UNIVERSITÀ DEGLI STUDI DI MILANO - BICOCCA
Facoltà di Scienze Matematiche, Fisiche e Naturali
SCUOLA DI DOTTORATO DI SCIENZE
Dottorato di Ricerca in Biotecnologie Industriali - XXII Ciclo

A dynamic perspective on cold-adapted enzymes at the molecular level.

Tutor: Prof. Luca DE GIOIA
Cotutor: Dr. Elena PAPALEO
Cotutor: Prof. Richard LAVERY
Coordinator: Prof. Marco VANONI

Ph.D. Dissertation
Marco PASI

Anno Accademico 2008-2009

Contents

List of Figures

List of Tables

Riassunto

Alcuni organismi, soprattutto unicellulari, si sono adattati a condizioni estreme di temperatura, e sono in grado di sopravvivere e proliferare in tali ambienti grazie anche all'ottimizzazione del loro repertorio enzimatico. Enzimi adattati al freddo (psicrofili) sono invariabilmente caratterizzati da notevole attività catalitica alle basse temperature, necessaria per far fronte alla riduzione della velocità delle reazioni chimiche in quelle condizioni, e da scarsa stabilità alle alte temperature. Numerosi studi atti ad elucidare i meccanismi di tale adattamento a livello molecolare concordano nel correlare l'elevata costante di velocità osservata in enzimi psicrofili ad un abbassamento dell'entalpia di attivazione della reazione, ottenuto dal punto di vista strutturale tramite una diminuzione del numero di interazioni stabilizzate entalpicamente che devono essere scisse per raggiungere lo stato di transizione. Dato che tali interazioni sono anche coinvolte nella stabilizzazione della struttura proteica, la loro diminuzione ha l'effetto di causare un aumento della flessibilità strutturale. D'altro canto la scarsa termostabilità degli enzimi psicrofili riflette un edificio proteico spiccatamente flessibile. La flessibilità riveste dunque un ruolo fondamentale nell'adattamento al freddo degli enzimi, e il suo studio tramite simulazioni di Dinamica Molecolare offre una descrizione dettagliata del fenomeno tenendo rigorosamente conto del suo aspetto dinamico.

L'α-amilasi cloruro-dipendente isolata dal batterio psicrofilo *Pseudoalteromonas haloplanktis* è ad oggi l'enzima psicrofilo meglio caratterizzato, sia dal punto di vista strutturale che da quello chimicofisico. Il confronto con un suo omologo mesofilo isolato dal pancreas di *Sus scrofa* ha permesso di ipotizzare quali fossero le strategie strutturali di adattamento di questa α-amilasi, e in particolare di individuare alcune mutazioni puntiformi, atte ad aggiungere interazioni intramolecolari deboli presenti nell'enzima mesofilo e lontane dal sito attivo, in grado di modificare k_{cat}, K_m e termostabilità di tali mutanti. Per approfondire le differenze nelle proprietà dinamiche in questi enzimi e chiarirne i determinanti strutturali, simulazioni di Dinamica Molecolare sono state effettuate sulle forme *wild-type* (wt) e su 7 forme mutanti dell'α-amilasi

psicrofila (4 mutanti singoli, un doppio, un quintuplo e un ettuplo mutante). Le simulazioni sono state analizzate con particolare attenzione alle caratteristiche strutturali (struttura secondaria, accessibilità al solvente, interazioni intramolecolari, con cofattori e con il solvente) e al loro comportamento dinamico (persistenza delle strutture secondarie, moti collettivi e correlati e flessibilità molecolare in diverse scale temporali). I risultati mostrano come la gran parte delle regioni ad elevata flessibilità siano localizzate nei dintorni della tasca di legame per il substrato, coinvolgendo in particolare i $\beta - \alpha$-loop che si estendono dalla struttura a barrel del dominio principale. Come già precedentemente riportato, questi loop risultano sovradimensionati nell'enzima mesofilo (PPA) rispetto al suo omologo psicrofilo (AHA), e dunque estesi a maggiori distanze dal sito attivo; le nostre analisi dimostrano che le dinamiche più rilevanti in PPA coinvolgono principalmente tali loop, con il risultato che la regione nelle immediate vicinanze del sito attivo è più flessibile in AHA. Inoltre, precedenti studi cristallografici sulle forme *free* e *bound* degli enzimi suggeriscono l'importanza di residui chiave nel legame e nel processamento di piccoli polisaccaridi; le dettagliate analisi di tali elementi cataliticamente attivi qui proposte hanno rivelato moti concertati e probabilmente rilevanti dal punto di vista funzionale. Questo suggerisce l'esistenza di un network di movimenti intrinseci nei due enzimi necessario per la loro attività. Il confronto con le simulazioni effettuate sulle forme mutanti di AHA mostrano come alcune delle interazioni ripristinate siano in grado di modificare in maniera globale il carattere dinamico di AHA, e in particolare delle regioni ad elevata flessibilità, nella direzione dell'enzima mesofilo, rivelando che le mutazioni hanno effetti distali e particolarmente rilevanti nella zona del sito attivo. Le differenze riscontrate sono lievi, e si apprezzano in un contesto di elevata conservazione delle dinamiche essenziali, a riprova dell'estrema robustezza dei moti globali negli enzimi.

Studi comparativi dell'adattamento enzimatico alle basse temperature hanno dimostrato l'esistenza di caratteri strutturali generalmente associati ad enzimi psicrofili, tra i quali la preferenza per amminoacidi piccoli e la diminuzione del numero di interazioni intramolecolari deboli. Ciononostante, la formulazione di una teoria unificata dell'adattamento al freddo sembra impossibile, in quanto le strategie adottate dagli enzimi per incrementare la loro efficienza catalitica alle basse temperature variano tra famiglie enzimatiche. Sulla base di queste

osservazioni, risulta dunque auspicabile minimizzare la deriva genetica tra gli enzimi posti a confronto in qualunque studio comparativo di questo tipo, in maniera da massimizzare la possibilità di identificare differenze effettivamente legate all'adattamento termico. Per questo motivo le due isoforme di serin-proteasi espresse dal Salmone Atlantico, delle quali una presenta caratteristiche di attività e stabilità tipiche di un enzima psicrofilo, risultano essere un buon modello di studio in questo senso. Le proprietà dinamiche dei due enzimi sono state confrontate sulla base di simulazioni di Dinamica Molecolare, insieme ad una seconda tripsina psicrofila isolata da *Oncorhynchus keta*. Metodi di valutazione della convergenza del campionamento e della significatività statistica della flessibilità sono stati applicati per confermare la validità delle conclusioni tratte sulla base delle simulazioni. I risultati del confronto indicano che all'interno della famiglia delle serin-proteasi *chymotrypsin-like*, enzimi adattati alle basse temperature presentano elevata rigidità strutturale in regioni lontane dal sito catalitico e maggiore flessibilità in regioni vicine al centro reattivo, in accordo con precedenti studi. Inoltre, le differenze in flessibilità riscontrate tra tripsine adattate a temperature diverse sono localizzate in regioni che corrispondono alle regioni dove sono concentrate le differenze in flessibilità tra l'elastasi psicrofila di Salmone Atlantico e una sua omologa mesofila di bovino. Uno studio filogenetico delle due famiglie di serin-proteasi ha permesso di dimostrare che la separazione tra le due famiglie è precedente alla separazione tra le forme adattate alle basse temperature, ponendo le basi per ipotizzare che elastasi e tripsine psicrofile abbiano indipendentemente adottato strategie molecolari simili per ottimizzare la loro flessibilità.

L'evidenza di convergenza evolutiva nella famiglia delle serina proteasi pone l'accento sull'importanza di estendere studi comparativi di questo tipo a piu' famiglie enzimatiche, al fine di ottenere un modello di adattamento al freddo sufficientemente dettagliato da permettere la progettazione razionale di enzimi con differente adattamento termico. Studi ad ampio spettro che possano prendere in considerazione un numero sufficiente di enzimi da poter individuare dei tratti di validità generale sono resi difficili dall'ingente richiesta di risorse computazionali necessarie a trattare ciascun enzima in maniera significativa e in dettaglio atomico. Modelli a più bassa risoluzione, riducendo in maniera sostanziale il numero di gradi di libertà, consentono di estendere

lo studio delle proprietà dinamiche di biomacromolecole a scale temporali e spaziali difficilmente accessibili in dettaglio atomico, aprendo la possibilità a confronti su un gran numero di enzimi. Vari approcci (collettivamente chiamati "coarse-grained") sono stati proposti negli ultimi anni per studiare in maniera accurata ed efficace le proprietà dinamiche delle proteine, accomunati dalla fondamentale necessità di "integrare" un gran numero di gradi di libertà in un numero minore. L'entità di tale semplificazione è proporzionale alla capacità del modello di descrivere in tempi ragionevoli fenomeni su più larga scala, ma è inversamente proporzionale alla sua risoluzione. Con lo scopo di determinare il livello di semplificazione più adatto allo studio dell'adattamento degli enzimi alle basse temperature, diversi approcci "coarse-grained" sono stati utilizzati per confrontare a diversi livelli di risoluzione le proprietà dinamiche di tripsine omologhe adattate a temperature differenti. Lo studio delle relazioni tra i risultati di tali confronti e i risultati delle simulazioni di Dinamica Molecolare in dettaglio atomico ha permesso di chiarire i limiti di un paricolare approccio "coarse-grained" in questo contesto e di porre le basi per una sua eventuale estensione.

Pubblicazioni

This thesis has lead to the following publications:

Chapter 2 E. Papaleo, M. Pasi, L. Riccardi, I. Sambi, P. Fantucci and L. De Gioia Protein flexibility in psychrophilic and mesophilic trypsins. Evidence of evolutionary conservation of protein dynamics in trypsin-like serine-proteases. *FEBS Letters*, 582:1008–1018, 2008

Chapter 4 M. Pasi, L. Riccardi, P. Fantucci, L. De Gioia and E. Papaleo Dynamic Properties of a Psychrophilic α-amylase in Comparison with a Mesophilic Homologue. *J Phys Chem B*, 113(41):13585–13595, 2009

Chapter 1

General Introduction

1.1 Cold Adapted Enzymes

1.1.1 Cold adapted organisms

If the vast extent of permanently cold habitats are considered, most parts of the planet are exposed to low temperatures, often below 0 °C [1]. In 1887, Forster reported that microorganisms isolated from a fish could grow well at 0 °C, focusing the attention on the growth and reproduction of bacteria at low temperatures. This was the first claim and afterwards several organisms, prokaryotic but also eukaryotic, have been found to have successfully colonised low-temperature habitats [2].

Evolution has allowed these cold-adapted organisms, called psychrophiles (in Greek the word *psychro* means cold, so psychrophilic equals to cold loving), not merely to survive, but also to successfully proliferate at low temperatures, often displaying metabolic fluxes at low temperatures that are more or less comparable to those exhibited by closely related mesophiles ("intermediate"-loving organisms) living at moderate temperatures [3–6]. This ability depends on a series of features, genotypic and/or phenotypic, involved in overcoming the negative effects of low temperatures [7].

In fact, temperature is one of the most important environmental factors for life: a reduction in temperature slows down most physiological processes, changes the way proteins interact, reduces membrane fluidity, hinders growth rates of organisms, provokes increased viscosity in water, induces a reduction in salt solubility and an increase in gas solubility. Decreases in the pH of biological buffers are also

noticed at low temperatures, affecting both protein solubility and the charge of amino acids, histidine residues in particular. Another consequence of the exposure to low temperature is a strong inhibition of chemical reaction rates: these can be reduced 30- to 80-fold when the medium temperature is decreased from 37 °C to 0 °C [8]. In order to overcome the detrimental effect of low temperatures, psychrophiles have developed several adaptive strategies from the level of individual types of molecules to that of the whole organism. For instance, cold-adaptation has led to the regulation of membrane fluidity, the synthesis of specialised molecules known as cold-shock proteins, cryo-protectors and antifreeze molecules, the regulation of ion channels permeability, micro-tubules polymerisation, and the modification of enzyme kinetics [7].

1.1.2 Biocatalysis at low temperatures

In permanently cold habitats, low temperatures have constrained psychrophiles to develop enzymatic tools to sustain metabolic rates compatible to life and comparable to those of temperate organisms at their respective physiological temperatures. The temperature dependence of chemical reactions is commonly described by the Arrhenius equation:

$$k = Ae^{-E_a/RT} \tag{1.1}$$

in which k is the rate constant, A is the pre-exponential factor (related to steric factors and molecular collision frequency), E_a is the activation energy, R is the gas constant (8.314 kJ mol^{-1} K^{-1}) and T is the absolute temperature. Thus, any decrease in temperature will induce an exponential decrease in the reaction rate.

Therefore, according to (1.1), at very low temperatures (0-4 °C), insufficient kinetic energy is available for the system to overcome reaction barriers. Psychrophilic organisms have evolved several strategies to compensate for the very slow metabolic rates that would occur at low temperatures as a result of this kinetic effect. These include an energetically expensive strategy of increasing enzyme concentration [9], seasonal expression of iso-enzymes in fish and nematodes [10–12], and the evolution of enzymes in which reaction rates tend to become temperature independent and approach diffusion control [7]. The majority of cold-adapted enzymes are characterised by a shift in apparent T_{opt}

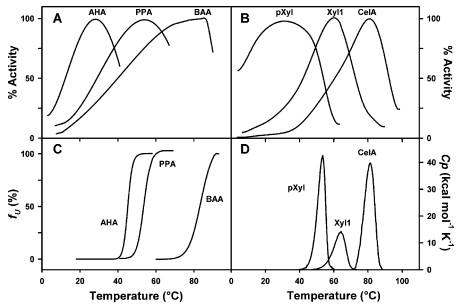


Figure 1.1: Temperature dependence of residual activity (%) (A, B) and conformational stability as recorded by fluorescence emission (f_U , C) and DSC (C_p , D) of psychrophilic, mesophilic and thermophilic enzymes. AHA, psychrophilic α-amylase; PPA, mesophilic α-amylase; BAA, thermostable α-amylase; pXyl, psychrophilic xylanase; Xyl1, mesophilic xylanase; CelA, thermophilic endoglucanase. Note that the maximal activity of the psychrophilic enzymes AHA and pXyl is reached at temperatures lower than those of any significant conformational event. Images are taken from reference [7].

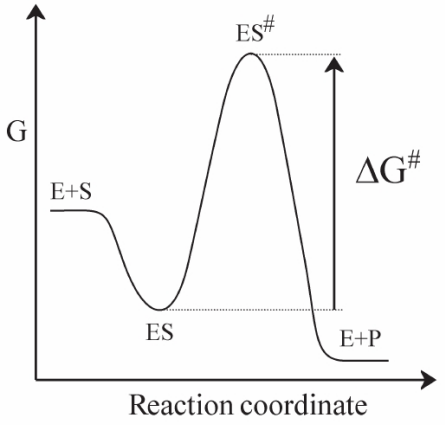


Figure 1.2: Reaction coordinate diagram in transition state theory, showing the Gibbs free energy changes during a catalysed reaction. E: enzyme; S: substrate; P: product; $\Delta G^{\ddagger}$: activation free energy. Image taken from reference [8].

(optimum temperature of activity) to a low temperature with a concomitant decrease in stability (Fig.1.1).

Moreover, cold-adapted enzymes tend to exhibit a high reaction rate (up to 10-fold higher k_{cat} values at low temperature compared to heat-stable homologues). According to the transition-state theory (*TST*, Fig.1.2), k_{cat} is related to temperature and thermodynamic activation parameters [7, 8, 13, 14] through the following equation, equivalent to

the Arrhenius equation:

$$k_{cat} = \kappa\left(\frac{k_B T}{h}\right) e^{-\frac{\Delta G^{\ddagger}}{RT}}, \tag{1.2}$$

where κ is the transmission coefficient generally close to unity, k_B is the Boltzmann constant ($1.38 \cdot 10^{-23}$ J K^{-1}) and h is the Planck constant ($6.63 \cdot 10^{-34}$ Js). Increased catalytic efficiency can be achieved decreasing the activation free energy ($\Delta G^{\ddagger}$) barrier between the ground state and the transition state ($TS^{\ddagger}$) [7]. $\Delta G^{\ddagger}$ is the result of two components:

$$\Delta G^{\ddagger} = \Delta H^{\ddagger} - T\Delta S^{\ddagger}, \tag{1.3}$$

where $\Delta H^{\ddagger}$ is the activation change in enthalpy, $\Delta S^{\ddagger}$ is the activation change in entropy and T is the absolute temperature. In order to consider the effects of $\Delta S^{\ddagger}$ and $\Delta H^{\ddagger}$ on k_{cat}, (1.2) and (1.3) can be combined to obtain:

$$k_{cat} = \kappa\left(\frac{k_B T}{h}\right) e^{-\{(\Delta H^{\ddagger}/RT)+(\Delta S^{\ddagger}/R)\}}, \tag{1.4}$$

From (1.4) it is clear that either $\Delta S^{\ddagger}$ has to increase, or $\Delta H^{\ddagger}$ has to decrease, or both, in order to increase k_{cat} values. Almost all cold-adapted enzymes studied to date present a lower $\Delta H^{\ddagger}$ when compared to the mesophilic counterparts. As a result, the reaction rate tends to be less temperature-dependent, and the high reaction rate (k_{cat}) is maintained at low temperatures [8, 13]. This decrease in $\Delta H^{\ddagger}$ is structurally accomplished through the reduction in the number and strength of enthalpy-driven interactions that need to be broken during transition-state formation. This implies a more flexible structure of the active-site in cold-adapted enzymes, and is the first insight suggesting that activity and stability are linked in the process of thermal adaptation. As a consequence of active-site flexibility, the ground-state enzyme-substrate complex (ES) resides in a wider distribution of conformational states than the activated enzyme-substrate complex, and this is bound to have effects on the reaction's activation entropy. Depending on the reaction, $\Delta S^{\ddagger}$ can be negative or positive as this term also includes contributions from the redistribution of water molecules. However, the key point is that the difference in the activation entropy between a mesophilic and a psychrophilic enzyme is always negative [8, 14, 15]. The gain in k_{cat} would be massive if the decrease in $\Delta H^{\ddagger}$ was not accompanied by a de-

crease in $\Delta S^{\ddagger}$ [8]. However, in practice, very large increases in k_{cat} are not observed in enzymes from psychrophiles because of an enthalpy-entropy compensation [16–18], which implies that a decrease in $\Delta H^{\ddagger}$ accompanied by a decrease in $\Delta S^{\ddagger}$ produces an overall small change in $\Delta G^{\ddagger}$.

1.1.3 Kinetic parameters at low temperatures: the contribution of k_{cat} and K_m to cold adaptation

The activity of enzymes from psychrophiles measured at their environmental temperature (0-4 °C) is generally lower than that for homologous enzymes from mesophiles at their environmental temperature (around 37 °C). This has been shown for AHA [7, 14, 19] and a xylanase [20]. Although psychrophilic enzymes are indeed active at their environmental temperature, it appears that higher activities could be achievable, and that adaptation is therefore to be considered incomplete [7, 14, 21]. Actually, the k_{cat} of a cold-adapted enzyme [22] could be enhanced by simultaneously decreasing $\Delta H^{\ddagger}$ and increasing $\Delta S^{\ddagger}$. This may be achieved by an indirect increase in the entropy of the system via water dislocation [23]. For example, in glucanases, a chain of well-ordered water molecules is located across the active site. Some water molecules are displaced when the substrate or the activated substrate binds to the active site [24]. If more water molecules are released upon biding of the transition-state substrate to the enzyme than upon binding of the ground-state substrate, then there will be considerable entropic benefits for the formation of a transition-state enzyme-substrate complex [23, 25–29]. Decrease in $\Delta S^{\ddagger}$ may be kept as low as possible in cold-adapted enzymes by manipulating the active site in order to enable the substrate to displace a grater number of water molecules in its activated form. It has also been suggested that the same goal could be achieved by adding a small amount of rigidity to part of the active site, while maintaining constant values of $\Delta H^{\ddagger}$ [8].

The strength of the *ES* interaction may decrease (electrostatic interactions) or increase (hydrophobic interactions) with increasing temperature. The former are exothermic, whereas the latter are formed endothermically, within the temperature range of 0-30 °C. However, both types of interactions are affected by modifications in water structure that occur as a function of temperature [30]. In cold-adapted enzymes,

K_m will be determined by the contributions of the various bond types involved in the ES interactions. Enzymes with low K_m have more negative binding energy (ΔG_{ES}) than those with higher K_m, in accordance with the relationship:

$$\Delta G_{ES} = -RTln(1/K_m). \tag{1.5}$$

As a result of higher ES affinity, the reaction falls into a deeper free energy well (more negative ΔG_{ES}) from where it has to climb in order to reach the transition state [11, 14, 15]. An enzyme can increase k_{cat} by lowering $\Delta G^{\ddagger}$. This can be achieved by stabilising the activated substrate in the transition state or by destabilising the ES complex (increasing K_m) [31]. There is experimental evidence that the substrate, in its transition-state but not in its ground-state form, interacts via strong electrostatic bonds and strong hydrogen bonds with the enzyme's active site [22]. As a result, compared to the ground-state substrate, enzyme affinities for the activated substrate are expected to increase much more steeply with decreasing temperature [23]. A recent computational study of the reaction of differently thermal adapted citrate synthases has suggested that the psychrophilic enzyme is less pre-organised towards electrostatic stabilisation of the activated ES complex with respect to the heat-stable homologues [32]. On the other hand, the increased flexibility of the cold-active enzyme, in regions far from the active site, allows it to accomodate the activated ES complex at a lower energetic cost, counterbalancing the decrease in stabilising enzyme-substrate electrostatic interactions, so that the in this case the enthalpic gain that results in enhanced catalytic activity is to be ascribed especially to processes occurring inside the psychrophilic enzyme.

The majority of cold-adapted enzymes have a higher k_{cat} and K_m than their thermostable counterparts (Tab.1.1), with the exception of enzymes, which work at $[S]$ close to K_m [15]. This relationship is well illustrated for α-amylases, wherein mutants of AHA tend to exhibit proportional decreases in k_{cat} and K_m [14, 33, 34], and has been recently seen in other contexts [35]. To facilitate substrate binding at a low energy cost, the active site of cold-adapted enzymes tends to be larger and more accessible to the substrate [22]. In addition to increase K_m, these structural features have been found to cause a reduction in substrate specificity [22].

Table 1.1: Catalytic kinetic and thermodynamic activation parameters, for selected psychrophile/mesophile pairs. Adapted from reference [22].

Enzyme	micro-organism	T_{opt} (°C)	k_{cat} (T) (min^{-1} (°C))	K_m (substrate) (mM)	$\Delta H^{\ddagger a}$	$\Delta S^{\ddagger b}$	$\Delta G^{\ddagger a}$
α-amylase (P)	*Pseudoalteromonas haloplanktis*	28	41820 (25)	0.23 (pNME)	35	-81	58
α-amylase (M)	*Bovine*	54	17460 (25)	0.06 (pNME)	46	-43	59
Trypsin (P)	*Gadus morhua*	50	240 (25)	0.08 (amide)	32	-57	49
Trypsin (M)	*Bovine*	47	120 (25)	0.7 (amide)	53	-25	60
Subtilisin (P)	*Bacillus. sp. TA41*	40	1920 (5)	0.026 (amide)	36	-92	62
Subtilisin (M)	*Bacillus subtilis*	-	1000 (5)	0.006 (amide)	46	-70	66
EF-2 GTPase (P)	*Methanococcoides burtonii*	-	2.0 (40)	0.02 (GTP)	75	-35	86
EF-2 GTPase (M)	*Methanosarcina thermophila*	-	3.7 (40)	0.03 (GTP)	88	13	84
Isocitrate Dehydrogenase (P)	*Colwellia maris*	20	1.0 (15)	0.062 (isocitrate)	6.5	-189	61
Isocitrate Dehydrogenase (M)	*Azotobacter vinelandii*	42	1.2 (15)	0.008 (isocitrate)	14.7	-157	60
Glutamate Dehydrogenase (P)	*Chaenocephalus aceratus*	30	228 (5)	2.0 (glutamate)	34	-114	65
Glutamate Dehydrogenase (M)	*Bovine*	30	54 (5)	0.87 (glutamate)	59	-32	68
Chitobiase (P)	*Arthrobacter sp.TAD20*	-	5880 (15)	0.027 (NPAG)	45	-51	60
Chitobiase (M)	*Serratia marcescens*	-	1080 (15)	0.038 (NPAG)	72	28	64
Chitinase A (P)	*Arthrobacter sp.TAD20*	-	102 (15)	( )	60	-31	69
Chitinase A (M)	*Serratia marcescens*	-	235 (15)	( )	74	25	67

(^{a}) $kJ\ mol^{-1}$;
(^{b}) $Jmol^{-1}K^{-1}$.

1.1.4 The activity-flexibility-thermolability trilogy

Enzyme catalysis generally involves the "breathing" of particular regions of the enzyme, enabling the accommodation of the substrate. The ease of such molecular movements may be one of the determinants of catalytic efficiency. Therefore optimising the function of an enzyme at a given temperature requires a proper balance between two opposing factors: structural rigidity, allowing the retention of a specific 3D conformation at the physiological temperature, and flexibility, allowing the protein to perform its catalytic function [36, 37]. Low temperatures tend to improve the compactness of a protein by disabling the breathing. Since heat-adaptation is generally correlated with the rigidification of a protein, psychrophily, at the opposite end of the temperature scale, should be characterised by an increase of the plasticity or flexibility of appropriate parts of the molecular structure in order to compensate for the lower thermal energy provided by the low temperature habitat. This plasticity would enable a good complementarity with the substrate at a low energy cost, thus explaining the high specific activity of psychrophilic enzymes. In return, this flexibility would be responsible for the weak thermal stability of the psychrophilic en-

zymes. While the decreased stability of psychrophilic enzymes does support this hypothesis, there is, however, no direct experimental evidence of an increase in flexibility.

Indeed, the flexibility of a protein, especially that related to activity and/or stability, remains a difficult parameter to determine experimentally, as the increase in flexibility can be limited only to a small but crucial part of the protein. Flexibility is complex to define: a protein in its native state does not present a single conformation. In the context of temperature adaptation, the notion of flexibility has to be defined as a parameter related directly to the specific activity of the enzyme. If flexibility can be described in terms of dynamic motion that is related to a specific time-scale, it can also be considered as a concept related to the amplitude of the deformation of the protein conformation at a given temperature. Techniques that provide some evaluation of the relative flexibility are: relative resistance to proteolysis associated with amino-acid sequencing to identify more flexible parts of a protein; fluorescence polarisation giving access to the freedom of orientation; hydrogen/deuterium (H/D) exchange experiments monitored by NMR or FTIR spectroscopy, and molecular dynamics simulations [7]. Nevertheless, the H/D exchange technique suffers from the disadvantage of being a measure of the accessibility of deeply buried residues and thus does not detect local flexibility, in particular that associated with the active site which is generally quite accessible [38]. These exchange rates could be in disagreement with the type of flexibility related to catalytic efficiency at low temperatures and additional measurements over a wide range of timescales are needed to distinguish the flexibility of psychrophilic, mesophilic and thermophilic enzymes. So far, conformational flexibility of psychrophilic enzymes has been strongly supported by dynamic fluorescence quenching experiments, using acrylamide as the quencher [39–42]. In this technique, the decrease of fluorescence arising from diffusive collisions between the quencher and the fluorophore (tryptophan residues), reflects the ability of the quencher to penetrate the structure and can consequently be viewed as an index of protein permeability. Recently, fluorescence quenching performed on the psychrophilic α-amylase from *P. haloplanktis* (AHA) and its pig pancreatic mesophilic (PPA) and *Bacillus amyloliquefaciens* thermostable homologues (BAA) revealed that the variation of fluorescence quenching between 10 and 37 °C decreases in the order psychrophile → mesophile

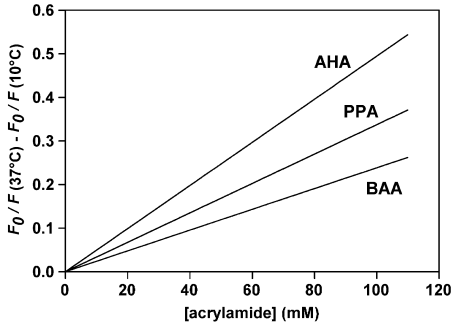


Figure 1.3: Conformational flexibility of psychrophilic, mesophilic and thermophilic enzymes. Curves representing the acrylamide quenching of the tryptophan fluorescence of psychrophilic (AHA), mesophilic (PPA) and thermostable (BAA) α-amylase. Differences of fluorescence ratios F_0/F at 37 and 10°C were plotted as a function of acrylamide concentration (mM), to take into account differences in the number of tryptophan residues. Image taken from reference [7].

→ thermophile [7, 39] (see Fig.1.3).

Thus, the increase in protein permeability in a temperature range where the native state prevails is much larger for AHA, intermediate for PPA, and low for BAA, indicating a strong correlation between the stability and conformational flexibility. The temperature factors (*B*-factors) derived by the crystal structures are measurements of atomic fluctuations in the crystal, but are not well suited for comparison between individually refined protein structures. The main reason for this is that the crystallographic temperature factors are highly biased by external factors such as data collection protocols and methods, data processing and scaling, as well as refinement protocols. In addition, atomic fluctuations in a crystal are highly influenced by intermolecular close contacts, and long-range effects from such interactions are in most cases difficult to estimate. In this method, as in the H/D exchange experiments, increased attention must be paid on where the analysed atoms are located in the structure [7]. Remarkable differences are seen for particular regions, such as for residues involved in catalysis or particular protein domains, for some of the low temperature adapted enzymes [43]. Increased local mobility may affect both activity and stability, but no definite conclusions can be drawn just from X-ray structure analysis. Neutron scattering analysis gives insight into the fast atomic thermal motions in macromolecules on the picosecond to nanosecond timescale, acting as a "lubricant" for larger conformational changes

[44]. One advantage of neutron scattering analyses is that they provide a direct method for measuring static and dynamic flexibility [45] and they allow analysis of numerous sample preparations [46]. Results indicate that flexibility is required for biological activity and flexibility increases from thermophiles to psychrophiles. The low stability of psychrophilic enzymes in comparison with their mesophilic homologues has been demonstrated by the drastic shift of their apparent optimal temperature of activity, the low resistance of the protein to denaturing agents and the high propensity of the structure to unfold at moderate temperatures. The decreased stability of psychrophilic enzymes, in addition to their increased low temperature activity, suggests that there is a direct link between activity and stability: the maintenance of activity at low temperature requires the weakening of intra-molecular forces which results in reduced stability. Therefore, the reduced stability of the enzyme may not necessarily arise from a general reduction in the strength of intra-molecular forces, but from weakened interactions in one or few important regions of the structure (localised flexibility) [7, 47–50]. The increased flexibility may reduce the energy barrier of the rate-governing shifts in conformation and, thereby, increase k_{cat}. On the other hand, determination of a causal relationship is difficult, and in this context it has been suggested that lower stability of psychrophilic enzymes may be due to random genetic drift as a consequence of the lack of evolutionary pressure for thermostable proteins at low temperatures [51]. In addition, global stability and localised flexibility of enzymes may have evolved independently [52].

Flexibility is the main adaptive character of psychrophilic enzymes, responsible for the decrease of activation enthalpy $\Delta H^{\ddagger}$ that efficiently increases k_{cat} at low temperatures. However, acquirement of flexibility in regions that are not directly involved in catalytic motions has no effect on $\Delta H^{\ddagger}$, but on the contrary may decrease $\Delta S^{\ddagger}$, therefore impairing the gain in k_{cat}. Accordingly, if negative values of $\Delta\Delta H^{\ddagger}_{p-m}$ (difference in activation entropy change between psychrophile ($_p$) and mesophile ($_m$)) arise form local flexibility in parts of the structure involved in the catalytic movements, the opposite negative effect of $\Delta\Delta S^{\ddagger}_{p-m}$ on k_{cat} can be attenuated by local rigidity in regions of the enzyme independent of the catalytic motions. Therefore, local rigidity may be a positive factor in the adaptation of psychrophilic enzymes if such independent regions do exist. In general, two types of adaptation already emerged.

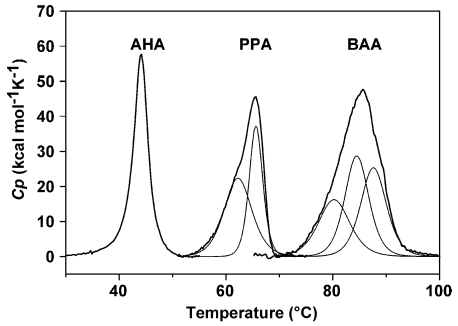


Figure 1.4: Thermal unfolding of three α-amylase recorded by differential scanning calorimetry (DSC). When compared with the psychrophilic AHA (from *Pseudoalteromonas haloplanktis*), the heat-stable enzymes PPA (from pig pancreas) and BAA (from *Bacillus amyloliquefaciens*) are characterised by higher T_m (top of the transition) and ΔH_{cal} (area under the transition) values, by a flattening of the transition and by the occurrence of calorimetric domains indicated by deconvolutions in thin lines. Picture taken from reference [7].

The enzyme could have evolved towards the lowest possible stability of its native state, as exemplified by the cold-active α-amylase from *P. haloplanktis* (Fig.1.4). On the other hand, only one region of the protein has acquired a higher flexibility, resulting in the appearance of a distinct heat-labile thermodynamic domain, as exemplified by a psychrophilic phosphoglycerate kinase [53]. The latter behaviour has substantiated the concept of local flexibility as an adaptive strategy of psychrophilic enzymes. A first explanation may be found in the structural motion required by catalysis. Indeed, the above mentioned psychrophilic enzymes displaying local flexibility/rigidity, rely on a hinge-bending mechanisms or on marked structural rearrangements. By contrast, the conformational changes associated with activity are extremely discrete in mesophilic α-amylases [54] and in the uniformly heat-labile psychrophilic α-amylases [8]. In this case, improving the overall flexibility could be the unique but very efficient parameter prone to adaptation of the enzyme to low temperatures. It follows that domain movements may be a prerequisite or a favourable background for the acquirement of local flexibility/rigidity. Another avenue for investigations, aimed at explaining whether the two types of conformational stability involve the necessity to maintain K_m values compatible with the cell metabolism. Indeed, lowering the magnitude of $\Delta S^{\ddagger}$ also restricts the number of conformational states available for the enzyme-substrate complex ES and hence improves the affin-

Table 1.2: Thermodynamic parameters for the irreversible heat inactivation of activity and for irreversible unfolding.

		Inactivation				Unfolding			
		ΔG^*	ΔH^*	$T\Delta S^*$	ΔS^*	ΔG^*	ΔH^*	$T\Delta S^*$	ΔS^*
α-amylase[a]	AHA	20.5	172.3	151.8	479	20.2	109.7	89.2	286
	PPA	21.5	153.0	131.5	395	21.5	84.7	63.2	190
	BAA	22.9	58.6	35.7	101	22.9	74.2	51.3	145
xylanase[b]	pXyl	19.6	109.4	89.8	265.6	19.6	98.9	79.3	234.5
	CelA	27.5	72.9	45.4	134.2	27	82.4	55.4	163.9

([a]) As the three α-amylases (psychrophilic AHA, mesophilic PPA and thermophilic BAA) are denatured in different temperature ranges, activation data are reported for an identical rate constant, $k = 0.05s^{-1}$.
([b]) Parameters recorded at 65°C, corresponding to a $k_{inactivation} = 1.4s^{-1}$ and $1.210^{-5}s^{-1}$ for psychrophilic pXyl and thermophilic CelA, respectively, and a $k_{unfolding} = 1.7s^{-1}$ and $1.910^{-5}s^{-1}$ for pXyl and CelA, respectively.

ity for substrates. As a matter of fact, all enzymes displaying local flexibility/rigidity are intra-cellular enzymes relying on K_m to regulate their activity (LDH, PGK, CS) or have adapted through strong K_m improvement, such as trypsins [55]. On the other hand, the cold-active α-amylase works under saturating substrate concentrations. This enzyme does not rely on K_m to a great extent under physiological conditions, therefore avoiding the need for a stable domain. Some significant advances in the understanding of the activity-stability relationship have also been drawn from the analysis of heat inactivation and irreversible unfolding of some psychrophilic enzymes compared to some of their mesophilic and thermophilic counterparts [39–41]. Both heat inactivation of activity and irreversible unfolding of the structures display the same trends.

An increase in activation free-energy from psychrophilic to heat-stable enzymes is observed and the same value for the denaturing rate constant k is reached at increasing temperatures. The small differences in ΔG^* of inactivation and unfolding values however arise from large differences in the enthalpic and entropic contributions. Indeed, the lowest ΔG^* values for the psychrophilic enzyme corresponds to the largest ΔH^* and $T\Delta S^*$ contributions, and conversely for the heat-stable homologues. The decrease of ΔH^* as enzyme stability increases mainly reflects the decrease in cooperativity of inactivation and of unfolding: indeed, the heat-labile enzymes denature over a shorter temperature range, leading to a steep slope of Arrhenius plots and subsequently to high activation energy E_a and ΔH^*. Such high cooperativity probably originates from the lower number of interactions that need to be broken

in order to disrupt the active conformation. In this formalism, $\Delta S^* > 0$ suggests that the randomness of the activated transition state increases before irreversible inactivation or unfolding. Accordingly, the transition state of the psychrophilic enzymes is reached through a larger entropy variation than that of heat-stable homologues, reflecting an increased disorder, which is the main driving force of heat denaturation. In addition, thermophilic enzymes seem to resist denaturation before the concomitant irreversible loss of activity and conformation (low ΔS^*, see also Fig.1.1).

As previously noted, some authors consider that the instability of psychrophilic enzymes is due to random genetic drift, as a consequence of the lack of evolutionary pressure for stable enzymes in low temperature environments [51]. Recently, random mutagenesis techniques using essentially error-prone polymerase chain reaction (*PCR*) amplification, DNA shuffling or, in a few cases, mutating strains [56–58] have been used to study the thermal adaptation of different enzymes to understand the relationship, if there is any, between stability, activity and flexibility. Basically, random mutations are introduced in a gene and the resulting library is screened for the acquisition or improvement of a specific property such as thermostability or improved activity. This approach detects mutations responsible for immediate and drastic adaptation in response to a strong selection imposed in the laboratory and based on a physicochemical property. In contrast, natural evolution is a very slow process driven by complex environmental and metabolic constraints, and enabling the viability or the performance of a living organism in a specific environment. When random mutants were only screened for high activity at low temperatures during direct evolution [57], the selected enzymes invariably displayed a decreased stability as well as a higher supposed flexibility, i.e., the canonical properties of psychrophilic enzymes. By contrast, when the selection pressure was thermal stability, mutants with higher stability and unchanged or enhanced catalytic activity were found, in contradiction with the hypothesis that a high stability induces a low catalytic efficiency. For some enzymes, early generation of stabilised mutants exhibited a decreased activity, as observed in nature, while several generations of mutants were required to obtain both a high thermostability and high activity. If it appears possible, at least in the laboratory, to uncouple activity and stability [59], the fact is, that enzymes displaying a high stabil-

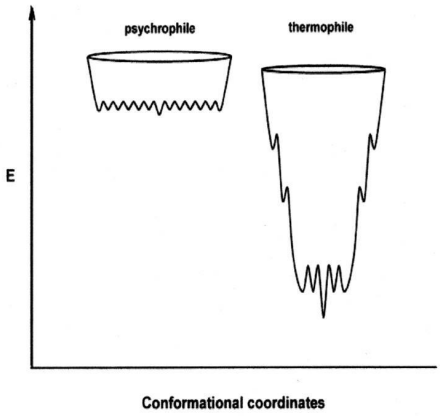


Figure 1.5: Proposed model of folding funnels for psychrophilic and thermophilic enzymes. In these schematic energy landscapes, the conformational stability (E) is represented as a function of the conformational diversity. Picture taken from reference [7].

ity together with a high flexibility or activity do not exist in nature, except for thermophilic enzymes that catalyse the conversion of labile metabolic intermediates [60]. Moreover, in psychrophilic organisms a selective pressure towards a low stability of proteins does exits; it is exerted by the low environment temperature which has to select molecular structures flexible enough to ensure efficient catalysis. Therefore, improvement of activity at low temperatures associated with a loss of stability appears to be the most frequent and accessible strategy at the level of psychrophilic proteins. As a consequence, the low stability of cold-adapted enzymes may be the simplest way to acquire activity in the absence of selection for stable proteins but the occurrence of heat-stable psychrophilic enzymes should not necessarily be ruled out.

1.1.5 The folding funnel model

D'Amico and coworkers [39] have proposed to integrate the kinetic, structural and biophysical data available on enzymes adapted to extreme temperatures in a model based on a new point of view, using the folding funnel model used to describe folding-unfolding reactions. The energy landscape of psychrophilic and thermophilic enzymes is described in Fig.1.5.

The height of the funnel, i.e., the free energy of folding, corresponds to the conformational stability and the upper edge of the funnel is occu-

pied by the unfolded state in random coil conformation. Accordingly, the edge of the funnel for the psychrophilic proteins is slightly larger (broader distribution of the unfolded state). When the folding of the poly-peptide occurs, the free energy level decreases, as well as the size of the conformational ensemble. However, thermophilic proteins show intermediate states corresponding to local energy minima. These minima are responsible for the roughness of the funnel slopes and for the reduced cooperativity of the folding-unfolding reaction. By contrast, the structural elements of psychrophilic proteins generally unfold cooperatively without intermediates, as a result of fewer stabilising interactions: therefore the funnel slopes are steep and smooth. The bottom of the folding funnel for a very stable and rigid thermophilic protein can be depicted as a single global minimum or as having only a few minima with high energy barriers between them [61]. On the contrary, the bottom for an unstable and flexible psychrophilic protein is rugged and depicts a large population of conformers with low energy barriers between them. Rigidity of the native state is therefore a direct function of the energy barrier height [61]. In this context, the activity-stability relationship in these extremozymes depends on the properties of the base of the folding funnel. Indeed, it has been argued that upon substrate binding to the association-competent sub-population, the equilibrium between all conformers is shifted towards this sub-population, leading to the active conformational ensemble [61, 62]. In the case of the rugged bottom of the folding funnel of psychrophilic enzymes, this equilibrium shift only requires a modest free energy change and a low enthalpy change for inter-conversion of the different conformations of the numerous conformational isomers existing in the wide conformer ensemble. The low energy barriers across a spectrum of micro-states would result in an overall decreased activation energy and a high k_{cat}. Moreover, the increased flexibility would cause the cold-adapted enzyme to spend more time in conformations that are not optimal for substrate binding and result in high K_m [63]. It would be insightful to test this model by measuring the activity and activation energies of individual molecules [64–66] of sets of cold-adapted enzymes and their thermostable homologues.

Table 1.3: Crystal structures of psychrophilic enzymes.

Enzyme	*Organism*	Growth temperature (°C)	Res. (Å)	PDB code
α-amylase	*Pseudoalteromonas haloplanktis*	4	2	1AQH
citrate synthase	*Antarctic bacterium DS2-3R*	5	2.09	1A59
elastase	*Salmo salar*	4	1.61	1ELT
glyceraldehyde-3-phosphate dehydrogenase	*Homarus americanus*	20	2.8	4GPD
Malate dehydrogenase	*Aquaspirillium arcticum*	4	1.9	1B8P
Alkaline phosphatase	*Pandalus borealis*	2	1.92	1K7H
Triose-phosphate isomerase	*Vibrio marinus*	15	2.7	1AW1
Trypsin	*Salmo salar*	4	1.8	2TBS
Trypsin	*Oncorhynchus keta*	5	1.8	1MBQ
Trypsin	*Atlantic cod*	4	1.85	2EEK
adenylate kinase	*Bacillus globisporius*	5	2.25	1S3G
chitinase	*Arthrobacter Tad20*	5	1.74	1KFW
uracil-DNA-glycosylase	*Gadus morhua*	5	1.9	1OKB
metallo-protease	*Pseudomonas tac II*	4	2.5	1O0T
catalase	*Vibrio salmonicida*	10	1.97	2ISA
subtilisine-like serine-protease	*Vibrio sp PA-44*	?	1.84	1SH7
Sphericase subtilisin-like	*Bacillus subtilis psycrophile*	?	1.4	2GKO
aspartate carbamoyl-transferase	*Moritella profunda*	2	2.85	2BE7
cellulase (endoglucanase)	*Pseudoalteromonas haloplanktis*	4	1.6	1TVP
xylanase	*Pseudoalteromonas haloplanktis*	4	1.2	1H12
pepsin	*Gadus morhua*	5	2.16	1AM5
phenylalanine hydroxylase	*Colwellia Psychrerythraea*	8	1.5	2V27
isocitrate dehydrogenase	*Desulfotalea Psychrophila*	10	2.3	2UXR
beta-lactamase class C	*Pseudomonas fluorescens*	25	2.26	2QZ6
endonuclase I	*Vibrio salmonicida*	10	1.5	2PU3
aminopeptidase	*Colwellia Psychrerthraea*	8	2.79	3CIA
lactate-dehydrogenase	*Champsocephalus gunnari*	1 - 2	2.35	2V65
lipase	*Photobacterium lipolyticum*	25-28	2.2	2ORY
pheromone en-6	*Antarctic Ciliate Euplotes nobilii*	3 - 4	NRM	2JMS
lipase B	*Candida antarctica*	10	2.5	1LBT

1.1.6 Structural adaptation to low temperatures

Refined 3D structures (see Tab.1.3) and structural models of psychrophilic enzymes have been useful in compiling an inventory of the structural factors possibly involved in cold-adaptation and also responsible for the low thermal stability [7].

It was found that all structural factors currently known to stabilise mesophilic enzymes could be attenuated in strength and number in psychrophiles and only minor structural modifications are needed to adapt a mesophile to cold temperatures. Frequently, the catalytic cavity seems to be larger and more accessible to ligands in psychrophilic enzymes than in mesophilic enzymes [43, 67]. This is achieved by the deletion of residues in loops bordering the active site, by the distinct conformation of these loops or by the replacement of bulky side-chains with smaller groups at the entrance of the active site [68]. This improved accessibility is though not only to be responsible for the accommodation of the substrate at low energy cost, but also to facili-

tate the release and exit of the reaction products [7, 33]. The residues and the geometry of the active site are conserved across broad taxonomic groups and from one extreme of temperature adaptation to the other. It appears that substrate binding is so specific that any alteration in residues involved in catalysis almost invariably leads to a reduction in catalytic efficiency, if not a complete loss of function. Thus, the locations of the residue substitutions described above, both for psychrophiles and thermophiles, must occur at some distance from substrate binding sites, and must alter function by changing the relative mobility of structures, rather than attributes of the binding surface [33]. In some enzymes, the electrostatic potential around the active-site region is also modified in order to attract the oppositely charged ligands and to channel the substrate towards the catalytic cavity [7, 43, 69, 70]. In all cases, the differences were caused by discrete substitutions in non-conserved charged residues resulting in local electrostatic potentials differing in both sign and magnitude and could be responsible for a better substrate binding by facilitating the interaction of oppositely charged ligands with the surface of the enzyme. The typical configuration of the active sites in cold-adapted enzymes suggests that these enzymes could display a broader specificity, due to the fact that diverse substrates having slightly distinct conformations or sizes can fit and bind to the active site [71]. As a second consequence, substrates should bind less firmly in the binding site, giving rise to higher K_m values as reported for cold active enzymes devoid of adaptive mutations within the catalytic centre. Although factors specific to each enzyme, or at least enzymatic family, constrain the set of available adaptive strategies, some seemingly universal structural factors involved in cold-adaptation have been provided by comparative statistical studies between psychrophilic and heat-stable enzymes [68, 72]. Compared to their mesophilic counterparts, cold adapted enzymes often feature increased numbers of glycine, alanine and serine residues, and in general of small residues, which promote flexibility. On the other hand, they exhibit a reduction in *(i)* the number of proline residues, which constrain the structure, especially in loops, *(ii)* of charged and polar residues, especially arginine, which mediates both ion pairs and hydrogen bonds, but also lysine and glutamate, and *(iii)* of large aliphatic residues as isoleucine and leucine, which stabilise the protein structure through hydrophobic interactions. Furthermore, and in part as a re-

sult of the aforementioned amino-acidic preferences and avoidances, they exhibit fewer hydrogen bonds, especially between side chains, ion pairs and aromatic interactions, and lower hydrophobicity. It should be noted in this respect that the hydrophobic effect itself is weakened by low temperatures, so the decrease in hydrophobic residues and the loss of hydrophobicity may simply be an artifact of its reduced ability to provide intra-molecular stabilisation [7]. In fact, reduction of hydrophobicity has also been identified in the inter-subunit interfaces of oligomeric proteins [5, 73], balanced by an increase in the number of ion pairs [74, 75]. Other features include reduction of side chain solvent accessible surface and increased apolar exposed surface, weakening of charge-dipole interactions in α-helices and decreased strength of anchoring of N and C-termini; floppy protein extremities appear in fact to be preferential sites for unfolding initiation. Increased surface charge has been observed in some cases [76], expected to improve interactions with the solvent and in turn enhance the flexibility or the "breathing" of the external shell of cold-active enzymes. As previously stated, not all of the above structural adjustments are employed in each psychrophilic enzyme, even though they often indeed manifest statistical significance. Interestingly, the same structural factors have been implicated in the stability of thermophilic proteins [36, 37] suggesting that there is a continuum in the strategy of protein adaptation to temperature [7]. These studies have provided valuable information regarding the importance of certain types of amino-acid substitutions in adaptation to extreme temperatures.

1.1.7 The family-centred approach to cold adaptation

Even if some common trends are observed, comparative studies of psychrophilic, mesophilic and thermophilic enzymes report that each protein family adopts different structural strategies to adapt to low temperature [14]. In the same way as it seems to be the case for thermophilic proteins, it can probably be concluded that structural rationalisation for cold-adaptation can not be completely generalised. Adaptation to low temperatures, the process of achieving the universally observed reduced stability, which presumably results in an increased flexibility, seems to be very strongly constrained by the specific features of each enzymatic family. In addition, the differences in unfolding free en-

ergies between cold active and the corresponding mesophilic enzymes are usually small and often caused by subtle structural changes, that cannot easily be distinguished from phylogenetic noise. For these reasons, comparative studies should minimise genetic drift between compared enzymes in order to maximise the possibility to identify differences that are actually related to cold adaptation. In this sense, the numerous studies on enzymes from cold-adapted fish have been relatively successful due to their particular close homology with mammalian counterparts [69].

On the other hand, if restricting comparisons to single enzymatic families is a requirement for significance, the necessity arises for large-scale comparisons among many of these small groups, combining the results in order to obtain a global picture of cold adaptation, able to provide the necessary insight to guide the rational design of thermally adapted biocatalysts.

1.1.8 Cold-adapted enzymes applications

Driven by increasing industrial demands for biocatalysts that could cope with industrial process conditions, considerable efforts have been devoted to the search of new enzymes [77]. Compared with organic synthesis, biocatalysis often provides far better chemical precision, enabling stereoselective processes at high yields with few side-products and a lower environmental burden [77]. Despite the fact that up to date more than 3000 different enzymes have been identified and many of them have found biotechnological and industrial applications, the present enzyme toolbox is still not sufficient to cover all the demands. A major cause is the fact that many available enzymes do not resist industrial reaction conditions. As a result, the characterisation of micro-organisms that are able to thrive in extreme environments has received great attention, in the hope they may be the source of novel enzymes [77, 78]. It is likely that in the near future, the potential value of cold-adapted enzymes will rise to a greater annual market than for thermostable enzymes, representing today about US$ 250 million [79]. Cold-adapted enzymes offer are particularly suited for industrial use, owing to their properties: they avoid the requirement for expensive heating steps, provide increased reaction yields, accommodate a high level of stereospecificity, minimise the number of chemical reactions

and undesirable side-products that are more likely to occur at higher temperatures and exhibit thermal lability for rapid and easy inactivation of the enzyme when required [67]. The ability to heat-inactivate cold-active enzymes at low temperatures has particular relevance to the food industry where it is important to prevent any modification of the original heat-sensitive substrates and products. This is also a benefit in sequential process (e.g. molecular biology) where the actions of an enzyme need to be terminated before the next process is undertaken; with cold-adapted enzymes this might be accomplished by heat inactivation rather than chemical extraction. Cold-active enzymes may also find application in mixed aqueous-organic or non-aqueous solvents for the purpose of organic synthesis thanks to their enhanced flexibility, which counteracts the rigidifying effects of low water activity in organic solvents [67].

1.2 Theoretical Study Of Enzymes

1.2.1 Protein dynamics: basic concepts and historical background

The native state of a protein is often considered as the conformation in which the protein features a fully folded structure which will in turn allow the protein to deliver its function. Indeed functionality in the great majority of cases strongly depends on the protein's ability to modify its structure, in order for example to accommodate or enclose the substrate, or to undergo activating conformational rearrangements. Proteins are in fact dynamic entities, experiencing constant remodelling of their structure according to a series of energetic equilibria. Therefore the native state is more accurately defined as a statistical distribution of micro-states with conformations differentiated by several levels of local unfolding [80, 81]. As a consequence enzyme molecules present different affinities for substrate [82], which may explain experimental observed molecular sub-populations with discrete substrate affinities in a "homogeneous" enzyme preparations [64].

Studying protein dynamics is a task very few experimental methods are up to, often because of issues related to time-scale and analysis resolution. *Molecular Dynamics* (*MD*) simulation techniques provide the possibility to obtain dynamic information on large protein systems

with atomic detail in the picosecond to microsecond time scale.

The Molecular Dynamics method was first introduced by Alder and Wainwright in the late 1950's to study the interactions of hard spheres. In 1964 Rahman carried out the first simulation using a realistic potential for liquid argon. The first molecular dynamics simulation of a realistic system was done by Rahman and Stillinger in their ~2ps simulation of liquid water in 1974 . The first protein simulations was reported in 1977 and consisted of a 9.2ps simulation of the bovine pancreatic trypsin inhibitor (*BPTI*) in vacuum. Since then the seemingly never ending improvements in computer hardware have allowed simulation of bigger systems, such as large proteins in explicit water, for longer time-intervals: the μs barrier was broken in 1998. Furthermore modern simulations use better parametrisations and fewer approximations, improving the treatment of long-range electrostatic forces, and address a variety of issues including the thermodynamics of ligand binding, the computation of free energies and the folding of small proteins.

1.2.2 Theoretical formalisms

Interactions between atoms can be rigorously modelled only by taking into account their electrons, and thus by using the quantum mechanical (*QM*) formalism, that requires the solution (exact or approximate) of the relativistic Schrödinger equation to obtain detailed information about a system's properties and their time-dependent evolution. Explicit treatment of electrons, however, is extremely consuming in terms of computational power, since complexity grows exponentially with the number of particles simulated. Therefore *QM* approaches are only affordable for small numbers of atoms, and cannot be applied to large protein systems. Nevertheless, modelling the simulated atoms using classical mechanics Newton's laws of motion, and thus considering all electrons to be in their quantomechanical ground state, yields an adequate description of the system, but prevents the possibility to correctly represent some processes, one of the most important being bond breaking and formation.

1.2.2.1 Classical mechanics

As said, the atoms composing a system can be treated, within very good approximation, as classical particles interacting via potential en-

ergy functions. A system of N particles at a given time τ is completely defined by the generalised coordinates $\mathbf{q}_i(\tau)$ and conjugate momenta $\mathbf{p}_i(\tau)$ of all the $i = \{1, \ldots, N\}$ particles.

Conservative systems are defined as allowing the existence a *potential energy* function $\mathcal{V}(\mathbf{q})$ such that forces acting on the particles are derivable from it by differentiation:

$$\mathbf{F}_i = -\frac{\partial \mathcal{V}}{\partial \mathbf{q}_i} \quad \mathcal{V} = \mathcal{V}(\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_N). \tag{1.6}$$

The equations that describe the time evolution of the system's particles (*equations of motion*), can be elegantly derived using Lagrange's formulation, which has the advantage of being invariant to the particular set of coordinates chosen. The classical Lagrangian $\mathcal{L}$ is defined as the difference between the system's kinetic and potential energies, expressed as functions of the generalised coordinates $\mathbf{q}_i$ and velocities $\dot{\mathbf{q}}_i$*

$$\mathcal{L}(\mathbf{q}, \dot{\mathbf{q}}) = \mathcal{K}(\dot{\mathbf{q}}) - \mathcal{V}(\mathbf{q}) = \sum_{i=1}^{N} \frac{1}{2} m_i \dot{\mathbf{q}}_i^2 - \mathcal{V}(\mathbf{q}). \tag{1.7}$$

In terms of the Lagrangian the classical equations of motion are:

$$\frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial \mathcal{L}}{\partial \dot{\mathbf{q}}_i} - \frac{\partial \mathcal{L}}{\partial \mathbf{q}_i} = 0. \tag{1.8}$$

The solution of equations of motion for a given initial condition $\{\mathbf{q}_i(0), \dot{\mathbf{q}}_i(0)\}$ gives the complete set of generalised coordinates and velocities for all particles at every time t. It should be noted that Lagrange's equations of motion, when the Cartesian coordinates are used, are equivalent to Newton's second law for many-particles systems.

Lagrange's equations of motion are a set of second order differential equations. The formulation can be simplified by defining the classical Hamiltonian for the system, which is a Legendre transformation of the Lagrangian applied in order to replace its velocity ($\dot{\mathbf{q}}_i$) dependence with

*the *dotted* notation, first introduced by Isaac Newton, is used to express time derivatives, so $\dot{\mathbf{q}}_i$ is equivalent to $\frac{d\mathbf{q}_i}{dt}$.

a dependence on the generalised momenta $\mathbf{p}_i$, defined as:

$$\mathbf{p}_i = m_i \dot{\mathbf{q}}_i = \frac{\partial \mathcal{L}}{\partial \dot{\mathbf{q}}_i}. \tag{1.9}$$

The Hamiltonian is therefore defined as:

$$\begin{aligned} \mathcal{H}(\mathbf{q}, \mathbf{p}) &= \sum_{i=1}^{N} \dot{\mathbf{q}}_i \mathbf{p}_i - \mathcal{L}(\mathbf{q}, \dot{\mathbf{q}}) \\ &= 2 \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} - \mathcal{L}(\mathbf{q}, \dot{\mathbf{q}}) \\ &= 2\mathcal{K}(\mathbf{p}) - \mathcal{K}(\dot{\mathbf{q}}) + \mathcal{V}(\mathbf{q}) \\ &= \mathcal{K}(\mathbf{p}) + \mathcal{V}(\mathbf{q}) \end{aligned} \tag{1.10}$$

and is seen to be the total energy of a conservative system. Since in conservative systems total energy **E** is constant, if follows that the Hamiltonian will be constant:

$$\mathcal{H}(\mathbf{q}(t), \mathbf{p}(t)) = \mathcal{H}(\mathbf{q}(0), \mathbf{p}(0)) = \mathbf{E}.$$

Hamilton's equations of motion of a system are obtained by partial differentiation of the Hamiltonian:

$$\begin{cases} \dfrac{\partial \mathcal{H}}{\partial \mathbf{p}_i} &= \dot{\mathbf{q}}_i, \\ \dfrac{\partial \mathcal{H}}{\partial \mathbf{q}_i} &= -\dot{\mathbf{p}}_i. \end{cases} \tag{1.11}$$

Solving Hamilton's equations of motions for each simulation time-step yields the time-dependent evolution of the system in terms of its particles' generalised coordinates and conjugate momenta.

1.2.2.2 Potential and forcefields

Calculation of the potential energy $\mathcal{V}(\mathbf{q})$ from the generalised coordinates is the most important aspect of *MD* simulation, since the better $\mathcal{V}(\mathbf{q})$ approximates the real interactions between the simulated particles, the more reliable will be the description the simulation provides of both dynamics and equilibrium properties of the system. Potential energy calculation is usually accomplished using simple potential energy

functions $\mathcal{V}(\mathbf{q},\mathbf{k})$, which depend on the generalised coordinates of the simulated atoms nuclei ($\mathbf{q}$), and which are parameterised ($\mathbf{k}$) in order to reproduce the effects of electrons. A set of $\mathcal{V}$ functions and $\mathbf{k}$ parameters for each function, structured so that interactions between different types of atoms are handled suitably, is usually referred to as *Force Field*. Clearly it is not possible to parameterise the interactions between the atoms of every existing molecule; therefore the *Force Field* correctly handles the interactions between a great number of small groups of atoms, and applies the same parameterisation to each group in any context it may be found. Several *Force Fields* have been developed and optimised to study a range of different molecules, including biomolecules, organic molecules, polymers and materials [83–85].

Interaction handled by *Force Fields* can be divided in *(i)* bonded interactions and *(ii)* non-bonded interactions.

Bonded Interactions These interactions describe bond stretching, bond angle bending and torsions, which involve 2, 3 and 4 atoms respectively, and are parameterised in the *Force Field* on the basis of these atoms' types. Bonded interactions have often been modelled with harmonic potential functions neglecting cross terms, such as stretch-bend contributions. The harmonic potential is described by Hooke's law for harmonic oscillators:

$$V = \frac{1}{2}\sum_i k_i(x_i - x_i^0)^2 \tag{1.12}$$

and represents Taylor's expansion to the second degree of the potential energy function centred at a minimum in diagonal approximation for:

$$k_i = \left(\frac{\partial^2 V}{\partial x_i^2}\right)_0 \tag{1.13}$$

The *force constants* k_i and the *reference values* x_i^0 in (1.12) are parameters the *Force Field* provides for each type of interaction between each atom type. Addition of higher degree terms to Taylor's expansion of the potential, yields non-harmonic potentials featuring greater accuracy. Approximation can be improved also by considering cross terms, which account for the coupling of interactions, such as bend-bend or

stretch-bend mixed terms.

Non-bonded Interactions These interactions comprise an electrostatic term, a repulsion term and a dispersion term, the last two referring to Van der Waals interactions contribution. The first term is handled with Coulomb's potential, while the Van der Waals interactions terms are commonly modelled with the Lennard-Jones potential:

$$V_{LJ}(r_{ij}) = \frac{C_{ij}^{(12)}}{r_{ij}^{12}} - \frac{C_{ij}^{(6)}}{r_{ij}^{6}} \tag{1.14}$$

Non-bonded interactions produce non-zero contributions to potential at any non-infinite distance, but eventually become negligible within overall approximation. In fact, the most time-consuming step of *MD* simulations is the calculation of non-bonded interactions: to reduce computational load non-bonded interactions are truncated at a defined cut-off. Truncation effectively causes aberrations at the cut-off boundary; therefore methods have been devised to minimise discontinuities in potential, such as switch and shift functions, which act in different ways to achieve the same goal of creating a slower slope of the potential from the value at the cut-off radius to zero. Furthermore, long-range electrostatics (outside cut-off radius) can be handled with fast algorithms as the Ewald Summation, *PME* or *PPPM*.

1.2.2.3 Integration

In a *MD* simulation the force acting on each atom changes whenever the particle changes its position, or whenever any of the other atoms with which it interacts changes its position. Under such circumstances the equations of motion cannot be solved analytically, but have to be integrated numerically using finite difference methods. An integration method developed by Verlet [86], consists in the direct solution of Newton's second law:

$$\mathbf{q}_i(t+\delta t) = 2\mathbf{q}_i(t) - \mathbf{q}_i(t-\delta t) + \delta t^2 \ddot{\mathbf{q}}_i(t) + O\left(\delta t^4\right), \tag{1.15}$$

where $\ddot{\mathbf{q}}_i(t)$ is the acceleration acting on the i-th particle at time t and $O(\delta t^4)$ are the neglected terms of forth and higher orders.

Verlet's algorithm is properly centred, i.e. $\mathbf{q}_i(t-\delta t)$ and $\mathbf{q}_i(t+\delta t)$

play symmetrical roles in (1.15), making it reversible in time, and, given conservative forces, it fulfils the conservation of linear momentum.

The Leap-Frog algorithm [87] is an optimised form of Verlet's algorithm which uses coordinates $\mathbf{q}_i$ at time t and velocities $\dot{\mathbf{q}}_i$ at time $t - \frac{\delta t}{2}$. From the positions, forces $\mathbf{F}_i(t)$ are updated and used to compute $\dot{\mathbf{q}}_i\left(t + \frac{\delta t}{2}\right)$ and $\mathbf{q}_i\,(t + \delta t)$:

$$\dot{\mathbf{q}}_i\left(t + \frac{\delta t}{2}\right) = \dot{\mathbf{q}}_i\left(t - \frac{\delta t}{2}\right) + \ddot{\mathbf{q}}_i(t)\delta t + O(\delta t^2),$$
$$\mathbf{q}_i(t + \delta t) = \mathbf{q}_i(t) + \dot{\mathbf{q}}_i(t + \frac{\delta t}{2})\delta t + O(\delta t^2).$$

The current velocities may be calculated

$$\dot{\mathbf{q}}_i(t) = \frac{1}{2}\left(\dot{\mathbf{q}}_i(t - \frac{\delta t}{2}) + \dot{\mathbf{q}}_i(t + \frac{\delta t}{2})\right).$$

The main advantage of the Leap-Frog algorithm is the explicit treatment of velocities in the calculation of coordinates, minimising the computational loss of precision.

After each integration step the coordinates of the system are recorded as frames of a classical trajectory in phase space, which represents the pathway the system has followed in the hyper-dimensional *Potential Energy Surface*.

1.2.2.4 Statistical mechanics

The thermodynamic properties observed in experiments are actually collective properties of the very large number of atoms present in the examined sample. *MD* simulations on the other hand generate information of a single copy of the analysed system, and therefore the mathematical expressions of statistical mechanics are required to relate the distribution and motion of the atoms and molecules of a system to macroscopic properties.

A system composed of N particles can be described by $3N$ coordinates q_i and $3N$ conjugate momenta p_i (see 1.2.2.1), thus the phase space of a generic system has $6N$ dimensions, and represents all the possible configurations the system can attain. Depending on simulation conditions, the trajectory frames are sampled on a subset of phase-space, the *ensemble*, in which the conditions of the system's *thermodynamic state*

are satisfied. The thermodynamic states of a system are defined by a small set of parameters, and on the basis of these various ensembles have been characterised:

- *Microcanonical (NVE)*: features constant number of particles (N), volume (V) and energy (E);
- *Canonical (NVT)*: features constant number of particles (N), volume (V) and temperature (T);
- *Isobaric-Isothermal (NpT)*: features constant number of particles (N), pressure (p) and temperature (T);
- *Grand canonical (μVT)*: features constant chemical potential (μ), volume (V) and temperature (T);

According to the laws of statistical mechanics, experimental observables may be calculated, for ergodic processes, as time averages from the simulation: in similar calculations all the frames of the trajectory are considered simultaneously and represent the statistical ensemble, so that time averages are also called ensemble averages. Thus it is clear that simulation trajectories can be used to measure both the time-evolution of *dynamical* properties of the system, and macroscopic observables from the entire set of sampled conformations.

1.2.3 MD simulation techniques

To improve significance of simulations, a number of methods were developed to extend the basic concept of *MD*, and giving the possibility to extend sampling potential of algorithms, or to decrease the degree of approximation. Some of these methods, which are now routinely used in productive simulations, are described in the following sections.

1.2.3.1 Boundary conditions

The size of the model systems is small compared to real macroscopic systems. From this point of view a simulated system contains a relatively small number of particles. When simulating an isolated system many of the atoms experience a large boundary surface to a vacuum environment and such a simulation would not be relevant to

study phenomena taking place in the bulk. To avoid this problem, periodic boundary conditions are used. The atoms are placed in a simulation box that is surrounded by translated copies of itself. The inner cell is then surrounded by a periodic 3-dimensional array. When an atom crosses the boundary it is replaced by an image atom that enters from the opposite side. Thus, the number of particles within the central box remains constant.

1.2.3.2 Coupling algorithms

Coupling algorithms are simulation devices which allow to maintain constant values of certain parameters, in order, for example, to maintain the conditions of a certain thermodynamic state which in turn enable correct ensemble sampling. Simulations performed in the *Isobaric-isothermal* ensemble (see 1.2.2.4) require constant values of both temperature and pressure; these are achieved by coupling the simulated system with an external immutable reservoir of heat and pressure, so that changes in the system temperature or pressure are balanced by the external reservoir towards its reference value. Algorithmically this effect can be implemented as extra steps in the integration algorithm (*weak* coupling), or alternatively a term taking into account this effect can be added to the system's Hamiltonian (*extended ensemble* coupling).

GROMACS implements an algorithm of each class for temperature coupling and for pressure coupling: Berendsen's *weak* coupling schemes are available for both parameters, while Nosé-Hoover and Parrinello-Rahman *extended ensemble* algorithms are used for temperature ad pressure respectively.

In this study, pressure and temperature were maintained constant by using Berendsen's algorithms, which work in conceptually similar ways, and having both the characteristic of rapidly obtaining convergence of the coupled properties.

The temperature coupling works by correcting the deviation of instant temperature T from the reference value T_0 according to:

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau}, \tag{1.16}$$

which corresponds to an exponential decay of temperature deviation with time constant τ. Temperature correction is obtained by scaling velocities of the system, and varying the frequency with which this

scaling is performed the strength of the algorithm can be modulated.

The pressure coupling scheme works in a very similar way, correcting deviation of instant pressure **P** from reference value **P** according to:

$$\frac{d\mathbf{P}}{dt} = \frac{\mathbf{P}_0 - \mathbf{P}}{\tau_P}. \tag{1.17}$$

Pressure corrections are executed by isotropically scaling the box vectors and the system coordinates.

1.2.3.3 Constraint algorithms

The dynamic behaviour of any molecule is a complex combination of different motions, featuring different typical frequencies. In order to obtain a reliable description of the simulated system, the integration time-step should be set to values (much) smaller than the period of the highest frequency motion occurring. In biological macromolecules, the bond vibrations are the highest frequency motions, limiting the integration time-step to values below 1 fs. On the other hand the major conformational changes arise from low frequency motions, which are uncoupled with bond vibrations.

Significant gain in simulation speed is achieved by constraining bond lengths so that their vibrations can be completely neglected, allowing to increase integration time step to values up to 2 fs. Algorithms like SHAKE [88] or LINCS [89] are suitable for this purpose. In particular LINCS is a non-iterative two steps algorithm that resets bond lengths after an unconstrained update, and is faster and more stable than SHAKE.

1.3 MD Simulation Analysis

Analysis of *MD* simulations makes use of two different sets of tools and techniques, the first to assess stability and reliability of the simulated trajectories, by computing properties and comparing them with experimental data; the second to analyse the ensemble extracting information on the simulated system's dynamical behaviour or ensemble properties.

1.3.1 Trajectory stability

This first phase of trajectory analysis addresses two main issues: *(i)* stability of the simulation and *(ii)* establishment of equilibrium (convergence).

Stability In this context the term stability refers to the absence of procedural artefacts arising from incorrect parameterisation or modelling of the simulation conditions. Several classes of properties can be monitored to estimate stability; properties of physical nature, such as potential and kinetic energies, volume, pressure, temperature and density, are expected to show low values of fluctuation and total drift. Parameters of conformational nature include the *Radius of Gyration*, which represents a rough estimate of structural compactness, and distances of cofactors from their coordinating residues.

Convergence In the initial phase of every simulation the system is considered to be in a state of non-equilibrium, during which its properties fluctuate markedly or drift while the system adjusts to the simulation conditions: only at the end of this equilibration phase is the system considered to be exhibiting reliably its properties: converged portions of the trajectory are called *equilibrium trajectories*. The most accurate method to assess trajectory convergence is the analysis of the time-dependent evolution of protein main-chain atoms *Root Mean Square Deviation* in respect to the initial conformation (*RMSD*):

$$\mathrm{RMSD}(t) = \left[\frac{1}{\sum_{i=1}^{N} m_i} \sum_{i=1}^{N} m_i \|r_i(t) - r_i(t_0)\|^2 \right]^{\frac{1}{2}}, \tag{1.18}$$

where m_i is mass of atom i and $r_i(t)$ represents the coordinates of atom i at time t. *RMSD* is calculated for each conformation of the trajectory after performing a least-squares fit to the reference structure (the starting conformation).

1.3.2 Trajectory analysis

This second phase of analysis initially addresses the issue of conformational sampling, by comparing each sampled structure with each

other, calculating the pairwise *RMSD* to obtain *RMSD-matrices*, which give insight in the configuration visited by the trajectory. To identify the presence of subgroups featuring marked similarity in the entire ensemble the *RMSD-matrices* are subjected to clustering. With this method similar structures are identified independently of their distance in time in the merged trajectory, leading to important information on the resampling of regions of the phase space.

Subsequently information on structural properties are extracted which may be connected with cold-adaptation, increasing their significance by modifying their values in order to allow easier comparisons, or by organising them in ordered representations. Flexibility is an important issue when analysing cold-adaptation, and protein main-chain atoms *Root Mean Square Fluctuation* with respect to the ensemble's average configuration (*RMSF*) is a well established flexibility index. Reliability of this index can be evaluated by comparing residue-profiles of *RMSF* with the crystallographic temperature factors, or with simulation-computed *Squared Generalised Order Parameters*.

Other potentially interesting parameters are secondary structure content and hydrogen bonding patterns, which give insight in the structural order of the system, and *Solvent Accessible Surface*, which allows the characterisation of the protein's interactions with the environment.

Bibliography

[1] N. J. Russell. Cold adaptation of microorganisms. *Philos Trans R Soc Lond B Biol Sci*, 326(1237):595–608, discussion 608–11, Jan 1990.

[2] R. Margesin and F. Schinner. Biodegradation and bioremediation of hydrocarbons in extreme environments. *Appl Microbiol Biotechnol*, 56(5-6):650–663, Sep 2001.

[3] R. Y. Morita. Psychrophilic bacteria. *Bacteriol Rev*, 39(2):144–167, Jun 1975.

[4] P. W. Mohr and S. Krawiec. Temperature characteristics and Arrhenius plots for nominal psychrophiles, mesophiles and thermophiles. *J Gen Microbiol*, 121(2):311–317, Dec 1980.

[5] N. J. Russell. Toward a molecular understanding of cold activity of enzymes from psychrophiles. *Extremophiles*, 4(2):83–90, Apr 2000.

[6] Jody W Deming. Psychrophiles and polar regions. *Curr Opin Microbiol*, 5(3):301–309, Jun 2002.

[7] D. Georlette, V. Blaise, T. Collins, S. D'Amico, E. Gratia, A. Hoyoux, J-C. Marx, G. Sonan, G. Feller, and C. Gerday. Some like it cold: biocatalysis at low temperatures. *FEMS Microbiol Rev*, 28(1):25–42, Feb 2004.

[8] T. Lonhienne, C. Gerday, and G. Feller. Psychrophilic enzymes: revisiting the thermodynamic parameters of activation may explain local flexibility. *Biochim Biophys Acta*, 1543(1):1–10, Nov 2000.

[9] D. L. Crawford and D. A. Powers. Evolutionary adaptation to different thermal environments via transcriptional regulation. *Mol Biol Evol*, 9(5):806–813, Sep 1992.

[10] J. Baldwin and P. W. Hochachka. Functional significance of isoenzymes in thermal acclimatization. Acetylcholinesterase from trout brain. *Biochem J*, 116(5):883–887, Mar 1970.

[11] G. N. Somero. Proteins and temperature. *Annu Rev Physiol*, 57:43–68, 1995.

[12] G. B. Jagdale and R. Gordon. Effect of temperature on the activities of glucose-6-phosphate dehydrogenase and hexokinase in entomopathogenic nematodes (Nematoda: Steinernematidae). *Comp Biochem Physiol A Physiol*, 118(4):1151–1156, Dec 1997.

[13] Khawar Sohail Siddiqui, Ricardo Cavicchioli, and Torsten Thomas. Thermodynamic activation properties of elongation factor 2 (EF-2) proteins from psychrotolerant and thermophilic Archaea. *Extremophiles*, 6(2):143–150, Apr 2002.

[14] G. Feller. Molecular adaptations to cold in psychrophilic enzymes. *Cell Mol Life Sci*, 60(4):648–662, Apr 2003.

[15] G. Feller and C. Gerday. Psychrophilic enzymes: molecular basis of cold adaptation. *Cell Mol Life Sci*, 53(10):830–841, Oct 1997.

[16] R. Barnes, H. Vogel, and I. Gordon. Temperature of compensation: significance for virus in- activation. *Proc Natl Acad Sci U S A*, 62(1):263–270, Jan 1969.

[17] P. S. Low and G. N. Somero. Temperature adaptation of enzymes: a proposed molecular basis for the different catalytic efficiencies of enzymes from ectotherms and endotherms. *Comp Biochem Physiol B*, 49(2):307–312, Oct 1974.

[18] George N Somero. Protein adaptations to temperature and pressure: complementary roles of adaptive changes in amino acid sequence and internal milieu. *Comp Biochem Physiol B Biochem Mol Biol*, 136(4):577–591, Dec 2003.

[19] G. Feller, T. Lonhienne, C. Deroanne, C. Libioulle, J. Van Beeumen, and C. Gerday. Purification, characterization, and nucleotide sequence of the thermolabile alpha-amylase from the antarctic psychrotroph Alteromonas haloplanctis A23. *J Biol Chem*, 267(8):5217–5221, Mar 1992.

[20] Tony Collins, Marie-Alice Meuwis, Ingeborg Stals, Marc Claeyssens, Georges Feller, and Charles Gerday. A novel family 8 xylanase, functional and physicochemical characterization. *J Biol Chem*, 277(38):35133–35139, Sep 2002.

[21] L. Zecchinon, P. Claverie, T. Collins, S. D'Amico, D. Delille, G. Feller, D. Georlette, E. Gratia, A. Hoyoux, M. A. Meuwis, G. Sonan, and C. Gerday. Did psychrophilic enzymes really win the challenge? *Extremophiles*, 5(5):313–321, Oct 2001.

[22] Khawar Sohail Siddiqui and Ricardo Cavicchioli. Cold-adapted enzymes. *Annu Rev Biochem*, 75:403–433, 2006.

[23] R. Wolfenden and M. J. Snider. The depth of chemical time and the power of enzymes as catalysts. *Acc Chem Res*, 34(12):938–945, Dec 2001.

[24] G. W. Harris, J. A. Jenkins, I. Connerton, and R. W. Pickersgill. Refined crystal structure of the catalytic domain of xylanase A from Pseudomonas fluorescens at 1.8 A resolution. *Acta Crystallogr D Biol Crystallogr*, 52(Pt 2):393–401, Mar 1996.

[25] G. E. Lienhard. Enzymatic catalysis and transition-state theory. *Science*, 180(82):149–154, Apr 1973.

[26] Mark J Snider, Danijela Lazarevic, and Richard Wolfenden. Catalysis by entropic effects: the action of cytidine deaminase on 5,6-dihydrocytidine. *Biochemistry*, 41(12):3925–3930, Mar 2002.

[27] S. S. Licht, C. C. Lawrence, and J. Stubbe. Thermodynamic and kinetic studies on carbon-cobalt bond homolysis by ribonucleoside triphosphate reductase: the importance of entropy in catalysis. *Biochemistry*, 38(4):1234–1242, Jan 1999.

[28] R. B. Loftfield, E. A. Eigner, A. Pastuszyn, T. N. Lövgren, and H. Jakubowski. Conformational changes during enzyme catalysis: role of water in the transition state. *Proc Natl Acad Sci U S A*, 77(6):3374–3378, Jun 1980.

[29] Y. X. Fan, P. McPhie, and E. W. Miles. Regulation of tryptophan synthase by temperature, monovalent cations, and an allosteric ligand. Evidence from Arrhenius plots, absorption spectra, and primary kinetic isotope effects. *Biochemistry*, 39(16):4692–4703, Apr 2000.

[30] Sandeep Kumar, Chung-Jung Tsai, and Ruth Nussinov. Maximal stabilities of reversible two-state proteins. *Biochemistry*, 41(17):5359–5374, Apr 2002.

[31] D. C. Carlow, S. A. Short, and R. Wolfenden. Role of glutamate-104 in generating a transition state analogue inhibitor at the active site of cytidine deaminase. *Biochemistry*, 35(3):948–954, Jan 1996.

[32] Sinisa Bjelic, Bjorn O. Brandsdal, and Johan Aqvist. Cold adaptation of enzyme reaction rates. *BIOCHEMISTRY*, 47(38):10049–10057, SEP 23 2008.

[33] Salvino D'Amico, Paule Claverie, Tony Collins, Daphné Georlette, Emmanuelle Gratia, Anne Hoyoux, Marie-Alice Meuwis, Georges Feller, and Charles Gerday. Molecular basis of cold adaptation. *Philos Trans R Soc Lond B Biol Sci*, 357(1423):917–925, Jul 2002.

[34] S. D'Amico, C. Gerday, and G. Feller. Structural determinants of cold adaptation and stability in a large protein. *J Biol Chem*, 276(28):25791–25796, Jul 2001.

[35] Katrín Gudjónsdóttir and Bjarni Asgeirsson. Effects of replacing active site residues in a cold-active alkaline phosphatase with those found in its mesophilic counterpart from escherichia coli. *FEBS J*, 275(1):117–127, Jan 2008.

[36] R. Jaenicke. Protein stability and molecular adaptation to extreme conditions. *Eur J Biochem*, 202(3):715–728, Dec 1991.

[37] R. Jaenicke. Protein folding: local structures, domains, subunits, and assemblies. *Biochemistry*, 30(13):3147–3161, Apr 1991.

[38] A. Svingor, J. Kardos, I. Hajdú, A. Németh, and P. Závodszky. A better enzyme to cope with cold. Comparative flexibility studies on psychrotrophic, mesophilic, and thermophilic IPMDHs. *J Biol Chem*, 276(30):28121–28125, Jul 2001.

[39] Salvino D'Amico, Charles Gerday, and Georges Feller. Temperature adaptation of proteins: engineering mesophilic-like activity and stability in a cold-adapted alpha-amylase. *J Mol Biol*, 332(5):981–988, Oct 2003.

[40] Daphné Georlette, Benjamin Damien, Vinciane Blaise, Eric Depiereux, Vladimir N Uversky, Charles Gerday, and Georges Feller. Structural and functional adaptations to extreme temperatures in psychrophilic, mesophilic, and thermophilic DNA ligases. *J Biol Chem*, 278(39):37015–37023, Sep 2003.

[41] Tony Collins, Marie-Alice Meuwis, Charles Gerday, and Georges Feller. Activity, stability and flexibility in glycosidases adapted to extreme thermal environments. *J Mol Biol*, 328(2):419–428, Apr 2003.

[42] J. P. Chessa, I. Petrescu, M. Bentahir, J. Van Beeumen, and C. Gerday. Purification, physico-chemical characterization and sequence of a heat labile alkaline metalloprotease isolated from a psychrophilic Pseudomonas species. *Biochim Biophys Acta*, 1479(1-2):265–274, Jun 2000.

[43] R. J. Russell, U. Gerike, M. J. Danson, D. W. Hough, and G. L. Taylor. Structural adaptations of the cold-active citrate synthase from an Antarctic bacterium. *Structure*, 6(3):351–361, Mar 1998.

[44] D. J. Bicout and G. Zaccai. Protein flexibility from the dynamical transition: a force constant analysis. *Biophys J*, 80(3):1115–1123, Mar 2001.

[45] G. Zaccai. How soft is a protein? A protein dynamics force constant measured by neutron scattering. *Science*, 288(5471):1604–1607, Jun 2000.

[46] Moeava Tehei, Bruno Franzetti, Dominique Madern, Margaret Ginzburg, Ben Z Ginzburg, Marie-Thérèse Giudici-Orticoni, Mireille Bruschi, and Giuseppe Zaccai. Adaptation to extreme environments: macromolecular dynamics in bacteria compared in vivo by neutron scattering. *EMBO Rep*, 5(1):66–70, Jan 2004.

[47] B. O. Brandsdal, E. S. Heimstad, I. Sylte, and A. O. Smalas. Comparative molecular dynamics of mesophilic and psychrophilic protein homologues studied by 1.2 ns simulations. *J Biomol Struct Dyn*, 17(3):493–506, Dec 1999.

[48] Elena Papaleo, Laura Riccardi, Chiara Villa, Piercarlo Fantucci, and Luca De Gioia. Flexibility and enzymatic cold-adaptation: a comparative molecular dynamics investigation of the elastase family. *Biochim Biophys Acta*, 1764(8):1397–1406, Aug 2006.

[49] Anita-Elin Fedøy, Nannan Yang, Aurora Martinez, Hanna-Kirsti S Leiros, and Ida Helene Steen. Structural and functional properties of isocitrate dehydrogenase from the psychrophilic bacterium desulfotalea psychrophila reveal a cold-active enzyme with an unusual high thermal stability. *J Mol Biol*, 372(1):130–149, Sep 2007.

[50] R. Chiuri, G. Maiorano, A. Rizzello, L. L. del Mercato, R. Cingolani, R. Rinaldi, M. Maffia, and P. P. Pompa. Exploring local flexibility/rigidity in psychrophilic and mesophilic carbonic anhydrases. *Biophys J*, 96(4):1586–1596, Feb 2009.

[51] K. Miyazaki, P. L. Wintrode, R. A. Grayling, D. N. Rubingh, and F. H. Arnold. Directed evolution study of temperature adaptation in a psychrophilic enzyme. *J Mol Biol*, 297(4):1015–1026, Apr 2000.

[52] P. A. Fields, B. D. Wahlstrand, and G. N. Somero. Intrinsic versus extrinsic stabilization of enzymes: the interaction of solutes

and temperature on A4-lactate dehydrogenase orthologs from warm-adapted and cold-adapted marine fishes. *Eur J Biochem*, 268(16):4497–4505, Aug 2001.

[53] P. L. Privalov. Cold denaturation of proteins. *Crit Rev Biochem Mol Biol*, 25(4):281–305, 1990.

[54] M. Qian, R. Haser, G. Buisson, E. Duée, and F. Payan. The active center of a mammalian alpha-amylase. Structure of the complex of a pancreatic alpha-amylase with a carbohydrate inhibitor refined to 2.2-A resolution. *Biochemistry*, 33(20):6284–6294, May 1994.

[55] H. Outzen, G. I. Berglund, A. O. Smalas, and N. P. Willassen. Temperature and pH sensitivity of trypsins from Atlantic salmon (Salmo salar) in comparison with bovine and porcine trypsin. *Comp Biochem Physiol B Biochem Mol Biol*, 115(1):33–45, Sep 1996.

[56] A. Greener, M. Callahan, and B. Jerpseth. An efficient random mutagenesis technique using an E. coli mutator strain. *Mol Biotechnol*, 7(2):189–195, Apr 1997.

[57] M. Roovers, R. Sanchez, C. Legrain, and N. Glansdorff. Experimental evolution of enzyme temperature activity profile: selection in vivo and characterization of low-temperature-adapted mutants of Pyrococcus furiosus ornithine carbamoyltransferase. *J Bacteriol*, 183(3):1101–1105, Feb 2001.

[58] Pietro Gatti-Lafranconi, Serena M Caldarazzo, Alessandra Villa, Lilia Alberghina, and Marina Lotti. Unscrambling thermal stability and temperature adaptation in evolved variants of a cold-active lipase. *FEBS Lett*, 582(15):2313–2318, Jun 2008.

[59] Anna Gudný Sigurdardóttir, Jóhanna Arnórsdóttir, Sigrídur H Thorbjarnardóttir, Gudmundur Eggertsson, Karsten Suhre, and Magnús M Kristjánsson. Characteristics of mutants designed to incorporate a new ion pair into the structure of a cold adapted subtilisin-like serine proteinase. *Biochim Biophys Acta*, 1794(3):512–518, Mar 2009.

[60] R. Sterner, G. R. Kleemann, H. Szadkowski, A. Lustig, M. Hennig, and K. Kirschner. Phosphoribosyl anthranilate isomerase from

Thermotoga maritima is an extremely stable and active homodimer. *Protein Sci*, 5(10):2000–2008, Oct 1996.

[61] S. Kumar, B. Ma, C. J. Tsai, N. Sinha, and R. Nussinov. Folding and binding cascades: dynamic landscapes and population shifts. *Protein Sci*, 9(1):10–19, Jan 2000.

[62] B. Ma, S. Kumar, C. J. Tsai, Z. Hu, and R. Nussinov. Transition-state ensemble in enzyme catalysis: possibility, reality, or necessity? *J Theor Biol*, 203(4):383–397, Apr 2000.

[63] George N Somero. Adaptation of enzymes to temperature: searching for basic "strategies". *Comp Biochem Physiol B Biochem Mol Biol*, 139(3):321–333, Nov 2004.

[64] Q. Xue and E. S. Yeung. Differences in the chemical reactivity of individual molecules of an enzyme. *Nature*, 373(6516):681–683, Feb 1995.

[65] D. Craig, E. A. Arriaga, P. Banks, Y. Zhang, A. Renborg, M. M. Palcic, and N. J. Dovichi. Fluorescence-based enzymatic assay by capillary electrophoresis laser-induced fluorescence detection for the determination of a few beta-galactosidase molecules. *Anal Biochem*, 226(1):147–153, Mar 1995.

[66] H. P. Lu, L. Xun, and X. S. Xie. Single-molecule enzymatic dynamics. *Science*, 282(5395):1877–1882, Dec 1998.

[67] Ricardo Cavicchioli, Khawar S Siddiqui, David Andrews, and Kevin R Sowers. Low-temperature extremophiles and their applications. *Curr Opin Biotechnol*, 13(3):253–261, Jun 2002.

[68] G. Gianese, F. Bossa, and S. Pascarella. Comparative structural analysis of psychrophilic and meso- and thermophilic enzymes. *Proteins: Structure Function and Genetics*, 47(2):236–249, 2002.

[69] A. O. Smalas, H. K. Leiros, V. Os, and N. P. Willassen. Cold adapted enzymes. *Biotechnol Annu Rev*, 6:1–57, 2000.

[70] B. O. Brandsdal, A. O. Smalas, and J. Aqvist. Electrostatic effects play a central role in cold adaptation of trypsin. *FEBS Lett*, 499(1-2):171–175, Jun 2001.

[71] I. Tsigos, K. Velonia, I. Smonou, and V. Bouriotis. Purification and characterization of an alcohol dehydrogenase from the Antarctic psychrophile Moraxella sp. TAE123. *Eur J Biochem*, 254(2):356–362, Jun 1998.

[72] Raghu Prasad Rao Metpally and Boojala Vijay B. Reddy. Comparative proteome analysis of psychrophilic versus mesophilic bacterial species: Insights into the molecular basis of cold adaptation of proteins. *BMC GENOMICS*, 10, JAN 8 2009.

[73] SY Kim, KY Hwang, SH Kim, HC Sung, YS Han, and YJ Cho. Structural basis for cold adaptation - Sequence, biochemical properties, and crystal structure of malate dehydrogenase from a psychrophile Aquaspirillium arcticum. *JOURNAL OF BIOLOGICAL CHEMISTRY*, 274(17):11761–11767, APR 23 1999.

[74] Daniele Tronelli, Elisa Maugini, Francesco Bossa, and Stefano Pascarella. Structural adaptation to low temperatures - analysis of the subunit interface of oligomeric psychrophilic enzymes. *FEBS JOURNAL*, 274(17):4595–4608, SEP 2007.

[75] Alessandro Siglioccolo, Francesco Bossa, and Stefano Pascarella. structural adaptation of serine hydroxymethyltransferase to low temperatures. *Int J Biol Macromol*, Oct 2009.

[76] J. C. Marx, V. Blaise, T. Collins, S. D'Amico, D. Delille, E. Gratia, A. Hoyoux, A. L. Huston, G. Sonan, G. Feller, and C. Gerday. A perspective on cold enzymes: current knowledge and frequently asked questions. *Cell Mol Biol (Noisy-le-grand)*, 50(5):643–655, Jul 2004.

[77] Bertus van den Burg. Extremophiles as a source for novel enzymes. *Curr Opin Microbiol*, 6(3):213–218, Jun 2003.

[78] Babu Joseph, Pramod W Ramteke, and George Thomas. Cold active microbial lipases: some hot issues and recent developments. *Biotechnol Adv*, 26(5):457–470, 2008.

[79] C. Gerday, M. Aittaleb, M. Bentahir, J. P. Chessa, P. Claverie, T. Collins, S. D'Amico, J. Dumont, G. Garsoux, D. Georlette,

A. Hoyoux, T. Lonhienne, M. A. Meuwis, and G. Feller. Cold-adapted enzymes: from fundamentals to biotechnology. *Trends Biothecnol*, 18(3):103–107, 2000.

[80] Y. Bai, T. R. Sosnick, L. Mayne, and S. W. Englander. Protein folding intermediates: native-state hydrogen exchange. *Science*, 269(5221):192–197, Jul 1995.

[81] J. O. Wooll, J. O. Wrabl, and V. J. Hilser. Ensemble modulation as an origin of denaturant-independent hydrogen exchange in proteins. *J Mol Biol*, 301(2):247–256, Aug 2000.

[82] B. Ma, S. Kumar, C. J. Tsai, and R. Nussinov. Folding funnels and binding mechanisms. *Protein Eng*, 12(9):713–720, Sep 1999.

[83] M. Karplus and J. Kuriyan. Molecular dynamics and protein function. *Proc Natl Acad Sci U S A*, 102(19):6679–6685, May 2005.

[84] H. Goldstein. *Classical Mechanics*. Addison-Wesley, Reading, Massachusetts, 1950.

[85] Landau L. and Lifshitz E. *Mechanics*. Pergamon Press, 1961.

[86] L. Verlet. Computer 'experiments' on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Phys Rev*, 159:98–103, 1967.

[87] R. W. Hockney. The potential calculation and some applications. *Methods comput Phys*, 9:136–211, 1970.

[88] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints - molecular dynamics of n-alkanes. *J Comp Phys*, 23(3):327–341, 1977.

[89] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. fraaije. LINCS: A linear constraint solver for molecular simulations. *J Comput Chem*, 18(12):1463–1472, 1997.

Chapter 2

Dynamic Properties of a Cold-active α-amylase

2.1 Introduction

The cold-active α-amylase from the Antarctic bacterium *Pseudoalteromonas haloplanktis* (namely AHA) is one of the best characterised psychrophilic enzymes. Its biochemical properties and X-ray structure have been extensively studied in the context of enzyme catalysis at low temperatures [1–4]. In particular, AHA was one of the first cold-adapted enzymes for which the X-ray structure was solved [4], revealing that the enzyme consists of three domains (Fig.2.1). Domain A is at the N-terminus and is composed of a $(\beta/\alpha)_8$-barrel forming one side of the catalytic cleft, domain B protrudes from domain A and forms the other side of the active cleft, and domain C is the C-terminal domain. A Ca^{2+} ion is bound between domains A and B and a Cl^- ion which is necessary for catalysis is bound close to the active site [5]. The reaction pathway of enzymes from the α-amylase family retains the anomeric configuration and has been demonstrated in recent years [5]. It involves a double displacement mechanism with the formation of a covalent glycosyl-enzyme intermediate. The enzyme is a glycosylhydrolase and features two acid catalytic residues: the proton donor (Glu200), the catalytic nucleophile (Asp174) and a second conserved aspartic acid residue (Asp264), which possibly stabilises the protonated state of the gluatamic acid side-chain.

This psychrophilic enzyme has evolved low conformational stabil-

ity and high flexibility through a reduction in the number and type of inter- and intra-molecular interactions [2, 6] and the enhanced flexibility has resulted in decreased enthalpy of activation with a concomitant improvement in k_{cat} [7, 8]. Site-directed mutagenesis studies of AHA have revealed a global weakening of intra-molecular interactions leading to an overall decrease in thermostability [7–9]. AHA is the largest multi-domain protein (~50 kDa) that has been shown to unfold/refold reversibly by a cooperative two-state mechanism [8] (see Fig.1.4). Recently, with transverse urea gradient gel electrophoresis (*TUG-GE*) two unfolding transitions were identified, demonstrating that in contrast to thermal unfolding, urea-induced unfolding is a three-state process [10]. The stability data support a model of cooperative unfolding of structures forming the active site and independent unfolding of the other more stable protein domains.

The X-ray structure of AHA is similar to that of α-amylases from mammalian species (e.g. Pig Pancreatic α-amylase, PPA) and may have been acquired in different organisms through horizontal gene transfer. As it is typical for mesophilic counterparts of cold-adapted enzymes, the pig enzyme is inherently more stable than AHA [11]. Structural differences have been highlighted between the two enzymes: AHA shows fewer intramolecular interactions, especially salt bridges and interactions mediated by arginine, reduced compactness of the hydrophobic core and of inter-domain interfaces, incresed hydrophobicity of the solvent accessible surface, and a different architectures of the ion-binding sites, resulting in lower values for the Ca^{2+} and Cl^{-} binding constants [2, 3]. Furthermore, loops surrounding the active site are shorter in AHA than in PPA, possibly contributing to different accessibility of the substrate-binding cleft [12].

A study of the overall flexibility of AHA has been recently presented in the context of 1 ns MD simulations of several psychrophilic enzymes [13]. This study provided the first evidence of the high flexibility of the insertion loops in PPA, i.e., the loops bordering the active site that are longer in the mesophilic enzyme than in AHA, but failed to find significant differences in flexibility in the conserved regions. However, the authors concluded that longer trajectories are necessary in order to provide a better statistical insight and elucidation of structural flexibility of selected enzymes. In fact, it is well known that the analysis of long multiple trajectories helps in the identification of recurring

features and can allow to avoid artifacts arising from the simulation procedure [14, 15].

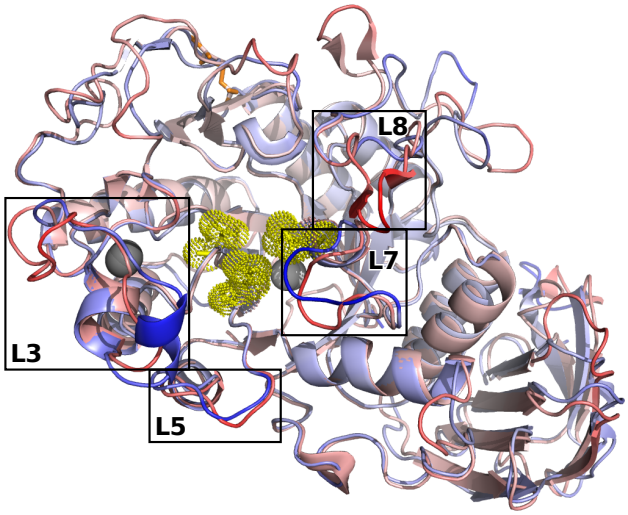
The aim of this study is to better elucidate structural and molecular features of the cold-adapted α-amylase in comparison with its mesophilic homologue. MD is a suitable tool to evaluate flexibility and correlate flexibility with protein structure and function [16–21] In addition, MD simulations can be used to rationalize experimental results, allowing the interpolation or extrapolation of empirical data into regions hardly accessible otherwise [19, 20, 22].

Given the importance of the simulation time scale and conformational sampling [14, 20], several MD simulations of the psychrophilic and mesophilic enzymes were carried out in explicit solvent at room temperature (298 K), allowing the collection of more than 60 ns of explicit-solvent MD simulations on the two system. These trajectories have been analyzed in terms of secondary structure content, solvent accessibility, intramolecular interactions and molecular flexibility, revealing a striking degree of similarity of the dynamical properties of the enzymes; comparison of flexibility of the active-site region allows us to suggest possible adaptive traits and to identify concerted motions in the active site of the two enzymes which may be linked to catalysis.

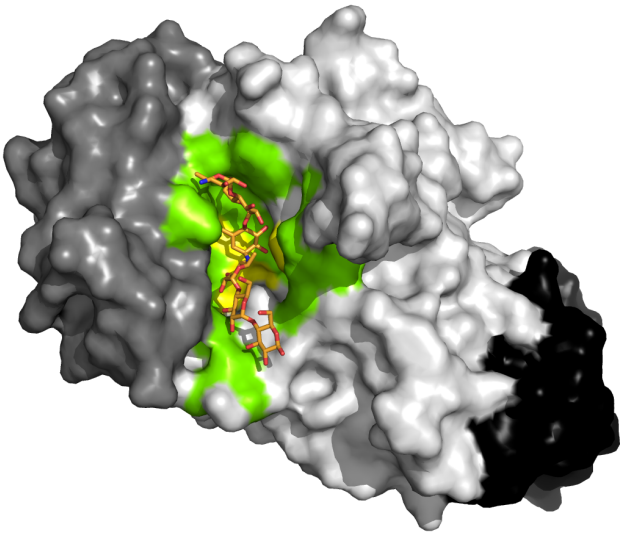
2.2 Methods

2.2.1 Molecular dynamics simulations

MD simulations, and some analyses, were performed using a modified version[23, 24] of the GROMOS-87[25] united-atom force-field with the GROMACS software version 3.2 (www.gromacs.org). The X-ray structures of native porcine (PPA) and *Pseudoalteromonas haloplanktis* (AHA) α-amylase (PDB entries 1PIF [26] and 1AQH [4], respectively) were used as the starting point for the simulations. After a careful visual inspection of the structures, minor corrections were made in order to obtain a perfect match with the primary sequence of the enzymes, available as SwissProt entries P00690 and P29957, respectively. In particular, in 1AQH the missing side-chain of E118 was restored by *in-silico* mutagenesis with PyMOL (www.pymol.org), while in 1PIF the N-terminal pyroglutamic acid, which is due to a physiological post-translational modification [27], was replaced with a proline residue,



(a)



(b)

Figure 2.1: (a) Cartoon representation of the crystal structures of PPA and AHA, superimposed and colored in shades of red and blue (respectively), with color intensity proportional to flexibility; (b) surface representation of PPA, showing a substrate (from crystal structure 1PIG after superposition of the two proteins); the substrate-binding groove is colored green, while the three domains are colored white (domain A), gray (domain B) and black (domain C). The chloride (white) and calcium (gray) ions are shown as spheres, the C70-C115$^{\mathbf{P}}$ disulphide bridge, missing in AHA is shown as orange sticks and the catalytic triad is shown as yellow dots and sticks.

which is structurally very similar and has the advantage of being more suitable to being treated by the force-field.

Protein structures, including crystallographic water molecules and the calcium ion, were energy minimized in vacuum using the Steepest Descent (SD) method (100 steps), and subsequently soaked in a dodecahedral box of SPC water molecules [28] so that all protein atoms are at a distance equal or greater than 0.5 nm from the box edges.

The ionization state of charged residues was set to be consistent with neutral pH: Lys and Arg residues were positively charged, whereas Asp and Glu were negatively charged. The tautomeric form of histidine residues was derived using GROMACS tools (www.gromacs.org) and confirmed by visual inspection of the molecular environment of each histidine. With the above assumptions, both AHA and PPA are characterized by negative overall charge, and in order to neutralize the overall charge of the system, a number of water molecules equal to the protein net charge was replaced by sodium ions.

Initially, solvent molecules were relaxed by molecular mechanics (SD method, 1000 steps). The optimization step was followed by 35 ps MD at 298 K (time step 0.001 ps) while restraining protein atomic positions using a harmonic potential. During equilibration, the coupling constant to the external bath [29] was set to 0.001 ps.

In the attempt to obtain shorter initial equilibration periods of the productive simulation trajectories, the system was slowly driven to the target temperature (298 K) and pressure (1 bar) through a series of short equilibration simulations (thermalization and pressurization steps).

MD simulations were performed in the NPT ensemble at 298 K, applying periodic boundary conditions and using an external bath[29] with a coupling constant of 0.1 ps. Pressure was kept constant (1 bar) by scaling the box vectors and the time-constant for pressure coupling was set to 0.5 ps. Periodic boundary conditions were used with a box size of 534.8 nm^2 and 539.0 nm^2, containing 15090 and 15224 SPC[28] water molecules for AHA and PPA respectively. The LINCS algorithm [30] was used to constrain bond lengths, allowing the use of a 2 fs time step. Electrostatic interactions were calculated using the Particle-mesh-Ewald (PME) summation scheme [31], with a cutoff of 0.12 nm for the separation of the direct and reciprocal space sum cutoffs. Van der Waals and Coulomb interactions were truncated at 0.8 nm. The non-bonded pair list was updated every 10 steps and conformations were stored

every 2 ps.

To improve conformational sampling, six independent simulations (*replicas*) were carried out for each protein system, initializing the MD runs with different sets of initial atomic velocities taken from a Maxwellian distribution. All *replicas* have been conducted for 6 ns, except trajectories showing the slowest convergence times (above 2.5 ns, *replicas* 2, 5 and 6 of AHA), which have been elongated to 11 ns.

Extending the set on which ensemble average properties are calculated increases the statistical significance of the computed values, and the probability that identified features are not simulation artifacts, but actual recurrent characteristics of the system.

2.2.2 Stability of MD trajectories

Stability of the computed trajectories was evaluated using different criteria, in order to identify equilibrium portions which were subsequently concatenated to compute ensemble properties. In particular, the time-evolution of energetic parameters was found to be stable throughout the entire simulation time in all the trajectories, constantly fluctuating around an average value (not shown).

The root mean square deviation (rmsd), which is a crucial parameter to evaluate the stability of MD trajectories, was computed for main-chain atoms with respect to the starting structure of the simulations, and its time-evolution (Fig.2.2) was monitored along with other relevant structural properties, such as the protein radius of gyration and the distance between the calcium and chloride ions and their coordinating residues.

For each protein, the equilibrated portions of the six *replicas* were joined to obtain a concatenated metatrajectory, which is representative of different directions of sampling around the starting structures. PPA trajectories were considered to have established equilibrium between 0.2 ns and 2.0 ns, whereas AHA trajectories generally required longer equilibration times (0.9 ns to 4.0 ns). After concatenating the equilibrium portions of the trajectories, the resulting analysis ensembles consisted of 36.50 ns and 29.74 ns for the psychrophilic and mesophilic α-amylases, respectively.

The extent of conformational sampling was investigated by computing rmsd matrices, in which the rmsd values computed comparing

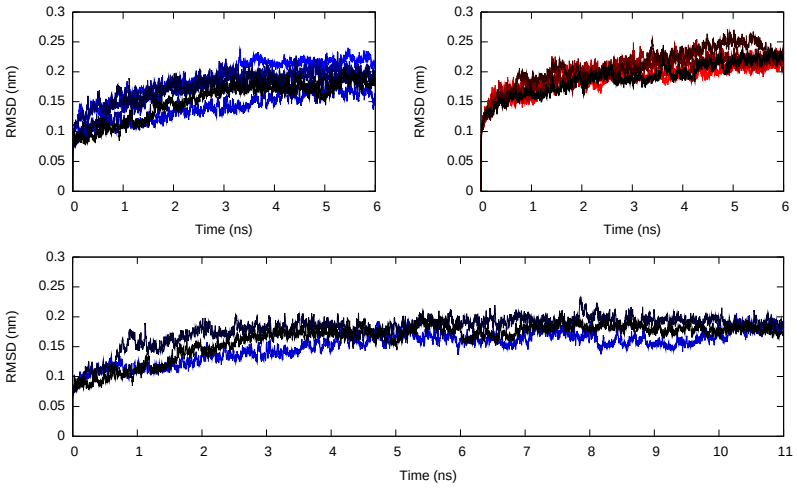


Figure 2.2: Convergence of trajectories: time evolution of main chain rmsd of *replicas* of AHA (6 ns-long, top left, 11 ns-long, bottom) and PPA (top right). *Replica-1, 2, 3, 4, 5, 6* are colored with different shades of blue (AHA) and red (PPA).

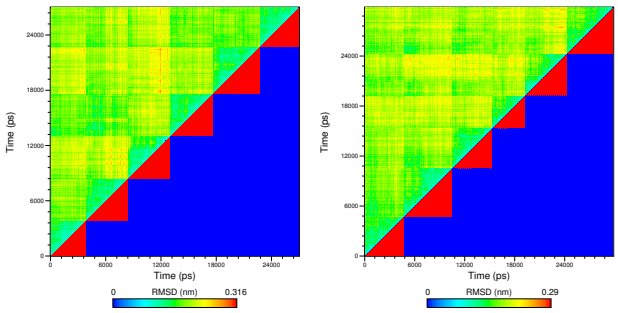


Figure 2.3: Cluster analysis: rmsd and clustering matrices of AHA (left) and PPA (right), using 6 ns-long *replicas*. The top-left half of the matrices shows pairwise main chain rmsd, while the bottom-right half graphically represent clustering results: red triangles indicate structures clustered together.

all possible pairs of frames are organized in a two-dimensional map which allows an evaluation of the re-sampling of similar substructures, were computed on the concatenated trajectories; rmsd matrices were processed using the Linkage clustering algorithm (www.gromacs.org): conformations from each *replica* cluster separately (2.3), thus showing that initializing the simulations with different atomic velocities has caused the systems to converge to different basins, widening the available sampling of the conformational space.

2.2.3 Molecular and structural properties

The secondary structure content was calculated for all stored conformations using the DSSP program, which defines secondary structural elements based on hydrogen-bonding patterns and geometrical features [32]. Comparison of the secondary structure patterns of the two enzymes was performed using the ensemble averages divided by the number of amino-acids considered, for the whole enzyme as well as for single protein domains (Tab.2.1).

To take into account the dynamic properties of secondary structures, their time-evolution was analyzed in detail by evaluating the most frequently attained secondary structure for each residue in order to obtain a residue-dependent persistence degree of secondary structure profile (PDSSP). In order to better identify the differences in secondary structure persistence between the mesophilic and psychrophilic α-amylases, the PDSSPs were mapped on the 3D-structures using a color code.

Intramolecular hydrogen bonds (H-bonds) were evaluated using both a donor-acceptor distance cutoff ($r_{DA} \leq 0.35\,\mathrm{nm}$) and acceptor-donor-hydrogen angular cut-off ($\alpha_{ADH} \leq 30\,^{\circ}$); ensemble averages were normalized using the total number of donors and acceptors.

Surrounding hydrophobicity (H_p) was computed for each residue as:

$$H_\mathrm{p}(i) = \sum_{j \neq i}^{N} H(r_d - r_{ij}) h_j$$

where the summation runs over the whole protein, r_{ij} is the distance between the C_α atoms of residues i and j, $H(x)$ is the Heaviside step function [$H(x) = 1$ if $x \geq 0$, and zero otherwise], r_d is the distance cut-off and h_j is the hydrophobicity index[33] of residue j in $kcal\ mol^{-1}$. Residue-profiles were obtained as ensemble averages. The distance cut-off was set to 0.8 nm, previously identified ([34] and references therein) as sufficient to characterize the hydrophobic behavior and both local and non-local interactions of amino-acids, and has been used in several other studies.

Relative solvent accessible surface (SAS) was calculated for each stored conformation using the NACCESS program [35].

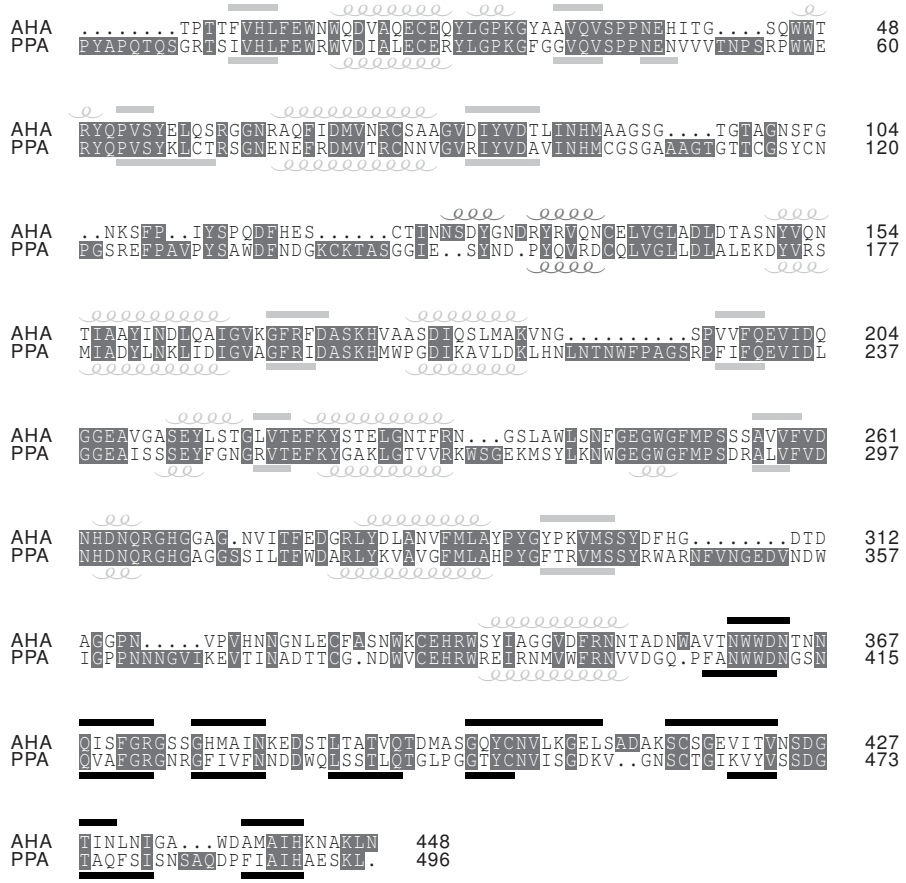


Figure 2.4: Structural alignment between PPA and AHA. The alignment was created using the TEXshade package[37].

2.2.4 Comparison between simulations

Residue-based comparison of the two proteins was carried out by matching residues according to a structural alignment between the crystal structures of the two enzymes performed with the DALI algorithm [36], and then integrated with available structural and functional data (Fig.2.4). In particular the correct matching of the catalytic triad, the residues binding the ionic cofactors and those contributing to disulphide bridges was taken into account.

Amino-acid numbering in the text is labeled $^{\mathbf{A}}$ or $^{\mathbf{P}}$ referring to the

sequence of AHA or PPA, respectively. Domains A, B and C of AHA are composed of residues $1-86/147-356^{\mathbf{A}}$, $87-146^{\mathbf{A}}$ and $357-448^{\mathbf{A}}$, respectively, whereas domains A, B and C of PPA are composed of residues $1-98/170-404^{\mathbf{P}}$, $99-169^{\mathbf{P}}$ and $405-496^{\mathbf{P}}$.

The visual analysis of protein structures was carried out using PyMOL (http://www.pymol.org).

2.2.5 Flexibility and dynamics

The root mean square fluctuation (RMSF) per residue was calculated on C_α atoms, after projection of the trajectories on the *essential subspace* (see on). The RMSF profiles were calculated on the whole concatenated trajectories (*long-time* RMSF profiles) or as the average of non-overlapping time-windows of 100 ps (*short-time* RMSF profiles). Flexibility of matching residues was compared by computing the Pearson's correlation coefficient (r) and the slope of the linear regression line (m) considering only matching residues in the alignment (excluding gaps).

Generalised order parameters (S^2) which are a measure of the degree of spatial restriction of motion have been calculated for the main-chain N-H bonds from the equilibrium trajectories as an approximation of the asymptotic value of the bond rotational autocorrelation function [38, 39], according to:

$$S^2 = \lim_{t\to\infty} C_{\mathrm{I}}(t);$$

where $C_{\mathrm{I}}(t)$ is defined as:

$$C_{\mathrm{I}}(t) = \langle P_2(\mu(\tau)\cdot\mu(\tau+t))\rangle.$$

$\mu(t)$ is the unit vector that describes the orientation of the N-H bond at time t, measured in the molecule-fixed frame, and $P_2(x)$ is the second Legendre polynomial. S^2 can take values ranging from 0, corresponding to completely isotropic motions, to 1, if the motions are completely restricted.

The N-H bond autocorrelation function $C_{\mathrm{I}}(t)$ was computed for each residue (with the exception of the N-terminal and proline residues). In particular, S^2 order parameters were calculated as the average value of $C_{\mathrm{I}}(t)$ starting from $t = 1$ ns. To evaluate the signifi-

cance of the obtained order parameters, the drift of the portions of $C_I(t)$ used to calculated S^2 was estimated by computing the skewness of its distribution. The generalised order parameters were calculated on the equilibrated portions of the *replicas*, and then averaged for each residue.

Residue-dependent C_α B-factor profiles were obtained extracting the atomic B-factors of the C_α atoms for each residue from the PDB files.

Correlation plots were obtained by first computing correlation matrices[40] $\hat{C}_{i,j}$ for C_α atoms, using averaging windows of 1 ns. Correlations were then plotted on the three-dimensional structures by connecting atoms i and j with lines, which thickness is proportional to $\hat{C}_{i,j}$. Since no negative values of $\hat{C}_{i,j}$ were found in the analyses, only the most significant ($\hat{C}_{i,j} > 0.4$ nm) long-range ($|i-j| > 12$) positive correlations are shown.

Backbone dihedral rotational freedom was evaluated by measuring the interquantile range between the 5^{th} and 95^{th} percentiles of the ensemble distribution of ϕ and ψ values for each considered residue, after normalizing the angles to fit in the $0°$-$360°$ range. This procedure was chosen in order both to obtain a significant estimation of the greater part of the statistical dispersion (90% of the sample is taken into account) and to disregard outliers.

Measurements of χ_1 on crystal structures were performed using the PDB module of the Biopython library. The available structures of AHA and PPA were classified as free or bound based on visual inspection of the active sites of each crystal structure: the presence of a substrate, transition-state analogue or inhibitor bound in the substrate-binding groove was sufficient to define the structure as bound.

2.2.6 Essential dynamics analysis

The essential dynamics (ED) analysis reveals high-amplitude concerted motions in the simulated trajectories, based on the calculation of the eigenvectors of the covariance matrix of the atomic positional fluctuations. The covariance matrix ($C_{i,j}$) was calculated for protein C_α atom and considering the equilibrated portion of the trajectories: this has been shown to give a correct picture of the essential subspace [41]. ED analysis was performed for all C_α atoms, and for C_α atoms of regions $259-282^{\mathbf{A}}$ and $295-319^{\mathbf{P}}$.

After removal of the translational and rotational degrees of freedom, C was constructed and then diagonalized to obtain eigenvectors and eigenvalues, which give information on correlated motions throughout the protein. By ordering the eigenvectors according to decreasing values of the eigenvalues, the most relevant eigenvectors can be detected.

ED analysis provided further information on conformational sampling. The number of eigenvectors required to obtain the description of 70% of the motion is 12 and 15 for AHA and PPA, respectively; the corresponding subspace is termed *essential subspace*, and is sufficient to describe protein motions at a reasonable level of accuracy [41]. Motions described by the *essential subspace* have been mapped on average 3D-structures by representing with ellipsoids (at probability 0.1) the computed anisotropic temperature factors obtained from the per-residue, C_α anisotropic U-tensors of cross-correlations of motion along the first 15 and 12 eigenvectors for AHA and PPA, respectively. The average 3D-structure is defined as the sampled conformation with the least distance (computed as main-chain rmsd) from the algebraic ensemble average of atomic positions.

Generally, the first three eigenvectors describe a consistent part of the motion and can be used as a reference subspace to analyze protein dynamics. Therefore, the projection of simulation frames in this reference subspace (2.5) was analyzed, showing a good re-sampling of similar conformations and indicating that the essential subspace is relatively well-sampled for both AHA and PPA when concatenated trajectories are considered.

The cosine content c_i of the ith principal component (p_i) was calculated according to the equation:

$$c_i = \frac{2}{T}\left(\int cos\,(i\pi t)\,p_i(t)dt\right)^2\left(\int p_i^2(t)dt\right)^{-1} \tag{2.1}$$

where T is the time of simulation. c_i is a negative index of the similarity between the dynamics of the system and those of a random diffusion process [42]; c_i is an absolute measure, which can be extracted from one covariance analysis and can take values between 0 (no cosine) and 1 (a perfect cosine). It has been demonstrated that insufficient sampling can lead to a behavior that resemble a functional motion, but actually

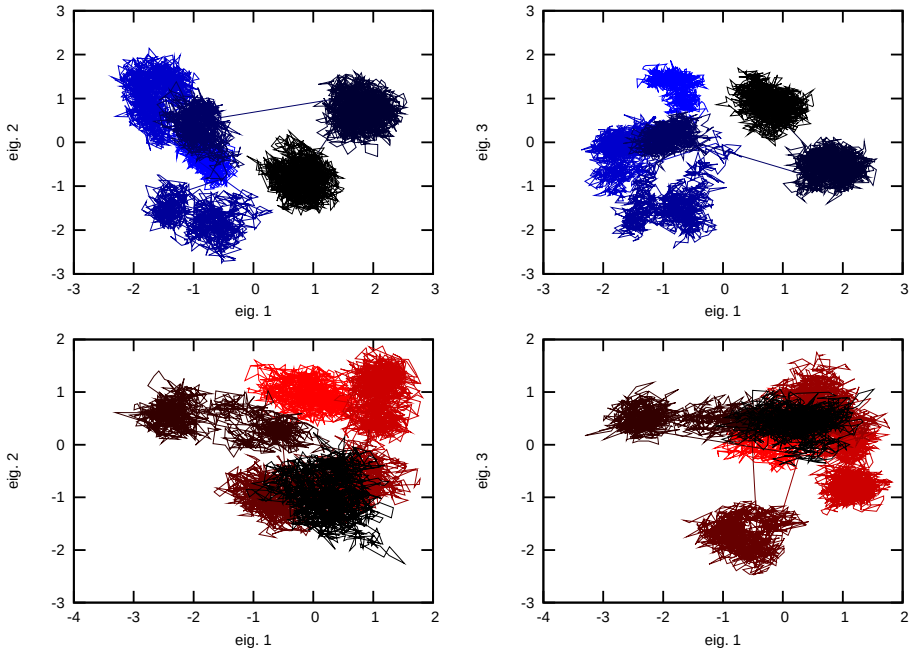


Figure 2.5: Projection of the simulation frames in the 3D-subspace formed by the first three eigenvectors for AHA (top) and PPA (bottom). The equilibrium conformations relative to each *replica* are shown with different shades of blue (AHA) and red (PPA). Displacements are shown in nm.

describes a random motion. The evaluation of cosine contribution for the first eigendirections is sufficient to give a reliable idea of the protein behavior [42].

Calculation of c_i on the principal components of protein motion (2.6), shows that single simulations are often characterized by high cosine content on the first eigenvectors, while the corresponding concatenated trajectories have lower or almost null values of c_i and therefore adequately represent significant, non-diffusive motions.

Similarity between trajectories has been computed as the root mean square inner product (*rmsip*) of the first 10 eigenvectors, ordered by decreasing eigenvalues, obtained diagonalizing the covariance matrix calculated from each of the two trajectories, according to:

$$RMSIP = \left(\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\eta_i \cdot \nu_j \right)^2 \right)^{\frac{1}{2}} \tag{2.2}$$

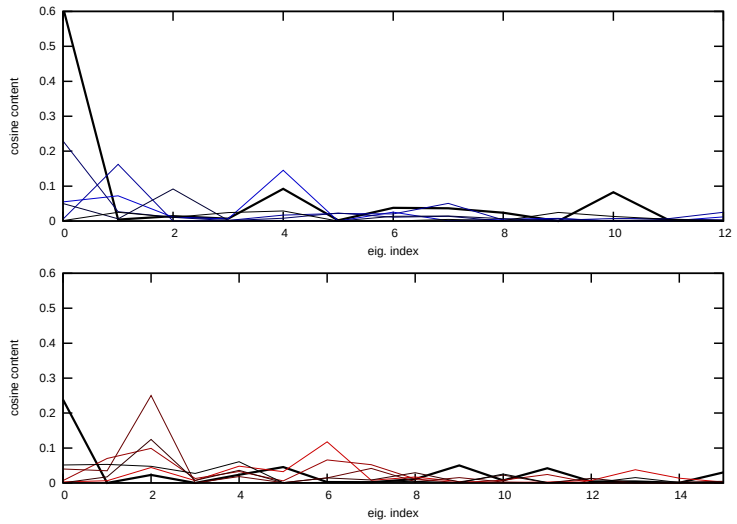


Figure 2.6: Cosine content of the first eigenvectors for AHA (top) and PPA (bottom) simulations. The cosine content calculated for the single *replicas* and the concatenated trajectories are shown as colored lines and thick black lines, respectively. The cosine content of the first 12 and 15 eigenvectors are shown for AHA and PPA, respectively.

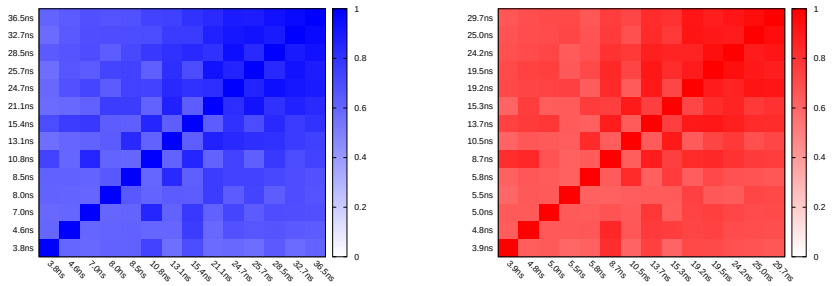


Figure 2.7: Overlap between portions of the metatrajectory of AHA (left) and PPA (right), computed as the RMSIP between the subspaces described by the first 10 eigenvectors of the covariance matrix computed on each portion. The 14 portions correspond to (in order): 5 single replicas, 2 replica pairs, 2 replica triplets, 2 replica quadruplets, 2 replica quintuplets and total metatrajectory.

where η_i and v_j are the ith eigenvector relative to the first trajectory and the jth eigenvector relative to the second trajectory and $N = 10$.

Convergence of the dynamical properties of the simulated enzymes between *replicas* is a fundamental issue when performing ED analysis. This issue has been addressed by plotting the projections of *replicas* on

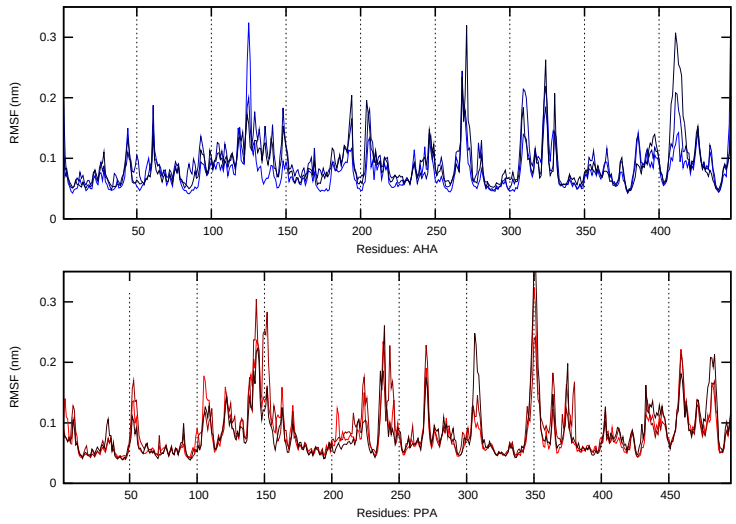


Figure 2.8: Comparison of RMSF profiles computed on the equilibrated portions of the single *replicas*, for AHA (top) and PPA (bottom); the profiles are colored from black (first *replica*) to blue or red (last *replica*, AHA and PPA respectively) .

the first three eigenvectors of the ED analysis on the concatenated metatrajectory (Fig.2.5), by measuring per-*replica* RMSF in the *essential subspace* (Fig.2.8) and by computing the similarity between the subspace described by the first 10 eigenvectors obtained performing ED analysis on portions of the metatrajectory of different lengths, generated by combining different sets of *replicas* (Fig.2.7). Single *replicas* span different, partially overlapping regions of the conformational space, indicating good sampling of the essential subspace (Fig.2.5), and display similar fluctuation intensities (Fig.2.8) that correlate well with the picture obtained from the full metatrajectory (Fig.2.12). Nevertheless at least 20-25 ns and 15-20 ns for AHA and PPA respectively (i.e. the combination of 3 to 4 *replicas*) are needed to reach a reasonable degree of convergence of the directions of principal motions (Fig.2.7).

2.3 Results and Discussion

2.3.1 Overall structural properties

Secondary structure plays a central role in the determination of protein architecture and function, therefore only minor changes are expected to be tolerated in the secondary structure of homologous pro-

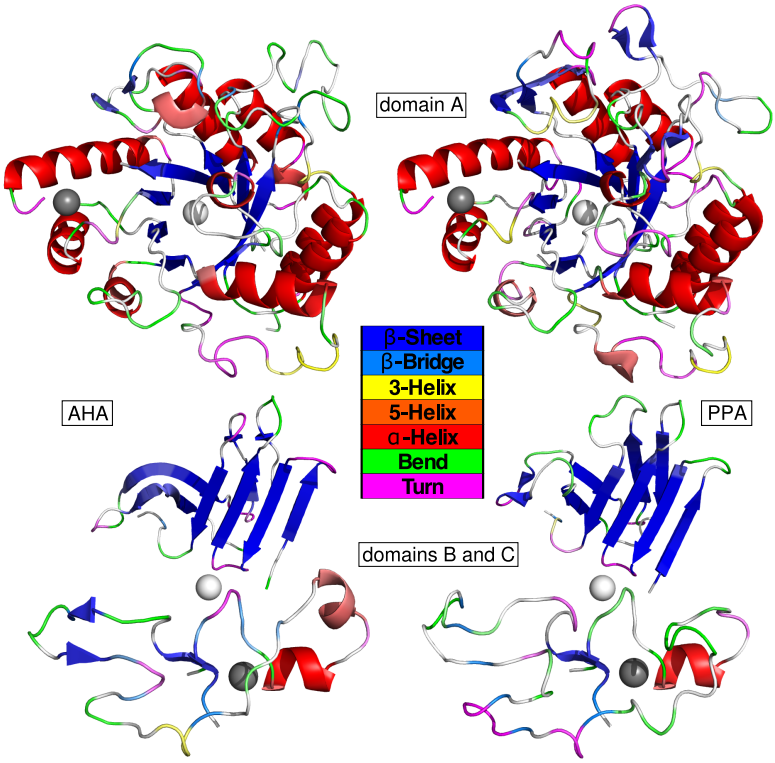


Figure 2.9: 3D representation of the secondary structure persistence in AHA (left) and PPA (right). The secondary structural classes of DSSP definition are adopted are represented with the color code indicated in the picture. Secondary structure persistence is shown for domain A (top) and domains B and C (bottom), separately.

teins which are adapted to different temperature ranges [12, 43]. Visual inspection of the 3D crystallographic structures clearly shows that the psychrophilic enzyme features shorter β-α loops and a less structured (β,α)-barrel, lacking especially the inner β-structures [12].

Analysis of the 3D representation of PDSSP (Fig.2.9, see Methods) shows that during the simulations, in the catalytic domain, the mesophilic enzyme features a greater average percentage of residues participating in the β-strands composing the core of the catalytic domain, the inner barrel, and greater random coil content ascribable to the longer, poorly structured β-α loops. The differences in these loops sizes may also account for the fact that AHA shows a higher α-helix content, while in PPA the likelihood of turns increases indicating that the longer loops are not completely unstructured, but fail to form real

Secondary Str.		PPA	AHA
		$\langle x \rangle \pm \sigma$ [%]	$\langle x \rangle \pm \sigma$ [%]
Total	β-strand	103.090 ± 9.115 [20.784%]	94.622 ± 7.783 [21.121%]
	β-bridge	15.529 ± 3.699 [3.131%]	15.550 ± 3.514 [3.471%]
	α-helix	100.097 ± 4.248 [20.181%]	106.616 ± 4.734 [23.798%]
	Bend	75.198 ± 6.724 [15.161%]	73.368 ± 5.897 [16.377%]
	Turn	66.671 ± 6.870 [13.442%]	47.823 ± 6.499 [10.675%]
	Coil	118.534 ± 6.277 [23.898%]	96.919 ± 5.675 [21.634%]
Domain A	β-strand	50.815 ± 6.396 [15.260%]	39.267 ± 4.434 [13.266%]
	β-bridge	5.245 ± 2.830 [1.575%]	5.662 ± 2.243 [1.913%]
	α-helix	95.137 ± 3.802 [28.570%]	98.451 ± 4.505 [33.260%]
	Bend	39.115 ± 5.360 [11.746%]	48.701 ± 4.389 [16.453%]
	Turn	43.801 ± 5.800 [13.153%]	28.040 ± 5.686 [9.473%]
	Coil	84.637 ± 4.452 [25.417%]	66.354 ± 3.650 [22.417%]
Domain B	β-strand	3.870 ± 2.479 [5.451%]	7.958 ± 3.217 [13.263%]
	β-bridge	6.395 ± 1.853 [9.007%]	6.147 ± 2.287 [10.245%]
	α-helix	4.960 ± 1.808 [6.986%]	8.163 ± 2.849 [13.605%]
	Bend	19.706 ± 2.681 [27.755%]	10.801 ± 2.685 [18.002%]
	Turn	12.702 ± 2.999 [17.890%]	6.334 ± 2.809 [10.557%]
	Coil	22.843 ± 2.338 [32.173%]	18.342 ± 2.691 [30.570%]
Domain C	β-strand	48.282 ± 5.679 [52.480%]	46.676 ± 5.743 [50.735%]
	β-bridge	1.267 ± 0.979 [1.377%]	1.778 ± 1.504 [1.933%]
	α-helix	0.000 ± 0.000 [0.000%]	0.002 ± 0.084 [0.002%]
	Bend	15.488 ± 3.457 [16.835%]	10.988 ± 2.841 [11.943%]
	Turn	8.181 ± 2.224 [8.892%]	11.941 ± 2.650 [12.979%]
	Coil	17.675 ± 3.653 [19.212%]	20.292 ± 3.844 [22.057%]

Table 2.1: Ensemble average of the secondary structure content in AHA and PPA, according to the DSSP software [32]. The mean values ($\langle x \rangle$) and standard deviations (σ) are shown along with the ratio between $\langle x \rangle$ and the total number of residues considered (%).

α-helices (Tab.2.1, Fig.2.9). Domain B of the mesophilic enzyme confirms this last observation, with a greater turn content, and appears to be less orderly structured that in the psychrophilic enzyme; the opposite holds for domain C, which shows a higher content of β-strand in the mesophilic enzyme.

Modulation of the number and type of intramolecular interactions may have provided to cold-adapted enzymes a simple but effective mean to achieve greater structural flexibility, which might be important in order to maintain high catalytic efficiency at low temperatures [43, 44]. Hydrogen bonds (H-bonds) are among the strongest non-

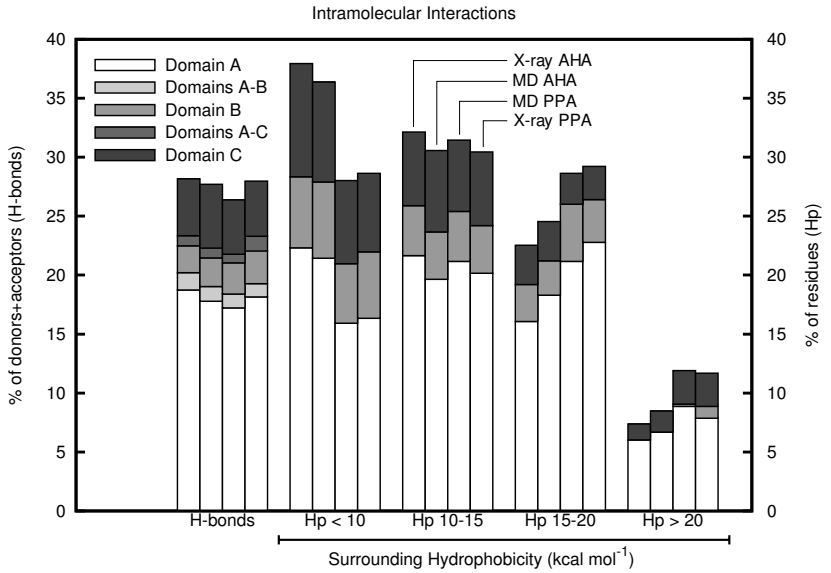


Figure 2.10: Hydrogen bonds and surrounding hydrophobicity (H_p) are described as histograms representing the percentage of donors+acceptors involved in H-bonds (first set, y axis on the left) or of residues with surrounding hydrophobicity in the indicated range (H_p, all other sets, y axis on the right); each set shows values computed for the crystal structure of AHA (1AQH, first), the MD simulations of AHA (second) and PPA (third), and the crystal structure of PPA (1PIF, fourth).

bonded interactions and the inverse proportionality of their energy to temperature makes them an interesting issue in cold-adaptation. The clustering of hydrophobic side chains within the protein core is another major driving force in the stability of the folded conformation. Comparison of these weak intramolecular interactions between the two α-amylases has been carried out by computing the number of H-bonds, and the number of residues featuring different ranges of H_p values (the residues surrounding hydrophobicity computed using experimental information, see Methods) for each of the three domains and for domain interfaces, both on crystal structures and on the merged MD trajectories (see Fig.2.10).

Our analyses show a slight decrease in the overall and per-domain number of H-bonds in the mesophilic α-amylases (Fig.2.10, first set), especially in domains A and C, while greater divergence can be identified in the extension and stability of hydrophobic clusters. AHA shows a greater number of residues with low surrounding hydropho-

bicity (H_p<10 kcal mol^{-1}, Fig.2.10 second set), when compared with the mesophilic enzyme, while PPA reveals a rise in the percentage of residues in highly hydrophobic environments (H_p>15 kcal mol^{-1}, Fig.2.10 third and fourth sets), especially in domain A. These findings correlate with the observed higher stability of the barrel's core β-strands, and suggest the weakening of core hydrophobicity as a possible adaptive trait.

2.3.2 Analysis of structural flexibility

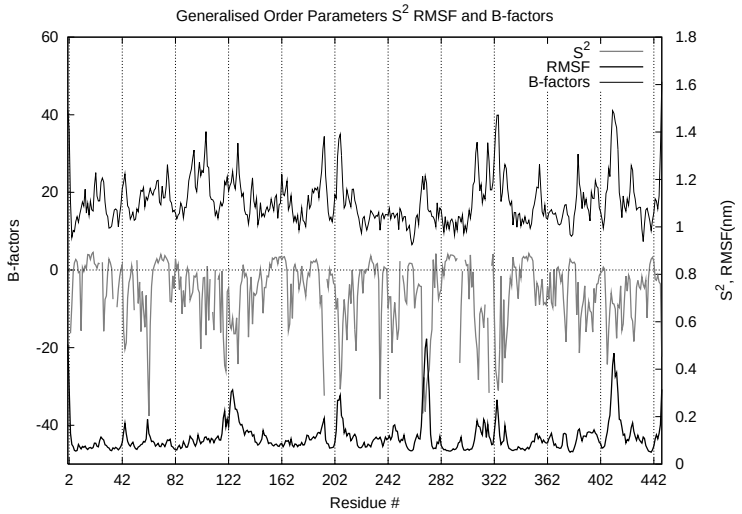
Low stability and high catalytic efficiency of cold adapted enzymes at low temperatures have been related to a flexible protein structure [45–47]. Enhanced flexibility is generally found either scattered throughout the entire protein edifice, or localized in particular regions [12, 43, 48].

In order to evaluate and compare flexibility in mesophilic and psychrophilic α-amylases, the root mean square fluctuation of C_α atoms (RMSF) with respect to the ensemble average structure was adopted as flexibility index. The validity of this flexibility index is supported by the comparison with crystallographic B-factor profiles and computed main-chain amide Generalized Order parameters[38] (S^2, Fig.2.11).

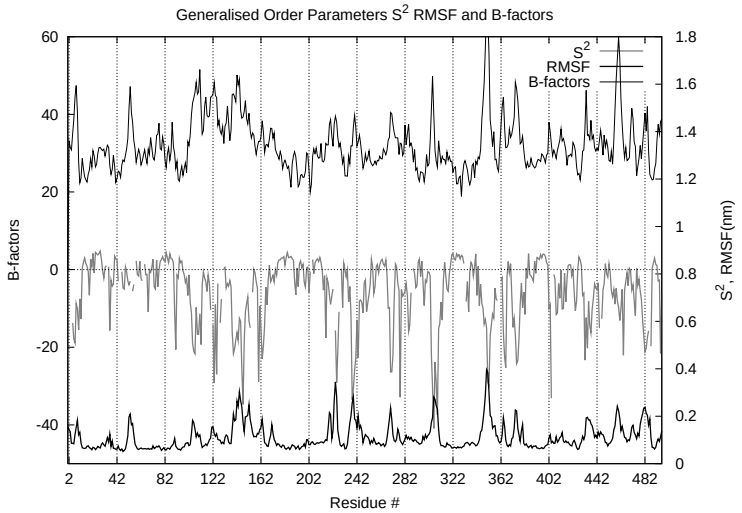
Comparison of RMSF, B-factors and S^2 parameters (2.11) demonstrates a good qualitative correspondence among the different flexibility indexes, the greatest divergence being identifiable in the temperature factors profiles, and probably ascribable to additional effects produced by the crystallization procedure, such as crystal-packing forces and intermolecular interactions with the nearest neighbors [40].

Secondary structure elements generally exhibit low flexibility in both proteins (Fig.2.12): peaks in the profiles are located in correspondence of poorly structured loop regions, and although differing remarkably in intensity, they often overlap when the two profiles are aligned. Analysis of profiles taking into account the position of secondary structure elements suggests that not only the latter are conserved in the two proteins, but also their degree of flexibility is maintained, especially if the structural elements composing the $(\beta/\alpha)_8$ barrel are considered.

The wider conformational sampling available allows us to extend the description made of these α-amylases flexibility by Spiwok *et al.*



(a)



(b)

Figure 2.11: Comparison of flexibility indexes in AHA (a) and PPA (b). RMSF, main chain N-H S^2 and B-factors are shown in thick black, gray and thin black lines, respectively.

[13]. A picture, similar to that previously described, results from using 100 ps windows (Fig.2.12a and Figure 2a from ref. [13]); this is remarkable given the differences in the simulation procedure and starting structures, especially for PPA .

Albeit slight differences in the alignment used for the comparison, regions featuring low flexibility are generally very similar, while greater divergence can be found in some peak values. When *long-time* fluctuations are considered (Fig.2.12a), along with a scattered increase in flexibility in the regions featuring low *short-time* RMSF and generally stable secondary structures, significant enhancements can be seen for most peaks. Detailed comparison of *long-time* fluctuations between the two enzymes (Fig.2.12a) allows the identification of four regions featuring significantly higher flexibility in AHA, and three regions fluctuating more in PPA (Tab.2.2). It is interesting to note that many of the highest peaks of the mesophilic enzyme flexibility profile correspond to insertions, i.e. gaps in the aligned sequence of AHA (Tab.2.2).

Structural localization of the regions listed in Tab.2.2 shows that differences in *long-time* flexibility between the two α-amylases strikingly cluster around the active-site and in the substrate-binding groove. For this reason, and to address the analysis towards the marked differences in the catalytic behavior of the two enzymes, the next sections present a detailed description of the character of these two enzymes active site flexibility and dynamics.

2.3.3 Overall flexibility of the active-site

The eight loops that protrude from the $(\beta/\alpha)_8$ barrel structure above the active site undergo particularly intense motions during the simulations, and represent the most flexible regions surrounding the catalytic triad. Among these, those featuring the most pronounced fluctuations are the portions of loops $\beta_3-\alpha_3$, $\beta_5-\alpha_5$ and $\beta_7-\alpha_7$ that project towards the catalytic triad forming a three-sided cleft on one side of the substrate-binding groove (**L3**, **L5** and **L7**, respectively, see Fig.2.1 and Tab.2.2 for the definition of the abbreviations). Furthermore, all three have been found to be in contact with co-crystallized substrate analogues, and a possible role in the catalytic cycle of both α-amylases has been suggested for the rearrangement of $\beta_7-\alpha_7$ in the process of binding and release of the substrate [27, 49–53].

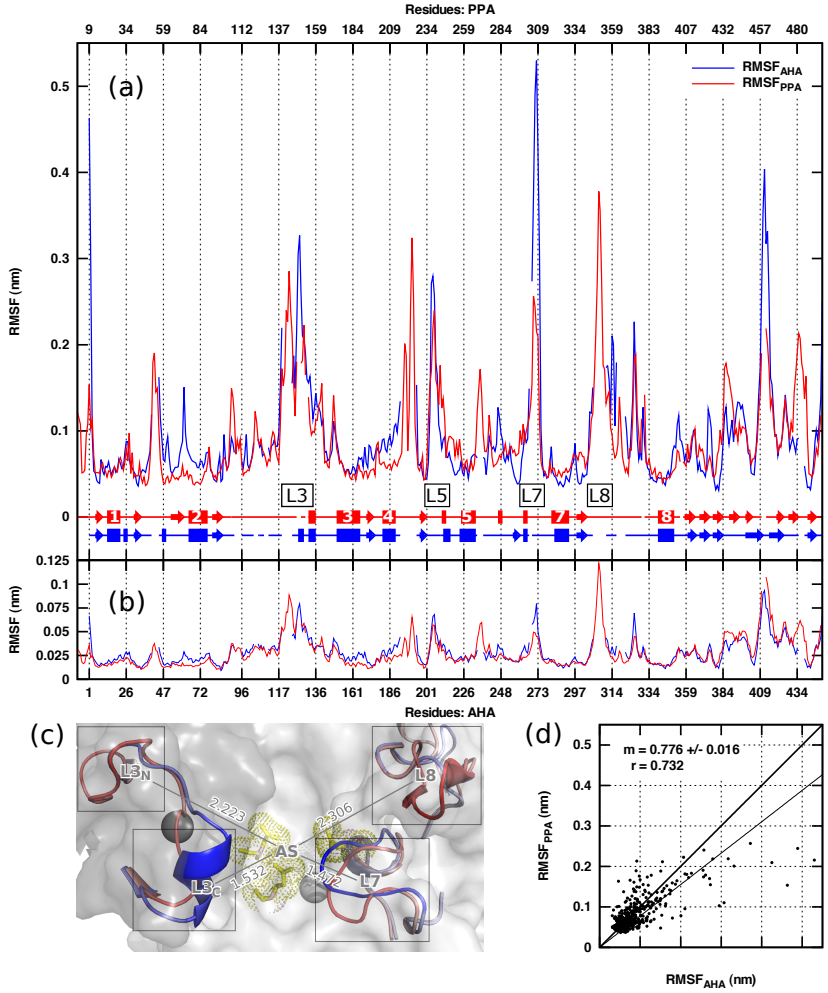


Figure 2.12: Aligned RMSF profiles as a function of alignment position computed on the metatrajectory (a) and on 100 ps intervals (b). The RMSF profiles of AHA (blue line) and PPA (red line) are shown as thick lines, broken in correspondence of gaps in the alignment; the four highly flexible loops mentioned in the text are outlined, and the most persistent secondary structures are represented schematically. (c) Close-up view of the two enzymes substrate binding groove (represented as the transparent surface of PPA), where distances between the catalytic triad (*AS*, show in yellow) and the most flexible regions in the surroundings are highlighted: $\mathbf{L3}_{\mathrm{N}}$ (PPA), $\mathbf{L3}_{\mathrm{C}}$ (AHA), **L7** (AHA) and **L8** (PPA). The backbone colored as in Fig.2.1. Distances are ensemble averages computed between the centres of mass of the groups ($\mathbf{L3}_{\mathrm{N}}$-*AS*: 2.223 nm, $\mathbf{L3}_{\mathrm{C}}$-*AS*: 1.532 nm, **L8**-*AS*: 2.306 nm, **L7**-*AS*: 1.472 nm). (d) Correlation of flexibility of corresponding residues for AHA and PPA according to the structural alignment: the regression (m) and correlation (r) coefficients are displayed in the plot.

AHA	PPA	Domain	Seq. (AHA)	Seq. (PPA)	Abbreviation
57 – 62	69 – 74	A	L**QS**R**G**G	L**CT**R**S**G	
92 – 95	**104 – 107**	B	**AGSG**	**GSGA**	
122 – 129	148 – 153	B	I**NNSD**Y**GN**	I**E--S**Y**ND**	**L3**$_C$
268 – 274	305 – 311	A	**GHG**G**AGN**	**HGA**G**GSS**	**L7**
384 – 392	**432 – 440**	C	**E**D**ST**L**TA**T**V**	**D**D**WQ**L**SS**T**L**	
408 – 415	456 – 461	C	**ELSADAK**S	**DKV--GN**S	
432 – 437	**478 – 486**	C	**NIGA---WD**	**SISNSAQDP**	
	52 – 55	A		**TNPS**	
	140 – 145	B		**KCKTA**	**L3**$_N$
	217 – 226	A		**LNTNWFPAGS**	
	269 – 271	A		**WSG**	
	347 – 354	A		**NFVNGEDV**	**L8**

Table 2.2: Regions featuring relevant differences in flexibility between the two α-amylases: bold intervals indicate which enzyme features higher fluctuations, while bold amino-acid names highlight substitutions between the two enzymes; in the lower part of the table, the highly flexible insertions of PPA are listed.

L3 features low structural similarity between the two enzymes and the greatest fluctuational mobility inside domain B: in AHA the C-terminal region of **L3** (**L3**$_C$) is longer and carries a stable, short and highly mobile α-helix, while in PPA it is preceded by a highly flexible coil region (the N-terminal region of **L3**, **L3**$_N$), which has no correspondence in the structure of the psychrophilic enzyme (Fig.2.12c). The comparison of the intensity and localization of **L3** motions shows that in AHA high flexibility is localized nearer to the catalytic triad than in PPA (see Fig.2.12c).

L5 ($203-206^{\mathbf{A}}$, $236-239^{\mathbf{P}}$), immediately follows the catalytic $E200^{\mathbf{A}}$ ($E233^{\mathbf{P}}$), and features great conservation both in structure and intensity of fluctuations (Fig.2.12). Significant differences between the two enzymes can be identified in the correlation of motions between **L5** and **L3**, greater in AHA as shown in Fig.2.13b, where relevant correlations between C_α atoms of the two loops have been mapped on the structure of the two enzymes (see Methods). This marked divergence in the correlation pattern may be connected with the absence, in AHA, of the $C70$-$C115^{\mathbf{P}}$ disulphide bridge on the opposite side of domain B.

L7 immediately follows a short helical structure carrying the third acidic catalytic residue ($D264^{\mathbf{A}}$, $D300^{\mathbf{P}}$), is most flexible region of the cold-active enzyme and features the greatest fluctuational differences

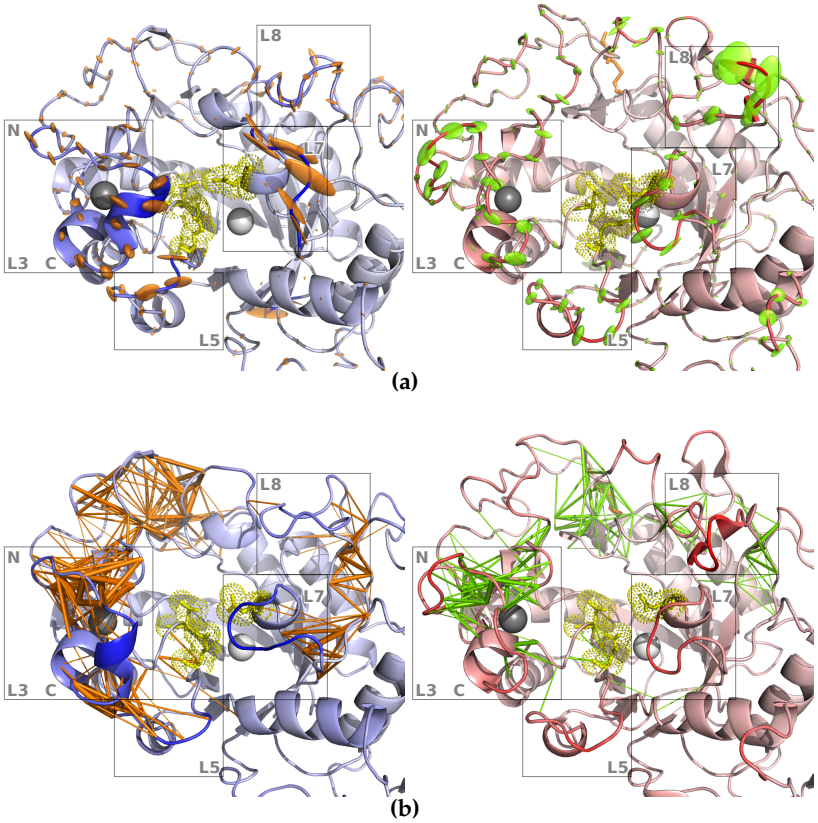


Figure 2.13: (a) Computed anisotropic temperature factors for the *essential subspace* are shown as ellipsoids for C_α atoms, mapped on the average structure of the simulated ensembles (see Methods); (b) correlation plots showing significant positive correlations between C_α atoms of domains A and B (see Methods). Left and right panels picture AHA and PPA, respectively. The backbone is colored as in Fig.2.1.

between the two enzymes (Fig.2.12a). Dominant motions of **L7** involve oscillations to and from the active site groove, with amplitudes strongly enhanced in AHA (Fig.2.13a). Interestingly in PPA, immediately above and in direct contact with **L7**, a highly flexible insertion is found (**L8**), which is the most flexible region of the mesophilic α-amylase (Fig.2.12c).

L3 and **L7** are the two most mobile loops surrounding the active site, and in both cases it is clear that highly flexible insertions in PPA (**L3**$_N$ and **L8**) are found in the immediate proximity of conserved regions (**L3**$_C$ and **L7**); we have shown that these more conserved regions are characterized by higher fluctuations in AHA, so that mobility of

the substrate-binding site in PPA is shifted further away from the catalytic triad (Fig.2.12c). The insertions in PPA appear to be of key importance in determining the overall flexibility of the mesophilic enzyme, as shown in Fig.2.12a, where gaps in the aligned sequence of AHA correspond to peaks in the RMSF profile of the mesophilic α-amylase. Furthermore, while the average flexibility of the two enzymes is comparable, if only matching residues are taken into account according to the structural alignment (i.e. insertions are ignored), the psychrophilic enzyme shows greater flexibility scattered throughout the entire protein, as shown in Fig.2.12d. Finally, analysis of the SAS of PPA shows that insertions exhibit on average more than double the SAS of PPA as a whole. Although a clear establishment of the role of the mesophilic enzyme longer loops and their flexibility in temperature-adaptation is complicated by the evolutionary distance between the two organisms from which the two α-amylases have been isolated [13, 54], taken together these results suggest that dominant motions in PPA, when compared to its cold-active homologue, are displaced to few highly flexible, highly solvent accessible loop regions, and thus further away from the active site.

2.3.4 Dynamic properties of loop7

The availability of crystallographic studies of both AHA and PPA in complex with oligosaccharidic substrate analogues (bound forms) has allowed us to analyze active site motions in the free form in terms of its opening and closing, and to make some insightful considerations on its dynamics in solution as observed during our simulations. These studies[49, 52] have shown that in both enzymes substrate binding causes significant displacements of **L7** towards the substrate, causing an apparent closure of the active-site groove[55, 56]. Furthermore, a $\sim 60^\circ$ rotation of the side-chain of D300$^{\mathbf{P}}$ around the C_α-C_β dihedral (χ_1) is observed in PPA, possibly leading the catalytic aspartate to its catalytically active conformation. Conversely, in all available crystal structures of AHA, bound and free, the side-chain of D264$^{\mathbf{A}}$ is very similar to that found in the bound forms of PPA. Therefore we have focused our analysis on *(i)* the conformational fluctuations of **L7** and *(ii)* the rotation of the catalytic aspartate side-chain around dihedral χ_1.

The dynamics of **L7** in the free form have been analyzed using

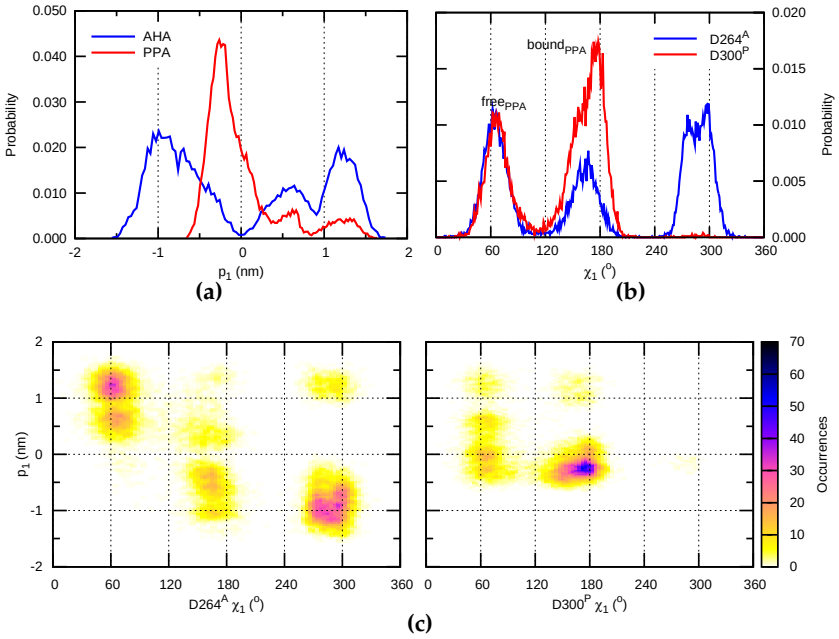


Figure 2.14: Distributions of (a) values of the first principal component of **L7** motion and (b) χ_1 dihedral of D264$^{\mathbf{A}}$/D300$^{\mathbf{P}}$ are joined (c) for AHA (left) and PPA (right); bin sizes have been set to on hundredth of the range of each data set.

the first principal component (p_1) of motion of the loop region (see Methods). In both enzymes fluctuations along p_1 identify an active-site opening/closing oscillation where positive values correspond to more open conformations; motions along p_1 describe 67.2% and 40.1% of total variance for AHA and PPA respectively. The extent of **L7** motions differs substantially between the two α-amylases, as previously described (Fig.2.12), and the distribution of p_1 values in the simulated ensemble (Fig.2.14a) shows that the high RMSF observed for **L7** in AHA arises from a multimodal distribution of sampled conformations, while in PPA **L7** motions are more restricted.

We have measured χ_1 dihedral values for all available crystal structures of the two enzymes, both in the presence (bound) or absence (free) of a substrate or inhibitor. Our measurements (Fig.2.15a) show that free structures of PPA exhibit χ_1 values around 60° ($free_{PPA}$); all structures of AHA and bound structures of PPA, instead, feature χ_1 values around 150° ($bound_{PPA}$).

This measurement was extended to 73 available crystal structures

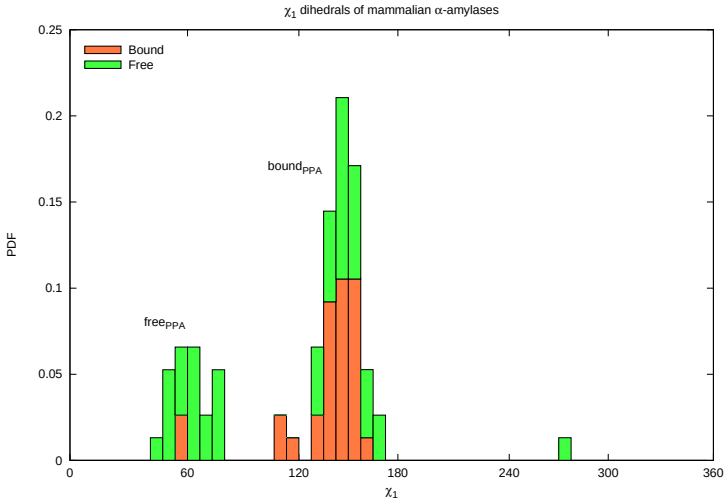
of α-amylases, of which 4 have been excluded from further analyses:

- 1BVN (PPA): a large proteinaceous inhibitor (Tendamistat) is bound in the active-site cleft, causing a large displacement of **L7** far from the active site: it features a χ_1 value intermediate between the $bound_{PPA}$ and $free_{PPA}$ values;
- 1DHK (PPA): the large proteinaceous bean lectin-like inhibitor is bound in the active-site cleft, and **L7** is trapped in a closed-like conformation: χ_1 is in range with other free structures;
- 1WO2 (PPA): the structure, featuring a tri-saccharide bound in the active site, contains 2 alternative states for ASP300$^{\mathbf{P}}$: one referring to the enzyme coordinating a Cl^- ion in the active site, featuring a $bound_{PPA}$-like χ_1 value, and one to the enzyme without the Cl^- ion, featuring a $free_{PPA}$-like value;
- 1BOI (AHA): it has been proposed that the χ_1 value of 75.2° may be a consequence of differences in the pH of the crystallization medium[3].

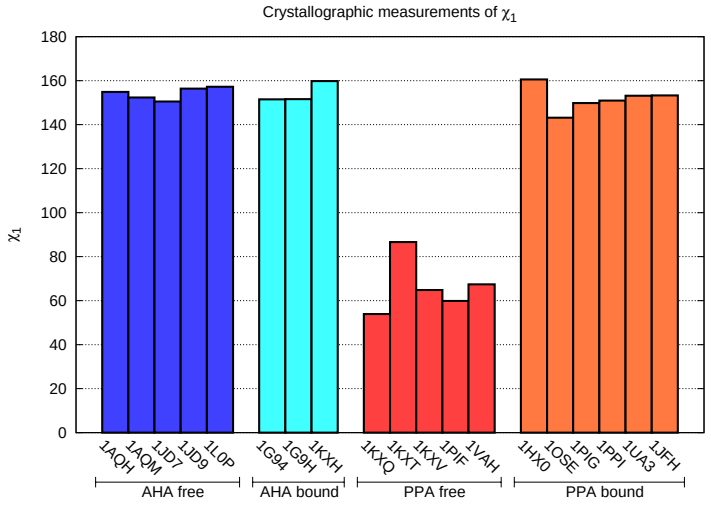
The distribution of the remaining 69 measured values is displayed in Fig.2.15b, and shows that bound structures strongly tend to adopt $bound_{PPA}$ χ_1 values, while free structures have a lower degree of selectivity for $free_{PPA}$ χ_1 values. The single outlier at ~282° is a crystal structure of Cl^--dependent human pancreatic α-amylase mutant N298S (1XGZ), which impairs Cl^--binding[57].

The distribution of χ_1 dihedral values in the simulated ensemble (Fig.2.14b) shows that in both enzymes, conformations of the catalytic aspartic span three basins, two of which are in good agreement with crystallographic results and have been termed $free_{PPA}$ and $bound_{PPA}$, in relation to values observed in crystal structures of PPA. The distribution also shows that in our simulations D264$^{\mathbf{A}}$ possesses greater rotational freedom than that exhibited in available crystal structures, and even greater than D300$^{\mathbf{P}}$.

The joint distributions of p_1 and χ_1 (Fig.2.14), instrumental in relating the behavior of **L7** and the catalytic aspartic-acid residue in the simulated ensembles, show that in AHA (Fig.2.14c, left) motions along p_1 strongly correlate with specific D264$^{\mathbf{A}}$ side-chain rotamers: values of χ_1 spanning the $free_{PPA}$ basin match positive values of p_1, while nega-



(a)



(b)

Figure 2.15: χ_1 measurements for all available crystal structures of AHA and PPA (a), and for 69 crystal structures of α-amylases, shown as column-stacked histograms (b): structures have been grouped on the basis of the presence (bound) or absence (free) of a substrate, transition-state analogue or inhibitor in the active site.

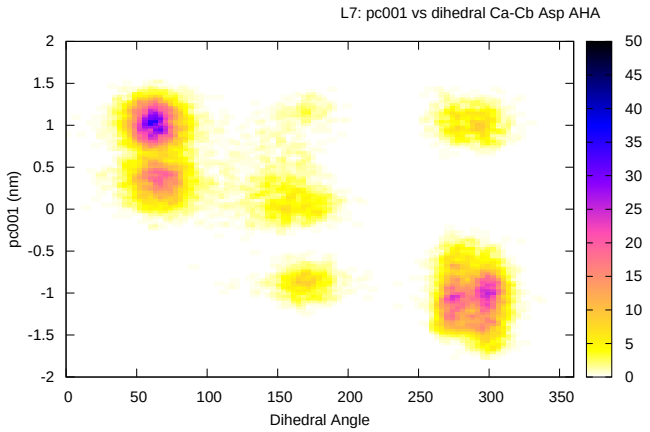


Figure 2.16: Joint distribution of values of the first principal component of **L7** motion and χ_1 dihedral of D264$^{\mathbf{A}}$, as computed on a 28.6 ns metatrajectory obtained joining replicas 2, 4, 5 and 6, of which 3 are 11 ns long; bin sizes are the same used in 2.14.

tive values of p_1 correspond to the other two χ_1 basins, including the $bound_{PPA}$ basin.

In order to assess if this behavior has converged in our simulations and it is not the consequence of the particular combination of *replicas*, we have repeated the analysis ona 28.6 ns metatrajectory composed of the 4 replicas featuring the most clear-cut convergence, both in terms of rmsd and of cosine content. The results obtained (Fig.2.16) are analogous to those shown in Fig.2.14c, and allow to drive similar conclusions to those discussed here.

These data suggest the presence of a strong interdependence between motions of **L7** and rotation of D264$^{\mathbf{A}}$ around χ_1, and in particular that more open conformations of the active site correlate with values of χ_1 corresponding to the $free_{PPA}$ basin; a similar correlated behavior cannot be identified for PPA (Fig.2.14c, right).

The observed dynamics of the active site of AHA suggest that the **L7** region, including catalytic D264$^{\mathbf{A}}$, undergoes conformational changes in the absence of substrate which may be related to the motions involved in reaching the transition state upon substrate binding. The same does not hold for PPA, and the low mobility of **L7** in the mesophilic enzyme is consistent with the computed high energy barrier between its *free* and *bound* microstates [56]. These marked differences in the extent and character of **L7** motions in AHA and PPA have

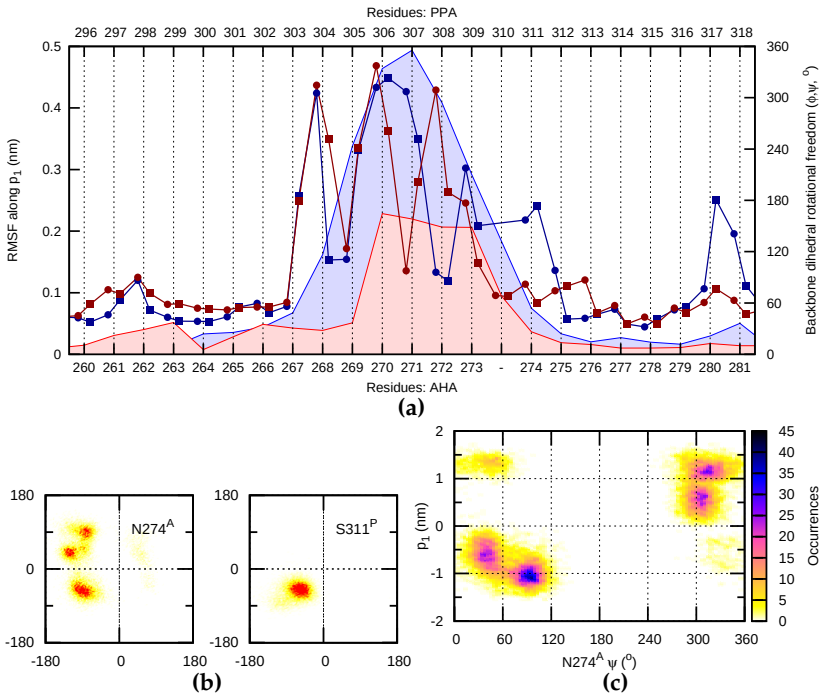


Figure 2.17: (a) Profiles of dihedral rotational freedom (linespoints) of ϕ (circles) and ψ(squares) for AHA (blue) and PPA (red) are plotted with the RMSF along the first principal component of **L7** motion p_1 (shaded curves); (b) Ramachandran plots of residues N274$^{\mathbf{A}}$ (left) and S311$^{\mathbf{P}}$ (right); (c) Joint distribution of N274$^{\mathbf{A}}$ ψ values and motions along p_1; bin sizes are the same used in 2.14.

prompted us to analyze the mechanistic determinants of this asymmetry. The structure and localization of **L7** in both enzymes suggests that it may be categorized as an Ω loop, as shown for other α-amylases[55]. The movement of these Ω loop has been suggested to rely on backbone rotations of hinge residues located at their extremities; we have thus proceeded to characterize the hinges of **L7** movement in AHA and PPA, by measuring both C_α fluctuations and backbone dihedral (ϕ and ψ) rotational freedom of **L7** and its immediate surroundings.

Fig.2.17a shows both the C_α RMSF along p_1 and an estimation of the statistical dispersion of ϕ and ψ values in the simulated ensemble (see Methods). The plot shows that C_α flexibility profiles follow a bell-shaped curve, which is preceded by a few residues that feature marked rotational freedom, in particular R267$^{\mathbf{A}}$ G268$^{\mathbf{A}}$ and H269$^{\mathbf{A}}$ (R303$^{\mathbf{P}}$, G304$^{\mathbf{P}}$ and H305$^{\mathbf{P}}$); these residues are well conserved in Cl^--

dependent α-amylases, and could work as the N-terminal hinges of **L7** motion. Globally, ϕ angles show good correspondence between the two α-amylase, the most apparent differences involving ϕ dihedrals of G271$^{\mathbf{A}}$ and A272$^{\mathbf{A}}$ (A307$^{\mathbf{P}}$ and G308$^{\mathbf{P}}$), and probably due to the inversion of the position of an Alanine and Glycine residues in the two α-amylases sequence. Greater differences are present in the ψ profiles, and particularly on the C-terminal side of the peak in the flexibility profiles, in correspondence of residue N274$^{\mathbf{A}}$, which features higher rotational freedom than S311$^{\mathbf{P}}$, as pointed out by the comparison of the two residues Ramachandran plots (Fig.2.17b): while N274$^{\mathbf{A}}$ interconverts between three distinct backbone conformations differing particularly for ψ values, S311$^{\mathbf{P}}$ only adopts ϕ,ψ values consistent with α-helical conformations. Furthermore, the joint distribution of p_1 and N274$^{\mathbf{A}}$ ψ (Fig.2.17c) shows that the three backbone rotamers of N274$^{\mathbf{A}}$ ψ correspond well with the peaks in the distribution of p_1 values in AHA (see Fig.2.14a), suggesting that N274$^{\mathbf{A}}$ could act as the C-terminal hinge of **L7** motion, and that its higher backbone rotational freedom could account for enhanced mobility of **L7** in AHA.

2.4 Conclusions

With the aim of contributing to the elucidation of the structural and dynamical features relevant for enzyme cold adaptation in the α-amylase family, we adopt MD simulations to investigate and compare psychrophilic and mesophilic homologues: our extensive sampling of the two α-amylases phase space allows us to draw significant conclusions concerning the nanosecond-scale behavior of these enzymes at room temperature, and to precisely identify the regions where the most striking differences in the dynamic properties of these two enzymes are localized.

We find that in both enzymes, highly flexible regions and the most noticeable differences in flexibility are located mainly in the surroundings of the substrate-binding groove and consist of the $\beta-\alpha$-loops that protrude from the main domain's barrel structure. It has previously been reported that these loops are longer in the mesophilic enzyme, and thus feature regions which have no counterpart in the structural alignment of the two proteins (i.e. insertions) which extend further away from the core of the barrel, where the active site is located. Our find-

ings show that the essential dynamics of PPA primarily involve these highly solvent accessible insertion regions, leaving the immediate surroundings of the active site, which feature greater structural similarity between the two α-amylases, comparatively more flexible in the psychrophilic enzyme. Increased active-site flexibility has often been considered as an adaptive trait in the context of temperature adaptation, and the likely interplay of other evolutionary phenomena related to the different nature of evolution in prokaryotes and eukaryotes [13, 54] does not, in our opinion, necessarily rule out its possible relevance in this context.

Furthermore, previous crystallographic studies of the substrate-bound and free conformations of the two enzymes suggested the importance of key structural rearrangements in the binding and processing of small substrates [49, 52]. In addition, it has been suggested for AHA that the kinetic parameters of substrate hydrolysis and the thermodynamic parameters of ligand binding are linked to improved molecular motions of the residues involved in substrate binding and catalysis[47]. Our detailed analysis of their active site has allowed us to identify concerted motions of a conserved highly flexible Ω loop involved in substrate binding (**L7**) and a catalytic aspartic residue ($D264^{\mathbf{A}}$/$D300^{\mathbf{P}}$), whose rotation towards the catalytically competent conformation strongly correlates with loop closure in AHA. We were also able to identify the hinge residues of **L7** movement, where the most striking differences between the two enzymes are found, and in particular an asparagine ($N274^{\mathbf{A}}$) whose backbone dihedral rotation correlates with **L7** movement, and whose greater rotational freedom may account for the enhanced motions of **L7** in AHA. These results substantiate the presence of concerted and arguably functionally relevant motions in both enzymes even in the absence of substrate, suggesting that an interplay of motions intrinsic in the two enzymes may be important in performing their activity [58, 59], and that modulation of these dynamic properties could be significant in the context of their thermal adaptation.

Bibliography

[1] G. Feller, T. Lonhienne, C. Deroanne, C. Libioulle, J. Van Beeumen, and C. Gerday. Purification, characterization, and nucleotide se-

quence of the thermolabile alpha-amylase from the antarctic psychrotroph Alteromonas haloplanctis A23. *J Biol Chem*, 267(8):5217–5221, Mar 1992.

[2] G. Feller, F. Payan, F. Theys, M. Qian, R. Haser, and C. Gerday. Stability and structural analysis of alpha-amylase from the antarctic psychrophile Alteromonas haloplanctis A23. *Eur J Biochem*, 222(2):441–447, Jun 1994.

[3] N. Aghajari, G. Feller, C. Gerday, and R. Haser. Structures of the psychrophilic Alteromonas haloplanctis alpha-amylase give insights into cold adaptation at a molecular level. *Structure*, 6(12):1503–1516, Dec 1998.

[4] N. Aghajari, G. Feller, C. Gerday, and R. Haser. Crystal structures of the psychrophilic alpha-amylase from Alteromonas haloplanctis in its native form and complexed with an inhibitor. *Protein Sci*, 7(3):564–572, Mar 1998.

[5] S. D'Amico, C. Gerday, and G. Feller. Structural similarities and evolutionary relationships in chloride-dependent alpha-amylases. *Gene*, 253(1):95–105, Jul 2000.

[6] G. Feller, D. d'Amico, and C. Gerday. Thermodynamic stability of a cold-active alpha-amylase from the Antarctic bacterium Alteromonas haloplanctis. *Biochemistry*, 38(14):4613–4619, Apr 1999.

[7] S. D'Amico, C. Gerday, and G. Feller. Structural determinants of cold adaptation and stability in a large protein. *J Biol Chem*, 276(28):25791–25796, Jul 2001.

[8] Salvino D'Amico, Jean-Claude Marx, Charles Gerday, and Georges Feller. Activity-stability relationships in extremophilic enzymes. *J Biol Chem*, 278(10):7891–7896, Mar 2003.

[9] Salvino D'Amico, Charles Gerday, and Georges Feller. Dual effects of an extra disulfide bond on the activity and stability of a cold-adapted alpha-amylase. *J Biol Chem*, 277(48):46110–46115, Nov 2002.

[10] Khawar S Siddiqui, Georges Feller, Salvino D'Amico, Charles Gerday, Laura Giaquinto, and Ricardo Cavicchioli. The active site is

the least stable structure in the unfolding pathway of a multidomain cold-adapted alpha-amylase. *J Bacteriol*, 187(17):6197–6205, Sep 2005.

[11] Salvino D'Amico, Charles Gerday, and Georges Feller. Temperature adaptation of proteins: engineering mesophilic-like activity and stability in a cold-adapted alpha-amylase. *J Mol Biol*, 332(5):981–988, Oct 2003.

[12] D. Georlette, V. Blaise, T. Collins, S. D'Amico, E. Gratia, A. Hoyoux, J-C. Marx, G. Sonan, G. Feller, and C. Gerday. Some like it cold: biocatalysis at low temperatures. *FEMS Microbiol Rev*, 28(1):25–42, Feb 2004.

[13] Vojtech Spiwok, Petra Lipovová, Tereza Skálová, Jarmila Dusková, Jan Dohnálek, Jindrich Hasek, Nicholas J Russell, and Blanka Králová. Cold-active enzymes studied by comparative molecular dynamics simulation. *J Mol Model*, 13(4):485–497, Apr 2007.

[14] B. Hess. Convergence of sampling in protein simulations. *Physical Review E*, 65:031910, 2002.

[15] L. S. Caves, J. D. Evanseck, and M. Karplus. Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin. *Protein Sci*, 7(3):649–666, Mar 1998.

[16] Elena Papaleo, Laura Riccardi, Chiara Villa, Piercarlo Fantucci, and Luca De Gioia. Flexibility and enzymatic cold-adaptation: a comparative molecular dynamics investigation of the elastase family. *Biochim Biophys Acta*, 1764(8):1397–1406, Aug 2006.

[17] Alessandro Pandini, Giancarlo Mauri, Annalisa Bordogna, and Laura Bonati. Detecting similarities among distant homologous proteins by comparison of domain flexibilities. *Protein Eng Des Sel*, 20(6):285–299, Jun 2007.

[18] Magne Olufsen, Arne O Smalås, Elin Moe, and Bjørn O Brandsdal. Increased flexibility as a strategy for cold adaptation: a comparative molecular dynamics study of cold- and warm-active uracil DNA glycosylase. *J Biol Chem*, 280(18):18042–18048, May 2005.

[19] Manuel Rueda, Carles Ferrer-Costa, Tim Meyer, Alberto Pérez, Jordi Camps, Adam Hospital, Josep Lluis Gelpí, and Modesto Orozco. A consensus view of protein dynamics. *Proc Natl Acad Sci U S A*, 104(3):796–801, Jan 2007.

[20] M. Karplus and J. A. McCammon. Molecular dynamics simulations of biomolecules. *Nat Struct Biol*, 9(9):646–652, 2002.

[21] G. Dodson and C. S. Verma. Protein flexibility: its role in structure and mechanism revealed by molecular simulations. *Cell Mol Life Sci*, 63(2):207–219, Jan 2006.

[22] Wilfred F van Gunsteren, Dirk Bakowies, Riccardo Baron, Indira Chandrasekhar, Markus Christen, Xavier Daura, Peter Gee, Daan P Geerke, Alice Glättli, Philippe H Hünenberger, Mika A Kastenholz, Chris Oostenbrink, Merijn Schenk, Daniel Trzesniak, Nico F A van der Vegt, and Haibo B Yu. Biomolecular modeling: Goals, problems, perspectives. *Angew Chem Int Ed Engl*, 45(25):4064–4092, Jun 2006.

[23] Aldert R. van Buuren, Siewert Jan Marrink, and Herman J. C. Berendsen. A molecular dynamics study of the decane/water interface. *The Journal of Physical Chemistry*, 97(36):9206–9212, September 1993.

[24] Alan E. Mark, Steven P. van Helden, Paul E. Smith, Lambert H. M. Janssen, and Wilfred F. van Gunsteren. Convergence properties of free energy calculations: .alpha.-cyclodextrin complexes as a case study. *Journal of the American Chemical Society*, 116(14):6293–6302, July 1994.

[25] W. F. van Gunsteren and H. J. C. Berendsen. *Gromos-87 manual.* Biomos BV Nijenborgh 4, 9747 AG Groningen, The Netherlands, 1987.

[26] M. Machius, L. Vértesy, R. Huber, and G. Wiegand. Carbohydrate and protein-based inhibitors of porcine pancreatic alpha-amylase: structure analysis and comparison of their binding characteristics. *J Mol Biol*, 260(3):409–421, Jul 1996.

[27] G. D. Brayer, Y. Luo, and S. G. Withers. The structure of human pancreatic alpha-amylase at 1.8 a resolution and comparisons with related enzymes. *Protein Sci*, 4(9):1730–1742, Sep 1995.

[28] H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, and J. Hermans. Intermolecular forces. In *Adv Appl Lipid Res*, volume 331–332. Reidel, 1981.

[29] H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, A. Di Nola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *J Chem Phys*, 81(8):3684–3690, 1984.

[30] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. fraaije. LINCS: A linear constraint solver for molecular simulations. *J Comput Chem*, 18(12):1463–1472, 1997.

[31] T. Darden, D. York, and L. Pedersen. Particle mesh ewald: An n-log(n) method for ewald sums in large systems. *J Chem Phys*, 98:10089–10092, 1993.

[32] W. Kabsch and C. Sander. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12):2577–2637, Dec 1983.

[33] D. D. Jones. Amino acid properties and side-chain orientation in proteins: a cross correlation appraoch. *J Theor Biol*, 50(1):167–183, Mar 1975.

[34] Chie Motono, M. Michael Gromiha, and Sandeep Kumar. Thermodynamic and kinetic determinants of thermotoga maritima cold shock protein stability: a structural and dynamic analysis. *Proteins*, 71(2):655–669, May 2008.

[35] S.J. Hubbard and J.M. Thornton. 'naccess', computer program, department of biochemistry and molecular biology, university college london. 1993.

[36] L. Holm and C. Sander. Mapping the protein universe. *Science*, 273(5275):595–603, Aug 1996.

[37] E. Beitz. Texshade: shading and labeling of multiple sequence alignments using latex2 epsilon. *Bioinformatics*, 16(2):135–139, Feb 2000.

[38] Giovanni Lipari and Attila Szabo. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. theory and range of validity. *J. Am. Chem. Soc.*, 104(17):4546 – 4559, 1982.

[39] DC Chatfield, A Szabo, and BR Brooks. Molecular dynamics of staphylococcal nuclease: Comparison of simulation with n-15 and c-13 nmr relaxation data. *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY*, 120(21):5301–5311, JUN 3 1998.

[40] P. H. Hünenberger, A. E. Mark, and W. F. van Gunsteren. Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations. *J Mol Biol*, 252(4):492–503, Sep 1995.

[41] A. Amadei, A. B. M. Lissen, and H. J. C. Berendsen. Essential dynamics of proteins. *Proteins*, 17:412–425, 1993.

[42] B. Hess. Similarities between principal components of protein dynamics and random diffusion. *Physical Review E*, 62:8438–8448, 2000.

[43] Khawar Sohail Siddiqui and Ricardo Cavicchioli. Cold-adapted enzymes. *Annu Rev Biochem*, 75:403–433, 2006.

[44] Richard A Goldstein. Amino-acid interactions in psychrophiles, mesophiles, thermophiles, and hyperthermophiles: insights from the quasi-chemical approximation. *Protein Sci*, 16(9):1887–1895, Sep 2007.

[45] P. A. Fields, B. D. Wahlstrand, and G. N. Somero. Intrinsic versus extrinsic stabilization of enzymes: the interaction of solutes and temperature on A4-lactate dehydrogenase orthologs from warm-adapted and cold-adapted marine fishes. *Eur J Biochem*, 268(16):4497–4505, Aug 2001.

[46] Khawar Sohail Siddiqui, Anne Poljak, Michael Guilhaus, Georges Feller, Salvino D'Amico, Charles Gerday, and Ricardo Cavicchioli. Role of disulfide bridges in the activity and stability of a cold-active alpha-amylase. *J Bacteriol*, 187(17):6206–6212, Sep 2005.

[47] S. D'Amico, J. S. Sohier, and G. Feller. Kinetics and energetics of ligand binding determined by microcalorimetry: insights into active site mobility in a psychrophilic alpha-amylase. *J Mol Biol*, 358(5):1296–1304, May 2006.

[48] T. Lonhienne, C. Gerday, and G. Feller. Psychrophilic enzymes: revisiting the thermodynamic parameters of activation may explain local flexibility. *Biochim Biophys Acta*, 1543(1):1–10, Nov 2000.

[49] M. Qian, R. Haser, G. Buisson, E. Duée, and F. Payan. The active center of a mammalian alpha-amylase. Structure of the complex of a pancreatic alpha-amylase with a carbohydrate inhibitor refined to 2.2-A resolution. *Biochemistry*, 33(20):6284–6294, May 1994.

[50] M. Qian, R. Haser, and F. Payan. Carbohydrate binding sites in a pancreatic alpha-amylase-substrate complex, derived from x-ray structure analysis at 2.1 a resolution. *Protein Sci*, 4(4):747–755, Apr 1995.

[51] N. Ramasubbu, V. Paloth, Y. Luo, G. D. Brayer, and M. J. Levine. Structure of human salivary alpha-amylase at 1.6 a resolution: implications for its role in the oral cavity. *Acta Crystallogr D Biol Crystallogr*, 52(Pt 3):435–446, May 1996.

[52] Nushin Aghajari, Michel Roth, and Richard Haser. Crystallographic evidence of a transglycosylation reaction: ternary complexes of a psychrophilic alpha-amylase. *Biochemistry*, 41(13):4273–4280, Apr 2002.

[53] Narayanan Ramasubbu, Chandran Ragunath, and Prasunkumar J Mishra. Probing the role of a mobile loop in substrate binding and enzyme activity of human salivary amylase. *J Mol Biol*, 325(5):1061–1076, Jan 2003.

[54] Burkhard Rost. Did evolution leap to create the protein universe? *Curr Opin Struct Biol*, 12(3):409–416, Jun 2002.

[55] G. André and V. Tran. Putative implication of alpha-amylase loop 7 in the mechanism of substrate binding and reaction products release. *Biopolymers*, 75(2):95–108, Oct 2004.

[56] Srinath Cheluvaraja, Mihail Mihailescu, and Hagai Meirovitch. Entropy and free energy of a mobile protein loop in explicit water. *J Phys Chem B*, 112(31):9512–9522, Aug 2008.

[57] Robert Maurus, Anjuman Begum, Hsin-Hen Kuo, Andrew Racaza, Shin Numao, Carsten Andersen, Jeppe W Tams, Jesper Vind, Christopher M Overall, Stephen G Withers, and Gary D Brayer. Structural and mechanistic studies of chloride induced activation of human pancreatic alpha-amylase. *Protein Sci*, 14(3):743–755, Mar 2005.

[58] J. L. Radkiewicz and C. L. Brooks. Protein dynamics in enzymatic catalysis: Exploration of dihydrofolate reductase. *J Am Chem Soc*, 122(2):225–231, 2000.

[59] Thomas H Rod, Jennifer L Radkiewicz, and Charles L Brooks. Correlated motion and the effect of distal mutations in dihydrofolate reductase. *Proc Natl Acad Sci U S A*, 100(12):6980–6985, Jun 2003.

Chapter 3

Effect of Mutation on the Cold-active α-amylase AHA

3.1 Introduction

As previously stated, AHA is one of the best characterised psychrophilic enzymes, and its qualities have made it a model test case for the study of cold adaptation [1–4]. The long-lasting availability of crystallographic structures of the psychrophilic enzyme [5] and of close heat-adapted homologues [6–9], along with the great number of in depth characterisations of its properties [10–16] all contribute to this special role. In particular, site-directed mutagenesis studies of AHA [11, 17, 18] have been among the most insightful in helping understand the relationship between activity, structural stability and flexibility. In these studies, a wide range of single and multiple mutants were constructed with the aim of restoring weak intra-molecular interactions which are present in AHA's close mesophilic homologue, PPA. The role of intramolecular weak interactions is accepted as being of prime importance in the process of thermal adaptation of proteins, from psychrophiles [19] to hyperthermophiles [20, 21]. Experimental characterisation of these mutants has shown that restoring weak interactions in AHA most often produced enzymes featuring both thermal stability and catalytic efficiency shifted towards values typical of the mesophile. The majority of studied mutants feature increases in the T_m (melting temperature) and ΔH_{cal} (calorimetric enthalpy) of unfolding, as measured from Differential Scanning Calorimetry (DSC) experi-

ments, along with higher values of the k_{cat}/K_m ratio, which results from a decrease in both k_{cat} and K_m [11, 17, 18]. On the basis of these observations, the authors proposed that structural stabilisation may result from improved rigidity of the whole molecule. Furthermore, since all introduced mutations lie outside the substrate-binding region, the observed changes in catalytic activity provide direct evidence of the underlying structural relationship between activity and stability. The authors hypothesised that the global rigidification has the effect of modulating the plasticity of the active site: reducing its flexibility, in fact, would increase the activation energy [22], leading to a reduction of the k_{cat} (see 1.1.4), and constrain the ground-state *ES* complex to a narrower distribution of conformational states, thus lowering K_m.

In this study, 7 of the previously characterised mutants featuring the most clear-cut effects on activity and thermal stability were constructed *in silico*, and used as starting structures for MD simulations to study flexibility of the mutants in comparison with the wild-type enzymes. The naming scheme of mutants is explained in Tab.3.1.

Tab.3.2 shows results of the experimental characterisation for the wild-type and selected mutant enzymes. The entity of the effects of mutations vary, and are generally stronger for multiple mutants, thus showing some degree of additivity, especially in T_m values. The mutations also have clear effects on the cooperativity of thermal folding/unfolding process: some single and double mutants (AHAQI AHATV and AHASS in Tab.3.2) cause the unfolding process to change from the typical two-state transition seen for AHA [10, 23] to a three-state transition (illustrated by the flattening of the DSC thermogram, see Fig.1.4 for a relevant example), in which two regions of the proteins (calorimetric domains) unfold at different temperatures (T_m), and feature different stability (ΔH_{cal}). It should be noted that multiple calorimetric domains are characteristic of the mesophilic and thermophilic enzymes [10]. Tab.3.2 shows that the T_m of the second domain is always higher than that of AHA, underlining the stabilisation induced by the mutations; the lower T_m of the first domain, instead, may be the consequence of strain imposed in one domain that is not compensated in other protein regions. It is interesting to note that multiple mutants containing the aforementioned mutations along with others (AHA5 and AHA5SS), show compensatory effects on stability, balancing the proposed strain to restore the two-state unfolding process. All

selected mutants show increased k_{cat}/K_m with respect to AHA, resulting from both lower K_m and k_{cat}, towards the values recorded for PPA. Two single mutants (AHAVF and AHATV) are the only exceptions to this trend, featuring k_{cat} values slightly higher than AHA.

Table 3.1: Naming scheme for selected mutants.

Name	Mutation(s)
AHANR	N12R
AHAQI	Q164I
AHAVF	V196F
AHATV	T232V
AHASS	Q58C, A99C
AHA5	N150D, Q164I, V196F, T232V, K300R
AHA5SS	Q58C, A99C, N150D, Q164I, V196F, T232V, K300R

More than 250 ns of explicit-solvent, multi-*replica* MD simulations have been performed on the 7 mutants; comparison of these trajectories with those of AHA and PPA has shown that restored weak interactions are capable of modifying the dynamic character of highly flexible regions in the direction of the mesophilic enzyme. These effects concern both the localisation and the intensity of fluctuations, and reveal that the mutations produce distal effects on flexibility which are especially significant in the active-site region. These differences are nevertheless subtle, and single out in a context of high conservation of the essential structural and dynamical features, as is expected given the high similarity of the systems under investigation.

3.2 Methods

In order to allow direct comparison of the mutants with AHA and PPA, simulations and analyses have been performed following the methods described in 2.2. Therefore, in this section only the procedures differing from those already described will be detailed.

Table 3.2: Experimental results for mutants considered.

Mutant	k_{cat} (s^{-1})	K_m (μM^{-1})	k_{cat}/K_m ($s^{-1}\mu M^{-1}$)	T_m °C	ΔH_{cal} $kcal\ mol^{-1}$
AHA	697±33	234±18	2.98	44.0	214
AHASS	340±13	112±6	3.04	1) 44.0	78
				2) 44.3	163
				tot.	241
AHAVF	752±5	223±12	3.37	45.4	234
AHANR	670±27	178±6	3.76	43.3	193
AHA5SS	331±15	86±10	3.87	51.8	277
AHAQI	514±18	118±11	4.36	1) 42.7	87
				2) 45.2	117
				tot.	204
AHATV	735±16	164±22	4.48	1) 42.7	102
				2) 45.7	104
				tot.	206
AHA5	308±8	66±5	4.67	49.4	229
PPA	291±8	65±4	4.48	1) 61.7	147
				2) 65.6	148
				tot.	295

3.2.1 Molecular dynamics simulations

MD simulations were performed using the exact same setup followed for wild-type α-amylases. The modified X-ray structure of AHA used for wild-type simulations was used as the starting point for *in-silico* mutagenesis: using PyMOL (www.pymol.org), residues were substituted to obtain the 7 mutants described in Tab.3.1. The same preparation protocol was followed, with the addition of an initial 1000 steps SD minimisation of the structure in vacuum, restraining the position of all atoms except the mutated residues with a harmonic potential. Six 6ns *replicas* were performed for each mutant, initializing the MD runs with different sets of initial atomic velocities taken from a Maxwellian distribution.

Stability and convergence of simulations were evaluated using the same procedure described for AHA and PPA. Simulations were considered to have reached equilibrium between 1ns and 2.5ns; two simula-

tions (third *replica* of AHA5SS and second *replica* of AHAQI) featuring long equilibration times were elongated to 9 ns. All analyses have been performed on just over 200 ns of equilibrated simulations.

3.2.2 Trajectory Analysis

All analyses have been carried out as described in 2.2.

The environment of mutated residues, defined as a sphere of radius 0.6 nm centred at its centre of mass, has been monitored throughout the simulations: the persistence degree of each residue in the surroundings of the mutated residue is the fraction of steps in which each residue appears inside the sphere.

The presence of salt bridges was monitored during the simulations; a salt bridge is defined as a pair of oppositely charged amino acids whose charged groups are less than 0.4 nm apart [24] for more than 35% of the simulation. The persistence cutoff was chosen very restrictive to overcome the strong similarities in ionic-interaction patterns among the enzymes under investigation. Salt bridges have been represented on 3D structures by connecting the C_α atoms of the interacting residues with lines.

The root mean square fluctuation (RMSF) per residue was calculated on C_α atoms, after projection of the trajectories on the *essential subspace*; the number of eigenvectors spanning the *essential subspace* is 9 for AHAQI, 10 for AHANR and AHA5SS, 11 for AHA5 and AHASS, 13 for AHAVF and AHATV.

3.3 Results

The structure of the studied α-amylase has been described in 2.1 as being composed of three domains, of which domain A is the central domain featuring a $(\beta/\alpha)_8$barrel structure made up of 8 core β-strand (β_1 to β_8) arranged in a cylinder-like structure surrounded by 8 solvent-exposed α-helices (α-helix1 to α-helix8). At the centre of the barrel lies the chloride-binding domain, and catalytic site, surrounded by the long loops connecting the strands with the helices ($\beta-\alpha$-loops). The longest of these loops ($\beta_3-\alpha_3$) is considered the second domain in α-amylase structure, domain B: the calcium binding site is found at the interface between the first two domains, that represent the two sides

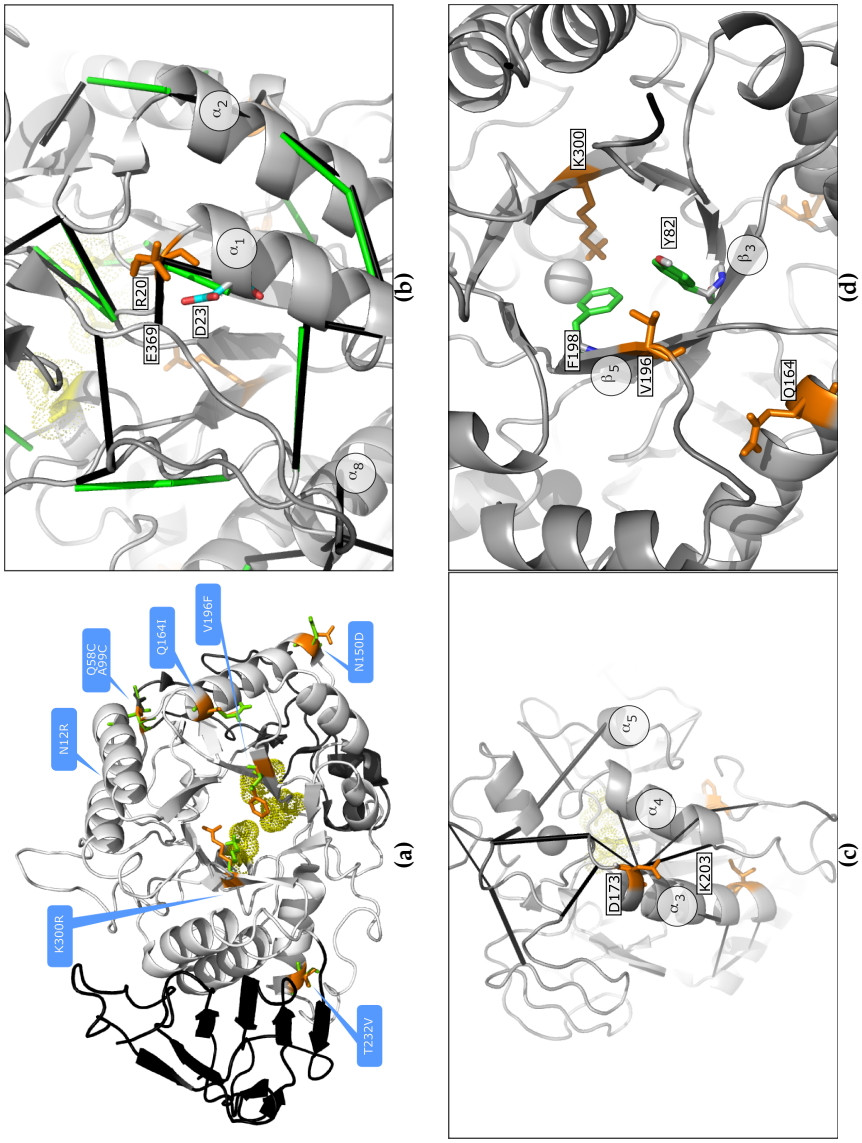


Figure 3.1: Three-dimensional localisation of mutations: (a) all studied mutations, (b) R12 (orange), D15 (cyan) and salt bridges of PPA (black) and AHANR (green); (c) D150 (orange, top) and salt bridges of PPA (black); (d) V196 (orange, center), Y82 and F198 (green).

of the substrate-binding site. The third domain, domain C, lies on the opposite side of domain A with respect to domain B. The position of the mutations on the 3D structure of AHA is shown in Fig.3.1a.

The first step in this study is to evaluate whether the mutations have effectively restored the targeted weak interactions, by monitoring the persistence of residues in the surroundings of mutations (see 3.2). The mutations K300R, A99C and Q58C have not been considered in this analysis, since they are meant to change the coordination of the chloride ion (K300R), and to engineer a disulphide bridge present in PPA.

3.3.1 Environment of the Mutations

3.3.1.1 N12R

The N12R* mutation restores a salt bridge between R12 ($R20^{\mathbf{P}}$), on loop $\beta_1 - \alpha_1$, and D15 ($D23^{\mathbf{P}}$), at the beginning of α-helix1. The two residues are located on the surface of the enzyme, and their side-chains face the same direction, pointing out towards the solvent (Fig.3.1b). In PPA $R20^{\mathbf{P}}$ forms a salt bridge with both $D23^{\mathbf{P}}$ and $E369^{\mathbf{P}}$, located at the C-terminal end of loop $\beta_8 - \alpha_8$, and interacts via the aliphatic carbons of its side-chain with other hydrophobic residues of $\beta_8 - \alpha_8$ ($366 - 370^{\mathbf{P}}$, see Fig.3.2a). The R12-D15 salt-bridge in the AHANR mutant is as persistent as its corresponding bridge in PPA; local hydrophobic interactions are also restored (with V318 and V320), while a hydrogen bond with Q66 present in AHA is lost (Fig.3.2a). The other salt bridge present in PPA cannot be established due to the Glu to Pro substitution, leaving the region less rigid.

3.3.1.2 N150D

The N150D mutation restores a surface salt bridge between D150, at the beginning of α-helix3, and K190 on the end of α-helix4. In PPA, the corresponding $D173 - K213^{\mathbf{P}}$ bridge represents one end of a complex network of 9 salt-bridges which extends through domain B (Fig.3.1c), connecting it with α-helix3, α-helix4, α-helix5 and loop $\alpha_4 - \beta_5$. The network relies especially on the hub role of $R176^{\mathbf{P}}$ (α-helix3), which establishes 4 salt bridges, the strongest of which with $D173^{\mathbf{P}}$. In AHA a

*amino-acid numbering follows the sequence of AHA and its mutants, where not otherwise specified

number of substitutions, and particularly R176$^{\mathbf{P}}$ to Q153, disallow the formation of this network. Nevertheless the two mutants bearing the N150D mutation (AHA5SS and AHA5), feature a D150-K190 salt bridge which is even more persistent than its corresponding in PPA (Fig.3.2b), possibly because of the absence of the other interactor.

3.3.1.3 Q164I

The Q164I mutation reinforces a hydrophobic cluster present between the end of α-helix3 and the beginning of β_4, which also involves $\alpha_3 - \beta_4$ and $\alpha_4 - \beta_5$. Q164, and the corresponding I187$^{\mathbf{P}}$ is located at the end of α-helix3, and its side-chain points towards the β-barrel. The region of the mutation features very high surrounding hydrophobicity (H_p) even in AHA, where the presence of a polar sidechain in an apolar environment is expected to have a destabilising effect. Clear increases in H_p are seen in the mutants which bear this mutation (AHAQI, AHA5SS and AHA5, see Fig.3.4) both locally (α-helix3, 160-170) and far ($\alpha_4 - \beta_5$, 190-200), resulting in a global increase of the percentage of residues with high surrounding hydrophobicity ($H_\mathrm{p} > 20 kcal\ mol^{-1}$) at the expense of the medium-high ($15 kcal\ mol^{-1} < H_\mathrm{p} < 20 kcal\ mol^{-1}$, Fig.3.3).

3.3.1.4 V196F

Mutation V196F is aimed at recreating two aromatic interactions between F196, located at the beginning of β_5, and residues Y82 and F198, on β_3 and β_5 respectively. The three residues all point towards the inside of the β-barrel, and are located not far away from residue Q164 (Fig.3.1d). In PPA, the corresponding F229$^{\mathbf{P}}$ is part of a large cluster of aromatic and hydrophobic residues which interact with the N-terminal (Fig.3.2d); the same does not hold for AHA, where the N-terminal is 8 residues shorter. The mutation is expected to have an effect on the packing of residues in its surrounding, owing to the greater size of the phenylalanine sidechain and its greater hydrophobicity; mutants bearing this mutation (AHAVF, AHA5SS and AHA5) show a local increase in the surrounding hydrophobicity (Fig.3.4).

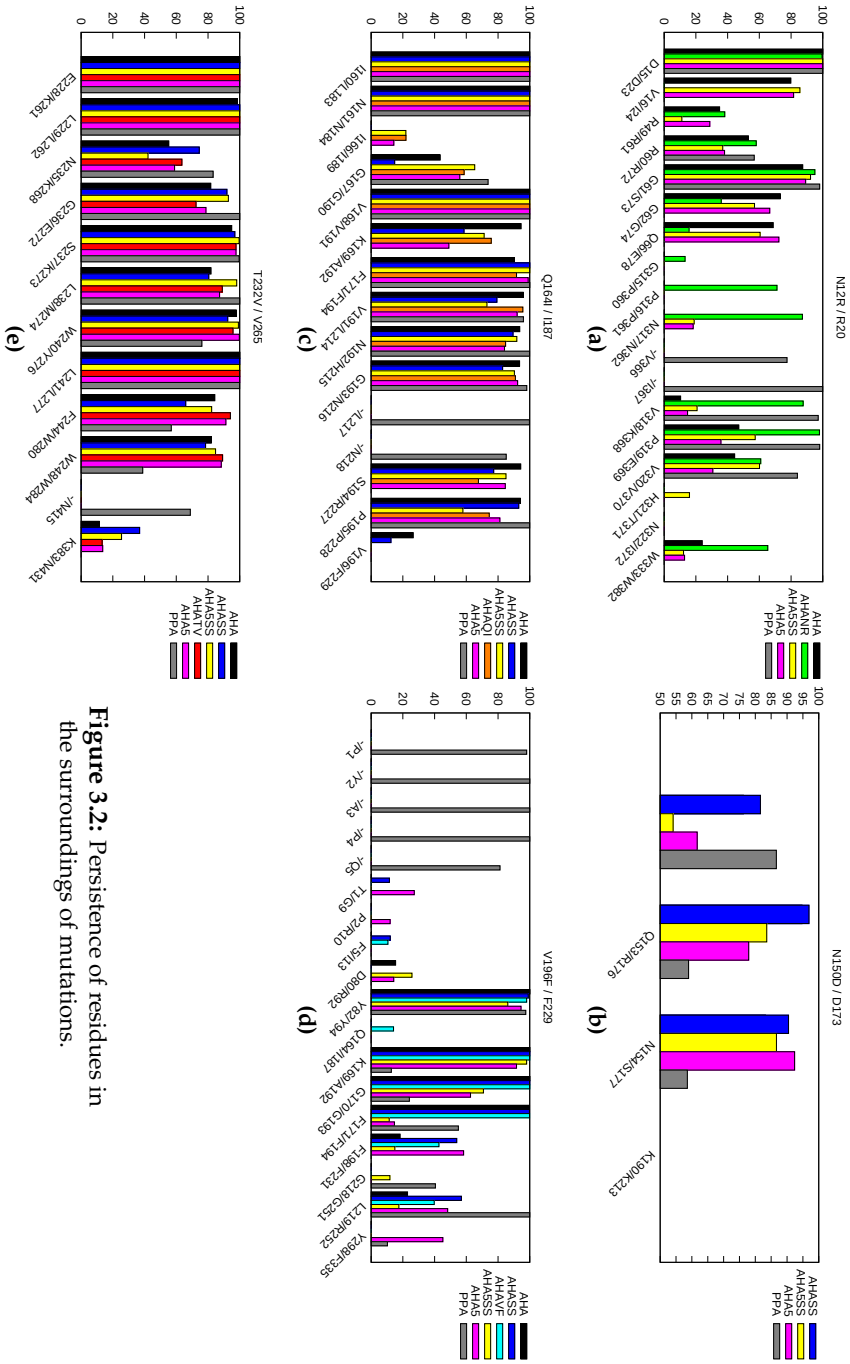


Figure 3.2: Persistence of residues in the surroundings of mutations.

3.3.1.5 T232V

The T232V mutation strengthens a hydrophobic cluster between α-helix6 and $\alpha_6 - \beta_7$, not far from the interface between domain A and domain C; the mutated residue is located at the top of α-helix6, its side-chain pointing towards the beginning of $\alpha_6 - \beta_7$. Mutants bearing the mutation (AHATV, AHA5SS and AHASS) show an increase in the persistence of hydrophobic residues in the surroundings of the mutation, especially residues L238, F244 and W248 which are located at the beginning, middle and end of $\alpha_6 - \beta_7$ (Fig.3.2e). These differences correlate with local increases in the surrounding hydrophobicity, especially in regions 230-240 and around 250 (Fig.3.4).

3.3.2 Overall structural properties

Few differences are seen when comparing secondary structural elements of the two wild-type enzymes; in the same way, mutants generally show very similar secondary structure content to AHA (Tab.3.3). The few clear-cut difference involve domain C, where all mutants show a greater percentage of residues in *bend* conformation, at the expense of β structures, suggesting a general destabilisation of secondary structural elements in the C-terminal domain.

Furthermore, Fig.3.3 shows that AHA and its mutants display very similar content of hydrogen bonds, and a similar distribution of surrounding hydrophobicity. All mutants except AHAVF show a slightly reduced content of residues with low surrounding hydrophobicity, especially AHA5SS, and an increase in residues with medium-low H_p; these differences can be chiefly ascribed to domain A. AHAVF, AHATV, AHA5 and again especially AHA5SS show a clearly greater content of residues with medium-high H_p, while only AHAQI and AHAVF shows a some difference in content of residues with high H_p, which in the former is larger, and in the latter smaller than that of AHA.

These observations point towards the fact that mutants may have strengthened their hydrophobic core, but only AHA5SS, and to a lesser extent AHA5, show clear-cut differences with AHA in the direction of values found in PPA.

Considering the surrounding hydrophobicity on a per-residue basis, as in the profile in Fig.3.4, confirms the observations made on the basis of the distributions, and allows to localise differences with respect

Table 3.3: Average secondary structure content of the wild-type and mutant enzymes.

		$\bar{x} \pm \sigma$(%)						
		AHA	AHA5S	AHA5SS	AHAQI	AHATV	AHA5	PPA
Prot.	*Coil*	21.6±1.3	21.4±1.4	21.7±1.6	21.9±1.3	21.4±1.3	21.5±1.3	23.9±1.3
	β-sheet	21.1±1.7	19.9±1.9	20.0±1.7	19.3±1.9	20.0±1.7	19.3±1.6	20.8±1.8
	β-bridge	3.5±0.8	3.7±0.9	3.3±0.8	3.8±1.0	3.2±0.9	3.7±0.8	3.1±0.7
	Bend	16.4±1.3	16.7±1.3	16.4±1.4	16.6±1.4	16.2±1.5	16.9±1.4	15.2±1.4
	Turn	10.7±1.5	11.0±1.7	11.7±1.8	10.9±1.5	11.4±1.5	10.7±1.6	13.4±1.4
	α-helix	23.8±1.0	24.3±1.3	23.5±1.2	24.2±1.2	24.2±1.0	24.2±1.2	20.2±0.8
Dom A	*Coil*	22.4±1.2	22.0±1.4	22.2±1.5	22.0±1.4	22.1±1.4	22.2±1.3	25.4±1.4
	β-sheet	13.3±1.5	13.0±1.6	13.5±1.5	12.9±1.4	13.2±1.6	12.7±1.4	15.3±1.9
	β-bridge	1.9±0.7	1.8±0.7	1.3±0.8	2.1±0.7	1.4±0.8	1.8±0.8	1.6±0.8
	Bend	16.5±1.5	15.5±1.6	14.8±1.6	15.4±1.7	15.0±1.5	15.4±1.6	11.7±1.6
	Turn	9.5±1.9	10.9±2.1	11.6±2.2	10.5±2.0	11.1±1.9	10.2±2.1	13.2±1.7
	α-helix	33.3±1.5	33.6±1.8	32.9±1.8	33.4±1.8	33.2±1.3	33.6±1.5	28.6±1.1
Dom B	*Coil*	30.5±4.5	29.8±4.8	29.8±4.7	31.3±4.2	29.2±4.3	29.5±5.0	32.1±3.2
	β-sheet	13.3±5.3	14.2±5.5	15.0±5.0	13.8±5.3	15.0±5.5	14.2±5.8	5.5±3.5
	β-bridge	10.2±3.8	10.2±3.8	9.3±3.8	9.8±4.2	9.5±4.3	9.7±4.2	9.0±2.7
	Bend	18.0±4.5	16.7±4.0	18.8±4.2	15.8±3.7	16.0±4.2	17.2±4.3	27.7±3.8
	Turn	10.5±4.7	10.0±4.3	9.8±4.5	10.5±4.5	10.3±4.3	10.8±4.7	17.9±4.2
	α-helix	13.7±4.7	15.3±4.5	13.2±4.7	16.0±3.8	17.0±3.3	14.5±4.3	7.0±2.5
Dom C	*Coil*	22.1±4.1	22.6±3.7	23.0±3.7	23.0±4.0	22.8±4.0	23.2±3.4	19.2±4.0
	β-sheet	50.8±6.2	44.7±6.0	43.0±5.4	42.8±6.0	44.2±5.9	42.6±4.8	52.5±6.2
	β-bridge	2.0±1.6	3.3±1.5	3.9±1.5	3.8±2.0	3.3±1.6	3.8±1.5	1.4±1.1
	Bend	12.0±3.0	17.6±3.0	17.3±3.3	18.2±3.0	17.1±3.3	18.4±2.9	16.8±3.8
	Turn	12.9±2.8	10.7±2.6	11.4±3.3	11.0±2.9	11.5±2.7	11.0±2.7	8.9±2.4
	α-helix	0.0±0.1	0.0±0.4	0.0±0.4	0.0±0.4	0.1±0.5	0.1±0.5	0.0±0.0

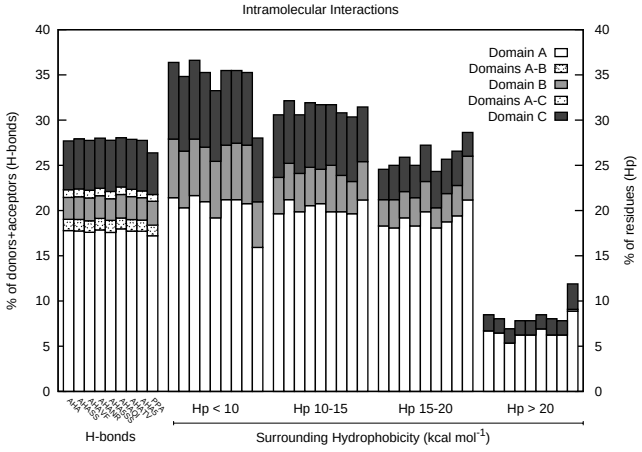


Figure 3.3: Hydrogen bonds and surrounding hydrophobicity (H_p) are described as histograms representing the percentage of donors+acceptors involved in H-bonds (first set, y axis on the left) or of residues with surrounding hydrophobicity in the indicated range (H_p, all other sets, y axis on the right); each set shows values computed for AHA, PPA and the mutants, and plotted as a function of the sequence of AHA (top labels) and PPA (bottom labels).

to mutations. Differences between AHA and the mutants are found in the surroundings of those mutations aimed at strengthening hydrophobic clusters, and will be discussed in further detail (see 3.4). Otherwise, the profiles are generally very similar between mutants and AHA, and comparable to that of PPA, showing that regions featuring high hydrophobicity are conserved, and are probably encoded in the structure of the enzymes. Generally, high H_{p} values are found in β-strand and, to a lesser extent, on the side of α-helics facing the β-barrel which lies at the core of these α-amylases' fold; α-helix7 and α-helix8 are the only two α-helix featuring high H_{p} values throughout their structure. This suggests that these two structural elements may act as a connection between the barrel's β-strand and domain C, creating a single hydrophobic core that runs through the entire protein structure.

Greater heterogeneity uncovers in the comparative analysis of flexibility: the RMSF profiles shown in Fig.3.4 reveal that mutants exhibit a very wide range of fluctuational behaviours, especially in correspondence of consistent peaks of the wild-type profiles. Generally, as seen for wild-type enzymes, stable secondary structure elements feature low RMSF, so that conservation of secondary structure reflects in the conservation of their dynamic properties. Analysis of the single profiles of mutants shows that the most relevant fluctuations cluster in the $\beta - \alpha$ loops that surround the active site, suggesting that this is a robust feature encoded in these α-amylases structure. Mutants are nevertheless, on average, slightly more flexible than AHA, but consistently throughout the sequence, so that their profiles appear to be more delocalised. Relevant differences in RMSF will be discussed in detail in the next section.

3.4 Discussion

The discussion of the most relevant differences between the studied enzymes will be carried out by taking into consideration separately the three domains that compose the α-amylase structure, and the interfaces between them. In the following, the residue numbering of AHA, and consequently of mutants, will be used if not otherwise specified.

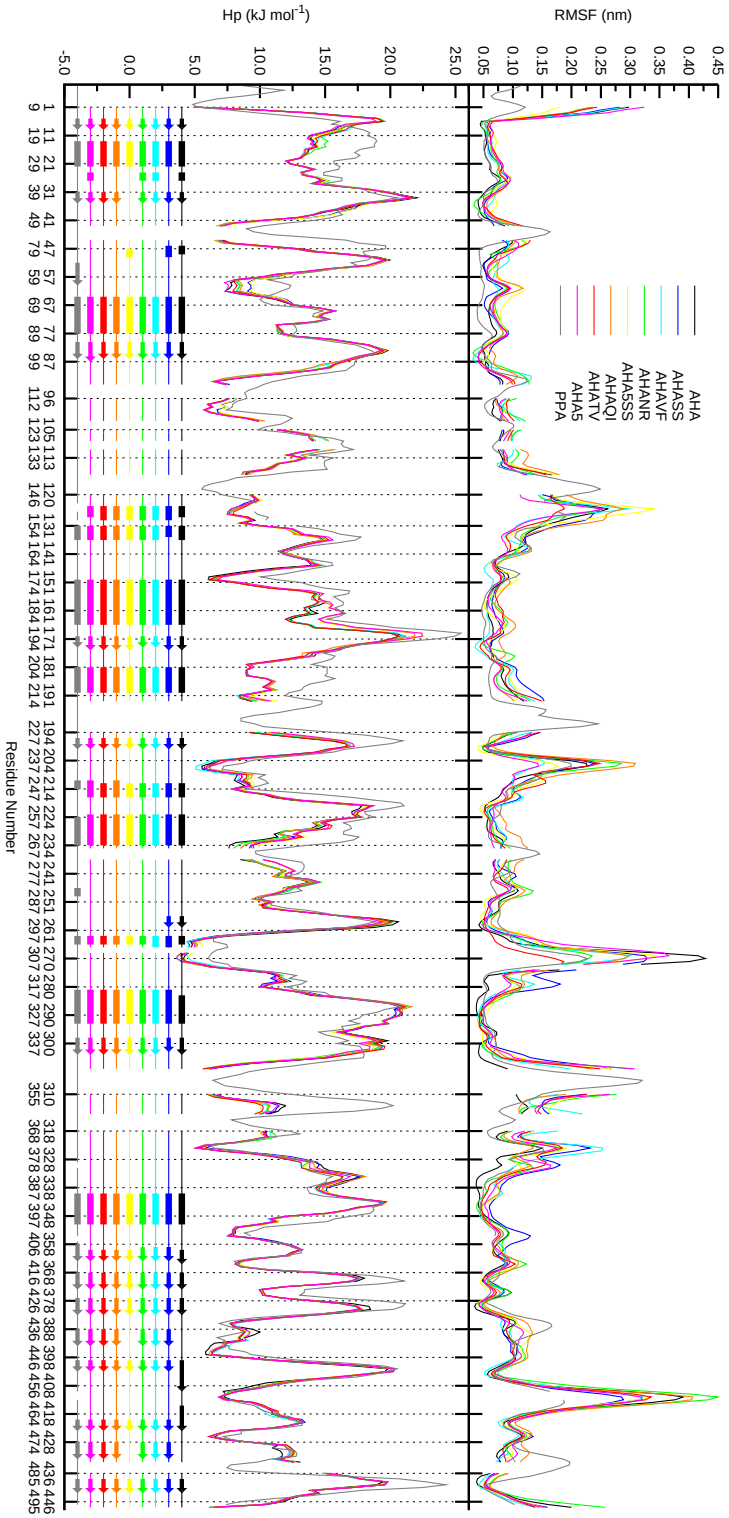


Figure 3.4: RMSF and surrounding hydrophobicity profiles, smoothed using window-averaging (window-size=5 residues). A schematic representation of the most frequently attained secondary structure is shown in the lower plot.

3.4.1 Domain A and the active site

Domain A is the central and largest domain of α-amylases; it accomodates the active site at the heart of its β_8/α_8 barrell structure, while the loops connecting β-strand and α-helix (loops $\beta_1-\alpha_1$ to $\beta_8-\alpha_8$) form the substrate-binding cleft.

It has been suggested that $\beta-\alpha$ loops have an important role in substrate recognition and processing in these and other homologous α-amylases; in particular loop $\beta_8-\alpha_8$ may be one of the first interactors with the incoming substrate, mediating enzymatic specificity [5], and loop $\beta_7-\alpha_7$ may rearrange upon substrate binding and release, contributing to its correct placement inside the active site. In fact, $\beta_7-\alpha_7$ is located in the immediate surroundings of the active site, few positions in sequence away from one of the catalytic residues, while $\beta_8-\alpha_8$ extends further away from the catalytic triad and above $\beta_7-\alpha_7$.

Loop $\beta_7-\alpha_7$ is well conserved, both in structure and sequence, among chloride-dependent α-amylases [25], while loop $\beta_8-\alpha_8$ has variable length in α-amylases from different species, and therefore displays different conformations in the respective crystal structures [5]. In PPA, $\beta_8-\alpha_8$ is 12 residues longer than in AHA owing to two insertions, both of which adopt β-hairpin structures.

L7 (residues 260-280) is the part of loop $\beta_7-\alpha_7$ that protrudes towards the active site, between β-strand7 and the short α-helix that carries the catalytic D264 (Fig.3.5a). As previously shown, it is the most flexible part of AHA, and the region showing the greatest differences between the two wild-type enzymes: AHA has a peak with a wide base (10 residues) and a sharp apex (G270); PPA has a much lower peak, with a wide apex ($308-311^{\mathbf{P}}$, see Fig.3.4 and Fig.3.5c). We have previously shown that differences in the localisation of high flexibility in this region result in AHA being more flexible than PPA in the immediate vicinity of the active site 2.3.3.

The studied mutants exhibit high flexibility in the **L7** region, in correspondence with the peak in the wild-type enzymes, although the height, the position and the shape of the peaks vary (Fig.3.5c). All mutants display less flexibility than AHA, and often amino-acid residues showing peak fluctuations in the region are located one or two positions further towards the N-terminal with respect to AHA. Nevertheless, considering the entire $\beta_7-\alpha_7$ loop, flexibility of the mutants is com-

parable to that of AHA, or even enhanced (in the case of multiple mutants and AHAQI), but more evenly distributed throughout the entire region. In fact, flexibility profiles of mutants show two minor peaks, not present in the profiles of AHA and PPA, located few positions in sequence before and after the main peak, corresponding to two short α-helical segments.

Multiple mutants show higher flexibility than the other mutants in the main peak, while AHANR and AHATV show very low flexibility and broad peaks, the former showing even lower flexibility than PPA. It is interesting to note that mutants displaying two calorimetric domains mutants (which will be referred to as ΔH_2, see Tab.3.2), are characterised by "broken" peaks in correspondence of residues (H269 for AHAQI and AHATV and G271 for AHASS) which are conserved among the two wild-type forms, and are thought to make hydrogen-bonded contacts with the substrate [6].

All mutants except AHATV show a secondary peak around N265 that corresponds with the short α-helix that carries the catalytic D264.

All mutants also show a lower peak in the 274-282 region, involving the first residue of α-helix7. This α-helix is 2 residues longer in all mutants, as it is in PPA: while in AHA the 2 residues (G281 and R282) adopt a bend conformation with high persistence ($\sim 80\%$), in the mutants they adopt α-helical structure with lower persistences ($\sim 45\% \rightarrow 80\%$). The highest stability of the α-helix is that of AHAQI, and might be explained by the higher persistence of a network of 6 salt bridges involving E279, D280, R282 and D285, which connect the final part of $\beta_7 - \alpha_7$ and the first residues of α-helix7 to $\beta_8 - \alpha_8$ and domain C (Fig.3.5b), and which is present only in part in the other enzymes.

On the other hand, the least stable α-helices are those of AHA5SS and AHASS: for the latter this instability correlates with its displaying the highest and broadest of flexibility peaks in this region. The case of AHA5SS is even more striking considering that it carries the mutations of both AHAQI and AHASS, and nevertheless displays such a different behaviour from both, confirming the supposed presence of compensating effects of multiple mutations [26].

Just above, and in direct contact with **L7**, **L8** (residues 305-320) is the part of $\beta_8 - \alpha_8$ that protrudes towards the active site; it encompasses both the insertions that make $\beta_8 - \alpha_8$ longer in PPA with respect to AHA (Fig.3.5). In particular, the first and longest of the two inser-

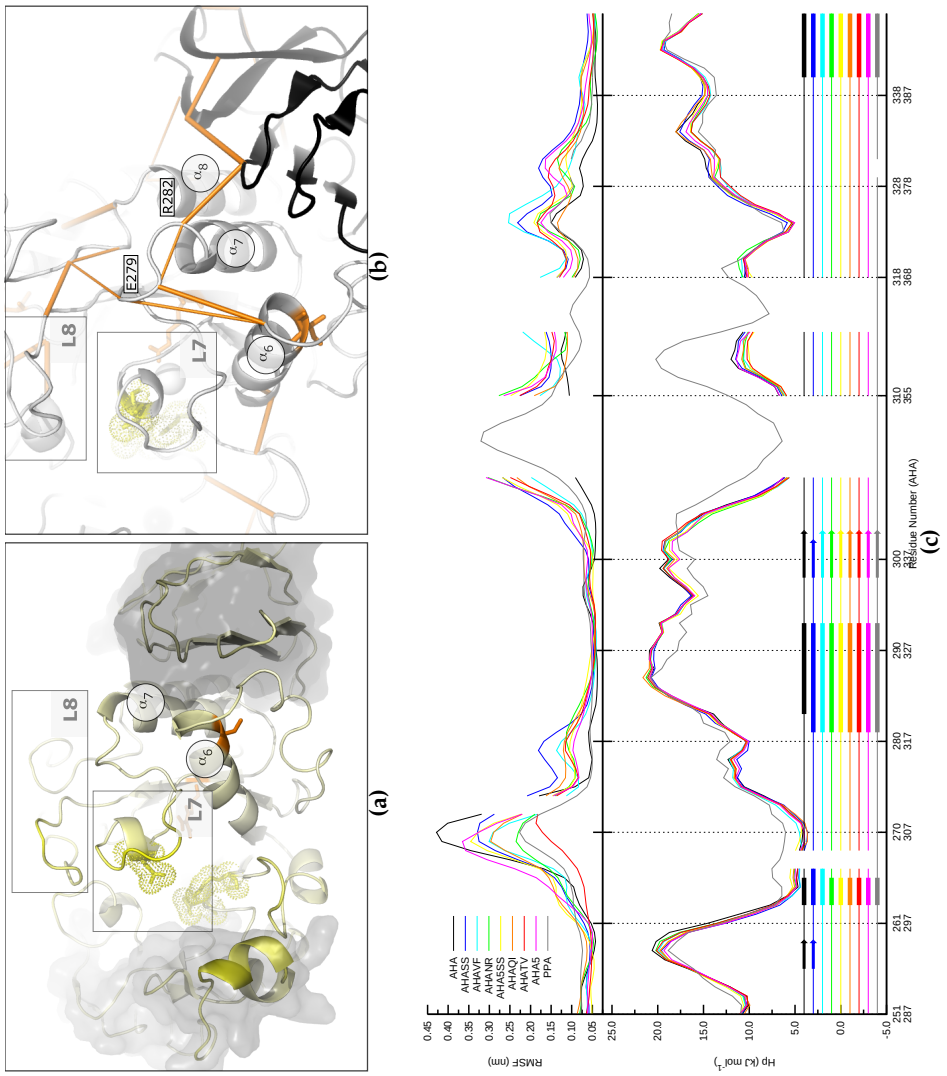


Figure 3.5: (a) Side view of AHA showing the position of **L7** and **L8**; domain B (left) and domain C (right) are shown as surface; the catalytic residues are shown as yellow balls and sticks. (b) Network of salt bridges around the beginning of α-helix7 in AHAQI. (c) RMSF and H_{p} profiles of the L7/L8 region.

tions, and the nearest to the active site, is the most flexible region of PPA. The studied mutants show a global increase in flexibility in this region, in particular in correspondence with the peaks in PPA. A first peak, around 305-312, corresponds to the first insertion: while AHA is very rigid, all mutants display higher fluctuations, and in particular AHA5 and AHANR show a flexibility peak which is comparable in height to that of PPA (Fig.3.5c).

The two wild-type enzymes show similar flexibility at a second peak around residue G324: in this region all mutants show increased fluctuations, and a secondary, lower peak, around A330, where a significant decrease of the surrounding hydrophobicity is evident for all mutants. AHASS is the most flexible in both peaks, while AHAVF, which is nearly as flexible in the first peak, totally lacks a second one. The other two ΔH_2 mutants interestingly display two peaks which are very similar in height, and AHATV displays a "broken" peak (in correspondence of A330).

Given the difference in length and structure between $\beta_8 - \alpha_8$ in PPA and AHA, it is striking to notice that mutants manage to attain a flexibility similar to that of the mesophile, especially since all mutations are located far from the region, exposing the ability of single amino-acidic changes to modulate significantly the flexibility of distant regions of AHA.

3.4.2 Domain B and the interface with domain A

Domain B is the longest $\beta - \alpha$ loop in both AHA and PPA (44 and 67 residues long, respectively); it connects β-strand3 and α-helix3, and although its starting and end points are rather near in the structure, domain B extends making contacts with $\beta - \alpha$ loops 2, 4 and 5, covering one side of the central barrel and closing the substrate-binding cleft.

On one side of domain B, where it interacts with $\beta_2 - \alpha_2$, are located residues Q58 and A99, which correspond to residues C70$^{\mathbf{P}}$ and C115$^{\mathbf{P}}$ in the structural alignment. These two cysteines form one of the five disulphide bridges in PPA; this bridge is lost in AHA, and engineered back in mutants AHASS and AHA5SS. In these mutants, the bond bridges the end of $\beta_2 - \alpha_2$ and domain B in correspondence with the loop between a high-persistence β-hairpin. These mutations have subtle local effects on surrounding hydrophobicity, and cause small changes on

the flexibility of the two residues, which feature low fluctuations in all studied enzymes. A stronger increase of H_p is produced on the residues of $\beta_2 - \alpha_2$ that lie nearest to the disulphide bridge (P37-I41), possibly acting to stabilise a nearby short α-helix (W46-Y50) (Fig.3.6a) which is destabilised or adopts a 3-helix conformation in PPA and other mutants. This stabilising effect could be mediated by the stabilisation of a short β-strand (E39-H40), which is directly affected by the increase in H_p and which is part of a β-hairpin structure that encompasses the loop that carries the α-helix (Fig.3.6a). The persistence of the α-helix correlates with the stability of the two β-strand, and AHASS and AHA5SS are the only two mutants featuring β and α-helical structures in this region as stable as they are AHA. The β-hairpin in AHA5SS is even more stable than in AHA, suggesting a role for compensating effects on the structural and dynamic properties in this region.

On the other side, domain B interacts with $\beta_4 - \alpha_4$, with which it forms the calcium binding site, and with $\beta_5 - \alpha_5$ (Fig.3.6b).

L5 (residues 200-210) is the part of loop $\beta_5 - \alpha_5$ that protrudes towards the active site. **L5** is preceded by the catalytic E200, which features very low fluctuations in all analysed simulations and, 5 residues further back in the sequence, by the $\alpha_4 - \beta_5$ loop which is longer and much more flexible in PPA. A hydrophobic cluster between α-helix3, $\alpha_3 - \beta_4$, β_4 and $\alpha_4 - \beta_5$ is reinforced by mutations Q164I and V196F (Fig.3.4), while mutation N150D allows the establishment of the D150-K190 salt bridge between α-helix3 and α-helix4. Accordingly, mutants bearing all three these mutations (AHA5 and AHA5SS) show the least flexibility in the **L5** region (Fig.3.6c). Interestingly though, AHAVF also shows similar flexibility, indicating that the effect of local hydrophobic core strengthening has fundamental importance in determining rigidification. The other two mutations have the effect of displacing peak fluctuation one or two positions in sequence towards the C-terminal, causing the 190-200 region to be more rigid, owing to salt-bridge formation on one hand, and to an increase in H_p in this region which is only present in mutants bearing the Q164I mutation.

3.4.3 Domain C and the interface with domain A

On the opposite side of domain A to where domain B is located, a domain composed of 8 β-strand structured in a greek-key motif is

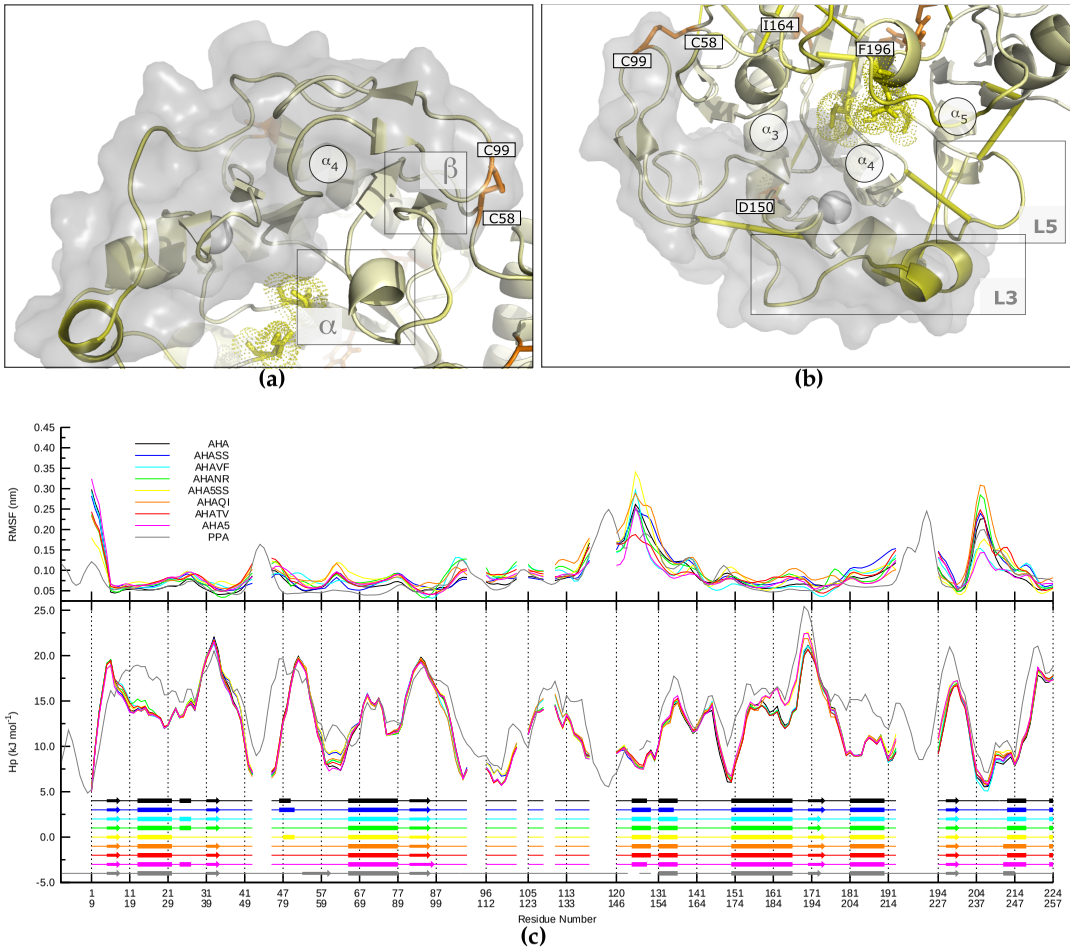


Figure 3.6: (a)Secondary structure around the engineered disulphide bridge. The W46-Y50 α-helix (α) and the E39-H40 β-strand (β) are highlighted; (b) Structure of **L3** and **L5**. (c) RMSF and Hp profiles of the domain B region.

tightly packed against α-helices 6, 7 and 8 of the $(\beta/\alpha)_8$ barrel; its interface with domain A is rich in aromatic and hydrophobic residues which contribute to their binding, and mediate the extension of the hydrophobic core of domain A towards domain C.

Domain C is very similar between AHA and PPA, both in terms of structure and of dynamical properties: the main differences between the wild-type enzymes are found in the regions where also the mutants show the greatest divergences.

The sandwich-like structure of domain C features in both AHA and PPA strong asymmetry between the its two composing β-sheet: the buried β-sheet constitutes the interface with domain A and is stable and ordered, while the solvent-exposed β-sheet is highly disordered, and it is split in 2 separate 2-stranded β-sheet featuring few mainchain hydrogen-bonded contacts between them (marked $\beta 1$ and $\beta 2$ in Fig.3.7a,(b)); $\mathbf{C}\beta_4$ and $\mathbf{C}\beta_7$ (388-393 and 428-433) make up the $\beta 1$ β-sheet. The two strands are highly persistent in PPA, while in AHA both are less ordered; in particular some residues at the centre of the sheet adopt β structuring with very low persistence, owing to the lack of mainchain hydrogen-bonded contacts between the two strands. Interestingly, all mutants display ordered β-strand, featuring a very similar hydrogen-bonding pattern to that of PPA; the only exception is AHA5SS, which has a similar behaviour to that of AHA. These observations are unexpected in two ways: first, this region doesn't carry any of the studied mutations, and nevertheless shows great structural and dynamical differences in mutated forms, demonstrating that point mutations have global effects, affecting regions very distant from the site of mutation. Secondly, the very different behaviour of AHA5SS from that of mutants bearing only some of its mutations confirms the importance of compensatory effects of multiple mutations.

The $\beta 2$ sheet, composed of $\mathbf{C}\beta_5$ and $\mathbf{C}\beta_6$ (399-409 and 415-423), adopts a strange distorted shape in AHA, and encompasses a short loop featuring the highest flexibility in domain C. The β-strands are instead broken in PPA in correspondence to their bending point in AHA, so that two distinct β-sheet form, roughly at $^\circ 90$ one from the other. Interestingly, all mutants show very similar behaviour to PPA. The short loop is also less flexible in PPA than in AHA, and shorter by 2 residues; all mutants also display a reduction in flexibility with respect to AHA, except for AHANR and AHAQI. It is also interesting to note that in all

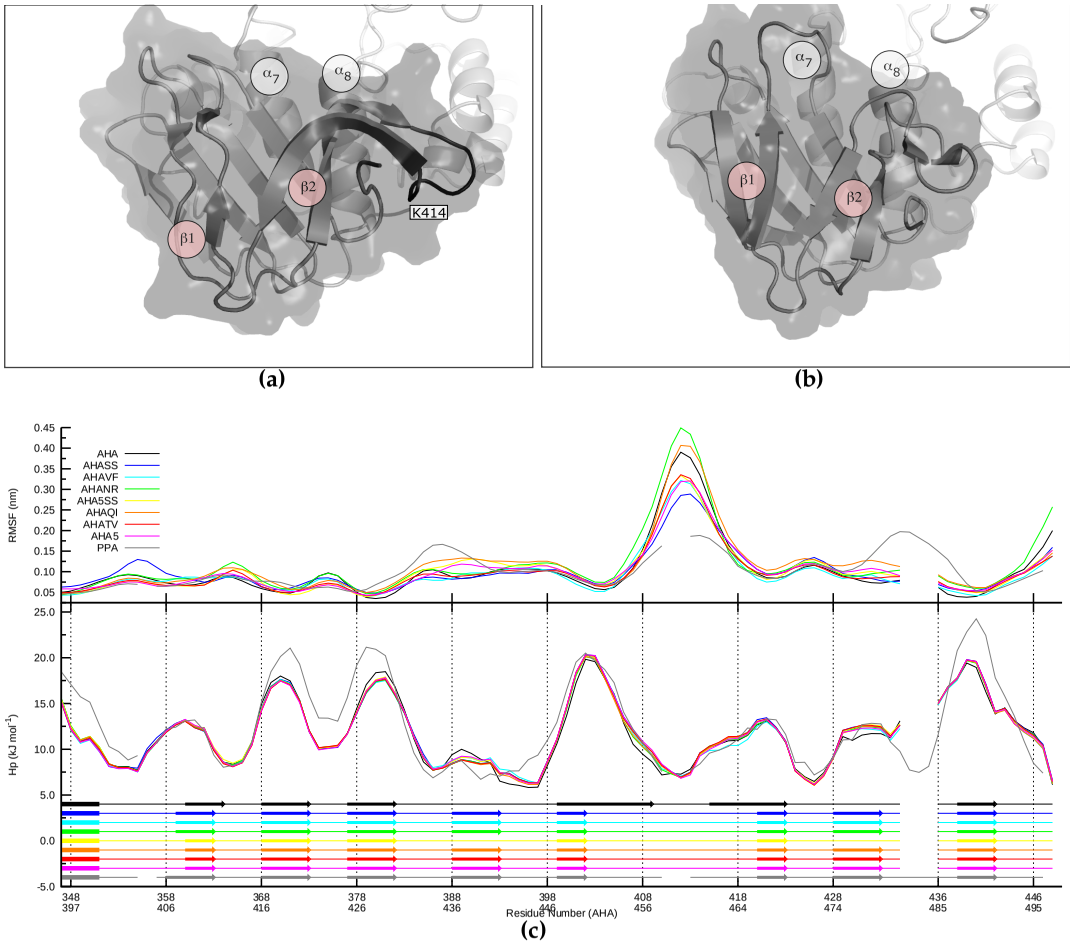


Figure 3.7: Secondary structure of domain C in AHA (a) and PPA (b); the two strands described in the text are highlighted in red to distinguish them from the barrel's β-strand. (c) RMSF and Hp profiles of domain C.

mutants, residue K414 of this loop establishes a stable salt bridge with the negatively charged C-terminal: the higher flexibility of AHANR corresponds to a lower persistence of this interaction.

3.5 Conclusions

Intramolecular weak interactions have a fundamental role in stabilising protein structures. Special attention has been given to their role in the context of thermal adaptation of proteins [19–21]. On the basis of the experimental characterisation available of mutants of AHA aimed at restoring weak interactions present in its close mesophilic homologue PPA, this study is endeavoured in providing molecular explanation to the observed variations in thermal stability and catalytic activity of these mutants with respect to the wild-type enzymes.

The analyses presented have shown that substitutions are capable of eliciting effects in agreement with the restoring or strengthening of target interactions, modifying the dynamic environment of mutated residues towards that of corresponding residues in the mesophilic enzyme. However, fundamental differences both in the structure and the sequence of the two wild-type enzymes make it impossible for these mutations alone to reproduce the full set of stabilising local interactions seen in PPA.

Nevertheless, these local effects are able to modify the dynamic character of the mutants in a global sense, producing complex distal effects that underlie the intimate interplay of weak interactions and hydrophobicity with flexibility and secondary structure. Considering the minor entity of the modifications introduced, substituting at the most 7 residues in a large enzyme made up of 448 residues, it is striking to observe the wide range of different dynamical behaviours these mutants exhibit. Differences among mutants, and between mutants and AHA are in fact at least of the same size, if not greater than those observed between AHA and PPA. This observation suggests two considerations: on one hand, the dynamic properties of the two wild-type enzymes are extremely similar, as confirmed by the fact short MD simulations were unable to distinguish significant differences between the two [27]. Secondly, these mutations are extremely effective in modifying the global properties of AHA, as demonstrated by their experimental characterisation [11, 17], showing that in the case of multiple mutants the "con-

version" of activity and thermal stability to mesophilic values is almost complete.

This ability of the mutations to elicit in the psychrophilic enzyme behaviour similar to the mesophilic is often confirmed in the results of MD simulations, albeit the comparison is restricted exclusively to structural and fluctuational properties. We see that the mutants exhibit secondary structure persistence, flexibility and, to a lesser extend, hydrophobicity values between those seen for the two wild-type enzymes. In particural, the effect of flexibility is largely apparent in regions featuring high flexibility in AHA and PPA, which chiefly cluster in regions near the active site and substrate-binding groove.

In conclusion, the distal character of the effects of introduced mutations raises a serious challenge for modelling techniques to interpret their underlying mechanical bases. Nevertheless these results provide atomic-detailed evidence of the ability of the selected mutations to globally modulate the dynamic properties of AHA, and offer unprecedented insight in the way this modulation takes place in a large protein.

Bibliography

[1] Salvino D'Amico, Paule Claverie, Tony Collins, Daphné Georlette, Emmanuelle Gratia, Anne Hoyoux, Marie-Alice Meuwis, Georges Feller, and Charles Gerday. Molecular basis of cold adaptation. *Philos Trans R Soc Lond B Biol Sci*, 357(1423):917–925, Jul 2002.

[2] G. Feller. Molecular adaptations to cold in psychrophilic enzymes. *Cell Mol Life Sci*, 60(4):648–662, Apr 2003.

[3] D. Georlette, V. Blaise, T. Collins, S. D'Amico, E. Gratia, A. Hoyoux, J-C. Marx, G. Sonan, G. Feller, and C. Gerday. Some like it cold: biocatalysis at low temperatures. *FEMS Microbiol Rev*, 28(1):25–42, Feb 2004.

[4] Khawar Sohail Siddiqui and Ricardo Cavicchioli. Cold-adapted enzymes. *Annu Rev Biochem*, 75:403–433, 2006.

[5] N. Aghajari, G. Feller, C. Gerday, and R. Haser. Structures of the psychrophilic Alteromonas haloplanctis alpha-amylase give

insights into cold adaptation at a molecular level. *Structure*, 6(12):1503–1516, Dec 1998.

[6] M. Qian, R. Haser, G. Buisson, E. Duée, and F. Payan. The active center of a mammalian alpha-amylase. Structure of the complex of a pancreatic alpha-amylase with a carbohydrate inhibitor refined to 2.2-A resolution. *Biochemistry*, 33(20):6284–6294, May 1994.

[7] G. D. Brayer, Y. Luo, and S. G. Withers. The structure of human pancreatic alpha-amylase at 1.8 a resolution and comparisons with related enzymes. *Protein Sci*, 4(9):1730–1742, Sep 1995.

[8] M. Qian, R. Haser, and F. Payan. Carbohydrate binding sites in a pancreatic alpha-amylase-substrate complex, derived from x-ray structure analysis at 2.1 a resolution. *Protein Sci*, 4(4):747–755, Apr 1995.

[9] N. Ramasubbu, V. Paloth, Y. Luo, G. D. Brayer, and M. J. Levine. Structure of human salivary alpha-amylase at 1.6 a resolution: implications for its role in the oral cavity. *Acta Crystallogr D Biol Crystallogr*, 52(Pt 3):435–446, May 1996.

[10] G. Feller, D. d'Amico, and C. Gerday. Thermodynamic stability of a cold-active alpha-amylase from the Antarctic bacterium Alteromonas haloplanctis. *Biochemistry*, 38(14):4613–4619, Apr 1999.

[11] Salvino D'Amico, Jean-Claude Marx, Charles Gerday, and Georges Feller. Activity-stability relationships in extremophilic enzymes. *J Biol Chem*, 278(10):7891–7896, Mar 2003.

[12] Khawar Sohail Siddiqui, Anne Poljak, Michael Guilhaus, Georges Feller, Salvino D'Amico, Charles Gerday, and Ricardo Cavicchioli. Role of disulfide bridges in the activity and stability of a cold-active alpha-amylase. *J Bacteriol*, 187(17):6206–6212, Sep 2005.

[13] Khawar S Siddiqui, Georges Feller, Salvino D'Amico, Charles Gerday, Laura Giaquinto, and Ricardo Cavicchioli. The active site is the least stable structure in the unfolding pathway of a multidomain cold-adapted alpha-amylase. *J Bacteriol*, 187(17):6197–6205, Sep 2005.

[14] S. D'Amico, J. S. Sohier, and G. Feller. Kinetics and energetics of ligand binding determined by microcalorimetry: insights into active site mobility in a psychrophilic alpha-amylase. *J Mol Biol*, 358(5):1296–1304, May 2006.

[15] Khawar Sohail Siddiqui, Anne Poljak, Michael Guilhaus, Davide De Francisci, Paul M G Curmi, Georges Feller, Salvino D'Amico, Charles Gerday, Vladimir N Uversky, and Ricardo Cavicchioli. Role of lysine versus arginine in enzyme cold-adaptation: Modifying lysine to homo-arginine stabilizes the cold-adapted alpha-amylase from Pseudoalteramonas haloplanktis. *Proteins*, May 2006.

[16] Soundararajan Srimathi, Gurunathan Jayaraman, Georges Feller, Bengt Danielsson, and Paranji R Narayanan. Intrinsic halotolerance of the psychrophilic alpha-amylase from pseudoalteromonas haloplanktis. *Extremophiles*, 11(3):505–515, May 2007.

[17] S. D'Amico, C. Gerday, and G. Feller. Structural determinants of cold adaptation and stability in a large protein. *J Biol Chem*, 276(28):25791–25796, Jul 2001.

[18] Salvino D'Amico, Charles Gerday, and Georges Feller. Dual effects of an extra disulfide bond on the activity and stability of a cold-adapted alpha-amylase. *J Biol Chem*, 277(48):46110–46115, Nov 2002.

[19] G. Gianese, F. Bossa, and S. Pascarella. Comparative structural analysis of psychrophilic and meso- and thermophilic enzymes. *Proteins: Structure Function and Genetics*, 47(2):236–249, 2002.

[20] S Chakravarty and R Varadarajan. Elucidation of factors responsible for enhanced thermal stability of proteins: A structural genomics based study. *BIOCHEMISTRY*, 41(25):8152–8161, JUN 25 2002.

[21] K Suhre and JM Claverie. Genomic correlates of hyperthermostability, an update. *JOURNAL OF BIOLOGICAL CHEMISTRY*, 278(19):17198–17202, MAY 9 2003.

[22] T. Lonhienne, C. Gerday, and G. Feller. Psychrophilic enzymes: revisiting the thermodynamic parameters of activation may explain local flexibility. *Biochim Biophys Acta*, 1543(1):1–10, Nov 2000.

[23] Sandeep Kumar, Chung-Jung Tsai, and Ruth Nussinov. Maximal stabilities of reversible two-state proteins. *Biochemistry*, 41(17):5359–5374, Apr 2002.

[24] Elena Papaleo, Magne Olufsen, Luca De Gioia, and Bjørn O Brandsdal. Optimization of electrostatics as a strategy for cold-adaptation: a case study of cold- and warm-active elastases. *J Mol Graph Model*, 26(1):93–103, Jul 2007.

[25] S. D'Amico, C. Gerday, and G. Feller. Structural similarities and evolutionary relationships in chloride-dependent alpha-amylases. *Gene*, 253(1):95–105, Jul 2000.

[26] Salvino D'Amico, Charles Gerday, and Georges Feller. Temperature adaptation of proteins: engineering mesophilic-like activity and stability in a cold-adapted alpha-amylase. *J Mol Biol*, 332(5):981–988, Oct 2003.

[27] Vojtech Spiwok, Petra Lipovová, Tereza Skálová, Jarmila Dusková, Jan Dohnálek, Jindrich Hasek, Nicholas J Russell, and Blanka Králová. Cold-active enzymes studied by comparative molecular dynamics simulation. *J Mol Model*, 13(4):485–497, Apr 2007.

Chapter 4

Evolutionary Conservation of Protein Dynamics in Trypsin-like Serine Proteases.

4.1 Introduction

Comparative analysis of psychrophilic, mesophilic and thermophilic enzymes suggested that different solutions to cold-adaptation have been "discovered" by different proteins. The differences in unfolding free energies between psychrophilic and mesophilic enzymes are usually small and caused by subtle structural and sequence changes with cooperative and additive effects. In fact, it has been observed that some weak intramolecular interactions are missing in cold-adapted enzymes [1–3]. The investigation of phylogenetically coherent groups of organisms and highly conserved proteins, is crucial to identify amino acid changes which are unambiguously linked to thermal adaptation [1, 4]. Critical changes for thermal adaptation can be hidden amid those produced by genetic drift and other effectors of natural selection. In this context, the observation that the Atlantic salmon expresses two main trypsin isoforms, an anionic form characterised by a typical cold-adapted behaviour (pAST, psychrophilic Atlantic salmon trypsin [5]) and a cationic mesophilic isoform (mAST, mesophilic Atlantic salmon trypsin [6]) is particularly interesting. Therefore, pAST and mAST are an excellent model system to study enzyme cold-adaptation, because of the high sequence similarity and since they belong to the same or-

ganism, as well as since trypsins are biochemically and structurally well characterised [5–8]. pAST was found to catalyse reactions up to 40 times more efficiently at lower temperatures and it presents a lower unfolding free-energy than mesophilic trypsins [7].

Flexibility plays a crucial role in biological systems and it is one of the main traits of cold-adapted enzymes [1–3, 9–18]. It has been suggested that cold-adapted enzymes are characterised by high flexibility of the whole structure or of distinct regions [1–3]. Experimental evaluation of flexibility is generally difficult, making atomistic simulations a suitable alternative. In this context, it is relevant to note that in order to carefully trace local differences in flexibility, a wide sampling of the conformational space is required [9–11, 19]. The analysis of multiple trajectories can help in the identification of recurring features and avoiding artifacts arising from the simulation procedure. However, MD simulations longer than 1-2 ns have been reported only for a few cold-adapted enzymes [11, 12, 17]. In light of the above observations, to shed light on the molecular features responsible for cold-adaptation in serine proteases, we have carried out comparative MD simulations of a mesophilic (mAST) and a psychrophilic (pAST) trypsin isolated from the same organism, in order to filter out as much as possible differences which are unrelated to cold-adaptation. To efficiently sample the potential energy surface, multiple MD simulations were carried out in explicit solvent, collecting 40 ns trajectories for each protein. Results were compared to previous data from MD simulations of psychrophilic and mesophilic elastases [11], which are homologous to trypsins, to evaluate whether there are specific sites within the serine-protease fold which can be considered key determinant of cold-adaptation, and consequently evaluate the possibility of molecular evolutionary convergence in the regions in which adaptive changes occur [20].

4.2 Methods

4.2.1 Molecular dynamics simulations

MD simulations were performed using the 3.3 version of the GROMACS software (www.gromacs.org), implemented on a parallel architecture, using the GROMOS96 force field. Along with mAST and pAST, another psychrophilic *Chum salmon* trypsin) [21], sharing high sequence

identity with pAST, has been considered to validate the collected data.

The X-ray structures of pAST, mAST and pCST (PDB entries 2TBS [5], 1A0J [6] and 1MBQ [21], respectively) were used as starting points for the MD simulations. Protein structures, including the crystallographic water molecules and calcium ions, were soaked in a dodecahedral box of SPC water molecules [22] and simulated using periodic boundary conditions. All the protein atoms are at a distance equal or greater than 0.5 nm from the box edges. The ionisation state of residues was set to be consistent with neutral pH and the tautomeric form of histidine residues was derived using GROMACS tools and confirmed by visual inspection. Initially, the system was relaxed by molecular mechanics (steepest descent, 10000 steps). The optimisation step was followed by 50 ps MD at 300 K (time step 1 fs) while restraining protein atomic positions using a harmonic potential. During equilibration, the coupling constant to the external bath was set to 1 fs [23]. The system was slowly driven to the target temperature (300 K) and pressure (1 bar) through a series of short equilibration simulations.

Productive MD simulations were performed in the NPT ensemble at 300 K, using an external bath with a coupling constant of 0.1 ps. Pressure was kept constant (1 bar) by modifying box dimensions and the time-constant for pressure coupling was set to 1 ps [23]. The LINCS [24] algorithm was used to constrain bond lengths, allowing the use of a 2 fs time step. Electrostatic interactions were calculated using the Particle-mesh Ewald (PME) [25] summation scheme. Van der Waals and Coulomb interactions were truncated at 1.0 nm. The non-bonded pair list was updated every 10 steps and conformations were stored every 2 ps.

To improve conformational sampling, independent 10 ns simulations (*replicas*) were carried out for each protein system, initialising the MD runs with different atomic velocities taken from a Maxwellian distribution.

4.2.2 Analysis of MD trajectories

The mainchain root-mean-square deviation (rmsd) was computed using starting structures as references. For each protein, the stable region of the four *replicas* were joined in a concatenated trajectory, which is representative of different directions of sampling around the starting

structure. The secondary structures were determined using DSSP [26]; solvent accessible surface (SAS) was calculated using NACCESS [27]. The root-mean-square fluctuation (RMSF) per residue was calculated on alpha carbons (C_α), considering different sets of protein conformations: the merged trajectory, from the combination of only three or two of the *replicas* of a same system, from the single trajectories, and from merged trajectories projected on relevant subsets of eigenvectors, and in particular onto the first three, or on the set whose eigenvalues cumulatively sum up to more than 70% of total variance, and the. Trajectories named Δ_i or $\Delta_i\Delta_j$ refer to concatenated trajectories where *replica i*, or *replicas i* and *j* are excluded (with $1 \leq i, j \leq 4$).

Generalised order parameters (S^2), which are a measure of the degree of spatial restriction of motion, have been calculated for main chain N-H bonds from the equilibrium trajectories as an approximation of the asymptotic value of the bond rotational auto-correlation function [28]. S^2 range from 0, corresponding to completely isotropic motions, to 1, if the motions are completely restricted. Details on the computation have been described (2.2.5).

To evaluate the dynamic behaviour of the surroundings of an amino-acid x, residues located within 0.5nm from the centre of mass of x were monitored during MD simulations. In particular, the persistence degree of each residue was defined as the ratio between the number of steps in which a residue appears in the surroundings of x on the total number of steps.

Essential dynamics (ED) is based on diagonalisation of the covariance matrix calculated on C_α-atoms, constructed from atomic displacements; it results in a set of eigenvectors that describe directions of protein motion. Details on essential dynamics analysis have been described in greater detail in 2.2.6.

An unrooted phylogenetic tree was obtained using the PHYLIP package (version 3.67) [29] on the basis of a ClustalW [30] alignment of 54 sequences (31 trypsins and 23 elastases). The Fitch-Margoliash method (FITCH program) was applied to distance data computed with the PROTDIST program, choosing the best tree among 100 obtained randomising the input sequence order (J option); bootstrap analysis with 100 replicates was performed using BOOTSTRAP and CONSENSE; default values were used when not otherwise specified.

4.3 Results

4.3.1 Stability of the simulations and evaluation of conformational sampling

Convergence and stability of the MD trajectories were evaluated monitoring some structural and energetic parameters. The following analyses have been carried out discarding the portions of each replica required to reach stable mainchain rmsd values, to ensure that calculated parameters reflect intrinsic properties of each system. In order to gain insights into the configurations visited by the system, a cluster analysis on the rmsd matrices and an essential dynamics (ED) analysis, were carried out. Conformations from each independent replica clustered separately, with few exceptions, suggesting that different initial velocities caused the trajectories to sample different fer en conformational basins.

Mainchain rmsd values fall in the $[0.21, 0.28]nm$ range, when cluster average structures are compared, indicating that the overall 3D structure is well conserved in the different replicas. The ED analysis gave further information about conformational sampling: 11 (mAST), 13 (pAST) and 12 (pCST) eigenvectors were required to explain 70% of the total motion. Generally, the first three eigenvectors describe ac on sis tent part oft he motion, and they could be used as a reference subspace to analyse protein dynamics. The projection of simulation frames int he 3 D-reference subspace was analysed (Fig.4.1), confirming and clarifying results obtained from cluster analysis.

In particular, a wide sampling of the conformational space with resampling of similar conformations was achieved in our simulations, showing that the essential subspace is well explored when concatenated trajectories are considered. To further evaluate the sampling efficiency, we have also computed the cosine-content of the principal components of protein motion (Fig.4.2). Single simulations are often characterised by large c values in the first eigenvectors, resembling a random diffusion motion, while the corresponding concatenated trajectories have lower or null c and therefore adequately represent essential and significant protein motions.

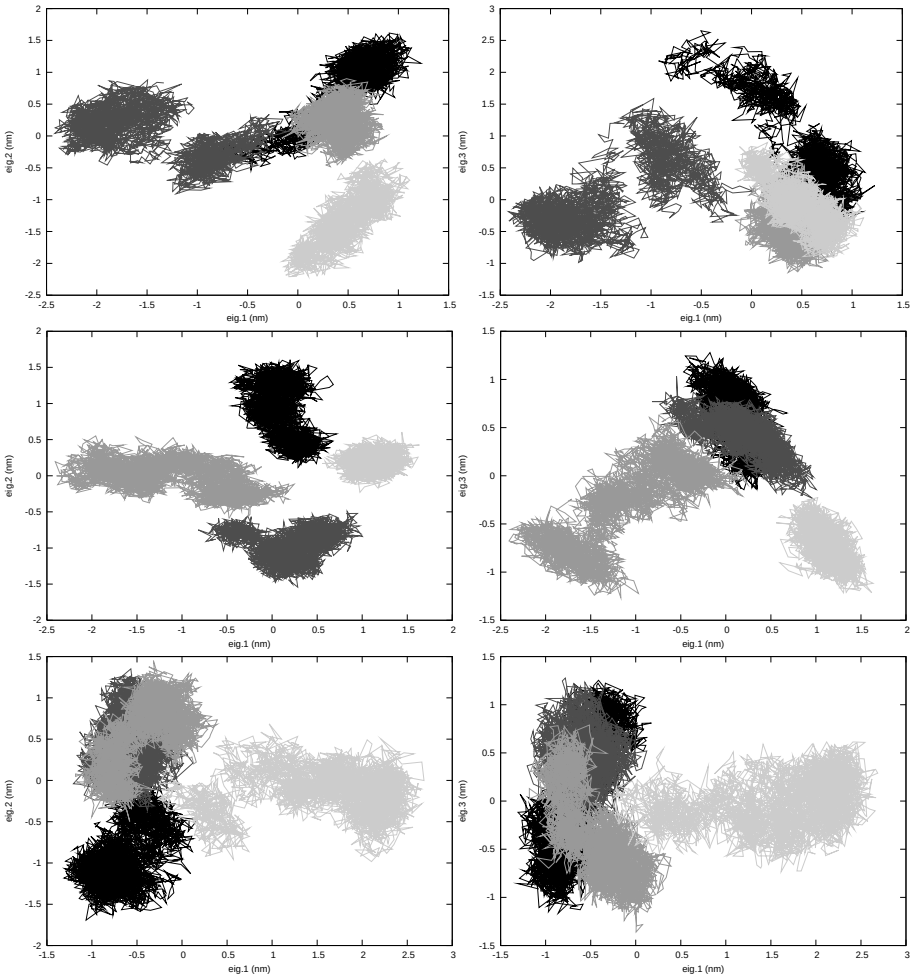


Figure 4.1: Projection of the simulation frames in the 3D-subspace formed by the first three eigenvectors for mAST (top), pAST (middle) and pCST (bottom). The equilibrium conformations relative to each *replica* are shown with different shades of grey, from white (*replica* 1) to black (*replica* 4).

Table 4.1: Average SS content in mAST, pAST and pCST trajectories and in the corresponding X-ray structures

	mAST		pAST		pCST	
	X-ray	MD	X-ray	MD	X-ray	MD
Coil	53 (23.7%)	65.2±4.7 (29.2%)	55 (24.7%)	63.4±4.7 (28.8%)	54 (24.5%)	63.4±4.7 (28.8%)
β-strand	71 (31.8%)	70.5±4.0 (31.6%)	72 (32.4%)	68.6±4.0 (31.2%)	70 (31.8%)	68.6±4.0 (31.2%)
β–bridge	8 (3.6%)	7.3±2.2 (3.3%)	8 (3.6%)	5.6±2.4 (2.5%)	8 (3.6%)	5.6±2.4 (2.5%)
Bend	37 (16.6%)	33.9±5.0 (15.2%)	31 (13.9%)	41.1±4.8 (18.7%)	34 (15.4%)	41.1±4.76 (18.7%)
Turn	30 (13.4%)	24.7±4.4 (11%)	35 (15.7%)	23.6±4.8 (10.7%)	34 (15.4%)	23.6±4.8 (10.7%)
α-helix	17 (7.6%)	16.2±1.5 (7.3%)	15 (6.7%)	13.9±0.78 (6.3%)	14 (6.4%)	13.9±0.8 (6.3%)
3_{10}-helix	7 (3.1%)	5.3±2.3 (2.4%)	6 (2.7%)	3.9±2.3 (1.8%)	6 (2.7%)	3.9±2.3 (1.8%)

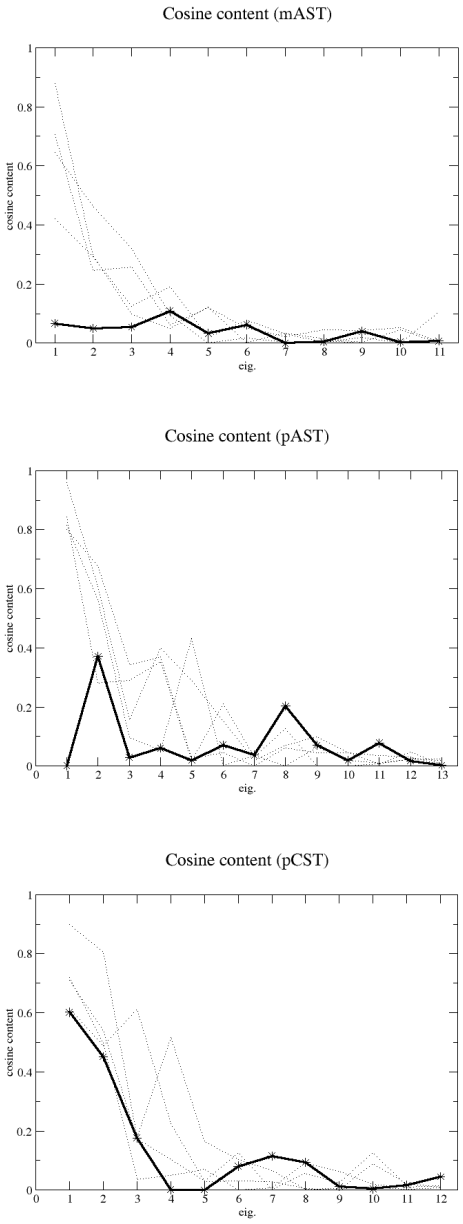


Figure 4.2: Cosine content of the first eigenvectors for mAST (left), pAST (centre) and pCST (right) simulations. The cosine content calculated for the single *replicas* and the concatenated trajectories are shown as grey points and black thick lines, respectively. The cosine content of the first 11, 13 and 12 eigenvectors, i.e., eigenvectors in the essential subspace, are shown for mAST, pAST and pCST, respectively.

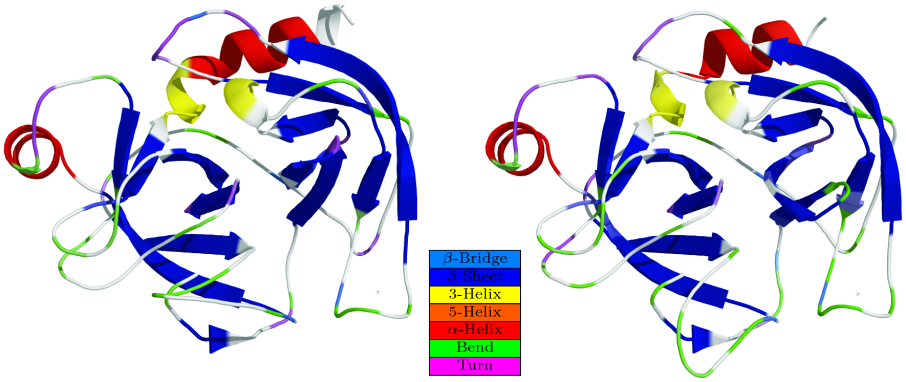


Figure 4.3: 3D representation of the SS persistence in mAST (left) and pAST (right).

4.3.2 Overall structural properties

Secondary structure (SS) elements often play a central role for protein structure and few changes are expected to be found when comparing SS elements across homologous enzymes which are adapted to different temperatures [1]. The psychrophilic trypsins show similar average SS content during the simulations and only subtle differences can be highlighted when compared tot he mesophilic enzyme (Tab.4.1): psychrophilic trypsins exhibit a lower content of β_bridges and 3_{10}-helics and a greater number of residues in bend-like conformations than mAST.

In order to take into account dynamic properties, the time evolution of SS elements was analysed and the most frequently attained SS was evaluated for each residue, obtaining a residue-dependent profile of the persistence of the SS elements. A 3D representation of the SS persistence analysis is shown in Fig.4.3, confirming that α-helix and β-strand are conserved among psychrophilic and mesophilic trypsins, with the exception of the α_2-helix at the C-terminal extremity, which is shorter in psychrophilic enzymes (Fig.4.3). Remarkably, the α_2-helix amino acids are strictly conserved int he psychrophilic trypsins, whereas amino acid substitutions are observed in mAST (R235 $\rightarrow$ S/N, S236 $\rightarrow$ D, I238 $\rightarrow$ L, S243 $\rightarrow$ A). Moreover, the mAST region 231-234, which precedes the α_2-helix, is characterised by a 3_{10}-helix conformation and features N233 $\rightarrow$ I and Y234 $\rightarrow$ F substitutions.

The observation that the α_2-helix is involved in few inter domain

Table 4.2: Molecular environment of the α_2-helix residues in mAST, pAST and pCST. The persistence degrees are indicated in parentheses. The amino acid composition is shown using the single-letter code for mAST, pAST and pCST (first, second and third column, respectively) and amino-acid substitutions between the mesophilic and psychrophilic trypsins are labelled (*).

			mAST		pAST		pCST	
			X-ray	MD	X-ray	MD	X-ray	MD
R235*	S*	N*	L123	L123 (66.3%)	L123	( )	L123	( )
			P124	P124 (69.6%)	P124	P124 (47.9%)	P124	P124 (69.4%)
			S125	S125 (85.2%)	T125	T125 (7.9%)	S125	S125 (26.2%)
			S126	( )		( )		( )
W237	W	W	I89	I89 (97.6%)	I89	I89 (83.9%)	I89	I89 (88.5%)
			M90*	M90* (94.5%)	R90*	R90* (79.3%)	R90*	R90* (85.1%)
			H91	H91 (99.8%)	H91	H91 (98.9%)	H91	H91 (97.3%)
				P92 (32.2%)		P92 (42.1%)		P92 (47.4%)
			I103	I103 (93.8%)	I103	I103 (88.9%)	I103	I103 (87.6%)
			M104	M104 (47.3%)	M104	M104 (35.3%)	M104	M104 (40.1%)
			L105	L105 (99.2%)	L105	L105 (97.4%)	L105	L105 (99.0%)
I238*	L*	L*		( )	V47	V47 (51.2%)	V47	V47 (61.7%)
				( )	I103	I103 (33.7%)	I103	I103 (36.9%)
			L105	L105 (94.2%)	L105	L105 (72.1%)	L105	L105 (84.2%)
			L123	L123 (84.6%)		( )		L123 (33.4%)
				( )		( )		P124 (32.2%)
T241	T	T	W51	W51 (63.0%)	W51	W51 (78.1%)	W51	W51 (75.8%)
			I89	I89 (96.8%)	I89	I89 (94.9%)	I89	I89 (94.8%)
			L105	L105 (65.3%)	L105	L105 (82.0%)	L105	L105 (79.5%)
M242	M	M	I47*	I47* (84.1%)	V47*	V47* (52.3%)	V47*	V47 (71.3%)
			S48*	S48* (76.7%)	N48*	N48* (80.6%)	N48*	N48 (83.9%)
			W51	W51 (99.9%)	W51	W51 (82.5%)	W51	W51 (86.3%)
				( )	L105	L105 (47.8%)		L105 (63.1%)
				( )		L123 (60.9%)		L123 (60.1%)
N245*	Y*	-	W51	W51 (98.1%)	W51	W51 (91.5%)	-	- ( )
			K87*	K87* (82.6%)	R87*	R87* (61.9%)		( )
			I89	I89 (71.7%)	I89	I89 (33.6%)		( )
			K107	K107 (74.0%)	K107	K107 (45.3%)		( )

stabilising interactions in the psychrophilic trypsins might have functional relevance. In fact, the α_2-helix is an important structural element of the serine-protease fold, being an inter-domain connecting region between the C- and N-terminal domains [8], and has also been suggested to influence the relative inter domain movement, acting as a lock for the double-domain structure of serine-pro teases [5, 8]. Residues 236, 239, 240, 243, 244 are solvent exposed ($SAS > 50\%$) and not involved in side-chain-mediated interactions with other protein regions. Notably, the S243 residue of mAST is replaced by alanine in the psychrophilic trypsins, increasing hydrophobicity at the protein surface.

The S236 residue of mAST is located in a molecular environment rich in negatively-charged residues and is replaced by aspartate, increasing the anionic character of psychrophilic trypsins. The side chains of the other α_2-helix and 3_{10}-helix residues are oriented towards the N-terminal domain or the inter domain loop (Tab.4.2) and only small differences were observed. However, the area of the surface at the interface between the C-terminal helix and the N-terminal domain is larger in mesophilic trypsins.

4.3.3 Protein flexibility

In order to evaluate and compare flexibility in mesophilic and psychrophilic trypsins, the RMSF per residue was calculated and adopted as flexibility index. Other descriptors of backbone motion are the generalised order parameters (S^2) and the crystallographic B-factors (Fig.4.4). B-factors are usually considered as a qualitative a ti index of flexibility, but they cannot be used consistently toe valuate differences in flexibility among similar structures, since their values can be affected by additional factors besides internal fluctuations [31].

The RMSF values exhibit a good qualitative correspondence tot he other flexibility indexes (Fig.4.4), indicating a consistent description of protein flexibility. Protein regions characterised by largest RMSF values generally correspond to homologous loops in mesophilic and psychrophilic trypsins, whereas the intensity of the fluctuations can be very different (Fig.4.4).

In order to verify its consistency, the RMSF of each protein has been computed considering different sets of protein conformations. The Pearson correlation coefficient was calculated comparing the different RMSF data sets relative to the same system (Fig.4.5). The correlation coefficients among different RMSF profiles decrease when then umber of sampled conformations in the MD ensemble decreases, whereas when a sufficient number of conformations is considered, the correlation coefficient is close to 1. These results indicate that a MD ensemble generated by a wide sampling of the conformational space is required to obtain a consistent picture of protein flexibility (Fig.4.5). If RMSF profiles of mesophilic and psychrophilic trypsins obtained from short MD ensembles were compared, the results could be poorly significant.

To better highlight regions characterised by different flexibility in

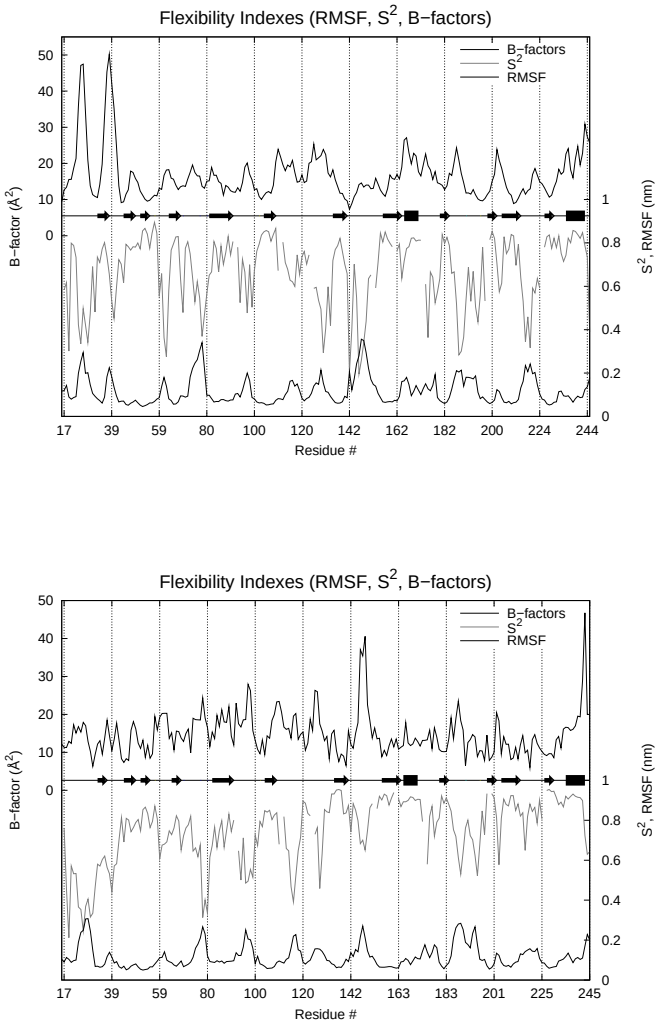


Figure 4.4: Comparison of flexibility indexes in mAST (top) and pAST (bottom). RMSF, mainchain N-H S^2 and crystallographic B-factors are shown in thick black, grey and thin black lines, respectively.

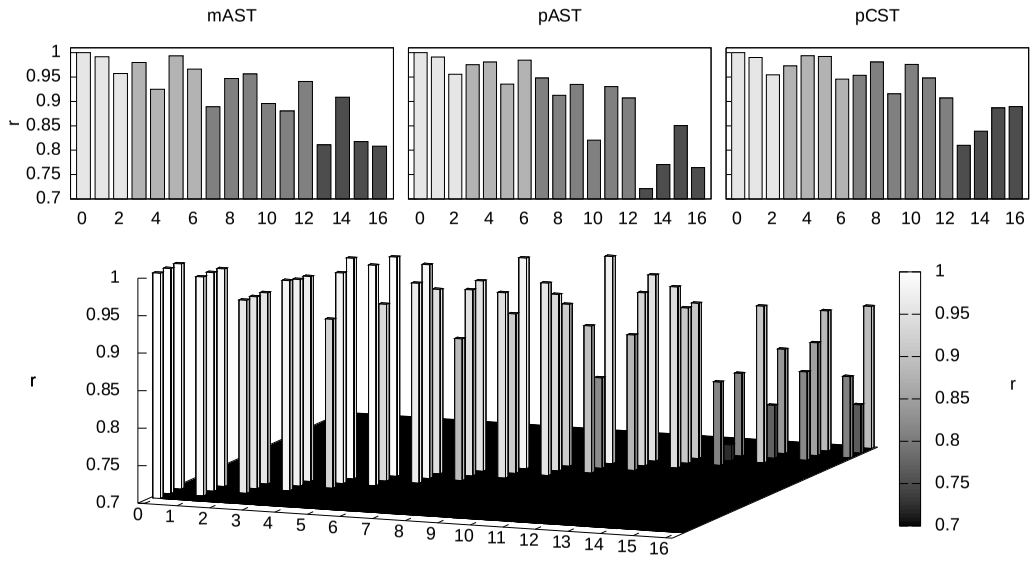


Figure 4.5: Correlation coefficients (r) between RMSF profiles calculated for different subsets of the same concatenated trajectory, plotted as histograms. The histograms show how the RMSF profiles correlate with profile 0 (relative to the total concatenated trajectory). The histograms in the lower panel are shown in the same order (third axis, from near to far) as in the upper panel; their slanted disposition on the third axis is exclusively for clarity. Each number on the x-axis indicates a RMSF profile calculated for a different subset of the same concatenated trajectory. [Concatenated trajectory (0), ED trajectory including the eigenvectors which describe the 70% of the protein motion (1), ED trajectory including the first three eigenvectors (2), Δ_1 trajectory (3), Δ_2 trajectory (4), Δ_3 trajectory (5), Δ_4 trajectory (6), $\Delta_1\Delta_2$ trajectory (7), $\Delta_1\Delta_3$ trajectory (8), $\Delta_1\Delta_4$ trajectory (9), $\Delta_2\Delta_3$ trajectory (10), $\Delta_2\Delta_4$ trajectory (11), $\Delta_3\Delta_4$ trajectory (12), *replica* 1 (13), *replica* 2 (14), *replica* 3 (15), *replica* 4 (16)].

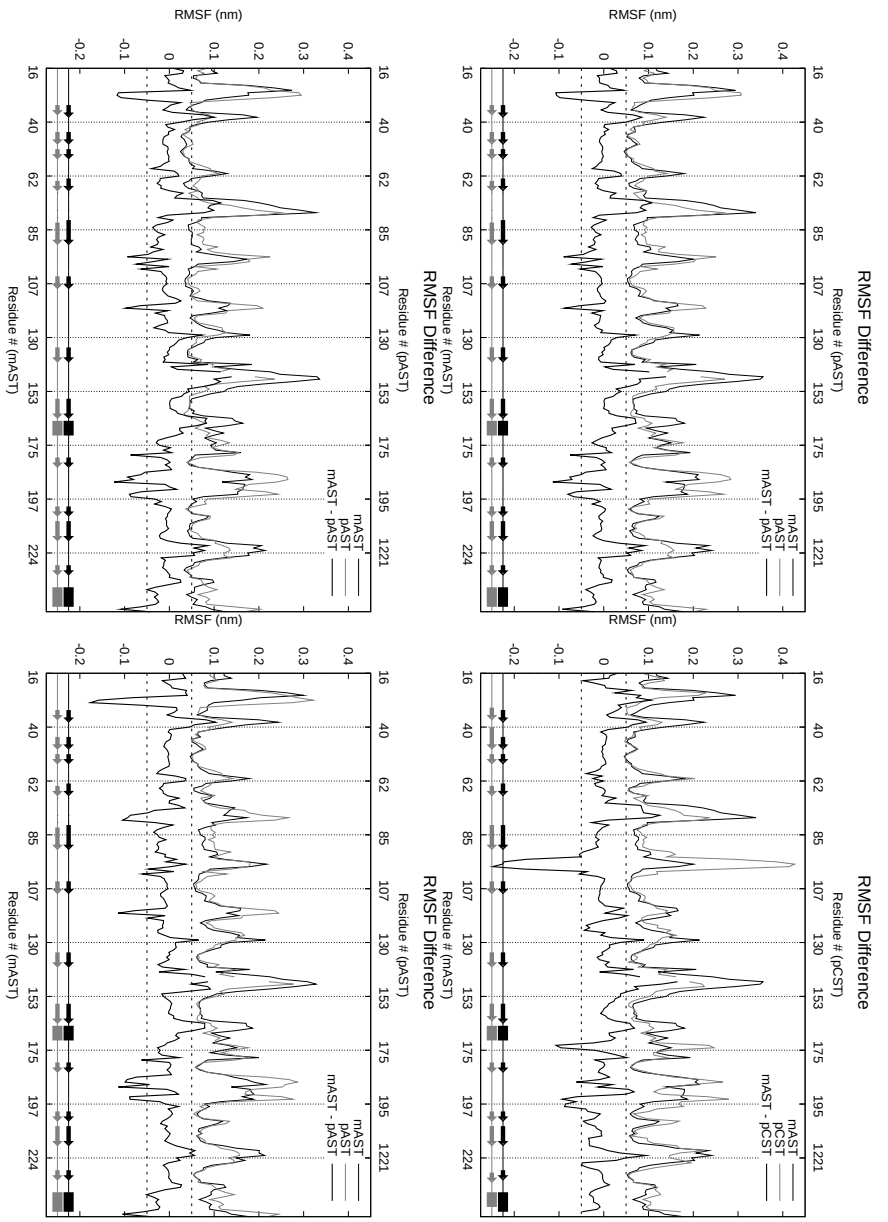


Figure 4.6: RMSF and RMSF-diff as a function of residue number. The RMSF profiles of pAST or pCST (grey) and mAST (black) are shown as thin lines, whereas the difference profiles are shown as thick lines. The profiles calculated on the concatenated trajectories (top), the ED-trajectory (bottom left) and a $\Delta\Delta$–trajectory (bottom right) are shown.

Sequence number in mAST	RMSF-diff range (nm)	Aa composition (psychrophiles)	Aa composition (mesophile)
23-25-26-27-28	(0.10/ 0.05)	**K**-**Y**-S-**Q**-**A**	**R**-**N**-S-**A**-**S**
73-74-75-76-77-78	(0.05/0.10)	I-**K**-V-**T**-E-G	I-**A**-V-**N**-E-G
93-94-95-96-97-98-99-101	(0.15/ 0.05)	**N**-Y-**S**-S-**Y**-N-**I**-N	**S**-Y-**N**-S-**R**-N-**L**-N
116-117-118	(0.10/ 0.05)	**T**-Y-V	**S**-Y-V
129-130	(0.05/0.10)	A-**P**	A-**S**
144-145-147-148-149-150-151	(0.10/0.15)	**T**-**M**-S-S-**T**-**A**-**D**	**L**-**S**-S-S-**S**-**N**-**Y**
164-165-166	(0.05/0.10)	S-**Y**-S	S-**S**-S
173-174-179	(0.10/ 0.05)	P-G-**A**P-G-**N**	
221-222-223-224-225	(0.05/0.10)	G-C-A-**E**-**P**	G-C-A-**Q**-**R**
244-245	(0.10/ 0.05)	S-**Y**	S-**N**

Table 4.3: Protein regions with significantly different RMSF values and amino-acid substitutions between psychrophilic and mesophilic trypsins (RMSF-diff lower or greater than 0 indicates higher flexibility in psychrophilic or mesophilic trypsins, respectively). The amino-acid composition is shown using the single-letter code and amino-acid substitutions between the mesophilic and psychrophilic trypsins are in boldface.

mesophilic and psychrophilic trypsins, the RMSF profiles of the psychrophilic trypsins were subtracted from the mAST corresponding profile (Fig.4.6), determining the correspondence between residues of mesophilic and psychrophilic trypsins from a structural alignment obtained with the Dali algorithm [32].

The regions characterised by RMSF-diff values lower than 0.05 nm (indicating greater flexibility int he psychrophilic trypsins) or higher than 0.05 nm (greater flexibility in mAST) have been further analysed evaluating amino acid composition and localisation in the 3D structure (Tab.4.3 and Fig.4.7a).

Notably, the regions characterised by different flexibility in psychrophilic and mesophilic trypsin sf it well with previously identified regions, which in a psychrophilic elastase isolated from the Atlantic salmon (pSE) and a homologous mesophilic elastase (mPE) are characterised by diverse flexibility [11] (Fig.4.7b). In particular, the mesophilic trypsin mAST and the mesophilic porcine elastase mPE are characterised by enhanced flexibility, when compared to the psychrophilic counterparts, in scattered regions distant from the active site, such as the inter domain loop, where proline residues, conserved in the psychrophilic serine-proteases (P130, P120) are substituted by threonine and serine residues, respectively (Tab.4.3 and Fig.4.7). The mesophilic enzymes are also more flexible int he calcium binding loop (73-78) and in the autolysis loop (144-151), which exhibit different backbone con-

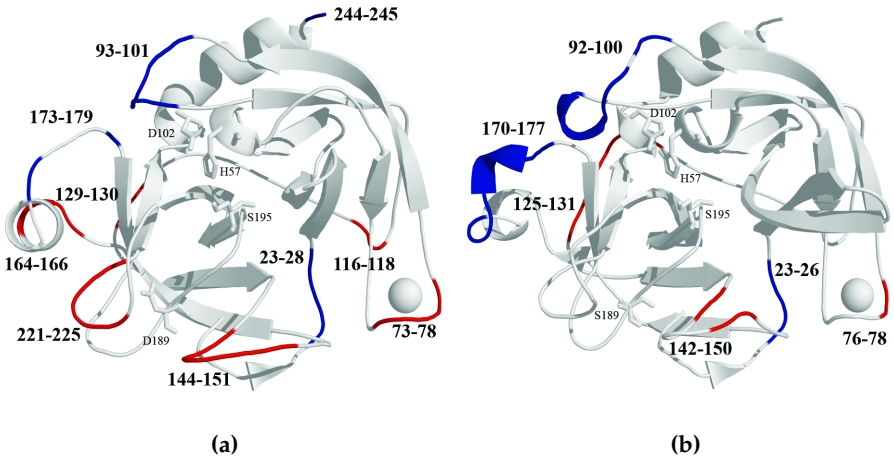


Figure 4.7: Regions characterised by greater flexibility in psychrophilic (blue) or mesophilic (red) trypsins (a),and elastases [11] (b) are highlighted. The SS-elements, calcium ion, and catalytic triad are shown as ribbons, spheres, and sticks, respectively.

formations in the mesophilic and psychrophilic trypsins [6].

The psychrophilic trypsin pAST and the psychrophilic salmon elastase (pSE) present enhanced localised flexibility in the proximity of the catalytic site and the specificity pocket (Tab.4.3 and Fig.4.7). In particular, amino acids 93-101 and 173-179 (trypsin numbering, corresponding to pSE regions 92-100 and 170-177), which belong to loop regions interacting with each other, are characterised by larger flexibility int he psychrophilic enzymes (Fig.4.7). Notably, the 92-100 residues in elastase shave been previously suggested to influence the access of the substrate to the catalytic site [11].

Interestingly, the enhanced flexibility of the 92-100 loop could be related to the lack of hydrophobic or polar interactions, in cold-adapted elastases [13] and trypsins, respectively. In both cases, the lack of interaction is due to the presence of short side-chain residues, as glycines or serines. Moreover, residues of the cold-adapted serine-proteases located in the 92-100 loop present a greater number of interactions with the solvent than the mesophilic counterpart, which could be implicated in the enhanced structural flexibility. The highly flexible region corresponding to amino acids 23-28 (trypsin numbering) is located at the N-terminal extremity.

The observed localisation of highly flexible protein portions in cor-

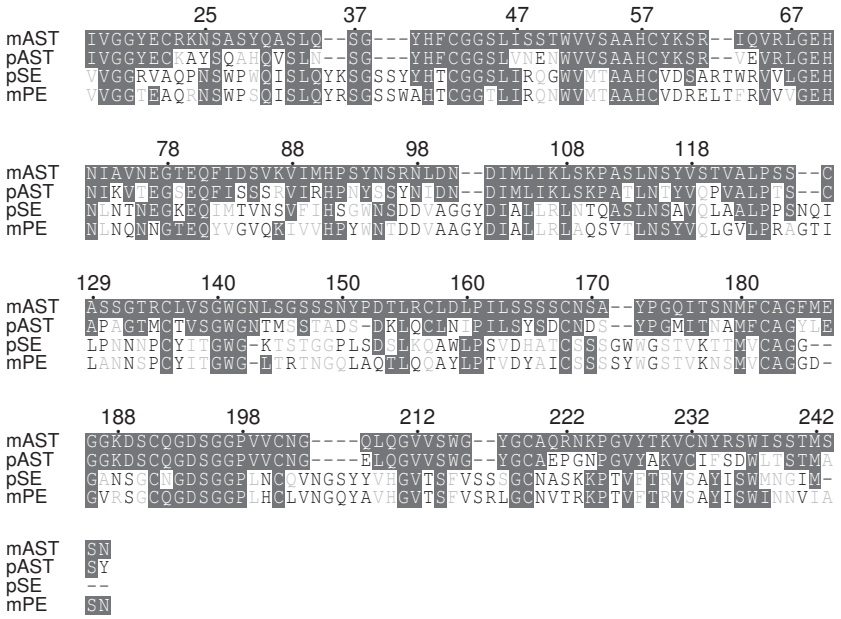


Figure 4.8: The multiple sequence alignment was obtained with ClustalW and processed with TeXShade [http://homepage.uni-tuebingen.de/beitz/tse.html]. Only the sequences of proteins analysed in this work are shown for sake of clarity.

responding regions of the investigated trypsins and elastases prompted us to carryout a phylogenetic analysis from an alignment of trypsin and elastase sequences (Fig.4.9 and Fig.4.8). The sequence identity between elastases and trypsins is about 35%; whereas the identity of mAST vs. pAST, or mPE vs. pSE, is about 70% (Fig.4.8). Results show that the separation between the investigated psychrophilic and mesophilic trypsins, as well as between psychrophilic and mesophilic elastases, is successive to the separation of trypsins from elastases. This separation is very clear from the phylogenetic tree of Fig.4.9a, where the bootstrap values shown denote good statistical support for the topological ordering of some branches between trypsins and elastases. Furthermore, in-depth analysis of the 100 bootstrap replicates (Fig.4.9b) shows that the average distances between two isoforms of the same enzyme ($\langle d(mAST,pAST)\rangle = 0.4282 \pm 0.0539$ and $\langle d(mPE,pSE)\rangle = 0.4362 \pm 0.0529$) are significantly lower than the average distances between enzymes featuring the same temperature adaptation

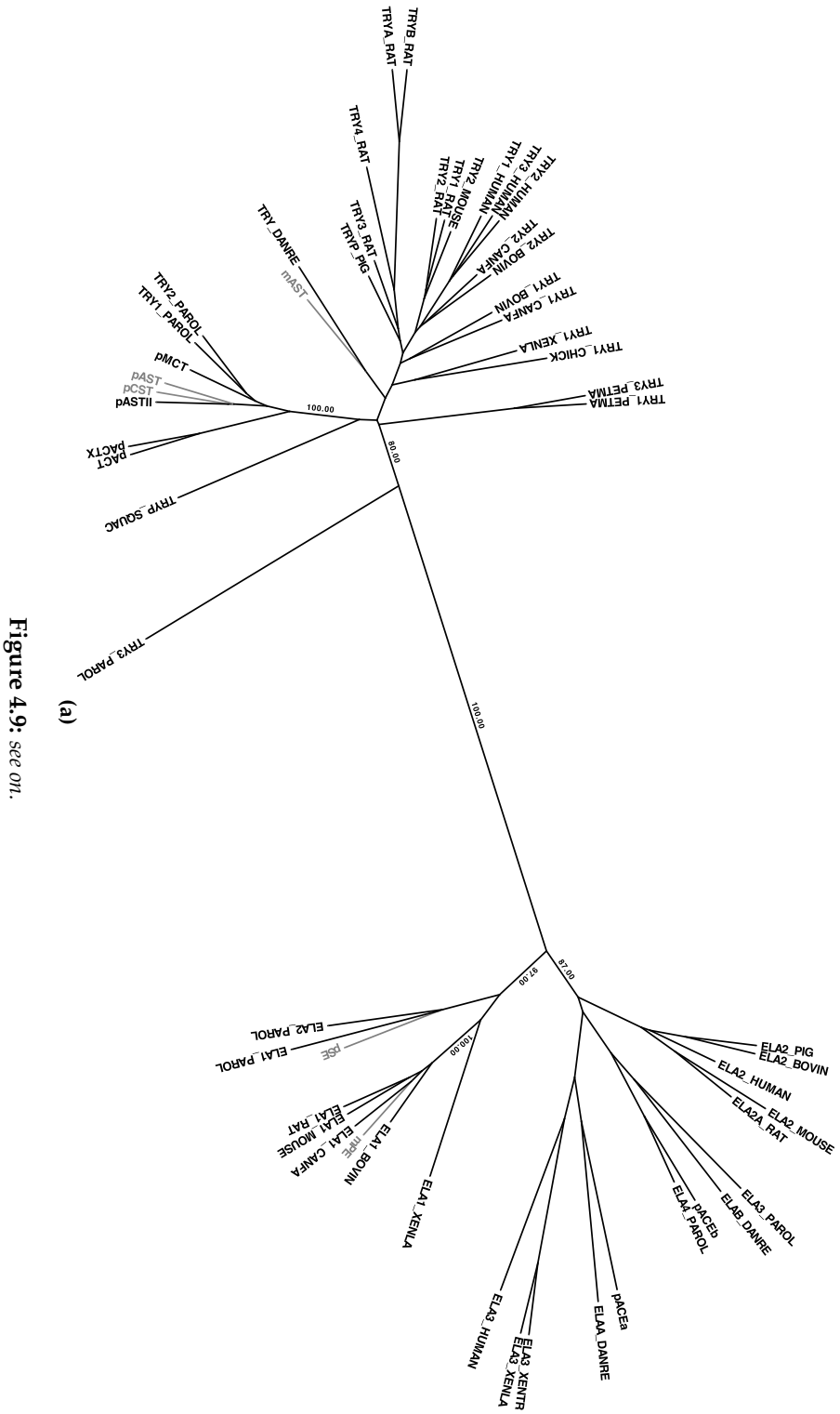
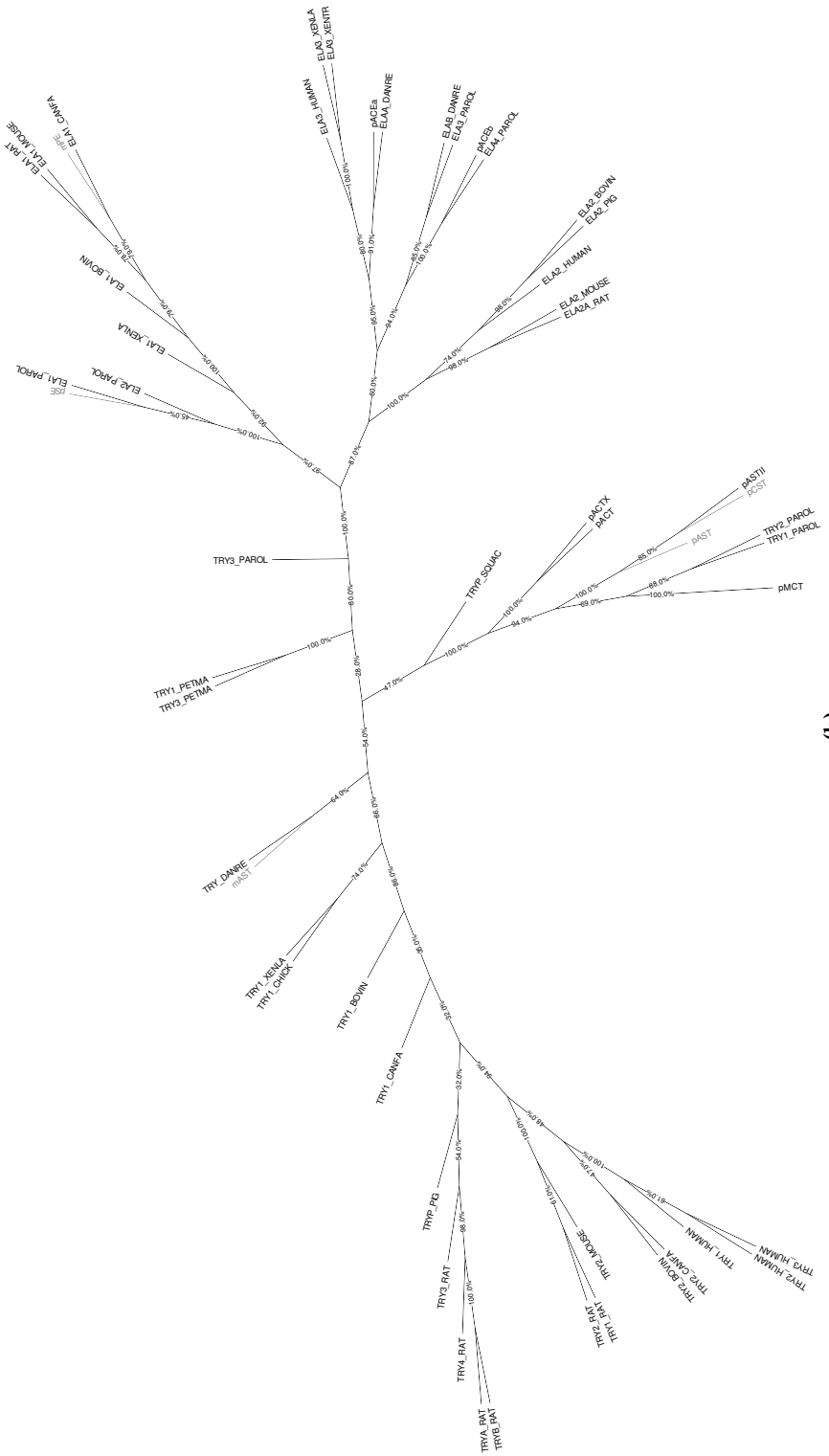


Figure 4.9: *see on.*

(a)



(b)

Figure 4.9: *see on.*

Figure 4.9: The unrooted phylogenetic tree (top), with some relevant bootstrap values specified on the internal branches. The unrooted consensus tree (bottom) resulting from the bootstrap analysis with all branch lengths (bootstrap values) shown. In both trees, terminal branches have been corrected to improve readability. Trees were visualised using Treeillustrator (Genohm BVBA, University of Ghent, http://www.genohm.com). The calculations were performed with the PHYLIP package (see Methods). Enzymes studied in this work are displayed in grey. Sequence names starting with "p" refer to enzymes, which have been demonstrated to be psychrophilic in the literature [33, 34] and the suffix "T" or "E" refers to trypsin and elastases, respectively (e.g. pAST, psychrophilic Atlantic salmon trypsin; pSE, psychrophilic Atlantic salmon elastase). The names of terminal nodes represent their entry names in the SwissProt Database with some exceptions: mAST (P35033), pAST (P35032), pASTII (P35031), pACT (psychrophilic Atlantic cod trypsin, P16049), pACTX (Q91041), pMCT (psychrophilic Maori cod trypsin, Q92099), pCST (Q8AV11), pACEa (psychrophilic Atlantic cod elastase, Q91039), pACEb (Q92077), mPE (P00772), pSE (Q7SIG3), TRY1_PAROL (Q9W7Q7), TRY2_PAROL (A7LD78), TRY3_PAROL (Q9W7Q5), TRY1_PETMA (O42608), TRY3_PETMA(O42159),TRY_DANRE (Q8AV83), ELA1_XENLA (Q8QGF6), ELA3_XENLA (GenBankID:148235116), ELA1_MOUSE (Q91X79), ELA2_MOUSE (P05208), ELA1_PAROL (Q9W7Q2), ELA2_PAROL (Q9W7Q1), ELA3_PAROL (Q9W7Q0), ELA4_PAROL (Q9W7P9), ELA2_HUMAN (P08217),ELA3_HUMAN (P09093), ELA2_PIG (P08419), ELA2_BOVIN (Q29461), ELA3_XENTR (GenBankID:56971746), ELAA_DANRE (Q53CG5), ELAB_DANRE (Q804G1).

($\langle d(mAST, mPE)\rangle = 1.1931 \pm 0.1196$ and $\langle d(pAST, pSE)\rangle = 1.2558 \pm 0.1141$). The non-parametric two sample Kolmogorov-Smirnov test confirms the two former distances are significantly separated from the two latter (with p-values $< 1.6e-45$). Therefore, the analysis of MD trajectories combined with the phylogenetic results indicates that psychrophilic trypsins and elastases have independently discovered the same molecular strategy to adapt to low temperatures.

4.4 Discussion

The picture emerging from the investigation of psychrophilic enzymes suggests that not all evolutionary strategies available to face the negative effects of low temperatures are used simultaneously by cold-adapted enzymes [1, 3, 35]. However, experimental and theoretical data are still needed to support and refine current hypotheses and to increase our knowledge on the molecular basis of cold-adaptation. In this context, the systematic investigation of different enzymatic families becomes crucial to unravel the cold-adaptive strategies discovered

by specific families. The present investigation, together with previous studies on elastases [11, 12], clarifies how distinct members of the serine protease family have faced the negative effect of low temperatures.

Recently, some studies have investigated whether cold adapted enzymes are characterised by increased overall flexibility or whether they are characterised by distinct regions featuring high local flexibility. For example, studies on wild-type and mutant forms of a psychrophilic α-amylase (AHA) showed that this enzyme undergoes a two-state reversible unfolding, supporting the view that flexibility int he structure of AHA is uniform [36]. On the contrary, the enhanced local flexibility of a loop involved in DNA recognition in uracil DNA glycosylases (UDG) correlates well with the $kcal\ mol^{-1}$ values of different mutants of psychrophilic and mesophilic UDGs [14]. Our results are consistent with a scenario in which both local rigidity in regions far from the functional site and improved flexibility of regions not directly involved in catalysis but close to functional sites can be a positive factor int he cold-adaptation of psychrophilic serine-proteases, in agreement with previous suggestions obtained comparing short MD simulations of pAST and a mammalian trypsin [13].

In the previous MD investigation of cold-adapted trypsins [13], the authors did not observed relevant difference in overall flexibility between pAST and a mammalian homologous, hypothesising the relevance of local difference in flexibility. However, they stated that the short-time scale available with MD simulations in the past years, limited the configurational sampling and that only longer and multiple simulations could increase the consistency of local differences in flexibility. Therefore, the wide conformational sampling, the longer and multiple MD simulations, the detailed description of protein flexibility and the comparison between isoforms isolated from the same organisms have allowed us to fully elucidate structural flexibility and to accurately map on the 3D structures the regions characterised by differential flexibility in the cold-adapted trypsins.

Interestingly, most of the amino acid differences between mesophilic and psychrophilic trypsins are localised in the loops close to the catalytic site or the specificity pocket and they could be strongly related to cold-adaptation. Furthermore the α_2-helix, which in the psychrophilic trypsins mediates fewer stabilising inter domain interactions between the C- and the N-terminal domain, is localised in the proximity

of the catalytic site and of the 90-99 loop. Our results are also expected to drive site-direct mutagenesis aimed at highlighting correlation between flexibility, kinetic parameters and thermal stability. In particular, site-direct mutagenesis in the 93-101 and 173-179 loops and in the α_2-helix, aimed at restoring the interactions and characteristics observed in mAST, should allow to evaluate the role of specific amino acids on the k_{cat} and K_m values and on thermal stability.

Generally, evolutionary divergence is functionally constrained so that properties more relevant for function diverge more slowly [15]. Thus, if protein dynamics is relevant for function it should be evolutionary constrained and several evidences of evolutionary selection for specific dynamical characteristics have been recently reported [10, 15, 16]. In particular, backbone protein flexibility profiles diverge slowly and are conserved at the protein family and super family levels, providing indirect evidences of the conservation of protein dynamics [15]. Moreover, the comparison of different ubiquitins revealed a higher degree of conservation of the dynamic properties for proteins which are both structurally and functionally related, suggesting that the optimisation and conservation of protein motions are crucial for specific biological activity [16]. In this context, the observation that the evolutionary separation between psychrophilic and mesophilic trypsins, as well as between psychrophilic and mesophilic elastases, is successive to the separation of trypsins from elastases, indicates that psychrophilic trypsins and elastases independently discovered the same solution to optimise protein flexibility at low temperatures, and is a remarkable example of molecular evolutionary convergence. In a broader context, our results suggest that enzymes sharing the same function and 3D fold may have to adopt similar strategies to optimise structural stability and flexibility in order to elicit their biological function under the challenging conditions of low temperature habitats.

Bibliography

[1] Khawar Sohail Siddiqui and Ricardo Cavicchioli. Cold-adapted enzymes. *Annu Rev Biochem*, 75:403–433, 2006.

[2] P. A. Fields, B. D. Wahlstrand, and G. N. Somero. Intrinsic versus extrinsic stabilization of enzymes: the interaction of solutes

and temperature on A4-lactate dehydrogenase orthologs from warm-adapted and cold-adapted marine fishes. *Eur J Biochem*, 268(16):4497–4505, Aug 2001.

[3] G. Feller. Molecular adaptations to cold in psychrophilic enzymes. *Cell Mol Life Sci*, 60(4):648–662, Apr 2003.

[4] T Thomas and R Cavicchioli. Archaeal cold-adapted proteins: structural and evolutionary analysis of the elongation factor 2 proteins from psychrophilic, mesophilic and thermophilic methanogens. *FEBS LETTERS*, 439(3):281–286, NOV 20 1998.

[5] A. O. Smalås, E. S. Heimstad, A. Hordvik, N. P. Willassen, and R. Male. Cold adaption of enzymes: structural comparison between salmon and bovine trypsins. *Proteins*, 20(2):149–166, Oct 1994.

[6] H. K. Schrøder, N. P. Willassen, and A. O. Smalås. Structure of a non-psychrophilic trypsin from a cold-adapted fish species. *Acta Crystallogr D Biol Crystallogr*, 54(Pt 5):780–798, Sep 1998.

[7] H. Outzen, G. I. Berglund, A. O. Smalas, and N. P. Willassen. Temperature and pH sensitivity of trypsins from Atlantic salmon (Salmo salar) in comparison with bovine and porcine trypsin. *Comp Biochem Physiol B Biochem Mol Biol*, 115(1):33–45, Sep 1996.

[8] H. K. Leiros, N. P. Willassen, and A. O. Smalås. Structural comparison of psychrophilic and mesophilic trypsins. elucidating the molecular basis of cold-adaptation. *Eur J Biochem*, 267(4):1039–1049, Feb 2000.

[9] Manuel Rueda, Carles Ferrer-Costa, Tim Meyer, Alberto Pérez, Jordi Camps, Adam Hospital, Josep Lluis Gelpí, and Modesto Orozco. A consensus view of protein dynamics. *Proc Natl Acad Sci U S A*, 104(3):796–801, Jan 2007.

[10] M. Karplus. Molecular dynamics simulations of biomolecules. *Acc Chem Res*, 35:321–233, 2002.

[11] Elena Papaleo, Laura Riccardi, Chiara Villa, Piercarlo Fantucci, and Luca De Gioia. Flexibility and enzymatic cold-adaptation:

a comparative molecular dynamics investigation of the elastase family. *Biochim Biophys Acta*, 1764(8):1397–1406, Aug 2006.

[12] Elena Papaleo, Magne Olufsen, Luca De Gioia, and Bjørn O Brandsdal. Optimization of electrostatics as a strategy for cold-adaptation: a case study of cold- and warm-active elastases. *J Mol Graph Model*, 26(1):93–103, Jul 2007.

[13] B. O. Brandsdal, E. S. Heimstad, I. Sylte, and A. O. Smalas. Comparative molecular dynamics of mesophilic and psychrophilic protein homologues studied by 1.2 ns simulations. *J Biomol Struct Dyn*, 17(3):493–506, Dec 1999.

[14] Magne Olufsen, Arne O Smalås, Elin Moe, and Bjørn O Brandsdal. Increased flexibility as a strategy for cold adaptation: a comparative molecular dynamics study of cold- and warm-active uracil DNA glycosylase. *J Biol Chem*, 280(18):18042–18048, May 2005.

[15] Sandra Maguid, Sebastián Fernández-Alberti, Gustavo Parisi, and Julián Echave. Evolutionary conservation of protein backbone flexibility. *J Mol Evol*, 63(4):448–457, Oct 2006.

[16] Neelan J Marianayagam and Sophie E Jackson. Native-state dynamics of the ubiquitin family: implications for function and evolution. *J R Soc Interface*, 2(2):47–54, Mar 2005.

[17] Magne Olufsen, Bjørn Olav Brandsdal, and Arne Oskar Smalås. Comparative unfolding studies of psychrophilic and mesophilic uracil dna glycosylase: Md simulations show reduced thermal stability of the cold-adapted enzyme. *J Mol Graph Model*, 26(1):124–134, Jul 2007.

[18] Vojtech Spiwok, Petra Lipovová, Tereza Skálová, Jarmila Dusková, Jan Dohnálek, Jindrich Hasek, Nicholas J Russell, and Blanka Králová. Cold-active enzymes studied by comparative molecular dynamics simulation. *J Mol Model*, 13(4):485–497, Apr 2007.

[19] B. Hess. Convergence of sampling in protein simulations. *Physical Review E*, 65:031910, 2002.

[20] Glenn C Johns and George N Somero. Evolutionary convergence in adaptation of proteins to temperature: A4-lactate dehydro-

genases of Pacific damselfishes (Chromis spp.). *Mol Biol Evol*, 21(2):314–320, Feb 2004.

[21] Eiko Toyota, Kenneth K S Ng, Shiro Kuninaga, Haruo Sekizaki, Kunihiko Itoh, Kazutaka Tanizawa, and Michael N G James. Crystal structure and nucleotide sequence of an anionic trypsin from chum salmon (oncorhynchus keta) in comparison with atlantic salmon (salmo salar) and bovine trypsin. *J Mol Biol*, 324(3):391–397, Nov 2002.

[22] H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, and J. Hermans. Intermolecular forces. In *Adv Appl Lipid Res*, volume 331–332. Reidel, 1981.

[23] H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, A. Di Nola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *J Chem Phys*, 81(8):3684–3690, 1984.

[24] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. fraaije. LINCS: A linear constraint solver for molecular simulations. *J Comput Chem*, 18(12):1463–1472, 1997.

[25] T. Darden, D. York, and L. Pedersen. Particle mesh ewald: An n-log(n) method for ewald sums in large systems. *J Chem Phys*, 98:10089–10092, 1993.

[26] W. Kabsch and C. Sander. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12):2577–2637, Dec 1983.

[27] S.J. Hubbard and J.M. Thornton. 'naccess', computer program, department of biochemistry and molecular biology, university college london. 1993.

[28] Giovanni Lipari and Attila Szabo. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. theory and range of validity. *J. Am. Chem. Soc.*, 104(17):4546 – 4559, 1982.

[29] J. Felsenstein. Phylip-phlygeny inference package (version 3.2). *Cladistics*, 5:164–166, 1989.

[30] J. D. Thompson, D. G. Higgins, and T. J. Gibson. Clustal w: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, 22(22):4673–4680, Nov 1994.

[31] P. H. Hünenberger, A. E. Mark, and W. F. van Gunsteren. Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations. *J Mol Biol*, 252(4):492–503, Sep 1995.

[32] L. Holm and C. Sander. Mapping the protein universe. *Science*, 273(5275):595–603, Aug 1996.

[33] H. K. Leiros, N. P. Willassen, and A. O. Smalås. Residue determinants and sequence analysis of cold-adapted trypsins. *Extremophiles*, 3(3):205–219, Aug 1999.

[34] E. Gudmundsdóttir, R. Spilliaert, Q. Yang, C. S. Craik, J. B. Bjarnason, and A. Gudmundsdóttir. Isolation and characterization of two cDNAs from Atlantic cod encoding two distinct psychrophilic elastases. *Comp Biochem Physiol B Biochem Mol Biol*, 113(4):795–801, Apr 1996.

[35] G. Gianese, F. Bossa, and S. Pascarella. Comparative structural analysis of psychrophilic and meso- and thermophilic enzymes. *Proteins: Structure Function and Genetics*, 47(2):236–249, 2002.

[36] S. D'Amico, C. Gerday, and G. Feller. Structural determinants of cold adaptation and stability in a large protein. *J Biol Chem*, 276(28):25791–25796, Jul 2001.

Chapter 5

Zooming Out on Cold-adaptation of Trypsins

5.1 Introduction

The evolutionary convergence of the cold-adaptive patterns found in the elastases family stresses the importance of extending comparative analyses aimed at unveiling the mechanical bases of cold-adaptation, to large numbers of members of different enzymatic families in order to gain significance and an evolutionary perspective, while maintaining the molecular-level detail. The presented approach, based on comparative all-atomic Molecular Dynamics (ATMD) simulations, has the drawback of being computationally demanding, thus not very efficient in studying the mechanical properties of proteins, and differences thereof among large sets of homologues.

Many approaches have been put forward, especially in recent years, to study protein mechanics and dynamics overcoming the restrictions imposed by the computational cost of ATMD, especially in the study of large-scale motions of proteins and biomolecular assemblies [1, 2]. These approaches are based on the common idea of using simplified descriptions of molecules through the "integration" of a large number of degrees of freedom into a few, a procedure customarily called "coarse graining". The level of this simplification ranges from the "united atom" approach, where only the hydrogen atoms are eliminated, to meso-scale models with interacting sites representing whole proteins. Once the level of simplification is set, the challenge is to es-

tablish how the beads, i.e. the "atoms" in the coarse-grained representation, interact. The set of functions and parameters that describe these interactions is called force-field. In order for the model to be accurate, the force-field must provide a description of the interaction between the beads that effectively reproduces the relevant properties of the simplified system, depending on its application. In particular, a critical issue is the combination of accuracy and predictive power in the force-field, and this is intimate related with its transferability, that is its ability to retain accuracy on systems with different compositions and different configurations. In general, as the number of beads decreases, the simulation is less expensive and the system that can be simulated is larger. On the other hand, the difficulty of creating an accurate and transferable force-field increases in the same direction.

Coarse-grained force-fields are often built based on a single configuration, as in Elastic Network Models (ENM) models: in these models, the system is represented by a network of beads, often one bead per amino acid, connected by elastic springs. The connectivity of the network is based on the starting structure: beads lying below a cutoff distance are linked with springs. In particular, in the homogeneous ENM (hENM) all springs in this representation are identical, irrespective of the physical and chemical details of protein structure [3, 4]. Other ENM approaches assign different strengths to springs based on bead distance [5]. Normal mode analysis (see 5.4) based on ENM models has had remarkable success in describing protein motion [1].

Of particular interest in this context, the method recently developed by Prof. Lavery and coworkers [6–9] is targeted to investigating the mechanical properties of proteins, combining high accuracy and high speed. The method is capable of analysing the mechanical properties of proteins on a residue-by-residue basis by estimating the energy necessary to displace each residue with respect to the rest of the protein, thereby obtaining effective force constant spectra, which are an expression of the rigidity of each residue within the protein edifice.

Speedup with respect to ATMD approaches is obtained primarily by employing the multi-bead coarse-graining scheme proposed by Zacharias, which has already proved effective in protein-protein docking studies [10, 11]. In this model, each amino acid has one pseudoatom at the C_α position; small side chains (excepting glycine) have a second pseudoatom at the geometric centre of the heavy atoms of the side

chain, while larger side chains (Arg, Gln, Glu, His, Lys, Met, Trp, Tyr) have a pseudoatom at the centre of the C_β-C_γ bond and a third pseudoatom at the geometrical cent of the heavy atoms of the side chain atoms beyond C_γ [11]. Pseudoatoms lying below a chosen cutoff distance in the starting structure interact with a set of quadratic springs with uniform force constants, according to the hENM description.

Further speedup comes from substituting Newtonian dynamics with stochastic Brownian dynamics (BD), which ignores inertial effects and treats solvent only through its hydrodynamic drag. The effective force constants K_i are calculated for each residue i by analysing the fluctuations of the mean distance d_i of i from each other residue which occur within a BD simulation, according to the formula:

$$K_i = \frac{3k_BT}{\left\langle (d_i - \langle d_i \rangle)^2 \right\rangle},$$

where $\langle \cdot \rangle$ denotes the average over the simulation, k_B is the Boltzmann constant and T is the temperature of the simulation.

Previous studies employing this method have shown that force constants are highly heterogeneous, even within the same protein, and are sensitive to the global structure, in contrast with B-factors which are strongly dominated by the local structure surrounding each residue [12]. Furthermore, enhanced rigidity as expressed in these force constant spectra has been found to be localised in cofactor binding and active sites [7, 8], and the method has revealed accurate enough to identify long-range effects on flexibility of point mutations, providing a link between dynamics and catalytic activity in a large protein [13]. More recently, the method was applied to study the resistance of GFP to deformation in different directions, finding good agreement with experimental values obtained by single-molecule force spectroscopy pulling experiments [14], and predicting a greater anisotropy of responses than seen experimentally [15].

In this chapter, the possibility to use this approach in the context of the study of the mechanical bases of enzymatic cold-adaptation will be evaluated by comparing in terms of flexibility and rigidity Brownian dynamics and Molecular dynamics simulations using three different resolution levels, from a one-bead coarse-graining scheme to all-atomic representation, and three different force-fields.

5.2 Models and Simulations

Four different simulations are taken into account in the next sections, and they will be referred to as ENMBD, ENMMD, CGMD and ATMD.

ENMBD refers to 500ps-long Brownian Dynamics simulations of a hENM, in the context of the approach described in the introduction. The integrator and analysis tool is a Fortran90 program written by R. Lavery and S. Sacquin-Mora. The simulation was obtained by integrating 50000 steps with time-step of 10 fs at 300 K. The distance cutoff for linking pseudoatoms in the network was set to 9 Å, and 6 kcal mol^{-1} $Å^{-2}$ was used as the force constant for the harmonic interactions. The most computationally expensive operation in such a simulation is that of taking into account the approximate effect of hydrodynamic interactions with the solvent by constructing the configuration-dependent diffusion tensor, which is a third-order tensor that describes hydrodynamic drag for each pair of pseudoatoms, and that can be computed as a function of solvent viscosity, pseudoatomic radii and atomic positions [7, 16]. To speed up simulations, the diffusion tensor is updated every 1000 steps; viscosity of the solvent was set to 1 cP. The simulations were run using as the starting structure the same used for ATMD simulations: for details on the preparation of ATMD simulations see 4.

ENMMD refers to 100ns-long Molecular Dynamics simulations of the same hENM model used for the ENMBD simulations. The simulations were run using the Elastic Network toolkit (tENt), a general purpose extension to the Molecular Modelling Tool Kit (MMTK) library [17], written by M. Pasi. MMTK is a full-featured Python library to perform molecular simulations: the library is extremely flexible owing to its hierarchical and modular structure, so that it allows to be easily extended to perform virtually any modelling-related task. The most important feature of MMTK is that it brings together parallelisation-ready fast C code for time-critical operations with the simplicity of a scripting language (Python) to perform all other tasks. ENMMD simulations were run in vacuum at 300 K, after gradually heating the system to the target temperature, with a time-step of 20 fs. The parameters used to build the hENM are the same described for ENMBD simulations. With this setup, tENt was able to compute almost 2500 integration steps per

second on a single-core desktop workstation.

CGMD refers to 100ns-long Molecular Dynamics simulations of a one-bead-per-aminoacid model with a detailed, force-field-like potential originally employed in the study of the structural transition of HIV1 protease [18, 19]. The method was used as implemented in the RedMD molecular simulation suite [20]. The potential energy function used in this approach includes a harmonic pseudo-bonded term, a bistable pseudo-bond-angle term that accounts for secondary-structure transitions, a pseudo-dihedral term and a non-bonded term which separately describes with a Morse potential local non-bonded interactions, whose optimal distance depends on the starting structure, and non-local non-bonded interactions. Parameterisation of the latter term was carried out by iteratively refined Boltzmann Inversion of the pair radial distribution function computed on a crystal structure [21] to match experimental structural data. Simulations were run in vacuum at 300 K, with a time-step of 20 fs, coupling the system with a constant temperature bath using the Berendsen [22] algorithm with a coupling constant of 20 ps. With this setup, the RedMD suite, run on a parallel architecture, was able to compute 220 integration steps per core per second using 8 cores.

In all simulations, except ATMD, the calcium ions have not been included. Force-constant spectra were computed from the trajectories using the formula provided in the introduction; residue profiles were obtained by averaging the pseudo-atomic values for each residue, except in the case of CGMD-derived spectra which were directly computed on C_α. RMSF profiles were in all cases calculated for C_α, after C_α-based least-square superposition, using routines from GROMACS. The residue profiles shown in the following use residue numbering [1-223], instead of the PDB numbering used in 4, which is based on a common numeration scheme adopted for serine proteases. Difference profiles and spectra were obtained by subtracting corresponding values according to an alignment of the three sequences.

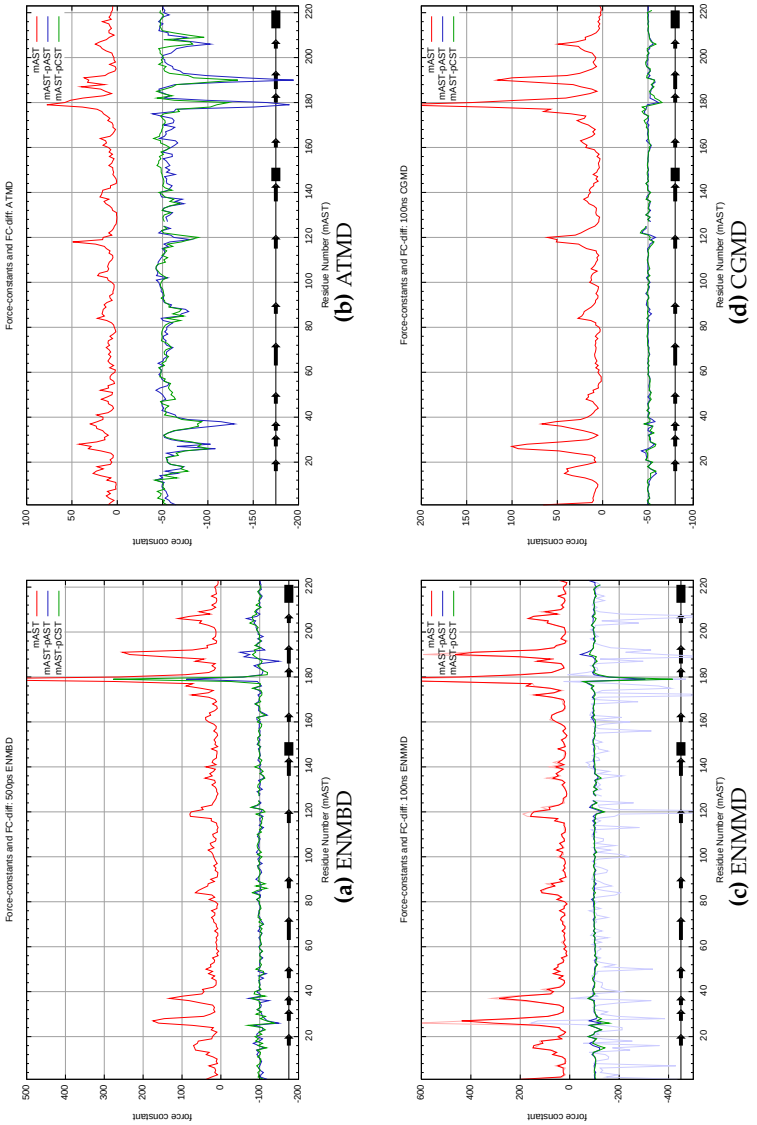
5.3 Fluctuational Analysis

In this section, a detailed comparison will be carried out among four different simulation approaches, in order to investigate the differences in the description they provide of the dynamic properties of

three trypsins, mAST, pAST and pCST, the first mesophilic, the other two psychrophilic (see chap.4, pag.106). This comparison will make use of rigidity (force-constant) and flexibility (RMSF) profiles, relating the differences the hENM approach is able to identify between psychrophilic and mesophilic trypsins (Fig.5.1 and Fig.5.2), to those found using ATMD simulations, and to the results in 4.

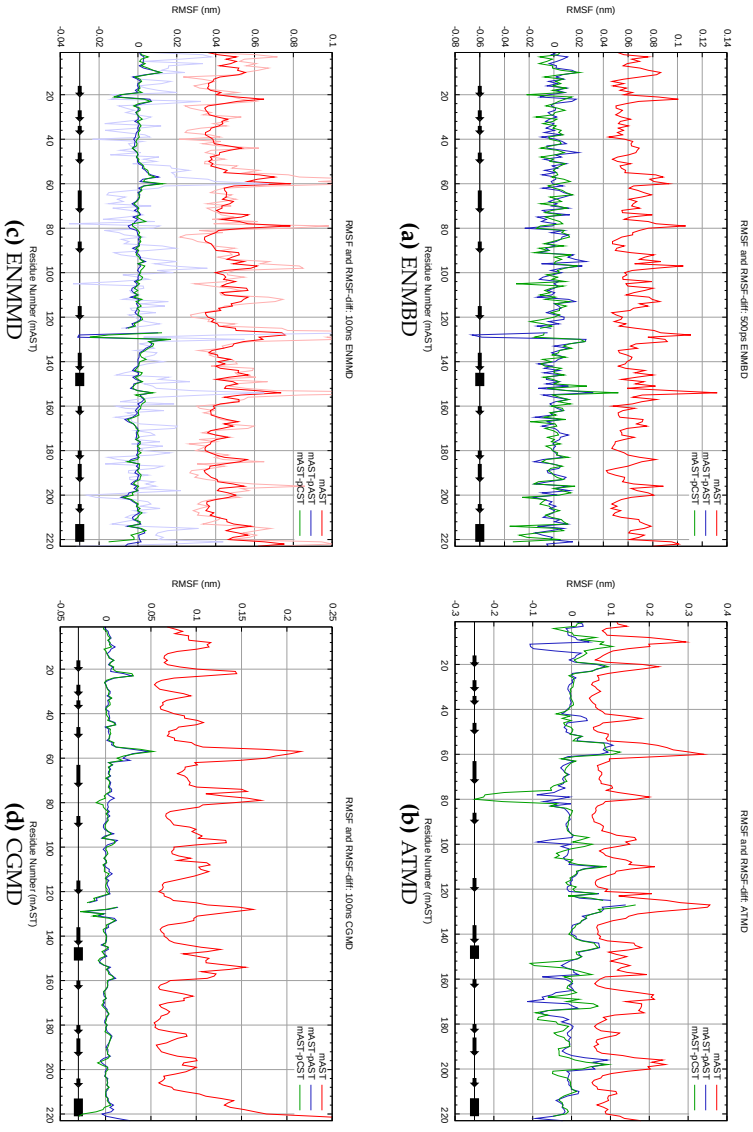
The force-constant spectra derived from ATMD simulations (Fig.5.1b) show global rigidification of the psychrophilic trypsins with respect to the mesophilic, quite consistently throughout the entire structure. The position of the main peaks in the mAST profile is roughly the same in the corresponding spectrum obtained from the ENMBD simulation (Fig.5.1a, pag.137), which is striking considering the differences between the two approaches. The relative heights of the peaks are substantially varied, and the three enzymes show overall greater rigidity in the ENMBD simulation than in the former. Furthermore, the psychrophilic enzymes exhibit very similar rigidity to their mesophilic homologue, and the few differences are found in correspondence with the highest peaks of the mAST profile. The strongest feature of the ENMBD spectrum is a peak of enhanced rigidity in the C-terminal (centred at residue 199), where the ATMD spectrum predicts the opposite trend.

Force-constant spectra aim at capturing protein rigidity; the complementary picture is illustrated in Fig.5.2, where C_α-RMSF profiles are plotted. Interestingly, comparing ATMD- and ENMBD-derived profiles, the latter features on average less than half the flexibility, confirming that in the ENMBD representation, the proteins are more restrained and fluctuate less than in ATMD. Ignoring the difference in scale, there is good agreement between the two mAST profiles in terms of the position and, to a lesser extent, relative height of the main peaks. The clear-cut differences between the meso- and psychrophilic enzymes seen in the ATMD profile are difficult to demarcate in the noisy ENMBD profiles; the difference profiles of pAST and pCST are very different, and in both cases their highest peaks point in opposite directions with respect to the corresponding peaks in the ATMD plot. Collectively, these observations suggest that the ENMBD representation is less sensitive than ATMD to the subtle structural and dynamic differences between mesophilic and psychrophilic trypsins.



(a) ENMBD **(b)** ATMD **(c)** ENMMD **(d)** CGMD

Figure 5.1: The force-constant spectrum of mAST (red) is plotted with the difference spectra of pAST (blue) and pCST (green), obtained by subtracting the psychrophiles' spectra from that of the mesophile. In the ENMMD plot, analogous profiles computed on the first 500ps of the simulation are shown in lighter colours. Force constants are in units of $kcal\ mol^{-1}$ Å^{-2}. To help analysis, difference profiles have been shifted by 100 (plots on the left) or 50 (plots on the right) $kcal\ mol^{-1}$ Å^{-2} towards more negative values. In each plot, a schematic representation of the most frequently attained secondary structure per residue is provided in black.



(a) ENMBD **(b)** ATMD **(c)** ENMMD **(d)** CGMD

Figure 5.2: The RMSF profile of mAST (red) is plotted with the difference profiles of pAST (blue) and pCST (green), obtained as in Fig.5.1. In the ENMMD plot, analogous profiles computed on the first 500 ps of the simulation are shown in lighter colours.

The ruggedness of the ENMBD-derived flexibility profiles possibly indicates that RMSF is a slower-converging property than the rigidity force constants, as suggested by the better convergence observed in the longer ENMMD simulation (Fig.5.2c, pag.138). Taking this third simulation into consideration may also be helpful in clarifying whether the two issues risen in this comparison, namely, restrained fluctuations and reduced sensitivity of ENMBD are due to the use of stochastic dynamics or to the coarse-grained representation of protein structure and the use of the hENM potential.

The profiles obtained from ENMMD simulations (Fig.5.2c and Fig.5.1c) exhibit reduced flexibility and enhanced rigidity with respect to both ATMD and ENMBD, showing that the extended sampling using Newtonian dynamics results in smaller and more consistent fluctuations than those seen in ENMBD simulations. This effect can more clearly be ascribed to extended sampling for the flexibility profile than for the force-constant spectrum, as pointed out by the ruggedness of the superposed short-time profiles. Interestingly though, in both plots, difference profiles suggest the three proteins converge to more similar dynamic properties on longer time-scales, and the fewer differences that can be identified between mesophilic and psychrophilic trypsins in ENMMD plots are only partially in greater agreement with ATMD.

Despite the differences found in this detailed comparison of fluctuations in these three trypsins, the most striking feature of figures Fig.5.2 and Fig.5.1 is that the description of their dynamics obtained from these four simulations, which result from approaches that differ dramatically in many fundamental aspects, and encompass almost two orders of magnitude of computational cost, ultimately converge to a similar picture of the flexibility and rigidity of the three enzymes, confirming the general validity of coarse-grained approaches for the study of the dynamic properties of proteins, and underlining the robustness of the relation between protein structure and dynamics [23–26]. This study pushes these assumptions further, raising a fundamental question on the sensitivity of coarse-grained methods to the subtle mechanical differences between enzymes with different thermal adaptation. The results shown so far suggest that hENM approaches may not be the best suited for this task, and that a description which goes beyond the structure may be required to gain effective insight into cold-adaptation.

The fourth and last approach included in this comparison provides

a more detailed description of the inter-pseudoatomic potential, distinguishing between bonded and non-bonded interactions, but adopts a coarser, one-bead-per-amino-acid representation of the structure. Figures Fig.5.2d and Fig.5.1d display the results of the fluctuational analysis performed on CGMD simulations. Overall, the flexibility and rigidity profiles show that the CGMD approach yields fluctuations which lie between those seen using hENM and ATMD. This observation confirms an inherent limitation in the sampling capabilities of ENM-based approaches: they provide efficient exploration of the system's phase space only in the immediate vicinity of the energy minimum, which is by definition the starting structure on the basis of which the connectivity of the elastic network is defined, so that large-scale structural transitions are disallowed until such connectivity is redefined. The inability to "jump" between the multiple substates (local minima) that are a natural consequence of the ruggedness of the potential energy surface of proteins, also makes hENM approaches strongly biased towards the starting structure. This limitation is overcome by the CGMD approach, but apparently its increased sampling efficiency does not correspond, in this case, with enhanced sensitivity: difference profiles obtained from CGMD simulations recognise few minor differences between mesophilic and psychrophilic enzymes both in terms of flexibility and, especially, rigidity.

In order to understand at a residue-by-residue level how much the harmonic approximation is responsible for the differences seen between ATMD and hENM-based simulations, the ATMD trajectories have been analysed in terms of the harmonic behaviour of selected inter-C_α distances.

5.4 Harmonicity in the molecular dynamics of trypsins

The theoretical study of protein dynamics has often strongly benefited from the use of simplifications and approximations that make it feasible to tackle such a complex phenomenon with available computational resources. One of the most widely used simplifications is the harmonic approximation, which finds its most important application in normal mode analysis (NMA), which allows the study of macromolec-

ular dynamics by analytic means, within the harmonic approximation. The fundamental restriction of normal mode analysis is its limitation to study the dynamics in a single harmonic potential well. The construction of the harmonic potential well is based on the calculation of the second derivative of the potential energy in correspondence of a minimum, so that the resulting harmonic model is an approximation of a single conformational substate, and is accurate only for small fluctuations around the energy minimum.

Nevertheless, such models have been routinely used in the study of much larger amplitude motions, and often the 10-20 slowest normal modes of any of the system's minima are sufficient to describe the large-scale conformational changes in remarkable agreement with experimental data [4, 27]. Many studies have been carried out with the aim of untangling this apparent contradiction, often by comparing NMA with quasi-harmonic or essential-dynamics analysis (based on principal component analysis, PCA) performed on ATMD simulations [28–31]. These studies have been successful in characterising the self-similar multi-scale organisation of protein free-energy landscape, whereby motions in the local minima are very similar in shape among each other and to the motions involved in large-scale conformational rearrangements. These studies have also shown that the greater part of protein fluctuations is contributed by a few anharmonic, large-scale collective motions which represent the transitions between the substates (local minima on the free-energy surface), often involving substantial rearrangements of the protein backbone. On the other hand, a great number of local motions featuring harmonic character, and which make each a small contribution to total fluctuation, correspond to oscillations within each substate. Models and approaches have been proposed to explicitly take into account these features in studying protein dynamics [5, 31–33].

In this study, the harmonicity of protein backbone motions as described by long, multi-*replica* ATMD simulations will be evaluated by characterising the fluctuations of those inter-C_α distances which would correspond to quadratic bonds in the hENM description, where their behaviour is harmonic. This allows to make a direct comparison between the two approaches, obtaining residue-by-residue information on the adequacy of the harmonic potential. This study could also be the starting point for further analyses aimed at the parametrisa-

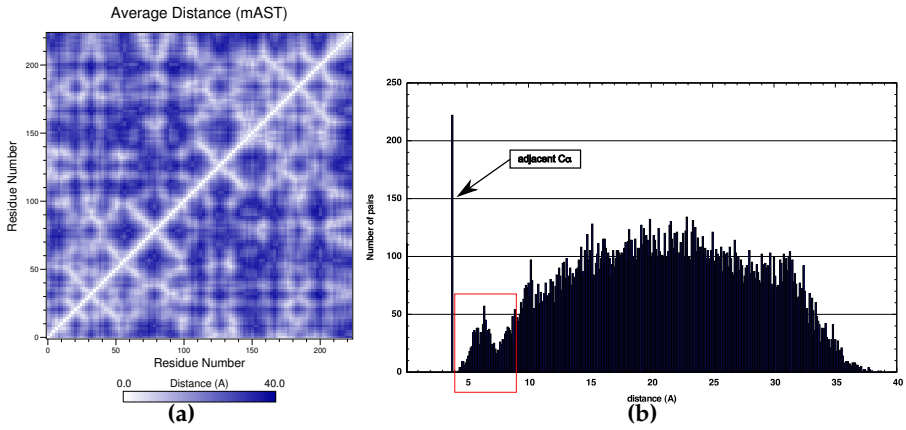


Figure 5.3: The average distance matrix (left) computed from *replica1*, from which the distance distribution is obtained (right), showing the adjacent C_α peak; in red, the range of pairs studied.

tion of an elastic network model that introduces heterogeneity both in the strength and the functional form of inter-pseudoatom interactions, based not only on geometrical features of the structure, but also on other properties, e.g., secondary structure and solvent accessibility.

5.4.1 Methods

The analyses have been performed on the equilibrated portions of 4 ATMD *replicas* of mAST; details on the simulations have been previously described (see 4.2.1). The set of inter-C_α distances to monitor was determined based on the way hENM are built: in the hENM approaches studied in the previous section, pseudoatom pairs lying nearer than 9 Å are linked by quadratic springs to build the elastic network; since adjacent C_α, i.e., C_α belonging to sequentially contiguous residues, lie at a distance of ~ 3.8 Å (see Fig.5.3), the range of interest was set to [4,9]Å.

In the first stage of the study, 10 pairs of C_α were selected to perform the analysis (Tab.5.1, pag.145). These pairs were chosen on the basis of their average distance in *replica* 1: 3 pairs were chosen that had average distance around 5 Å, 3 around 6.5 Å and 4 around 8 Å. Since one of the goals of this study is to correlate harmonicity with other structural parameters, the choice was made in order for the pairs to be well spread in terms of distance in sequence, average secondary structure

and average solvent-accessible surface (SAS) of the residues bearing the C_α. The reported distance and SAS values have been measured as the per-*replica* ensemble averages, while the adopted per-residue secondary structure is that most frequently attained during each *replica*. Secondary structural elements have been classified as customary: α (α-helices), β (β-strands) and l (*loops*). SAS values have been divided into 3 categories: *LA* ($SAS \leq 10\%$), *AA* ($10 < SAS \leq 40\%$) and *HA* ($SAS > 40\%$). Combining these residue-based categorisations in pairs, 6 categories are possible for both secondary structure (*SS type*) and SAS (*SAS type*). Further details on the calculation of these properties have been provided previously (see 2.2).

The time-course of the inter-C_α distances were measured on the single *replicas* and on the merged trajectory obtained concatenating the 4 *replicas*. A histogram of these time-courses was then computed to obtain the distance distributions. All histograms were built in the range $[3,25]$Å, to take into account large fluctuations of distance values; no values exceeding these limits have been recorded in the simulations. The bin size for histograms was set to 0.1 Å. Distributions were then characterised using standard statistical descriptors; in particular, the third and fourth central moments of the distributions were calculated (skewness, γ_1 and excess kurtosis, γ_2) to obtain a quantitative description of the shape of the distributions, and of their similarity to a gaussian curve. For simplicity, excess kurtosis will be referred to as kurtosis in the text. In order to directly compare with hENM-based approaches, the same procedure was applied on a 500 ps ENMBD simulation of mAST (see 5.2).

In the second stage of this study, the described analysis procedure was extended to all C_α pairs featuring average distance in the $[4,9]$Å range. The resulting large number of distributions of inter-C_α distance fluctuations were analysed using a straightforward procedure to estimate modality, i.e., the presence of more than one peak. To this aim, the shape of the distribution is studied by analysing its first derivative. Each peak of the distribution is a local maximum, a stationary point in correspondence of which the first derivative is zero.

The approximate first derivative was calculated using the Savitzky-Golay filter as implemented in the Matlab Signal Processing Toolbox, using an order 2 polynomial for least-squares fitting. The raw data is usually very noisy, and the local fluctuations in the distributions, due

to insufficient sampling, could be recognised as local maxima in this procedure. For this reason the choice of the filter's smoothing frame size is important to obtain both reliability and accuracy. Based on visual inspection of numerous distance distributions, the parameter has been set to 11. This choice allows an approximate ($\sqrt{11} \simeq$) 3-fold increase in the signal to noise ratio of the distributions. Furthermore, given that distribution points are 0.1 Å apart, the method gives best results on data featuring a single inflection in each $11 * 0.1 = 1.1$ Å interval [34].

Stationary points in the distributions are identified as those points in correspondence of which the derivative changes sign. Since the interpolation range is chosen greater than the actual fluctuation range, the derivative function is expected to start positive (increasing probability towards mean/first mode), and end negative. Thus an odd number of stationary points is expected, possible values being 1 (unimodal), 3 (bimodal) or more: all distributions with more than 1 stationary point have been selected as multimodal. To avoid interpreting often occurring small peaks at the extremes of the distributions (outliers) as local maxima, the derivative calculation was performed on the 0.5-99.5% interpercentile range of distance distributions.

5.4.2 Results and Discussion

If two isolated particles linked by a spring that obeys Hooke's law were to be considered, then the fluctuations of one particle with respect to the other would occur on a parabolic (quadratic) potential, and the distribution of inter-particle distances would be a perfect gaussian curve. The comparison of inter-C_α distance fluctuations from ATMD with those from a network composed of a great number of these harmonic oscillators, coupled together in a complex way (hENM), will be made chiefly by estimating "how gaussian" the distributions of inter-C_α distances are. To this end, the third and fourth central moments of the distributions (skewness and kurtosis) are especially useful, given that values of these two metrics are known for perfect gaussians. Skewness (γ_1) is a measure of the degree of asymmetry of a distribution; gaussians being perfectly symmetric have $\gamma_1 = 0$, while positive γ_1 values correspond to a bell-shaped curve with a longer tail on the right, and viceversa. Kurtosis (γ_2) measures the degree of "peakedness" of a distribution; gaussians have $\gamma_2 = 0$, while positive kurtosis indicates a

Table 5.1: Summary of the properties of the 10 pairs in the first test-set. SAS values are averages over the 4 *replicas*, and secondary structure percentages are average percentages over the 4 *replicas*. * Residues 146 and 150 belong to the same α-helix.

range (Å)	aa_1 $\#_1$-$\#_2$ aa_2 -	*SS type*	SS_1 $\langle\%_1\rangle$	SS_2 $\langle\%_2\rangle$	*SAS type*	SAS_1 (%)	SAS_2 (%)	ΔSAS	$\Delta\#$ (Å)	$d_{1,2}$
5	D 167-200 R	*l*/*l*	*l*	*l*	*HA*/*HA*	54.9	65.2	10.3	33	4.71
5	V 36-87 L	β/β	β (100)	β (100)	*LA*/*LA*	2.4	0.0	2.4	51	4.97
5	S 144-146 S	α/*l*	*l*	α (99)	*AA*/*HA*	32.2	68.7	36.5	2	5.43
6.5	S 146-150 S	α/α*	α (99)	α (96)	*HA*/*HA*	68.7	72.1	3.4	4	6.19
6.5	Q 15-100 V	*l*/*l*	*l*	*l*	*LA*/*LA*	3.9	2.4	1.5	85	6.28
6.5	S 17-24 F	β/β	β (97)	β (93)	*LA*/*HA*	10.0	57.0	47.	7	6.9
8	F 24-175 G	β/*l*	β (93)	*l*	*HA*/*HA*	57.0	60.9	3.9	151	8.29
8	L 87-216 I	α/β	β (100)	α (100)	*LA*/*LA*	0.0	1.9	1.9	129	7.65
8	S 107-213 R	α/*l*	*l*	α (99)	*AA*/*HA*	77.7	34.9	42.8	106	8.52
8	N 184-187 L	β/*l*	*l*	β (100)	*LA*/*HA*	81.1	0.9	80.2	3	8.07

taller and thinner peak and negative values correspond to flatter distributions.

Fig.5.4 shows the distributions of distances for the 10 selected C_α pairs (see also Tab.5.1). The plots show that inside our limited selection, considering the merged ATMD trajectory, not all pairs confirm the clear trend observed for atomic positional fluctuations (see chap.5.3, pag.135, in particular Fig.5.2), i.e., much greater fluctuations in ATMD with respect to ENMBD simulations.

In fact, in 4 cases, the width of the peaks is greater in the latter (pairs 36-87, 144-146, 146-150 and 184-187), and almost identical in a fifth case (pair 87-216). In all of these cases, ATMD distance distributions feature a consistent, well-definite unimodal peak both for single *replicas* and the merged trajectory. It is quite striking to observe that all these pairs connect residues that participate in or immediately precede (residues 144 and 184) stable secondary structures (Tab.5.1). Pair 17-24 also belongs to stable secondary structures, and shows a quite consistent behaviour among *replicas*, especially 1, 2 and 4, and very similar fluctuations in the ENMBD and merged ATMD simulations. These observations give a first important hint on how secondary structuring can modulate backbone dynamics.

In the other four cases, the distributions are very different both in shape and position among *replicas*, suggesting that global structural features have a strong influence on inter-C_α distance fluctuations. This influence may be the consequence of differences, among various substates, in the forces that restrain a pair to a particular range of distances, modifying the "peakedness" of the distributions (pairs 24-175 and 107-

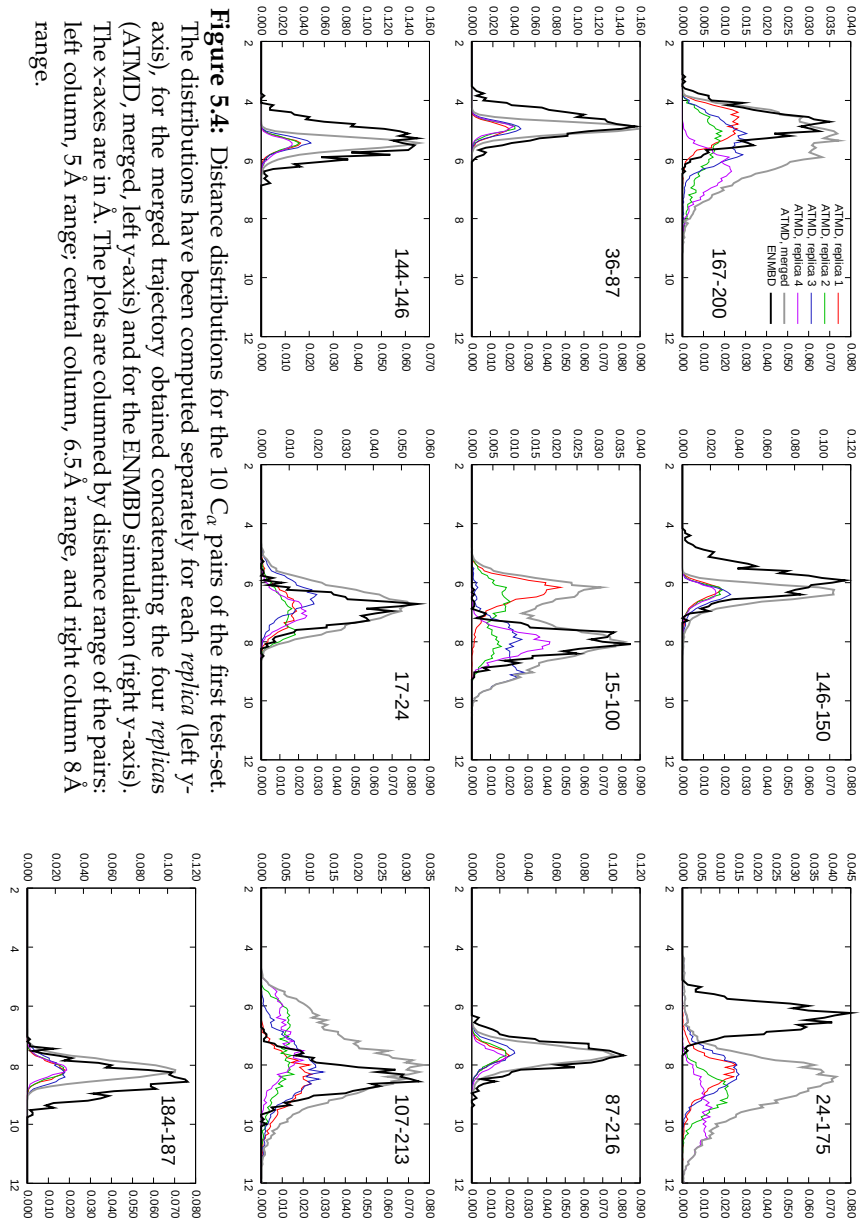


Figure 5.4: Distance distributions for the 10 C_α pairs of the first test-set. The distributions have been computed separately for each *replica* (left y-axis), for the merged trajectory obtained concatenating the four *replicas* (ATMD, merged, left y-axis) and for the ENMBD simulation (right y-axis). The x-axes are in Å. The plots are columned by distance range of the pairs: left column, 5 Å range; central column, 6.5 Å range, and right column 8 Å range.

213) or their mean values (pair 167-200), and even switching the selectivity of a pair between distinct substates (pair 15-100), resulting in a multimodal distribution of distances in the merged trajectory.

Generally, for these last four cases, one pair member lacking secondary structure results in a wide range of different fluctuational behaviours, of which the lack of resemblance to a gaussian distribution can often be inferred by eye. A quantitative description of this resemblance is shown in Fig.5.5, where for each of the distributions of Fig.5.4, statistical descriptors have been computed.

As expected, distributions computed on the ENMBD simulation very closely resemble gaussians, as shown by their low γ_1 and γ_2 values, all below 0.5, while ATMD-derived distributions generally deviate more from normality. The previously observed inter-*replica* heterogeneity is even more apparent in the quantitative analysis, and the comparison of the histograms with the corresponding plots in Fig.5.4 strongly advocates the use of a quantitative strategy for the evaluation of harmonicity. The statistical descriptors are sensitive enough to identify differences among *replicas* which were hardly recognisable from the distributions (e.g., pair 144-146), and to reveal that a harmonic potential is unable to provide an accurate description even of some of the fluctuations that result in well-defined unimodal distributions (e.g., pair 146-150).

The analysis of this small but highly heterogeneous test set suggests that consistent behaviour among *replicas*, and to a lesser extent, the harmonic character of inter-C_α distances feature good correlation with secondary structure, but not with solvent accessibility nor average distance. Although data are largely insufficient to draw conclusions of general validity on these relationships, these results exemplify the sensitivity of the proposed statistical analysis of distributions. To obtain this quantitative information with sufficient statistical significance, this analysis has been extended to a much larger test-set.

A total of 1520 C_α pairs were found to have average distance in the [4,9]Å range in at least one of the 4 *replicas*. The corresponding set for the ENMBD simulation counts 1374 pairs; 91.5% of these pairs coincide in the two sets. The same analyses described earlier were carried out on the distributions, with the addition of a procedure to evaluate their modality (see the Methods section, 5.4.1). To assess reliability of this procedure, the results for 45 pairs chosen at even intervals along the

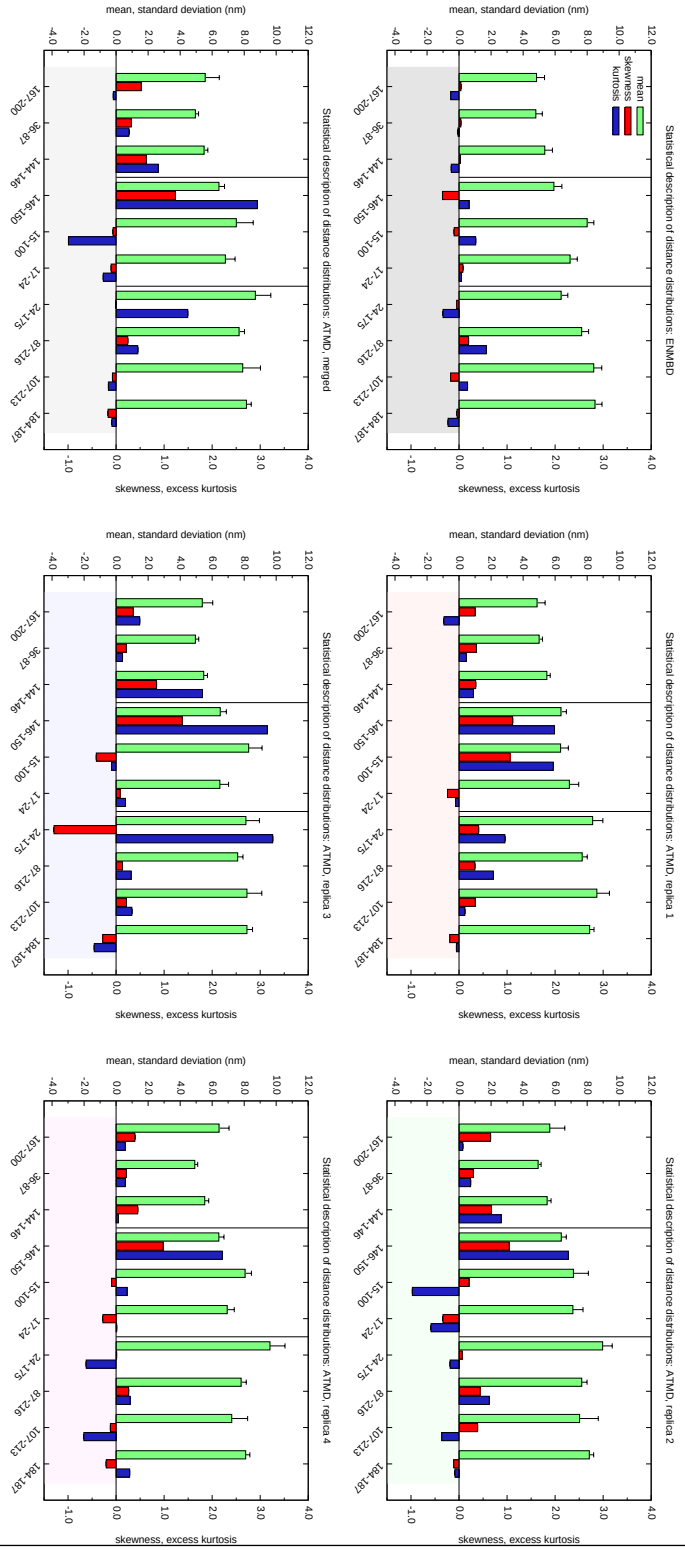


Figure 5.5: Statistical description of distance distributions for the 10 pairs in the first test-set, divided by simulation. Distribution mean and standard deviation are represented with green boxes with errorbars (left y axis), while skewness and excess kurtosis are represented with red and blue boxes (right y axis). The negative plane of the y axis is shaded with the colours used to distinguish simulations in Fig.5.4.

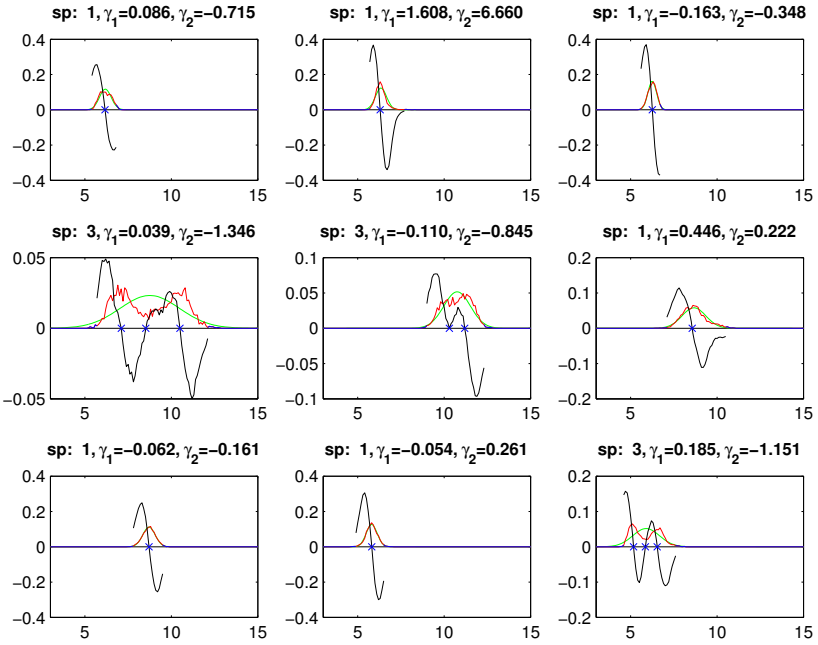


Figure 5.6: The performance of the test for modality is shown on 9 C_α pairs; distributions of their distance computed on the first *replica* (red) are plotted along with a normal distribution with the same mean and standard deviation as the raw data (green), the approximate first derivative (black) computed on the 5-95 interpercentile range and the stationary points marked off by the algorithm (blue X). For each distribution, the number of stationary points found (**sp**) and the values of skewness and kurtosis are reported. The horizontal axes are in Å.

1520 were visually inspected (see Fig.5.6), confirming that the procedure correctly discriminates the stationary points of the distributions.

The number of pairs whose distance features unimodal distribution has been computed for the 4 *replicas*, and results are reported in Tab.5.2, providing an indication on how often inter-C_α distances sample distinct substates in single ATMD *replicas*. Harmonic behaviour can be ruled out for between 7 and 15% of pairs in each *replica* based solely on the modality of their distance distributions. Comparing the identified pairs among *replicas* helps in strengthening significance, allowing the discrimination of pairs that exhibit uniform behaviour. Only 5 pairs are consistently multimodal in all *replicas*, while more than a quarter of all pairs (407) are multimodal in at least one *replica*. These figures confirm the previously observed heterogeneity among *replicas*, and the influence of global structure on the shape of distance distri-

set	unimodal	γ_1<.3, γ_2<.3	γ_1<.4, γ_2<.4	γ_1<.31, γ_2<.54	γ_1<.5, γ_2<.5
replica 1	1340	447	671	607	824
replica 2	1293	478	635	586	759
replica 3	1412	484	688	622	845
replica 4	1397	568	764	696	894
$\mathbb{T}$	1514	967	1184	1112	1289
$\mathbb{C}$	1112	134	242	217	361
$\mathbb{N}$	6	553	336	408	231

Table 5.2: Number of pairs characterised as featuring unimodal distributions on the basis of the derivative analysis (second column), and number of pairs featuring normal distribution according to various skewness and kurtosis cut-offs (third to sixth columns). The total number of pairs taken in consideration is 1520; it should be noted that the search whose results are reported in the last four columns was performed on the pairs in the first column. The numbers are reported separately for each *replica*; the number of pairs matching the criteria in at least one of the *replicas* ($\mathbb{T}$), those matching in all *replicas* ($\mathbb{C}$) and those not matching in all *replicas* ($\mathbb{N}$) are also reported. For values reported in the fifth column, it holds $\mathbb{T} \equiv \mathbb{U}$, $\mathbb{C} \equiv \mathbb{H}$ and $\mathbb{N} \equiv \mathbb{A}$.

butions. Even in the most stringent scenario, almost 75% of pairs are found to be unimodal, and are thus left to be characterised by means of the statistical descriptors. Handling this large set imposes the formulation of some assumptions on the definition of normality for the purpose of this study, and in particular the designation of a cut-off for skewness and kurtosis that can best estimate normality of the distributions.

Using a single cutoff value for both skewness and kurtosis could represent a simple strategy to identify normally-distributing pairs. To assess whether this could be feasible, the number of pairs with unimodal distance distributions displaying absolute values of γ_1 and γ_2 below various cutoff values are plotted in Fig.5.7 for the 4 *replicas* and the ENMBD, ENMMD and CGMD simulations studied in the previous sections (5.2 and 5.3, see also Tab.5.2).

In the case of ENMBD, most pairs feature low values of the two parameters, and a cutoff of 0.5 identifies almost 95% of the total number of unimodal pairs, which in this case coincides with the totality of analysed pairs (1374, represented with a horizontal line). It is therefore not surprising that the initial 10-pairs test-set consistently exhibited γ_1 and γ_2 values below 0.5. Very similar considerations can be made for ENMMD, the other hENM-based simulation. The curves computed from ATMD simulations always lie below the ENMBD curve, showing that at any cutoff value, the number of gaussian-distributing pairs

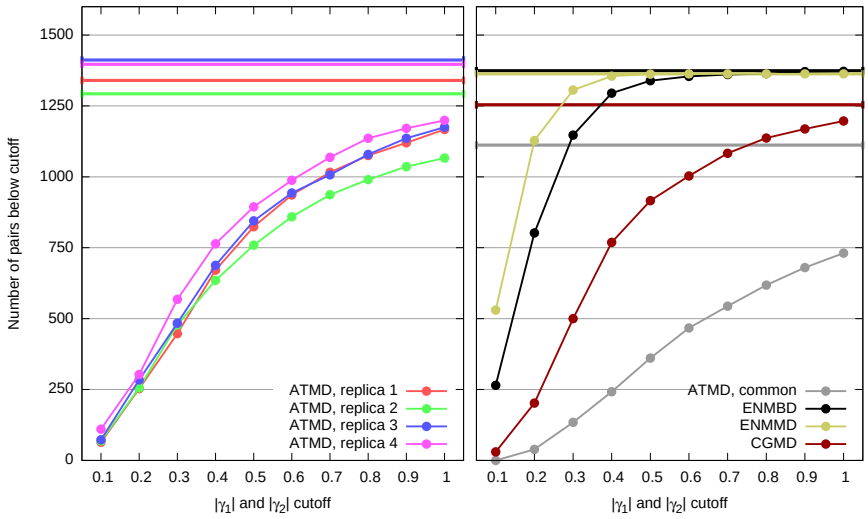


Figure 5.7: The number of pairs featuring absolute values of γ_1 and γ_2 below various cutoff values are plotted for *replicas* on the left, and for the ENMBD, ENMMD and CGMD simulations on the right, along with the number of common pairs among the 4 sets on the left, i.e., the number of pairs exhibiting γ_1 and γ_2 below each cutoff consistently in the 4 *replicas* (set $\mathbb{C}$ in Tab.5.2). The horizontal lines represent the total number of unimodal pairs, among which the selection on the basis of the cutoff is made (first column of Tab.5.2).

is lower in the former, and represents a much smaller fraction of the total number of unimodal pairs. The CGMD simulation has a behaviour which is strikingly similar to that of single ATMD *replicas*; although the curves are not directly quantitatively comparable, since the number of pairs studied in the former simulation is lower than in any of the latter, this shows that also in this aspect CGMD the simulation exhibits an intermediate behaviour between the all-atomic and elastic descriptions, as previously observed in the fluctuational analysis (see 5.3). In both cases, the number of identified pairs increases monotonically with increasing cutoff, and even almost linearly when the *common* set is taken into account, i.e., the pairs that are consistently identified at a given cutoff value in all *replicas*. Thus using a combined cutoff no value is clearly more discriminating, suggesting that, at least in this direction, normal pairs are not noticeably separated from others.

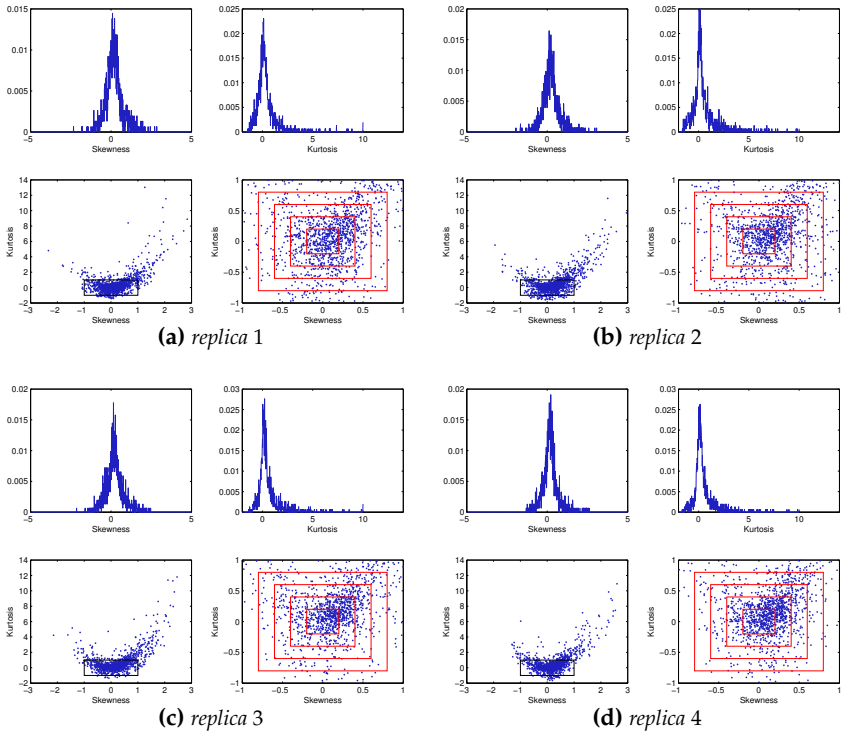
To investigate the features of the single central moments, the relationship between skewness and kurtosis has been evaluated for all pairs by computing their distributions and scatter plots for each *replica*

(Fig.5.8). The plots are very similar among *replicas*; the distributions of both parameters have peaks around 0, and are right-skewed, i.e., show a preference for positive values, especially kurtosis. This feature is also evident from the V-shaped scatter plots, which clearly expose the asymmetry with respect to kurtosis, and suggest that distributions of inter-C_α distances in the simulations feature only a limited subset of the possible (γ_1,γ_2) combinations. The scatter plots also show that the immediate surroundings of the origin are more densely populated, and extend in two almost symmetric tails, illustrating that, in absolute terms, the two parameters increase concurrently, and that distance distributions are more often less "peaked" than gaussians ($\gamma_2 > 0$), thus showing a preference for outliers, and in particular for distances greater than the average value ($\gamma_1 > 0$). This suggests that inter-C_α distances in ATMD simulations are less restrained than in an elastic network, and tend to have greater freedom to extend, as they would in a Morse-shaped potential energy well.

To relate to the results obtained using absolute value cutoffs (Fig.5.7), a closer view of the scatter plots around the origin is provided, and decorated with rectangles which encompass the pairs identified at some of the absolute value cutoffs used. It is apparent that the increasing number of pairs identified at increasing cutoff values results from a non-uniform distribution of pairs in the range, as a consequence of the aforementioned asymmetric spread of the two parameters.

On the other hand, hENM-based simulations show a very different disposition of skewness and kurtosis values, reflecting the harmonic character of distance fluctuations (Fig.5.9). The distributions are centred at 0, but tighter and less skewed than in ATMD, resulting in a narrow cloud in the scatter plots; comparing ENMBD and ENMMD suggests that greater sampling is instrumental in obtaining an even more consistent picture. The clouds also exhibit strong symmetry about both axes, pointing out that the use of an absolute value cutoff is a reasonable choice; their different extension in the two direction though recommends the use of two different cutoffs for the two central moments. Therefore, the cutoff values 0.31 and 0.54 for γ_1 and γ_2 respectively have been computed as the average absolute values of the 1^{st} and 99^{th} percentiles of their distributions as extracted from the ENMBD simulation.

These cutoff values have been used to determine normality among



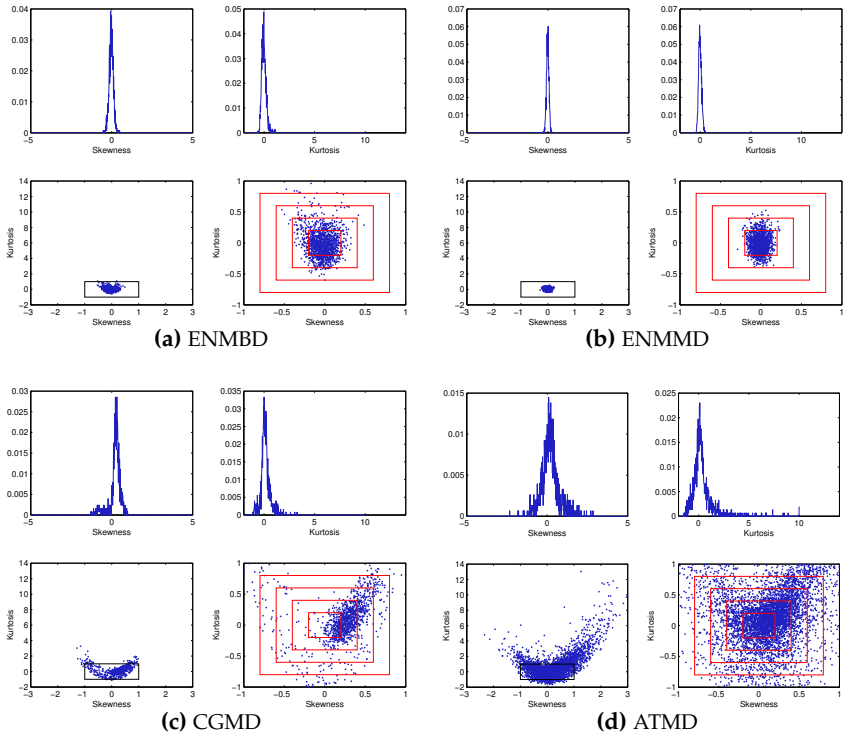
(a) *replica* 1 **(b)** *replica* 2

(c) *replica* 3 **(d)** *replica* 4

Figure 5.8: Analysis of the relationship between skewness and kurtosis of distance distributions, plotted for the 4 *replicas* (top, *replicas* 1 and 2, bottom, *replicas* 3 and 4). For each *replica*, 4 plots are shown: top row, the distributions of skewness (left) and kurtosis (right); bottom row, scatter plot of their relationship (left), and zoom of the immediate surroundings of the origin (right). The latter plots are decorated with red rectangles, which encompass the pairs identified at some of the absolute value cutoffs used in Fig.5.7.

the previously determined unimodal distance distributions; the selected pairs (see Tab.5.2) have then been compared in terms of secondary structure, solvent accessibility and distance.

The results of this comparison are shown in Fig.5.10, where pairs have been divided in groups on the basis of the properties of the residues bearing their constituting C_α, 6 for secondary structure, 6 for SAS and 12 for inter-C_α distance (see caption of Fig.5.10 for details). The histograms illustrate, separately for each *replica*, the quantities and percentages pertaining to 4 different sets of pairs: *(i)* the total number of unimodal pairs exhibiting the particular set of values of the studied property ($\mathbb{U}$, very-low-saturation-coloured boxes); of these, *(ii)* the number of pairs that have been identified as behaving harmonically



(a) ENMBD **(b)** ENMMD

(c) CGMD **(d)** ATMD

Figure 5.9: Analysis of the relationship between skewness and kurtosis of distance distributions, plotted for the ENMBD (top left), ENMMD (top right), CGMD (bottom left) and ATMD (bottom right) simulations. The last plot is obtained combining the four plots in Fig.5.8. See caption of that figure for details on the plots.

according to our criteria (i.e., unimodal distribution with absolute γ_1 and γ_2 values below the chosen cutoffs, $\mathbb{H}_r$, low-saturation-coloured boxes); among these *(iii)* the number of pairs that feature harmonic behaviour consistently in *all replicas* ($\mathbb{H}$, full-saturation-coloured boxes) and finally, *(iv)* the number of pairs that, on the contrary, have been rejected by our algorithm consistently in *all replicas* ($\mathbb{A}$, dark coloured boxes). These latter values have been plotted in the negative direction for clarity, and also to represent visually the fact that while for the first three sets it holds $\mathbb{U} \supseteq \mathbb{H}_r \supseteq \mathbb{H}$, the fourth set is unrelated to the second and third, but $\mathbb{A} \subseteq \mathbb{U}$.

The histograms show that sets $\mathbb{U}$ and $\mathbb{H}_r$ exhibit consistent behaviour among *replicas*; nevertheless, comparison of sets $\mathbb{H}_r$ and $\mathbb{H}$ implies that small fractions of the pairs in $\mathbb{H}_r$ are common to all *replicas*, confirming the previously observed inter-*replica* variability of har-

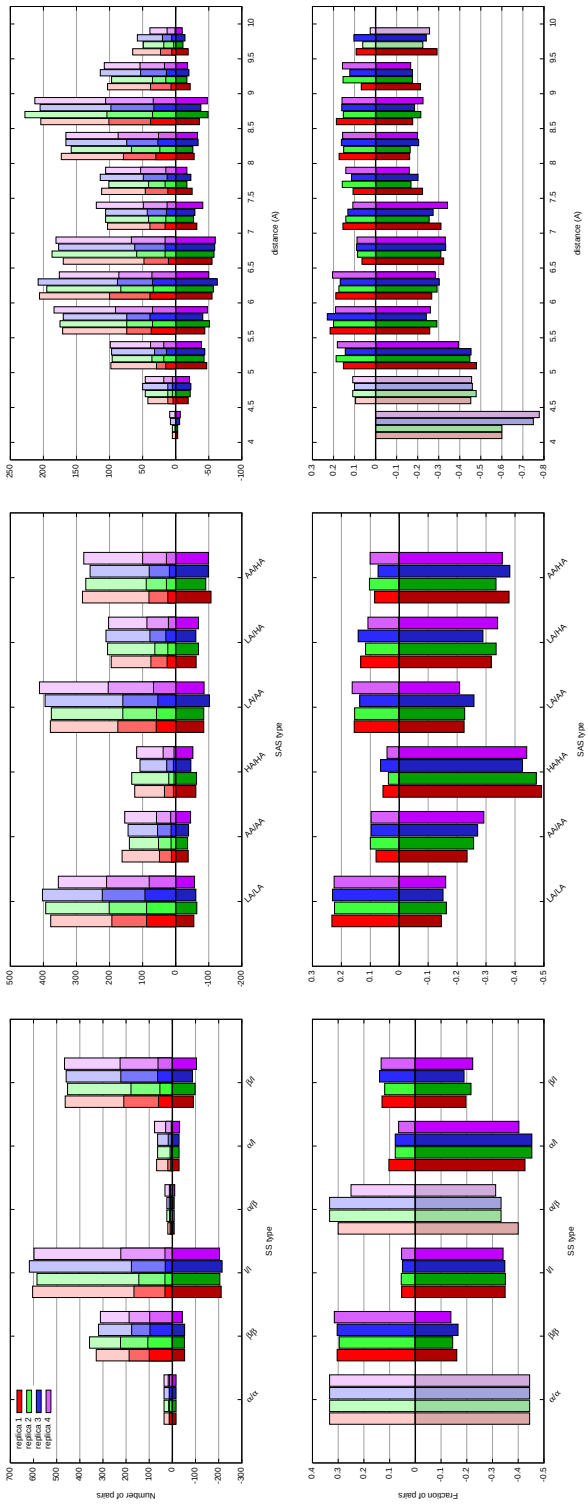


Figure 5.10: Comparison of harmonicity of distance fluctuations with structural properties of mAST. The simple *SS type* and *SAS type* categorisation of pairs has been employed (see the Methods section 5.4.1). The range of average distances has been divided into 12 categories, 0.5 Å-wide. Histograms in the top row display the number of pairs $n_{\mathbb{U}}, n_{\mathbb{H}_r}, n_{\mathbb{H}}$ and $n_{\mathbb{A}}$ of sets $\mathbb{U}$, $\mathbb{H}_r$, $\mathbb{H}$ and $\mathbb{A}$, separated in the relevant *SS type*, *SAS type* and distance range groups (see text for definitions of the sets). Histograms in the bottom row show the ratios $n_{\mathbb{H}}/n_{\mathbb{U}}$ (positive axis) and $n_{\mathbb{A}}/n_{\mathbb{U}}$ (negative axis); boxes are coloured with low-saturated colours when $n_{\mathbb{U}} < 50$.

monicity (see also Tab.5.2). Therefore, to provide a significant description of the relation between harmonicity and the considered structural properties, only the features of set $\mathbb{H}$ will discussed in detail, and compared to set $\mathbb{A}$ to assess differential propensity of pairs to consistently display harmonic or anharmonic behaviour.

The studied enzyme, mAST, contains only two α-helical structures, both in peripheral position with respect to the central β structures that constitute the main features of its fold. The number of pairs involving these structures is therefore low, and the fraction of harmonic pairs computed for α-helices is poorly significant. Analysis of the other secondary structural elements shows that almost 30% of pairs between β structures feature harmonic behaviour, and about the same fraction of pairs between *loop* structures feature anharmonic behaviour. Considering the striking preference of *loop* structures for anharmonicity (for l/l, $\mathbb{A}$ is between 6 and 7 times more populated than $\mathbb{H}$), and adding to the picture pairs between differently structured residues, apparently confirms that one unstructured pair element is sufficient to shift the pair's behaviour towards anharmonicity; further insight into this hypothesis will be provided below. Overall, these observations provide a clear and significant demonstration of the role of secondary structure, and in this case especially β-strands, in determining backbone fluctuational character, confirming the hypotheses formulated analysing the 10-pair test-set.

On the contrary, the existence of a relation with SAS was unclear from the aforementioned small test-set. Extension to the entire set of inter-C_α pairs shows that indeed such a relation exists, and that increasing SAS the fraction of harmonic pairs decreases while that of anharmonic pairs increases, so that pairs between low accessibility residues (LA/LA) have more than 3 times more frequently harmonic behaviour than pairs between highly accessible residues (HA/HA), which in turn have more than 3 times more frequently anharmonic behaviour than LA/LA. This trend holds, although to a lesser extent (the extreme clusters have a ratio of about 1.5) when analysing mixed pairs, i.e., pairs between residues belonging to different SAS categories, and suggests an approximately linear (inverse) proportionality between harmonicity and SAS.

The last property taken into consideration in this comparative analysis is average distance between the C_α constituting the pair; differently

from the other analysed properties, this one is characteristic of the pair, and not of the residues bearing its constituent C_α. Although pair distance fluctuations have been monitored on a much wider range, the histograms in Fig.5.10 consider only average distances between 4 and 10 Å, since pairs above this value are too dispersed to allow significant analysis; furthermore, more than 90% of the total pairs are accounted for in these histograms.

The distribution of $\mathbb{U}$ along the average distance axis is bimodal, with a first peak in the interval [6.0,6.5]Å, and a second, slightly higher, in [8.5,9.0]Å. The distribution of both harmonic and anharmonic pairs grossly retain this shape, the latter exhibiting higher values than the former throughout the entire distance range. There seems thus not to be a distance range at which harmonic behaviour is more frequent, as was previously seen, e.g., for the β/β and LA/LA groups. However, interesting variations are seen in the relative occurrence of harmonicity at different average distances.

The low population in the first two intervals, and in the last one for *replicas* 2 and 4, hinders significance of the corresponding intervals in the fraction histogram; nevertheless, this plot shows that in the first third of the distance range (up to 6 Å), the fraction of harmonic pairs increases, while anharmonic pairs decrease. This gives rise to the global maximum of the harmonic fractions distribution ([5.5,6.0]Å), in correspondence with a local minimum of the anharmonic fractions. This trend is inverted in the following two intervals, so that on the right side of the first peak in the distribution of $\mathbb{U}$ ([6.0,6.5]Å), the ratio of anharmonic to harmonic pairs is doubled ([6.5,7.0]Å) with respect to its left side ([5.5,6.0]Å). In the rest of the distance range (from 7.5 Å on) the number of harmonic and anharmonic pairs more closely follow the shape of the distribution of $\mathbb{U}$, as shown by their relative amounts converging to similar and almost constant values of about 20%.

These observations set forth that a sizable and complex relation exists between harmonic behaviour and average distance of C_α pairs which deserves further characterisation, and are striking in that they suggest that proximal pairs have greater preference for anharmonic behaviour.

As a final effort in this study, the results of the harmonicity analysis are transformed in order for them to be visualised in terms of mAST's sequence and three-dimensional (3D) structure. For each residue, the

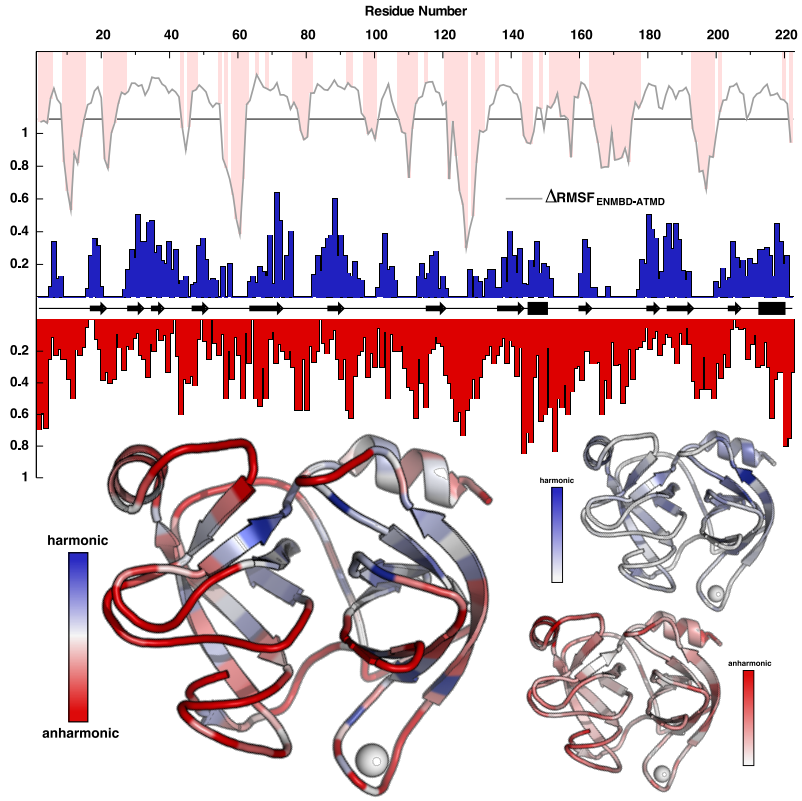


Figure 5.11: The fraction of harmonic ($n_{\mathbb{H}}/n_{\mathbb{U}}$, blue histogram) and anharmonic ($n_{\mathbb{A}}/n_{\mathbb{U}}$, red histogram) pairs for each residue is plotted along the sequence of mAST. Between the histograms, a schematic representation of the most frequently attained secondary structure per residue is provided. The two histograms are mapped on the 3D structure of mAST using two accordingly coloured gradients (bottom right). Normalised $r_{\mathbb{H}/\mathbb{A}}$ values are mapped on the 3D structure using a combined gradient (bottom left). The ΔRMSF difference profile, obtained subtracting the ENMBD-derived profile from that of ATMD, is plotted in grey; a horizontal black line shows the position of the 0.1 Å value in the difference profile. See text for the description of the profile's shading.

fraction of pairs in which it features that are harmonic and anharmonic has been calculated, and plotted as a function of residue number (Fig.5.11). The plot gives an approximate relative idea of how harmonic or anharmonic is the behaviour of each residue. It is clear that there is no preferential localisation along the sequence, and peaks are well distributed throughout the abscissa. A higher fraction of harmonic pairs is consistently found in correspondence with (as seen in Fig.5.10), and in the immediate surroundings of stable secondary structural elements,

while the largest peaks in the anharmonicity profile are often located in correspondence with *loops*.

The two profiles are also shown mapped on the 3D structure of mAST (PDB entry 1A0J [35], chain A), using two different gradients (Fig.5.11, bottom right); a third structure is shown, coloured according to the ratio $r_{\mathbb{H}/\mathbb{A}}(i) = n_{\mathbb{H}}(i)/(n_{\mathbb{A}}(i) + n_{\mathbb{H}}(i))$, where $n_{\mathbb{H}}(i)$ and $n_{\mathbb{A}}(i)$ are the numbers of harmonic and anharmonic pairs involving residue i. $r_{\mathbb{H}/\mathbb{A}}$ ranges from 0, indicating that all pairs are anharmonic, to 1 when all pairs are harmonic. The 3D representation shows that higher $r_{\mathbb{H}/\mathbb{A}}$ values cluster on the β-strands at the core of the protein, on the sides of the β-barrels facing each-other and that form the interface between the two domains of mAST. On the contrary, low $r_{\mathbb{H}/\mathbb{A}}$ values appear most often in the periphery of the structure: the dependence of anharmonicity on SAS is further evidenced by the striking observation that the two helices consistently feature lower $r_{\mathbb{H}/\mathbb{A}}$ values on their solvent-facing side. The character of the fluctuations seems thus to change dramatically when moving from the inside towards the outside of mAST: whether this effect is due to changes in the surrounding hydrophobicity or in local density, which are known to exist between the surface and the core of globular proteins, or even to a specific role of the solvent is an issue of great interest.

Detailed inspection of the 3D distribution of $r_{\mathbb{H}/\mathbb{A}}$ allows to notice that often high-$r_{\mathbb{H}/\mathbb{A}}$ regions are localised on corresponding stretches of contiguous β-strands in the β_barrels. Since the main feature of β-strand in strands is that of being cross-linked by hydrogen bonds, this observation suggests a possible role of these interactions in determining harmonic fluctuational behaviour, and opens the issue of the importance of non-covalent weak interactions in locally shaping the effective inter-C_α potential: this issue certainly deserves further study.

Finally, in order to relate these findings with the differences in fluctuations observed at the beginning of this chapter, the RMSF difference (ΔRMSF) between profiles obtained from the ENMBD and ATMD simulations is shown in Fig.5.11: lower (more negative) values of the difference profile indicate that the ATMD trajectory features greater flexibility for the residue. The area above the plot is shaded in correspondence of those residues that feature a value of $r_{\mathbb{H}/\mathbb{A}} \leq 0.25$. In other words, those residues that are involved three times more often in anharmonic-behaving than in harmonic-behaving pairs, and that may

therefore be considered "intrinsically" anharmonic. It is striking to note that all peaks in the difference profile greater than 1 Å are found in correspondence with highly anharmonic residues, definitely pointing out that the differences in the entity of the fluctuations that can be described by the ENMBD and ATMD approaches are to a large extent ascribable to the harmonic character of those fluctuations.

5.5 Conclusions

Life at the molecular level is inherently a multi-scale phenomenon. The overwhelmingly complex network of finely regulated processes that sustain life, spanning orders of magnitude both in time and space from electron transfer in enzyme active sites to mitosis, achieves its perfection through seamless communication between all time and length-scales. Its definition as multi-scale originates from the fact that a detailed analysis of these mechanisms must be made by cutting layers through the continuity to choose the perfect balance between size and accuracy.

Making such a choice may reveal itself as a task on its own, and this seems to be the case in the study of cold-adaptation of enzymes, which is both a local and global process in which the entirety of the protein may have a role, and which has been shown to feature heterogeneous patterns that can thus be expected to be strongly constrained by sequence and structure, as has been shown for elastases. In such a perspective, large-scale, evolutionary-based comparative analyses of the molecular-detailed dynamical properties of homologous enzymes adapted to different temperatures may be the only tool available to unveil the mechanical determinants of enzymatic cold adaptation.

For these reasons it would be relevant to devise a simplified but accurate approach to perform these comparative analyses at a higher throughput than is currently feasible; a similar effort can be envisaged today owing to the ever increasing number of detailed comparisons available providing a solid base for the development. Furthermore, the ongoing outstanding interest that has recently been enjoyed by simplified physical methods to study the dynamics of biomolecules ensures constructive collaboration with a heterogeneous and motivated community on a topic under active development.

The aim of this chapter is to shed light on the applicability of a

particular "coarse-grained" model to the study of the mechanical determinants of enzymatic cold-adaptation. The comparison between an elastic-network-based approach and all-atom molecular dynamics simulations carried out in this chapter has shown that the former consistently yields lower fluctuations. The provided description of protein flexibility is qualitatively similar to that at full atomic detail, but when analyses are performed that require quantitative accuracy, such as the comparison between homologous enzymes adapted to different temperature conditions, the approach reveals it is not sensitive enough, at least in the context of this study. Furthermore, detailed analysis of harmonicity in ATMD has shown that observed differences in the atomic fluctuations between hENM-based approaches and ATMD may effectively be due to the anharmonic character of a significant fraction of these fluctuation, but has also found significant correlation between harmonicity, or the lack thereof, and the structural properties with which it has been compared, especially secondary structuring and structural burial/solvent accessibility, at least in the case of the mesophilic Atlantic Salmon trypsin. Extending this analysis to other enzymes featuring diverse folds and functions, may provide sufficient information of general validity to guide the formulation of an extended network approach that takes advantage of more than just the topological information encoded in the structure to improve its accuracy and sensitivity.

Bibliography

[1] Valentina Tozzini. Coarse-grained models for proteins. *Curr Opin Struct Biol*, 15(2):144–150, Apr 2005.

[2] Valentina Tozzini. Multiscale modeling of proteins. *Acc Chem Res*, Sep 2009.

[3] MM Tirion. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *PHYSICAL REVIEW LETTERS*, 77(9):1905–1908, AUG 26 1996.

[4] A. R. Atilgan, S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin, and I. Bahar. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys J*, 80(1):505–515, Jan 2001.

[5] K. Hinsen, A. J. Petrescu, S. Dellerue, M. C. Bellissent-Funel, and G. R. Kneller. Harmonicity in slow protein dynamics. *Chemical Physics*, 261(1-2):25–37, 2000. 31 ELSEVIER SCIENCE BV AMSTERDAM Sp. Iss. SI 368MT.

[6] Isabelle Navizet, Fabien Cailliez, and Richard Lavery. Probing protein mechanics: residue-level properties and their use in defining domains. *Biophys J*, 87(3):1426–1435, Sep 2004.

[7] Sophie Sacquin-Mora and Richard Lavery. Investigating the local flexibility of functional residues in hemoproteins. *Biophys J*, 90(8):2706–2717, Apr 2006.

[8] Sophie Sacquin-Mora, Emilie Laforet, and Richard Lavery. Locating the active sites of enzymes using mechanical properties. *Proteins*, 67(2):350–359, May 2007.

[9] Richard Lavery and Sophie Sacquin-Mora. Protein mechanics: a route from structure to function. *J Biosci*, 32(5):891–898, Aug 2007.

[10] Karine Bastard, Chantal Prévost, and Martin Zacharias. Accounting for loop flexibility during protein-protein docking. *Proteins*, 62(4):956–969, Mar 2006.

[11] Martin Zacharias. Protein-protein docking with a reduced protein model accounting for side-chain flexibility. *Protein Sci*, 12(6):1271–1282, Jun 2003.

[12] Bertil Halle. Flexibility and packing in proteins. *Proc Natl Acad Sci U S A*, 99(3):1274–1279, Feb 2002.

[13] Sophie Sacquin-Mora, Pierre Sebban, Valérie Derrien, Bernhard Frick, Richard Lavery, and Christiane Alba-Simionesco. Probing the flexibility of the bacterial reaction center: the wild-type protein is more rigid than two site-specific mutants. *Biochemistry*, 46(51):14960–14968, Dec 2007.

[14] Hendrik Dietz, Felix Berkemeier, Morten Bertz, and Matthias Rief. Anisotropic deformation response of single protein molecules. *Proc Natl Acad Sci U S A*, 103(34):12724–12728, Aug 2006.

[15] Sophie Sacquin-Mora and Richard Lavery. Modeling the mechanical response of proteins to anisotropic deformation. *Chemphyschem*, 10(1):115–118, Jan 2009.

[16] D. L. Ermak and J. A. McCammon. Brownian dynamics with hydrodynamic interactions. *Journal of Chemical Physics*, 69(4):1352–1360, 1978.

[17] K. Hinsen. The molecular modeling toolkit: A new approach to molecular simulations. *Journal of Computational Chemistry*, 21(2):79–85, 2000. 23 JOHN WILEY & SONS INC NEW YORK 271NR.

[18] V. Tozzini and J. A. McCammon. A coarse grained model for the dynamics of flap opening in hiv-1 protease. *Chemical Physics Letters*, 413(1-3):123–128, 2005.

[19] Valentina Tozzini, Joanna Trylska, Chia en Chang, and J. Andrew McCammon. Flap opening dynamics in hiv-1 protease explored with a coarse-grained model. *J Struct Biol*, 157(3):606–615, Mar 2007.

[20] Adam Górecki, Marcin Szypowski, Maciej Długosz, and Joanna Trylska. Redmd–reduced molecular dynamics package. *J Comput Chem*, 30(14):2364–2373, Nov 2009.

[21] Joanna Trylska, Valentina Tozzini, and J. Andrew McCammon. Exploring global motions and correlations in the ribosome. *Biophys J*, 89(3):1455–1463, Sep 2005.

[22] H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, A. Di Nola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *J Chem Phys*, 81(8):3684–3690, 1984.

[23] Alejandra Leo-Macias, Pedro Lopez-Romero, Dmitry Lupyan, Daniel Zerbino, and Angel R Ortiz. An analysis of core deformations in protein superfamilies. *Biophys J*, 88(2):1291–1299, Feb 2005.

[24] Sandra Maguid, Sebastián Fernández-Alberti, Gustavo Parisi, and Julián Echave. Evolutionary conservation of protein backbone flexibility. *J Mol Evol*, 63(4):448–457, Oct 2006.

[25] Alessandro Pandini, Giancarlo Mauri, Annalisa Bordogna, and Laura Bonati. Detecting similarities among distant homologous proteins by comparison of domain flexibilities. *Protein Eng Des Sel*, 20(6):285–299, Jun 2007.

[26] Sandra Maguid, Sebastian Fernandez-Alberti, and Julian Echave. Evolutionary conservation of protein vibrational dynamics. *Gene*, 422(1-2):7–13, Oct 2008.

[27] Jianpeng Ma. Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. *Structure*, 13(3):373–380, Mar 2005.

[28] S. Hayward, A. Kitao, and N. Gō. Harmonic and anharmonic aspects in the dynamics of bpti: a normal mode analysis and principal component analysis. *Protein Sci*, 3(6):936–943, 1994.

[29] S. Hayward, A. Kitao, and N. Gō. Harmonicity and anharmonicity in protein dynamics: a normal mode analysis and principal component analysis. *Proteins*, 23(2):177–186, 1995.

[30] D. Janezic, R. M. Venable, and B. R. Brooks. Harmonic-analysis of large systems .3. comparison with molecular-dynamics. *Journal of Computational Chemistry*, 16(12):1554–1566, 1995. 25 JOHN WILEY & SONS INC NEW YORK TE611.

[31] F. Pontiggia, G. Colombo, C. Micheletti, and H. Orland. Anharmonicity and self-similarity of the free energy landscape of protein g. *Physical Review Letters*, 98(4):4, 2007. Pontiggia, F. Colombo, G.

Micheletti, C. Orland, H. 23 AMERICAN PHYSICAL SOC COLLEGE PK 130HD.

[32] A. Kitao, S. Hayward, and N. Go. Energy landscape of a native protein: Jumping-among-minima model. *Proteins-Structure Function and Genetics*, 33(4):496–517, 1998. 73 WILEY-LISS NEW YORK 144TQ.

[33] A. Amadei, B. L. de Groot, M. A. Ceruso, M. Paci, A. Di Nola, and H. J. C. Berendsen. A kinetic model for the internal motions of proteins: Diffusion between multiple harmonic wells. *Proteins-Structure Function and Genetics*, 35(3):283–292, 1999. 32 WILEY-LISS NEW YORK 188TX.

[34] Abraham. Savitzky and M. J. E. Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8):1627–1639, July 1964.

[35] H. K. Schrøder, N. P. Willassen, and A. O. Smalås. Structure of a non-psychrophilic trypsin from a cold-adapted fish species. *Acta Crystallogr D Biol Crystallogr*, 54(Pt 5):780–798, Sep 1998.