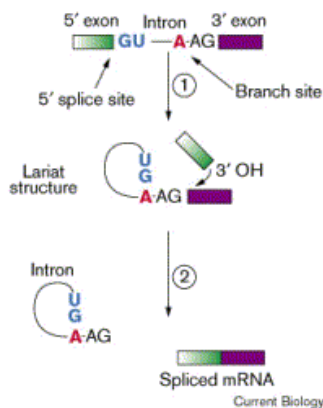


Figure 9. Splicing of mRNA precursors involves two successive transesterification reactions. (1) the 2'OH of a specific adenosine (red) at the branch site attacks the 5' splice site (blue). This reaction releases the 5' exon (green; with a 3' OH terminus) and leaves the 5' end of the intron (blue) joined by a 2'- 5' phosphodiester bond to the branch site adenosine (red); this intron-3' exon intermediate is therefore in the form of a lariat. (2) the 3' OH of the 5' exon intermediate (green) attacks the 3' splice site, producing the spliced mRNA and lariat-shaped intron products. (Newman, 1998)

The [U4/U6·U5] triple snRNP

A major component of the spliceosome is the [U4/U6·U5] tri-snRNP. The three-dimensional structure of complex B, determined at 40 Å resolution by cryo-electron microscopy, is similar in size and shape to the isolated tri-snRNP, indicating that the main constituent of complex B is the tri-snRNP (Boehringer *et al.*, 2004). The tri-snRNP is especially protein rich. At least 29 proteins have been identified in the human tri-snRNP (Makarova *et al.*, 2002), most of which are highly conserved and orthologous proteins have been identified and isolated from yeast tri-snRNP (Gottschalk *et al.*, 1999). Upon integration of the tri-snRNP into the prespliceosome, extensive remodelling of RNA/RNA- and RNA/protein interactions is initiated; resulting in the formation of the catalytically active spliceosome. These rearrangements involve different parts of the tri-snRNP and are highly conserved from yeast to vertebrates (Mougin *et al.*, 2002). This indicates that the structures of RNAs and RNA-protein interactions in the tri-snRNP are designed to accomplish the correct rearrangement of RNA-RNA base pairing in the spliceosomal complex during the assembly, activation and splicing reaction. Although spliceosomal activation is relatively well characterised at the RNA level, the role of proteins and RNPs is still poorly understood. Thus a thorough understanding of the architecture of the tri-snRNP is of great importance to elucidate the mechanisms underlying the structural changes involved in activation of the spliceosome. In each round of splicing the components of the tri-snRNP are reassembled to form a stable 25S complex. The assembly of the tri-snRNP is considered to take place in the Cajal bodies (Schaffert *et al.*, 2004), nuclear compartments characterised by the protein coilin. The association of the components of the tri-snRNP is mainly mediated by protein/protein interactions. The interactions that mediate the formation of the tri-snRNP are as yet poorly understood. Still, utilising X-ray crystallography and cryo-electron-microscopy along with classical biochemical experiments, much progress has been made in recent years analysing the components of the tri-snRNPs and identifying their roles. There among is a 15.5 kDa protein that is considered to initiate the assembly pathway of the U4/U6 di-snRNP. The structure of the 15.5 kDa was determined by X-ray crystallography (Vidovic *et al.*, 2000). It binds to the 5' stem loop of the U4 snRNA and serves as a platform for the subsequent association of the remaining U4/U6 RNP proteins. The binding of the 15.5 kDa protein to the U4 RNA is indispensable for the binding of a 61 kDa protein and a subsequent integration of a heteromeric complex consisting of a 20 kDa (CypH), 60 kDa (hPrp4)

and a 90 kDa (hPrp3) protein (Nottrott *et al.*, 1999; Nottrott *et al.*, 2002). The 61 kDa protein, hereafter termed p61, has been shown to be crucial for the formation of the [U4/U6·U5] tri-snRNP both *in vitro* (Makarova *et al.*, 2002) and *in vivo* (Schaffert *et al.*, 2004).

Protein p61

The protein p61 is homologous to the *Saccharomyces cerevisiae* pre RNA processing factor, Prp31p (Maddock *et al.*, 1996) and Spp13p from *Schizosaccharomyces pombe* (Kaufer & Potashkin, 2000). A central domain of p61 also shares homology with Nop56 and Nop58 (Makarova *et al.*, 2002), constituents of the box C/D snoRNP (small nucleolar ribonucleoprotein particles), which functions in directing the processing of ribosomal RNA (rRNA). As mentioned above, p61 interacts with the U4/U6 snRNP only in the presence of a 15.5 kDa protein. The 15.5 kDa protein is suggested to induce conformational changes in the secondary structure of U4 RNA required for the recognition and binding of p61 (Nottrott *et al.*, 2002; Cojocaru *et al.*, 2005). Based on the observation that p61 remained attached to U4/U6 di-snRNP at salt concentrations where the U5 snRNP dissociated from the tri-snRNP, it was defined as U4/U6 snRNP specific (Makarova *et al.*, 2002). Furthermore, a 14 amino acid segment of p61, within the Nop56/Nop58 homologous region, was shown to bind to U4 snRNA with cross linking experiments (Nottrott *et al.*, 2002). Interestingly, the 15.5 kDa protein is also a component of the box C/D snoRNP and binds there to an RNA motif almost identical to the 15.5 kDa binding site in U4 snRNA (Watkins *et al.*, 2000). Hence, with the identification of these proteins and their protein-RNA interactions, a link has been found between the pre-mRNA and pre-rRNA processing machineries supporting ideas about common ancestry of the U4 RNP and box C/D snoRNPs.

Protein p61 as well as its yeast homolog Prp31p have been shown by immunodepletion- and complementation studies to be crucial for the assembly of the tri-snRNPs and the subsequent splicing reaction *in vitro* (Weidenhammer *et al.*, 1997; Makarova *et al.*, 2002). Furthermore, Schaffert *et al.* (2004) demonstrated with iRNA experiments the detrimental effects of knocking down p61 for tri-snRNP assembly *in vivo*. Based on yeast-two-hybrid studies, p61 is suggested to specifically interact with a 102 kDa U5 snRNP specific protein (Prp6p yeast homolog) and hence bridge the U4/U6 and U5 snRNPs in the tri-snRNP (Makarova *et al.*, 2002). A knowledge of the three-

dimensional structures of these proteins would greatly enhance our understanding of the interactions involved in the assembly of the tri-snRNP. Moreover, p61 has an important medical aspect since it is linked to autosomal dominant retinitis pigmentosa, a disease that leads to degeneration of the photoreceptors in the eye (Vithana *et al.*, 2001).

With the aim of crystallising and finally determining the structure of p61 with X-ray crystallography, the p61 gene was expressed in *E. coli*, the recombinant protein was purified and set up for crystallisation trials.

RNA editing

RNA editing is a term that was first used in the mid 1980s to describe a novel phenomenon of insertions and deletions of uridine residues in mitochondrial RNAs of trypanosomes. *Trypanosoma* along with *Leishmania* and *Crithidia* and a number of other less familiar genera belong to a large group of parasitic protozoa collectively called the kinetoplastid protozoa. Kinetoplastid protozoa are characterised by a unique cellular organelle that stains strongly with certain dyes. Due to its location at the base of the flagellum, the kinetoplastid was first considered to be involved in the flagellar movement of the protozoa and therefore the name: kineto – movement, plastid – particle containing pigment. In fact the kinetoplastid represents a portion of the single mitochondrion of the cell and the distinct staining is due to its content of a huge mass of mitochondrial DNA. The mitochondrial DNA in the kinetoplastid makes up a catenated DNA network (kDNA), which consists of two types of circles, the maxi- and minicircles (Shapiro & Englund, 1995). The maxicircles are homogenous circular DNA molecules varying in number, 20 – 50, and size, 23 – 36 kbp, in different species. Maxicircles are the functional equivalents of conventional mitochondrial DNA as they encode ribosomal RNAs and mitochondrial proteins, which are involved in energy metabolism, i.e. electron transport and oxidative phosphorylation in the inner membrane of the mitochondrion. The minicircles are found, depending on species, in 5,000 – 10,000 copies of 460 – 2,500 bp long heterogeneous DNA molecules. The genetic function of the minicircles was a mystery and remained unassigned for many years. It was not until the discovery of RNA editing that their role as encoding guide RNA (see below) was elucidated.

The first indications of posttranscriptional changes in trypanosome mitochondrial RNA sequences were the puzzling findings of apparent frame-shifts in the genes of the cytochrome oxidase subunit II (COII) and an NADH ubiquinone oxidoreductase subunit 7 (ND7) (de la Cruz *et al.*, 1984; Hensgens *et al.*, 1984). Benne *et al.* (1986) found four non-encoded uridylylates in the mature transcripts of COII of *Crithidia fasciculata* and *Trypanosoma brucei*. It was demonstrated that the addition of these four uridylylates overcomes the -1 reading frame-shift and results in a translatable transcript. This curious phenomenon was termed RNA editing. Soon, other examples of this very unusual gene organisation were found. Upon sequencing of the maxicircles from *T. brucei*, *L. tarentolae* and *C. fasciculata*, it was evident that the majority of genes had

frame-shifts or lacked initiation or termination codons. Furthermore, regions that encoded recognisable genes in *L. tarentolae* and *C. fasciculate* apparently did not contain any open-reading frames at all in *T. brucei* (Simpson *et al.*, 1987). Subsequent studies uncovered RNA editing, i.e. additions and deletions of uridylates, in 12 of 18 trypanosome mitochondrial mRNA transcripts (Stuart *et al.*, 1997). The extent of editing ranges from the insertion of four uridylates in the COII transcript as described above to the insertion and deletion of hundreds of uridylate residues. One of the most challenging cryptogenes (genes whose transcripts are edited within the coding sequence) was the COIII-gene in *T. brucei*, which is so extensively edited (pan-edited) that the mature mRNA is nearly twice the length of the gene (Feagin *et al.*, 1988). Analysis of the pan-editing of the COIII transcript in *T. brucei* led to the suggestion that RNA editing occurred in an overall 3' to 5' direction (Abraham *et al.*, 1988). This was supported by other studies, where hundreds of partially edited mRNAs of COIII and the moderately edited *Cyb* (apocytochrome) from mitochondrial extracts of *L. tarentolae* and *T. brucei* were sequenced and the sequence patterns analysed (Decker & Sollner-Webb, 1990; Sturm & Simpson, 1990). However, the basic question about the origin of the genetic information added in the editing process still remained unanswered and this posed paradox to the central dogma. To address this question, Blum *et al.* (1990a) performed a computer search on the *L. tarentolae* maxicircle looking for short DNA sequences that could account for edited mRNA sequences. Additional to the Watson Crick base pairing the search also allowed for G-U base pairing, which is known in rRNA and tRNA and this turned out to be the key in this case. Seven short sequences encoding four edited mRNAs were found scattered across the maxicircle (Blum *et al.*, 1990a). This was the first evidence of the existence of genetic information guiding RNA editing. The small mitochondrial RNA-transcripts carrying the editing information were termed guide RNAs (gRNAs). Subsequently, the minicircles were scanned in a similar manner resulting in the first evidence of genetic function of the minicircles as carrying sequence information for the editing of COIII in *L. tarentolae* (Sturm & Simpson, 1990). Following that, minicircle encoded gRNAs were found in *T. brucei* (Corell *et al.*, 1993) and other species as well. Analysis of the gRNA sequences revealed three functional elements: a 5' anchor region cognate to a region just downstream the pre-edited RNA sequence, the guiding sequence specifying the insertion and deletion of Us at the editing sites and a non-encoded oligo-U-tail at the 3' end. Concurrently to the discovery of gRNAs, enzymatic activities involved in

editing, TUTase-, RNA-ligase- and endonuclease activity, were detected in mitochondrial extracts. Based on the knowledge of the gRNAs, the 3' to 5' progression of editing and the presence of these enzymatic activities, a model, the “enzyme cascade” model (Figure 10), for the mechanism of RNA editing was proposed (Blum *et al.*, 1990a). According to the enzyme cascade model, RNA editing occurred in a series of enzyme-catalysed reactions, initiating with specific interactions of the anchor region of gRNA with the pre-edited mRNA. Subsequently, a specific endonuclease cleavage at the first mismatched RNA would create a substrate for the TUTase, which would add one or more uridine residues and finally the two ends of the mRNA would be religated by an RNA ligase (Blum *et al.*, 1990a). This model was consistent with observations but it did not give a satisfactory explanation of the function of the oligo-U-tail. An alternative model, the transesterification model (Figure 10), was proposed, in which the oligo-U-tail had more significance. The transesterification model predicted the existence of intermediate chimeric molecules consisting of gRNAs covalently bound to the pre-edited mRNA at editing sites by the oligo-U-tail, which served as a repository for uridine residues (Blum *et al.*, 1991; Cech, 1991).

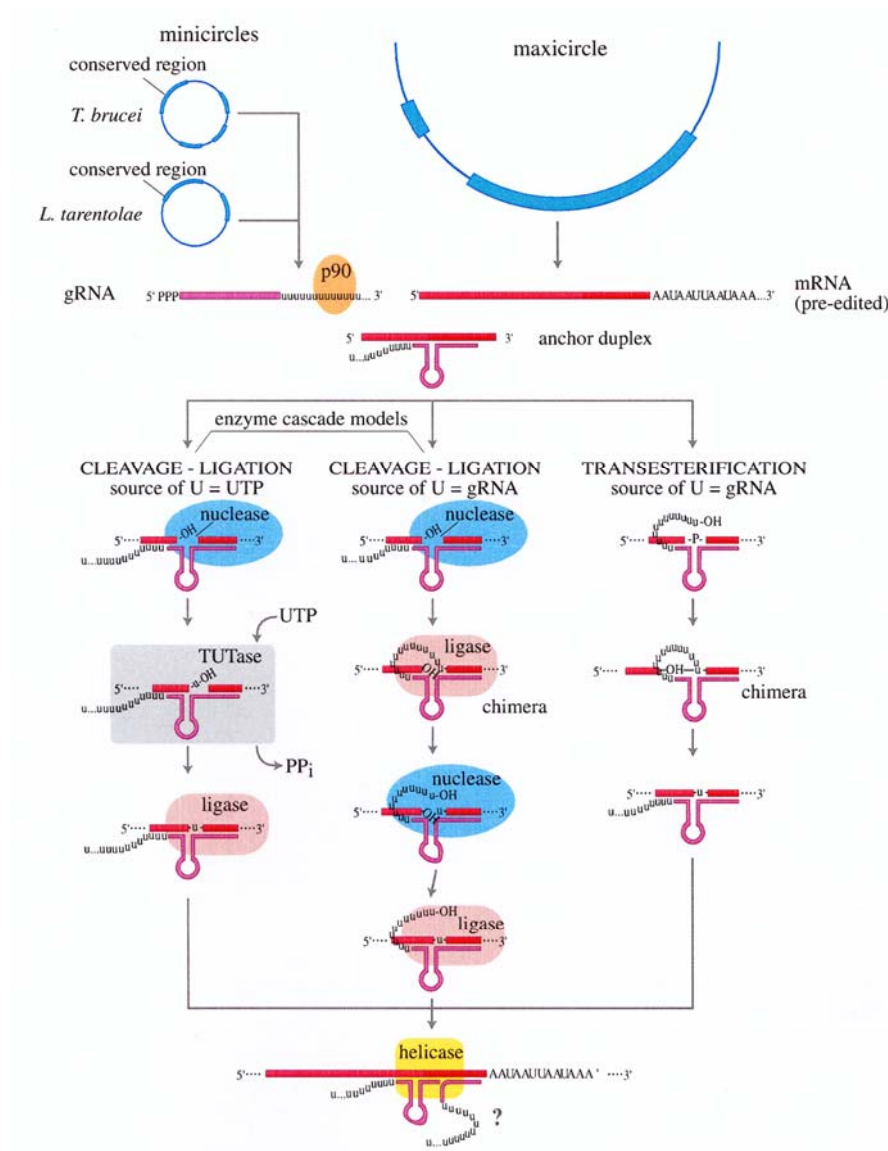


Figure 10. Theoretical models for the mechanism of RNA editing (Simpson, 2001).

Mechanism of RNA editing

The development of *in vitro* systems of RNA editing (Seiwert & Stuart, 1994; Kable *et al.*, 1996) has provided much insight into the molecular mechanism of RNA editing. The *in vitro* systems enabled experimental testing of the models proposed for RNA editing that led to the approval of the enzymatic cleavage-ligation model and disproved the alternative transesterification model (Estevez & Simpson, 1999; Simpson *et al.*, 2003). Although the crude principles of RNA editing were known, it was not until recently that a picture of the proteins and enzymes involved in the process started to emerge. This picture is rapidly becoming clearer, mainly due to the availability of the whole genome sequences of *T. brucei* and *L. tarentolae* and rapid gene identification techniques such as mass spectroscope (Simpson *et al.*, 2003). Figure 11 shows a current

model of the RNA editing process, which is considered to take place on a multiprotein complex termed the editosome. The editing occurs through the concerted action of a series of enzymes (Madison-Antenucci *et al.*, 2002; Stuart & Panigrahi, 2002; Simpson *et al.*, 2003). As proposed in the original model, the editing of mRNA is initiated upon hybridisation of the gRNA anchor region to mRNA just downstream of the editing sites. The non-encoded 3' oligo-U-tail is proposed to stabilise the gRNA-mRNA hybrid by interacting with the purin rich region upstream the editing site (Blum *et al.*, 1991; Leung & Koslowsky, 2001). The oligo-U-tail is dispensable *in vitro* (Burgess *et al.*, 1999) but some evidence indicates that it might be essential for *in vivo* editing (Aphasizhev *et al.*, 2002; Ernst *et al.*, 2003). The formation of the gRNA/mRNA anchor duplex targets cleavage of the pre-mRNA by an editing endonuclease. The particular endonuclease catalysing this cleavage has not yet been identified. The next step is the insertion or deletion of one or more uridylates determined by the guiding region of the gRNA. In the case of insertion, uridine residues are added by a 3' terminal uridine transferase, TUTase. Two candidates for catalysing this addition of uridine residues have been identified, one of which co-fractionates with the active editosome, the TbMP57 (*Trypanosoma brucei* Mitochondrial Protein of 57 kDa) (Ernst *et al.*, 2003). The other identified TUTase, TbMP108, was shown to be essential for editing (Aphasizhev *et al.*, 2002) but is not a stable component of the editosome and is suggested to have an alternate function in editing, for example addition of the oligo-U-tail to gRNA (Aphasizhev *et al.*, 2003). In the case of deletion, a 3' specific exonuclease is proposed to trim off the extra uridine residues in the mRNA. Finally, mRNA fragments are rejoined by an RNA ligase. Two potential RNA ligases, TbMP48 and TbMP52, have been identified; the latter was reported essential for editing activity (Schnauffer *et al.*, 2003).

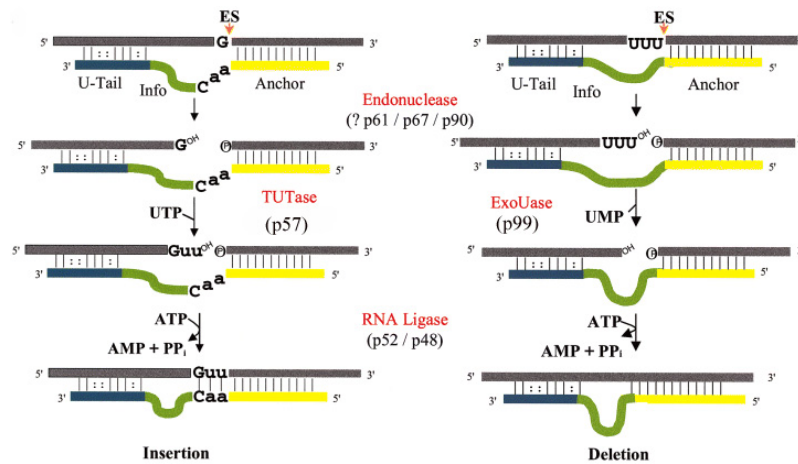


Figure 11. The catalysis of editing. The pre- mRNA (upper bar) forms an anchor duplex with the gRNA (lower bar), which has anchor, informational (info), and 3' U-tail regions. The pre-mRNA is cleaved by an endonuclease and Us are added by TUTase or removed by 3'-ExoUase at the 3'-end of the 5'-pre-mRNA fragment, depending on whether it is an insertion or deletion editing site, respectively. The processed fragment is rejoined to the 3'-fragment by an editing RNA ligase. The catalysts are components of the editosome. The figure was taken from Stuart & Panigrahi (2002) and changed according to more recent information.

Editosome

The existence of a complex catalysing RNA editing was deduced from the finding that enzymatic activities associated with editing activity sedimented together at ~20S, which was also the case with gRNAs and pre-mRNAs (Pollard *et al.*, 1992; Corell *et al.*, 1996). Analysis of the ~20S editosomal complex purified by several labs (Pollard *et al.*, 1992; Panigrahi *et al.*, 2001; Aphasizhev *et al.*, 2003; Schnauffer *et al.*, 2003) revealed a ~1.6 MDa multiprotein complex containing 7 – 20 proteins, depending on the purification procedure, capable of performing at least one round of editing *in vitro*. Thus, among the proteins associated with the editosome there must be enzymes that catalyse the core activities: endonuclease, TUTase, exoUase and ligase. However, protein identification turned out to be all but trivial, mainly due to low cellular abundance of the editosome, low sensitivity of *in vitro* editing assays and uncertainty regarding the editosomal specificity of enzyme assays (Panigrahi *et al.*, 2001). Furthermore, the fact that the functional state of the editosomal proteins is in the context of a large multisubunit complex poses difficulties in assigning a function to them. Via mass spectroscopy of the purified proteins, the first genes encoding editosomal proteins with known editing function were reported, *Trypanosoma brucei* RNA Editing Ligases, TBREL1 and TBREL2 (formerly TbMP52 and TbMP48) (Panigrahi *et al.*, 2001; Schnauffer *et al.*, 2001). TBREL1 and TBREL2 share sequence homology but only TBREL1 is essential for both editing and viability of the parasite (Schnauffer *et al.*,

2001). A 108 kDa mitochondrial 3'-TUTase was also found to be essential for editing and was considered to be the only TUTase in *T. brucei* mitochondria (Aphasizhev *et al.*, 2002). Hence, it was suggested to be responsible for U addition although it fractionates away from the 20S complex, which indicates that it is not a stable component of the editosome. Another 3'-TUTase, TbMP57, has now been reported as a component of the editosome and a most plausible U addition editing agent (Ernst *et al.*, 2003). The editing endo- and exonucleases are not yet known. However, several candidates have been proposed. According to sequence analysis of two related editosomal proteins, TbMP99 and TbMP100, they contain endonuclease/exonuclease/phosphatase family motifs, suggesting a nuclease function (Panigrahi *et al.*, 2003). An exoUase role for TbMP99 in RNA editing is supported in other studies (Schnauffer *et al.*, 2003). In addition, a set of five related proteins, TbMP61, TbMP67, TbMP90, TbMP46 and TbMP44, containing RNase III and/or RNA binding motif were identified as components of the editosomal complex and hence, at least some of them, are expected to have a nuclease function in RNA editing (Panigrahi *et al.*, 2003). The proteins TbMP18, TbMP24, TbMP42, TbMP63 and TbMP81 make up another group of interrelated editosomal components. The five protein orthologs share conserved C-terminal regions, characterised by an oligonucleotide binding motif, termed the OB-fold (see below). TbMP42, TbMP63 and TbMP81 contain two zinc finger motifs, a motif widely occurring in nucleic acid binding domains and also implicated in protein interactions (Alberts *et al.*, 1998). The roles of these enzymes are not as yet fully understood. Two of them TbMP63 and TbMP81 are proposed to function as coordinating factors regarding the order of catalytic steps in the editing process and to mediate substrate recognition and binding (Schnauffer *et al.*, 2003). A DEAD box RNA helicase, mHel61p is associated with the editosome although loosely or transiently bound (Stuart & Panigrahi, 2002). The helicase is not essential for RNA editing as shown by mutagenic experiments but absence of active enzyme results in diminished editing *in vivo* (Missel *et al.*, 1997). The role of the helicase in RNA editing may be to catalyse conformational changes associated with RNA binding or displacement of gRNAs during editing, perhaps by unwinding gRNA/pre-mRNA duplexes, which is a common activity of helicases (Missel *et al.*, 1997).

In recent years data have emerged indicating the physical and functional separation of insertion and deletion editing in distinct subunits of the editosome. Schnauffer *et al.*,

(2003) demonstrated the presence of two subcomplexes within the editosome. The subcomplex catalysing *in vitro* deletion editing contained TBREL1, TbMP63, TbMP99. The other subcomplex catalysing deletion editing consisted of TBREL2, TbMP81 and TBMP57 (Schnauffer *et al.*, 2003). This was in line with previous studies, which linked TBREL1 and TBREL2 to deletion- and insertion editing, respectively (Huang *et al.*, 2001; Cruz-Reyes *et al.*, 2002). Furthermore, the presence of pairs and sets of related proteins reinforces the idea of deletion and insertion editing being catalysed by distinct parts in the editosome (Panigrahi *et al.*, 2003).

TbMP42 and its relatives

Five related protein components of the *T. brucei* editosome, TbMP81, TbMP63, Tb42, TbMP24 and TbMP18, were identified by mass spectroscopy of purified editing complexes (Panigrahi *et al.*, 2001; Panigrahi *et al.*, 2003). All five proteins share sequence similarity in the C-terminus region. Secondary structure predictions of the C-terminal region revealed strong resemblance to an oligonucleotide binding motif, the OB-fold, a conserved module in the single-strand binding (SSB) family domain. The OB-fold is a structural domain characterised by seven- or five-stranded β -sheets that are coiled to form a closed β -barrel and is capped by an α -helix, which is typically located between the fourth and fifth strand (Murzin, 1993). The OB-fold is found in a wide variety of protein families that bind to single- or double stranded nucleic acids (Suck, 1997). The three largest proteins, TbMP81, TbMP63 and TbMP42 contain two C_2H_2 zinc finger motifs. This fold consists of an α -helix and two β -strands with the two cystein- and two histidine residues bound to a zinc ion in a tetrahedral manner, yielding a finger-like projection. C_2H_2 zinc fingers occur in a variety of proteins, including nucleic acid and chromatin binding transcription factors and translation initiation factors (Laity *et al.*, 2001). Although the majority of known zinc fingers associate with DNA, increasing evidence indicate that zinc fingers are also widely used to recognise RNA (Lu *et al.*, 2003) and have as well been implicated in protein interaction (Alberts *et al.*, 1998). As mentioned above, TbMP63 was found in relation with TBREL1 and the exoUase TbMP99 in the editosomal deletion subcomplex, while TbMP81 was associated with TBREL2 and the TUTase TbMP57 in the insertion subcomplex. Both subcomplexes contained small amounts of TbMP18 and “perhaps TbMP42” (Schnauffer *et al.*, 2003). In addition, TbMP81 was reported to stimulate the activity of the TUTase TbMP57 (Ernst *et al.*, 2003) and TbREL2 (Schnauffer *et al.*, 2003). The way TbMP63

and TbMP81 interact with proteins in different parts of the editosome as well their zinc finger- and presumed OB-fold motifs implies a nucleic acid binding- and protein interaction function for these proteins. Interestingly, the editing ligases, TBREL1 and TBREL2 do not contain OB-fold motifs, which are normally present in ligases. Thus, TbMP63 and TbMP81 and/or the related proteins might provide the OB-fold in *trans* to the editing ligases (Schnauffer *et al.*, 2003).

Ongoing research on TbMP42, by Göringer and colleges at the University of Darmstadt, has revealed that the protein has both endo- and exonuclease activity (RNA Meeting Göttingen 2003, DFG Meeting Halle 2004). To address the question of the importance of TbMP42 for editing, the effect of silencing TbMP42-gene expression on editing was tested on several mitochondrial transcripts (Poster, Brecht *et al.*, <http://biosun.bio.tu-darmstadt.de/goringer/publications/publ.html>). The results indicated that the absence of TbMP42 drastically reduces editing of some transcript but not all. Recombinant His-tagged TbMP42 was shown to bind nucleic acids in a zinc dependent manner. The binding appeared unspecific since the protein neither discriminated between double and single stranded substrate nor between RNA and DNA.

Trypanosoma brucei causes the African sleeping sickness and is transmitted to humans through the bite of the tsetse fly. The parasites multiply in the blood of mammals and cause several non-specific symptoms, such as fever, headache and joint pains. In the second stage the parasites infect the central nervous system resulting in severe symptoms and death. Over 60 million people in 36 countries on sub-Saharan Africa are threatened by the disease and every year it causes more than 60,000 deaths. RNA editing is essential for the survival of the bloodstream form of *T. brucei*. In order to identify potential drug targets and vaccines, it is of major importance to understand the mechanism of RNA editing in detail. TbMP42 is an active component of the RNA editosome with many curious features. It has both endo- and exonuclease activity and it has properties characteristic for nucleic binding proteins, such as zinc fingers and OB-fold. Furthermore, it has no sequence homology to known proteins others than its editosomal relatives. It would be of great interest to know the structural principles underlying the enzymatic and nucleic acid binding properties of TbMP42. In order address these questions, effort was made to crystallise the protein for structure determination.

Methods and materials

Materials

Chemicals

Standard chemical reagents were of analytical purity grade and were mainly obtained from the companies: Roth (address), MERCK (Darmstadt, Germany), Sigma-Aldrich (Deisenhofen, Germany) and Applichem (Darmstadt, Germany).

Equipment

Thermomixer Comfort from Eppendorf (Hamburg, Germany) and PCR-cycler from Biometra (Göttingen, Germany) were used for DNA amplification by PCR. Cell cultures were grown in an Applikon bioreactor (Schiedam, Netherlands) or using Incubators from Infors (Einsbach, Germany). Cells were lysed with Microfluidizer 110 S from Microfluidics (Newton, USA). Protein purification was performed on Äkta prime and Äkta purifier from Amersham Pharmacia Biotech. (Freiburg, Germany). Centrifugation was performed with Avanti Centrifuges from Beckman Coulter (Krefeld, Germany) and tabletop centrifuges from Eppendorf (Hamburg, Germany). Gels were visualised and photographed with the GelDoc system from BioRad (München, Germany). Ultraspec2100pro from Amersham Pharmacia Biotech. (Freiburg, Germany) was used for photo-spectrometry purposes. Rigaku Micromax 007 or -RUH3R rotating anode generators (Rigaku-MS, TX, USA) equipped with Mar-345 image plates from MarResearch (Eppendorf, Germany) were used for testing crystal diffraction and collecting diffraction data. Information for other apparatus is provided in the description of appropriate sections of methods

Buffers

The composition of buffers used in the experiments is provided in the description of the experimental conditions in appropriate sections of methods.

Enzymes and inhibitors

Enzymes for cloning purposes: restriction enzymes, T4-DNA ligase, Pfu- and Taq polymerases were obtained from New England Biolabs (Frankfurt, Germany) and Fermentas (Sankt Leon-Rot, Germany). For cleaving GST and MBP fusion tags from target proteins, Factor Xa from Novagen (Darmstadt, Germany) and home made TEV

and Prescission proteinases were used. PMSF was used to stop serine proteinase digestion and Complete EDTA-free Protease Inhibitor Cocktail Tablets from Roche Diagnostics, (Mannheim, Germany) were used for protease inhibition in cell lysates.

Kits

DNA cleanup and plasmid purification was carried out with kits from Qiagen (Hilden Germany): QIAquick Gel Extraction Kit, QIAquick PCR Purification Kit and QIAprep Spin Miniprep Kit.

Bacterial strains

<i>Escherichia coli</i> strains	Reference or source	Utility
XL1-Blue <i>recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac</i> [F' <i>proAB lacIqZΔM15 Tn10</i> (Tetr)]	(Bullock <i>et al.</i> , 1987) Novagen	Cloning
Top10 F' <i>mcrAΔ(mrr-hsdRMS mcrBC)Φ80lacZΔM15 ΔlacX74 recA1 deoR araD139 Δ(ara-leu)7697 galU galK rpsL</i> (Str ^R) <i>endA1 nupG</i>	Invitrogen	Expression of: pBAD-VPR
BL21(DE3) <i>E. coli</i> B, F ⁻ , <i>ompT</i> , r _B ⁻ m _B ⁻ , λDE3(<i>cIts857, ind1, Sam7, nin5, lacUV5-T7</i> RNA polymerase)	(Studier <i>et al.</i> , 1990) Novagen	Expression of: pET28b-p61, pGEX6P1-p61, pMal-TbMP42 pMal-TbMP42 pJOE-TbMP42-c pETM42-TbMP42 pETM42-c TbMP42-c
BL21(DE3) -RIL -RP -pLysS -pLysE		Expression tests
M15[REP4]	Qiagen	Expression of: pQE-60_ TbMP42
SG13009[REP4] <i>Nal^SStr^Srif^S, lac⁻ara⁻gal⁻mtl⁻F⁻recA⁻uvr⁺</i>	Qiagen	Expression tests

Vectors used for production of recombinant proteins in *E. coli*

Expression vector	Reference or Source	Utility
pBAD	Invitrogen	pBAD_VPR
pGEX6P-1	Amersham	pGEX6P-1_p61 pGEX6P-1_p61_mut pGEX6P-1_TbMP42
pET28a (+)	Novagen	pET28a_p61
pQE60	Qiagen	pQE-60_TbMP42
pJOE 37075.3	(Wilms <i>et al.</i> , 2001)	pJOE_TbMP42-Cterm
pMal-c2X	New England Biolabs	pMal_TbMP42 pMal_TbMP42
pETM-42	Heidelberg	pETM-42_TbMP42 pETM-42_TbMP42-Cterm

Medium and antibiotics

For preparation of LB and 2YT medium, the following ingredients were added to 1 l water (tap water or distilled water works fine).

	LB	2YT
Tryptone	10 g	16 g
Yeast extract	5 g	5 g
NaCl	10 g	5 g

For preparation of LB-agar plates, 6 g of agar were added to 400 ml LB-broth to get 1.5 % LB-agar. Medium was autoclaved at 121 °C for 20 minutes. Antibiotics were added to give the final concentrations as follows:

	Stock solution	Working concentration
Ampicillin	100mg/ml	100 µg/ml
Kanamycin	50mg/ml	50 µg/ml
Chloramphenicol	34 mg/ml	34 µg/ml
Streptomycin	10 mg/ml	10 µg/ml

Oligonucleotides

Primers for PCR and sequencing were ordered from MWG Biotech (München, Germany). TbMP42 DNA/RNA substrate was ordered PAGE purified from Purimex (Staufenberg, Germany)

p61 specific	
p61r236 Reverse from base 236	5' CTCCACTGGTCCCATCACTT 3'
p61f775 Forward from base 775	5' CTGTCGGGCTTCTCGTCTAC 3'
p61r775 Reverse from base 775	5' GTAGACGAGAAGCCCGACAG 3'
p61f1126 Forward from base 1126	5' AACCGTATGAGCTTCGGAGA 3'
p61r1228 Reverse from base 1228	5' GTGGCCTCGTTTACCTGTGT 3'
Amplification of the TbMP42-gene by PCR	
Tb08_3 Reverse C-terminal	5' GAAGATCTCACCCCTCAACACTGACCCACAG 3' <i>BglII</i>
Tb08_5 Forward N-terminal	5' CGTCATGAAGCGTGTTACTTCACATATTTTC 3' <i>BspHI</i>
Tb08_fw Forward N-terminal	5' CCGGATCCATGAAGCGTGTTACTT 3' <i>BamHI</i>
Tb08_rv Reverse C-terminal	5' GGTCTAGACTCGAGTTACACCCTCAACACTGACCCAC 3' <i>XbaI XhoI</i>
pJOE_tb_f Forward from base 748	5' CATGCCATGGGAACACAGTATGTTCA 3' <i>NcoI</i>
pJOE_tb_r Reverse C-terminal	5' CGGGATCCCACCCTCAACACTGACCCACA 3' <i>BamHI</i>
Tb08_bsmBIa Forward N-terminal	5' ACTTACCGTCTCCCATGAAGCGTGTTACTTCACATATTTTCG 3' <i>BsmBI</i>
Tb08_kpnI Reverse C-terminal	5' GGGGTACCTTACACCCTCAACACTGACCCA 3' <i>KpnI Acc65I</i>
TbMP42 specific	
Tb08_3 Forward from base 242	5' TGATGAACAAGGGGGGTAAG 3'
TbMP42 substrate	
Oligo_1 (25 nt DNA)	5' GGGAAATGTGATTTTTGCGAGTAGCC 3'
Oligo_2 (20 nt DNA)	5' GGCTACTCGCTCACATTCCC 3'
Oligo_3 (25 nt DNA/RNA hybrid)	5' GGGAAUGUGA(dU) ₅ GCGAGUAGCC 3'
Oligo_4 (20 nt RNA)	5' GGCUACUCGCUCACAUUCCC 3'
Vector specific primers for sequencing	
MALE Forward pMal, pETM42 and pJOE	5' GGTCGTCAGACTGTTCGATGAAGCC 3'
M13/pUC Reverse pMal	5' GTAAAACGACGGCCAGT 3'

Molecular biology techniques

Polymerase Chain Reaction

Genes and gene fragments were amplified by polymerase chain reaction (PCR). PCR reactions for cloning purposes were set up as following:

	Volume	Stock concentration	Amount
DNA-template	0.5 μ l	200-300 ng/ μ l	100-150 ng
Forward primer	5 μ l	4 pmol/ μ l	20 pmol
Reverse primer	5 μ l	4 pmol/ μ l	20 pmol
dNTPs	4 μ l	10 mM each	40 nmol
10 X PCR buffer	5 μ l	10 X	1 X
Polymerase	1 μ l	2.5 U/ μ l	2.5 U
H ₂ O	ad 50 μ l		

The temperature and time length of the reaction steps were adjusted in each experiment regarding primer and template properties to obtain optimal results.

10 X PCR buffer:

100 mM Tris-HCl pH 8.8

500 mM KCl

15 mM MgCl₂

Ethanol precipitation

PCR products were purified with QIAquick PCR Purification Kit or ethanol precipitated by adding 2.5 fold volume of 96 % ethanol and sodium acetate to a final concentration of 0.3 M. The solution was incubated at room temperature for 20 minutes and centrifuged at maximum speed in a minicentrifuge. The pellet was washed with 70 % ethanol, centrifuged at maximum speed for 5 minutes and finally the pelleted DNA was dried in a speedvac (Eppendorf, Hamburg).

Plasmide preparation

QIAprep Spin Miniprep Kit was used for plasmide preparations from *E. coli* following the producer's instructions.

Agarose-gel electrophoresis

DNA probes were electrophoresed on 0.7-1 % agarose-gels for analysis of the size and amount of DNA fragments. Loading dye was added to DNA probes, which were then electrophoresed at 85 Volt in TBE or TAE-modified buffer, stained in electrophoresis buffer containing ca. 5×10^{-4} % ethidium bromide and finally visualised on a Gel Photo Apparatus (Biorad).

<u>6 X loading dye</u>	<u>TBE buffer:</u>	<u>TAE-modified buffer:</u>
10 mM Tris-HCl pH 7.6	90 mM Tris-borate pH 8.3	40 mM Tris-acetate pH 7.6
0.03 % bromophenol blue	2 mM EDTA	0.1 mM EDTA
0.03 % xylene cyanol FF		
60% glycerol		
60 mM EDTA		

Preparation of 1 l of 10 X buffers:

<u>TBE buffer</u>	<u>TAE-modified buffer</u>
108 g Tris-base	48.4 g Tris-base
55 g boric acid	11.4 ml glacial acetic acid
40 ml 0.5 M EDTA	2 ml 0.5 M EDTA

DNA restriction digest

Plasmids and DNA fragments were digested with restriction enzymes using appropriate buffers from New England Biolabs and MBI Fermentas following the producers' instructions.

Gel extraction from agarose gels

DNA samples were loaded on 1 % agarose gels in modified TAE buffer and electrophoresed. Bands were cut out of the gels with a scalpel and purified with QIAquick Gel Extraction Kit according to instructions from the producer.

Ligation

The most successful method for the ligation step of cloning was the "do not measure DNA concentration strategy". Purified plasmid from 5 ml culture and purified PCR product from one 50 μ l PCR reaction, were restriction digested and purified from agarose gels. DNA was eluted from the gel extraction columns with 20 μ l water and mixed in the three ligation reaction and one control reaction (C) as follows:

	Plasmid DNA (μ l)	Insert DNA (μ l)	10 x Ligase Buffer (μ l)	100 mM ATP (μ l)	20 U/ μ l Ligase (μ l)	H ₂ O (μ l)
1	5	4	2	1	1	7
2	5	7	2	1	1	4
3	5	9	2	1	1	2
C	5	0	2	1	1	11

The reactions were incubated overnight at 14 °C.

Competent cell preparation

The RbCl method for preparing competent cells gives better transformation efficiencies than the conventional CaCl₂ procedure for most strains. The following protocol can be used to obtain competent cells with a transformation efficient of 10⁶ - 10⁷ colonies per microgram of DNA.

A single colony from an agar plate was inoculated into 2 ml LB medium and grown overnight at 37 °C. The overnight culture was diluted in 100 ml and grown to OD₆₀₀ = 0.5. Cells were cooled on ice and centrifuged 10 minutes at 3000 xg and 4 °C. Cells were resuspended in 40 ml cold TBF1 buffer, incubated on ice for 5 minutes and centrifuged for 10 minutes at 3000 xg and 4 °C. The cell pellet was gently resuspended in 4 ml cold TBF2 buffer and aliquoted into pre-chilled eppendorfs. The competent cells were finally frozen with liquid nitrogen and stored at -80°C.

TBF1:

30 mM KOAc
100 mM RbCl₂
10 mM RbCl₂
50 mM MnCl₂
15 % (v/v) glycerol
adjusted to pH 5.8 with acetic acid

TBF2:

10 mM MOPS
75 mM CaCl₂
10 mM RbCl₂
15 % (v/v) glycerol
adjusted to pH 6.5 with KOH

The buffers were mixed from separately autoclaved stock solutions and sterile filtered.

Transformation

100-200 ng of plasmid or, in the case of cloning procedures, the entire ligation-reaction was added to 50 μ l of competent cells. Following 30 minutes incubation on ice, cells were heat-shocked at 42 °C for 30 s. 150 μ l of LB-medium were added and incubated further 20 minutes at 37 °C. Finally the cells were plated on LB-agar plates containing the appropriate antibiotic to select for transformed cells.

Identification of positive clones

Bacciscreen

Colonies were picked with a sterile pipette tip which was then let touch the interior of a PCR tube and streaked on an LB-agar plate containing the appropriate antibiotic. A master mix solution was then added to the PCR tubes. After amplification, the PCR product was electrophoresed on 1 % agarose gel and visualised by ethidium bromide staining. A master mix solution contained per tube:

	Volume	Stock concentration	Amount
Forward primer	2 μ l	4 pmol/ μ l	8 pmol
Reverse primer	2 μ l	4 pmol/ μ l	8 pmol
dNTPs	0.4 μ l	25 mM each	10 nmol
10X PCR buffer	1 μ l	10 X	1 X
Taq polymerase	0.2 μ l	2.5 U/ μ l	0.5 U
H ₂ O	ad 1 0 μ l		

Restriction digest test

For confirmation of positive clones found in bacciscreens or to search for positive clones in cases where the bacciscreen failed to give reliable results, a restriction digest was performed on the plasmids isolated from putative positive colonies. Plasmids were digested with the restriction enzymes used in the cloning procedure. The resulting fragments were electrophoresed in 1 % agarose and visualised by ethidium bromide staining.

Sequencing

Sequencing was performed in the Laboratory of Prof. Dr. Pieler on an ABIPrism 3100 sequencer.

SDS-PAGE

Protein samples were analysed for size and purity by SDS PAGE (sodium-dodecyl-sulfate polyacrylamide gel electrophoresis) using gel apparatus from Biometra (Göttingen, Germany). Samples were mixed 1:1 with loading buffer. Pelleted bacteria from liquid cultures were suspended in $0.2 * OD_{600} * V$ (volume of culture) of loading buffer. For instance 0.5 ml of culture with $OD_{600} = 1$ was solved in 100 μ l loading buffer. Samples were inhibited with 0.1 mM PMSF (in case of VPR and protease cleavage experiments) and incubated at 95 °C for 1 minute. Electrophoresis was performed at 30 mA for 50 – 80 minutes depending on the acrylamide concentration in

the gels. The gels were soaked in staining solution, heated for 1 minute in a microwave oven, incubated at room temperature for 20 minutes and rinsed with water.

Separating gel

7 – 15 % Acrylamide, 0.2 – 0.4 % Bisacrylamide
375 mM Tris/HCl pH 8.8
0.1 % (w/v) SDS
0.1 % (v/v) TEMED
0.05 % (w/v) Ammonium persulphate

Stacking gel

4 % Acrylamide, 0.2 – 0.4 % Bisacrylamide
125 mM Tris/HCl pH 6.8
0.1 % (w/v) SDS
0.1 % (v/v) TEMED
0.05 % (w/v) Ammonium persulphate

Running buffer

192 mM Glycine
25 mM Tris/HCl pH 8.3
0.1 % SDS

2X Loading buffer

62.5 mM Tris pH 6.8
70 mM SDS
50 % (v/v) Glycerol
0.1 % (w/v) Bromophenol blue
5 % β -Mercaptoethanol

Staining solution

10 % (v/v) Ethanol
5 % (v/v) Acetic acid
0.002 % (w/v) Coomassie G/R250

Handy protocols for mixing the solutions can be found at <http://www.nwfsc.noaa.gov/protocols/sds-page.html>

Western Blotting

PVDF membranes were wet in methanol. Whatman-paper was cut in the size of the gel (8 peaces on each side of the gel) and wet in BTB-buffer. Metal plates of the Semi-dry (Biorad) were wet with BTB-buffer. Gels were run at 10 V, 100 – 150 mA for 1 hour. After rinsing membranes with water, they were rinsed three times with PBS-T-buffer and 1X 15 minutes in PBS-T-buffer. Blocking was performed with incubation in 5 % skim milk in PBS-T for one hour at room temperature or at 4 °C overnight. Washing with PBS-T followed, three times rinse and 1X incubation for 15 minutes. Antibodies were diluted in PBS-T and swirled on the membrane for one hour at room temperature or at 4 °C overnight. Washing with PBS-T followed, three times rinse and 1X incubation for 15 minutes. ECL detection reagents were mixed (1:1), poured onto the washed membrane and incubated for exactly one minute. The membrane was wrapped in plastic foil, placed with a film in a developing cassette and developed.

10X BTB-stock (Blot Transfer Buffer)

56 g (50 mM) Tris
30 g (40 mM) Glycine
37.5 ml 10 % (w/v) SDS0 (0.0375 %)
H₂O ad 1 l

1X BTB

20 ml 10X BTB
140 ml H₂O
40 ml methanol fresh from 4 °C

10X PBS (pH 7.4) stock

80 g NaCl
2 g KCl
11.5 g Na₂HPO₄*7H₂O

PBS-T

1X PBS + 0.02 % Tween 20

Crystallisation methods

Crystal growth

The sitting drop vapour diffusion technique was used for crystallisation. In this technique a drop containing protein, buffer, precipitants and eventually additives is allowed to equilibrate with a larger reservoir containing similar buffers and precipitants in higher concentrations. Since the vapour pressure is lower for the reservoir solution there is a net transfer of water through the vapour phase from the protein droplet to the reservoir. Initially the protein in the droplet is in either a metastable or undersaturated state with respect to protein nucleation. The loss of water from the droplet increases the concentrations of both protein and precipitant. The solution equilibrates to achieve supersaturating concentrations of protein, which results in nucleation (if one is lucky) or precipitation (Figure 12).

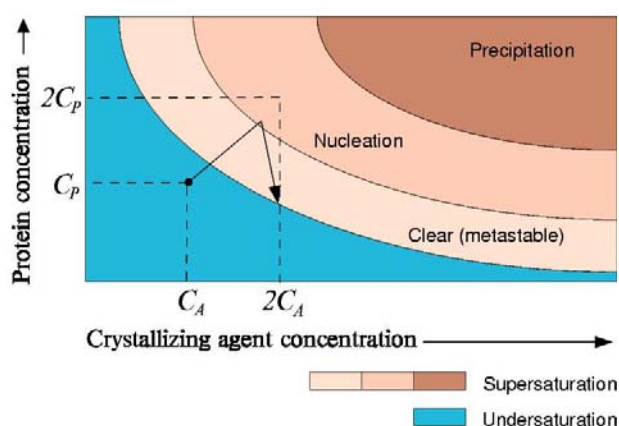


Figure 12. Phase diagram of a vapour diffusion experiment, where crystals grow. The figure was taken from: <http://www-structmed.cimr.cam.ac.uk>

For crystallisation setups, 24-well sitting drop plates from Hampton Research (CA, USA) were used and the plates were sealed with Crystal Clear Sealing Tape, also from Hampton Research. A typical procedure for initial screening was to mix 1 μ l of 5 - 10 mg/ml protein solution with 1 μ l of a 500 μ l reservoir solution. For initial screens the Hampton Crystal Screens 1 and 2 and the Footprint screens at 4 $^{\circ}$ C and 20 $^{\circ}$ C were used first. Depending on how the drops looked, measures were taken for further crystallisation trials. Other crystal screens used were the Hampton Research Screens: Lite, Cryo, Natrix, PEG/Ion and the Additive Screen 1-2, the Jena Bioscience Screens 1 - 10 and the Magic Screens.

Seeding

Seeding is a technique used to manipulate the protein's position in phase space. The energy barrier to protein to crystallize is the formation of the critical nucleus. Once the nucleation energy barrier has been overcome, the crystals can grow. Spontaneous nucleation, however, often occurs at high degrees of supersaturation favouring production of numerous small crystals or aggregates. This can impede further crystal growth. Because conditions for nucleation and crystal growth differ, it can be of advantage to optimize these processes separately. Thus, transferring nuclei from a nucleation condition into a supersaturated condition, which favours crystal growth, can result in great improvement of crystal quality. Three main methods of seeding are used to improve crystal quality:

- i. Streak seeding: small nuclei are transferred with the help of a whisker to an equilibrated protein solution
- ii. Microseeding: crystals are crushed and transferred into a stabilising solution. After serial dilution of the seed stock a whisker or a pipette is used to inoculate a pre equilibrated drop
- iii. Macroseeding: a crystal is picked, washed and transferred into an equilibrated protein solution optimised for crystal growth.

Only streak seeding and microseeding were used here in order to improve crystals. The precipitant and/or the protein concentrations of the solution seeded into were lowered relative to the nucleation conditions. For microseeding experiments a series of diluted seeding stock, 1:100, 1:200, 1:500, 1:1000, was prepared and 1 μ l was transferred into a pre-equilibrated drop.

Protein preparation, crystallisation and structure determination of the *Vibrio*-proteinase

Expression and purification of VPR

The production of the *Vibrio*-proteinase for crystallisation preparations was based on the previously established expression system (Arnórsdóttir *et al.*, 2002) and the purification protocol according to the one described for the proteinase from *Vibrio* strain PA44 (Kristjánsson *et al.*, 1999), with the following modifications. Expression of the *Vibrio*-proteinase gene cloned in the pBAD TOPO vector was carried out in 12 l cultures of the *E. coli* strain Top10 at 18 °C in a bioreactor (Applikon Biotechnology). Cells were harvested 12 h after induction with 0.025% L - arabinose and addition of CaCl₂ to a final concentration of 10 mM. For one preparation, cell pellet from 6 l culture was suspended in 300 - 400 ml of basic buffer (buffer A: 25 mM Tris, pH 8.0 containing 10 mM CaCl₂) and disrupted by running it 5 times with 5 minutes intermediate incubations on ice, through a microfluidiser (Microfluidics™) at 550 kPa pressure. The crude cell extract was centrifuged at 15 000 xg for 15 minutes at 4 °C. The protein in the supernatant was precipitated by a 75% saturation of ammonium sulphate and centrifuged at 15 000 xg for 30 minutes at 4 °C. The pellet was redissolved in buffer A containing 1 M (NH₄)₂SO₄ and centrifuged at 100 000 xg for 1 hour at 4 °C to remove insoluble impurities. The subsequent purification steps were carried out at 4 °C using the Äkta system (Amersham Biosciences, Freiburg, Germany). The protein solution was loaded on a Phenyl-Sepharose column (16/10 Amersham Biosciences) equilibrated with buffer A containing 1 M (NH₄)₂SO₄. Elution was achieved by a 20 column volume gradient of 1 - 0 M (NH₄)₂SO₄ and fractions were tested for activity with succinyl-AlaAlaProPhe-*p*-nitroanilide. The fractions containing proteolytic activity were pooled and applied to a 2 ml N-carbobenzoxy-D-phenylalanyl-triethylenetetramine-Sepharose column (Fujiwara & Tsuru, 1976) equilibrated with buffer A. After a washing step with 0.5 M NaCl, the *Vibrio*-proteinase was eluted with buffer A containing 2 M GdmCl. Fractions of 2.5 ml were collected into tubes containing 2 ml of 3 M (NH₄)₂SO₄ in buffer A. The pooled fractions containing proteolytic activity were loaded onto a 5 ml Phenyl-Sepharose column (Hitrap Phenyl FF, Amersham Bioscience) equilibrated with buffer A containing 1 M (NH₄)₂SO₄ and eluted with a 20 column volume gradient of 1 - 0 M (NH₄)₂SO₄. The purified 40 kDa

Vibrio-proteinase was concentrated to 3 - 6 mg/ml by means of salting out with 75% saturated ammonium sulphate, adding 3 parts of a saturated ammonium sulphate solution to 1 part of protein solution. The solution was centrifuged and the precipitate resuspended with buffer A to a concentration of 5 mg/ml. At this point the protein was divided into aliquots, flash cooled in liquid nitrogen and stored at $-80\text{ }^{\circ}\text{C}$. Aliquots containing the purified 40 kDa *Vibrio*-proteinase were incubated at $40\text{ }^{\circ}\text{C}$ for 50 minutes to give the mature 30 kDa enzyme, which was then inhibited with phenylmethanesulfonylfluoride (PMSF) in a final concentration of 1 mM and applied onto a Superdex 75 column (HR 10/30, Amersham Biosciences) equilibrated with 10 mM Tris pH 8.0 and 10 mM CaCl_2 . Fractions containing the 30 kDa *Vibrio*-proteinase were pooled and concentrated in centrifugal concentrators (Centricon and Minicon from Millipore) for crystallisation trials.

Activity assay

The activity of VPR was assayed using succinyl-AAPF-*p*-nitroanilide as a substrate. The substrate was solved in dimethylsulfoxide to a 25 mM substrate solution, which can be stored at 4°C . The assays were performed at 0.3 mM substrate concentration in assay buffer (usually 120 μl of 25 mM substrate solution were added to $\sim 10\text{ ml}$ assay buffer) at room temperature. Typically 50 μl of enzyme solution were added to 950 μl of substrate solution and the increase in absorbance at 410 nm ($\epsilon = 8480\text{ M}^{-1}\text{cm}^{-1}$) was measured.

Assay buffer:

0.1 M Tris/HCl pH 8.6
10 mM CaCl_2

Crystallisation and data collection

The recombinant *Vibrio*-proteinase was crystallised using the sitting drop method. The protein solution used in the initial crystallisation trials was 2.5 mg/ml protein in 10 mM Tris/Cl pH 8.0 and 10 mM CaCl_2 . A promising condition was found using the Hampton Crystal Screen 1 condition 41 (10 % 2-propanol, 20 % PEG 4000, 0.1 M HEPES pH 7.5) where clusters of needles grew overnight. After variations of temperature, pH and concentrations of the precipitant and protein solutions, well diffracting crystals were obtained by mixing in equal volumes of a protein solution of 6 mg/ml and a precipitant solution containing 15 % PEG 4000, 10 % isopropanol, 0.1 M Tris/Cl pH 8.0 at $20\text{ }^{\circ}\text{C}$. Data used for structure determination were collected at 100 K using the mother liquor as

cryoprotectant on a Rigaku Micromax 007 rotating anode generator (Rigaku-MSX, TX/USA) operating at 40 kV and 20 mA equipped with a Mar-345 image plate detector (MarResearch, Eppendorf, Germany). The crystal to detector distance was 250 mm and 1° oscillation images were collected with 20 minutes exposure time. Diffraction data were processed using the programs DENZO and SCALEPACK (Otwinowski & Minor, 1997) and Molecular replacement using the CCP4 suite. A high-resolution dataset was obtained at the BW7B beamline at EMBL outstation DESY Hamburg. Data collection statistics for the synchrotron data, which was used to build the structure of the *Vibrio*-proteinase, 1SH7, are shown in Table 1.

Structure solution and refinement

The structure of the *Vibrio*-proteinase was solved by molecular replacement using the program MolRep (Vagin & Teplyakov, 1997). A homology model of the *Vibrio*-proteinase based on the known structure of proteinase K (PDB ID: 1IC6, Betzel *et al.*, 2001) was used as a search model. The structure was refined with Refmac5 (Winn *et al.*, 2001). A random set of 10% of reflection was excluded from refinement to monitor R_{free} (Brünger, 1992). Model building was done in XtalView (McRee, 1999). Water molecules were assigned with ARP/wARP (Lamzin & Wilson, 1993) using standard parameters. Refinement statistics are shown in Table 1.

Structure analysis

Superposition of structures was performed with LSQMAN (Kleywegt, 1996). Salt bridges were found using WHATIF (Rodriguez *et al.*, 1998), excluding His and with a distance cut-off of 4 Å between charged atoms. Hydrogen bonds were defined with HBPLUS (McDonald & Thornton, 1994). Surface areas were calculated using the WHATIF-server (<http://swift.cmbi.kun.nl/WIWWWI/>), which uses probe radius 1.4 Å. Electrostatic potentials were calculated with Delphi (Honig & Nicholls, 1995). Graphics were made with Pymol (DeLano, W. L. (2002). The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA, USA).

Protein preparation for crystallisation trials of p61

Expression and purification of p61

BL21 cells transformed with pGEX6P-1-p61 were grown overnight at 30 °C in LB-amp containing 1 % (w/v) glucose. The overnight culture was diluted approximately 10-fold in 2YT-amp containing 1 % glucose, grown at 15°C to $OD_{600} = 1$ and finally induced with 0.5 mM IPTG. The cells were harvested 12-15 hours after induction. Cell pellet from 1 l culture was suspended in 40 ml lysis buffer. After 1 hour rolling at 4 °C, the solution was submitted to two freeze/thaw cycles and sonification (2 X 20 pulses). The crude cell lysate was centrifuged at 27 000 xg at 4 °C for 30 minutes. The supernatant was applied to 2 ml of GSH-sepharose beads equilibrated with binding buffer and rolled at 4 °C overnight. The beads were washed 3-5 times with binding buffer (depending on amount of protein in washing fractions measured with Bradford reagent). In the effort to clear the beads of a protein suggested to be DnaK, additional washing steps were performed with DnaK wash buffer. Finally, 50 µg of Prescission proteinase was added to the beads suspension and incubated at 4 °C for 45 minutes. The supernatant of the beads containing the cleaved p61 was collected. Prescission proteinase digestion was repeated twice and the fractions collected were pooled.

Lysis buffer:

50 mM Tris pH 7.7
150 mM NaCl
10 mM MgCl₂
2 mM DTT
1 mg/ml lysozyme
5µg/ml DNase I
1 tablet inhibition cocktail

Binding buffer:

50 mM Tris pH 7.7
150 mM NaCl
10 mM MgCl₂
2 mM DTT

DnaK wash buffer:

50 mM Tris pH 7.3
60 mM KC l
10 mM MgCl₂
2 mM ATP
2 mM DTT

RNase stability test

RNaseA was incubated at 95 °C for 10 minutes. After slow cooling to room temperature, 0.5 µg of the RNase were added to 15 µg protein samples with and without protease inhibition cocktail and incubated in 50 mM Tris pH 7.7, 300 mM NaCl, 10 mM MgCl₂, 2 mM DTT at room temperature for 1 hour. BSA (bovine serum albumin) serving as a control was treated with RNase in the same manner. After incubation samples were centrifuged at 11 000 xg for 5 minutes and samples of supernatant were analysed by SDS-PAGE.

Protein preparation for crystallisation trials TbMP42

Cloning, expression and purification of TbMP42

TbMP42-His₆ from pQE-60

A preculture of *E. coli* cells of the strain M15[REP4] transformed with the expression vector pQE-60_TbMP42 was grown in LB-amp/kan at 30 °C overnight. The preculture was diluted twenty-fold with LB-amp/kan and grown further at 37°C. At OD₆₀₀ = 0.6 the expression of TbMP42 was induced with 0.5 mM IPTG. 3 hours after induction the cells were harvested by centrifugation at 4800 xg for 15 minutes. Cell pellet from 1 l culture was solved in 30 ml Extraction buffer 1 containing 4 M urea, stirred on ice for 30 minutes and centrifuged at 10 000 xg for 15 minutes. The supernatant was discarded. The pellet was resuspended in 30 ml Extraction buffer 2 containing 8 M urea, stirred on ice for 30 minutes and centrifuged at 10 000 xg for 15 minutes. The supernatant was collected. The extraction with 8 M urea was repeated twice. The resulting three supernatant solutions were combined, centrifuged at 100 000 xg for 30 minutes and applied on a 5 ml Hitrap chelating Ni⁺² column equilibrated with Extraction buffer 2 at 4° C. After a washing step with wash-buffer (Extraction buffer 2 adjusted to pH 5.9) the protein was eluted by lowering the pH to 4.5 (Extraction buffer 2 adjusted to pH 4.5). The eluted fractions were collected into tubes containing 1 M Tris pH 8 (1/10 of the fractions volume). The fractions containing TbMP42-His₆ were pooled, the protein concentration was adjusted to 1 mg/ml and the solution was dialysed at 4 °C overnight against approximately twohundred-fold volume of refolding buffer. The dialysis was repeated with fresh buffer for 2 hours. After dialysis the solution was centrifuged at 100 000 xg for 30 minutes at 4 °C and applied on a gel filtration column.

Extraction buffer 1:

0.1 M NaH₂PO₄
0.01 M Tris pH 8
4 M urea

Extraction buffer 2:

0.1 M NaH₂PO₄
0.01 M Tris pH 8
8 M urea

Refolding buffer:

20 mM HEPES pH 7.4
100 mM KCl
1 mM ZnSO₄

TbMP42-GST from pGEX6P-1

The TbMP42-gene was amplified from pQE-60-TbMP42 with PCR using the primers: Tb08_fw (5'-CCGGATCCATGAAGCGTGTTACTT-3') and Tb08_rv (5'-GGTCTAGACTCGAGTTACACCCTCAACACTGACCCAC 3'). Underlined are the restriction sites for *Bam*HI and *Xho*I, respectively, which were used for cloning the gene into pGEX6P-1.

A preculture of BL21(DE3)pLysS transformed with pGEX6P-1-TbMP42 was grown in LB-amp at 30 °C overnight, diluted twenty-fold in LB-amp and grown further at 15 °C. Ethanol to an end concentration of 1% was added when the culture had reached OD₆₀₀ = 1 and following that (30 minutes later), expression was induced with 0.5 mM IPTG. The cells were harvested by centrifugation at 4 800 xg for 15 minutes after 12-15 hours induction at 15 °C. Cell pellet from 1 l culture was suspended in 40 ml lysis buffer and disrupted by running it 3 times with 5 minutes intermediate incubations on ice, through a microfluidiser (Microfluidics™) at 550 kPa pressure. The crude cell lysate was incubated with 1 µl benzonase and 1 tablet of inhibition cocktail for 30 minutes at 4 °C. After centrifugation at 15 000 xg for 30 minutes, the lysate was filtrated and applied on GSH-sepharose column equilibrated with binding buffer. The protein was eluted with the same buffer containing 20 mM reduced glutathione (GSH). The fractions containing TbMP42-GST were pooled, prescission proteinase was added and the solution was dialysed at 4 °C overnight against dialysis buffer.

Lysis buffer:

50 mM HEPES pH 7
150 mM NaCl
2 mM DTT
0.1 mM ZnCl₂

Binding buffer:

50 mM HEPES pH 7
150 mM NaCl
2 mM DTT

Elution buffer:

50 mM HEPES pH 7
150 mM NaCl
2 mM DTT
20 mM GSH

Dialysis buffer:

50 mM HEPES pH 7
300 mM NaCl
2 mM DTT
1mM EDTA

TbMP42-MBP from pMal-c2X

The TbMP42-gene was amplified from pQE-60-TbMP42 with PCR using the primers: Tb08_fw (5'-CCGGATCCATGAAGCGTGTTACTT-3') and Tb08_rv (5'-GGTCTAGA CTCGAGTTACACCCTCAACACTGACCCAC 3'). Underlined are the restriction sites for *Bam*HI, *Xba* I and *Xho*I, respectively. Attempts to clone the amplified gene into pMal-c2X using *Bam*HI and *Xba*I failed, probably due to insufficient *Xba*I digestion. Therefore, the insert was restriction digested with *Bam*HI/*Xho*I and cloned into pMal-c2X that was restriction digested with *Bam*HI/*Sal*I, which is possible because *Sal*I and *Xho*I have identical overhangs.

A preculture of BL21 transformed with pMal-c2X-TbMP42 was grown in LB-amp at 30 °C overnight, diluted twenty-fold in LB-amp and grown further at 15 °C. Expression was induced at OD₆₀₀ = 0.7 with 0.5 mM IPTG. The cells were harvested by centrifugation at 4 800 xg for 15 minutes after 12-15 hours induction at 15 °C. Cell pellet from 1 l culture was suspended in 40 ml lysis buffer and disrupted by running it 3 times with 5 minutes intermediate incubations on ice, through a microfluidiser (Microfluidics™) at 550 kPa pressure. The crude cell lysate was incubated with 1 µl benzonase and 1 tablet of inhibition cocktail for 30 minutes at 4 °C. After centrifugation at 15 000 xg for 30 minutes, the lysate was filtrated and applied on amylose-resin column equilibrated with binding buffer. The protein was eluted with the same buffer containing 10 mM maltose. The fractions containing TbMP42-MBP were pooled, concentrated (Vivaspin) to 5 ml and applied on superdex 200 (16/60) equilibrated with gel filtration buffer.

Lysis buffer:

50 mM Tris pH 8.5
200 mM NaCl
10 mM MgCl₂
5 mM β-mercaptoethanol
0.5 mM ZnCl₂

Binding buffer:

50 mM Tris pH 8.5
200 mM NaCl
10 mM MgCl₂
5 mM β-mercaptoethanol

Elution buffer:

50 mM Tris pH 8.5
200 mM NaCl
10 mM MgCl₂
5 mM β-mercaptoethanol
10 mM maltose

Gel filtration buffer:

10 mM Tris pH 8.5
10 mM MgCl₂
5 mM β-mercaptoethanol

Factor Xa digestion test - TbMP42-MBP from pMal-c2X

Aliquots containing 200 µl of 1 mg/ml TbMP42-MBP, directly from amylose elution, were incubated with 5 µl of 2 U/µl Factor Xa at 4°C and 25 °C. Samples were drawn and inhibited with 1 mM PMSF at time intervals: 0, 15, 45, 90, 120 minutes, 19 and 22 hours after addition of Factor Xa and analysed by SDS-PAGE.

TbMP42-MBP from pETM-42

The TbMP42-gene was amplified from pQE-60-TbMP42 with PCR using the primers: Tb08_bsmBIa 5'-ACTTACCGTCTCCCCATGAAGCGTGTTACTTCACATATTTTCG-3' and Tb08_kpnI 5'-GGGGTACCTTACACCCTCAACACTGACCCA-3'. Underlined are the restriction sites for *BsmBI* and *KpnI/Acc65I*, respectively. The TbMP42-gene contains a recognition site for *NcoI* in its sequence. In order to clone the gene using the *NcoI* restriction site on pETM42, the gene is amplified with a *BsmBI* recognition site. *BsmBI* cuts at positions 1/5 outside of the recognition site and thus can be used to create overhangs to utilise restriction sites irrespective of the presence of such sites in the sequence (Figure 13). Expression and purification were performed in the same manner as described above for TbMP42-MBP produced from pMal-2cX.

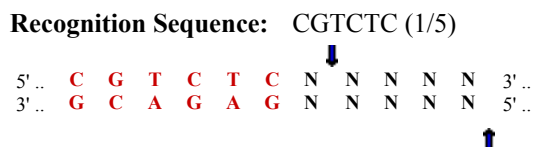


Figure 13. The *BsmBI* restriction site

TEV proteinase digestion test- TbMP42-MBP from pETM42

TbMP42-MBP eluate from amylose resin was adjusted to pH 8 and sodium citrate was added to an end concentration of 5 mM. Aliquots containing 200 µl of 1 mg/ml TbMP42-MBP were incubated with 25 µl, 50 µl and 75 µl of 1.6 mg/ml TEV proteinase at 4 °C, 20 °C and 34 °C. Samples were drawn and inhibited with 1 mM PMSF at 0.5, 1, 2, 4, etc. hours up to 24 hours after addition of proteinase.

TbMP42-MBP_Cterm from pETM-42

The TbMP42-gene contains an *NcoI* restriction site at base 750. A PCR product used for cloning the full length TbMP42-gene into pETM42 (page 52), was cut with *NcoI* and *Acc65I* to create a C-terminal fragment comprising amino acids 250-393. Expression was performed in the same manner as described above for TbMP42-MBP produced from pMal-2cX.

TbMP42-MBP_Cterm from pJOE 3075.3

The TbMP42-gene was amplified from pQE-60-TbMP42 with PCR using the primers: pJOE_tb_f 5'-CATGCCATGGGAACACAGTATGTTCA-3' and Tb08 pJOE_tb_r 5'-CGGGATCCCACCCTCAAACTGACCCACA-3'. Underlined are the restriction sites for *NcoI* and *BamHI*, respectively, which were used for cloning the gene into the corresponding restriction sites on pJOE 3075.5.

Results

The cold adapted *Vibrio*-proteinase

Preparation of the *Vibrio*-proteinase for crystallisation trials

The recombinant *Vibrio*-proteinase (VPR) was purified after expression in the *E. coli* strain TOP10 using the pBAD TOPO TA expression system. Other expression systems have failed to give active protein probably due to detrimental background expression (Arnórsdóttir *et al.*, 2002). The pBAD TOPO system provides an N-terminal leader peptide and a C-terminal His₆-tagged V5 epitope for detection and purification (Figure 14). The expression of the VPR-gene from the pBAD TOPO system is satisfactory for biochemical experiments but is not optimal for crystallisation preparations. First, the expression yields only moderate amounts of protein. Second, the expression product does not bind to His₆-tag affinity columns, which is probably due to autoproteolysis on the V5 epitope, cleaving off the His₆-tag. Thus, each preparation requires large amounts of expression culture and an elaborate purification procedure (Figure 15). For a preparation of 1 mg of pure VPR a minimum of 6 l expression culture was required.

```

ATG GGC TCT GGA TCC GGT GAT GAC GAT GAC AAG CTC GCC CTT PCR AAG
                                     GAG CGG GAA product TTC
Met Gly Ser Gly Ser Gly Asp Asp Asp Asp Lys                                     Lys

GGC GAG CTT GAA GGT AAG CCT ATC CCT AAC CCT CTC CTC GGT CTC GAT TCT
Gly Glu Leu Glu Gly Lys Pro Ile Pro Asn Pro Leu Leu Gly Leu Asp Ser

ACG CGT ACC GGT CAT CAT CAC CAT CAC CAT TGA
Thr Arg Thr Gly His His His His His His ***

```

Figure 14. Organisation of the insert region of the pBAD TOPO TA expression vector. An N-terminal 11 amino acid leader peptide (underlined) similar to a thioredoxin segment can enhance production of soluble heterologous target protein in *E. coli*. A C-terminal His₆-tagged V5 epitope (grey) contains a plausible VPR-digestion site since the VPR-construct from pBAD does not bind to His₆-affinity matrix.

VPR is isolated from *E. coli* cultures as a 40 kDa protein consisting of the catalytic domain and a C-terminal prosequence (Figure 7, page 16). Figure 15 shows an overview of the purification procedure of VPR. The 40 kDa VPR was purified by a combination of ammonium precipitation, hydrophobic interaction chromatography and affinity chromatography. Subsequently the C-terminal prosequence was cleaved off by mild heat treatment to yield the 30 kDa mature VPR, which was finally purified by size exclusion chromatography.

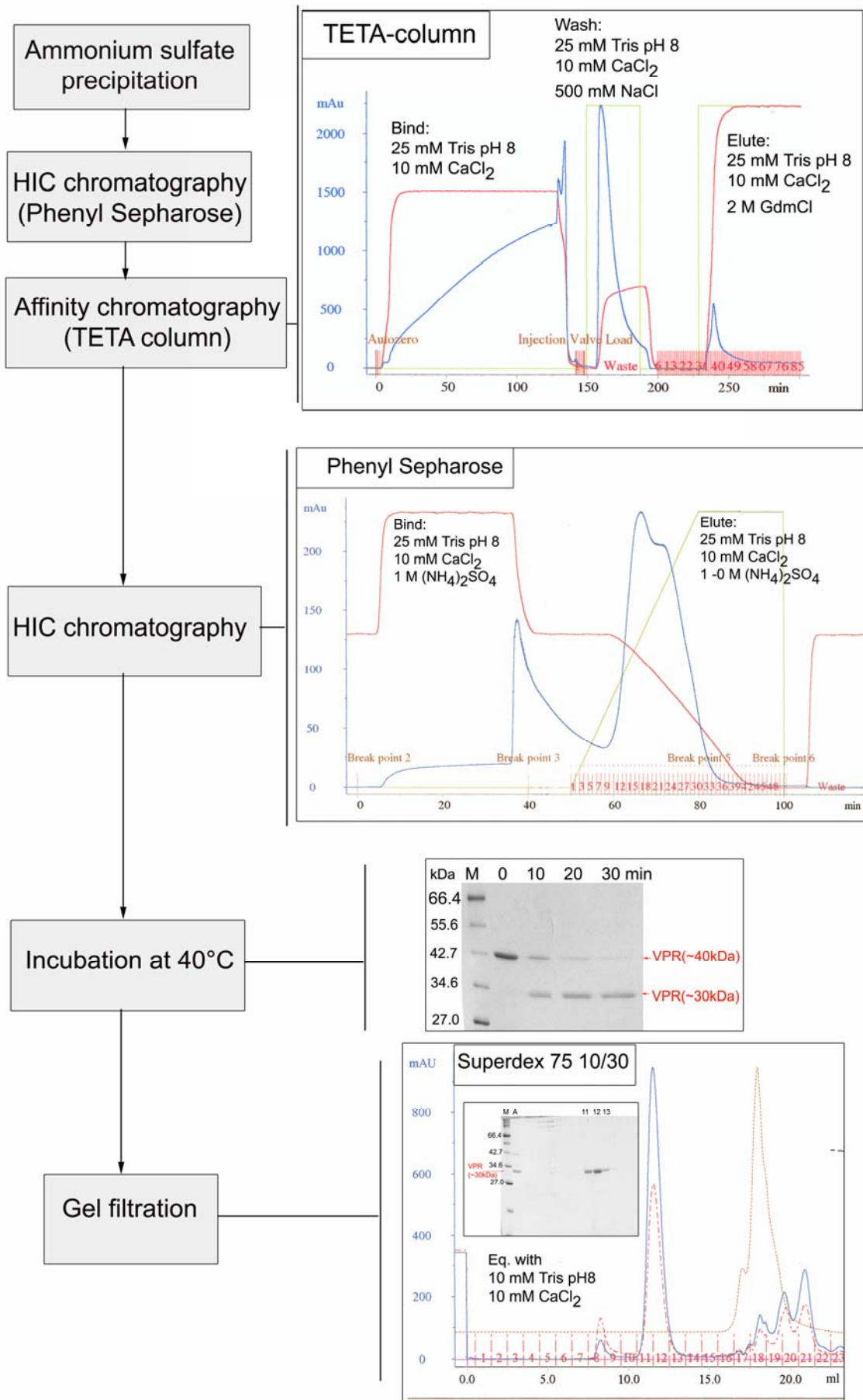


Figure 15. An overview of the purification procedure of VPR. After salting out by 75 % saturation of ammonium sulphate the precipitated proteins were redissolved in 1 M ammonium sulphate, centrifuged and applied on a phenyl sepharose column. The fractions containing VPR activity were applied on N-carbobenzoxy-D-phenylalanyl-triethylenetetramine-Sepharose column, generally called the TETA column. VPR binds tightly to the TETA column and has to be eluted with 2 M GdmCl into ammonium sulphate solution adjusted to have an end concentration of 1 M. The fractions containing VPR activity were applied on a 5 ml phenyl sepharose column and the eluted 40 kDa VPR was incubated at 40 °C to undergo autocatalysis that gives 30 kDa mature VPR. Finally the 30 kDa VPR was purified on the analytical superdex 75 10/30 column.

Crystallisation of the *Vibrio*-proteinase

In the very first crystal screen set up for VPR, needles grew overnight at 20 °C in condition 43 (10 % 2-propanol, 20 % PEG 4000, 0.1 M HEPES pH 7.5) of the commercial Hampton Crystal Screen 1 (Figure 16, left). In 2 – 3 days, the needles transformed into platelets that diffracted to 3 Å on a rotating anode; a relatively high resolution regarding the crystals' small size (longest axis < 100 µm). The one-dimensional crystal needles were reproduced without problems at different temperatures (4 °C, 10 °C, 20 °C) over a range of pH 7.0 - 8.5 and varying protein and precipitant concentrations but the transformation into three-dimensional crystals did not occur. After countless unsuccessful attempts to reproduce and improve the crystals with seeding techniques, three-dimensional crystals (Figure 16, left) surprisingly grew from needles in a drop identical to others, which had previously not yielded three-dimensional crystals (20 °C, 10 % 2-propanol, 15 % PEG 4000, 0.1 M Tris pH 8.0). These crystals diffracted to 2.4 Å on a rotating anode and to 1.84 Å on the synchrotron at DESY in Hamburg.

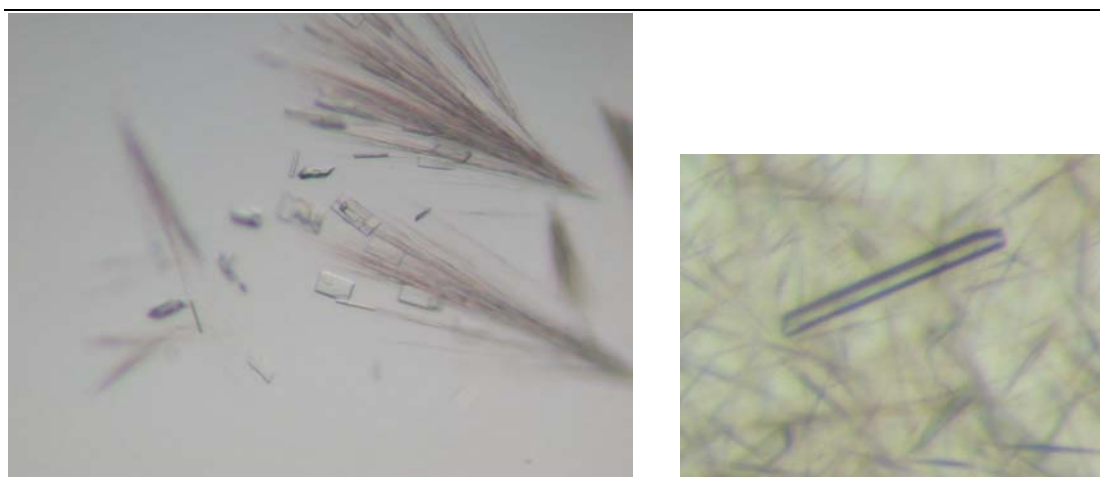


Figure 16. VPR crystals. The needles seen on the left, grew overnight at 20 °C in 10 % 2-propanol, 20 % PEG 4000, 0.1 M HEPES pH 7.5, transformed into platelets in 2-3 days. In spite of their small size (longest axis < 100 µm), they diffracted to 3 Å on a rotating anode. The crystal on the right (long axis ~200 µm) grew out of a crowd of needles at 20 °C in 10 % 2-propanol, 15 % PEG 4000, 0.1 M Tris pH 8.0 and diffracted to 1.84 Å on the BW7B beamline at DESY in Hamburg.

Structure determination of the *Vibrio*-proteinase

The VPR crystals belong to space group $P2_1$ with unit cell dimensions of $a = 43.2 \text{ \AA}$, $b = 36.9 \text{ \AA}$, $c = 140.5 \text{ \AA}$ and $\beta = 97.8^\circ$. The Matthews coefficient (Matthews, 1968) ($V_m = 1.9 \text{ \AA}^3/\text{Da}$) suggested two molecules in the asymmetric unit with a solvent content of 36.3 %. The low solvent content fits with the relatively good diffraction regarding the small crystal size. The structure was determined by molecular replacement using a homology model based on the known structure of proteinase K (PDB accession code, 1IC6) as a search model. The crystallised 30 kDa catalytic domain of VPR encompasses amino acids 140 - 420 of the 530 amino acid long pre-pro-enzyme (Arnórsdóttir *et al.*, 2002). The model was refined at a resolution of 1.84 \AA with an R-factor of 14.1 % and an R-free value of 19.6 % (Table 1). The Ramachandran plot geometry of the model is satisfactory for this resolution, with 90 % of the residues within the most favoured regions, 10 % in the additionally allowed and only one residue in the generously allowed region (Figure 17).

Table 1. Data collection and refinement statistics for VPR (PDB code: 1SH7). Numbers in parenthesis refer to the highest resolution shell.

Data collection	
Resolution range (\AA)	40-1.81 (1.87-1.81)
Space group	$P2_1$
Unit cell parameters	$a = 43.2 \text{ \AA}$ $b = 36.9 \text{ \AA}$ $c = 140.5 \text{ \AA}$ $\beta = 97.80^\circ$
Number of reflections	135,690
Unique reflections	37,893
Completeness (%)	93.2 (50.4)
R_{sym}^a (%)	9.3 (50.3)
Average I/σ	13.0 (2.7)
Refinement Statistics	
Resolution range (\AA)	30-1.84 (1.88-1.84)
$R_{\text{cryst}}/R_{\text{free}}^b$ (%)	14.1(22.6)/19.6(29.8)
Rms deviation from ideality	
Bonds (\AA)/angles ($^\circ$)	0.014/1.521
Average B values (\AA^2)	
Protein/water/PMSF/ Ca^{+2}	13.3/25.4/34.1/11.9
Ramachandran plot ^c	
Most favoured, additional, generously allowed (%)	89.9/9.9/0.2

^a $R_{\text{sym}} = 100 \cdot \sum_h \sum_i |I_i(h) - \langle I(h) \rangle| / \sum_h I(h)$, where $I_i(h)$ is the i th measurement of the h reflection and $\langle I(h) \rangle$ is the average value of the reflection intensity.

^b $R_{\text{cryst}} = \sum |F_o - F_c| / \sum |F_o|$, where F_o and F_c are the observed and calculated structure factors respectively. R_{free} is R_{cryst} with 10 % of test set structure factors.

^c Calculated with PROCHECK (Laskowski *et al.*, 1993)

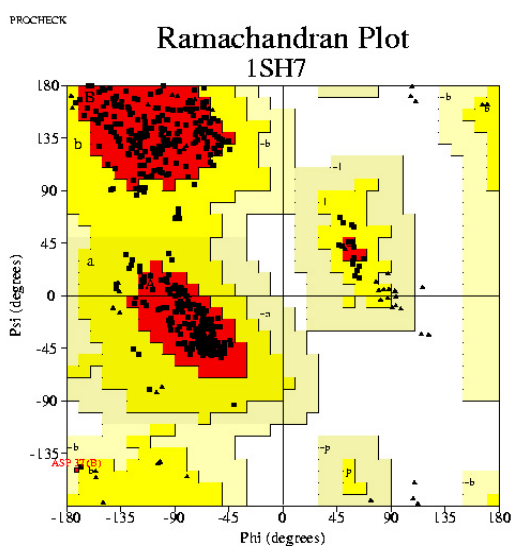


Figure 17. Ramachandran plot of 1SH7, the model of the *Vibrio*-proteinase refined at 1.84 Å.

Crystal structure of the *Vibrio*-proteinase

Figure 18 shows the three-dimensional structure of the *Vibrio*-proteinase, hereafter referred to as 1SH7 according to its PDB accession code. The structure exhibits the α/β scaffold characteristic of subtilisin-like serine proteinases. It consists of six α -helices, one 3/10 helix, a β -sheet made of seven parallel strands and two β -sheets made of two antiparallel strands each. The structure determination confirms the presence of three previously predicted disulfide bonds, Cys67-Cys99, Cys163-Cys194 and Cys277-Cys281 (Arnórsdóttir *et al.*, 2002). Three calcium-binding sites are found in 1SH7, two of which had been predicted based on sequence alignments and one not described as yet in other subtilases (see below). The active site of 1SH7 consists of the catalytic triad: Asp37, His70 and, Ser220 and substrate recognition and binding sites that are well conserved among subtilases (Siezen *et al.*, 1991). The substrate-binding site in 1SH7 appears on the surface as a relatively distinct cleft, where the substrate is accommodated by forming a triple-stranded antiparallel β -sheet with the residues of the S4 and S3 binding sites (nomenclature of subsites, S4 –S2', is according to Schechter and Berger (1967). The bottom of the S1 substrate-binding pocket is made up of residues A154-A155-G156 and the oxyanion hole residue N157. The substrate binding cleft appears to be relatively open-mouthed with T105 at the rim of S4, whereas in many subtilases this site is occupied by a larger residue, typically a tyrosine (e.g. subtilisin and proteinase K), which is assumed to form a flexible lid on the S4 pocket (Takeuchi *et al.*, 1991).

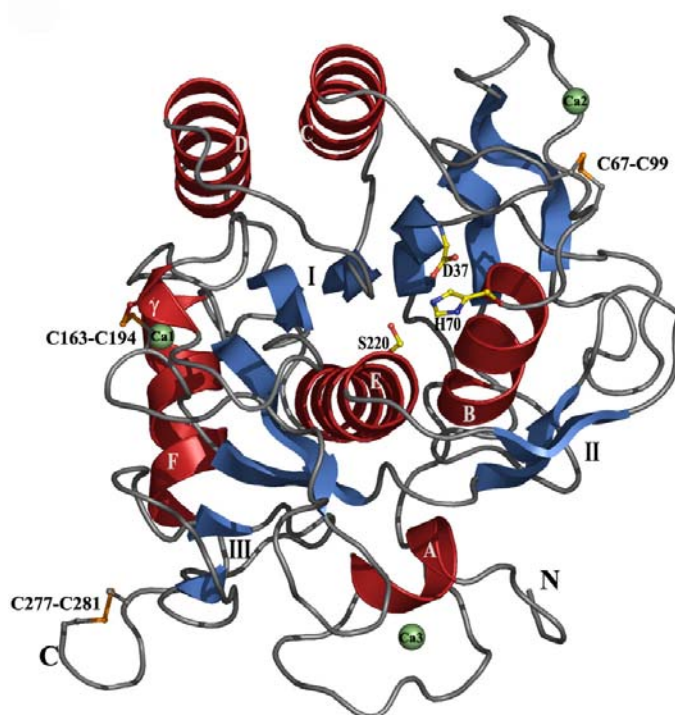


Figure 18. Model of the crystal structure of the *Vibrio*-proteinase. The residues of the catalytic triad, S220, H70 and D37 are shown in yellow, the calcium ions as green spheres and the disulfide bridges in orange. The helices and sheets are labelled according to the topology diagram shown in Figure 20.

Overall structure comparison of the cold adapted *Vibrio*-proteinase and related enzymes from meso- and thermophiles

A 0.98 Å resolution structure of proteinase K, PDB accession code: 1IC6 and a 1.37 Å resolution structure of thermitase, PDB accession code: 1THM, were used for structural comparison with 1SH7. The high resolution of all three structures allows a reasonable comparison with respect to the quality of the models. The pairwise least square superposition of the three structures, with a 3.5 Å distance cut-off revealed that 85 - 93 % of the C α -atoms lie at common positions and gave a root mean square deviation in the range of 0.84 - 1.21 Å (Table 2 and Figure 19). The structural resemblance regarding: RMS deviation, fraction of common C α -atoms and the amino acid sequence identity, is in the order: 1SH7-1IC6>1SH7-1THM>1IC6-1THM.

Table 2. Pairwise superposition of C α -atoms in 1SH7, 1IC6 and 1THM with a 3.5 Å distance cutoff.

	1SH7-1IC6	1IC6-1THM	1SH7-1THM
Number of residues	281-279	279-279	281-279
Aligned residues	261 (93%)	238 (85%)	246 (88%)
Identities	120 (43%)	86 (31%)	93 (33%)
Root mean square deviation	0.84 Å	1.21 Å	1.11 Å

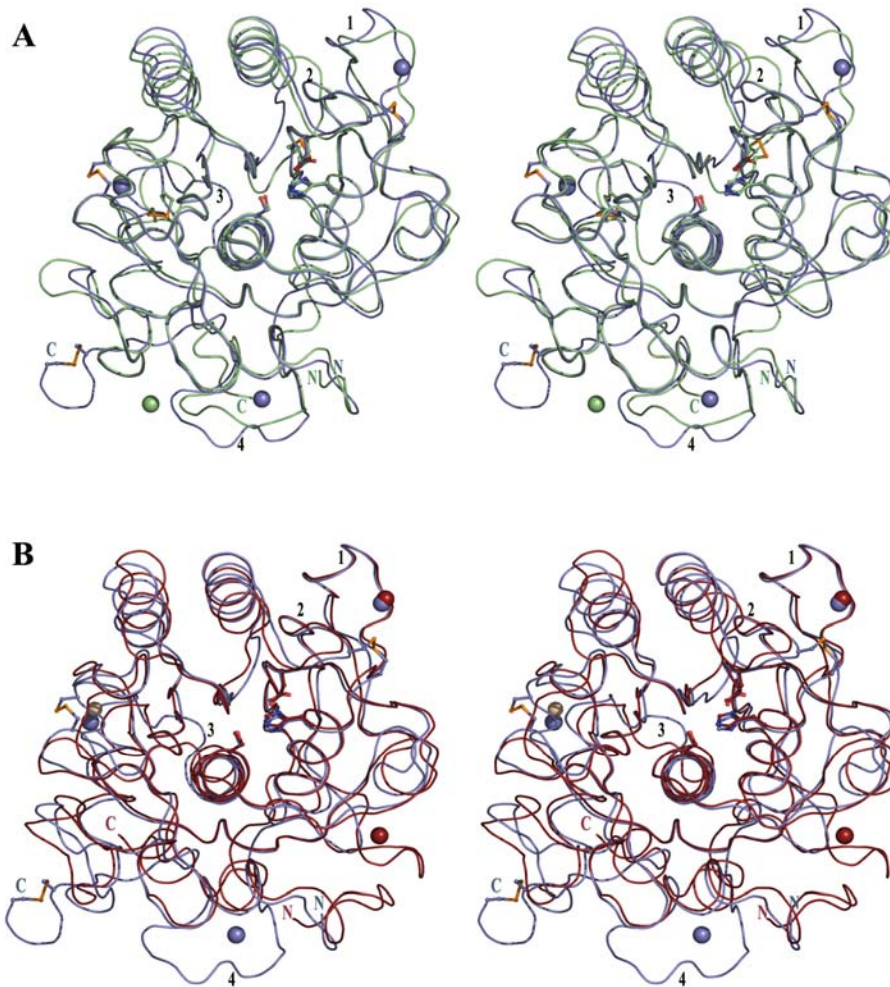
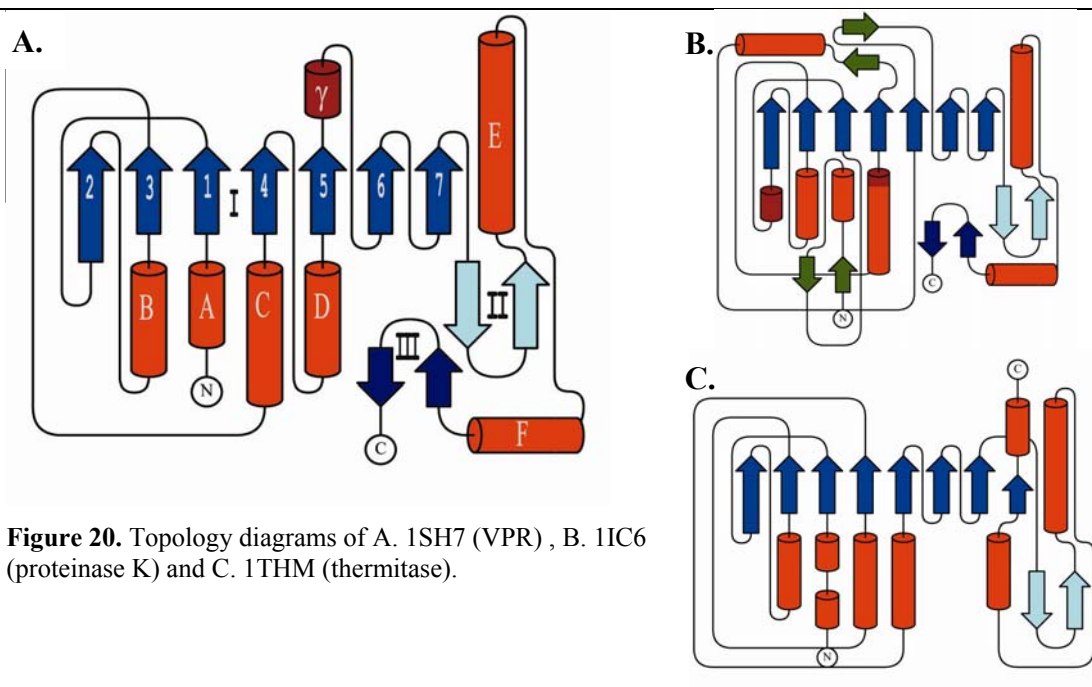


Figure 19. Stereo view of the superposition of the cold adapted *Vibrio*-proteinase (1SH7, blue) with (A) proteinase K (1IC6, green) and (B) thermitase (1THM, red). Calcium ions and a sodium ion (beige) bound to thermitase are shown as spheres. The numbering relates to the four regions that deviate due to multiple insertion and deletions as described in the text.

The distance deviations of the superposed structures and the locations of insertions and/or deletions are restricted to a few parts of the structure. The most distinct differences are seen in the regions of the N- and C-termini, where 1THM aligns poorly with both 1SH7 and 1IC6. The C-termini of 1IC6 and 1SH7 also diverge; the four last residues of 1IC6 are not equivalent to residues in 1SH7. Furthermore 1SH7 has an extended C-terminus relative to 1IC6. On Figure 19 the four regions in the structures that deviate considerably because of multiple residue insertions and deletions are marked as described below. First, a surface loop region, Phe57-Asn68 in 1SH7 does not align with 1IC6. This loop is identical in 1SH7 and 1THM and hosts a calcium-binding site that has been described as a medium strong calcium-binding site in thermitase (Gros *et al.*, 1991). Secondly, 1IC6 has an insertion relative to both 1THM and 1SH7 in an

extended surface loop, residues 119 - 125 in 1IC6. This surface loop in 1IC6 contains some plausible stabilising features, a disulfide bridge, Cys34-Cys123, and a salt-bridge, Asp117-Arg121. Third, a loop region connecting α -helices E, carrying the Ser of the catalytic triad, and the succeeding α -helix F is not well conserved among the enzymes and the structures are accordingly variable. Fourth, 1SH7 contains a new calcium-binding site. This part of the structure is noticeably different from the corresponding regions in proteinase K and thermitase. If the allowed distance between equivalent C α -atoms is defined as being within 2 Å, the ratio of common C α -atoms is still over 80 % between 1SH7 and the other two structures. The topology diagrams of the three enzymes show that they all share similar secondary structure organisation. Apart from the common features of subtilases, such as the 7-8 stranded beta sheet I (Figure 20) and most of the alpha helices, all three enzymes share a 2 stranded beta sheet II. In addition 1SH7 and 1IC6 both have a C-terminal beta sheet III. The high structural homology of these enzymes of different temperature origin gives an opportunity to examine structural features that might contribute to their different temperature adaptation.



Charged residues and ion pairs

Thermitase contains 30 charged side chains, whereas proteinase K and the *Vibrio*-proteinase contain 38 each. The *Vibrio*-proteinase differs from the enzymes compared in having a higher proportion of negatively charged side chains (Table 3). Charged residues reside on the protein surface in regions that are the least conserved. Superposition of 1SH7, 1IC6 and 1THM revealed that at seven sites, there are identically charged side chains in all three proteins. Additionally, each pair of enzymes, 1SH7-1IC6, 1SH7-1THM and 1IC6-1THM, has 4 - 6 side chains with the same charge in equivalent positions. Thus, the conservation of charged residues is comparable to the overall homology of these structures, being in the range of 30 – 40 %.

Table 3. Comparison of structural features of 1SH7, 1IC6 and 1THM.

	1SH7	1IC6	1THM
Number of charged residues	38	38	30
(D+E)/(R+K)	24/14	18/20	15/15
Number of non-compensated charged residues	23	23	15
(D+E)/(R+K)	(16/7)	(10/13)	(7/8)
Number of ion pairs ^{a)}	8	8	10
<i>Number of hydrogen bonds</i>			
mainchain-mainchain	152	157	161
mainchain-sideschain	87	68	76
sidechain-sideschain	23	10	30
total	262	235	267
Exposed surface area ^{b)} (Å ²)	10115	10079	9822
apolar ^{c)} (Å ²)	4989	5024	4732
Buried surface area ^{b)} (Å ²)	31695	32013	31714
apolar ^{c)} (Å ²)	18601	19288	19234

^{a)} An interaction is assigned to a salt bridge where distance between atoms of opposite charge is within 4 Å. Interactions involving histidine are not included.

^{b)} solvent accessible surface area for residues 1-275 of each enzyme

^{c)} carbon- and sulphur atoms

The tendency of an increased number of salt bridges in going from lower to higher temperature origin, frequently observed in comparisons of related structures of different temperature origin, cannot be confirmed for the enzymes compared in this study. Ionic interactions as defined here are restricted to two oppositely charged residues (Asp, Glu, Arg and Lys) within 4 Å distance. The meso- and psychrophilic structures have the same number of salt bridges and only two less than the thermophilic structure (Table 3). An important aspect of the plausible contribution of salt bridges to protein stability

resides in their location and distribution. Bae & Phillips (2004) recently defined as “critical ion pairs” for temperature adaptation, those pairs which are non-conserved between the structures compared and bridging residues of distant regions (more than ten residues) of the polypeptide chain. Four non-conserved ion pairs in 1IC6 link residues that are within four-residues’ distance in the polypeptide chain. In contrast, all the salt bridges in 1SH7 and all but one in 1THM, involve residues of more than ten-residues’ distance (Table 4).

Table 4. Listing of salt bridges and the shortest distances between charged atoms. Salt bridges are restricted to 4 Å distance between charged atoms of the residues: Asp, Glu, Arg and Lys. Conserved ion pairs are in the upper row. Critical ion pairs (Bae & Phillips, 2004) with respect to both of the compared enzymes are underlined.

1SH7		1IC6		1THM	
D56 - R95	2.99 Å			D57 - R102	2.97 Å
D59 - R95	3.03 Å			D60 - R102	3.00 Å
D183 - R10	2.74 Å	D187 - R12	2.77 Å	D188 - K17	3.91 Å
		^{a)} (E27 - 87	4.65 Å)	E28 - K95	2.79 Å
<u>D138 - R169</u>	3.02 Å	<u>E48 - R80</u>	3.93 Å	<u>D124 - K153</u>	3.20 Å
<u>E236 - R252</u>	2.83 Å	<u>E50 - R52</u>	2.95 Å	<u>D188 - R270</u>	2.81 Å
<u>E255 - K267</u>	2.81 Å	D98 - K94	2.75 Å	<u>D201 - R249</u>	3.41 Å
<u>D260 - R185</u>	2.92 Å	<u>D112 - R147</u>	2.76 Å	E253 - R249	2.98 Å
<u>D274 - R14</u>	3.28 Å	D117 - R121	2.94 Å	<u>D257 - R270</u>	2.80 Å
		D184 - R188	3.02 Å	<u>D257 - K275</u>	2.78 Å
		<u>D260 - R12</u>	3.02 Å		

^{a)} The criterion of conserved ion pairs is when the distance between corresponding charged residues is within 6 Å. Therefore, although not defined here as a salt bridge, this interaction excludes the corresponding ion pair in 1THM from being critical in this comparison.

The higher number of critical ion pairs in 1SH7 and 1THM, which contain seven such interactions each compared to three in 1IC6, supports evidence of the plausible significance of salt bridges in adaptation of enzymes to hot as well as to cold environments (Kumar & Nussinov, 2004). There is one common ion pair, Asp183-Arg10 (numbers relate to 1SH7), in all the three enzymes, connecting sites that are otherwise not well conserved in 1THM relative to 1SH7 and 1IC6. 1SH7 and 1THM share an ion pair arrangement, Asp56-Arg95 and Asp59-Arg95 (numbers relate to 1SH7), connecting the surface loop that hosts their common calcium-binding site to a site proximate to the substrate-binding site (Figure 21). Critical ion pairs are found in both 1IC6 and 1THM bridging the α -helices C and D, which are directly connected to the loops of substrate binding. In 1THM the ion pair network formed by Asp188-Arg270, Asp257-Arg270 and Asp257-Lys275 tethers the C-terminal. Such tethering has

been suggested to contribute to increased stability in other proteins (Leiros *et al.*, 2000). Thus, by observing single ion pair interactions, differences emerge that cannot be seen by the mere count of interactions. In the context of estimating the effect of salt bridges to protein stability, their accessibility to the solvent is highly important. We thus checked the solvent accessibility of the ion pairs forming salt bridges in the three protein structures, but such comparisons did not however reveal any trends in terms of temperature adaptation of the enzymes.

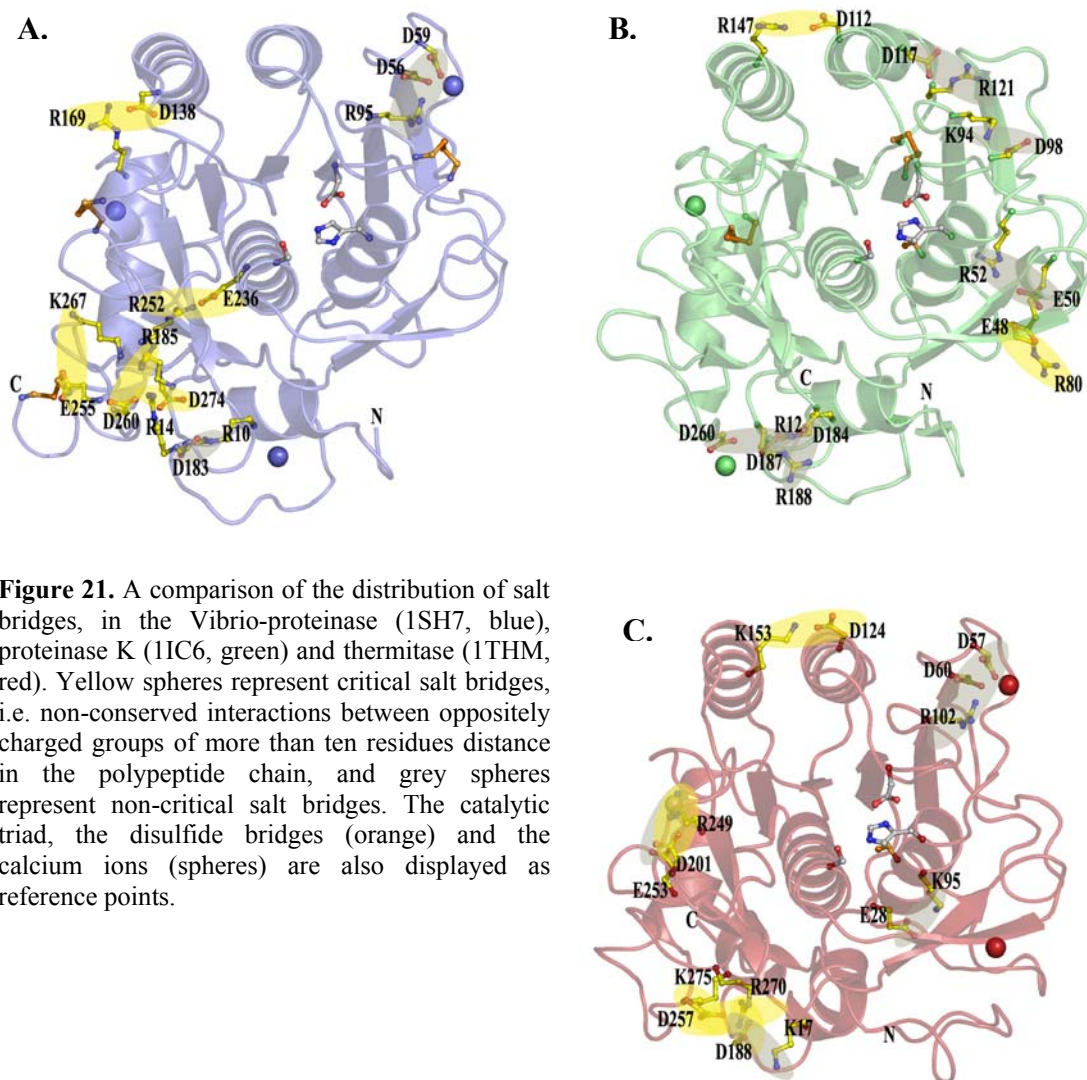


Figure 21. A comparison of the distribution of salt bridges, in the Vibrio-proteinase (1SH7, blue), proteinase K (1IC6, green) and thermitase (1THM, red). Yellow spheres represent critical salt bridges, i.e. non-conserved interactions between oppositely charged groups of more than ten residues distance in the polypeptide chain, and grey spheres represent non-critical salt bridges. The catalytic triad, the disulfide bridges (orange) and the calcium ions (spheres) are also displayed as reference points.

Hydrogen bonds

Due to their large number, hydrogen bonds play a substantial role in the stability of proteins. The number and type of hydrogen bonds are frequently reported as factors correlated to temperature adaptation of proteins (Querol *et al.*, 1996; Kumar *et al.*, 2000) but the evidence is far from being clear cut (Szilagyi & Zavodszky, 2000; Gianese *et al.*, 2002). The total number of hydrogen bonds in the cold adapted 1SH7 is higher than that of 1IC6 and comparable to the hydrogen bond content of 1THM (Table 3). Furthermore, the number of sidechain-sidechain- and mainchain-sidechain hydrogen bonds was found to be by far the lowest in the mesophilic structure, 1IC6.

Calcium binding sites

The presence of bound calcium ions is a feature shared among members of the subtilisin superfamily, where calcium-binding has proven to be essential for the correct folding and structural stability (Betzl *et al.*, 1990; Siezen & Leunissen, 1997). Considering the stabilising effect of binding metal ions in many proteins, it might be expected that increased affinity and number of bound metal ions should correlate with thermostability of proteins. Differences in stability and kinetic properties between meso- and psychrophilic enzymes have in fact been related to fewer or weaker metal ion binding sites in the latter (Davail *et al.*, 1994; Feller *et al.*, 1994; Almog *et al.*, 2003). In the case of thermitase, differences in calcium binding was considered as one of the major reasons for the enhanced stability of the enzyme as compared to its mesophilic counterparts (Teplyakov *et al.*, 1990). Surprisingly, three calcium-ions are found associated with the structure of 1SH7, whereas 1IC6 and 1THM have two each (Figure 18, Table 5). At one of the binding sites called Ca1, which is analogous to the known strong calcium-binding site Ca1 in proteinase K (Betzl *et al.*, 1988), the calcium ion in 1SH7 is coordinated by O δ 1 and O δ 2 of Asp196, the carbonyl-oxygen of Pro171 and Gly173 and four water molecules. According to sequence alignments, this site is well conserved among the members of the family of proteinase K, including the enzymes of thermo- and mesophilic origin most related to the *Vibrio*-proteinase. The second calcium-binding site in 1SH7 corresponds to the described, second or medium strength calcium-binding site, Ca2 of 1THM (Teplyakov *et al.*, 1990). O δ 1 and O δ 2 of Asp61, O δ 1 of Asp56, the carbonyl oxygen of Asp63 and three water molecules coordinate the calcium ion. According to sequence alignments this calcium binding site would also be present in the highly homologous proteinases from *Vibrio alginolyticus* and *Vibrio*

cholerae, but absent in the thermophilic proteinase from *Thermus* Rt41a and aqualysin I from *Thermus aquaticus*. The third, additional calcium-binding site of 1SH7, Ca3 (Figure 18), has not been found in known structures of proteinases so far. The calcium ion links the α -helix A and residues of the succeeding surface loop and it is coordinated by the side chain and carbonyl oxygen of Asp9, the side chains of Asp12, Gln13, Asp19, the carbonyl oxygen of Asn21 and one water molecule in a pentagonal bipyramidal manner (Figure 22A). Sequence alignments indicate that this new calcium-binding site is most likely present in the closest relatives (Figure 22B). The presence of calcium in this site was confirmed by anomalous density map (Figure 23). Calcium binding plays a critical role in the stability of the *Vibrio*-proteinase, as in the case of related enzyme (Kristjánsson, personal communication). From the structural comparisons carried out here it is difficult however to deduce how or whether differences in calcium binding sites contribute to temperature adaptation of the enzymes involved.

Table 5. Calcium binding sites, Ca1, Ca2 and Ca3, in the *Vibrio* proteinase, 1SH7, and the corresponding sites in proteinase K, 1IC6, and thermitase, 1THM, coordinating atoms and their distances from calcium ions.

	1SH7			1IC6			1THM		
Ca1	P171	O	2.4 Å	P175	O	2.4 Å			
	G173	O	2.4 Å	V177	O	2.4 Å			
	D196	Oδ1	2.5 Å	D200	Oδ1	2.7 Å			
	D196	Oδ2	2.3 Å	D200	Oδ2	2.5 Å			
	Wat	O	2.4 Å	Wat	O	2.6 Å			
	Wat	O	2.5 Å	Wat	O	2.5 Å			
	Wat	O	2.5 Å	Wat	O	2.5 Å			
	Wat	O	2.3 Å	Wat	O	2.5 Å			
Ca2	D56	Oδ2	2.3 Å				D57	Oδ2	2.4 Å
	D61	Oδ1	2.5 Å				D62	Oδ1	2.5 Å
	D61	Oδ2	2.6 Å				D62	Oδ2	2.6 Å
	D63	O	2.3 Å				T64	O	2.4 Å
	Wat	O	2.4 Å				Q66	Oε1	2.4 Å
	Wat	O	2.4 Å				Wat	O	2.3 Å
	Wat	O	2.4 Å				Wat	O	2.2 Å
Ca3	D9	O	2.4 Å						
	D9	Oδ1	2.4 Å						
	D12	Oδ1	2.4 Å						
	Q13	Oε1	2.3 Å						
	D19	Oδ1	2.5 Å						
	N21	O	2.4 Å						
	Wat	O	2.3 Å						

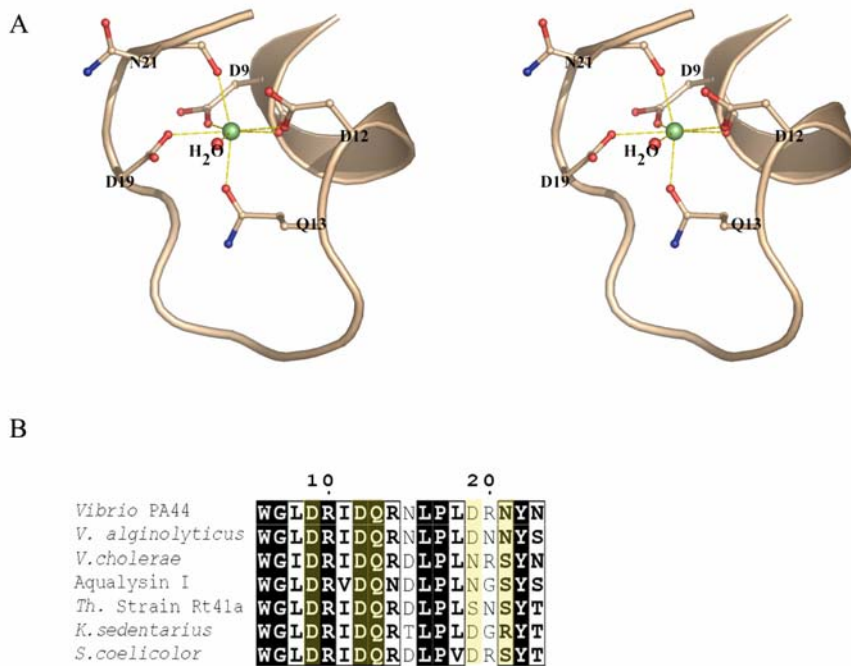


Figure 22. (A) A stereoview of the new calcium-binding site, Ca3, found in the structure of the *Vibrio*-proteinase. The calcium ion is coordinated in a pentagonal bipyramidal manner by the carboxyl groups of D9 and N21, the side chain oxygen atoms of D9, D12, Q13, D19 and one water molecule. (B) The sequence containing the residues forming Ca3 (shaded with yellow) in the *Vibrio*-proteinase is well conserved among the most related enzymes of meso- (proteinases from *Vibrio alginolyticus*, *Vibrio cholerae*, *Kytococcus sedentarius* and *Streptomyces coelicolor*) and thermophilic origin (aqualysin I from *Thermus aquaticus* and proteinase from *Thermus* sp. Rt41a).

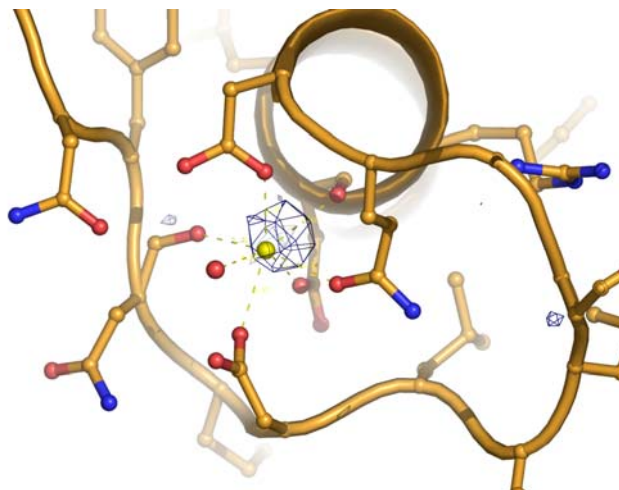


Figure 23. Anomalous map in the region of Ca3 in 1SH7. The calcium ion is drawn as a yellow sphere.

Disulfide bridges

There are three disulfide bridges in the structure of 1SH7 (Figure 18). In 1SH7 Cys67-Cys99 connects the loop carrying the Ca2 binding site and the loop containing

the residues of substrate binding pocket S4. The second disulfide bridge in 1SH7, Cys163-Cys194, bridges residues next to the Ca1 binding site and a region carrying residues of the substrate binding pocket S1. According to sequence alignment these two disulfide bridges are highly conserved among the most related enzymes of the *Vibrio*-proteinase including aqualysin I. The third disulfide bridge in 1SH7, Cys277-Cys281, is at the C-terminus. The structure of 1IC6 contains two disulfide bridges that, although non-identical to the ones found in 1SH7, also link parts of the structure directly connected to the substrate binding sites. There is no disulfide bridge in 1THM. The higher number of disulfides in 1SH7 relative to its related enzymes and the absence of such bonds in 1THM, does not give evidence of disulfides playing a critical role in the different temperature adaptation of the enzymes compared here. This is also in consent with what is seen in a psychrophilic subtilisin that contains the same or higher number of disulfide bridges as highly homologous mesophiles (Davail *et al.*, 1994; Almog *et al.*, 2003). Introduction of disulfide bridges by mutagenesis has only in rare cases resulted in increased stability (Bryan, 2000; D'Amico *et al.*, 2002). Based on a comparison on the reactivity towards sulfitolysis and dithiothreitol, the disulfide bridges of the *Vibrio*-proteinase had previously been suggested to be more accessible to solvent than proteinase K and the thermophilic aqualysin I (Kristjánsson *et al.*, 1999). This is confirmed by analysis of the surface accessibility of the disulfide bridges in the structures compared here, 1SH7 and 1IC6, and hence is assumed to apply as well to aqualysin I, which contains the two conserved disulfide bridges of 1SH7. The disulfide bridges in those enzymes are found in regions where many supposedly stabilising features, such as calcium-binding sites and ion pairs come together and they have both sequential and spatial proximity to parts involved in substrate binding. These facts among the observed differences regarding surface exposure of their disulfide bridges, while being absolutely crucial for the active conformation of the *Vibrio*-proteinase (Kristjánsson *et al.*, 1999), might have some relevance regarding temperature adaptation. First, this might reflect a tendency in the more stable enzymes to protect critical parts of the structure by decreasing their solvent accessibility. Second, the absence of disulfide bridges in THM is in line with the observed tendency of thermophilic enzymes to have a reduced occurrence of thermolabile residues (Vieille & Zeikus, 2001).

Surface properties and packing

The chemical properties of the groups comprising protein surfaces are expected to be important for adaptation of protein function both to high and low temperatures, as these determine the important interactions of the protein to water; interactions which are highly dependent on temperature as a result of changes in water structure (Makhatadze & Privalov, 1993; Privalov & Makhatadze, 1993; Tsai *et al.*, 2002). A larger fraction of polar surface area in a number of thermophilic proteins has been suggested to contribute to their increased stability (Vogt *et al.*, 1997; Haney *et al.*, 1999; Szilagyi & Zavodszky, 2000). In several cases, differences in surface charge distributions or an increase of nonpolar surface areas have been suggested to be relevant factors in adaptation to low temperatures (Davail *et al.*, 1994; Smalås *et al.*, 1994; Aghajari *et al.*, 1998; Russel *et al.*, 1998; Kim *et al.*, 1999). In citrate synthases adapted to different temperatures a clear trend was observed in reduced exposure of apolar surfaces in proceeding from psychrophile to hyperthermophile structures (Bell *et al.*, 2002). Thermo- and in particular hyperthermophilic proteins have been reported to have an improved packing and fewer and smaller cavities in their protein core relative to mesophiles (Szilagyi & Zavodszky, 2000). Other statistical approaches analysing structural parameters in large samples of dissimilar proteins regarding origin and temperature range, do not give evidence of significant trends, neither regarding polarity of protein surfaces or a different degree of packing (Kumar *et al.*, 2000; Gianese *et al.*, 2002; Kumar & Nussinov, 2004).

The cold adapted 1SH7 and the mesophilic 1IC6 have a larger solvent accessible surface area and a larger fraction of nonpolar surface area than 1THM (Table 3). Thus, among these enzymes the recurring trend in thermophilic enzymes to reduce exposed apolar surfaces is observed. The total area of buried surfaces is similar for the three enzymes, but their composition is different in that 1SH7 buries significantly less apolar surface as compared to either 1IC6 or 1THM. By the same token more buried surface area of the cold enzyme is polar than of either the meso- or thermophilic enzyme (Table 3). The larger buried apolar surface of 1IC6 and 1THM would be expected to contribute significantly to the higher stability of these enzymes through the hydrophobic effect. The effect of larger buried apolar surface can be estimated to be in the range 5.7 - 15.6 kcal/mol between 1IC6 and 1SH7 and 5.3 – 14.3 kcal/mol between 1SH7 and 1THM, when calculated as has been suggested by Criswell *et al.* (Criswell *et al.*, 2003).

Thus, the cold adapted enzyme would to a lesser extent be dependent on the hydrophobic effect for stability than its counterparts adapted to higher temperatures. In fact, Kristjánsson & Magnusson (2001) reached the same conclusion from their study of the effects on lyotropic salts on the stability of *Vibrio*-proteinase, proteinase K and the thermophilic homologue, aqualysin I. It remains debateable however, whether this observation, as well as reported cases of larger exposed apolar surfaces of cold enzymes, is merely a consequence of a diminished hydrophobic effect at low temperature, or if it is indeed a part of a molecular strategy of cold-adaptation. Due to ordering of water structure at low temperature (temperatures below approximately the temperature of maximum stability) the entropic penalty of exposing apolar surfaces is reduced and thus the hydrophobic effect (Tsai *et al.*, 2002). At these low temperatures destabilisation of the protein structure is therefore enthalpically controlled, both as a result of the ordered water structure (Tsai *et al.*, 2002), and through the interactions of the water with both apolar and polar groups of the protein (Makhatadze & Privalov, 1993; Privalov & Makhatadze, 1993; Graziano *et al.*, 1997; Robinson & Cho, 1999). Hence the entropically driven hydrophobic effect would be expected to contribute less to the overall stability of the proteins at low temperatures, or destabilize them locally or globally, which in effect may lead to more open and resilient structures.

A most noticeable difference in the surfaces of the proteins compared here resides in their different surface electrostatic potentials (Figure 24). Reflecting the different occurrence of noncompensated negative charges, as mentioned above and shown in Table 3, 1SH7 has large parts of the surface negatively charged in regions where 1IC6 and 1THM have less charged or positively charged surfaces. Furthermore the substrate-binding cleft of 1SH7 differs from that of 1IC6 and 1THM both in shape, being seemingly deeper and more distinct, and in being more negatively charged than the binding pockets of 1IC6 and 1THM (Figure 24). The biological implication of this difference with respect to different temperature adaptation is not clear. Interestingly however, the *Vibrio*-proteinase shares its anionic character with several other cold adapted enzymes (Leiros *et al.*, 2000; Arnórsdóttir *et al.*, 2002). More anionic charge of rat trypsinogen as compared to the bovine homologue has been suggested to be a source of increased flexibility of the former (Pasternak *et al.*, 1999). Also a group of highly flexible proteins, the natively unfolded proteins, are characterised by a large (predominantly negative) net charge (Uversky, 2002). It has been suggested that a

higher number of uncompensated charged residues on protein surfaces may contribute to cold adaptation by providing stronger interaction energy with the highly ordered water structure at low temperatures (Kumar & Nussinov, 2004). As reflected in a significant increase in surface tension and viscosity at low temperatures, water is optimally hydrogen bonded. The energetic cost of the dissolution of a protein under such conditions, arising from the unfavourable disruption of the optimised hydrogen bond network, may be offset by favourable electrostatic interactions of the charged groups with water at the protein surface (Kumar & Nussinov, 2004). Among amino acid residues only Arg is more soluble than Glu or Asp (Radzicka & Wolfenden, 1988). Thus, endowing the protein surface with their hydrophilic nature may be a way to enhance favourable electrostatic interaction with water at low temperature and at the same time it results in an anionic character, which may favour a more disordered or flexible structure.

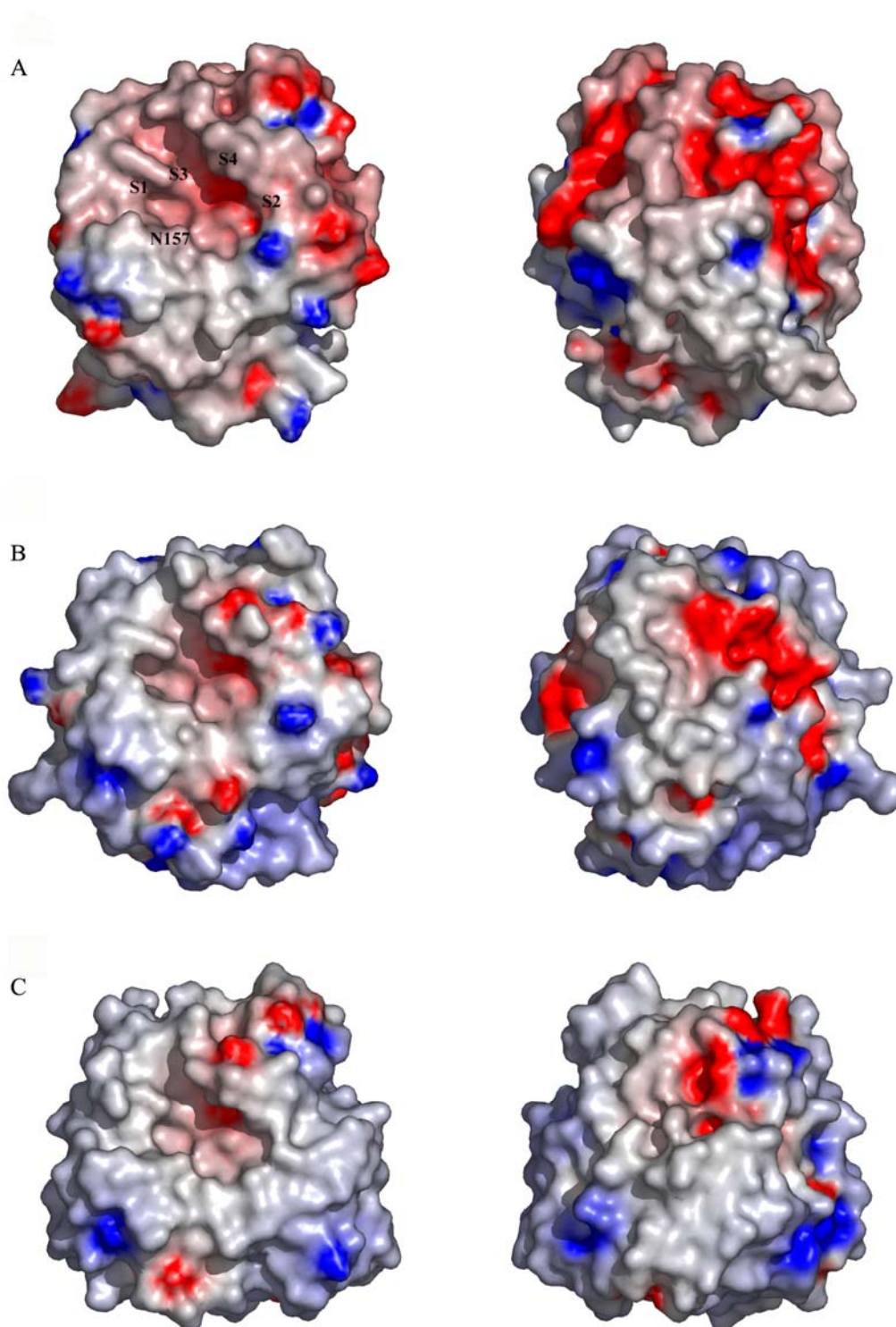


Figure 24. Comparison of the electrostatic surface potentials of (A) 1SH7, (B) 1IC6 and (C) 1THM. On the right hand side the molecules have been rotated 180° about the y-axis. The approximate locations of substrate binding pockets, S1-S4 (nomenclature according to Schechter & Berger, 1967) and the oxyanion hole residue, N157, are labelled on the surface of the *Vibrio*-proteinase (A). The positive potential is in blue and the negative potential is in red. The electrostatic surface potential was calculated with Delphi (Honig & Nicholls, 1995) and the graphical presentations were made in Pymol (DeLano, W. L. (2002). The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA, USA).

Protein preparation of p61, a component of the [U4/U6·U5] triple

snRNP

Expression and purification of p61

Expression vectors, pET28a and pGEX6P-1, carrying the p61-gene were obtained from Lührmann's research group at the Max Planck Institute in Göttingen. Initial expression tests using conventional expression methods in *E. coli* did not yield soluble protein. In order to improve the solubility of the protein, several alternative expression strategies were tried, such as expressing at temperatures as low as 15°C, varying inducer concentration and inducing at different cell densities. In addition, different *E. coli* strains were tested as well as strains containing extra copies of rare codon tRNA-genes (RIL and RP) and plasmids designed to enhance target protein solubility (pLysS and pLysE). The best results were obtained using RIL and RP cells and express at 15 °C overnight. However the yield was still so low that by SDS PAGE the protein bands corresponding to p61 were only vaguely or not visible by coomassie blue staining and Western blots had to be performed in order to analyse results of expression- and solubility tests. Since high amounts of pure protein are required for crystallisation trials these results were not satisfactory. A second clone of the p61-gene in the expression vector pGEX5X-1 was obtained from Bernhard Lagerbauer at Max Planck Institute in Martinsried. This clone contains four mutations in the nucleotide sequence yielding four amino acid substitutions as compared to the Lührmann-clone (Attachment A). The pGEX5X-1 construct contains a Factor Xa protease recognition site for cleavage of the GST-fusion protein from the target protein. More specific and less elaborate is the cleavage with Prescission protease, which is utilised in the pGEX6P-1 system. Therefore the p61-gene was subcloned into pGEX6P-1. Expression in the *E. coli* strain BL21 at 15 °C in 2YT medium yielded reasonable amounts of soluble protein (Figure 25).

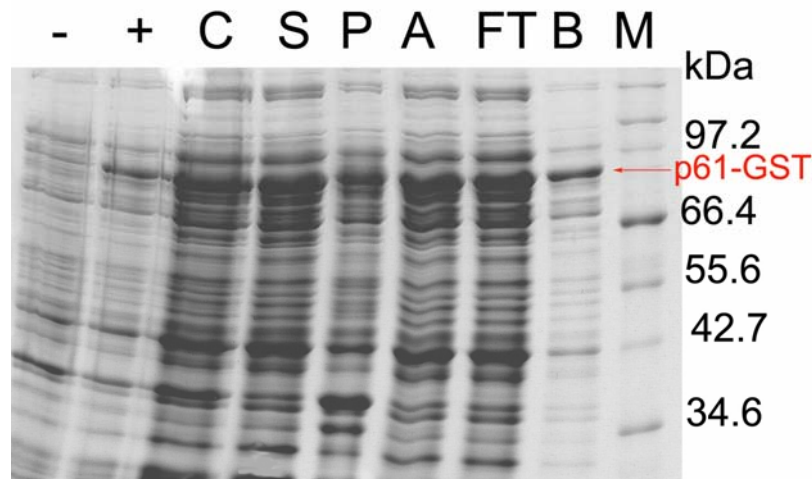


Figure 25. Expression of p61-gene from pGEX6P-1 in BL21(DE3) cells at 15°C in 2YT-medium and binding of the protein to GSH-beads. The lanes are labelled as follows: uninduced control (-), induced cell (+), crude cell lysate (C), soluble fraction (S), pelleted fraction (P), GSH bead after sample application (A), flow-through (FT), GSH-beads after wash (B). The 88 kDa p61-GST fusion protein is not visible in the uninduced control (-) but is seen after induction (+). Most of the p61-GST is in the soluble fraction of the cell lysate, but the binding to GSH-sepharose beads is poor as seen from the large amount of p61-GST in the non-bound fraction (FT).

The GST-tagged p61 was purified in batch with GSH-sepharose beads. The binding of the protein was incomplete (Figure 25). Precission protease cleavage of p61 from the GST-fusion protein still bound to the beads resulted in a relatively pure protein sample (Figure 26); however the 61 kDa protein was always accompanied with an approximately 70 kDa protein suggested to be DnaK. Efforts to get rid of the 70 kDa protein by additional washing steps (see page 48) were not successful (Figure 27). At this stage the amount of protein was approximately 2 mg from 1 l of expression culture.

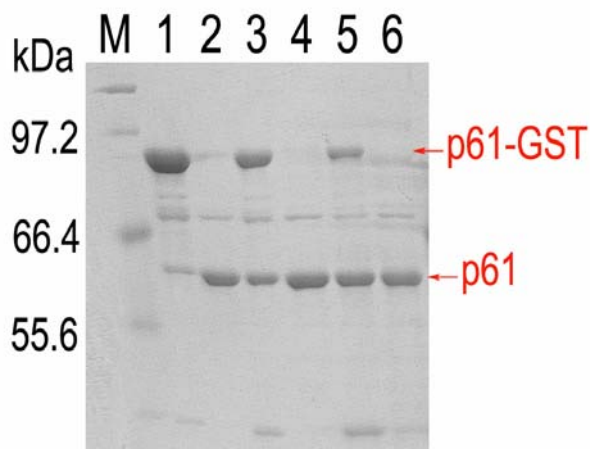


Figure 26. Separation of GST-fusion protein from p61B. GST-p61 bound to GSH-sepharose beads were incubated with Precission proteinase at 4 °C for 45 minutes and the supernatant of the beads containing the cleaved p61 was collected. This was repeated twice. Lane 1: sample of beads before Precission proteinase digestion, lane 2: supernatant of beads after 45 minutes incubation with Precission proteinase, lane 3: beads after first digestion, lane 4: supernatant after second digestion, lane 5: beads after second digestion, lane 6: supernatant after third digestion.

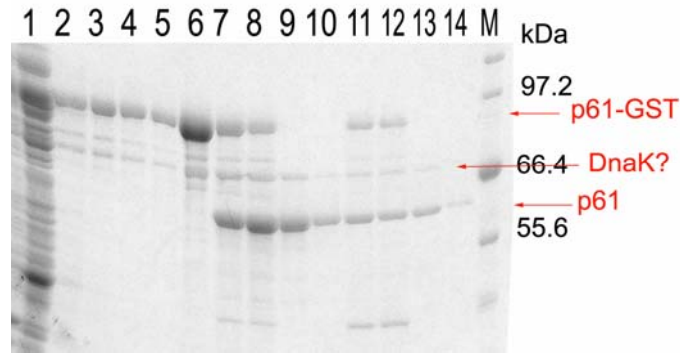


Figure 27. DnaK impurities.

According to wavelength scans performed on the protein samples after affinity purification, there were large amounts of nucleic acids present in the samples. In order to further purify the protein and separate the nucleic acids from it, the protein sample was applied on ion exchange and gel filtration columns. A large nucleic acid peak eluted from the anion exchange column but p61 did not elute in a distinct peak. Figure 28 shows a typical gel filtration profile; a large nucleic acid peak, containing only trace amount of p61, eluted in the void volume.

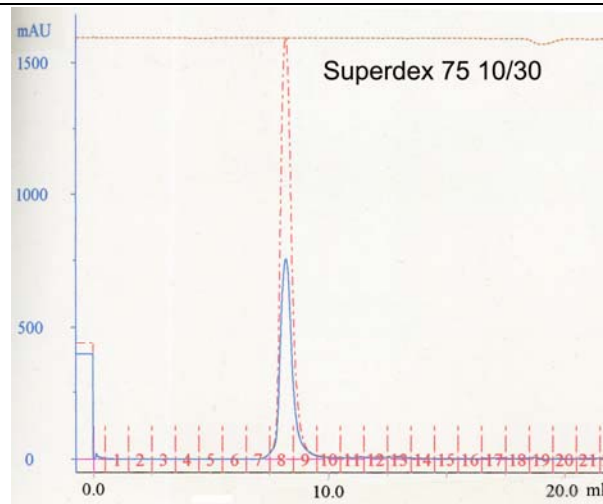


Figure 28. Superdex 75

Coexpression with 15.5 kDa protein

Based on the knowledge that the 15.5 kDa protein serves as a platform for the binding of p61 to the U4/U6 di-snRNP and some indications that 15.5 kDa might interact *in vitro* with p61 (Ralf Ficner, personal communication), attempts were made to enhance the solubility of the recombinant p61 by coexpression with the 15.5 kDa protein. The

15.5 kDa protein was overexpressed but p61 was not detectable with coomassie blue staining (Figure 29). Western blotting showed a protein-band of the size corresponding to p61 emerge after induction and the same protein bound to Talon beads (Figure 29). However, the amount of protein was far too little for further preparation for crystallisation trials.

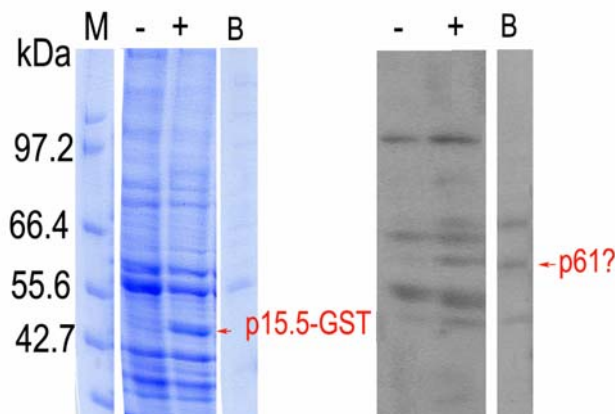


Figure 29. Coexpression of the genes encoding the 15.5 kDa protein and p61 from pGEX4T-2 and pET28a, respectively. No difference was seen between the uninduced control (-) and the induced culture regarding p61, whereas p15.5-GST was produced in large amounts. On a Western blot a protein band that could correspond to the p61 was seen in the induced culture and the same protein bound to Talon beads (B).

Pull down with the 15.5 kDa protein

The binding of the 15.5 kDa protein to p61 was tested by a pull down experiment. GST-tagged-p61 was coupled to GSH-sepharose and incubated with the purified 15.5 kDa. After incubation, the GSH-sepharose beads were centrifuged and samples were taken from supernatant and the pelleted beads. No complex formation between the 15.5 kDa protein and p61 was detected, since the 15.5 kDa protein was found almost exclusively in the supernatant (Figure 30).

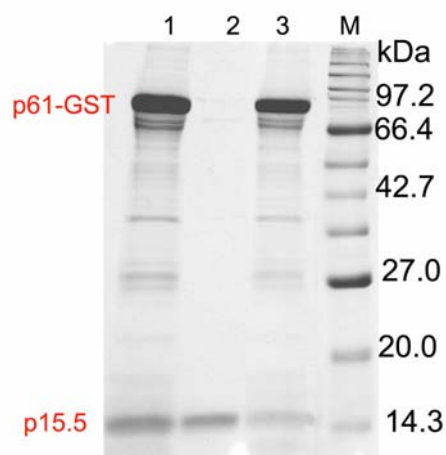


Figure 30. Binding test between p61 and p15.5. Lane 1: p61-GST was coupled to GSH-sepharose and incubated with the purified 15.5 kDa at 4 °C for 1 hour in 50 mM Tris pH 7.7, 100 mM NaCl, 10 mM MgCl₂, 2 mM DTT. Lane 2: supernatant of centrifuged beads. Lane 3: beads suspended in fresh buffer.

RNase stability test

Purification profiles from ion exchange and size exclusion chromatography indicated that p61 bound nucleic acids from *E. coli* during expression in an unspecific manner and that the stability of the protein in solution relied on this binding. If this was the case it meant that the protein sample was not homogenous, which is unfavourable for crystallisation. To test this, p61 was incubated with heat treated RNase. After digestion of RNA in p61 samples, no protein was detected in the solution (Figure 31). Hence, the stability of p61 is supposedly dependent on interactions with nucleic acids and/or protein binding partners.

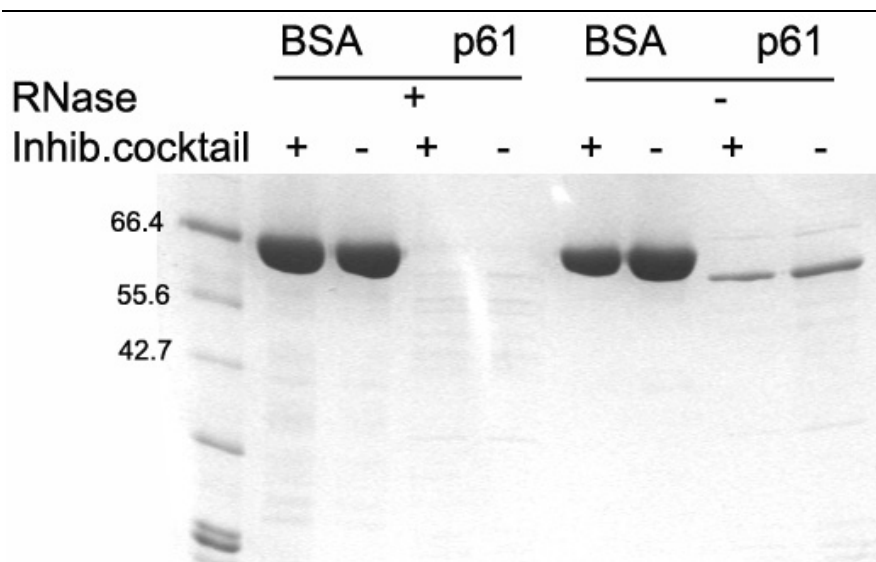


Figure 31. Protein p61 was incubated with heat treated RNaseA in 50 mM Tris pH 7.7, 100 mM NaCl, 10 mM MgCl₂, 2 mM DTT for one hour at room temperature with protease inhibition cocktail (lane 4) and without (lane 5). As a control BSA (bovine serum albumin) was treated with RNase in the same manner (lanes 2 and 3). Further controls are shown in lanes 6-9, where the proteins are incubated without RNase.

Crystallisation trials of p61

No crystals of p61 were obtained. In the majority of the crystallisation conditions tested, the protein precipitated immediately.

Protein preparation and crystallisation attempts on TbMP42, a component of the *Trypanosoma brucei* editosome

TbMP42 sequence

The TbMP42-gene is 1182 basepairs long encoding 393 amino acids protein. The cloned gene obtained from Dr. Göringer's lab in Darmstadt contains four mutations relative to the published sequence (Attachment B). The protein has a calculated molecular mass of 42 kDa but migrates at ~ 47 kDa on SDS PAGE. The sequence is predicted to contain an N-terminal mitochondrial import signal, two zinc fingers and a C-terminal OB-fold (Figure 32). The DNA sequence contains an *NcoI* restriction site, of which the ATG encodes Met250. This *NcoI* restriction site was used to construct a 143 amino acids C-terminal truncated form of TbMP42, containing the endo- and exonuclease activity.

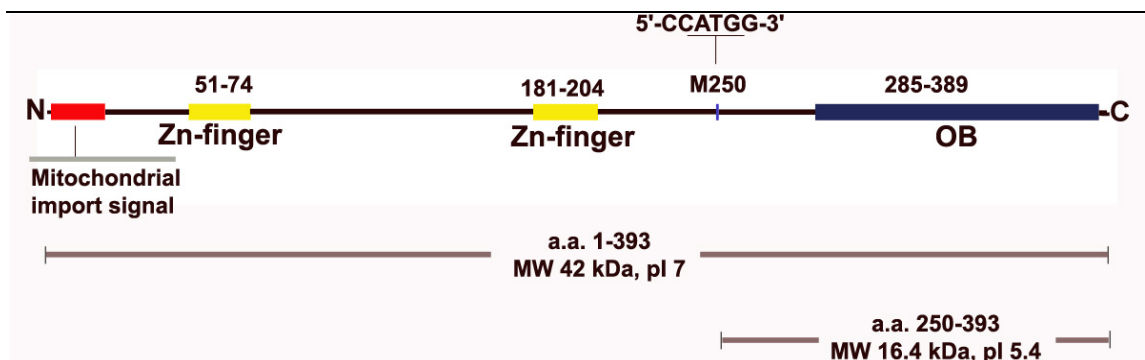


Figure 32. TbMP42 amino acid sequence properties.

BLAST searches to the NCBI (National Center for Biotechnology Information) non-redundant protein database revealed orthologous proteins from *Trypanosoma cruzi* and *Leishmania major*. Apart from these two orthologs and the other editosomal proteins (TbMP81, TbMP63, TbMP24 and TbMP18) no related proteins were found in the BLAST search. The sequence identities between TbMP42 and the orthologs from *T. cruzi* and *L. major* are 59 % and 49 %, respectively (Attachment B). The sequences which mark the zinc finger motifs are highly conserved and the C-terminal segments, comprising the exo-/endonuclease activity in TbMP42 are also similar in the three sequences (Figure 33).

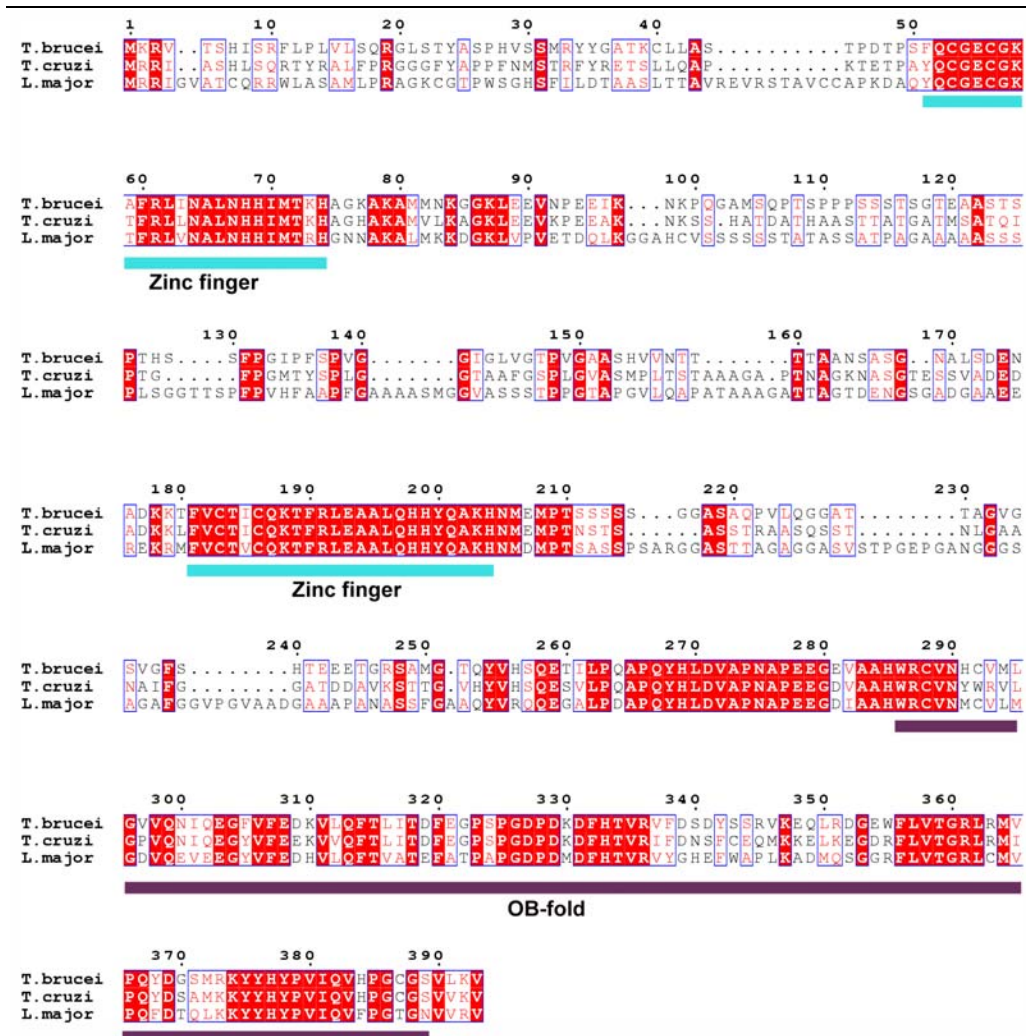


Figure 33. Sequence alignment of TbMP42 (*T. brucei*) and its relatives from *T. cruzi* and *L. major*. The sequences were aligned with ClustalW (<http://www.ebi.ac.uk/clustalw>) and processed with ESPrnt (<http://esprnt.ibcp.fr/ESPrnt/ESPrnt/>).

TbMP42 constructs

The TbMP42-gene was obtained cloned into the expression vector pQE-60. The expression of the TbMP42-gene from pQE-60 resulted in production of insoluble protein, which could be purified from inclusion bodies via IMAC (Immobilized Metal Affinity Chromatography) under denaturing conditions (see below) and renatured by dialysis. In order to produce TbMP42 in a soluble form with the help of fusion proteins, constructs of the protein fused to GST and MBP were generated. In addition to full-length constructs, C-terminal fragments of TbMP42 comprising the endo- and exonuclease activity (Michael Brecht, personal communication) were established. Figure 34 shows a summary of TbMP42 constructs used in the effort of making preparations of the protein suitable for crystallisation trials.

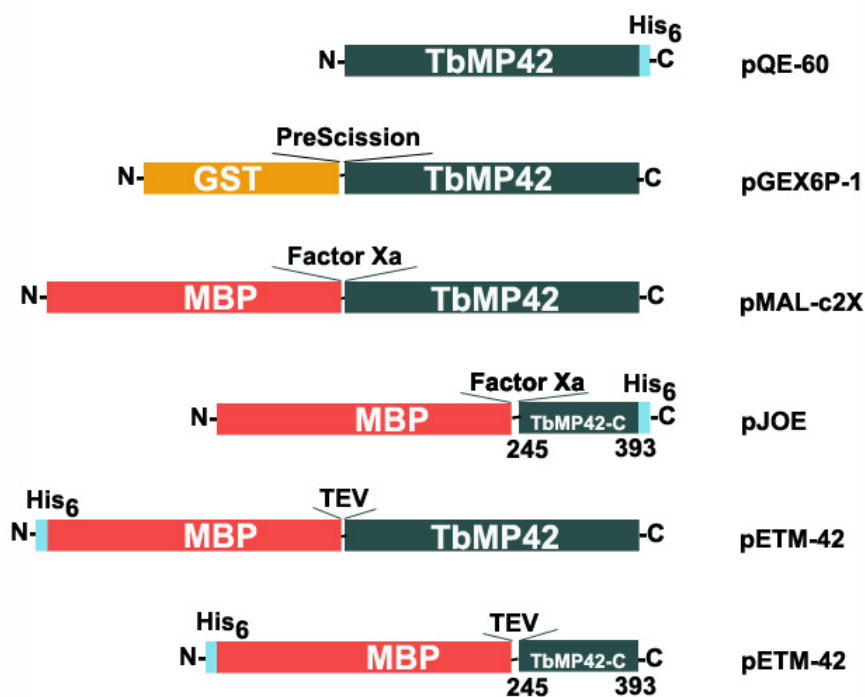


Figure 34. Summary of TbMP42 constructs used in the effort to get protein suitable for crystallisation trials. The TbMP42-gene was obtained cloned into the expression vector pQE-60, which provides His-tag for IMAC purification. The gene was amplified from the pQE-60 vector with primers designed for cloning the gene into the pGEX6P-1, providing a GST-fusion protein and a PreScission protease cleavage site, and pMal-c2X, providing a MBP-fusion protein and Factor Xa cleavage site. Due to problems regarding Factor Xa digestion, the gene was cloned into another system, pETM42, which utilises a His-tagged MBP-fusion protein that can be cleaved off with TEV-protease. Finally a fragment encoding residues 250-393 was cloned into the expression vectors pETM42 and pJOE, which provides MBP-fusion protein and a C-terminal His-tag.

Purification of TbMP42-His₆

The TbMP42-gene was overexpressed from pQE-60 yielding insoluble protein. Purification of the His-tagged protein under denaturing conditions resulted in high amounts (ca. 25 mg protein from 1 l of culture) of relatively pure (denatured) protein (Figure 35).

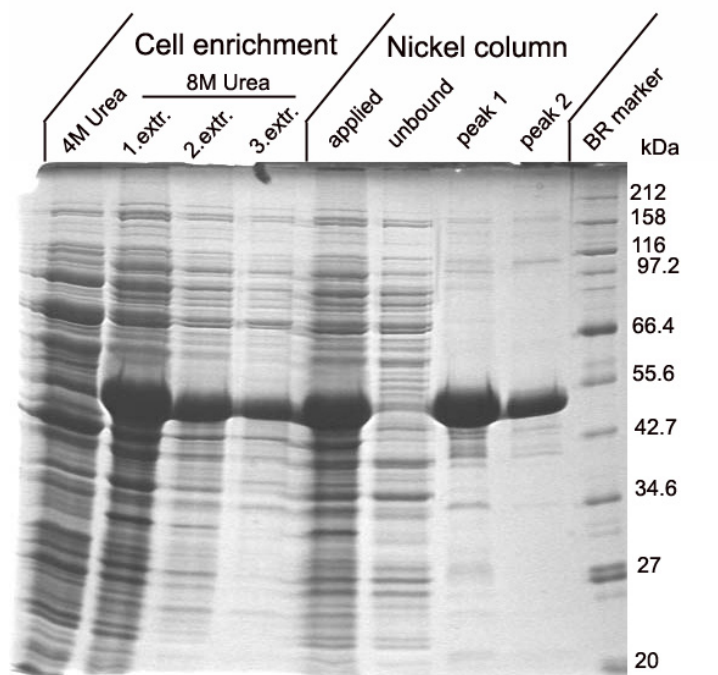


Figure 35. Purification of TbMP42-His₆ under denaturing conditions. The cell pellet was resuspended in a buffer containing 4 M urea, which extracted a large amount of host proteins but no TbMP42. The pelleted proteins remaining were extracted with 8 M urea and applied on a Hitrap Chelating Ni⁺² column at pH 8. After a washing step with buffer at pH 5.9, TbMP42 was eluted in two peaks by lowering the pH in a stepwise manner to pH 4.5.

The protein solution was dialysed against a buffer without a denaturing agent in order for renaturation to occur. Most of the protein remained in solution during and after the dialysis (Figure 36). Finally, the protein was applied on a gel filtration column. The profiles obtained from gel filtration experiments (Figure 37) showed that the protein formed soluble aggregates (particle size > 1000 kDa), since it eluted beyond the exclusion limit of any gel filtration column available, including Superose 6. In the effort of solving these aggregates, the buffers' salt concentrations and pH were varied (in the range of 0.1 – 2.0 M NaCl and pH 6 – 8.5, respectively), without success, however. The protein's tendency to aggregate is probably due to incorrect or at least incomplete folding during the renaturing step.

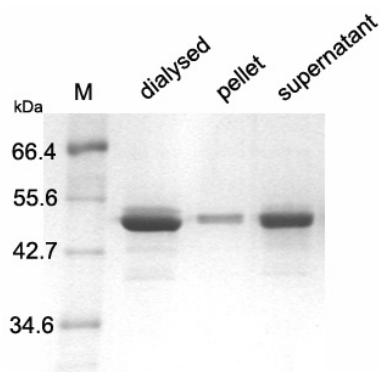


Figure 36. Dialysed protein solution. The protein purified under denaturing conditions was renatured by dialysis against a buffer without a denaturing agent, 20 mM HEPES pH 7.4, 100 mM KCl, 1 mM ZnSO₄. White precipitate, probably Zn(OH)₂ forms in the dialysed solution. The dialysed solution was centrifuged and the pellet and supernatant analysed by SDS PAGE. The protein is mainly in the soluble fraction.

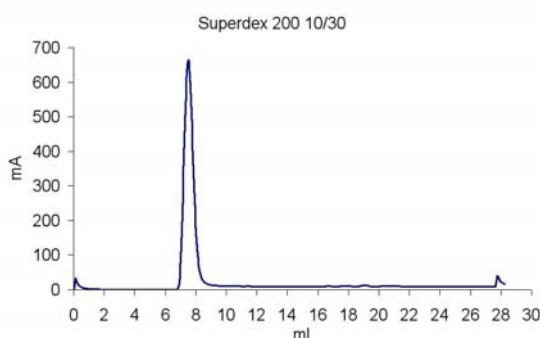


Figure 37. Gel filtration profile of renatured TbMP42-His₆ on Superdex 200 (10/30). The protein runs beyond the exclusion limit indicating that it forms aggregates. Identical profiles were obtained for Superdex 75 and Superose 6. Superose 6 can, according to producer (Amersham), separate proteins in the range of 0.5 – 5000 kDa.

Crystallisation trials of TbMP42-His₆

Most conditions of the crystallisation screens tested resulted in immediate brown precipitation, reflecting denatured protein. This is not surprising in view of other evidence (see above) of imperfect folding. In a few conditions, however, non-diffracting crystals were obtained (Figure 38). Attempts to optimise crystallisation conditions and improve crystals with seeding techniques were not successful.

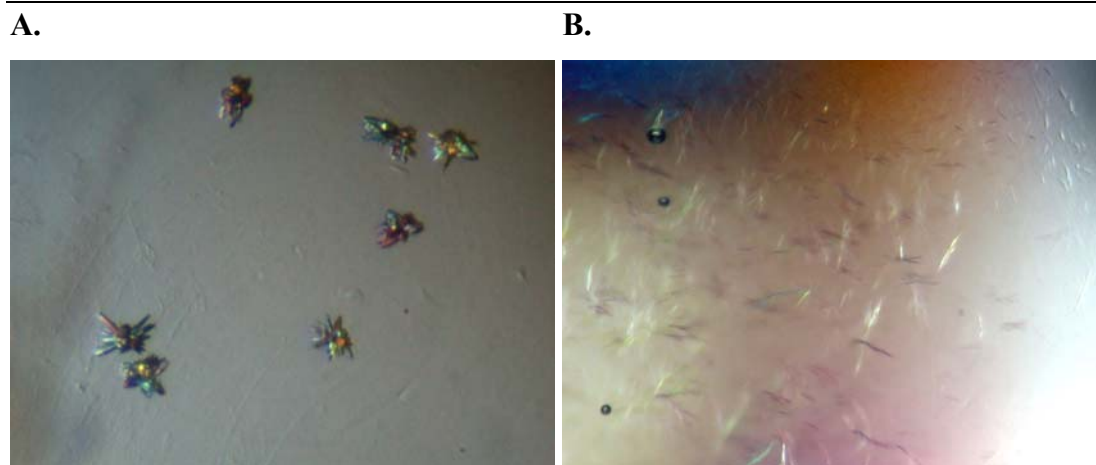


Figure 38. **A.** Clusters of small crystals grew at 20 °C in 18 % PEG 8000, 0.1 M cacodylate pH 6.5, 0.2 M ZnOAc. **B.** Crystalline material formed in 24 % PEG 1500, 20 % glycerol (condition 43 of Hampton Crystal Screen Cryo).

Purification of TbMP42-GST

Different expression conditions were tested for the production of TbMP42-GST from pGEX6P-1. These experiments included the testing of several *E. coli* strains, varying expression temperature and addition of ethanol, sucrose, sorbitol and ZnCl₂ to the expression cultures (Figure 39).

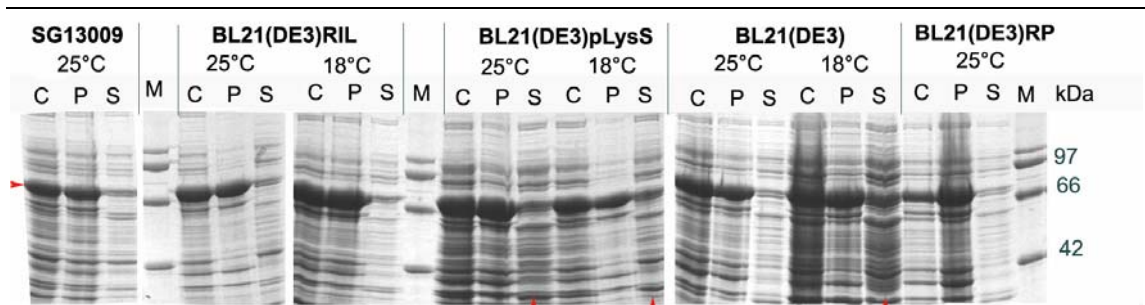


Figure 39. Expression tests with different *E. coli* strains at 18 °C and 25 °C. C is crude cell lysate, P is insoluble fraction of cell lysate and S is soluble fraction. TbMP42-GST is marked with a red arrow to the left. The protein is in most cases in the insoluble fraction but in BL21(DE2) and BL21(DE3)pLysS there is detectable amount in the soluble fraction.

The expression in BL21(DE2) and BL21(DE3)pLysS was further optimised yielding satisfactory amount of soluble protein (Figure 40). However, the binding of the GST-fusion protein to GSH was poor and PreScission proteinase cleavage failed (Figure 41). This might reflect incomplete folding of the protein, plausibly leading to soluble aggregates, in which the digestion site is not accessible for the PreScission proteinase.

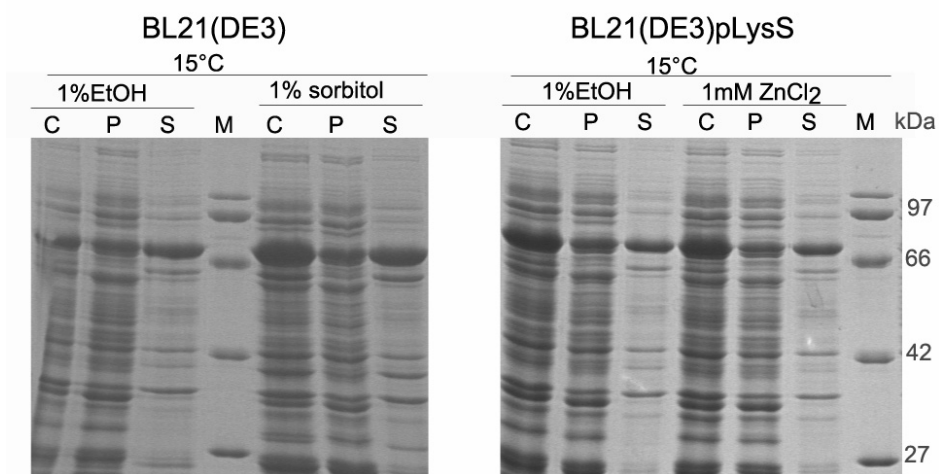


Figure 40. Optimisation of expression in the strains, BL21(DE2) and BL21(DE3)pLysS.

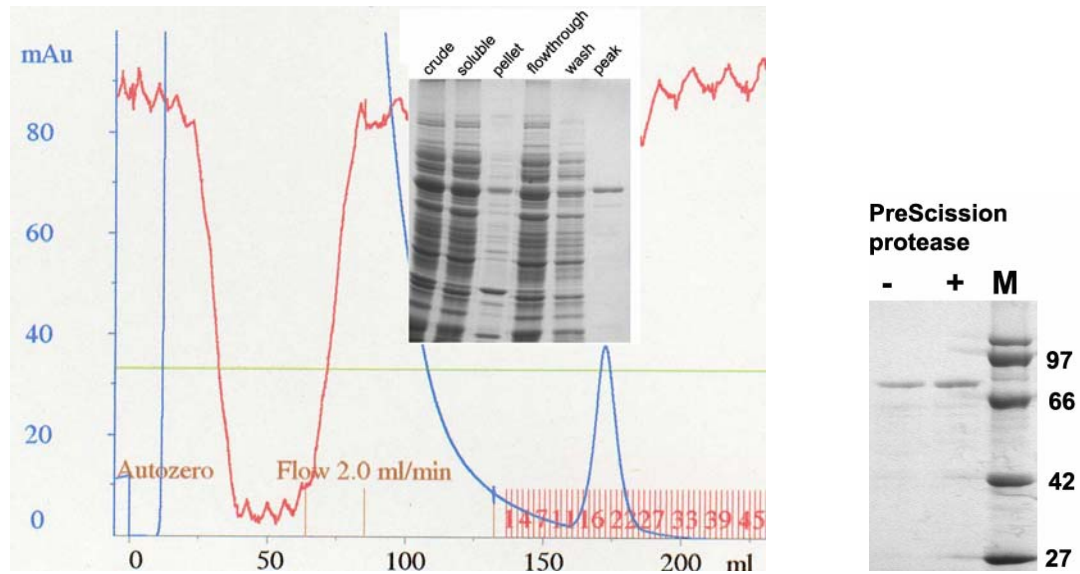


Figure 41. Purification of TbMP42-GST on GSH sepharose and subsequent PreScission protease cleavage. TbMP42-GST does not bind properly to the GSH sepharose as most of the protein is in nonbound fraction (flowthrough). PreScission protease fails to cleave of the GST fusion protein; no difference is seen on SDS PAGE before (-) and after (+) incubation with PreScission protease.

Purification of TbMP42-MBP

Overexpression of the TbMP42-gene from pMal-c2X resulted in large amounts of soluble TbMP42-MBP, which was purified on amylose-resin yielding ca. 30-40 mg of protein from 1 l culture (Figure 42).

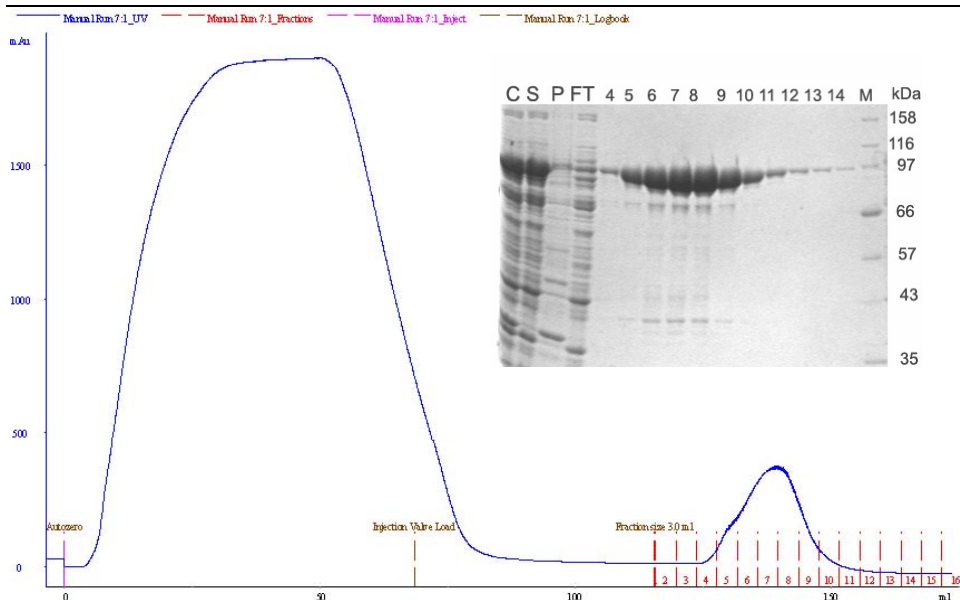


Figure 42. Purification of TbMP42-MBP on amylose-resin. The lanes on the SDS PAGE gel are labelled as follows: C is crude cell lysate, S is soluble fraction and P insoluble fraction of cell lysate, FT is the amylose column flowthrough, 4-14 are the corresponding fractions on the purification profile.

The TbMP42-MBP construct from pMal-c2X carries the recognition site IEGR, designed for specific cleavage of the fusion protein with Factor Xa. In fact, the meagre specificity of Factor Xa is well known and this was also the case here, where the proteinase apparently cleaved at two GR sites in the target protein (Figure 43). Attempts to separate the resulting protein fragments by ion-exchange-, hydrophobic interaction- and size exclusion chromatography were not successful. Since it was not possible to separate MBP from TbMP42 with Factor Xa, it was decided to try and crystallise the TbMP42-MBP chimera.

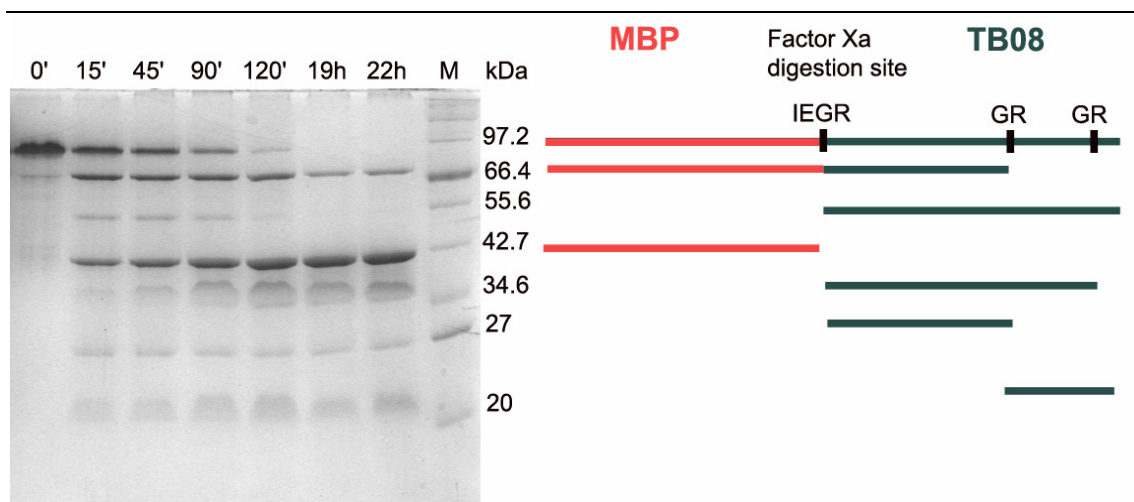


Figure 43. Factor Xa digestion. Incubation of TbMP42-MBP with Factor Xa results in generation of several fragments of TbMP42 due to unspecific cleavage of Factor Xa at two GR sites in TbMP42.

As a final purification step the protein was applied on a gel filtration column. The protein eluted beyond the exclusion volume at salt concentrations in the range of 0.2 - 2 M NaCl indicating that the protein sample was clumped in particles of >600 kDa. With decreasing salt concentration a shoulder of the main peak emerged, which became a distinct peak in a buffer containing no NaCl (Figure 44). This peak was very broad indicating that it might consist of different oligomers but its maximum corresponds approximately to a tetramer. The protein from the peak containing the presumed tetrameric TbMP42-MBP chimera was collected and concentrated for crystallisation trials.

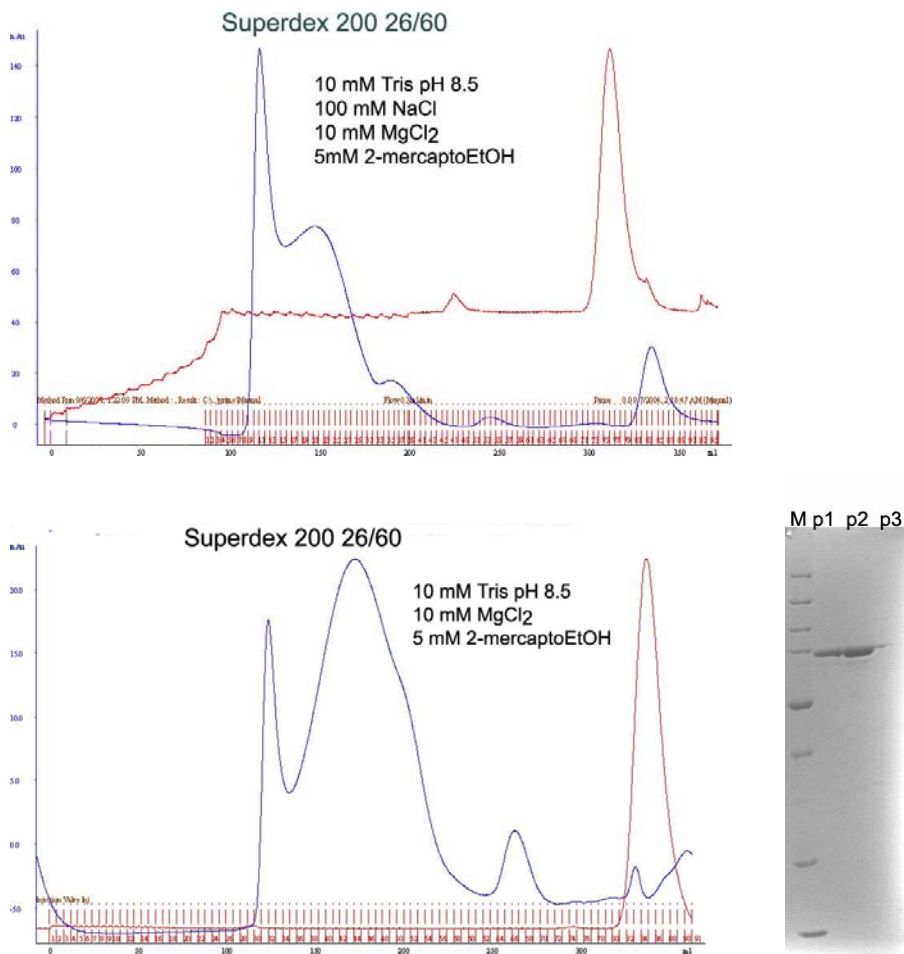


Figure 44. Purification profiles of TbMP42-MBP on Superdex 200. At 100 mM NaCl concentration there was a shoulder (upper profile) on the main peak which became more distinct in a gel filtration performed without NaCl (lower profile). The third peak presumably containing monomeric protein chimera also became more pronounced without NaCl. All three peaks, p1- the aggregated protein, p2 – the presumed tetramer and p3 the monomeric chimera, contained pure protein as seen by the SDS PAGE analysis. However, the peaks were considerably broad, which might indicate non-homogenous composition regarding oligomer size.

In order to overcome the problem of cleaving off the MBP fusion protein, the TbMP42-gene was cloned into still another vector, pETM42, which provides a TEV proteinase recognition site. The gene was overexpressed and the protein was purified in the same manner as described for the construct from pMal-c2X. No TEV proteinase cleavage was detected (Figure 45). The failure of TEV proteinase to cleave the fusion protein might be explained by poor exposure of the digestion site reflecting its tendency to aggregate, as is seen for other constructs as well.



Figure 47. Crystals obtained in condition B6 of JBScreen 1 (30 % PEG 400, 0.1 M Tris pH 8.5, 0.2 M Sodium Citrate). The crystals were too fragile to be mounted on the rotating anode. Numerous attempts to reproduce the crystals failed.



Figure 48. Crystals obtained in condition nr. 24 of the Magic Screen (30 % PEG 400, 0.1 M CAPS pH 10.5). The crystals were not big enough to be fished out of the drop. The crystals could not be reproduced.



Figure 49. Needle like crystals grew during initial screening in conditions: Natrix nr. 33 (30 % Hexanediol, 0.2 M Ammonium Chloride, 0.01 M Magnesium Chloride, 50 mM HEPES pH 7.0), Crystal Screen 2 nr. 19 (2.5 M 1,6 Hexanediol, 0.1 M Sodium Citrate pH 5.6) and nr. 39 (3.4 M 1,6 Hexanediol, 0.1 M Tris pH 8.5 and 0.2 M Magnesium Chloride). The crystals were well reproducible over a relatively wide range of pH (7-10) and with different additives. No condition was found however that gave three-dimensional crystals and seeding attempts were not successful.

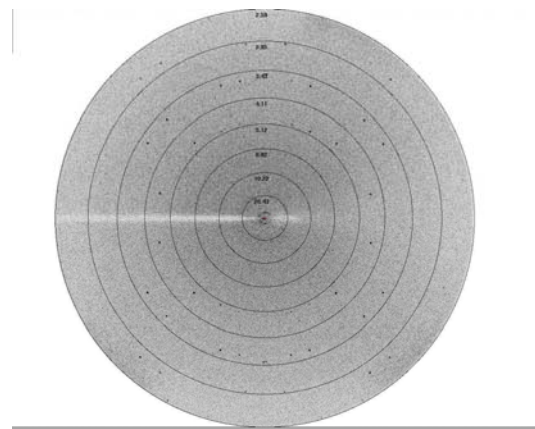
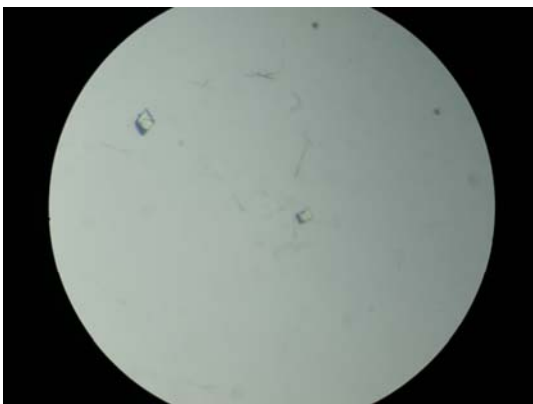


Figure 50. In drops containing both nucleic acid substrate and protein, both crystal-needles, identical to the ones obtained without substrate, and salt crystals grew. The diffraction pattern for the salt crystal is shown to the right. The low resolution limit (~ 5 Å) indicates a relatively big asymmetric unit for a salt.

TbMP42_C-terminal (250-393)-MBP

The gene-fragments encoding the C-terminal part (residues 250-393) of TbMP42, containing the endo/exonuclease activity, were cloned into pETM42 and pJOE. Expression-experiments exclusively yielded insoluble protein. This is in line with results of similar attempts performed by Göringer's research group in Darmstadt (Michael Brecht, personal communication).

Discussion

The cold adapted *Vibrio*-proteinase

Structure comparisons of the *Vibrio*-proteinase, proteinase K and thermitase

From the comparison of the three subtilases in this study, some structural differences were observed, which stand out and may be important for their temperature adaptation. First, whereas the overall exposed surface areas of the psychro- and the mesophilic enzymes are larger than for the thermophile enzyme, mainly as a result of larger area of apolar atoms, the meso- and thermophilic enzymes bury significantly more apolar surface in their folded structures as compared to the cold enzyme. Hence, the higher extent of hydrophobic interactions in the meso- and thermophilic proteins is suggested to contribute to the increased stability of the meso- and thermophilic enzymes relative to the cold adapted *Vibrio*-proteinase. This conclusion is in line with previous experimental results obtained on the effects of lyotropic salts on conformational stability of the *Vibrio*-proteinase, proteinase K and the thermophilic relative, aqualysin I, where the cold enzyme was shown to be less dependent on hydrophobic interactions for structural stability than its counterparts of higher temperature origin (Kristjánsson & Magnússon, 2001). Furthermore, this finding was supported by comparative sequence analysis (Arnórsdóttir *et al.*, 2002). These results also agree with the thermodynamics of the hydrophobic effect in protein stabilization, being enforced with rising temperature and thus stabilising structures at high temperatures, at least to a certain level, but diminishing in strength at lower temperatures. The diminished hydrophobic effect at low temperatures may account for the larger exposure of apolar surfaces observed in the *Vibrio*-proteinase and several other reported cold adapted enzymes (Feller *et al.*, 1994; Smalås *et al.*, 1994; Feller *et al.*, 1997; Aghajari *et al.*, 1998), relative to enzymes adapted to higher temperatures. Regarding the proposed role of increased exposed apolar surface areas as a mechanism of cold adaptation, it should be considered that interactions of such surfaces with water at low temperatures may be quite different to what could be observed at higher temperatures as a result of temperature dependence of the properties of water. According to Robinson & Cho (1999) polar surface groups give rise to a lower entropy and lower enthalpy in the surrounding water structure, whereas apolar groups would have the opposite effect. Whether this proposed effect of apolar groups, in promoting less order in the highly ordered water structure at low temperatures, influences protein motions remains to be determined.

Another surface property in which the *Vibrio*-proteinase differs with respect to the other enzymes compared in this study is its increased anionic character. Cold adapted enzymes are frequently found to be more anionic than their homologues adapted to higher temperatures. It is unclear however, whether this property makes any contribution to cold adaptation. Anionic character has been suggested to promote flexibility in trypsinogens, but a possible mechanism for this observation was not provided (Pasternak *et al.*, 1999). Kumar and Nussinov (2004) have pointed out the possible dual roles of electrostatics in adaptation of proteins to both high and low temperatures. In cold enzymes it was suggested that charges could ensure proper solvation against higher surface tension and viscosity characterising water at low temperatures and might also impart greater flexibility, especially in active site regions (Kumar & Nussinov, 2004). Interestingly, the analysis of the amount and pattern of electrostatic forces in the enzymes compared here support this view.

Interactions at the protein-water interface are crucial for both function and stability of proteins. These interactions are affected by temperature not least as a result of changes in the structure and consequently the properties of water. Thus, part of the molecular strategies in temperature adaptation of proteins must be expected to involve changes aimed at accommodating the temperature dependent changes in the structure and physical properties of water. Clearly more information is needed in this area to gain a better insight into the forces that facilitate cold adaptation in proteins.

Previous mutagenic experiments reviewed

Prior to the crystal structure determination of the *Vibrio*-proteinase (Arnórsdóttir *et al.*, 2005), the first experiments of an ongoing mutagenic study, aimed at shedding light on the molecular principles of cold adaptation in VPR, were performed (Arnórsdóttir and Kristjánsson, unpublished data). Based on comparisons of amino acid sequences and homology models¹ of VPR and the most related enzymes of meso- and thermophilic origin², VPR mutants were rationally designed. The mutants were designed with the aim of increasing structural stability relative to the wild type enzyme. Most notably, the comparison revealed frequent Ser_{VPR}→Ala_{therm} amino acid exchanges between the cold adapted VPR and the thermostable enzymes (Arnórsdóttir *et al.*, 2002). In eight sites

¹ Homology models were derived from the known structure of proteinase K, the 0.98 Å resolution 1IC6

² The proteinases from *V. alginolyticus*, *V. cholerae*, *Thermus* Rt41a and aqualysin I from *Thermus aquaticus* with 60-87 % sequence identity to VPR

there is Ser in VPR, where Ala occupies the equivalent site in aqualysin I. Such Ser to Ala exchanges occur at seven sites between VPR the *Thermus* Rt41a proteinase. Ser-Ala exchanges have been reported as highly correlated to thermostability in other comparative studies (Argos *et al.*, 1979; Peek *et al.*, 1992; Haney *et al.*, 1999). Another trend related to thermostability was seen in the number of proline residues in surface loops. The thermophilic enzymes contain five proline residues, which are absent in VPR, four of which are located in surface loops (Arnórsdóttir *et al.*, 2002). Three VPR mutants with a single Ser→Ala substitution were constructed and two mutants were made where a thermophile's proline site in a surface loop was imitated in VPR. All five mutants had increased stability over the temperature range measured (Figure 51), a higher value of $T_{50\%}$ (the temperature of 50 % reduction of activity in 30 minutes) and reduced catalytic efficiency in comparison to the wild type enzyme. The effects of the mutations were predicted and interpreted on molecular level with the help of computer graphics of homology models. The determination of the crystal structure of VPR now allows a more focused design of mutants and more reliable examination of the results of mutagenic experiments. The first site directed mutagenic experiment served partly to establish the method and were thus meant to be relatively "harmless".

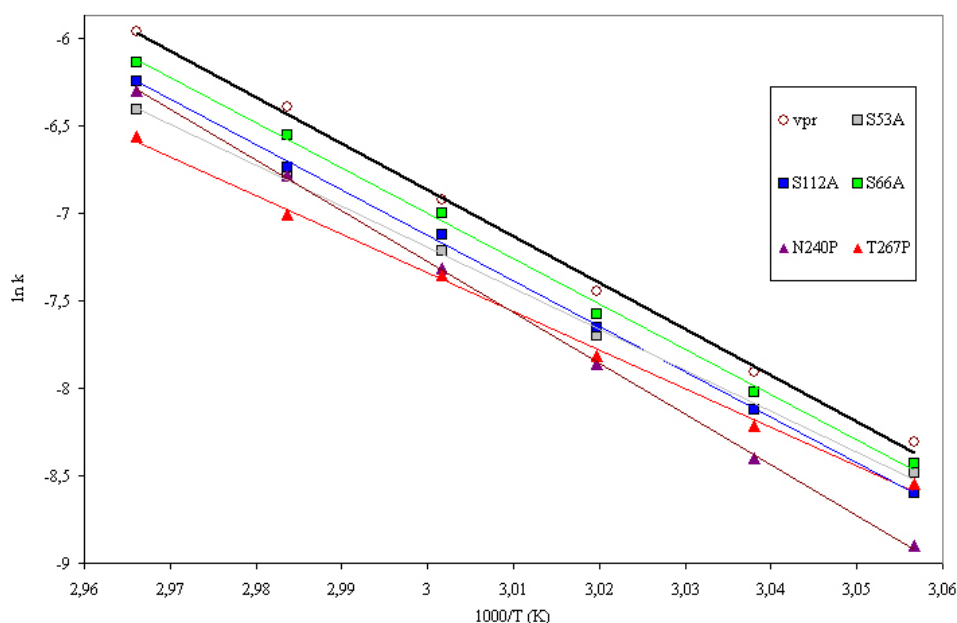


Figure 51. Arrhenius plot for thermal inactivation of the wild type VPR, three Ser→Ala mutants and two Xaa→Pro mutants. The enzymes, dissolved in 25 mM Tris, pH 8, containing 15 mM CaCl₂, 1 mM EDTA and 0.1 M NaCl, were heated at selected temperatures and the first-order rate constants of inactivation determined at each temperature for constructing the plot. The slope of the plot reflects activation energy of inactivation. Figure taken from Arnórsdóttir, Master Thesis, 2001.

Ser→Ala mutants

Ser51 in VPR was substituted with Ala. Ala occupies the equivalent site in most related enzymes, but in rare cases there is another hydrophobic residue found in this site. This site is located on a β -strand with little or no surface accessibility. The stability increase of the Ser51→Ala was related to disturbance of neighbouring hydrophobic interactions caused by the VPR_{wt} Ser in this site. The crystal structure of VPR supports this. The side chain of Ser51 is directed into a hydrophobic cluster and does not have any polar contacts to other groups.

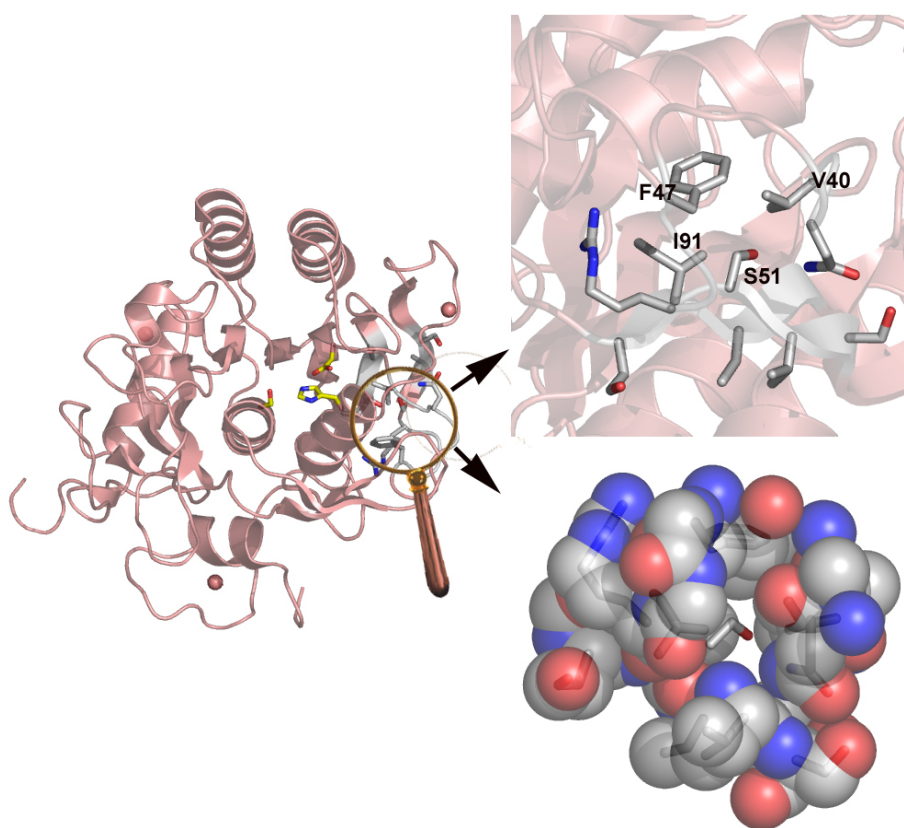


Figure 52. The Ser51 site. Shown are: the location of the mutated site in the molecule (left) with the catalytic triad residues (yellow) as reference points, a stick representation of the residues neighbouring the mutated Ser (upper right) and a sphere representation of the same residues where carbon atoms are coloured grey, nitrogen blue and oxygen red.

Ser64 is located on a loop connected to the α -helix carrying the active His (Figure 53). The loop projects to the surface and hosts a calcium binding site, which is also present in thermitase and probably in the closest relatives of VPR as well. A disulfide bridge connects this loop to another loop which hosts the residues of the S4 substrate binding site. VPR is the only enzyme among related enzymes that has Ser in this site; the other

enzymes have Ala or Gly in the equivalent site. Proteinase K has sequence deletions in this region relative to thermitase and VPR and does not contain a comparable loop structure. The Ser64→Ala mutant had a drastically reduced catalytic efficiency; mostly due to a large increase in K_m . Even though the crystal structure reveals that the loop was rather inaccurately modelled, it does not improve our understanding of the influence of this mutation on the substrate affinity of VPR. The effect of this mutation must be assigned to the location of the site in close proximity to important structural factors such as calcium binding, disulfide bridges and substrate binding sites.

Ser110 is on an α -helix connected to the loop that comprises the S4 substrate binding site (Figure 53). This site is occupied with Ala in most of the related enzymes. Ala is an α -helix stabilising residue, whilst Ser is regarded as α -helix destabilising. This has been supported by mutagenic studies, where Ala→Ser substitutions in α -helices most often result in destabilisation (Blaber *et al.*, 1993). Surprisingly, the Ser110→Ala mutation led to a drastic decrease in k_{cat} (80 %) but a moderate increase in $T_{50\%}$. This could indicate a local rigidifying effect hampering conformational changes required by the catalytic mechanism.

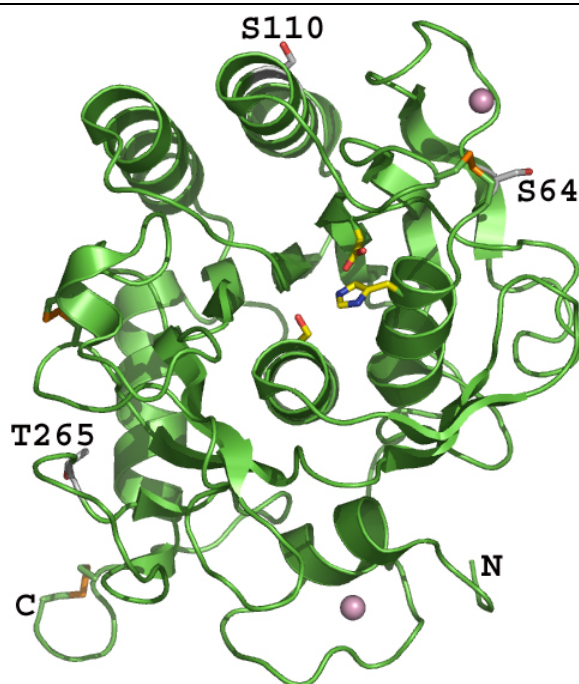


Figure 53. Locations of S64, S110 and T265 (grey) in the structure of the *Vibrio*-proteinase. Residues of the catalytic triad are yellow, the disulfide bridges are shown with orange sulphur atom and calcium ions are coloured pink.

Xaa→Pro mutants

The proline mutations had more pronounced effect on thermostability than the Ser to Ala changes. Furthermore, the stability of the proline mutants showed different temperature dependence as reflected in changes in activation energy (E_a) of heat induced inactivation (Figure 51).

Asn238 is located in a loop connected with α -helix E (Figure 18, page 59) that carries the active Ser. The activity loss of the Asn238→Pro mutant is more temperature dependent than the wild type enzyme and the other mutants. Thus, the stability increases considerably at low temperatures but approaches that of the wild type enzyme with increased temperature. By replacing Asn238 with Pro, hydrogen bonds made by Asn238 with backbone carbonyl-oxygens of Leu234 and Gln235 are lost. In addition, the surface becomes less polar in this area and hydrogen bonds between Asn and water molecules are lost. The changes in the surface potential and loss of hydrogen bonds may be related to the difference in temperature dependence of the thermostability of the mutant. Thus, the stability gain of rigidifying the loop might weigh more than the unfavourable loss of hydrogen bonds at low temperature. With increasing temperature the stabilising effect of proline might be more strongly counterbalanced by the loss of hydrogen bonding (Figure 54). Interpolation of the Arrhenius plots (Figure 51) shows that at ~ 79 °C, the mutant has the same half life as the wild type enzyme.

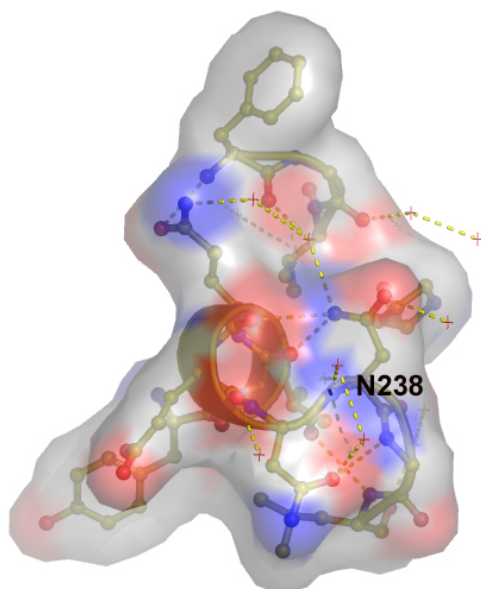


Figure 54. The Asn238 site. Polar interaction of Asn238 and neighbouring residues are shown with dashed lines. Water molecules are marked with red crosses.

Thr265 is on a loop connecting the strands in β -sheet III close to the C-terminal (Figure 53). The Thr265 \rightarrow Pro mutation has comparable effect on $T_{50\%}$ to that of the Asn238 \rightarrow Pro mutation. However, unlike Asn238 \rightarrow Pro, the Thr265 \rightarrow Pro mutant is less temperature dependent for activity loss than the wild type enzyme (Figure 51). At ~ 49 °C this mutant has a similar half life as the wild type enzyme but with increasing temperature its stability increases significantly relative to the wild type enzyme.

The different effects of the two proline mutations reflect different entropy and enthalpy contributions to E_a of inactivation. Thus, the enthalpy contribution appears to be more important in the Asn238 \rightarrow Pro mutant, whilst the entropy factor dominates in the E_a change in the Thr265 \rightarrow Pro mutant. This can be rationalised by examination of the locations and interactions at the mutated sites. In the case of the Asn238 \rightarrow Pro, hydrogen bonds were lost both between the Asn238 side chain and backbone atoms as well as with solvent molecules. The only residue that Thr265 can have polar interaction with is the flexible Lys267. The side chain nitrogen of Lys267 forms a hydrogen bond with O_γ of Thr265 only in one of the two molecules in the asymmetric unit. In the other molecule the side chain of Lys267 has another conformation and forms a hydrogen bond with the carbonyl oxygen of Asn250, which is located on α -helix F. In this mutant the stability increase due to reduced force of entropy driven unfolding is thus not as strongly counterbalanced by loss of polar interactions as is seen in the Asn238 \rightarrow Pro mutant. Furthermore, a proline at site 265 is likely to stabilise the proximate β -sheet.

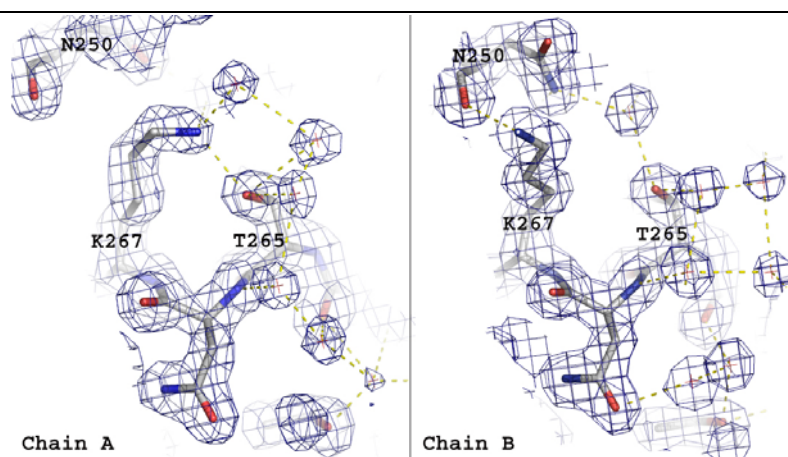


Figure 55. The Thr265 site. Thr265 forms hydrogen bonds with two water molecules. In one of the two molecules in the asymmetric unit Thr265 forms a hydrogen bond with Lys267 but. In the other (chain B) Lys265 forms a hydrogen bond with the carbonyl oxygen of Asn250. Polar interactions are shown as dashed lines and water molecules are marked as a cross.

Concluding remarks

The crystal structure of the cold adapted *Vibrio*-proteinase can be used to examine structural principles of temperature adaptation. First, it is a very important tool for rational design and interpretation of results in mutagenic experiments. Second, structure comparisons with related enzymes of different temperature origin can enhance our understanding of protein properties related to temperature adaptation. In this study, the cold adapted *Vibrio*-proteinase was compared to related enzymes of different temperature origin, with high overall structural resemblance. The sequence homology between the enzymes compared here is relatively low, however. It would be of interest to crystallise and determine the structures of the highly homologous thermophilic enzymes, aqualysin I and the *Thermus* Rt41a proteinase for further comparative studies. Protein preparation of the thermostable enzymes would supposedly be uncomplicated, at least compared to the elaborate purification of the labile *Vibrio*-proteinase. Also, the thermostable enzymes should be at least as likely to give well diffracting crystals as the *Vibrio*-proteinase. In fact, preliminary X-ray crystallographic analysis on aqualysin I was published in 1993 (Green *et al.*, 1993), but was never finished. Green *et al.*, (1993) reported that no molecular replacement solution was found using proteinase K as a search model. This would probably not be a problem now that the much more similar crystal structure of the *Vibrio*-proteinase is available.

The 61 kDa protein component of the [U4/U6·U5] triple snRNP

The human component of the [U4/U6·U5] tri-snRNP, p61, is involved in protein-protein and protein-RNA interactions in its physiological environment. The p61-gene was cloned into pGEXP-1 and expressed with a GST-fusion tag. Considerable achievements in optimising expression conditions and purification procedure to yield satisfactory amounts of soluble p61 were accomplished. However, crystallisation trials did not result in growth of protein crystals. Some evidence of features, highly unfavourable for protein crystallisation, appeared in the purification procedure. First, p61 could not be cleared of a protein band suggested being the chaperone DnaK, indicating that a fraction of p61 might not be properly folded. Second, p61 apparently binds *E. coli* nucleic acids in an unspecific manner, resulting in heterogeneous protein sample. In addition, this nucleic acid binding is apparently crucial for the stability of the recombinant p61. The dependence of p61 on protein and/or nucleic acids binding is not surprising, regarding the protein's functional role in mediating protein-protein and protein-RNA interactions in the spliceosomal assembly cycle. Thus, a reasonable measure to take would be to try and crystallise p61 in complex with a protein or an RNA binding partner. The specificity of p61-RNA binding has not been reported yet but it is known that p61 interacts with the U5 snRNP specific 102 kDa protein (yeast homolog Prp6p). Binding studies on full length and truncated forms of the proteins with the aim of identifying interacting domains could form the basis of subsequent co-crystallisation experiments.

The editosomal protein TbMP42

TbMP42 is one of a set of five interrelated protein components of the *Trypanosoma brucei* editosome, which are considered to mediate protein-protein and/or protein-RNA interactions during the mRNA editing cycle. TbMP42 contains two zinc fingers and is predicted to have an OB-fold on its C-terminal part. TbMP42 has endo- and exonuclease activity, which resides in the C-terminal part. The recombinant TbMP42 produced in *E. coli* forms inclusion bodies, but can be purified via IMAC under denaturing condition and renatured by dialysis to give active enzyme (Michael Brecht, personal communications). Expression at low temperatures, addition of chaperone inducing agents into expression cultures, as well as fusing TbMP42 with GST-tag or MBP yields soluble protein, however, with strong tendency to clump and form particles of undefined size. This was apparent from gel filtration elution profiles of TbMP42, both alone and as a fusion protein. Several indications emerged that the renatured TbMP42 as well as the protein fused with GST-tag or MBP might not be properly folded. Thus, prescission and TEV proteinases failed to cleave off fusion tags and Factor Xa cleaved into the target protein; a problem that can arise from improper folding of the target protein. Expression of a truncated sequence that encodes the endo- and exonuclease C-terminal domain of TbMP42 fused with MBP did not give soluble protein.

TbMP42 is in its physiological state a part of a protein complex and is considered to play a role in mediating protein-protein and/or protein-RNA interactions in the editosome (Stuart *et al.*, 2005). It might be crucial for the integrity of its functional state to interact with other editosomal proteins or RNA. The difficulties encountered in obtaining soluble correctly folded recombinant TbMP42 in *E. coli* could reflect the requirement of such binding partners. In addition, TbMP42 is dependent on zinc binding and there might be insufficient amount of zinc in the cells to allow proper folding of TbMP42 during overexpression. Addition of metal ions to expression cultures is in some cases indispensable for production of functional metal ion dependent proteins. This is for instance the case for the recombinant *Vibrio*-proteinase, dealt with in other parts of this thesis, where addition of calcium to the expression culture is crucial for production of active proteinase. Addition of zinc to a small test expression culture of TbMP42-GST gave good results regarding protein solubility (Figure 40, page 83) but scaling up the expression culture failed because of cell death and –lysis after induction.

Whether the *E. coli* cells could not tolerate the addition of zinc or were killed by the nuclease activity of TbMP42 remains speculative, since this was not examined further.

Initial crystallisation trials of TbMP42 fused with MBP resulted in identification of a few promising crystallisation conditions. In particular, conditions containing hexanediol gave reproducible crystal needles. Attempts to improve the crystals did not result in diffracting crystals. In order to increase the probability of obtaining diffracting crystals, the protein was incubated with double stranded non-hydrolysable DNA- and RNA/DNA hybrid substrates and screened for crystallisation conditions. Surprisingly, salt crystals grew parallel to the presumed protein crystal needles in many hexanediol conditions and in MPD conditions containing both protein and substrate. Noteworthy, the salt crystals did not appear in drops containing either the nucleic acid substrate or protein alone, nor did crystals grow in control drops where buffer alone was mixed with the crystallisation conditions.

In view of the exhaustive effort that has been made to crystallise the recombinant TbMP42 produced in *E. coli*, without success however, a new strategy should be considered. Among alternatives could be to change expression system to a eukaryotic host, such as yeast or baculovirus. Another possibility would be to change the source organism and try to crystallise one of the TbMP42 orthologs from *Trypanosoma cruzi* or *Leishmania major*.

Acknowledgements

“All men are islands. I like to think I’m Ibiza” (Will in Hornby’s *About a boy*), but Donne was right; we knew it all along: “No man is an island, entire of itself; every man is a piece of the continent, a part of the main....”

I am indebted to many people that made this dissertation possible and want to express my warm tanks to:

Prof. Ralf Ficner for giving me the opportunity to work in his group on interesting and challenging projects. I appreciate his goodwill and faith in ones competence, which characterised his supervision.

Oliver Einsle for being the Korreferent, for his help with the computerkram and for lending Magnús his bike. Achim Dickmanns in his dynamic and cheerful manner was always there for a good advice and constructive discussion. Markus Rudolph was indefatigable¹ in explaining the principles of crystallography and helping me with data collection and processing. I am very grateful for his kind support.

Tine Kragh Nielsen for being an Eins-A colleague and for being my friend. From her I have learnt many things, from basic methods in crystallography to much more important aspects of life, such as enjoying Aalborg Jule Akvavit.

Tine, Julia Wittmann and Anja Strasser for correcting the manuscript of this thesis. Them and other members of the groups of Ralf Ficner, Oliver Einsle and Markus Rudolph, in particular my lab fellows Christos and Winfried, I thank for providing a good working atmosphere. The friendly sequencer, Andreas, I thank for quick and high quality sequencing. I thank Ulrich Göringer and his group in Darmstadt for pleasant cooperation.

I very much enjoyed the collaboration with Magnús Már Kristjánsson and I thank him for his important contribution to this work.

Arnór and Teitur deserve big gold medals and trillion kisses for their patience and for being the best. My parents and Helmut I thank for their unconditional support and encouragement.

¹ seemingly incapable of tiring

References

- Abraham, J. M., Feagin, J. E. and Stuart, K. (1988). Characterization of cytochrome c oxidase III transcripts that are edited only in the 3' region. *Cell* **55**, 267-272.
- Aghajari, N., Feller, G., Gerday, C. and Haser, R. (1998). Structures of the psychrophilic *Alteromonas haloplanctis* α -amylase give insights into cold adaptation at a molecular level. *Structure* **6**, 1503-1516.
- Aghajari, N., Van Petegem, F., Villeret, V., Chessa, J. P., Gerday, C., Haser, R. and Van Beeumen, J. (2003). Crystal structures of a psychrophilic metalloprotease reveal new insights into catalysis by cold-adapted proteases. *Proteins* **50**, 636-647.
- Alberts, I. L., Nadassy, K. and Wodak, S. J. (1998). Analysis of zinc binding sites in protein crystal structures. *Protein Sci.* **7**, 1700-1716.
- Alfredsson, G. A., Gudmundsson, H. M., Xiang, J. Y. and Kristjánsson, M. M. (1995). Subtilisin-like serin proteases from psychrophilic marine bacteria. *J. Mar. Biotechnol.* **3**, 71-72.
- Almog, O., Gonzalez, A., Klein, D., Greenblatt, H. M., Braun, S. and Shoham, G. (2003). The 0.93 Å crystal structure of sphericase: a calcium-loaded serine protease from *Bacillus sphaericus*. *J. Mol. Biol.* **332**, 1071-1082.
- Alvarez, M., Zeelens, J. P., Mainfroid, V., Rentier-Delrue, F., Martial, J. A., Wyns, L., Wierenga, R. K. and Maes, D. (1998). Triose-phosphate isomerase (TIM) of the psychrophilic bacterium *Vibrio marinus*. *J. Biol. Chem.* **273**, 2199-2206.
- Aphasizhev, R., Aphasizheva, I. and Simpson, L. (2003). A tale of two TUTases. *Proc. Natl. Acad. Sci. U S A* **100**, 10617-10622.
- Aphasizhev, R., Sbicego, S., Peris, M., Jang, S. H., Aphasizheva, I., Simpson, A. M., Rivlin, A. and Simpson, L. (2002). Trypanosome mitochondrial 3' terminal uridylyl transferase (TUTase): the key enzyme in U-insertion/deletion RNA editing. *Cell* **108**, 637-648.
- Argos, P., Rossmann, M. G., Grau, U. M., Zuber, H., Frank, G. and Tratschin, J. D. (1979). Thermal stability and proteins structure. *Biochemistry* **18**, 5698-5703.
- Arnórsdóttir, J., Kristjánsson, M. M. and Ficner, R. (2005). Crystal structure of a subtilisin-like serine proteinase from a psychrotrophic *Vibrio* species reveals structural aspects of cold adaptation. *Febs J.* **272**, 832-845.
- Arnórsdóttir, J., Smáradóttir, R. B., Magnússon, Ó. T., Eggertsson, G., Thorbjarnardóttir, S. H. and Kristjánsson, M. M. (2002). Characterization of a cloned subtilisin-like serine proteinase from a psychrotrophic *Vibrio* species. *Eur. J. Biochem.* **269**, 5536-5546.
- Bae, E. and Phillips, G. N., Jr. (2004). Structures and analysis of highly homologous psychrophilic, mesophilic, and thermophilic adenylate kinases. *J. Biol. Chem.* **279**, 28202-28208.
- Baltimore, D. (1970). RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* **226**, 1209-1211.
- Beadle, B. M., Baase, W. A., Wilson, D. B., Gilkes, N. R. and Shoichet, B. K. (1999). Comparing the thermodynamic stabilities of a related thermophilic and mesophilic enzyme. *Biochemistry* **38**, 2570-2576.
- Bell, G. S., Russell, R. J., Connaris, H., Hough, D. W., Danson, M. J. and Taylor, G. L. (2002). Stepwise adaptations of citrate synthase to survival at life's extremes. From psychrophile to hyperthermophile. *Eur. J. Biochem.* **269**, 6250-6260.
- Benne, R., Van den Burg, J., Brakenhoff, J. P., Sloof, P., Van Boom, J. H. and Tromp, M. C. (1986). Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell* **46**, 819-826.

References

- Berget, S. M., Moore, C. and Sharp, P. A. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci. U S A* **74**, 3171-3175.
- Betzl, C., Gourinath, S., Kumar, P., Kaur, P., Perbandt, M., Eschenburg, S. and Singh, T. P. (2001). Structure of a serine protease proteinase K from *Tritirachium album* limber at 0.98 Å resolution. *Biochemistry* **40**, 3080-3088.
- Betzl, C., Pal, G. P. and Saenger, W. (1988). Three-dimensional structure of proteinase K at 0.15-nm resolution. *Eur. J. Biochem.* **178**, 155-171.
- Betzl, C., Teplyakov, A. V., Harutyunyan, E. H., Saenger, W. and Wilson, K. S. (1990). Thermitase and proteinase K: a comparison of the refined three-dimensional structures of the native enzymes. *Protein Eng.* **3**, 161-172.
- Blaber, M., Lindstrom, J. D., Gassner, N., Xu, J., Heinz, D. W. and Matthews, B. W. (1993). Energetic cost and structural consequences of burying a hydroxyl group within the core of a protein determined from Ala→Ser and Val→Thr substitutions in T4 lysozyme. *Biochemistry* **32**, 11363-11373.
- Bloch, E., Rachel, R., Burggraf, S., Hafenbradl, D., Jannasch, H. W. and Stetter, K. O. (1997). *Pyrolobus fumarii*, gen. and sp. nov., represents a novel group of archaea, extending the upper temperature limit for life to 113 degrees C. *Extremophiles* **1**, 14-21.
- Blum, B., Bakalara, N. and Simpson, L. (1990a). A model for RNA editing in kinetoplastid mitochondria: "guide" RNA molecules transcribed from maxicircle DNA provide the edited information. *Cell* **60**, 189-198.
- Blum, B., Sturm, N. R., Simpson, A. M. and Simpson, L. (1991). Chimeric gRNA-mRNA molecules with oligo(U) tails covalently linked at sites of RNA editing suggest that U addition occurs by transesterification. *Cell* **65**, 543-550.
- Boehringer, D., Makarov, E. M., Sander, B., Makarova, O. V., Kastner, B., Luhrmann, R. and Stark, H. (2004). Three-dimensional structure of a pre-catalytic human spliceosomal complex B. *Nat. Struct. Mol. Biol.* **11**, 463-468.
- Brünger, A. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472-475.
- Bryan, P., Wang, L., Hoskins, J., Ruvinov, S., Strausberg, S., Alexander, P., Almog, O., Gilliland, G. and Gallagher, T. (1995). Catalysis of a protein folding reaction: mechanistic implications of the 2.0 Å structure of the subtilisin-prodomain complex. *Biochemistry* **34**, 10310-10318.
- Bryan, P. N. (2000). Protein engineering of subtilisin. *Biochim. Biophys. Acta* **1543**, 203-222.
- Bullock, W. O., Fernandez, J. M. and Short, J. M. (1987). XL1-Blue: a high efficiency plasmid transforming *recA* *Escherichia coli* strains with beta-galactosidase selection. *Bio. Techniques.* **5**, 376-379.
- Burgess, M. L., Heidmann, S. and Stuart, K. (1999). Kinetoplastid RNA editing does not require the terminal 3' hydroxyl of guide RNA, but modifications to the guide RNA terminus can inhibit *in vitro* U insertion. *RNA* **5**, 883-892.
- Cech, T. R. (1991). RNA editing: world's smallest introns? *Cell* **64**, 667-669.
- Chow, L. T., Gelinas, R. E., Broker, T. R. and Roberts, R. J. (1977). An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* **12**, 1-8.
- Cojocaru, V., Nottrott, S., Klement, R. and Jovin, T. M. (2005). The snRNP 15.5K protein folds its cognate K-turn RNA: a combined theoretical and biochemical study. *RNA* **11**, 197-209.
- Corell, R. A., Feagin, J. E., Riley, G. R., Strickland, T., Guderian, J. A., Myler, P. J. and Stuart, K. (1993). *Trypanosoma brucei* minicircles encode multiple guide RNAs which can direct editing of extensively overlapping sequences. *Nucl. Acids Res.* **21**, 4313-4320.

References

- Corell, R. A., Read, L. K., Riley, G. R., Nellisery, J. K., Allen, T. E., Kable, M. L., Wachal, M. D., Seiwert, S. D., Myler, P. J. and Stuart, K. D. (1996). Complexes from *Trypanosoma brucei* that exhibit deletion editing and other editing-associated properties. *Mol. Cell Biol.* **16**, 1410-1418.
- Criswell, A. R., Bae, E., Stec, B., Konisky, J. and Phillips, G. N., Jr. (2003). Structures of thermophilic and mesophilic adenylate kinases from the genus *Methanococcus*. *J. Mol. Biol.* **330**, 1087-1099.
- Cruz-Reyes, J., Zhelonkina, A. G., Huang, C. E. and Sollner-Webb, B. (2002). Distinct functions of two RNA ligases in active *Trypanosoma brucei* RNA editing complexes. *Mol. Cell Biol.* **22**, 4652-4660.
- D'Amico, S., Gerday, C. and Feller, G. (2001). Structural determinants of cold adaptation and stability in a large protein. *J. Biol. Chem.* **276**, 25791-25796.
- D'Amico, S., Gerday, C. and Feller, G. (2002). Dual effects of an extra disulfide bond on the activity and stability of a cold-adapted alpha-amylase. *J. Biol. Chem.* **277**, 46110-46115.
- D'Amico, S., Marx, J. C., Gerday, C. and Feller, G. (2003). Activity-stability relationships in extremophilic enzymes. *J. Biol. Chem.* **278**, 7891-7896.
- Das, R., Zhou, Z. and Reed, R. (2000). Functional association of U2 snRNP with the ATP-independent spliceosomal complex E. *Mol. Cell* **5**, 779-787.
- Davail, S., Feller, G., Narinx, E. and Gerday, C. (1994). Cold adaptation of proteins. Purification, characterization, and sequence of the heat-labile subtilisin from the antarctic psychrophile *Bacillus* TA41. *J. Biol. Chem.* **269**, 17448-17453.
- de la Cruz, V. F., Neckelmann, N. and Simpson, L. (1984). Sequences of six genes and several open reading frames in the kinetoplast maxicircle DNA of *Leishmania tarentolae*. *J. Biol. Chem.* **259**, 15136-15147.
- Decker, C. J. and Sollner-Webb, B. (1990). RNA editing involves indiscriminate U changes throughout precisely defined editing domains. *Cell* **61**, 1001-1011.
- Deming, J. W. (2002). Psychrophiles and polar regions. *Curr. Opin. Microbiol.* **5**, 301-309.
- Ebeling, W., Hennrich, N., Klockow, M., Metz, H., Orth, H. D. and Lang, H. (1974). Proteinase K from *Tritirachium album* Limber. *Eur. J. Biochem.* **47**, 91-97.
- Eder, J., Rheinhecker, M. and Fersht, A. R. (1993). Folding of subtilisin BPN': role of the pro-sequence. *J. Mol. Biol.* **233**, 293-304.
- Edgcomb, S. P. and Murphy, K. P. (2000). Structural energetics of protein folding and binding. *Curr. Opin. Biotechnol.* **11**, 62-66.
- Ernst, N. L., Panicucci, B., Igo, R. P., Jr., Panigrahi, A. K., Salavati, R. and Stuart, K. (2003). TbMP57 is a 3' terminal uridylyl transferase (TUTase) of the *Trypanosoma brucei* editosome. *Mol. Cell* **11**, 1525-1536.
- Estevez, A. M. and Simpson, L. (1999). Uridine insertion/deletion RNA editing in trypanosome mitochondria--a review. *Gene* **240**, 247-260.
- Feagin, J. E., Abraham, J. M. and Stuart, K. (1988). Extensive editing of the cytochrome c oxidase III transcript in *Trypanosoma brucei*. *Cell* **53**, 413-422.
- Feller, G. (2003). Molecular adaptations to cold in psychophilic enzymes. *Cell Mol. Life Sci.* **60**, 648-662.
- Feller, G. and Gerday, C. (2003). Psychophilic enzymes: hot topics in cold adaptation. *Nat. Rev. Microbiol.* **1**, 200-208.

References

- Feller, G., Payan, F., Theys, F., Qian, M., Haser, R. and Gerday, C. (1994). Stability and structural analysis of alpha-amylase from the antarctic psychrophile *Alteromonas haloplanctis* A23. *Eur. J. Biochem.* **222**, 441-447.
- Feller, G., Zekhnini, Z., Lamotte-Brasseur, J. and Gerday, C. (1997). Enzymes from cold-adapted microorganisms. The class C beta-lactamase from the antarctic psychrophile *Psychrobacter immobilis* A5. *Eur. J. Biochem.* **244**, 186-191.
- Fields, P. A. (2001). Review: Protein function at thermal extremes: balancing stability and flexibility. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* **129**, 417-431.
- Fujiwara, K. and Tsuru, D. (1976). Affinity chromatography of several proteolytic enzymes on carbobenzoxy-D-phenylalanyl-triethylenetetramine-Sepharose. *Int. J. Peptide Protein Res.* **9**, 18-26.
- Gerday, C., Aittaleb, M., Arpigny, J. L., Baise, E., Chessa, J. P., Garsoux, G., Petrescu, I. and Feller, G. (1997). Psychrophilic enzymes: a thermodynamic challenge. *Biochim. Biophys. Acta* **1342**, 119-131.
- Gianese, G., Bossa, F. and Pascarella, S. (2002). Comparative structural analysis of psychrophilic and meso- and thermophilic enzymes. *Proteins* **47**, 236-249.
- Gottschalk, A., Neubauer, G., Banroques, J., Mann, M., Luhrmann, R. and Fabrizio, P. (1999). Identification by mass spectrometry and functional analysis of novel proteins of the yeast [U4/U6.U5] tri-snRNP. *Embo J.* **18**, 4535-4548.
- Graziano, G., Catanzano, F., Riccio, A. and Barone, G. (1997). A reassessment of the molecular origin of cold denaturation. *J. Biochem. (Tokyo)* **122**, 395-401.
- Green, P. R., Oliver, J. D., Strickland, L. C., Toerner, D. R., Matsuzawa, H. and Ohta, T. (1993). Purification, crystallization and preliminary X-ray investigation of aqualysin I, a heat-stable serine protease. *Acta Crystallog. sect. D* **49**, 349-352.
- Gros, P., Betzel, C., Dauter, Z., Wilson, K. S. and Hol, W. G. (1989). Molecular dynamics refinement of a thermitase-eglin-c complex at 1.98 Å resolution and comparison of two crystal forms that differ in calcium content. *J. Mol. Biol.* **210**, 347-367.
- Gros, P., Kalk, K. H. and Hol, W. G. J. (1991). Calcium binding in thermitase. Crystallographic studies of thermitase at 0, 5 and 100 mM calcium. *J. Biol. Chem.* **266**, 2953-2961.
- Haki, G. D. and Rakshit, S. K. (2003). Developments in industrially important thermostable enzymes: a review. *Bioresour. Technol.* **89**, 17-34.
- Haney, P. J., Badger, J. H., Buldak, G. L., Reich, C. I., Woese, C. R. and Olsen, G. J. (1999). Thermal adaptation analyzed by comparison of protein sequences from mesophilic and extremely thermophilic *Methanococcus* species. *Proc. Natl. Acad. Sci. U S A* **96**, 3578-3583.
- Hastings, M. L. and Krainer, A. R. (2001). Pre-mRNA splicing in the new millennium. *Curr. Opin. Cell Biol.* **13**, 302-309.
- Hensgens, L. A., Brakenhoff, J., De Vries, B. F., Sloof, P., Tromp, M. C., Van Boom, J. H. and Benne, R. (1984). The sequence of the gene for cytochrome c oxidase subunit I, a frameshift containing gene for cytochrome c oxidase subunit II and seven unassigned reading frames in *Trypanosoma brucei* mitochondrial maxi-circle DNA. *Nucl. Acids Res.* **12**, 7327-7344.
- Hochachka, P. W. and Somero, G. N. (1984). *Biochemical Adaptation*. Princeton, NJ, Princeton University Press.
- Honig, B. and Nicholls, A. (1995). Classical electrostatics in biology and chemistry. *Science* **268**, 1144-1149.
- Huang, C. E., Cruz-Reyes, J., Zhelonkina, A. G., O'Hearn, S., Wirtz, E. and Sollner-Webb, B. (2001). Roles for ligases in the RNA editing complex of *Trypanosoma brucei*: band IV is needed for U-deletion and RNA repair. *Embo J.* **20**, 4694-4703.

References

- Jaenicke, R. (2000). Do ultrastable proteins from hyperthermophiles have high or low conformational rigidity? *Proc. Natl. Acad. Sci. U S A* **97**, 2962-2964.
- Jaenicke, R. and Böhm, G. (1998). The stability of proteins in extreme environments. *Curr. Opin. Struct. Biol.* **8**, 738-748.
- Jany, K. D., Lederer, G. and Mayer, B. (1986). Amino acid sequence of proteinase K from the mold *Tritirachium album* Limber. *FEBS Lett.* **199**, 139-144.
- Jurica, M. S. and Moore, M. J. (2003). Pre-mRNA splicing: awash in a sea of proteins. *Mol. Cell* **12**, 5-14.
- Kable, M. L., Seiwert, S. D., Heidmann, S. and Stuart, K. (1996). RNA editing: a mechanism for gRNA-specified uridylyate insertion into precursor mRNA. *Science* **273**, 1189-1195.
- Kano, H., Taguchi, S. and Momose, H. (1997). Cold adaptation of a mesophilic serine protease, subtilisin, by *in vitro* random mutagenesis. *Appl. Microbiol. Biotechnol.* **47**, 46-51.
- Kaufner, N. F. and Potashkin, J. (2000). Analysis of the splicing machinery in fission yeast: a comparison with budding yeast and mammals. *Nucl. Acids Res.* **28**, 3003-3010.
- Kim, S. Y., Hwang, K. Y., Kim, S. H., Sung, H. C., Han, Y. S. and Cho, Y. (1999). Structural basis for cold adaptation. Sequence, biochemical properties, and crystal structure of malate dehydrogenase from a psychrophile *Aquaspirillum arcticum*. *J. Biol. Chem.* **274**, 11761-11767.
- Kleywegt, G. J. (1996). Use of non-crystallographic symmetry in protein structure refinement. *Acta Crystallog. sect. D* **52**, 842-857.
- Kristjánsson, M. M. and Ásgeirsson, B. (2002). Properties of extremophilic enzymes and their importance in Food Science and technology. In *Handbook of Food Enzymology*. J. R. Whitaker, A. G. J. Voragen and D. W. S. Wong. New York, Marcel Dekker, Inc.: 77-100.
- Kristjánsson, M. M. and Magnússon, Ó. T. (2001). Effect of lyotropic salts on the stability of a subtilisin-like proteinase from a psychrotrophic *Vibrio*-species, proteinase K and aqualysin I. *Protein and Peptide Lett.* **8**, 249-255.
- Kristjánsson, M. M., Magnússon, Ó. T., Gudmundsson, H. M., Alfredsson, G. A. and Matsuzawa, H. (1999). Properties of a subtilisin-like proteinase from a psychrotrophic *Vibrio* species. Comparison with proteinase K and aqualysin I. *Eur. J. Biochem.* **260**, 752-760.
- Kumar, S. and Nussinov, R. (2004). Different roles of electrostatics in heat and in cold: adaptation by citrate synthase. *Chembiochem* **5**, 280-290.
- Kumar, S., Tsai, C.-J. and Nussinov, R. (2000). Factors enhancing protein thermostability. *Protein Eng.* **13**, 179-191.
- Laity, J. H., Lee, B. M. and Wright, P. E. (2001). Zinc finger proteins: new insights into structural and functional diversity. *Curr. Opin. Struct. Biol.* **11**, 39-46.
- Lamzin, V. A. and Wilson, K. S. (1993). Automated refinement of protein molecules. *Acta Crystallog. sect. D* **49**, 129-147.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. and Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 293-291.
- Lee, Y. C., Miyata, Y., Terada, I., Ohta, T. and Matsuzawa, H. (1991). Involvement of NH₂-terminal pro-sequence in the production of active aqualysin I (a thermophilic serine protease) in *Escherichia coli*. *Agric. Biol. Chem.* **55**, 3027-3032.
- Leiros, H.-K. S., Willassen, N. P. and Smalås, A. O. (2000). Structural comparison of psychrophilic and mesophilic trypsins. Elucidating the molecular basis of cold-adaptation. *Eur. J. Biochem.* **267**, 1039-1049.

References

- Leiros, I., Moe, E., Lanes, O., Smalas, A. O. and Willassen, N. P. (2003). The structure of uracil-DNA glycosylase from Atlantic cod (*Gadus morhua*) reveals cold-adaptation features. *Acta Crystallog. sect. D* **59**, 1357-1365.
- Leung, S. S. and Koslowsky, D. J. (2001). RNA editing in *Trypanosoma brucei*: characterization of gRNA U-tail interactions with partially edited mRNA substrates. *Nucl. Acids Res.* **29**, 703-709.
- Lu, D., Searles, M. A. and Klug, A. (2003). Crystal structure of a zinc-finger-RNA complex reveals two modes of molecular recognition. *Nature* **426**, 96-100.
- Maddock, J., Roy, J. and Woolford, J., Jr (1996). Six novel genes necessary for pre-mRNA splicing in *Saccharomyces cerevisiae*. *Nucl. Acids Res.* **24**, 1037-1044.
- Madison-Antenucci, S., Grams, J. and Hajduk, S. L. (2002). Editing machines: the complexities of trypanosome RNA editing. *Cell* **108**, 435-438.
- Makarov, E. M., Makarova, O. V., Urlaub, H., Gentzel, M., Will, C. L., Wilm, M. and Luhrmann, R. (2002). Small Nuclear Ribonucleoprotein Remodeling During Catalytic Activation of the Spliceosome. *Science* **298**, 2205-2208.
- Makarova, O. V., Makarov, E. M., Liu, S., Vornlocher, H. P. and Luhrmann, R. (2002). Protein 61K, encoded by a gene (PRPF31) linked to autosomal dominant retinitis pigmentosa, is required for U4/U6*U5 tri-snRNP formation and pre-mRNA splicing. *Embo J.* **21**, 1148-1157.
- Makhatadze, G. I. and Privalov, P. L. (1993). Contribution of hydration to protein folding thermodynamics. I. The enthalpy of hydration. *J. Mol. Biol.* **232**, 639-659.
- Maroney, P. A., Romfo, C. M. and Nilsen, T. W. (2000). Functional recognition of 5' splice site by U4/U6.U5 tri-snRNP defines a novel ATP-dependent step in early spliceosome assembly. *Mol. Cell* **6**, 317-328.
- Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491-497.
- McDonald, I. K. and Thornton, J. M. (1994). Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**, 777-793.
- McRee, D. E. (1999). XtalView/Xfit--A versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156-165.
- Missel, A., Souza, A. E., Norskau, G. and Goringer, H. U. (1997). Disruption of a gene encoding a novel mitochondrial DEAD-box protein in *Trypanosoma brucei* affects edited mRNAs. *Mol. Cell Biol.* **17**, 4895-4903.
- Miyazaki, K., Wintrode, P. L., Grayling, R. A., Rubingh, D. N. and Arnold, F. H. (2000). Directed evolution study of temperature adaptation in a psychrophilic enzyme. *J. Mol. Biol.* **297**, 1015-1026.
- Mougin, A., Gottschalk, A., Fabrizio, P., Luhrmann, R. and Branlant, C. (2002). Direct probing of RNA structure and RNA-protein interactions in purified HeLa cell's and yeast spliceosomal U4/U6.U5 tri-snRNP particles. *J. Mol. Biol.* **317**, 631-649.
- Muller, A., Hinrichs, W., Wolf, W. M. and Saenger, W. (1994). Crystal structure of calcium-free proteinase K at 1.5-Å resolution. *J. Biol. Chem.* **269**, 23108-23111.
- Murzin, A. G. (1993). OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. *Embo J.* **12**, 861-867.
- Narinx, E., Baise, E. and Gerday, C. (1997). Subtilisin from psychrophilic antarctic bacteria: characterization and site-directed mutagenesis of residues possibly involved in the adaptation to cold. *Protein Eng.* **10**, 1271-1279.
- Newman, A. (1998). RNA splicing. *Curr. Biol.* **8**, R903-905.

References

- Nilsen, T. W. (2002). The spliceosome: no assembly required? *Mol. Cell* **9**, 8-9.
- Nottrott, S., Hartmuth, K., Fabrizio, P., Urlaub, H., Vidovic, I., Ficner, R. and Luhrmann, R. (1999). Functional interaction of a novel 15.5kD [U4/U6.U5] tri-snRNP protein with the 5' stem-loop of U4 snRNA. *Embo J.* **18**, 6119-6133.
- Nottrott, S., Urlaub, H. and Luhrmann, R. (2002). Hierarchical, clustered protein interactions with U4/U6 snRNA: a biochemical role for U4/U6 proteins. *Embo J.* **21**, 5527-5538.
- Otwinowski, Z. and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In *Methods in Enzymology*. R. M. Sweet. New York, Academic Press. **276**: 307-326.
- Paehler, A., Banerjee, A., Dattagupta, J. K., Fujiwara, T., Lindner, K., Pal, G. P., Suck, D., Weber, G. and Saenger, W. (1984). Three-dimensional structure of fungal proteinase K reveals similarity to bacterial subtilisin. *Embo J.* **3**, 1311-1314.
- Panigrahi, A. K., Gygi, S. P., Ernst, N. L., Igo, R. P., Jr., Palazzo, S. S., Schnauffer, A., Weston, D. S., Carmean, N., Salavati, R., Aebersold, R. and Stuart, K. D. (2001). Association of two novel proteins, TbMP52 and TbMP48, with the *Trypanosoma brucei* RNA editing complex. *Mol. Cell Biol.* **21**, 380-389.
- Panigrahi, A. K., Schnauffer, A., Carmean, N., Igo, R. P., Jr., Gygi, S. P., Ernst, N. L., Palazzo, S. S., Weston, D. S., Aebersold, R., Salavati, R. and Stuart, K. D. (2001). Four related proteins of the *Trypanosoma brucei* RNA editing complex. *Mol. Cell Biol.* **21**, 6833-6840.
- Panigrahi, A. K., Schnauffer, A., Ernst, N. L., Wang, B., Carmean, N., Salavati, R. and Stuart, K. (2003). Identification of novel components of *Trypanosoma brucei* editosomes. *RNA* **9**, 484-492.
- Pasternak, A., Ringe, D. and Hedstrom, L. (1999). Comparison of anionic and cationic trypsinogens: the anionic activation domain is more flexible in solution and differs in its mode of BPTI binding in the crystal structure. *Protein Sci.* **8**, 253-258.
- Peek, K., Daniel, R. M., Monk, C., Parker, L. and Coolbear, T. (1992). Purification and characterization of a thermostable proteinase isolated from *Thermus* sp. strain Rt41A. *Eur. J. Biochem.* **207**, 1035-1044.
- Perona, J. J. and Craik, C. S. (1995). Structural basis of substrate specificity in the serine proteases. *Protein Sci.* **4**, 337-360.
- Pollard, V. W., Harris, M. E. and Hajduk, S. L. (1992). Native mRNA editing complexes from *Trypanosoma brucei* mitochondria. *Embo J.* **11**, 4429-4438.
- Privalov, P. L. and Makhatadze, G. I. (1993). Contribution of hydration to protein folding thermodynamics. II. The entropy and Gibbs energy of hydration. *J. Mol. Biol.* **232**, 660-679.
- Querol, E., Perez-Pons, J. A. and Mozo-Villarias, A. (1996). Analysis of protein conformational characteristics related to thermostability. *Protein Eng.* **9**, 265-271.
- Radzicka, A. and Wolfenden, R. (1988). Comparing the polarities of the amino acids: side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol, and neutral aqueous solution. *Biochemistry* **27**, 1664-1670.
- Robinson, G. W. and Cho, C. H. (1999). Role of hydration water in protein unfolding. *Biophys. J.* **77**, 3311-3318.
- Rodriguez, R., China, G., Lopez, N., Pons, T. and G., V. (1998). Homology modeling, model and software evaluation: three related resources. *CABIOS* **14**, 523-528.
- Russel, R. J. M., Gerike, U., Danson, M. J., Hough, D. W. and Taylor, G. L. (1998). Structural adaptations of the cold active citrate-synthetase from an Antarctic bacterium. *Structure* **6**, 351-361.
- Schaffert, N., Hossbach, M., Heintzmann, R., Achsel, T. and Luhrmann, R. (2004). RNAi knockdown of hPrp31 leads to an accumulation of U4/U6 di-snRNPs in Cajal bodies. *Embo J.* **23**, 3000-3009.

References

- Schechter, I. and Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **27**, 157-162.
- Schnauffer, A., Ernst, N. L., Palazzo, S. S., O'Rear, J., Salavati, R. and Stuart, K. (2003). Separate insertion and deletion subcomplexes of the *Trypanosoma brucei* RNA editing complex. *Mol. Cell* **12**, 307-319.
- Schnauffer, A., Panigrahi, A. K., Panicucci, B., Igo, R. P., Jr., Wirtz, E., Salavati, R. and Stuart, K. (2001). An RNA ligase essential for RNA editing and survival of the bloodstream form of *Trypanosoma brucei*. *Science* **291**, 2159-2162.
- Seiwert, S. D. and Stuart, K. (1994). RNA editing: transfer of genetic information from gRNA to precursor mRNA *in vitro*. *Science* **266**, 114-117.
- Shapiro, T. A. and Englund, P. T. (1995). The structure and replication of kinetoplast DNA. *Annu. Rev. Microbiol.* **49**, 117-143.
- Sheridan, P. P., Panasik, N., Coombs, J. M. and Brenchley, J. E. (2000). Approaches for deciphering the structural basis of low temperature enzyme activity. *Biochim. Biophys. Acta* **1543**, 417-433.
- Siezen, R. J., de Vos, W. M., Leunissen, J. A. and Dijkstra, B. W. (1991). Homology modelling and protein engineering strategy of subtilases, the family of subtilisin-like serine proteinases. *Protein Eng.* **4**, 719-737.
- Siezen, R. J. and Leunissen, J. A. (1997). Subtilases: the superfamily of subtilisin-like serine proteases. *Protein Sci.* **6**, 501-523.
- Simpson, L. (2001). RNA editing. In *Frontiers of Science*, Academic Press. **1: The Origins of Life**: 510-516.
- Simpson, L., Neckelmann, N., de la Cruz, V. F., Simpson, A. M., Feagin, J. E., Jasmer, D. P. and Stuart, J. E. (1987). Comparison of the maxicircle (mitochondrial) genomes of *Leishmania tarentolae* and *Trypanosoma brucei* at the level of nucleotide sequence. *J. Biol. Chem.* **262**, 6182-6196.
- Simpson, L., Sbicego, S. and Aphasizhev, R. (2003). Uridine insertion/deletion RNA editing in trypanosome mitochondria: a complex business. *RNA* **9**, 265-276.
- Smalås, A. O., Heimstad, E. S., Hordvik, A., Willassen, N. P. and Male, R. (1994). Cold adaptation of enzymes: structural comparison between salmon and bovine trypsins. *Proteins* **20**, 149-166.
- Smalås, A. O., Leiros, H. K., Os, V. and Willassen, N. P. (2000). Cold adapted enzymes. *Biotechnol. Annu. Rev.* **6**, 1-57.
- Staley, J. P. and Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* **92**, 315-326.
- Stevens, S. W., Ryan, D. E., Ge, H. Y., Moore, R. E., Young, M. K., Lee, T. D. and Abelson, J. (2002). Composition and functional characterization of the yeast spliceosomal penta-snRNP. *Mol. Cell* **9**, 31-44.
- Stuart, K., Allen, T. E., Heidmann, S. and Seiwert, S. D. (1997). RNA editing in kinetoplastid protozoa. *Microbiol. Mol. Biol. Rev.* **61**, 105-120.
- Stuart, K. and Panigrahi, A. K. (2002). RNA editing: complexity and complications. *Mol. Microbiol.* **45**, 591-596.
- Stuart, K. D., Schnauffer, A., Ernst, N. L. and Panigrahi, A. K. (2005). Complex management: RNA editing in trypanosomes. *Trends Biochem. Sci.* **30**, 97-105.
- Studier, F. W., Rosenberg, A. H., Dunn, J. J. and Dubendorf, J. W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymology* **185**, 60-89.

References

- Sturm, N. R. and Simpson, L. (1990). Kinetoplast DNA minicircles encode guide RNAs for editing of cytochrome oxidase subunit III mRNA. *Cell* **61**, 879-884.
- Sturm, N. R. and Simpson, L. (1990). Partially edited mRNAs for cytochrome b and subunit III of cytochrome oxidase from *Leishmania tarentolae* mitochondria: RNA editing intermediates. *Cell* **61**, 871-878.
- Suck, D. (1997). Common fold, common function, common origin? *Nat. Struct. Biol.* **4**, 161-165.
- Svingor, A., Kardos, J., Hajdu, I., Nemeth, A. and Zavodszky, P. (2001). A "better" enzyme to cope with cold: Comparative flexibility studies on psychrotrophic, mesophilic and thermophilic IPMDHs. *J. Biol. Chem.* **276**, 28121-28125.
- Szilagyi, A. and Zavodszky, P. (2000). Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. *Structure* **8**, 493-504.
- Taguchi, S., Ozaki, A. and Momose, H. (1998). Engineering of a cold-adapted protease by sequential random mutagenesis and a screening system. *Appl. Environ. Microbiol.* **64**, 492-495.
- Takeuchi, Y., Noguchi, S., Satow, Y., Kojima, S., Kumagai, I., Miura, K., Nakamura, K. T. and Mitsui, Y. (1991). Molecular recognition at the active site of subtilisin BPN': crystallographic studies using genetically engineered proteinaceous inhibitor SSI (*Streptomyces* subtilisin inhibitor). *Protein Eng.* **4**, 501-508.
- Tepljakov, A. V., Kuranova, I. P., Harutyunyan, E. H., Vainstein, B. K., Frömmel, C., Höhne, W. E. and Wilson, K. S. (1990). Crystal structure of thermitase at 1.4 Å. *J. Mol. Biol.* **214**, 261-279.
- Terada, I., Kwon, S. T., Miyata, Y., Matsuzawa, H. and Ohta, T. (1990). Unique precursor structure of an extracellular protease, aqualysin I, with NH₂- and COOH-terminal pro-sequences and its processing in *Escherichia coli*. *J. Biol. Chem.* **265**, 6576-6581.
- Tindbaek, N., Svendsen, A., Oestergaard, P. R. and Draborg, H. (2004). Engineering a substrate-specific cold-adapted subtilisin. *Protein Eng. Des. Sel.* **17**, 149-156.
- Tsai, C. J., Maizel, J. V., Jr. and Nussinov, R. (2002). The hydrophobic effect: a new insight from cold denaturation and a two-state water structure. *Crit. Rev. Biochem. Mol. Biol.* **37**, 55-69.
- Uversky, V. N. (2002). What does it mean to be natively unfolded? *Eur. J. Biochem.* **269**, 2-12.
- Vagin, A. and Tepljakov, A. (1997). MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* **30**, 1022-1025.
- van den Burg, B. (2003). Extremophiles as a source for novel enzymes. *Curr. Opin. Microbiol.* **6**, 213-218.
- Van Petegem, F., Collins, T., Meuwis, M. A., Gerday, C., Feller, G. and Van Beeumen, J. (2003). The structure of a cold-adapted family 8 xylanase at 1.3 Å resolution. Structural adaptations to cold and investigation of the active site. *J. Biol. Chem.* **278**, 7531-7539.
- Vidovic, I., Nottrott, S., Hartmuth, K., Luhrmann, R. and Ficner, R. (2000). Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. *Mol. Cell* **6**, 1331-1342.
- Vieille, C., Burdette, D. S. and Zeikus, J. G. (1996). Thermozyms. *Biotechnol. Annu. Rev.* **2**, 1-83.
- Vieille, C. and Zeikus, G. J. (2001). Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol. Mol. Biol. Rev.* **65**, 1-43.
- Vithana, E. N., Abu-Safieh, L., Allen, M. J., Carey, A., Papaioannou, M., Chakarova, C., Al-Magthteh, M., Ebenezer, N. D., Willis, C., Moore, A. T., Bird, A. C., Hunt, D. M. and Bhattacharya, S. S. (2001). A human homolog of yeast pre-mRNA splicing gene, PRP31, underlies autosomal dominant retinitis pigmentosa on chromosome 19q13.4 (RP11). *Mol. Cell* **8**, 375-381.

References

- Vogt, G., Woell, S. and Argos, P. (1997). Protein thermal stability, hydrogen bonds, and ion pairs. *J. Mol. Biol.* **269**, 631-643.
- Watkins, N. J., Segault, V., Charpentier, B., Nottrott, S., Fabrizio, P., Bachi, A., Wilm, M., Rosbash, M., Branlant, C. and Lührmann, R. (2000). A common core RNP structure shared between the small nuclear box C/D RNPs and the spliceosomal U4 snRNP. *Cell* **103**, 457-466.
- Weidenhammer, E. M., Ruiz-Noriega, M. and Woolford, J. L., Jr. (1997). Prp31p promotes the association of the U4/U6 x U5 tri-snRNP with prespliceosomes to form spliceosomes in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* **17**, 3580-3588.
- Will, C. L. and Lührmann, R. (1997). Protein functions in pre-mRNA splicing. *Curr. Opin. Cell Biol.* **9**, 320-328.
- Will, C. L. and Lührmann, R. (2001). Spliceosomal UsnRNP biogenesis, structure and function. *Curr. Opin. Cell Biol.* **13**, 290-301.
- Wilms, B., Hauck, A., Reuss, M., Sylđatk, C., Mattes, R., Siemann, M. and Altenbuchner, J. (2001). High-cell-density fermentation for production of L-N-carbamoylase using an expression system based on the *Escherichia coli* rhaBAD promoter. *Biotechnol. Bioeng.* **73**, 95-103.
- Winn, M. D., Isupov, M. N. and Murshudov, G. N. (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallog. sect. D* **57**, 122-133.
- Wintrade, P. L., Miyazaki, K. and Arnold, F. H. (2000). Cold adaptation of a mesophilic subtilisin-like protease by laboratory evolution. *J. Biol. Chem.* **275**, 31635-31640.
- Yabuta, Y., Subbian, E., Oiry, C. and Shinde, U. (2003). Folding pathway mediated by an intramolecular chaperone. A functional peptide chaperone designed using sequence databases. *J. Biol. Chem.* **278**, 15246-15251.

Appendices

A. p61 sequences

BLAST 2 SEQUENCES RESULTS VERSION BLASTN 2.2.10 [Oct-19-2004]

```

Sequence 1 gi_18249846          Length 2153 (1 .. 2153)
(Makarova et al., 2002)
Sequence 2 Seq from Bernhard Lagerbauer Length 1500 (1 .. 1500)

Query: 358 atgtctctggcagatgagctcttagctgatctcgaagaggcagcagaagaggaggaagga 417
          |||
Sbjct: 1 atgtctctggcagatgagctcttagctgatctcgaagaggcagcagaagaggaggaagga 60

Query: 418 ggaagctatggggaggaagaagaggagccagcgatcgaggatgtgcaggaggagacacag 477
          |||
Sbjct: 61 ggaagctatggggaggaagaagaggagccagcgatcgaggatgtgcaggaggagacacag 120

Query: 478 ctggatctttccgggattcagtcgaagaccatcgccaagctatgggatagtaagatgttt 537
          |||
Sbjct: 121 ctggatctttccgggattcagtcgaagaccatcgccaagctatgggatagtaagatgttt 180

Query: 538 gctgagattatgatgaagattgaggagtatatcagcaagcaagccttcagaagtg 597
          |||
Sbjct: 181 gctgagattatgatgaagattgaggagtatatcagcaagcaagccttcagaagtg 240

Query: 598 atgggaccagtgaggccgcctgaataccgcgtcatcgtggatgccaacaacctgacc 657
          |||
Sbjct: 241 atgggaccagtgaggccgcctgaataccgcgtcatcgtggatgccaacaacctgacc 300

Query: 658 gtggagatcgaaaacgagctgaacatcatccataagtccatccgggataagtactcaaag 717
          |||
Sbjct: 301 gtggagatcgaaaacgagctgaacatcatccataagtccatccgggataagtactcaaag 360

Query: 718 agattccctgaactggagtccttggtccccaatgcactggattacatccgcacggtcaag 777
          |||
Sbjct: 361 agattccctgaactggagtccttggtccccaatgcactggattacatccgcacggtcaag 420

Query: 778 gagctgggcaacagcctggacaagtgaagaacaatgagaacctgcagcagatcctcacc 837
          |||
Sbjct: 421 gagctgggcaacagcctggacaagtgaagaacaatgagaacctgcagcagatcctcacc 480

Query: 838 aatgccaccatcatggtcgtcagcgtcaccgcctccaccaccaggggagcagctgtcg 897
          |||
Sbjct: 481 aatgccaccatcatggtcgtcagcgtcaccgcctccaccaccaggggagcagctgtcg 540

Query: 898 gaggaggagctggagcggctggaDgaggcctgcgacatggcgctggagctgaacgcctcc 957
          |||
Sbjct: 541 gaggaggagctggagcggctggaEgaggcctgcgacatggcgctggagctgaacgcctcc 600

Query: 958 aagcaccgatctacgagtatgtggagtcccggatgtccttcaIggcaccacaacctgtcc 1017
          |||
Sbjct: 601 aagcaccgatctacgagtatgtggagtcccggatgtccttcaTggcaccacaacctgtcc 660

Query: 1018 atcattatcggggcatccacggccccaagatcatgggtgtggccggcggcctgaccaac 1077
          |||
Sbjct: 661 atcattatcggggcatccacggccccaagatcatgggtgtggccggcggcctgaccaac 720
    
```


BLAST 2 SEQUENCES RESULTS VERSION BLASTP 2.2.10 [Oct-19-2004]

Sequence 1 gi_18249847 (Makarova et al., 2002) U4/U6 snRNP-associated 61 kDa protein [Homo sapiens]

Sequence 2 Seq from Bernhard Laggerbauer

```

Query: 1  MSLADELLADLEEAEEEEEGGSYEEEEPAIEDVQEETQLDLSGDSVKTIAKLWDSKMF 60
          MSLADELLADLEEAEEEEEGGSYEEEEPAIEDVQEETQLDLSGDSVKTIAKLWDSKMF
Sbjct: 1  MSLADELLADLEEAEEEEEGGSYEEEEPAIEDVQEETQLDLSGDSVKTIAKLWDSKMF 60

Query: 61 AEIMMKIEEYISKQAKASEVMGPFVEAAPEYRVIVDANNLTVEIENELNIIHKFIRDKYSK 120
          AEIMMKIEEYISKQAKASEVMGPFVEAAPEYRVIVDANNLTVEIENELNIIHKFIRDKYSK
Sbjct: 61 AEIMMKIEEYISKQAKASEVMGPFVEAAPEYRVIVDANNLTVEIENELNIIHKFIRDKYSK 120

Query: 121 RFPELESVPNALDYIRTVKELGNSLDKCKNNENLQQILT NATIMVSVTASTTQGQQLS 180
          RFPELESVPNALDYIRTVKELGNSLDKCKNNENLQQILT NATIMVSVTASTTQGQQLS
Sbjct: 121 RFPELESVPNALDYIRTVKELGNSLDKCKNNENLQQILT NATIMVSVTASTTQGQQLS 180

Query: 181 EEELERLEACDMALELNASKHRIYEVESRMSFAPNLSIIIGASTAAKIMGVAGGLTN 240
          EEELERLEACDMALELNASKHRIYEVESRMSFAPNLSIIIGASTAAKIMGVAGGLTN
Sbjct: 181 EEELERLEACDMALELNASKHRIYEVESRMSFAPNLSIIIGASTAAKIMGVAGGLTN 240

Query: 241 LSKMPACNIMLLGAQRKTLSGFSSTSVLPHTGYIYHSDIVQSLPPDLRRKAARLVAAKCT 300
          LSKMPACNIMLLGAQRKTLSGFSSTSVLPHTGYIYHSDIVQSLPPDLRRKAARLVAAKCT
Sbjct: 241 LSKMPACNIMLLGAQRKTLSGFSSTSVLPHTGYIYHSDIVQSLPPDLRRKAARLVAAKCT 300

Query: 301 LAARVDSFHSTEGKVGVELKDEIERKFDKWQEPVPVQVKPLPAPLDGQRKKRGGRRYR 360
          LAARVDSFHSTEGKVGVELKDEIERKFDKWQEPVPVQVKPLPAPLDGQRKKRGGRRYR
Sbjct: 301 LAARVDSFHSTEGKVGVELKDEIERKFDKWQEPVPVQVKPLPAPLDGQRKKRGGRRYR 360

Query: 361 KMKERLGLTEIRKQANRMSFGEEDDAYQEDLGFSLGHLGKSGSGRVQRQTQVNEATKARI 420
          KMKERLGLTEIRKQANRMSFGEEDDAYQEDLGFSLGHLGKSGSGRVQRQTQVNEATKARI
Sbjct: 361 KMKERLGLTEIRKQANRMSFGEEDDAYQEDLGFSLGHLGKSGSGRVQRQTQVNEATKARI 420

Query: 421 SKTLQRTLQKQSVVYGGKSTIRDRSSGTASSVAFTPLQGLEIVNPQAAEKKVAEANQKYF 480
          SKTLQRTLQKQSVVYGGKSTIRDRSSGTASSVAFTPLQGLEIVNPQAAEKKVAEANQKYF
Sbjct: 421 SKTLQRTLQKQSVVYGGKSTIRDRSSGTASSVAFTPLQGLEIVNPQAAEKKVAEANQKYF 480

Query: 481 SSMAEFLKVKGEKSGLMST 499
          SSMAEFLKVKGEKSGLMST
Sbjct: 481 SSMAEFLKVKGEKSGLMST 499
    
```

B. TbMP42 sequences

BLAST 2 SEQUENCES RESULTS VERSION BLASTN 2.2.10 [Oct-19-2004]

Sequence 1 [gi_14495339](#) (Panigrahi et al., 2001) Length 1182 (1 .. 1182)

Sequence 2 Sequence from Ulrich Göringer Length 1179 (1 .. 1179)
Identities = 1175/1179 (99%)

```

Query: 1   atgaagcgtgttacttcacatatatttcgcggtttttaccacttgtgttgcgcagcgtggc 60
          |||
Sbjct: 1   atgaagcgtgttacttcacatatatttcgcggtttttaccacttgtgttgcgcagcgtggc 60

Query: 61   ttgtcgacttacgcatcacctcatgtctcgtccatcggttactatggagctacaaaatgt 120
          |||
Sbjct: 61   ttgtcgacttacgcatcacctcatgtctcgtccatcggttactatggagctacaaaatgt 120
          |||
          M
          I

Query: 121  cttttggcttccacaccagacactccctccttcagtgcgggcagtgcggttaaggctttt 180
          |||
Sbjct: 121  cttttggcttccacaccagacactccctccttcagtgcgggcagtgcggttaaggctttt 180
          |||

Query: 181  cgtcttataaatgccttaaatacatcacattatgaccaagcacgcaggaaggcaaggct 240
          |||
Sbjct: 181  cgtcttataaatgccttaaatacatcacattatgaccaagcacgcaggaaggcaaggct 240
          |||

Query: 241  atgatgaacaagggggtaagttggaggaagtaaacctgaagaataaaaaacaaaccg 300
          |||
Sbjct: 241  atgatgaacaagggggtaagttggaggaagtaaacctgaagaataaaaaacaaaccg 300
          |||

Query: 301  caaggggcaatgtcgaaccaacatcacccaccgtcatcttcgacgtccggtacggag 360
          |||
Sbjct: 301  caaggggcaatgtcgaaccaacatcacccaccgtcatcttcgacgtccggtacggag 360
          |||

Query: 361  gccgcatcaacctcaccacgcattcaagctttcccggaataccgttttcaccgtgggt 420
          |||
Sbjct: 361  gccgcatcaacctcaccacgcattcaagctttcccggaataccgttttcaccgtgggt 420
          |||

Query: 421  ggcatagggctcgttggaaactccagtaggtgctgcctctcatgtcggttaatactactacc 480
          |||
Sbjct: 421  ggcatagggctcgttggaaactccagtaggtgctgcctctcatgtcggttaatactactacc 480
          |||

Query: 481  actgctgctaattctgcttctggaaatgccctttccgacgagaacgcgacaaaaaaca 540
          |||
Sbjct: 481  actgctgctaattctgcttctggaaatgccctttccgacgagaacgcgacaaaaaaca 540
          |||

Query: 541  tttgtctgtacgatttgtcagaaaacatttcggttgaagccgcccttcaacaccattac 600
          |||
Sbjct: 541  tttgtctgtacgatttgtcagaaaacatttcggttgaagccgcccttcaacaccattac 600
          |||

Query: 601  caggcgaagcacaacatggagatgccgacatcatcttcaagttctggcggcgccagtgt 660
          |||
Sbjct: 601  caggcgaagcacaacatggagatgccgacatcatcttcaagttctggcggcgccagtgt 660
          |||

Query: 661  caaccagtgtgcaggggtggcgccactacagcaggtgttggaaagtgtgggtttttcccat 720
          |||
Sbjct: 661  caaccagtgtgcaggggtggcgccactacagcaggtgttggaaagtgtgggtttttcccat 720
          |||

```


Appendices – B. TbMP42

Query: 721 actgaggaggaaactggcagatcagccatgggaacacaatagttcacagtcaagaaaca 780
|||||
Sbjct: 721 actgaggaggaaactggcagatcagccatgggaacacagtagttcacagtcaagaaaca 780

Query: 781 atactcccgcaagccccgcaatatcatcttgatgtcgcgccgaacgcacctgaggagggt 840
|||||
Sbjct: 781 atactcccgcaagccccgcaatatcatcttgatgtcgcgccgaacgcacctgaggagggt 840

Query: 841 gaagtggctgcacattggcgttgtgtcaatcattgtgttatggtggcggtggtgcaaac 900
|||||
Sbjct: 841 gaagtggctgcacattggcgttgtgtcaatcattgtgttatggtggcggtggtgcaaac 900

Query: 901 attcaggaagggtttgtattgaagacaaagttctccagtttacactcatcacagacttc 960
|||||
Sbjct: 901 attcaggaagggtttgtattgaagacaaagttctccagtttacactcatcacagacttc 960

Query: 961 gagggaccctctcctggggacccggataaggactttcatacagttcgtgtgtttgatagc 1020
|||||
Sbjct: 961 gagggaccctctcctggggacccggataaggactttcatacagttcgtgtgtttgatagc 1020

Query: 1021 gattatagttcgagggtgaaagagcaactccgtgacggtagtggttcctcgtcactggc 1080
|||||
Sbjct: 1021 gattatagttcgagggtgaaagagcaactccgtgacggtagtggttcctcgtcactggc 1080

Query: 1081 agactgcgcatggtgcctcaatatgatggttccatgcgaaagtattaccactatcctgtc 1140
|||||
Sbjct: 1081 agactgcgcatggtgcctcaatatgatggttccatgcgaaagtattaccactatcctgtc 1140

Query: 1141 atacaagtacaccgggctgtgggtcagtggtgaggtg 1179
|||||
Sbjct: 1141 atacaagtacaccgggctgtgggtcagtggtgaggtg 1179

Appendices – B. TbMP42

BLAST 2 SEQUENCES RESULTS VERSION BLASTP 2.2.10 [Oct-19-2004]

Sequence 1 gi_14495340 (Panigrahi et al., 2001) RNA-editing complex protein MP42
[Trypanosoma brucei]

Sequence 2 Sequence from Ulrich Göringer

Identities = 391/393 (99%), Positives = 393/393 (99%)

```
Query: 1 MKRVTSHISRFLPLVLSQRGLSTYASPHVSSRYYGATKCLLASTPDTPSFQCGECGKAF 60
          MKRVTSHISRFLPLVLSQRGLSTYASPHVSSRYYGATKCLLASTPDTPSFQCGECGKAF
Sbjct: 1 MKRVTSHISRFLPLVLSQRGLSTYASPHVSSRYYGATKCLLASTPDTPSFQCGECGKAF 60

Query: 61 RLINALNHHIMTKHAGKAKAMMNKGGKLEEVNPEEIKNKPQGAMSQTPSPPPSSSTSGTE 120
          RLINALNHHIMTKHAGKAKAMMNKGGKLEEVNPEEIKNKPQGAMSQTPSPPPSSSTSGTE
Sbjct: 61 RLINALNHHIMTKHAGKAKAMMNKGGKLEEVNPEEIKNKPQGAMSQTPSPPPSSSTSGTE 120

Query: 121 AASTSPTHSSFPGIPFSPVGGIGLVGTPVGAASHVVNTTTTAANSASGNALSDENADKKT 180
          AASTSPTHSSFPGIPFSPVGGIGLVGTPVGAASHVVNTTTTAANSASGNALSDENADKKT
Sbjct: 121 AASTSPTHSSFPGIPFSPVGGIGLVGTPVGAASHVVNTTTTAANSASGNALSDENADKKT 180

Query: 181 FVCTICQKTFRLEAALQHYYQAKHNMEMPTSSSSSGGASAPVLQGGATTAGVGSVGFSSH 240
          FVCTICQKTFRLEAALQHYYQAKHNMEMPTSSSSSGGASAPVLQGGATTAGVGSVGFSSH
Sbjct: 181 FVCTICQKTFRLEAALQHYYQAKHNMEMPTSSSSSGGASAPVLQGGATTAGVGSVGFSSH 240

Query: 241 TEEETGRSAMGTQYVHSQETILPQAPQYHLDVAPNAPEEGEVAAHWRCVNHCVMLGVVQN 300
          TEEETGRSAMGTQYVHSQETILPQAPQYHLDVAPNAPEEGEVAAHWRCVNHCVMLGVVQN
Sbjct: 241 TEEETGRSAMGTQYVHSQETILPQAPQYHLDVAPNAPEEGEVAAHWRCVNHCVMLGVVQN 300

Query: 301 IQEGFVFEDKVLQFTLITDFEGPSPGDPDKDFHTVTRVFDSDYSSRVKEQLRDGEWFLVTG 360
          IQEGFVFEDKVLQFTLITDFEGPSPGDPDKDFHTVTRVFDSDYSSRVKEQLRDGEWFLVTG
Sbjct: 301 IQEGFVFEDKVLQFTLITDFEGPSPGDPDKDFHTVTRVFDSDYSSRVKEQLRDGEWFLVTG 360

Query: 361 RLRMVPQYDGS MRKYYHYPIQVHPGCGSVLIV 393
          RLRMVPQYDGS MRKYYHYPIQVHPGCGSVLIV
Sbjct: 361 RLRMVPQYDGS MRKYYHYPIQVHPGCGSVLIV 393
```

Appendices – B. TbMP42

BLASTP 2.2.10 [Oct-19-2004]

Query= gi|14495340|gb|AAK64280.1|AF382335_1 RNA-editing complex protein MP42 [Trypanosoma brucei] (393 letters), Zinc fingers, OB-fold

>gi|38073431|gb|AAR10837.1| MP42 [Trypanosoma cruzi]

Length = 395

Score = 479 bits (1232), Expect = e-134

Identities = 239/401 (59%), Positives = 298/401 (74%), Gaps = 14/401 (3%)

```
Query: 1 MKRVTSHISRFLPLVLSQRGLSTYASPHVSSMRYYGATKCLLASTPDTPS FQCGECGKAF 60
M+R+ SH+S+ L RG YA P S R+Y T L A +TP++QCGECGK F
Sbjct: 1 MRRIASHLSQRTYRALFPRGGGFYAPPFNMSTRFYRETSLLQAPKTETPAYQCGECGKTF 60

Query: 61 RLINALNHHIMTKHAGKAKAMMNKGGKLEEVNPEEIKNKPPQGAMSQTPSPPPSSSTSGTE 120
RL+NALNHHIMTKHAG AKAM+ K GKLEEV PEE KNK A + T +++T T
Sbjct: 61 RLLNALNHHIMTKHAGHAKAMVLKAGKLEEVKPEEAKNKSSHA-TDATAASTTATGATM 119

Query: 121 AASTSPHSSFPPIPFSPVGGIGLVGTPVGAASHVVNTT-----TTAANSASG--NALS 172
+A+ PT FPG+ +SP+GG G+P+G AS + +T T A +ASG +++
Sbjct: 120 SATQIPT--GFGMTYSPLGGTAAFGSPLGVASMLPTSTAAAGAPTNAAGKNASGTESSVA 177

Query: 173 DENADKKT FVCTICQKTFRLEAALQHHYQAKHNMEMPTSSSSSGGASAQPVLQGGATTAG 232
DE+ADKK FVCTICQKTFRLEAALQHHYQAKHNMEMPT+S+S AS+ +T G
Sbjct: 178 DEDADKKL FVCTICQKTFRLEAALQHHYQAKHNMEMPTNSTS---ASSTRAASQSSTNLG 234

Query: 233 VGSVGFSHTEETGRSAMGTQYVHSQETILPQAPQYHLDVAPNAPEEGEVAAH WRCVNH 292
+ F ++ +S G YVHSQE++LPQAPQYHLDVAPNAPEEG+VAHWRCVN+
Sbjct: 235 AANAIFGGATDDAVKSTTGVHYVHSQESVLPQAPQYHLDVAPNAPEEGDVAHWRCVNYW 294

Query: 293 VMLGVVQNIQEGVFVFEKVLQFTLITDFEGPSPGDPDKDFHTVVRVFDSDYSSRVKEQLRD 352
+LG VQNIQEG+VFE+KV+QFTLITDFEGPSPGDPDKDFHTVR+FD+ + ++K++L++
Sbjct: 295 RVLGPVQNIQEGYVFEEKVVQFTLITDFEGPSPGDPDKDFHTVRIFDNSFCEQMKKELKE 354

Query: 353 GEWFLVTGRLRMVPQYDGS MRKYYHY PVIQVHPGCG SVLKV 393
G+ FLVTGRLRM+PQYD +M+KYYHY PVIQVHPGCGSV+KV
Sbjct: 355 GDRFLVTGRLRMIPQYDSAMKYYHY PVIQVHPGCGSVV 395
```

>gi|38073399|gb|AAR10821.1| MP42 [Leishmania major]

Length = 448

Score = 355 bits (910), Expect = 2e-96

Identities = 195/392 (49%), Positives = 244/392 (62%), Gaps = 45/392 (11%)

```
Query: 46 PDTPS FQCGECGKAFRLINALNHHIMTKHAGKAKAMMNKGGKLEEVNPEEIKNKPPQGAMS 105
P +QCGECGK FRL+NALNHHIMT+H AKA+M K GKL V +++K S
Sbjct: 58 PKDAQYQCGECGKTFRLVNALNHHIMTRHGNNAKALMKKDGKLVPEVDQLKGAHCVSS 117

Query: 106 QPTSPPPSSSTS---GTEAASTSP----THSSFP-----GIPFSPVGGIGLVGTPVGA 151
+S +SS + AAS+SP T S FP G + +GG+ TP G
Sbjct: 118 SSSSTATASSATPAGAAAASSSPLSGGTTSPFPVHFAAPFGAAAASMGGVASSSTPPGT 177

Query: 152 ASHVVNTTTTAANSASGNALS DENA-----DKKT FVCTICQKTFRLEAALQHHYQ 201
A V+ TAA A+ A +DEN +K+ FVCT+CQKTFRLEAALQHHYQ
Sbjct: 178 APGVLQAPATAAAGAT-TAGTDENGSGADGAAEEREKRMFVCTVCQKTFRLEAALQHHYQ 236

Query: 202 AKHNMEMPTSSSS-----SGGASA----QPVLQGGATTA---GVGSVGFSHTE 242
AKHNM+MPTS+SS +GGAS +P GG + GV V
Sbjct: 237 AKHNMDMPTSASSPSARGASTTAGAGGASVSTPGEPGANGGAGAFGGVPGVAADGAA 296

Query: 243 EETGRSAMGT-QYVHSQETILPQAPQYHLDVAPNAPEEGEVAAH WRCVNHCVMLGVVQNI 301
S+ G QYV QE LP APQYHLDVAPNAPEEG++AAHWRCVN CV++G VQ +
Sbjct: 297 APANASSFGAAQYVRQEGALPDAPQYHLDVAPNAPEEGDIAAHWRCVNMVLMGDVQEV 356

Query: 302 QEGVFVFEKVLQFTLITDFEGPSPGDPDKDFHTVVRVFDSDYSSRVKEQLRDGEWFLVTGR 361
+EG+VFED VLQFT+ T+F P+PGDPD DFHTVVR+ ++ + +K ++ G FLVTGR
Sbjct: 357 BEGYVFEDHVLQFTVATEFATPAPGDPDMDFHTVVRVYGHEFWAPLQADMQS GGRFLVTGR 416

Query: 362 LRMVPQYDGS MRKYYHY PVIQVHPGCG SVLKV 393
L MVPQ+D ++KYYHY PVIQV PG G+V++V
Sbjct: 417 LCMVPQFDTQLKYYHY PVIQVFPGTGNVVRV 448
```

Abbreviations

Å	Ångström
amp	Ampicillin
ATP	Adenosine TriPhosphate
B	Temperature Factor
bp	base pairs
BSA	Bovine Serum Albumin
C	Celsius
Cp	Heat capacity
Da	Dalton
DESY	Deutsche Elektronen-SYnchrotron
DNA	Deoxyribose Nucleic Acid
dNTP	Deoxy-Nucleoside-Triphosphat, N= A, T, G, C
dYT	double-strength Yeast extract-Tryptone medium
EDTA	Ethylene Diamine Tetra-Acetic acid
G	Gibbs free energy
gRNA	guide RNA
GSH	Reduced Glutathione
GST	Glutathion-S-Transferase
H	Enthalpy
HEPES	N-(2-HydroxyEthyl)-Piperazine-N'-(2-EthaneSulfonic acid)
IMAC	Immobilized Metal Affinity Chromatography
IPTG	IsoPropyl β -D-ThioGalactoside
iRNA	Interference RNA
kan	Kanamycin
kcat	rate constant
kDNA	kinetoplast DNA
Km	Michaelis Konstant
LB	Luria Bertani medium
M	Molarity
MBP	Maltose Binding Protein
MOPS	3-(N-MOrpholino)-PropaneSulfonic acid
mRNA	messenger RNA
MW	Molecular Weight
N	native
OD	Optical Density

Abbreviations

Pa	Pascal
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PDB	The RCSB Protein Data Bank
PMSF	Phenyl Methane Sulfonyl Fluoride
Prp	Pre RNA Processing
PVDF	Poly(Vinylidene Fluoride)
RCSB	Research Collaboratory for Structural Bioinformatics
R-factor	Residual factor
Rms	Root mean square
RNA	RiboNucleic Acid
rRNA	ribosomal RNA
S	Entropy, Svedberg
s	seconds
SDS PAGE	SodiumDodecylSulfate PolyAcrylamide Gel Electrophoresis
snoRNA	small nucleolar RNA
snRNA	small nuclear RNA
snRNP	small nuclear RiboNucleoProtein particles
sp.	species
T	Temperature
TAE	Tris – Acetate – EDTA
Taq	<i>Thermus aquaticus</i>
TBE	Tris – Borate – EDTA
TbMP	<i>Trypanosoma brucei</i> Mitochondrial Protein
TBREX	<i>Trypanosoma brucei</i> RNA Editing Ligases
TEMED	N,N,N',N'-TEtraMethylene EthyleneDiamine
TEV	Tobacco Etch Virus
Tm	melting Temperature
tRNA	transfer RNA
U	Unit, unfolded
v/v %	volume/volume percent
Vm	Matthews coefficient
VPR	<i>Vibrio</i> -proteinase
w/v %	weight/volume percent
xg	times gravity

Curriculum Vitae

Jóhanna Arnórsdóttir, born on January 10th 1969 in Reykjavík, Iceland

Education and professional experience

- 1985-1989 Abitur at the Gymnasium in Reykjavík
- 1989-1992 Bachelor of Science in Biology at the University of Iceland
- 1992-1993 Teachers Credentials for Secondary Level Education at the University of Iceland
- 1993-1994 Research assistant at the Faculty of Science at the University of Iceland
- 1994-1999 Teacher of Chemistry and Biology at the coordinated secondary school, Fjölbautaskólinn í Ármúla in Reykjavík
- 1999-2001 Master of Science Study in Biology at the University of Iceland under the supervision of Prof. Dr. Guðmundur Eggertsson, Sigríður Þorbjarnardóttir and Dr. Magnús Már Kristjánsson. Structural principles of cold adaptation of a subtilisin-like serine proteinase from a psychrotrophic *Vibrio*-species.
- Since 2001 PhD thesis in the Department of Molecular Structural Biology at the Georg-August Universität Göttingen under the supervision of Prof. Dr. Ralf Ficner.

Publications

Arnórsdóttir, J., Smáradóttir, R. B., Magnússon, Ó. T., Eggertsson, G., Thorbjarnardóttir, S. H. and Kristjánsson, M. M. (2002). Characterization of a cloned subtilisin-like serine proteinase from a psychrotrophic *Vibrio* species. *Eur. J. Biochem.* **269**, 5536-5546.

Arnórsdóttir, J., Kristjánsson, M. M. and Ficner, R. (2005). Crystal structure of a subtilisin-like serine proteinase from a psychrotrophic *Vibrio* species reveals structural aspects of cold adaptation. *Febs J.* **272**, 832-845.