UiT
THE ARCTIC
UNIVERSITY
OF NORWAY

Faculty of Science and Technology

Department of Chemistry

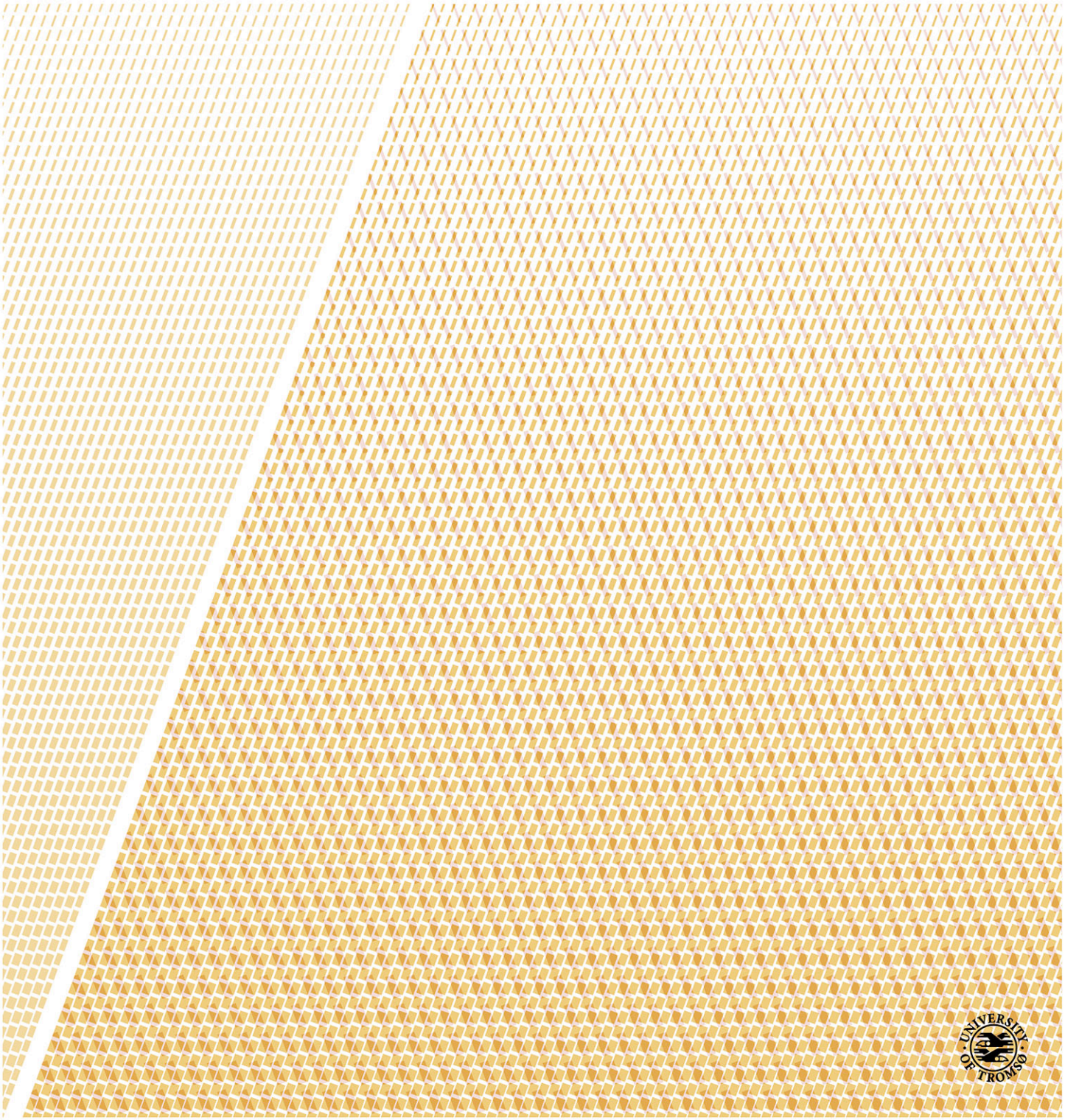
Surface properties tune thermal adaptation of enzymes

Computational studies of endonuclease A

—

Davide Michetti

A dissertation for the degree of Philosophiae Doctor – April 2017



Advisor

Prof. Bjørn Olav Brandsdal

Co-advisor

Prof. Nils Peders Willassen

Submitted in partial fulfillment of the requirements

for the degree of Philosophiae Doctor

to the

Faculty of Science and Technology

Uit - The Arctic University of Norway

Tromsø, February 2017

Abstract

Cold-adapted enzymes (CAEs or psychrophiles) display stunning properties thanks to adaptation to low temperature environments (< 15 °C). In particular, the lower thermostability and higher specific activity compared to warm-adapted (or mesophilic) homologs have attracted the interest of researchers and industries. In the case of basic research, this class of enzymes is an exciting model for studying the relationship between enzyme structure and activity/stability. Different structural and sequence hypotheses have been suggested to account for enzyme cold adaptation, but still a clear picture is lacking despite decades of studies in the field. However, it has been shown that psychrophilic reactions are less dependent on temperature through lower activation enthalpies.

In my PhD project I have studied the endonucleases A from *Aliivibrio salmonicida* (VsEndA, cold-adapted) and *Vibrio cholerae* (VcEndA, warm-adapted). The goal was to compare these two systems, in order to find structural differences that could explain the mechanism of cold adaptation in the psychrophilic EndA. These two enzymes have previously been studied experimentally and the crystal structures have been solved. We first employed Molecular Dynamics (MD) simulations to characterize and compare the dynamical behavior of the two homologs and the underlying interaction network. RMSF analysis did not show a significant increase in the flexibility of VsEndA compared to VcEndA and the active sites have similar mobility, contrary to general findings and assumptions in previous cold adaptation studies. There are, however, few regions with slightly increased RMSF values in both enzymes and in particular the psychrophilic C-terminal displays higher flexibility than the mesophilic one. Protein structure network (PSN) analysis revealed two clusters of ion-pairs in this region of VcEndA (E179-R222-E226 and D210-R225-E214). MD simulations of VsEndA incorporating the R222 salt-bridge network, display a decrease of the C-terminal RMSF

compared to the *wild-type* (WT) enzyme, suggesting a possible role in thermostability for this electrostatic cluster in the mesophilic variant. We furthermore aligned the sequences of VsEndA and VcEndA together with other homologs from different thermal environments, in order to pinpoint amino acid substitutions shared only by psychrophilic enzymes. The amino acid substitutions T120V, S141I and A166S emerged from these studies as potential important mutations for temperature adaptation.

In order to develop an Empirical Valence Bond (EVB) model to evaluate reaction thermodynamics, a clear understanding of the reaction mechanism and the conformations of the different states of the EndA reaction is fundamental. In this respect, we first carried out Quantum Mechanical (QM) studies of the enzymatic reaction, computing the relevant energetics. Our calculations confirmed a concerted mechanism. Subsequent EVB calculations of the reaction in VcEndA gave activation free energies consistent with experimental data. *In silico* calculations of Arrhenius plots by EVB simulations reproduced the activation parameters for the reaction of VsEndA and VcEndA, displaying the characteristic trends in activation enthalpy ($\Delta H^{\ddagger}$) and entropy ($\Delta S^{\ddagger}$) for cold-adapted enzymes. Furthermore, the lower $\Delta H^{\ddagger}$ in the psychrophilic enzyme was reflected in a lower internal energy change of the system along the reaction for the regions outside the active site. This observation was further tested in our EVB simulations by gradually imposing positional restraints on the outer parts of VsEndA structure and increasing in this way the rigidity of these regions. A lower surface mobility could, indeed, transform the psychrophilic enzyme activation parameters to resemble those of the mesophilic counterpart. We subsequently assessed with EVB the effect of mutations on the reaction catalyzed by VsEndA/VcEndA. The EVB models for the EndA homologs estimated an effect on activation parameters for the mutation S141I, located on the enzyme surface (previously identified by multiple sequence alignment). Altogether these data suggest that surface properties enable VsEndA to adapt to low temperatures.

Acknowledgments

My adventure in the far North began in 2011, one year before I started my PhD when as a fresh master student I decided to work on my thesis abroad. Through the guidance of my supervisor Prof. Luca Di Gioia and co-supervisor Elena Papaleo I was directed for the first time to the group of Prof. Bjørn Olav Brandsdal. My first thanks then goes to them who made possible the beginning of my scientific career. In my PhD years I had the fortune of maintaining a collaboration with them, and with the contribute and advice of Elena Papaleo I wrote my first paper.

This would not have been possible, however, without the help and trust of my supervisor Bjørn Olav Brandsdal. To him goes my deepest gratitude for his guidance and support during my PhD. For the 10˙000 I apologize for my constant late alarm in the morning. Secondly, big thanks to Geir Villy Isaksen for his continuous technical support and for all his scripts and softwares. My PhD would have lasted 10 years more without him. Thanks to the University of Tromsø and to Norway for paying my salary and to all CTCC for welcoming me between their lines. Thanks to Laura Liikanen, Ravdna Sarre, Davide Bon, Tor Arne Andberg and Glenn Robert Morello to make our office the best and funniest of all in the chemistry department.

Outside the work environment I want to acknowledge all my Tromsø family and in particular my roommates: Roberto Di Remigio, Thibaud Freyd, Eivind and Audun Johannson, Jean-Baptiste Kohel and Jennifer Rodriguez. I would not have made it in the cold North without you guys! Love you! Sorry again for all the "Olives"! "Merci beaucoup" to Alex..andre Descomps for his friendship and for helping me through the first year in Tromsø. To the Italian team in Tromsø thanks for the wonderful julebørds and for the football matches "della Nazionale". You brought a bit of Italy in Norway.

Last but not least, thanks to my parents who have supported me in all my studies and made me felt loved. Thanks also to my two sisters and my two dogs. Love you girls! Thanks to all my family to make me feel like I belong. A special mention for the two cousins Andrea Billy Malinverni and Giorgio Jimmy Amigoni, the brothers I never had. A deep gratitude goes to all my friends who have waited for my "I will finish soon" with difficulty.

Finally all my thanks and love go to my girlfriend Chiara Ceruti for bearing with me for a two years distance relationship and for saving me from "la selva oscura".

E per finire non mi resta che ringraziare un´ultima persona:

Grazie Graziella!

List of papers

The PhD thesis is based on the following papers:

I. **Davide Michetti**, Bjørn Olav Brandsdal, Davide Bon, Geir Villy Isaksen, Matteo Tiberti, Elena Papaleo. *A comparative study of cold- and warm-adapted endonucleases A using sequence analyses and molecular dynamics simulations*. Accepted by PlosOne

II. **Davide Michetti**, Glenn Robert Morello, Geir Villy Isaksen, Bjørn Olav Brandsdal. *An Empirical Valence Bond model for the reaction of the endonuclease A from Vibrio cholerae*. Manuscript

III. **Davide Michetti**, Geir Villy Isaksen, Bjørn Olav Brandsdal. *Enthalpy-entropy optimization to survive in the cold: A case study of endonuclease A*. Manuscript

Table of Contents

Acronyms

CAE: Cold-adapted enzyme

CES: Clustering-based Ensemble Similarity

DFT: Density Functional Theory

DRES: Dimensionality-reduction-based Ensemble Similarity

EndA: Endonuclease A

EVB: Empirical valence Bond

FEP: Free Energy Perturbation

FEP-US: FEP-umbrella sampling

KIE: Kinetic Isotope Effect

KDE: Kernel Density Estimation

L-J: Lennard-Jones potential

MD: Molecular Dynamics

MM: Molecular Mechanics

MSA: Multiple Sequence Alignment

NAC: Near Attack Conformation

PCA

PS: Product State

PSN: Protein Structure Network

QM: Quantum Mechanics

RDF: Radial Distribution Function

RMSD/RMSF: Root Mean Square Deviation/Fluctuation

RS: Reactant State

rmsip: Root Mean Square Inner Product

SPE: Stochastic Proximity Embedding

TS: Transition State

VcEndA: *Vibrio cholerae* endonuclease A

vdW: Van der Waals

VsEndA: *Aliivibrio salmonicida* endonuclease A

Vvn: *Vibrio vulnificus* nuclease

Introduction

Enzymes are widely applied as biological tools for industrial, medical and scientific research purposes. Nature offers a rich and broad variety of biocatalysts to select from for the appropriate properties and reaction for these applications. Nonetheless, in order to have full performance and yield, it is necessary to mutate the *wild-type* enzyme for improved features such as stability and/or activity. This task is very challenging, since enzymes are formed by hundreds of amino acids and their 3D shape is determined by even more numerous bonded and non-bonded interactions. Tailored-made enzymes can either be produced through directed evolution, i.e. random mutagenesis and selection of the optimal variants, or in a rational structure-guided manner. The first option has gained more success, despite being limited by the vastness of sequence space and the low frequency of effective mutations. Rational design, instead, requires detailed knowledge of the specific structure and catalytic mechanism. It often fails due to limited understanding of active site amino acids and how their interplay relates to activity and stability. To fill this gap, it is advantageous to study how nature optimizes enzymes for determined environments and reactions. Fortunately, there exist many examples in biological life where these biomolecules change their structure, and so their biochemical properties, to face harsh medium conditions: psychrophiles and thermophiles (adapted to low and high temperatures), halophiles (adapted to high salt concentration), alcalophiles and acidophiles (adapted to basic or acidic environment) etc.

Standard experimental techniques are employed to characterize these classes of enzymes structure (X-ray or NMR), kinetics (Steady State Kinetic analysis), reaction mechanism (Kinetic Isotope Effect or KIE) and thermodynamics (Arrhenius plot), temperature stability (Differential Scanning Calorimetry or Fluorescence quenching) and other medium optima (pH, ion preference etc.), in order to understand their properties. Even though

these methods are fundamental for describing macroscopic properties of cold-adapted enzymes, they cannot explain what happens in the "microscopic world". For this purpose, the only available options are computational methods such as MD simulations or Empirical Valence Bond (EVB). EVB, in particular, has the great advantage of combining structural sampling and free energy calculations. What makes it even more interesting is, however, the ability to derive activation parameters and study their change as a consequence of protein mutagenesis. Thus, it is possible to pinpoint the structural causes of enzyme cold adaptation in ways that no other technique can achieve.

In Chapter 1, I will explain what enzymes are and how they are studied through kinetics and Transition State Theory (TST). A short summary of the theories put forward since the beginning of the 20th century, to explain the origin of enzymatic catalysis will follow. In Chapter 2, the concept of cold-adapted enzymes will be introduced, together with the findings gathered in the last decades and the main conclusions that have been suggested to explain this phenomenon. The enzyme models that have been chosen to investigate psychrophilic enzymes will be presented in Chapter 3. An overview of the computational methods employed in my PhD project will then be given in Chapter 4. The main discoveries of the three articles will be discussed and will be related to literature findings in Chapter 5. Finally, a summary connecting the different findings and concluding remarks will end the thesis (Chapter 6-7).

1

Enzyme Catalysis

Enzymes (from Greek "*en*" and "*zyme*", meaning *in yeast*) are mostly proteins, but a small proportion can also be formed by RNA (RNAzymes) [1,2] or DNA (DNAzymes) [3]. The ability to catalyze chemical reactions with high efficiency and selectivity in biologically useful timescales, makes these biomolecules fundamental for life. Enzymes speed up reactions that "non-helped" would take years to a few seconds or less. They are, in fact, able to accelerate the reaction rate by more than 10 orders of magnitude with respect to the uncatalyzed reaction [4]. Their main catalytic role is to provide an optimal environment for the reaction to occur more rapidly, and, at the same time, to protect the cells from toxic and reactive reaction intermediates, as radicals or charged molecules. On the contrary of other type of catalysts, enzyme activity can be regulated by molecules known as inhibitors or activators. The reason for this is that they work together in synchronicity inside metabolic or transductional pathways, and the ability to tune their activity is fundamental for cellular or multi-cellular organisms organization. Another type of regulation is given by the allosteric effect, where binding of small molecules or effectors induces conformational changes in the enzyme, affecting substrate binding. Allosteric enzymes are often multi-subunit polypeptides, with multiple active sites, and are characterized by cooperative binding, where ligand binding in one subunit enhances the binding affinity in the others.

The binding of substrate and the following reaction take place in a designated part of the enzyme called the active site, where all the important residues for catalysis are located. This region is confined in a pocket of variable size (according to substrate) on the enzyme surface. The specificity for

the substrate is determined by the active site size and shape, its physicochemical properties and by the configuration of the amino acids coating it, in particular the catalytic one. Enzymes, in fact, are able to distinguish between very similar molecules and stereoisomers. Promiscuous enzymes are an exception to this rule, since they are able to bind different ligands and catalyze reactions other than the physiological one [5]. For example, the cytochrome P450, located in the human liver and involved in detoxification of blood molecules, is able to oxidize a broad spectrum of functional groups from structurally different substrates [6].

Enzymes often bind non-proteic biomolecules such as cofactors (ions), coenzymes (vitamin-derived NADH, CoA or non-vitamin ATP) and prosthetic groups (heme group), which assist in catalysis or in structural stability. They work in aqueous solutions and most of them function in mild pH and temperature conditions. Extremophilic enzymes constitute an exception, since they thrive at the limits of temperature for life (psychrophile or hyperthermophile), pH (acidophiles and alkalophiles), NaCl concentration (halophiles) [7] and in non-aqueous solutions (solvent-tolerant) [8].

1.1. Enzyme kinetics

Enzyme kinetics studies the rate and related constants of reactions, and it is the first stage in the characterization of enzymes. It is also a fundamental approach to study reaction mechanisms and to determine the change of the reaction rate upon variation of experimental parameters, such as pH or temperature. Enzyme kinetics is essential, particularly, for the analysis and comparison of enzymes adapted to different temperatures. This class of enzymes displays, in fact, distinct and consistent kinetics trends, as will be further discussed in Chapter 2.

The kinetics of substrate binding and catalysis can be summarized in the classical equation:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} ES \underset{k_{-2}}{\overset{k_2}{\rightleftarrows}} EP \underset{k_{-3}}{\overset{k_3}{\rightleftarrows}} E + P \qquad (1)$$

where the enzyme E binds the substrate S to form an ES complex, with rate of formation $\mathbf{k_1}$ and corresponding opposite rate of dissociation $\mathbf{k_{-1}}$. Once the substrate is bound, the reaction takes place to form an enzyme-product complex (EP), with equilibrium shifted to the right ($\mathbf{k_2 >> k_{-2}}$). Final release of the product regenerates the enzyme to its initial state (E + P), in a rapid step compared to the catalytic one ($\mathbf{k_3 >> k_2}$). Like all other catalysts, enzymes accelerate reaction rate without being consumed or permanently altered, and they increase reaction rates maintaining the chemical equilibrium between reactants and products.

The key factor affecting the rate of enzyme reaction is the concentration of the substrate, [S]. Studying the effect of [S] on the reaction rate is complicated, though, by the fact that it changes during catalysis due to its consumption. Enzyme kinetics is then generally experimentally determined measuring the initial rate, $\mathbf{V_0}$, at different starting substrate concentrations, when [S] >> [E]. Under these conditions, the substrate concentration is almost unchanged and a product is nearly undetectable. The related plot (see **Fig 1**) shows a saturation-kinetics in which, at low [S], $\mathbf{V_0}$ increases linearly, and there is still presence of free enzyme [E]; at higher concentrations, the initial velocity reaches asymptotically its highest value, known as $\mathbf{V_{max}}$, where all the enzyme is present in the bound form ES in Eq. 1, reaching a plateau.

The simplest and best-known model to reproduce this kinetic behavior for single-substrate monomeric enzyme reactions was studied by Michaelis and Menten in 1913. They proposed the reaction scheme in Eq. 2, where the first step of substrate binding is fast and reversible, while the second step is irreversible and rate-determining with first-order rate constant $\mathbf{k_{cat}}$ (corresponding to $\mathbf{k_2}$).

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} EP \rightarrow E + P \qquad (2)$$

In case the reaction proceeds through different intermediate states, $\mathbf{k_{cat}}$ will be a function of the several single rate constants, while in the case of no intermediates it will be equal to $\mathbf{k_2}$. When $\mathbf{V_0 = V_{max}}$, the system is in *steady-state*, since the concentration of the ES complex is constant and the initial rate depends on it.

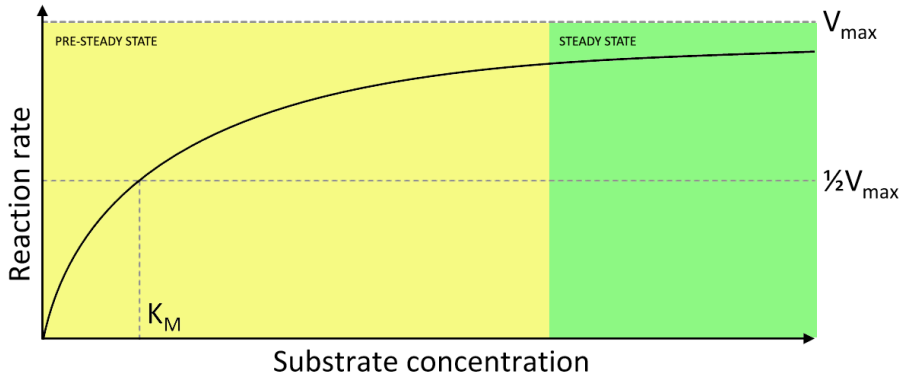


Fig 1: Variation of the initial reaction rate (V_0) with respect to the substrate concentration, [S]. *V_{max} is the maximum velocity that can be achieved and K_M is the substrate concentration where half of V_{max} is reached. Before the plateau, the ES complex concentration increases along the substrate one and the kinetics is in pre-steady state (yellow area). When V_0 reaches the plateau region, the concentration of the ES complex becomes approximately constant and the reaction is in a steady-state.*

The plot in **Fig 1** is reproduced by the Michaelis-Menten equation:

$$V_0 = \frac{V_{max}[S]}{K_M + [S]} \qquad (3)$$

where $\mathbf{K_M}$ is defined as the concentration at which the initial velocity is half of $\mathbf{V_{max}}$. The maximum speed is related to $\mathbf{k_{cat}}$ as follows:

$$V_{max} = k_{cat}[ES] \qquad (4)$$

Furthermore, $\mathbf{K_M}$ can be also expressed as:

$$K_M = \frac{k_{cat} + k_{-1}}{k_1} \quad (5)$$

Since $\mathbf{k_{cat}}$ is the rate-limiting step, then $\mathbf{k_{cat} << k_{-1}}$ and $\mathbf{K_M}$ can be approximated to $\mathbf{k_{-1}/k_1}$, which is the dissociation constant of the ES complex. The Michaelis constant can be then seen as an approximation of the enzyme affinity for the substrate.

1.2. Transition State Theory (TST)

TST was simultaneously developed by Eyring, Evans and Polanyi, in 1935, to link kinetic rate constants with thermodynamic activation parameters [9,10]. It was originally used to study reaction mechanisms in gas phase, while nowadays TST is also employed to study reactions in enzymes. According to TST, the reaction is described considering only two physical entities: the reagents (ground state) and the activated complex $[ES]^{\ddagger}$ (transition state). The TS lies on a saddle point at the highest point of the potential energy surface between the reactant and product state (**Fig 2**), and the barrier height is equal to the activation free energy $\Delta G^{\ddagger}$ (**Fig 2**). From a mathematical point of view, it is said that the TS misses one degree of freedom, which is the reaction coordinate (3N-7 instead of 3N-6 for reactant and product). The activated complex is not a chemical species with relevant stability (lifetime 10^{-13} s, which is the time for a single stretching vibration), but rather a distorted structure in between the structure of the reactants and that of the products. According to the Hammond postulate [11], the structure of the TS has a free energy closer to the reactant state for exothermic reactions, meaning that it resembles reactants more than products. Conversely, in the case of an

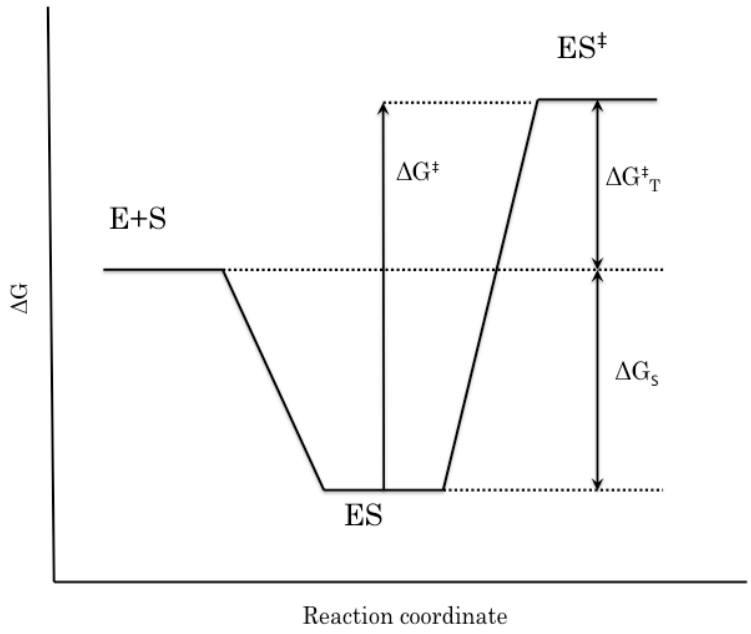


Fig 2: Free energy reaction diagram for an enzymatic reaction. *The different steps in substrate binding and catalysis are characterized. $\Delta G^{\ddagger}$ is related to the Michaelis-Menten kinetic turnover-number k_{cat}, while $\Delta G^{\ddagger}{}_T$ is related to the catalytic efficiency k_{cat}/K_M* [12]. *ΔG_S is the free energy of binding.*

endothermic reaction its free energy will be closer to product and it will resemble this state.

The fundamental postulate of TST is that the transition-state and ground-state populations are in thermodynamic equilibrium. Furthermore, barrier recrossing of the trajectory is not accepted before reaching thermal equilibrium with the product state. Assuming equilibrium between reactant and transition state, the equilibrium constant, K‡, will be:

$$K^{\ddagger} = \frac{[ES]^{\ddagger}}{[ES]} \qquad (6)$$

Reformulating Eq. 6 and considering that $\Delta G^{\ddagger}$ is equal to $-RT\ln(K^{\ddagger})$, we obtain:

$$[ES]^{\ddagger} = [ES]exp(-\Delta G^{\ddagger}/RT) \qquad (7)$$

where T is the temperature and R the gas constant. If we further combine Eq. 7 with $\mathbf{v = k_{cat}[ES] = k^{\ddagger}[ES]^{\ddagger}}$ (where v is the reaction rate and k‡ is the frequency of the bond vibration leading to product formation), then:

$$k_{cat} = k^{\ddagger} exp(-\Delta G^{\ddagger}/RT) \quad (8)$$

Finally, considering that the bond vibration frequency, v, of the TS is equal to $\mathbf{k_b T/h}$ and that $k^{\ddagger}=v\kappa$ we get:

$$k_{cat} = \kappa \frac{k_B T}{h} exp\left(\frac{-\Delta G^{\ddagger}}{RT}\right) \quad (9)$$

where κ is the transmission coefficient, which is a multiplicative factor equal to the fraction of TS that proceeds to product state, and it is generally close to 1. $\mathbf{k_B}$ is the Boltzmann constant and $\mathbf{h}$ is the Planck constant.
If the $\mathbf{\Delta G^{\ddagger}}$ is decomposed into its enthalpic and entropic contributions we get:

$$k_{cat} = \kappa \frac{k_B T}{h} exp\left(\frac{\Delta S^{\ddagger}}{R}\right) exp\left(\frac{-\Delta H^{\ddagger}}{RT}\right) \quad (10)$$

As explained by Fersht [12], the Michaelis constant $\mathbf{K_M}$ and the enzyme efficiency, given by $\mathbf{k_{cat}/K_M}$, can be related to thermodynamic activation parameters:

$$E + S \underset{\Delta G_{bind}}{\overset{K_M}{\rightleftarrows}} ES \xrightarrow[\Delta G^{\ddagger}]{k_{cat}} EP \quad (11)$$
$$E + S \underset{\Delta G^{\ddagger}_{enz}}{\overset{k_{cat}/K_M}{\rightleftarrows}} EP$$

The free energy of binding ($\mathbf{\Delta G_S}$ or $\mathbf{\Delta G_{bind}}$), the free energy of activation ($\mathbf{\Delta G^{\ddagger}}$) and $\mathbf{\Delta G_T^{\ddagger}}$ or $\mathbf{\Delta G_{enz}}$ (**Fig 2**), can be derived from the kinetic rate constants:

$$\Delta G_{bind} = RTln(K_M)\ ;\ \Delta G^{\ddagger}_{enz} = -RTln(k_{cat}/K_M) + RTln(kT/h) \quad (12)$$

As will be explained in the next section, TST terminology and concepts have been largely employed to explain enzyme catalysis.

1.3. The quest for the origin of enzymatic catalytic power

The statement made by Warshel in 2006 [13], is still a valid point of view to summarize the importance and difficulty in pinpointing the origin of enzyme efficiency, and the state of the art of the scientific effort:

"...In this respect it is crucial to understand what is the origin of the enormous catalytic power of enzymes, which remains one of the challenges of modern biophysics. Although many elements of this puzzle were elucidated by biochemical and structural studies, the source of the catalytic power of enzymes has not been widely understood and, clearly, has not been agreed upon by the scientific community..."

Attempts to explain the origin of the enzyme catalytic power goes back to 1894, when Emil Fischer proposed the famous *lock-and-key* model [14]. This theory was based on the observation of stereoselectivity for the enzyme invertase, which was able to hydrolyze α-glucosides, but not ß-glucosides. This led to the hypothesis of an enzyme action, in which intimate fit, like a *lock and a key*, between enzyme and substrate is a necessary requirement for catalysis.

Subsequently in 1946-48, Linus Pauling suggested that the origin of enzyme catalysis was due to a tighter binding of the enzyme in the TS, rather than the ground state, due to a major complementarity of the enzyme for the activated complex [15,16]. In 1958, Koshland, helped by newly solved X-ray crystal structures (myoglobin in 1958 [17] and haemoglobin in 1960 [18]), suggested a new model accounting for enzyme substrate specificity, which he called induced-fit model. The hypothesis he put forth was that the substrate, upon enzyme binding, would induce a conformational change to properly orient the catalytic residue for catalysis, whereas a non-substrate would not.

In early 1970s, Wolfenden [19] further formalized Pauling's theory, by comparing reactions catalyzed by enzymes and the corresponding uncatalyzed

reaction in water, showing that the former had a higher rate constant (k_{cat}) than the latter (k_{non}), and thus explaining the enzyme catalytic power. Here, the enzyme proficiency constant was also defined, which measures the capacity of an enzyme to lower the activation barrier compared to the uncatalyzed reaction in water. The catalytic proficiency is obtained by dividing the second-order constant enzymatic efficiency (k_{cat}/K_M), by the rate constant of the reaction in water (k_{non}) [20].

Jencks [21] theorized what he named the *Circe effect* (or ground-state destabilization), giving to substrate binding not only a role for specificity, but also for catalysis. According to this theory, the free energy of binding (ΔG_{bind}) can be utilized to provide extra driving force for catalysis, via 1) ES complex destabilization, relative to TS, and consequent higher free energy to lower $\Delta G^{\ddagger}$, or 2) entropy loss attained during binding, that is not paid for during TS formation (as opposed to the reaction in solution) [21]. This theory was one of the first attempts to explain how enzymes can lower $\Delta G^{\ddagger}$.

Warshel also pointed out the importance of comparing the enzymatic reaction and the one in solution, in order to understand the origin of enzymatic catalysis [22]. In particular, he stated that, to avoid any discussion between different mechanistic effects in the solution and the enzyme reaction, a *reference reaction* in water had to be compared to the enzymatic one [13,23]. In this reference reaction, all the catalytic residues present in the enzyme except the protein itself, are considered. In order to assess the acceleration of the reaction rate, the enzymatic free energy of activation, $\Delta G^{\ddagger}$, and the free energy of activation in water, $\Delta G^{\ddagger}_{cage}$, had to be compared. $\Delta G^{\ddagger}_{cage}$ is the free energy difference between the reactants, in the same water cage, and the TS (**Fig 3**).

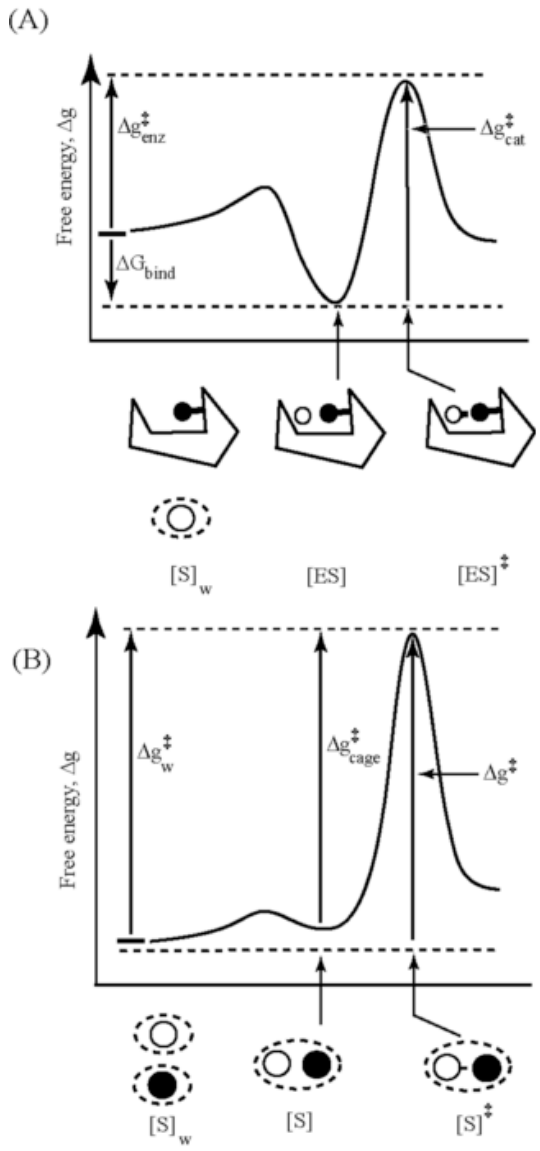


Fig 3: Schematic description of the free energy profile *for an enzymatic reaction (A) and for the corresponding reference reaction (B). Picture taken from ref* [13].

In 1976, Warshel and Levitt [22] conducted one of the first quantum mechanical computation on enzyme reactivity, and postulated an electrostatic TS stabilization for decreasing the reaction barrier in lysozyme. This study showed that the transition -state charged species (carbonium ion) was stabilized by electrostatic interactions, with Asp52, in lysozyme. Two years later, Warshel [24] explained how the lysozyme active site could give better electrostatic stabilization of the TS (ΔG_{Q_μ}) than bulk water. In fact, polar solvents spend about half of the energy gained from charge-dipole interaction in reorienting solvent molecules for proper dipole-dipole interactions with the

TS ($\Delta G_{\mu\mu}$). In this way, the transition-state free energy of solvation (ΔG_{sol}) is given by:

$$\Delta G_{sol} \cong \Delta G_{Q\mu} + \Delta G_{\mu\mu} \cong \Delta G_{Q\mu} - 1/2\,\Delta G_{Q\mu} = 1/2\,\Delta G_{Q\mu} \quad (13)$$

Warshel called the free energy $\Delta G_{\mu\mu}$ term "reorganization energy", and postulated that enzymes pay this energetic cost during the folding process, so that, upon TS formation, the active site dipoles are already properly oriented to interact with the substrate. He further formalized this theory for all enzymes [13,23].

In the late 80's, quantum tunneling was suggested for biologic reactions involving electron and hydrogen transfer [25,26]. This theory goes beyond TST, since light quantum particles, as hydrogen or an electron, can reach the product state without visiting the TS, but passing through the barrier. This phenomenon occurs when the probability of finding the particle in the reactant state, RS, is the same as in the product state, PS. Tunneling has a mass dependency and favors lighter isotopes.

In 90's, the existence of a so-called low-barrier hydrogen bonds (LBHBs) in proteins was postulated for enzyme catalysis [27–29]. This kind of HBs are formed when the pKa of the electronegative heteroatoms is similar, which allows them to equally share the hydrogen atom. LBHBs have shorter interaction distances than normal hydrogen-bond (less than 2.5 Å), they have stronger interaction, and are largely covalent. Furthermore, the active site facilitates the formation of low-barrier hydrogen bonds, since, after ligand binding, the involved heteroatoms are desolvated, and the dielectric constant is lowered (as opposed to bulk water) [30,31]. It was suggested that LBHBs can greatly accelerate enzymatic rate of proton transfer in acid-base reactions [30].

In 1996, Bruice *et al.* [32–35] introduced the term *Near Attack Conformation* (NAC) for reactions in solution, based on quantum- [36] and

molecular-mechanics [37] calculations. They later applied this concept for the enzymes dehydrogenase, dehalogenase and O-methyltransferase [38]. The NAC is a conformation in between RS and TS, closely resembling the last one, which an enzyme needs to pass before crossing the reaction barrier. Enzymes can accelerate the reaction rate by increasing the fraction of NACs.

The proposal of a dynamical contribution to enzyme catalysis was put forth in the late 1970's [39–42]. It has gained success [43–58] and generated a strong debate [53,59–61] between researchers for the past two decades. An enzymatic reaction is described through a complex and rugged free energy landscape, formed by different states and substates, barriers and transition states (**Fig 4**). The protein samples different conformations from substrate binding to product release through motions that range from femtoseconds (bond vibration), nanoseconds (loop movement) to milliseconds (active site opening-closing). Between these conformational changes, coordinated motions within the enzyme–substrate complex are thought to allow the reaction to proceed at a much faster rate compared to the reaction in solution. Emerging evidence indicates that enzymatic rates may be closely tied to the ability of enzymes to sample alternate structures, such that it allows the reactive environment to achieve structural and electrostatic complementarity to the TS along the reaction [44,45,48–51,54–56,62–65].

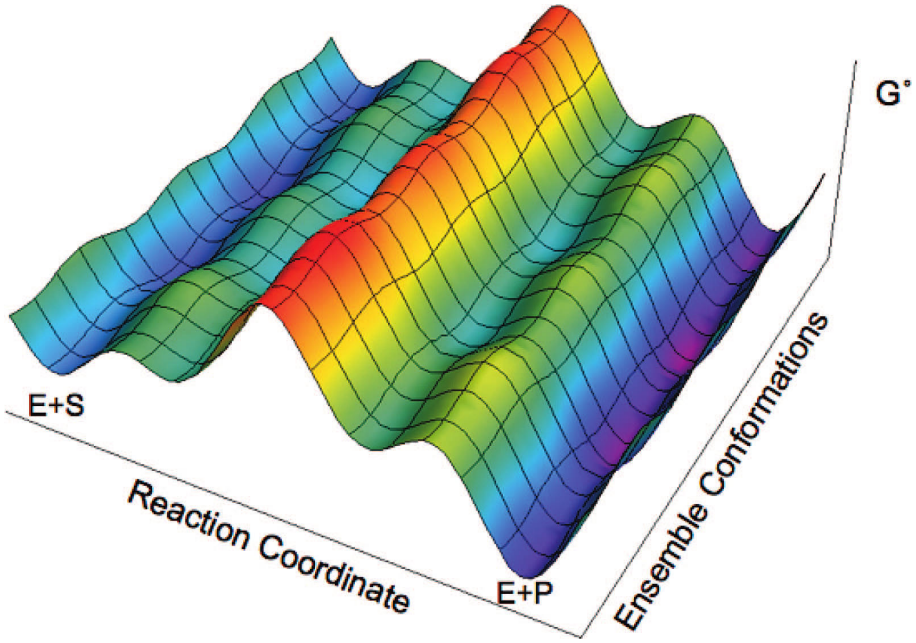


Fig 4: Schematic 3D representation of the standard free-energy landscape picturing the enzyme reaction. *This figure illustrates the complexity of the free-energy landscape characterizing the reaction from Michaelis complex (ES) to the enzyme product complex (EP). Picture taken from Benkovic et al.* [58].

In enzyme design a clear understanding of the theory underlying the enzymatic catalysis can guide in rationally changing the enzyme structure to achieve better activity. Furthermore it helps in explaining how certain mutation can affect the enzymatic reaction rate.

2

Cold-adapted enzymes

Life on earth has evolved to thrive in different temperature ranges, and some microorganisms (extremophiles) have adapted to its extremes: from -20, arctic sea ice [66], up to 121 ˚C, hydrothermal marine vents [67]. Microorganisms can be classified (**Fig 5**) according to the optimal temperature (T_{opt}) for growth:

- Psychrophile, T_{opt} <15 ˚C
- Psychrotolerant: T_{opt} 20-30 ˚C
- Mesophile: T_{opt} 37 ˚C
- Thermophile: T_{opt} 60 ˚C
- Hyperthermophile: T_{opt} > 80 ˚C

Extremophilic organisms have to cope with harsh conditions for survival, and have evolved specific mechanisms, from the cellular to the molecular level, to be able to grow and survive in thermal equilibrium with the environment. Low temperatures challenge microorganism with slow metabolic fluxes, due to the lower kinetic energy available, and a higher water viscosity. Microorganisms living at high temperatures, instead, have to cope with the instability of molecules and cellular compartments, in particular the cellular membrane.

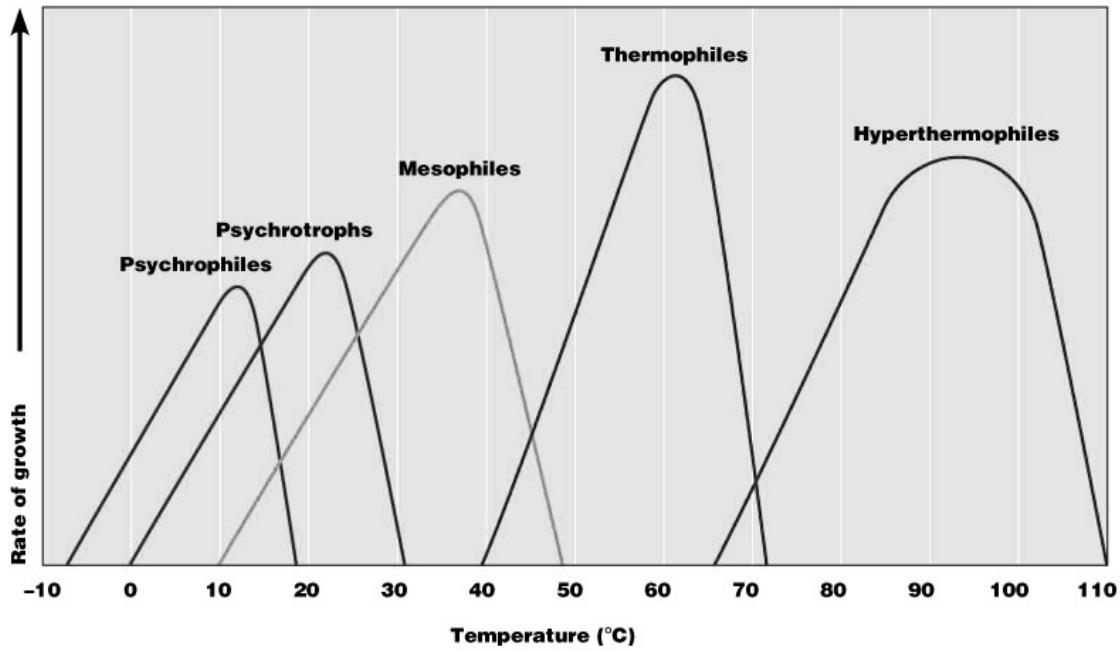


Fig 5: Microorganisms classification according to temperature of growth (°C).

A key issue with lowering the temperature is the exponential decrease in the enzymatic catalytic rates that follows from the Eyring equation in TST (Eq. 9). When the temperature is decreased from 37 °C to 0 °C, a 20- to 250-fold reduction in the enzymatic activity is typically observed [68]. Despite what is suggested by TST, CAEs display higher specific catalytic activity, when compared to their warm-active counterparts. CAEs achieve a higher catalytic efficiency (k_{cat}/K_M) by either increasing k_{cat}, at the expense of K_M, or by optimizing both kinetic parameters. Normally, though, the *turnover number* is increased, and the Michaelis constant is higher than warm-active enzymes [68–72]. Their maximal activity is shifted to lower temperatures than mesophilic enzymes, resulting in a lower thermostability, and unfolding at moderate temperatures. However, it is still not clear whether the thermolability is a consequence of higher activity or the lack of evolutionary pressure on maintaining stability. It seems reasonable to assume that CAEs lost thermostability following random genetic drift during evolution as opposed to mesophilic and thermophilic enzymes [73]. In order to achieve higher k_{cat}, psychrophilic enzymes decrease $\Delta H^{\ddagger}$ [74] making the reactions less temperature-dependent, rather than lowering $\Delta G^{\ddagger}$ instead (Eq. 10). The

enthalpic adjustment is however accompanied by a more negative activation entropy ($T\Delta S^{\ddagger}$), thereby counterbalancing the positive effect of $\Delta H^{\ddagger}$ on the free energy of activation. To date, all cold-active enzymes catalyze their reaction with a lower activation enthalpy and a more negative activation entropy compared to their mesophilic counterparts, thus representing the fingerprint of cold-active enzymes. [68–72]

Cold-adapted enzymes represent a very interesting subject for basic research to understand the relationship between stability, structure and activity. Furthermore, enzymes active at lower temperatures and with increased thermolability, constitute important targets for applications in different industrial areas (detergents, food and dairy products, wastewater treatment, bioremediation etc. [75,76]).

The main mechanism for adapting enzymes' structure and reaction to cold temperatures is proposed to originate from an increased flexibility of the cold-active enzymes, either localized in the proximity of the active site or in more distant parts of the structure [68,77–80]. No direct evidence has been, though, discovered to support this thoery, but it has rather been inferred from stability, kinetic and thermodynamic studies. The lower thermostability, higher K_M (lower substrate affinity) and more negative activation entropy would, in fact, suggest a major flexibility of psychrophilic enzymes with respect to their mesophilic counterparts [68,81]. Furthermore, a set of experimental techniques such as X-ray b-factors [82], hydrogen-deuterium exchange [83], fluorescence quenching [84,85], neutron scattering [86] and computational ones like molecular dynamics [87–91] have supported these hypotheses in certain enzyme families.

Most of the hypotheses put forth so far for the different families of CAE, have tried to find structural and sequence variations able to increase protein flexibility. Each enzyme family typically adopts its own strategy to achieve psychrophilic features, thus, a common theory for enzyme cold adaptation has not yet been formulated. Over the decades a series of possible structural

changes have been described (for reviews see [68,74,79–81,92,93]). The most obvious one is a decrease of all the stabilizing interactions: salt-bridges, hydrogen-bonds and aromatic interactions (π-π, cation-π or dipole-π). In particular, CAEs display a lower arginine content (low Arg/(Arg+Lys) ratio), as they can make more ionic and polar interactions with the guanidino group than lysines (triosephosphate isomerase [94], subtilisin [95], alanine dehydrogenase [96], cellulase [97], aminopeptidase [98], chitinase [99], acetate kinase [100], ß-mannanase [101], subtilisin S41 [95], α-amylase [102], malate dehydrogenase [103], phosphoglycerate kinase (PGK) [104], xylanase [105] and alkaline phosphatase [106]). Salt bridges have been shown to be one of the main factors affecting protein stability, and several comparative studies, based on X-ray structural data, have reported that the number of salt bridges is lower in cold-adapted enzymes. Quite interesting is α-amylase case, where the mutation of lysines into arginines transforms the cold- into warm-active enzyme, increasing thermostability and $\Delta H^{\ddagger}$, while decreasing both k_{cat} and K_M. [107]. Jónsdóttir *et al.* showed in a thermophilic subtilisin-like proteinase (aqualysin I) that only the key charged residues organized in extensive interaction network, are able to act on thermostability [107]. Disruption of six salt-bridges by single mutation of ten arginines to alanines in a mesophilic elastase, showed importance for stability and activity, but only for certain ion-pairs [108]. It is important to keep in mind that the strength of electrostatic interactions and the hydrophobic effect exhibit also temperature dependence. Hydrogen-bonds and ion-pairs become stronger with decreasing temperature, while the hydrophobic effect weakens (leading to cold-denaturation) [109]. Fewer aromatic interactions are displayed by cold-adapted subtilisin [95], β-lactamase [110] and α-amylase [111].

The hydrophobic core tends to be less compact and presents more cavities, thanks to the insertion of less hydrophobic and shorter residues [112–115] (lower fraction of large aliphatic residues ((Ile+Leu)/(Ile+Leu+Val) ratio). In this way the inner movement of side chains and the distance between amino acids increase, leading to scaffold destabilization and weaker hydrophobic

effect in folding. Bigger cavities in the core result in an increased accessibility to solvent water, and a major "wetting" inside the psychrophilic enzymes than the mesophiles [116]. The surface has more extended hydrophobic areas in the psychrophilic enzymes, thus decreasing the entropy of water molecules around it and surface solvation [99,101]. In certain cases, the surface hydrophilicity is increased, often with a larger number of acidic residues [95,117], improving interactions with the solvent and by destabilization of negative charges repulsion. Uracyl-DNA glycosylase [112], malate dehydrogenase [103], citrate synthase [113], cellulase [97] and trypsin [118,119] are characterized by marked differences in the electrostatic potential in proximity of the active site between cold- and warm-adapted enzymes, suggesting a facilitation in ligand binding.

A broader and more solvent accessible active site cleft favoring substrate-binding and product release, has been reported [74,81,92]. This is achieved by substitution of bulky residues for smaller ones or different conformation of binding site loops. Substitution of two alanines and insertion of a loop in the active site of a psychrophilic citrate synthase, mimicking the hyperthermophilic homolog, did not lead to changes in k_{cat}, but rather the affinity for one of the substrates (acetyl-CoA) [120]. Lower content of disulfide-bridges compared to the warm-adapted homologs has been reported for alkaline phosphatase [106] and lysozyme [121]. Quite surprisingly, cold-adapted α-amylase has two disulfide bridges more than the mesophilic one [102]. The mutation into cysteine in the cold-active GH5 cellulase to form an extra disulfide-bridge stabilizing a flexible domain, resulted in higher thermostability, leaving the kinetic parameters unaltered [122]. In some psychrophilic sequences there is a preference for a higher number of glycines [80,94], and lower number of prolines [102,103,106,113,123,124] inside loops than in the mesophilic homologs. Finally, longer surface loops have also been linked to increase flexibility in CAEs, thanks to a major amplitude of movement of the elongated part and connected secondary structures [97,113,125,126]. For the psychrophilic alpha-amylase and subtilisin S41, a

weaker binding of Ca^{2+} was detected [95,102]. Indeed, mutation of a glutamic acid to alanine, near the Ca^{2+} binding site of a mesophilic alkaline protease, decreases the ion affinity and increases k_{cat} and K_M [127].

Recent simulations of cold- and warm-adapted trypsin suggest that the protein rigidity well outside the active site controls the enthalpy – entropy balance, and thus the temperature adaptation [128,129]. Single point mutations in warm-adapted trypsin, disrupting hydrogen bonding networks and making the protein surface softer, was predicted to significantly alter the enthalpy – entropy balance, making the enzyme more cold-adapted-like [128]. In a very recent work, Isaksen *et al.* [129] showed computationally that the enthalpy – entropy balance was completely converted to those of cold-adapted trypsin, by gradually freezing parts of the surface residues in the warm-adapted trypsin.

3

Model systems

3.1. ββ-α metal dependent endonucleases

Endonuclease A (EndA) is a non-specific, metal-dependent bacterial nuclease which has a periplasmic or extracellular localization, i.e. fully exposed to external environment. It cleaves the phosphodiester bond within DNA or RNA without any sequence specificity, leaving a 3'-OH and a 5'-P product. A divalent cation, mostly Mg^{2+}, is employed for DNA binding, P-O3' polarization, and TS phosphoanion/product stabilization. Until now, the crystal structures of the homologs EndA from *Aliivibrio salmonicida* (VsEndA) [130], *Vibrio cholerae* (VcEndA) [131] and *vulnificus* [132] have been deposited in the PDB library.

EndA shares with other nucleases a ßß-α motif, formed by two antiparallel ß-strands, and a perpendicular α-helix (**Fig 6**). The ßß-α motif was first described by Kühlmann *et al.* [133], observing the conservation of this supersecondary structure also in the structure of the His-Cys box homing endonuclease from *Physarum polycephalum* (I-Ppoi) [134], the non-specific nuclease from *Serratia marcescens* (Smn) [135], the HNH DNAses Colicin-E7 and 9 [136,137], and the Endonuclease VII from the bacteriophage T4 (EndVII) [138]. Kühlmann *et al.* also called this motif ß-finger and lately also the name His-finger has been employed. The ß-finger was later discovered in the crystal structures of the HNH homing endonuclease I-HmuI from *Bacillus* phage SPO1 [139], the restriction endonuclease Hpy99I from *Helicobacter pylori* [140], Nuclease A from *Anabaena sp.* (NucA) [141], Endonuclease G from *Streptococcus pneumoniae* (EndoG) [142], Endonuclease A from *Streptococcus pneumoniae* (EndA) [143], and restriction endonuclease PacI

from *Pseudomonas alcaligenes* [144]. The ßß-α motif superfamily can be divided into three different families: a) His-Cys box nucleases (I-PpoI), b) HNH nucleases (colicinE7/9, I-HmuI and Hpy99I) and c) DRGH nucleases (EndA, EndoG, NucA, Smn and EndVII).

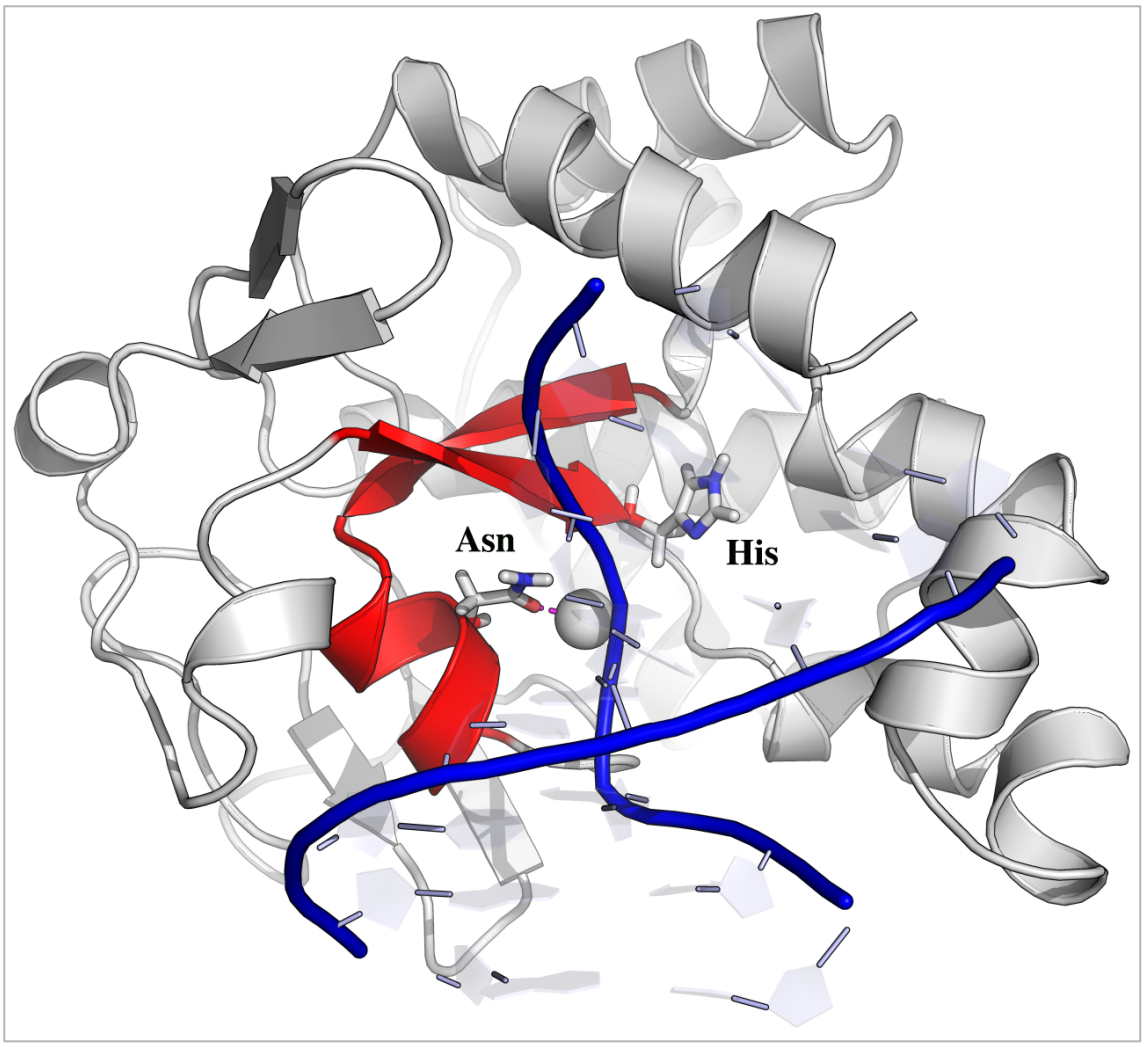


Fig 6: ßß-α motif (red) with the conserved asparagine (Asn), coordinating the Mg^{2+}, and the histidine (His). *VcEndA bound to DNA (ribbons in blue) is represented in the picture.*

The ßß-α motif is located at the center of the active site and it is involved in cation binding through a conserved asparagine, and in catalysis through a conserved histidine. The phosphodiester bond is cleaved through an acid-base mechanism, where a) a histidine abstracts a proton from a water in the bulk solvent, b) the hydroxide ion attacks the phosphate in a S_N2-like reaction, in

line with the P-O3' bond, and c) the leaving O3' oxygen is protonated. The TS structure is a trigonal bipyramidal phosphorane.

3.2. EndA from *Aliivibrio salmonicida* (VsEndA) and *Vibrio cholerae* (VcEndA)

VsEndA belongs to the salmon pathogen *Aliivibrio salmonicida,* which is a psychrophilic and halophilic gram-negative bacteria living in the ocean [145]. VcEndA is expressed by the notorious *Vibrio cholerae* (responsible for the cholera disease), which instead is a mesophilic and halotolerant (does not need NaCl for growth) gram-negative bacteria, preferring brackish/estuarine waters [145]. VsEndA displays the classical features of psychrophilic enzymes, with higher k_{cat} and K_M (5-37 °C), and a lower melting temperature (T_m) than VcEndA (see **Table 1**).

Table 1: Kinetic parameters at 15° and 25°C and temperature of melting (T_m) for VsEndA and VcEndA [146].

Parameters		VsEndA	VcEndA
k_{cat} (s^{-1})	15°	14.7	3.10
	25°	18.5	7.18
K_M (nM)	15°	202	131
	25°	169	156
T_m (°C)		44.8	52.8
$\Delta G^{\ddagger}$ ($kcal \cdot mol^{-1}$) 25°C		15.72	16.27
$\Delta H^{\ddagger}$ ($kcal \cdot mol^{-1}$) 25°C		7.93	17.63
$T\Delta S^{\ddagger}$ ($kcal \cdot mol^{-1}$) 25°C		–7.8	1.36

VsEndA has an optimal NaCl concentration for activity of 425 mM and an optimal pH of 8.5-9.0. VcEndA instead has an optimum of 175 mM [NaCl] and an optimal pH of 7.5-8.0. Regarding the reaction activation parameters, the psychrophile follows the CAEs trend, displaying lower $\Delta H^{\ddagger}$ and a more negative $T\Delta S^{\ddagger}$ than the mesophile [146] (**Table 1**), while maintaining similar $\Delta G^{\ddagger}$.

The structures of VsEndA and VcEndA are composed of 207 and 208 amino acids, respectively. In the psychrophile a lysine is inserted in position 52 (52A), while in the mesophile a proline and an asparagine are inserted in the C-terminal. VsEndA and VcEndA display 72% sequence identity (**Fig 7**), and the superimposed PDB structures (on the backbone) have 0.75 Å root mean square deviation (rmsd) [130,146].

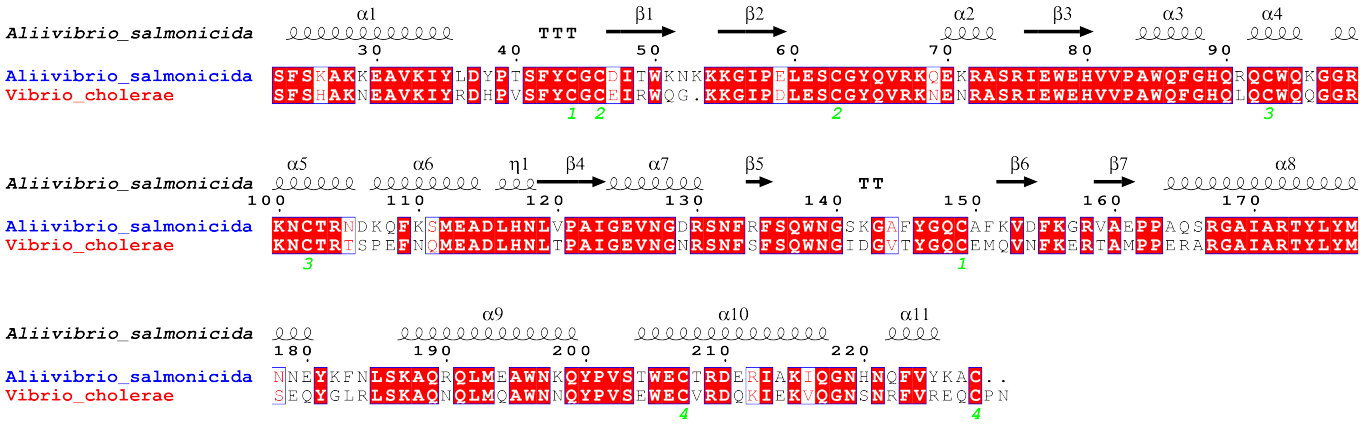


Fig 7: Sequence alignment of VsEndA (2PU3) and VcEndA (2G7F). *The green numbers locate the disulfide bridges. The secondary structures are based on VsEndA PDB X-ray (2PU3)* [130].

Together with the catalytic Mg^{2+}, a Cl^{-} ion is also deeply inserted in the protein core and most likely serves a structural role. In the comparison by Altermark *et al.* [130] of the crystal structures of VsEndA (2PU3) and VcEndA (2G7F), the most striking difference found was the number of lysines, twelve more in the cold-active. Consequently, the isoelectric point pI of VsEndA is higher than VcEndA, 9.6 and 8.6 respectively. The authors suggested both a possible role for cold- and salt-adaptation for the increased number of positively charged residue. In particular, the comparison of the electrostatic

surface potential calculated at optimal salt concentration, shows similar values. This would suggest that most of the lysines in VsEndA serve to make the binding of the negatively charged DNA feasible at higher NaCl concentration.

The number of disulfide bridges, 4, is the same in the two variants. Concerning the intramolecular interactions the cold-active enzyme has lower number of salt-bridges (–2), and a decreased amount of hydrogen bonds (–10). The number of glycines is decreased by one residue in VsEndA (contrary to the general findings in the literature), while the prolines are two less in VsEndA, although one is located at the second last residue in the C-terminal, and should not affect enzyme stability.

4

Computational Methods

In the next sections the different computational techniques I have employed in my project will be introduced. They are divided in two parts, where the first describes MD and related methods theory (sections 4.1-4.5). In the section 4.6 the EVB methodology will be introduced.

4.1. Molecular Mechanics

Molecular Mechanics (MM) allows to calculate *in silico* energies and conformations of molecules by modeling atoms and bonds as ball and spring, respectively. Such treatment of molecules with the exclusion of electrons from calculations is made possible by the Born-Oppenheimer approximation. This assumption decouples, in fact, electronic and nuclear motions, due to the large difference in mass and since the electrons rapidly follow the new position of the nuclei. MM calculates the ground-state molecular energies based only on the position of the nuclei. As such, MM is inherently not suitable for the calculation of properties dependent upon the electronic distribution.

The molecular energy is calculated in MM through empirical or semi-empirical energy functions depending on the atomic positions of the system $\mathbf{r}^N$, known as *force fields*:

$$U_{(r^N)} = \sum_{bonds} \frac{1}{2} k_b(b_i - b_0)^2 + \sum_{angles} \frac{1}{2} k_\theta(\theta_i - \theta_0)^2 + \sum_{impropers} \frac{1}{2} k_\varphi(\varphi_i - \varphi_0)^2 + \sum_{torsions} \frac{1}{2} V_n(1 + cos(n\omega - \gamma)) + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right) \quad (14)$$

The first four terms represent the bonded part of the equation, and calculate bond stretching, angle and out of plane bond bending and bond rotation or torsion. The harmonic potential is employed for the bond, angles and improper torsions, where $\mathbf{k_b}$, $\mathbf{k}_\theta$ and $\mathbf{k}_\phi$ are the relative force constants, and $\mathbf{b_0}$, θ_0 and ϕ_0 are the equilibrium values. The periodic torsional potential models the energy variation upon bond rotation. Here, $\mathbf{V_n}$ is the force constant, $\mathbf{n}$ is the multiplicity of the function (number of minima in the function), γ the phase shift (value of the torsion angle in the minima), and ω the torsion. The non-bonded part estimates the electrostatic and van der Waals (vdW) interactions between atoms separated by more than three bonds, through Coulombic and Lennard-Jones (12-6) potentials. In the L-J term, ε_{ij} represents the well depth of the interaction potential and σ_{ij} is the collision diameter, i.e. the interatomic distance where the interaction energy is zero. The $\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}$ term accounts for the repulsive part of the potential, while the $-\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}$ term for the attractive part. Bond, angle and improper terms are considered as 'hard' degrees of freedom, as the high values of their force constants makes it difficult to deviate from equilibrium values. Torsions and non-bonded terms account for most of the variation in the potential energy $\mathbf{U(r^N)}$.

The main force fields for biomolecules are CHARMM [147,148], Amber [149] and OPLS [150]. The potential functional form given in Eq. 14 is shared among these force fields, but each of them is characterized by additional terms accounting for special interactions (e.g. H-bonds). Force fields parameters are calibrated against experimental or computational data (QM analysis) of small molecules. One key attribute of force fields is their transferability, as their parameters are typically calibrated on a small set of molecules, which can then be applied to other larger molecules, including proteins, nucleic acids, lipids, carbohydrates etc. The final concept defining force fields is atom types. They are atom-related and they describe the atom number, hybridization and the local environment. Each atom type bears its own parameters.

4.2. Molecular Dynamics

In computational chemistry one of the key interests is finding minima on the potential energy surfaces. These minima are usually the most populated regions of the surface and here molecules are in their most stable conformation. When dealing with large biomolecules, the hypersurface can be rugged and can have many different local minima. For properties calculation, free energies for example, it is important to sample the phase space in order to locate the thermally accessible conformations of a biomolecule. Molecular dynamics (MD) is a widespread method to sample the phase space and it has been widely employed for studying proteins/enzyme, nucleic acids and membranes etc. [151–156].

MD generates a time dependent trajectory of configurations by numerically integrating Newton's equation of motion ($\mathbf{F=ma}$). This trajectory describes the positions and velocities of all the particles in the system at selected points in time [157]. By calculating the force acting on atom $\mathbf{i}$ ($\mathbf{F_i}$) at time $\mathbf{t}$, the acceleration, $\mathbf{a_i}$, of atom $\mathbf{i}$ can be derived. According to Newton's second law:

$$a_i(t) = \frac{F_i}{m_i} = -\frac{1}{m_i}\nabla_i U_{pot} \quad (15)$$

where $\mathbf{m_i}$ is the mass of the given atom. The force, $\mathbf{F_i}$, in Eq. 15 is calculated from the negative gradient of the potential energy function, $-\nabla_i \mathbf{U_{pot}}$, i.e. the function from Eq. 14. The atoms acceleration is then combined with velocity $\mathbf{v(t)}$ and position $\mathbf{r(t)}$, at time $\mathbf{t}$, to calculate the position and the velocity at time $\mathbf{t+\delta t}$. The integration proceeds in small timesteps $\mathbf{\delta t}$ during which the forces are assumed to be constant. There are different algorithms to integrate the equations and the most used ones are the Verlet [158] and the leapfrog [159] algorithms. The Verlet method considers the accelerations and positions at time $\mathbf{t}$ and the positions from the previous step $\mathbf{r(t-\delta t)}$, to calculate the new positions $\mathbf{r(t+\delta t)}$:

$$r_i(t+\delta t) = r_i(t) + \delta t v_i(t) + \frac{1}{2}\delta t^2 a_i(t) + \cdots \quad (16)$$

$$r_i(t-\delta t) = r_i(t) - \delta t v_i(t) + \frac{1}{2}\delta t^2 a_i(t) - \cdots \quad (17)$$

These two Taylor expansions that approximate the positions, velocities and accelerations, when added give rise to:

$$r_i(t+\delta t) = 2r_i(t) - r_i(t-\delta t) + \delta t^2 a_i(t) \quad (18)$$

The velocities can then be derived as:

$$v(t) = [r_i(t+\delta t) - r_i(t-\delta t)]/2\delta t \quad (19)$$

Since at time $\mathbf{t=0}$ the positions are known, but not the velocities, initial velocities are assigned randomly from the Maxwell-Boltzmann distribution, $\mathbf{P(v_i)}$, at a given temperature:

$$P(v_i) = \sqrt{\frac{m_i}{2\pi k_B T}} exp\left[-\frac{1}{2}\frac{m_i v_i^2}{k_B T}\right] \quad (20)$$

Thanks to statistical mechanics formulations, macroscopic thermodynamics properties can be computed from the simulation of a microscopic system. These properties are calculated as time averages, since their values fluctuate over time, and represent an average over all the microscopic sub-states from which the system is composed of:

$$A_{average} = \lim_{t\to\infty} \frac{1}{t} \int_{t=0}^{t} A(r^N(t), p^N(t))dt \quad (21)$$

where $\mathbf{A_{average}}$ is the average thermodynamic property over time. In theory, to acquire such a value, all the phase space (6N dimensional space formed by $\mathbf{3\ r^N}$ and $\mathbf{3\ p^N}$) should be sampled. According to the ergodic hypothesis, one can assume that the time average can be approximated by an ensemble average $\langle \boldsymbol{A} \rangle$, where an ensemble is a set of replicates of the system:

$$A_{average} = \langle A \rangle = \iint A(r^N, p^N)\, \rho(r^N, p^N) dr^N dp^N \quad (22)$$

where $\boldsymbol{\rho\ (r^N, p^N)}$ is the Boltzmann distribution:

$$\rho(r^N, p^N) = e^{\frac{(-E(r^N, p^N)/k_B T)}{Q}} \quad (23)$$

where, for a system of $\mathbf{n}$ atoms, $\mathbf{r^N}$ and $\mathbf{p^N}$ are the atomic positions and momenta, $\mathbf{E}$ is the state energy, $\mathbf{k_B}$ is the Boltzmann constant, and $\mathbf{Q}$ is the partition function.

$$Q = \iint \sum_{i=1}^{N} e^{-\beta E_i} \quad (24)$$

The ergodic hypothesis assumes that not all the states in the phase space contribute equally to $\mathbf{A_{average}}$, but rather the most populated states do. In the

computation of $\langle A \rangle$ is included the Boltzmann distribution to account only for the relevant states.

Eq. 22 can be further simplified in a clearer form:

$$\langle A \rangle = \frac{1}{M} \sum_{i=1}^{M} A(r^n, p^n) \qquad (25)$$

where $\mathbf{t_i}$ is the time at the different time step, and $\mathbf{M}$ is the total number of configurations in the trajectory.

4.3. Force field models to represent ions

It has been estimated that approximately 30–40% of proteins requires one or more metal ions to be able to carry out their biological function in cells [160,161]. Metal ions have a central role in protein stability/function, folding and ligand binding. Furthermore, they are essential factors for the stabilization of nucleic acids conformation. Divalent cations (mostly Mg^{2+}) are employed by most nucleases for hydrolysis of the phosphodiester bond [162].

An efficient treatment of metal ions for ligand coordination geometry and distances is essential in force fields. In MM, there are multiple ways to model metal ions in biomolecules and solution. The main ones are:

- The bonded model [163,164];
- The non-bonded soft-sphere model [165,166];
- The cationic dummy-atom model [167–170];

The bonded model treats the interaction and geometry of metal ions with coordinating residues through bond, angle, torsion, electrostatic and vdW terms. It has the disadvantage of not allowing geometry interconversion or ligand release, and it is characterized by a large number of parameters to optimize, which limits transferability. The non-bonded model is the simplest one since it treats the interactions simply through Coulombic and vdW terms and it does not impose any geometrical constraint. It is described by a point

charge and only the Lennard-Jones parameters need to be optimized. The main drawback is its inadequacy in simulating multinuclear metal centers [169] and in describing transition metals [171]. Finally, the cationic dummy-atom model defines the metal ion as a metal core surrounded by six cationic dummy-atoms, placed along the lines of coordinating dative bonds in an octahedral arrangement. Each dummy atom has a charge of **+q**, a minimal portion of the total mass and a negligible vdW potential. The metal core has a charge of **n-6q** (to ensure the total charge is **+n**), a larger part of the total mass and vdW potential. The octahedral geometry is fixed by bond, angle and torsion potentials between the dummy-atoms and the metal core. The coordination of the ligand is modeled only by the electrostatic interaction of the negatively charged atoms with the positively charged dummy-atoms.

The geometry of the dummy complex itself is kept rigid by the imposition of large force constants on the metal–dummy bonds. Overall rotation of the dummy-atom around the nucleus is allowed, and no internal forces are associated with the rotation, since there are no bonds between the dummy complex and the surrounding ligands.

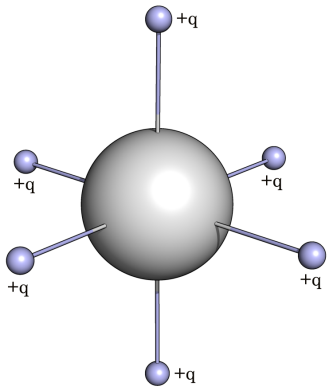


Fig 8: Cationic dummy-atom model. In grey the negatively charged metal core, while in light blue the positively charged dummy-atoms.

Delocalization of the ion charge away from the center is advantageous in the case of multi-nuclear metal centers, as it prevents excessive repulsion between the metal centers. The parameters are generally optimized through free energy perturbation and molecular dynamics (FEP/MD) to reproduce the

experimental or QM derived hydration free energy and radial distribution functions (RDF) in aqueous solution (M-O distances). In our case, the parameters had to be adapted from OPLS-AA to CHARMM. In fact, in these two force fields the definition of the L-J potential is different.

For OPLS-AA:

$$U_{L-J_{(OPLS)}} = \sum \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) \quad (26)$$

where the geometric parameters $\mathbf{A_{ij}}$ and $\mathbf{B_{ij}}$ are equal respectively to $\mathbf{4\varepsilon\sigma^{12}}$ and $\mathbf{4\varepsilon\sigma^{6}}$. $\mathbf{A_{ij}}$ and $\mathbf{B_{ij}}$ are combined by geometric rule $\mathbf{X_{ij} = (X_{ii}Xj_{jj})^{1/2}}$. In CHARMM instead the L-J potential is:

$$U_{L-J_{(CHARMM)}} = \sum \varepsilon_{ij} \left[\left(\frac{R\min_{ij}}{r_{ij}} \right)^{12} - \left(\frac{R\min_{ij}}{r_{ij}} \right)^{6} \right] \quad (27)$$

where $\boldsymbol{\varepsilon_{ij}}$ is given by the geometric rule, while $\mathbf{Rmin_{ij}}$, which is equal to $\boldsymbol{\sigma_{ij}}$, by the arithmetic rule ($\mathbf{Rmin_{ij} = [Rmin_i + Rmin_j]/2}$).

4.4. Methods for comparing ensembles similarity

4.4.1. Principal Component analysis (PCA)

Principal Component Analysis (PCA), or Essential Dynamics, is a multivariate statistical method applied to MD trajectories to reduce the number of dimensions describing the system dynamics [172]. PCA separates the configurational space sampled by MD in two subspaces: an essential one formed by few degrees of freedom and consisting of anharmonic motions, and one physically constrained subspace, considered less relevant for the protein dynamics. The essential subspace consists of a set of collective motions characterized by low frequency modes, i.e. large and slow conformational changes, as opposed to high frequency modes, i.e. local vibrations.

The trajectory is described by a set of vectors $\mathbf{x(t)}$, where $\mathbf{x}$ is a 3N-dimensional column vector representing the Cartesian coordinate of a system

formed by **N** atoms. In order to exclude overall translational and rotational motion, all the different configurations are fitted to a reference structure (which can be the first one or an average). The symmetric covariance matrix **C** (**3Nx3N**) of the positional deviations can then be calculated from:

$$C_{ij} = \frac{1}{s}\sum_{t=0}^{s+1} \langle\, (x_i(t) - \langle x_i \rangle)(x_j(t) - \langle x_j \rangle)^T \,\rangle \qquad (28)$$

where **S** is the total number of configurations in the trajectory, while $\mathbf{x_i}$ and $\mathbf{x_j}$ are the Cartesian coordinate for atom **i** and **j**. The brackets **< >** represent the average over time. By diagonalization of the covariance matrix of the Cα displacement, the essential dynamics of the simulation is obtained:

$$C = T\Lambda T^T \qquad (29)$$

where **T** is a matrix of column eigenvectors (η_i) and **Λ** is the diagonal matrix containing the corresponding eigenvalues. If **S** > **3N** there will be in total **3N** eigenvalues. Each eigenvalue is a representation of the mean squared displacements of atoms along the corresponding eigenvector. If one plots the eigenvalues against the respective eigenvectors (scree plot), one can see that few eigenvectors have high eigenvalues. This means that most of the system variation is contained in the first few eigenmodes.

The trajectory **x(t)** can be projected on the eigenvectors giving rise to the principal components:

$$p(t) = T^T x(t) \qquad (30)$$

where **p(t)** are the new rotated coordinates and $\mathbf{T^T}$ is the matrix containing the eigenvectors. The projection can be done on single eigenvectors visualizing the corresponding eigenmode.

PCA is often employed for comparing the sampling of different ensembles belonging to the same system or to a similar one: multiple replicas, apo versus holo enzyme, open versus closed conformation, wild-type versus mutant

proteins etc. In fact, the ensembles to compare must have the same number of atoms for the covariance matrix and overall 3D geometry similarity.

The root mean square inner product (*rmsip*) can be used to measure the degree of overlap between the essential subspaces of two ensembles [173]. The inner product between the first ten eigenvectors is normally computed:

$$rmsip = \left(\frac{1}{10} \sum_{i=1}^{10} \sum_{j=1}^{10} (\eta_i \eta_j)^2 \right)^{1/2} \quad (31)$$

where η_i and η_j are the eigenvectors of the two ensembles. This value can range from 0 (when there is no correlation in the sampled phase space) to 1 (for completely overlapping simulations).

4.4.2. Clustering-based Ensemble Similarity (CES) and the Dimensionality-Reduction-based Ensemble Similarity (DRES)

Two different methods can be employed to calculate the overlap between two ensembles generated by MD, namely the *Clustering-based Ensemble Similarity* (CES) and the *Dimensionality-Reduction-based Ensemble Similarity* (DRES) [174]. At the basis of these two methods there is the estimation of the probability density of conformations sampled by two trajectories. The advantage of comparing probability distributions, as opposed to covariance matrices of atomic positional deviations, is that they not only can suggest if the conformations between two ensembles are similar, but also whether they occur at the same frequency.

Their algorithms consist of two steps: first the probability densities are derived and subsequently they are compared to define the similarities between the ensembles. The difference between the two probability distributions is calculated using distance measures, derived from information theory. The CES and DRES methods differ in the first step only. CES employs an Affinity Propagation (AP) clustering [175] algorithm to separate the conformations in

different clusters according to similarity. The Cα RMSD is used as criterion for comparison of the different structures of MD trajectories. The number of clusters is determined by the value **k**, also known as preference, which constitutes the likelihood of a conformation to act as a cluster centre (or *exemplar*). The preference value **k** is set equal for all conformations, so that they are all likely to become centre of the cluster. When **k** is set as the total number of structures for an ensemble, each conformation will be part of its own cluster. On the contrary, a value equal to zero means that all conformations will belong to the same cluster. The algorithm is initialized by setting all the data points as exemplar and then it iteratively selects the optimal exemplar and assigns them the other data points based on similarity. The final clusters population for an ensemble is taken as the probability density over the discrete number of clusters. The main limitations of the CES methods are that differences between structures in the same cluster are not taken into account, as well as similarity between those in different clusters. The DRES methodology employs the Stochastic Proximity Embedding (SPE) [176] to embed the ensemble conformational space in a lower dimensional subspace, while maintaining the similarity between different conformations. Since it is a stochastic algorithm, it should be run multiple times to assess the validity of the results. SPE represents each ensemble conformation as a d-dimensional vector and seeks a collection of vectors where the pairwise Euclidean distance ($\mathbf{d_{ij}}$) between the vectors resembles the Cα RMSD ($\mathbf{r_{ij}}$) of the represented structures. The method starts by initializing the d-dimensional vectors and by iteratively refining them, randomly picking two vectors and adjusting their d-coordinates to get $\mathbf{d_{ij}} \approx \mathbf{r_{ij}}$. The coordinates are optimized by minimization of the "stress" function:

$$S = \frac{\sum (d_{ij} - r_{ij})^2 / r_{ij}}{\sum r_{ij}} \qquad (32)$$

If $\mathbf{r_{ij}} > \mathbf{r_c}$ (RMSD cutoff defining local neighbors: 1.5 Å) and $\mathbf{d_{ij}} \geq \mathbf{r_{ij}}$, meaning nonlocal neighbors far on the d-dimensional map, the coordinates are kept and new points on the map are taken. If $\mathbf{r_{ij}} \leq \mathbf{r_c}$ or if $\mathbf{r_{ij}} > \mathbf{r_c}$ and $\mathbf{d_{ij}} < \mathbf{r_{ij}}$ the

coordinates are updated. Finally, a Kernel Density Estimation (KDE) [176] is employed to get the densities underlying d-dimensional subspace.

Given two ensembles (or statistical populations) **A** and **B**, the distance between their corresponding density functions $\mathbf{p_A}$ and $\mathbf{p_B}$ (derived from CES or DRES) can be computed through the *Kullback-Leibler divergence* [177]:

$$D_{KL}(p_A, p_B) = \int p_A(x) log \frac{p_A(x)}{p_B(x)} dx \quad (33)$$

where $\mathbf{D_{KL}(p_A, p_B)}$ is always positive and it is equal to zero only in case $\mathbf{p_A = p_B}$. The main problem with this distance measure is that it is not symmetric and so $\mathbf{D_{KL}(p_A, p_B) \neq D_{KL}(p_B, p_A)}$. In order to have a symmetric measure the *Jensen-Shannon divergence* is preferable:

$$D_{JS}(p_A, p_B) = 0.5(D_{KL}(p_A, (p_A, p_B)/2) + D_{KL}(p_B, (p_A, p_B)/2) \quad (34)$$

Both CES and DRES employ the $\mathbf{D_{JS}}$ (Eq.(*34*). The values span from 0 to log(2), where a 0 value means that the two distributions are identical. The software Encore is capable to carry out such analyses [178].

4.5. Protein Structure Network (PSN) analysis

The 3D structure of proteins is governed by non-bonded interactions (together with bonded ones) between side-chain/backbone atoms such as hydrogen bonds, salt-bridges, hydrophobic and aromatic interactions. These interactions influence protein stability/dynamics and their evolution is involved in phenomena such as folding/unfolding, ligand binding, amino acid substitutions, protein-protein interactions, allostery etc. Graph Theory has been extensively applied to characterize the overall interaction networks of proteins [179] in different types of studies: protein structural comparison [180], conformational transitions , drug discovery [181], structural adaptation to temperature [182], folding theory [183], allosteric regulation [184,185], protein-protein binding [186]. The advantage of employing Graph Theory to analyze biomolecular structures is the simplification of the complex 3D

organization into a mathematical framework, retaining all the connectivity information. This representation in the form of an adjacency matrix, keeping trace of interacting residues, allows analysis using an arsenal of network based mathematical concepts.

This mathematical framework treats protein residues as *nodes* (elements or vertices) and their non-bonded interactions as *edges* (connections). Two vertices $\mathbf{v_i}$ and $\mathbf{v_j}$ are defined as adjacent if there exists an edge $\mathbf{e_{ij}}$ connecting them, while two edges are adjacent if they share a common node. The degree of a vertex $\mathbf{v_i}$, $\mathbf{deg_i}$, denotes the number of vertices adjacent to it. Simple graphs are employed for the study of protein topologies, since there is no need for multiple edges between two vertices, as opposed to multigraphs. Furthermore, the graph is often weighted, meaning that edges are discriminated giving different weights to each of them. The weight is related to the strength, which can be expressed with different interaction measures: frequency of occurrence or energy interaction. Finally, the distance between two nodes, $\mathbf{l_{ij}}$, is defined in terms of number of edges connecting the two vertices. PSN analysis can be carried out on X-ray structures or NMR ensembles, but equilibrium (or average) properties are obtained from the study of ensembles generated by MD or Monte Carlo simulations, which explore the phase space of thermally accessible protein states.

In Graph Theory proteins are described as "small-world" networks, characterized by short average path length between each amino acid, mediated by key residues with high connectivity. This concept was theorized by Watts and Strogatz [187] and named by analogy with the small-world phenomenon, also known as six degrees of separation. In graph formality, these key residues are defined as hubs, i.e nodes with a high degree $\mathbf{deg_i}$. These hubs have been found to be hotspots for stability [182], protein folding [188], allostery [189] and protein-protein interactions [186], since they constitute central connections mediating the short path lengths between other vertices.

Together with hubs other network parameters defining topological organization can be studied: *clusters*, *cliques*, *communities* and *communication Paths* [190]. A connected set of nodes in a network is referred as a cluster (**Fig 9**b). Cliques are instead defined as a set of **n** nodes, where each node is connected to each other (**Fig 9**c). The conjunction of cliques through common nodes creates a community. Finally, communication paths are employed to see which is the shortest (by number of edges) route connecting two residues. The study of this parameter is particularly of interest in the case of allostery, where conformational change in one domain can induce variations in other parts of the protein. Also, the study of the effect induced by mutations over shortest paths can give insightful findings related to the role of certain amino acid substitutions.

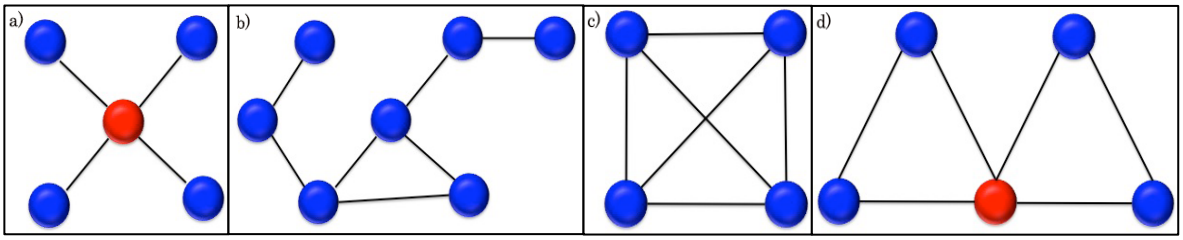


Fig 9: **Different network parameters** that can be studied to define the topological organization of a protein: a) hub, b) cluster, c) clique and d) community.

In the case of cold adaptation, comparison of protein networks from homologs with different temperature optima, i.e psychrophilic vs mesophilic-thermophilic, can give fundamental insight into structural adaptation to temperature changes. For example, clusters of hydrophobic residues in the protein core or on the surface, as well as clusters of salt-bridges, were found to affect protein stability.

There are different available software for studying PSN in biomolecules, as WORDOM [191], Vishgraph [190], Bio3d [192], GraProStr [193], NAPS [194], Cytoscape [195] etc.

4.6. Empirical Valence Bond (EVB)

EVB is a semi-empirical methodology that can be employed to calculate free energy profiles for reactions in solutions and enzymes [196,197]. It is calibrated to the energetics of a reference reaction (the equivalent uncatalyzed reaction in solution), where the data can come from experimental analysis or Quantum mechanical (QM) calculations. In particular, this technique is employed to study enzyme catalysis by computing the effect of the surroundings over the chemical reaction when bringing the reacting fragments from water to the enzyme. Normally the reaction energetics is computed by QM methods, such as Density Functional Theory (DFT) for a small model system. When dealing with whole proteins, the calculations become too computationally expensive for QM methods. EVB is then an affordable solution to account for the effect of the entire surroundings (Protein+Solvent) over the reacting fragments.

The EVB approach uses MD for sampling and employs the Free Energy Perturbation (FEP) technique to drive the reaction from reactant to product, passing through the transition state. As in Valence Bond (VB) theory, the reaction is described by a set of diabatic (or resonance) states, representing reactant, intermediates and product structures. By mixing this states the reaction potential energy surface (PES), and more importantly the Transition State, (TS) can be obtained. The energy of each state **i** is represented on the diagonal elements ($\mathbf{H_{ii}}$) of the EVB Hamiltonian matrix, and is given by the force field function in Eq. 35:

$$H_{ii} = \varepsilon_i = \alpha^i_{gas} + U^i_{bnd,rr} + U^i_{nb,rr} + U^i_{nb,rs} + U^i_{ss} \qquad (35)$$

where the constant term $\alpha^{i}_{\mathbf{gas}}$ is the gas-phase energy of diabatic state **i**, with all the fragments at infinite separation. The term $\mathbf{U_{bnd,rr}}$ represents the bonded interactions within the reacting atoms (bond, angles, torsions and impropers). In order to model bond formation and breaking, the Morse potential [198] is chosen for the reacting fragments, while for the rest of the

system the normal harmonic potential is employed. The non-bonded potentials $\mathbf{U_{nb,rr}}$, $\mathbf{U_{nb,rs}}$ and $\mathbf{U_{ss}}$ estimate the electrostatic and van der Waals interactions within the reacting atoms (**rr**), between the reacting fragments and the surrounding atoms (**rs**) and within the surroundings (**ss**). The off-diagonal term $\mathbf{H_{ij}}$, representing the adiabatic mixing, is descripted either by a constant or an exponential function of the distances between reacting atoms in Eq. 36:

$$H_{ij} = H_{ji} = A_{ij} e^{(-\mu_{ij} r_{ij})} \tag{36}$$

where $\mathbf{r_{ij}}$ is the distance between a pair of selected atoms (usually atoms involved in bond forming/breaking) and $\mathbf{A_{ij}}$ and μ_{ij} are parameters to fit on experimental or QM data. The off-diagonal $\mathbf{H_{ij}}$ and the gas α_i parameters are calibrated on the reference reaction free energies, and then are kept constant also in the enzyme simulations.

The adiabatic ground state energy $\mathbf{E_g}$ and the corresponding eigenvector $\mathbf{c_g}$ are obtained by solving the secular equation:

$$H_{EVB} c_g = E_g c_g \tag{37}$$

The free energy function, $\mathbf{\Delta g}$, is obtained by sampling the reactant state with MD simulation. In order to adiabatically drive the system to the product state through the TS, Free Energy Perturbation (FEP) needs to be coupled with MD. By using a mapping potential ε_m, which is composed by the EVB diagonal potentials, the system can be driven from reactant (ε_1) to product state (ε_2,) through transition-state, by the linear combination in Eq. 38:

$$\varepsilon_m = \varepsilon_1(1 - \lambda_m) + \varepsilon_2 \lambda_m \quad (0 \leq \lambda_m \leq 1) \tag{38}$$

where the factor λ_m is changed in $\mathbf{n}$ steps from 0 to 1. When $\lambda_m = 0$, the mapping potential match the reactant state and at 1 the product state.

The free energy $\Delta G(\lambda_m)$ associated with changing the factor λ_m from 0 to 1 is calculated through Eq. 39:

$$\Delta G(\lambda_m) = \sum_{m}^{n-1} \delta G(\lambda_m \rightarrow \lambda_{m+1}) \qquad (39)$$

The free energy on the ground state potential surface, $\Delta G(x)$, along the coordinate x, can then be calculated using FEP-umbrella sampling (FEP-US) in Eq. 40:

$$\Delta G(x) = \Delta G(\lambda_m) - RT \ln\langle \delta(x' - x) \cdot exp\{-[E_g(x') - \varepsilon_m(x')]/RT\}\rangle_m \qquad (40)$$

The reaction coordinate x corresponds to the energy gap ($\Delta\varepsilon = \varepsilon_1 - \varepsilon_2$) between the valence bond states.

4.6.1. Calculation of activation parameters

Often the study of enzymatic reaction cannot be limited to free energies (as for cold-active reactions), but the $\Delta G^{\ddagger}$ needs to be further decomposed in enthalpic and entropic contribution (Eq. 10). This can be done experimentally by calculating the reaction rate at different temperatures, and then by plotting $1/T$ versus $\Delta G^{\ddagger}/T$ (**Fig 10**).

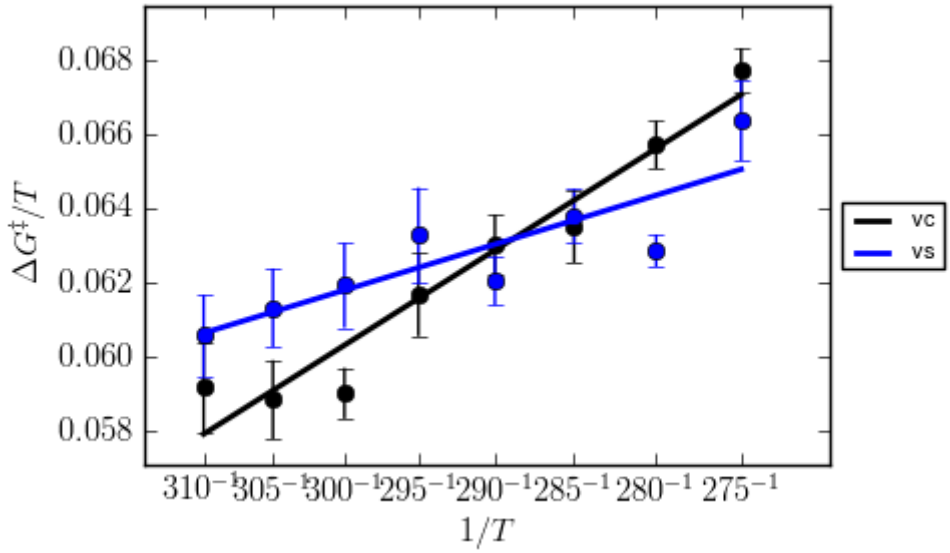


Fig 10: Example of Arrhenius plot. *The thermodynamics of VsEndA (blue) and VcEndA (black) are compared. Figure from paper III.*

The points in the plot are then connected together by linear regression and the $\Delta H^{\ddagger}$ value is derived from the line slope:

$$\Delta H^{\ddagger} = \frac{\sum_{i=1}^{n}(T_i^{-1} - \langle T^{-1}\rangle)(\Delta G_i^{\ddagger} \cdot T_i^{-1} - \langle \Delta G^{\ddagger} \cdot T^{-1}\rangle)}{\sum_{i=1}^{n}(T_i^{-1} - \langle T^{-1}\rangle)^2} \qquad (41)$$

while $\Delta S^{\ddagger}$ is calculated as follows:

$$\Delta S^{\ddagger} = \langle \Delta G^{\ddagger} \cdot T^{-1}\rangle - \Delta H^{\ddagger}\langle T^{-1}\rangle \qquad (42)$$

5

Summary of the papers

In the following section I will describe the main results from the papers in this thesis and I will outline the story line that connects them together.

The main goal of my PhD has been to seek the structural features that make VsEndA adapt to cold temperatures by means of MD and EVB simulations. As already stated in the Introduction, the most solid finding discovered in cold adaptation theory is that the lower activation enthalpy is responsible for counteracting the exponential decrease of k_{cat} with lowering the temperature. For this reason, we were particularly interested in those amino acid mutations able to change the balance $\Delta H^{\ddagger}$ / $\Delta S^{\ddagger}$.

To study enzyme cold adaptation we chose the psychrophilic and mesophilic endonuclease A, VsEnda and VcEndA, as model systems. Altermark *et al.* [146] and Niiranen *et al.* [199] characterized experimentally both VsEndA and VcEndA, collecting physiochemical properties and three crystal structures of the apo enzymes (2PU3 for VsEndA [130] and 2G7F/2G7E [146] for VcEndA).

Our starting point (paper I) was to carry out multiple sequence alignment (MSA) between VsEndA and closely related sequence homologs from both mesophilic and psychrophilic/psychrotolerant organisms. The goal was to isolate common amino acid substitutions within the psychrophilic/psychrotolerant sequences and at the same time to exclude substitutions caused by genetic drift or common with other mesophilic homologs. The standard procedure for sequence analysis is to align the studied protein with few homologs whose X-ray structure has been solved. By expanding the MSA with homologous sequences characterized by sparse identity scores, we have shown that the data are more statistically meaningful and clear. It should be noted that MSA can just be a preliminary and

approximate analysis, due to the inherent uncertainties associated with multiple subtypes for single microorganism and the unrefined classification of microorganisms according to T_{opt}. Nevertheless, the amino acid substitutions T120V, I141S and A166S (from VcEndA to VsEndA) displayed clear psychrophilic trends in this analysis (**Fig 11**), where the ones in positions 120 and 166 are located in the enzyme core, while 141 on loop3 is solvent-exposed. Secondly, we characterized the dynamical behavior of the two homologs by moving the systems with MD simulations. The RMSF profiles (**Fig 14**) showed that there were not significant rigidity/flexibility differences between the two enzymes (further discussions will follow in the next section). Finally, the study of the interaction network underlying the trajectories of the two systems was characterized with PSN analysis. The residues V120 (threonine in the mesophile) and A166 (serine in the psychrophile) form two hydrophobic hubs in the core of VsEndA and VcEndA, respectively. T120V in particular modulates the interaction pattern of Y43, which is also involved in binding the Mg^{2+} coordinating residue E79. The threonine in VcEndA creates an alternative hydrogen bond acceptor, possibly affecting the ion binding and/or catalysis. Two clusters of salt-bridges, centered on Arg222 and 225, are present in the mesophile and can possibly explain the increased rigidity of its C-terminal compared to that of VsEndA.

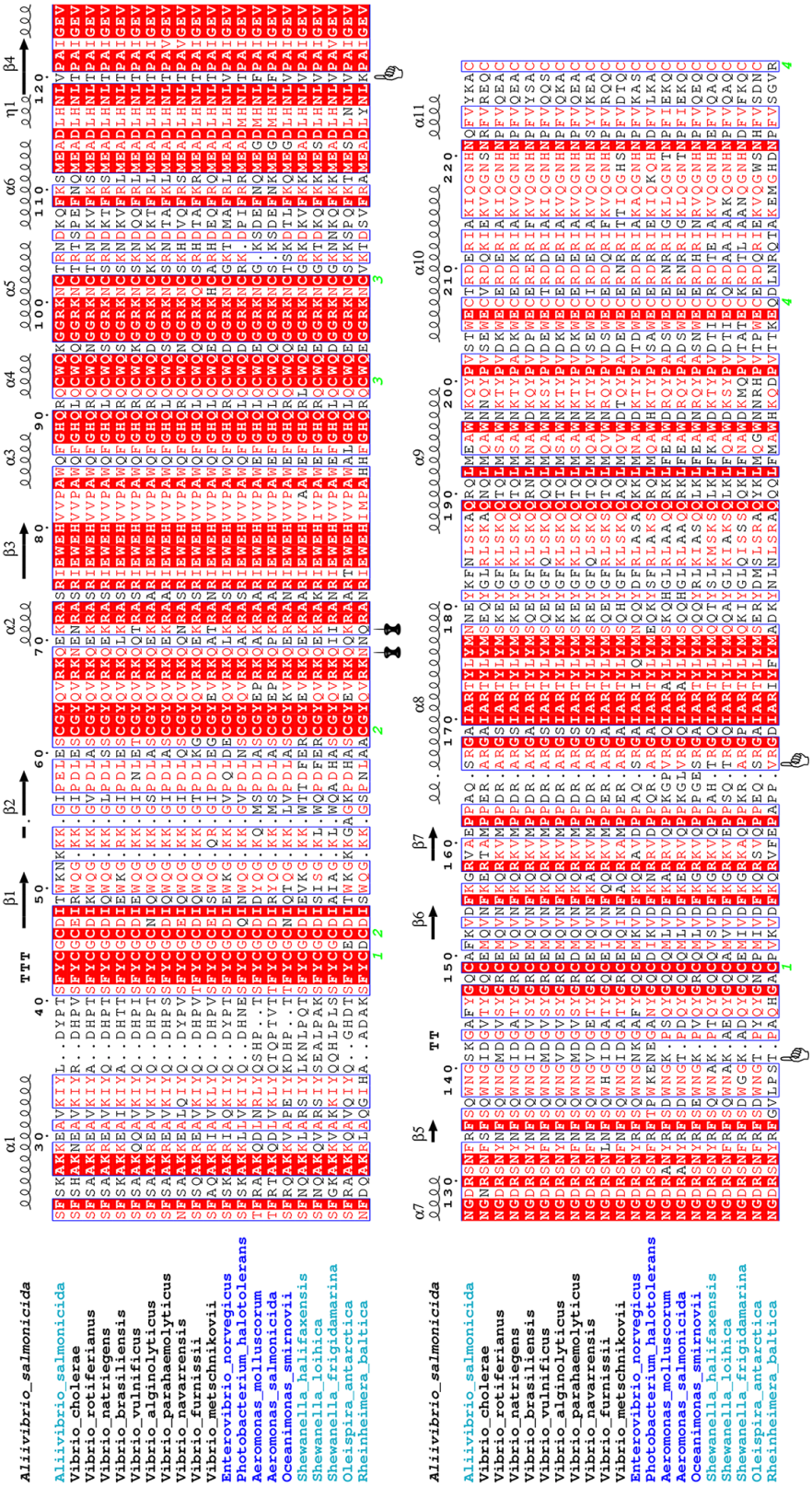


Fig 11: Multiple sequence alignment of psychrophilic, psychrotolerant and mesophilic homologs sequences of VsEndA. *The sequences of psychrophilic, psychrotolerant and mesophilic enzymes are highlighted in turquoise, blue and black, respectively. The hand icon marks the amino acid substitutions selected in this study, whereas the pin marks those analyzed by Niiranen et al.* [199]. *The green numbers highlight the disulfide bridges. The sequences are ranked by sequence identity. Figure from paper I.*

In the second paper, a successful EVB model was created for the reaction of VcEndA, which was able to reproduce the activation free energy of the simulated reactions. Since the protein structure was available only in the apo form, the reactant structures with enzyme-DNA complex had to be generated. The atomic coordinates for DNA were taken from the crystal structure of the homing endonuclease I-Ppoi (1CZ0 [200]), as suggested by the QM/MM work of Bueren-Calabuig *et al.* on Vvn [201]. This endonuclease shares only the ββ-α motif with EndA, but nevertheless the two active sites are perfectly superimposable with respect to the reacting fragments geometries and distances. To be able to reproduce a double coordination of the reactive phosphate to the Mg^{2+} (through O3' and OP1) the classical non-bonded ion model had to be dismissed in favor of the dummy-atom model [168]. QM optimizations of the reactant state and transition state revealed that the reaction mechanism proceeds through a concerted proton donation from nucleophilic water to the basic His80, attack of the OH^- on the phosphate and O3' release. Li *et al.* [132] suggested a different reaction mechanism for EndA, based on the crystal structure of Vvn. In their view the reactive phosphate was binding the Mg^{2+} only with the non-bridging oxygen OP1. We tested this hypothesis by EVB simulations and we showed that single coordination of the magnesium ion, through OP1, was not catalytically likely, since the computed activation free energies were between 40 and 60 kcal/mol.

In the last paper, we tested the EVB model to calculate the experimental activation parameters, $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$, for VsEndA and VcEndA. An Arrhenius plot was computed *in silico* by simulating the reaction at eight different temperatures, from 275K to 310K, and deriving $\Delta S^{\ddagger}$ and $\Delta H^{\ddagger}$ (as descripted in section 4.6.1 of the Methods). The models were indeed able to determine the activation parameters and, most importantly, showed for VsEndA a lower activation enthalpy and entropy than for VcEndA. Several mutations were then tested *in silico* for both systems: N71K (also experimentally mutated, with kinetic data available), T120V and S141I. The only mutation able to affect the activation parameters was S141I, partially "transforming" the cold-in warm-active enzyme and *vice versa* (**Fig 12**). Finally, the hypothesis of

Isaksen *et al.* [128,129] on "surface softness" and its effect on thermal adaptation was examined in VsEndA. By decreasing the radius of the simulated spheres and gradually constraining the outer parts of the enzyme, we were also able to show that the dynamics of the protein surface can tune the $\Delta H^{\ddagger}/\Delta S^{\ddagger}$ balance. Indeed, with a simulated sphere of radius 20 Å the psychrophilic activation parameters were almost completely converted to the mesophilic one (**Fig 12**).

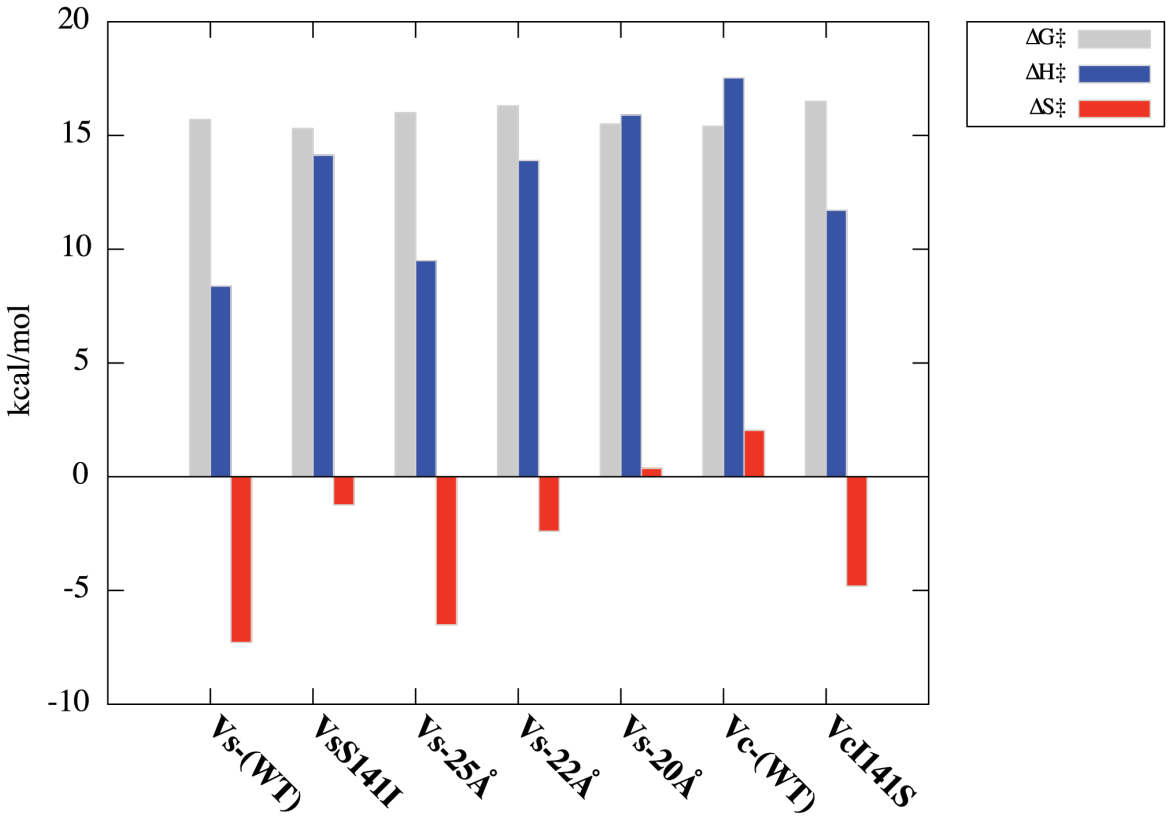


Fig 12: Activation parameters computed by in silico Arrhenius plots. *$\Delta G^{\ddagger}$, $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ for wild-type (WT), reduced radius (Vs-25Å and Vs-20Å) and for mutation S141I in VsEndA (Vs) simulations. In the case of VcEndA (Vc), the $\Delta G^{\ddagger}$, $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ are represented for wild-type (WT) and mutation I141S simulations. Figure from paper III.*

6

Discussion

In this section the main contributions from my PhD work will be examined in depth and will be correlated to other studies on cold adaptation from the literature.

Structural mechanisms determining enzymes cold adaptation. It is important to emphasize again that there is no common structural mechanism between the different enzymes classes to face lower temperatures. There are, however, trends that are shared between different protein families. This lack of homogeneity, in terms of adaptation, can only partially be explained by the evolutionary divergence of the different class of enzymes, but it is also affected by other factors. For instance, the enzymatic function dictates certain structural adjustments and discards others. In the case of EndA, a positively charged surface is required for binding the negatively charged DNA, and the psychrophilic variant has a more positive electrostatic surface [130]. This feature is achieved through extra lysines, and a positive imbalance of this residue has not yet been reported in the literature for other cold-adapted enzymes. The psychrophilic alkaline protease subtilisin S41, on the contrary, expresses eleven aspartic acids more than the mesophilic homolog, achieving in this way a lower pI and a more negative surface [95]. The same is observed for the cold-active cellulase, which displays a higher negatively charged surface due to a lower content of arginine/lysine and a higher number of aspartic acids (especially in the linker loop between the catalytic domain and the cellulose binding module) [97].

In the case of enzymes coordinating ions for structural stability or catalysis, the psychrophilic homologs are distinguished by a lower binding affinity than

mesophiles, as in the case of α-amylase [102] and subtilisin S41 [95]. Endonuclease A features, indeed, both a structural Cl^- and a catalytic Mg^{2+}, but unfortunately no ion binding assessment has been carried out to compare the affinity between VsEndA and VcEndA [146,199].

Another factor affecting the selection of the structural mechanism for low temperatures is the number of domains or monomers building a protein. In fact, in multidomain or multimeric enzymes the different properties are often located at the interface between monomers or domains in cold- and warm-adapted variants (citrate synthase [113], isocitrate dehydrogenase [202] and cellulase [97]).

Finally, enzymes with multiple environmental adaptations prove to be the most enigmatic and difficult cases to study, since two or more adaptational pressures act synchronously and distinguishing the respective factors becomes challenging. VsEndA and VcEndA belong to this class of proteins since they are at the same time psychrophilic/mesophilic and halophilic/halotolerant. Thus, the amino acid substitutions between the two homologs are caused either by temperature and/or salt adaptation. Altermark *et al.* [130], in particular, discussed the increased number of lysines in VsEndA compared to VcEndA, arguing that an increased positively charged electrostatic surface enables VsEndA to bind DNA at higher salt-condition. However, a possible role in cold adaptation was not excluded. When studying enzymes adapted to different environments careful attention must be paid in the selection of the optimal experimental and computational analyses to distinguish the two phenomena.

Lower content of salt-bridges in CAEs. One structural mechanism is frequently encountered in many cold-adapted enzymes (but not common to all) is a lower content of arginines and salt-bridges than their mesophilic counterparts. An increased number of ion-pair networks has been linked to the higher thermostability of warm-adapted enzymes, rather than to kinetic constants [74]. VsEndA has indeed four arginines less than VcEndA and it

lacks two salt-bridges compared to the warm-active variant [130]. In particular, two ion-pair networks located in the C-terminal of VcEndA are suggested to affect the thermostability of the mesophile. RMSF analysis pointed to a lower flexibility of this region in VcEndA (**Fig 14**). Furthermore, MD simulations of VsEndA mutated in the C-terminal (N179E, Q222R and K226E) to carry the salt-bridges anchoring two domains together in VcEndA, showed a decreased RMSF in this region with respect to the wild-type enzyme.

It is important to highlight the importance of carrying out MD simulations for studying salt-bridge networks, since as pointed out by Olufsen *et al.* [203] structural analysis on X-ray structures can be misleading. The crystallization conditions can, in fact, affect presence and geometry of ionic interactions.

Increased flexibility in CAEs. The main theory accounting for cold adaptation has been suggested to be an increased global or local flexibility of psychrophilic compared to mesophilic enzymes. Feller *et al.*, in particular, attributed this property to the active site [81,93]. Despite this simple and clear hypothesis, no direct experimental proof has been published so far. H/D exchange [83,204] and fluorescence quenching [98,123,205] are the preferred techniques to compare psychrophilic and mesophilic enzymes flexibility, but they cannot conclude if there is a difference in flexibility itself. That is, both analyses test the accessibility of the enzyme core to a fluorophore (fluorescence quenching) or to deuterium oxide (D_2O, H/D exchange). The degree of permeability of these chemical probes is not a direct evidence of protein dynamics. A more definitive proof may be obtained by NMR relaxation data, but unfortunately, due to the protein length limitation for this technique, only studies on cold-shock proteins [206] are available nowadays.

Siglioccolo *et al.* [82] compared the normalized B-factors in a set of X-ray structures from psychrophilic, mesophilic and thermophilic enzymes, concluding that the first shows increased static flexibility. We studied the B-factors from the PDB structures of VsEndA (2PU3) [130] and VcEndA (2G7F) [146] (normalizing the values according to Siglioccolo *et al.* [82] formula) and

concluded that no meaningful and correct comparison could be conducted (**Fig 13**). In fact, part of the B-factor profile of the warm-adapted EndA was increased compared to that of the cold-adapted one, and, most importantly, the shape of the two plots were not matching. That is, the flexibility profiles of homolog proteins should have an overall similar outline, except few regions with different peaks height. X-ray structures analysis can be inexact, since comparison of PDB entries with different crystallization procedure and resolution can produce experimental artifacts not related to flexibility.

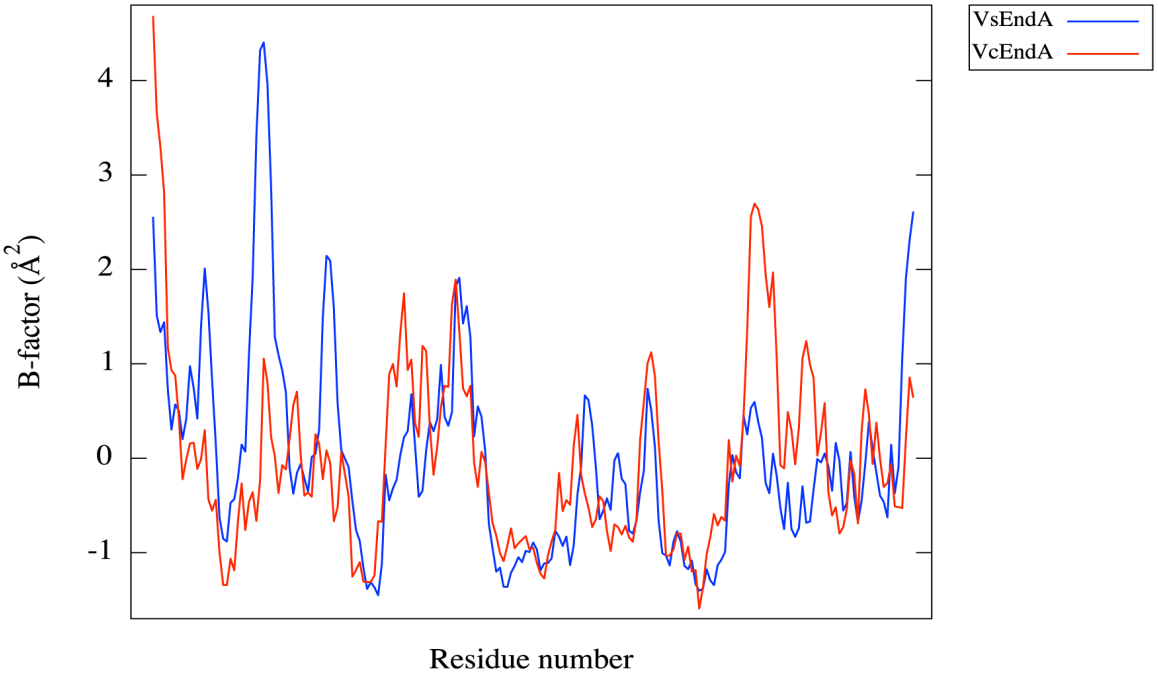


Fig 13: Normalized b-factors profiles for the PDB crystal structures of VsEndA [130] ***and VcEndA*** [131]. *The b-factors have been normalized according to the formula by Siglioccolo et al.* [82]*: B'=(B-<B>)/σB. B' is the normalized value, <B> is the average one and σB is the standard deviation.*

Contrary to experimental techniques, MD simulations give the opportunity to directly evaluate enzyme dynamics and to compare the flexibility between different systems. In our case, MD simulations and RMSF analysis of the homologs VsEndA and VcEndA did not display any increased flexibility of the active site. Moreover, globally the dynamics of the two variants was very

similar and it is very difficult to state that VsEndA is more flexible than VcEndA. Nevertheless, both have regions with modest differences in RMSF values (**Fig 14**). VcEndA shows higher peaks close to the active site on α1 (+ 0.3 Å), α3-6 (+ 0.4 Å) and loop3 (+ 0.2 Å). VsEndA is, on the contrary, more flexible (+ 0.2 Å) in a large part of the profile at the C-terminal, although distant to the active site.

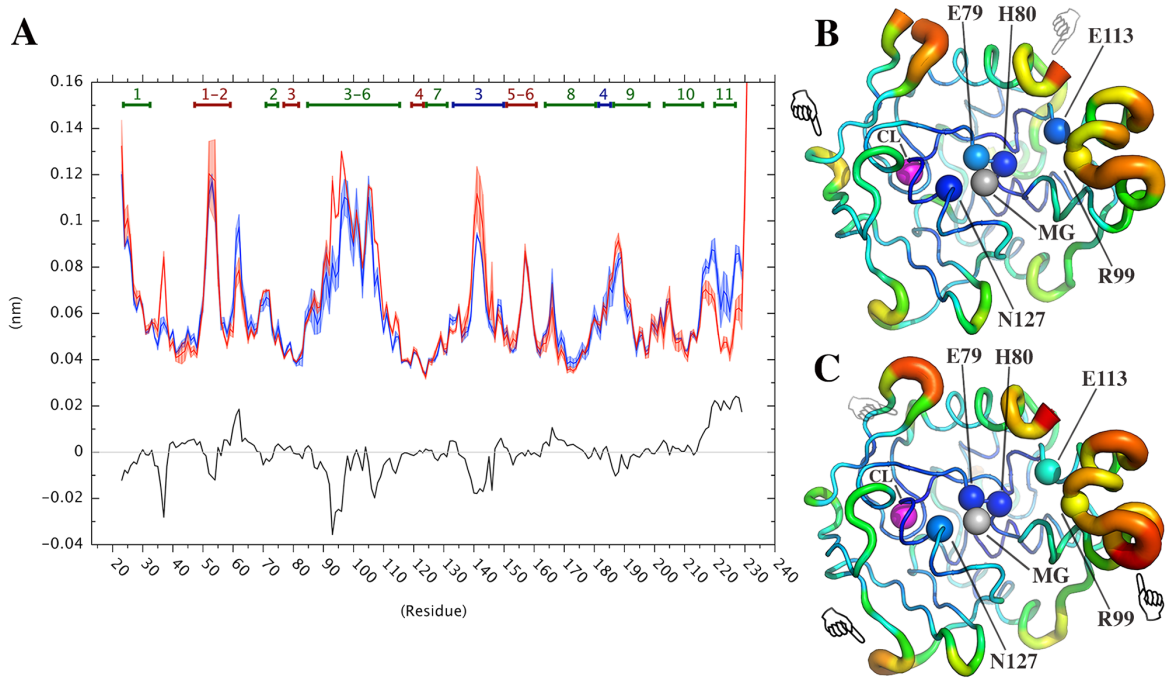


Fig 14: RMSF profiles for the MD simulations of VsEndA (blue) and VcEndA (red). *A) The RMSF profiles are averaged over three simulations per enzyme. The shaded area around the RMSF profile represents the standard deviation. The black line is difference in values between VsEndA and VcEndA (RMSF[Vs]-RMSF[Vc]). B,C) The RMSF is plotted on the 3D structure of VsEndA (B) and VcEndA (C) with the putty representation in Pymol.*

These results are in contradiction to the general theory, which describe CAEs as more flexible than the mesophilic homologs in order to adapt the activity to low temperature. Feller *et al.* [68,81,93], in particular, located the increased flexibility to the active site of psychrophilic enzymes.

In order to discuss RMSF profiles, it is important to define a threshold where a region can be said effectively to be more flexible than the other. Indeed, in our analysis most of the differences are below 0.1 Å and the largest one (in VcEndA) is 0.4 Å. Papaleo *et al.* [90], instead, found in trypsin the

largest fluctuations to be 0.1 Å, while Isaksen *et al* [128,129] defined the highest RMSF to be around 0.5 Å in trypsin. Furthermore, it is important to keep in mind that the comparison of RMSF profiles is affected by MD simulation schemes, trajectory length and RMSF calculation procedure. Most of the studies, even recent ones, of cold adaptation carried out MD simulations from few ns (particularly early one) to few dozens ns [207]. In such short trajectory lengths it is difficult to observe major conformational changes. In our case, simulations needed at least 10-20 ns to equilibrate, resulting in broad conformational changes reflected in a RMSD jump from 1 to 1.5 Å. Moreover, each system replicate (4 in total) lasted 500 ns. Simulation length above 100 ns is therefore preferable to characterize full protein flexibility scale.

Surface softness tunes $\Delta H^{\ddagger}/\Delta S^{\ddagger}$. Isaksen *et al.* [128,129] proposed a different theory than flexibility for the catalytic adaptation to low temperatures. Here, the surface "softness" is responsible for tuning the balance between $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ in the reaction of trypsin. The term softness is not related to the concept of structural flexibility, but rather to the change of potential energy ($\Delta U^{\ddagger}$) along the reaction coordinate (from reactant to transition state). The activation enthalpy can in fact be decomposed into different terms:

$$\Delta H^{\ddagger} = \Delta U_{rr}^{\ddagger} + \Delta U_{rs}^{\ddagger} + \Delta U_{ss}^{\ddagger} \qquad (43)$$

where the subscripts r represents the reacting fragments and s the surroundings (including both protein and solvent atoms). Isaksen *et al.* [128] suggested that the potential energy for the interactions of the active site surroundings ($\Delta U^{\ddagger}_{ss}$), is weaker for the cold-active enzyme than the warm-active. This potential energy variation can be seen as a force constant placed over the reactive fragments surroundings, which it "softer" in the cold-adapted enzyme than the warm-adapted, resulting in a lower $\Delta H^{\ddagger}$. This hypothesis was computationally proved by freezing atoms outside the catalytic site in the EVB

calculations. By imposing positional force constants on the cold-adapted trypsin surface, the activation enthalpy and entropy were increased, resembling the warm-adapted trypsin. Also in the case of the EndA homologs we discovered that the potential energy of the surroundings, $\Delta U^{\ddagger}ss$, was decreased in the psychrophile compared to the mesophile. Freezing the outer parts of the enzyme indeed made the calculated $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ closer to the values in VcEndA, corroborating the hypotheses that the rigidity of the surface can influence the reaction free energies. Furthermore, the substitution I141S was the only one tested able to transform the activation parameters of VsEndA towards those of VcEndA and *vice versa*. The residue in position 141 is located on the enzyme surface, on the contrary of position 71 and 120, which were the other substitutions assessed by EVB simulations that are located in the active site and the enzyme core, respectively.

The origin of the catalytic power of EndA. The last point of this discussion is not directly related to cold adaptation, but is a comment on the main findings on the EndA reaction mechanism from paper II. In the last decade there has been a heated debate between researchers supporting dynamical effect in catalysis and those supporting the reorganization theory hypothesis [52,53,60]. In the case of endonucleases VsEndA and VcEndA (also Vvn) we provide strong evidence that the origin of the enzyme catalytic power, as opposed to the reference reaction, lies in the electrostatic stabilization of the P-O3' bond by the Mg^{2+} ion. EVB calculated activation free energies, with sub-optimal coordination of the magnesium ion with PO_4^- without the leaving oxygen in 3', are almost matching the barrier of the reaction in water. As pointed out by Warshel *et al.* [13], the catalytic effect by ions accounts for the reorganization energy theory that he postulated earlier. In fact, the cost of positioning and orienting the cationic dipole is already paid upon folding of the endonuclease.

7

Concluding Remarks

The study of cold-adapted enzymes can give valuable answers to three key points in enzyme design and protein engineering, which is the relationship between structure and 1) activity, 2) thermostability and 3) the connection between activity-flexibility-thermostability. The knowledge of how to change the 3D conformation of enzymes in order to improve their reaction rate or stability is of immense importance for biotechnological applications. In this thesis we have however concentrated on the structural feature that enable VsEndA to adapt its activity to low temperatures. In particular the focus was on those substitutions able to alter the $\Delta H^{\ddagger}/\Delta S^{\ddagger}$ balance.

Based on MD simulations and multiple sequence alignment, we were able to pinpoint amino acid substitutions possibly linked with thermal adaptation. PSN and RMSF analyses furthermore allowed us to suggest structural effects related to this residues change. EVB calculations showed that the enzyme surface rigidity is responsible for decreasing the activation enthalpy in VsEndA compared to VcEndA, as in the case of trypsin [128,129]. Our hypothesis was further corroborated by substitution of the solvent exposed residue in position 141. *In silico* mutation of these residue in VsEndA with the residue present in VcEndA and in the opposite direction, resulted in the largest variation of the activation parameters. Currently, the surface rigidity theory has been evaluated only in the cases of endonuclease A and trypsin. In order to corroborate this assumption, additional test cases need to be found and assessed.

EVB proved to be a unique tool in calculating activation parameters in enzyme reactions, a feature nowadays lacking in other *in silico* methods. In

fact QM and QM/MM techniques despite being more accurate, lack the sampling power of the PES needed for calculations of such macroscopic values. There are several cases in the literature where this methodology has been applied to successfully compute $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ for biocatalysis [128,129,208,209]. EVB proves to be very useful for studying cold-adapted reaction, since they are characterized by a different activation enthalpy and entropy balance and simulations of mutant enzyme can highlight the effect on the activation parameters.

In order to finally validate our *in silico* findings mutational studies should be carried out in VsEndA/VcEndA for the amino acid substitution S141I. Finally further analysis should be carried out to highlight and test other possible sequence substitutions in EndAs able to affect thermostability and/or catalysis.

References

1. Cech TR. The Generality of Self-Splicing RNA: Relationship to Nuclear mRNA Splicing. Cell. 1986;44:207–10.

2. Kruger K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE, Cech TR. Self-Splicing RNA: Autoexcision and Autocyclization of the Ribosomal RNA Intervening Sequence of Tetrahymena. Cell. 1982;31:147–57.

3. Breaker RR, Joyce GF. A DNA enzyme that cleaves RNA. Chem Biol. 1994;1:223–9.

4. Miller BG, Wolfenden R. Catalytic Proficiency: The Unusual Case of OMP Decarboxylase. Annu Rev Biochem. 2002 Jun;71(1):847–85.

5. Atkins WM. Biological messiness vs. biological genius: Mechanistic aspects and roles of protein promiscuity. J Steroid Biochem Mol Biol. 2016;8(5):583–92.

6. Guengerich FP. Mechanisms of cytochrome P450 substrate oxidation: MiniReview. J Biochem Mol Toxicol. 2007;21(4):163–8.

7. Demirjian DC, Morís-Varas F, Cassidy CS. Enzymes from extremophiles. Current Opinion in Chemical Biology. 2001. p. 144–51.

8. De Bont JA m. Solvent-tolerant bacteria in biocatalysis. Trends in Biotechnology. 1998.

9. Eyring H. The Activated Complex in Chemical Reactions. J Chem Phys. 1935;3(2):107–15.

10. Evans MG, Polanyi M. Some applications of the transition state method to the calculation of reaction velocities, especially in solution. Trans Faraday Soc. 1935;31:875.

11. Fersht AR. Structure and mechanism in protein science. A guide to enzyme catalysis and protein folding. New York: W.H Freeman & Company; 1999. 631 p.

12. Fersht AR. Catalysis, binding and enzyme-substrate complementarity. Proc R Soc Lond B. 1974;187:397–407.

13. Warshel A, Sharma PK, Kato M, Xiang Y, Liu H, Olsson MHM. Electrostatic basis for enzyme catalysis. Chemical Reviews. 2006. p. 3210–35.

14. Fischer E. Einfluss der Configuration auf die Wirkung der Enzyme. Berichte der Dtsch Chem Gesellschaft. WILEY-VCH Verlag; 1894 Oct;27(3):2985–93.

15. Pauling L. Molecular Architecture and Biological Reactions. Chem Eng News. American Chemical Society; 1946 May 25;24(10):1375–7.

16. Pauling L. Nature of forces between large molecules of biological interest. Nature. 1948;161(4097):707–9.

17. Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. Nature. 1958;181(4610):662–6.

18. Perutz MF, Rossmann MG, Cullis AF, Muirhead H, Will G, North ACT. Structure of Hæmoglobin: A Three-Dimensional Fourier Synthesis at 5.5-Å. Resolution, Obtained by X-Ray Analysis. Nature. 1960;185(4711):416–22.

19. Wolfenden R. Analog approaches to the structure of the transition state in enzyme reactions. Acc Chem Res. 1972 Jan;5(1):10–8.

20. Radzicka A;, Wolfenden R. A proficient enzyme. Science (80-). 1995;267(5194):90–2.

21. Jencks WP. Binding Energy, Specificity, and Enzymic Catalysis: The Circe Effect. In: Advances in Enzymology and Related Areas of Molecular Biology. 1975. p. 219–410.

22. Warshel A, Levitt M. Theoretical Studies of Enzymic Reactions : Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. J Mol Biol. 1976;103:227–49.

23. Warshel A. Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites. Journal of Biological Chemistry. 1998.

24. Warshel A. Energetics of enzyme catalysis. Proc Natl Acad Sci U S A. 1978;75(11):5250–4.

25. Grant KL, Klinman JP. Evidence That both Protium and Deuterium Undergo Significant Tunneling in the Reaction Catalyzed by Bovine Serum Amine Oxidase. Biochemistry. 1989;28:6597–605.

26. Cha Y, Murray CJ, Klinman JP. Hydrogen Tunneling in Enzyme Reactions. Science (80-). 1989;243(4896).

27. Frey P, Whitt S, Tobin J. A low-barrier hydrogen bond in the catalytic triad of serine proteases. Science (80-). 1994;264(5167).

28. Cleland W, Kreevoy M. Low-barrier hydrogen bonds and enzymic catalysis. Science (80-). 1994;264(5167).

29. Gerlt J a, Gassman PG. Understanding the rates of certain enzyme-catalyzed reactions: Proton abstraction from carbon acids, acyl transfer reactions, and displacement reactions of phosphodiesters. Biochemistry. 1993;32(45):11943–52.

30. Cleland WW, Frey PA, Gerlt JA. The Low Barrier Hydrogen Bond in Enzymatic Catalysis. Biochemistry. 1998;273:25529–32.

31. Ishikita H, Saito K. Proton transfer reactions and hydrogen-bond networks in protein environments. J R Soc Interface. 2014;11(91):1–17.

32. Hur S, Bruice TC. Comparison of formation of reactive conformers (NACs) for the Claisen rearrangement of chorismate to prephenate in water and in the E. coli mutase: The efficiency of the enzyme catalysis. J Am Chem Soc. 2003;125(19):5964–72.

33. Hur S, Bruice TC. Enzymes do what is expected (chalcone isomerase versus chorismate mutase). J Am Chem Soc. 2003;125(6):1472–3.

34. Hur S, Bruice TC. The mechanism of catalysis of the chorismate to prephenate reaction by the Escherichia coli mutase enzyme. Proc Natl Acad Sci U S A. 2002;99(3):1176–81.

35. Bruice TC, Benkovic SJ. Chemical basis for enzyme catalysis. Biochemistry. 2000. p. 6267–74.

36. Lightstone FC, Bruice TC. Separation of Ground State and Transition State Effects in Intramolecular and Enzymatic Reactions. 2. A Theoretical Study of the Formation of Transition States in Cyclic Anhydride Formation. J Am Chem Soc. 1997;119(39):9103–13.

37. Lightstone FC, Bruice TC. Ground State Conformations and Entropic and Enthalpic Factors in the Efficiency of Intramolecular and Enzymatic Reactions. 1. Cyclic Anhydride Formation by Substituted Glutarates, Succinate, and 3,6-Endoxo-Δ4-tetrahydrophthalate Monophenyl Esters. J Am Chem Soc. 1996;118:2595–605.

38. Bruice TC, Lightstone FC. Ground state and transition state contributions to the rates of intramolecular and enzymatic reactions. Accounts of Chemical Research. 1999. p. 127–36.

39. Kurplus M, McCammon JA. Dynamics of Proteins: Elements and Function. Annu Rev Biochem. 1983 Jun;52(1):263–300.

40. Gavish B, Werber MM. Viscosity-dependent structural fluctuations in enzyme catalysis. Biochemistry. 1979;18(7):1269–75.

41. McCammon J a, Wolynes PG, Karplus M. Picosecond dynamics of tyrosine side chains in proteins. Biochemistry. 1979;18(6):927–42.

42. Careri G, Fasella P, Gratton E. Enzyme Dynamics: The Statistical Physics Approach. Annu Rev Biophys Bioeng. 1979 Jun;8(1):69–97.

43. Radkiewicz JL, Brooks CL. Protein dynamics in enzymatic catalysis: Exploration of dihydrofolate reductase. J Am Chem Soc. 2000;122(2):225–31.

44. Agarwal PK, Billeter SR, Rajagopalan PTR, Benkovic SJ, Hammes-Schiffer S. Network of coupled promoting motions in enzyme catalysis. Proc Natl Acad Sci U S A. 2002;99(5):2794–9.

45. Hammes GG. Multiple conformational changes in enzyme catalysis. Biochemistry. 2002. p. 8221–8.

46. Benkovic SJ, Hammes-Schiffer S. A perspective on enzyme catalysis. Science (80-). 2003;301(5637):1196–202.

47. Luo J, Bruice TC. Anticorrelated motions as a driving force in enzyme catalysis: the dehydrogenase reaction. Proc Natl Acad Sci U S A. 2004;101(36):13152–6.

48. Hammes GG. How Do Enzymes Really Work? J Biol Chem J Biol Chem. 2008;283(283):22337–46.

49. Hammes GG, Chang Y-C, Oas TG. Conformational selection or induced fit: a flux description of reaction mechanism. Proc Natl Acad Sci U S A. 2009;106(33):13737–41.

50. Nashine VC, Hammes-Schiffer S, Benkovic SJ. Coupled motions in enzyme catalysis. Current Opinion in Chemical Biology. 2010. p. 644–51.

51. Hammes GG, Benkovic SJ, Hammes-Schiffer S. Flexibility, diversity, and cooperativity: Pillars of enzyme catalysis. Biochemistry. 2011;50(48):10422–30.

52. Hanoian P, Liu CT, Hammes-Schiffer S, Benkovic S. Perspectives on Electrostatics and Conformational Motions in Enzyme Catalysis. Acc Chem Res. 2015 Feb 17;48(2):482–9.

53. Kohen A. Role of Dynamics in Enzyme Catalysis: Substantial versus Semantic Controversies. Acc Chem Res. 2015 Feb 17;48(2):466–73.

54. Singh P, Abeysinghe T, Kohen A. Linking protein motion to enzyme catalysis. Molecules. 2015. p. 1192–209.

55. Callender R, Dyer RB. The dynamical nature of enzymatic catalysis. Acc Chem Res. 2015;48(2):407–13.

56. Taraphder S, Maupin CM, Swanson JMJ, Voth GA. Coupling Protein Dynamics with Proton Transport in Human Carbonic Anhydrase II. J Phys Chem B. 2016;acs.jpcb.6b02166.

57. Nú Ñ Ez S, Wing C, Antoniou D, Schramm VL, Schwartz SD. Insight into Catalytically Relevant Correlated Motions in Human Purine Nucleoside Phosphorylase. J Phys Chem A. 2006;110:463–72.

58. Benkovic SJ, Hammes GG, Hammes-Schiffer S. Free-Energy Landscape of Enzyme Catalysis. Biochemistry. 2008;47:3317–3321.

59. Kamerlin SCL, Warshel A. At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis. Proteins: Structure, Function and Bioinformatics. 2010. p. 1339–75.

60. Warshel A, Bora RP. Perspective: Defining and quantifying the role of dynamics in enzyme catalysis. J Chem Phys J Chem Phys J Chem Phys J Chem Phys J Chem Phys J Chem Phys. 2016;144(144).

61. Adamczyk AJ, Cao J, Kamerlin SCL, Warshel A. Catalysis by dihydrofolate reductase and other enzymes arises from electrostatic preorganization, not conformational motions. Proc Natl Acad Sci U S A. 2011;108(34):14115–20.

62. Agarwal PK, Doucet N, Chennubhotla C, Ramanathan A, Narayanan C. Conformational Sub-states and Populations in Enzyme Catalysis. In: Methods in Enzymology. 2016. p. 273–97.

63. Hanoian P, Liu CT, Hammes-Schiffer S, Benkovic S. Perspectives on electrostatics and conformational motions in enzyme catalysis. Accounts of Chemical Research. 2015. p. 482–9.

64. Kohen A. Kinetic isotope effects as probes for hydrogen tunneling, coupled motion and dynamics contributions to enzyme catalysis. Prog React Kinet Mech. 2003;28(2):119–56.

65. Yang Y, Pan L, Lightstone FC, Merz KM. Chapter One – The Role of Molecular Dynamics Potential of Mean Force Calculations in the Investigation of Enzyme Catalysis. In: Methods in Enzymology. 2016. p. 1–29.

66. Junge K, Eicken H, Deming JW. Bacterial Activity at -2 to -20°C in Arctic Wintertime Sea Ice. Appl Environ Microbiol. 2004;70(1):550–7.

67. Kashefi K. Extending the Upper Temperature Limit for Life. Science (80-). 2003;301(5635):934–934.

68. Feller G, Gerday C. Psychrophilic enzymes: hot topics in cold adaptation. Nat Rev Microbiol. 2003 Dec;1(3):200–8.

69. Smalås AO, Leiros H-KS, Os V, Willassen NP, Leiros I. Cold adapted enzymes. Biotechnol Annu Rev. 2000;6:1–57.

70. Marx JC, Blaise V, Collins T, D'Amico S, Delille D, Gratia E, et al. A perspective on cold enzymes: current knowledge and frequently asked questions. Cell Mol Biol (Noisy-le-grand). 2004 Jul;50(5):643–55.

71. Somero GN. Adaptation of enzymes to temperature: searching for basic strategies". Comp Biochem Physiol B Biochem Mol Biol. 2004;139(3):321–33.

72. Collins T, Roulling F, Piette F, Marx J-C, Feller G, Gerday C, et al. Fundamentals of Cold-Adapted Enzymes. In: Psychrophiles: from Biodiversity to Biotechnology. Berlin, Heidelberg: Springer Berlin Heidelberg; 2008. p. 211–27.

73. Miyazaki K, Wintrode PL, Grayling RA, Rubingh DN, Arnold FH. Directed evolution study of temperature adaptation in a psychrophilic enzyme. J Mol Biol. 2000;297(4):1015–26.

74. Siddiqui KS, Cavicchioli R. Cold-Adapted Enzymes. Annu Rev Biochem. 2006;75:403–33.

75. Siddiqui KS. Some like it hot, some like it cold: Temperature dependent biotechnological applications and improvements in extremophilic enzymes. Biotechnol Adv. 2015;33:1912–22.

76. Cavicchioli R, Charlton T, Ertan H, Mohd Omar S, Siddiqui KS, Williams TJ. Biotechnological uses of enzymes from psychrophiles. Microb Biotechnol. 2011;4(4):449–60.

77. Low PS, Bada JL, Somero GN. Temperature Adaptation of Enzymes: Roles of the Free Energy, the Enthalpy, and the Entropy of Activation. Proc Natl Acad Sci USA. 1973;70(2):430–2.

78. Low PS, Somero GN. Adaptation of Muscle Pyruvate Kinases to Environmental Temperatures and Pressures. J Exp Zool. 1976;198(1):1–11.

79. Russell NJ, Russell NJ. Toward a molecular understanding of cold activity of enzymes from psychrophiles. Extremophiles. 2000;4:83–90.

80. Fields PA. Review: Protein function at thermal extremes: balancing stability and flexibility. Comp Biochem Physiol Part A. 2001;129.

81. Struvay C, Feller G. Optimization to Low Temperature Activity in Psychrophilic Enzymes. Int J Mol Sci. 2012;13:11643–65.

82. Siglioccolo A, Gerace R, Pascarella S. "Cold spots" in protein cold adaptation: Insights from normalized atomic displacement parameters (B'-factors). Biophys Chem. 2010;153(1):104–14.

83. Liang ZX, Tsigos I, Lee T, Bouriotis V, Resing KA, Ahn NG, et al. Evidence for increased local flexibility in psychrophilic alcohol dehydrogenase relative to its thermophilic homologue. Biochemistry. 2004;43(46):14676–83.

84. Chiuri R, Maiorano G, Rizzello A, Del Mercato LL, Cingolani R, Rinaldi R, et al. Exploring Local Flexibility/Rigidity in Psychrophilic and Mesophilic Carbonic Anhydrases. Biophysj. 96:1586–96.

85. Sonan GK, Receveur-Brechot V, Duez C, Aghajari N, Czjzek M, Haser R, et al. The linker region plays a key role in the adaptation to cold of the cellulase from an Antarctic bacterium. Biochem J. 2007;407:293–302.

86. Tehei M, Franzetti B, Madern D, Ginzburg M, Ginzburg BZ, Giudici-Orticoni M-T, et al. Adaptation to extreme environments: macromolecular dynamics in bacteria compared in vivo by neutron scattering. EMBO Rep. 2004;5:66–70.

87. Dick M, Weiergräber OH, Classen T, Bisterfeld C, Bramski J, Gohlke H, et al. Trading off stability against activity in extremophilic aldolases. Nat Publ Gr. 2015;

88. Tiberti M, Papaleo E. Dynamic properties of extremophilic subtilisin-like serine-proteases. J Struct Biol. 2011;174:69–83.

89. Kundu S, Roy D. Comparative structural studies of psychrophilic and mesophilic protein homologues by molecular dynamics simulation. J Mol Graph Model. 2009;27(8):871–80.

90. Papaleo E, Pasi M, Riccardi L, Sambi I, Fantucci P, Gioia L De, et al. Protein flexibility in psychrophilic and mesophilic trypsins. Evidence of evolutionary conservation of protein dynamics in trypsin-like serine-proteases. FEBS. 2008;582(6):1008–1018.

91. Olufsen M, Smalås AO, Moe E, Brandsdal BO. Increased flexibility as a strategy for cold adaptation: A comparative molecular dynamics study of cold- and warm-active uracil DNA glycosylase. J Biol Chem. 2005;280(18):18042–8.

92. Papaleo E, Tiberti M, Invernizzi G, Pasi M, Ranzani V. Molecular Determinants of Enzyme Cold Adaptation: Comparative Structural and Computational Studies of Cold- and Warm-Adapted Enzymes. Curr Protein Pept Sci. 2011;12(657–683).

93. Feller GN. Protein stability and enzyme activity at extreme biological temperatures. J Phys Condens Matter. 2010;22(32):323101.

94. Rentier-Delrue F, Mande SC, Moyens S, Terpstra P, Mainfroid V, Goraj K, et al. Cloning and Overexpression of the Triosephosphate isomerase Genes from Psychrophilic and Thermophilic Bacteria: Structural Comparison of the Predicted Protein Sequences. Journal of Molecular Biology. 1993. p. 85–93.

95. Davail S, Feller G, Narinx E, Gerday C. Cold adaptation of proteins. Purification, characterization, and sequence of the heat-labile subtilisin from the Antarctic psychrophile Bacillus TA41. J Biol Chem. 1994 Jul 1;269(26):17448–53.

96. Galkin A, Kulakova L, Ashida H, Sawa Y, Esaki N. Cold-adapted alanine dehydrogenases from two Antarctic bacterial strains: Gene cloning, protein characterization, and comparison with mesophilic and thermophilic counterparts. Appl Environ Microbiol. 1999;65(9):4014–20.

97. Garsoux GE, Lamotte J, Gerday C, Feller G. Kinetic and structural optimization to catalysis at low temperatures in a psychrophilic cellulase from the Antarctic bacterium Pseudoalteromonas haloplanktis. Biochem J. 2004;384:247–53.

98. Huston AL, Haeggström JZ, Feller G. Cold adaptation of enzymes: Structural, kinetic and microcalorimetric characterizations of an aminopeptidase from the Arctic psychrophile Colwellia psychrerythraea and of human leukotriene A 4 hydrolase. BBA - Proteins Proteomics. 2008;1784:1865–72.

99. Ramli ANM, Mahadi NM, Shamsir MS, Rabu A, Joyce-Tan KH, Murad AMA, et al. Structural prediction of a novel chitinase from the psychrophilic Glaciozyma antarctica PI12 and an analysis of its structural properties and function. J Comput Aided Mol Des. 2012;26(8):947–61.

100. Tang MAK, Motoshima H, Watanabe K. Cold adaptation: Structural and functional characterizations of psychrophilic and mesophilic acetate kinase. Protein J. 2014;33(4):313–22.

101. Parvizpour S, Razmara J, Jomah AF, Ramli ANM, Md Illias R, Shamsir MS. Structural and functional analysis of a novel psychrophilic b-mannanase. J Comput Aided Mol Des. 2014;28:685–98.

102. Feller G, Payan F, Theys F, Qian M, Haser R, Gerday C. Stability and structural analysis of α-amylase from the artarctic psychrophile Alteromonas haloplanctis A23. Eur J Biochem. Blackwell Publishing Ltd; 1994 Jun;222(2):441–7.

103. Kim SY, Hwang KY, Kim SH, Sung HC, Han YS, Cho Y. Structural basis for cold adaptation. Sequence, biochemical properties, and crystal structure of malate dehydrogenase from a psychrophile Aquaspirillium arcticum. J Biol Chem. 1999 Apr 23;274(17):11761–7.

104. Bentahir M. Structural, Kinetic, and Calorimetric Characterization of the Cold-active Phosphoglycerate Kinase from the Antarctic Pseudomonas sp. TACII18. J Biol Chem. 2000;275(15).

105. Van Petegem F, Collins T, Meuwis MA, Gerday C, Feller G, Van Beeumen J. The structure of a cold-adapted family 8 xylanase at 1.3 Å resolution. Structural adaptations to cold and investigation of the active site. J Biol Chem. 2003 Feb 21;278(9):7531–9.

106. Ásgeirsson B, Nielsen BN, Højrup P. Amino acid sequence of the cold-active alkaline phosphatase from Atlantic cod (Gadus morhua). Comp Biochem Physiol B Biochem Mol Biol. 2003 Sep;136(1):45–60.

107. Jónsdóttir LB, Ellertsson B, Invernizzi G, Magnúsdóttir M, Thorbjarnardóttir SH, Papaleo E, et al. The role of salt bridges on the temperature adaptation of aqualysin I, a thermostable subtilisin-like proteinase. Biochim Biophys Acta - Proteins Proteomics. 2014;1844(12):2174–81.

108. Bian F, Yue S, Peng Z, Zhang X, Chen G, Yu J, et al. A comprehensive alanine-scanning mutagenesis study reveals roles for salt bridges in the structure and activity of Pseudomonas aeruginosa elastase. PLoS One. 2015;10(3).

109. Pucci F, Rooman M. Physical and molecular bases of protein thermal stability and cold adaptation. Curr Opin Struct Biol. Elsevier Ltd; 2017;42(February):117–28.

110. Feller I G, Zekhnini Z, Lamotte-Brasseur J, Gerday C. Enzymes from cold-adapted microorganisms The class C ß-lactamase from the antarctic psychrophile Psychrobacter imrnobilis A5. Eur J Biochem. 1997;244:186–91.

111. Feller G. Review Molecular adaptations to cold in psychrophilic enzymes. C Cell Mol Life Sci. 2003;60:648–62.

112. Udg U, Dna-repair EC. research papers The structure of uracil-DNA glycosylase from Atlantic cod (Gadus morhua) reveals cold-adaptation features research papers. Acta Crystallogr D Biol Crystallogr. 2003 Aug;59(Pt 8):1357–65.

113. Russell RJ, Gerike U, Danson MJ, Hough DW, Taylor GL. Structural adaptations of the cold-active citrate synthase from an Antarctic bacterium. Structure. 1998;6(3):351–61.

114. Aghajari N, Feller G, Gerday C, Haser R. Crystal structures of the psychrophilic alpha-amylase from Alteromonas haloplanctis in its native form and complexed with an inhibitor. Protein Sci. Cold Spring Harbor Laboratory Press; 1998 Jun;7(6):564–72.

115. Leiros HK, Willassen NP, Smalås AO. Residue determinants and sequence analysis of cold-adapted trypsins. Extremophiles. 1999 Aug;3(3):205–19.

116. Paredes DI, Watters K, Pitman DJ, Bystroff C, Dordick JS, D'Amico S, et al. Comparative void-volume analysis of psychrophilic and mesophilic enzymes: Structural bioinformatics of psychrophilic enzymes reveals sources of core flexibility. BMC Struct Biol. 2011;11(1):42.

117. Violot S, Aghajari N, Czjzek M, Feller G, Sonan GK, Gouet P, et al. Structure of a full length psychrophilic cellulase from Pseudoalteromonas haloplanktis revealed by X-ray diffraction and small angle X-ray scattering. J Mol Biol. 2005;348(5):1211–24.

118. Gorfe AA, Brandsdal BO, Leiros HK, Helland R, Smalås AO. Electrostatics of mesophilic and psychrophilic trypsin isoenzymes: qualitative evaluation of electrostatic differences at the substrate binding site. Proteins. 2000 Aug 1;40(2):207–17.

119. Brandsdal BO, Smalås AO, Åqvist J. Electrostatic effects play a central role in cold adaptation of trypsin. FEBS Lett. 2001 Jun 15;499(1–2):171–5.

120. Gerike U, Danson MJ, Hough DW. Cold-active citrate synthase: mutagenesis of active-site residues. Protein Eng. 2001;14(9):655–61.

121. Kyomuhendo P, Myrnes B, Brandsdal BO, Smalås AO, Nilsen IW, Helland R. Thermodynamics and structure of a salmon cold active goose-type lysozyme. Comp Biochem Physiol - B Biochem Mol Biol. 2010 Aug;156(4):254–63.

122. Badieyan S, Bevan DR, Zhang C. Study and design of stability in GH5 cellulases. Biotechnol Bioeng. 2012;109(1):31–44.

123. Georlette D, Damien B, Blaise V, Depiereux E, Uversky VN, Gerday C, et al. Structural and functional adaptations to extreme temperatures in psychrophilic, mesophilic, and thermophilic DNA ligases. J Biol Chem. 2003;278(39):37015–23.

124. Smalås AO, Heimstad ES, Hordvik A, Willassen NP, Male R. Cold adaption of enzymes: Structural comparison between salmon and bovine trypsins. Proteins Struct Funct Genet. 1994 Oct;20(2):149–66.

125. Wallon G, Lovett ST, Magyar C, Svingor A, Szilagyi A, Zàvodszky P, et al. Sequence and homology model of 3-isopropylmalate dehydrogenase from the psychrotrophic bacterium Vibrio sp. I5 suggest reasons for thermal instability. Protein Eng. 1997;10(6):665–72.

126. Riise EK, Lorentzen MS, Helland R, Smalås AO, Leiros H-KS, Willassen NP. The first structure of a cold-active catalase from Vibrio salmonicida at 1.96 Å reveals structural aspects of cold adaptation. Acta Cryst. 2007;63:135–48.

127. Liu Y, Zhang T, Zhang Z, Sun T, Wang J, Lu F. Improvement of cold adaptation of Bacillus alcalophilus alkaline protease by directed evolution. J Mol Catal B Enzym. 2014;106:117–23.

128. Isaksen GV, Åqvist J, Brandsdal BO. Protein surface softness is the origin of enzyme cold-adaptation of trypsin. PLoS Comput Biol. 2014;10(8):e1003813.

129. Isaksen GV, Åqvist J, Brandsdal BO. Enzyme surface rigidity tunes the temperature dependence of catalytic rates. Proc Natl Acad Sci U S A. 2016;113(7822–7).

130. Altermark B, Helland R, Moe E, Willassen NP, Smalås AO. Structural adaptation of endonuclease I from the cold-adapted and halophilic bacterium Vibrio salmonicida. Acta Crystallogr D Biol Crystallogr. 2008;64(Pt 4):368–76.

131. Altermark B, Smalås AO, Willassen NP, Helland R. The structure of Vibrio cholerae extracellular endonuclease I reveals the presence of a buried chloride ion. Acta Crystallogr D Biol Crystallogr. 2006;62(Pt 11):1387–91.

132. Li C, Ho L, Chang Z, Tsai L, Yang W. DNA binding and cleavage by the periplasmic endonuclease Vvn from Vibrio Vulnificus : A novel structure with an old active site. Microbiology. 2003;22(15):1–40.

133. Kühlmann UC, Moore GR, James R, Kleanthous C, Hemmings AM. Structural parsimony in endonuclease active sites: Should the number of homing endonuclease families be redefined? FEBS Letters. 1999. p. 1–2.

134. Flick KE, Jurica MS, Monnat RJ, Stoddard BL. DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-PpoI. Nature. 1998;394(6688):96–101.

135. Miller MD, Tanner J, Alpaugh M, Benedik MJ, Krause KL. 2.1 Å structure of Serratia endonuclease suggests a mechanism for binding to double-stranded DNA. Nat Struct Biol. Nature Publishing Group; 1994 Jul;1(7):461–8.

136. Kleanthous C, Kühlmann UC, Pommer a J, Ferguson N, Radford SE, Moore GR, et al. Structural and mechanistic basis of immunity toward endonuclease colicins. Nat Struct Biol. 1999;6(3):243–52.

137. Chak K-F, Safot MK, Kutt W-Y, Hsieh S-Y, Yuant HS, Dickerson RE. The crystal structure of the immunity protein of colicin E7 suggests a possible colicin-interacting surface (protein-protein interaction/E-group colicins). Biochemistry. 1996;93:6437–42.

138. Raaijmakers H, Vix O, Tö I, Golz S, Rries Kemper B, Suck D. X-ray structure of T4 endonuclease VII: a DNA junction resolvase with a novel fold and unusual domain-swapped dimer architecture. EMBO J. 1999;18(6):1447–58.

139. Shen BW, Landthaler M, Shub DA, Stoddard BL. DNA binding and cleavage by the HNH homing endonuclease I-HmuI. J Mol Biol. 2004;342(1):43–56.

140. Sokolowska M, Czapinska H, Bochtler M. Crystal structure of the bba-Me type II restriction endonuclease Hpy99I with target DNA. Nucleic Acids Res. 2009;37(11):3799–810.

141. Ghosh M, Meiss G, Pingoud A, London RE, Pedersen LC. The nuclease A – inhibitor complex is characterized by a. J Biol Chem. 2007;282(8):5682–5690.

142. Loll B, Gebhardt M, Wahle E, Meinhart A. Crystal structure of the EndoG/EndoGI complex: Mechanism of EndoG inhibition. Nucleic Acids Res. 2009;37(21):7312–20.

143. Moon AF, Midon M, Meiss G, Pingoud A, London RE, Pedersen LC. Structural insights into catalytic and substrate binding mechanisms of the strategic EndA nuclease from Streptococcus pneumoniae. Nucleic Acids Res. 2011;39(7):2943–53.

144. Shen BW, Heiter DF, Chan S-H, Wang H, Xu S-Y, Morgan RD, et al. Unusual Target Site Disruption by the Rare-Cutting HNH Restriction Endonuclease PacI. Struct Des. 2010;18:734–43.

145. Brenner DJ, Krieg NR, Staley JT. Bergey´s manual of systematic bacteriology. 2nd ed. George Garrity, editor. New York, NY: Springer US; 2004.

146. Altermark B, Niiranen L, Willassen NP, Smalås AO, Moe E. Comparative studies of endonuclease I from cold-adapted Vibrio salmonicida and mesophilic Vibrio cholerae. FEBS J. 2007;274:252–63.

147. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B. 1998;102(18):3586–616.

148. Foloppe N, Mackerell AD. All-Atom Empirical Force Field for Nucleic Acids: I. Parameter Optimization Based on Small Molecule and Condensed Phase Macromolecular Target Data. J Comput Chem. 2000;21(2):86–104.

149. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, et al. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. J Am Chem Soc. 1995;117:5179–97.

150. Jorgensen WL, Maxwell DS, Tirado-Rives J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. J Am Chem Soc. 1996 Jan;118(45):11225–36.

151. Auffinger P, Westhof E. Simulations of the molecular dynamics of nucleic acids. Curr Opin Struct Biol. 1998 Apr;8(2):227–36.

152. Schleif R. Modeling and Studying Proteins with Molecular Dynamics. Methods Enzymol. 2004;383:28–47.

153. Lindhal E, Sansom M. Membrane proteins: molecular dynamics simulations. Curr Opin Struct Biol. 2008 Aug;18(4):425–31.

154. Grossfield A. Recent progress in the study of G protein-coupled receptors with molecular dynamics computer simulations. Biochim Biophys Acta - Biomembr. 2011 Jul;1808(7):1868–78.

155. Hertig S, Latorraca NR, Dror RO. Revealing Atomic-Level Mechanisms of Protein Allostery with Molecular Dynamics Simulations. Liu J, editor. PLOS Comput Biol. 2016 Jun 10;12(6):e1004746.

156. Mocci F, Laaksonen A. Insight into nucleic acid counterion interactions from inside molecular dynamics simulations is “worth its salt.” Soft Matter. 2012;8(36):9268.

157. Leach AR. Molecular modelling : principles and applications. Longman; 1996. 595 p.

158. Verlet L. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. Phys Rev. American Physical Society; 1967 Jul 5;159(1):98–103.

159. Hockney RW. The potential calculations and some applications. Methods Comput Phys. 1970;9(9):136–211.

160. Andreini C, Bertini I, Rosato A. Metalloproteomes: a bioinformatic approach. Acc Chem Res. 2009 Oct 20;42(10):1471–9.

161. Andreini C, Bertini I, Cavallaro G, Holliday GL, Thornton JM. Metal ions in biological catalysis: from enzyme databases to general principles. J Biol Inorg Chem. 2008 Nov;13(8):1205–18.

162. Mishra NC. Nucleases : molecular biology and applications. Wiley-Interscience; 2002. 314 p.

163. Hoops SC, Anderson KW, Merz KM. Force field design for metalloproteins. J Am Chem Soc. American Chemical Society; 1991 Oct;113(22):8262–70.

164. Hancock RD. Molecular mechanics calculations and metal ion recognition. Acc Chem Res. 1990;23(8):253–7.

165. Stote RH, Karplus M. Zinc binding in proteins and solution: A simple but accurate nonbonded representation. Proteins Struct Funct Genet. Wiley Subscription Services, Inc., A Wiley Company; 1995 Sep;23(1):12–31.

166. Åqvist J. Ion water interaction potentials derived from free energy perturbation simulations. J Phys Chem. 1990 Oct;94(21):8021–4.

167. Åqvist J, Warshel A. Computer simulation of the initial proton transfer step in human carbonic anhydrase I. J Mol Biol. 1992;224(1):7–14.

168. Duarte F, Bauer P, Barrozo A, Amrein BA, Purg M, Åqvist J, et al. Force field independent metal parameters using a nonbonded dummy model. J Phys Chem B. 2014;118(16):4351–62.

169. Oelschlaeger P, Klahn M, Beard WA, Wilson SH, Warshel A. Magnesium-cationic dummy atom molecules enhance representation of DNA polymerase beta in molecular dynamics simulations: improved accuracy in studies of structural features and mutational effects. J Mol Biol. NIH Public Access; 2007 Feb 16;366(2):687–701.

170. Pang YP, Xu K, Yazal JE, Prendergas FG. Successful molecular dynamics simulation of the zinc-bound farnesyltransferase using the cationic dummy atom approach. Protein Sci. Wiley-Blackwell; 2000 Oct;9(10):1857–65.

171. Åqvist J, Warshel A. Free energy relationships in metalloenzyme-catalyzed reactions. Calculations of the effects of metal ion substitutions in staphylococcal nuclease. J Am Chem Soc. American Chemical Society; 1990 Apr;112(8):2860–8.

172. Amadei A, Linssen ABM, Berendsen HJC. Essential Dynamics of Proteins. PROTEINS Struct Funct Genet. 1993 Dec;17(4):412–25.

173. Hess B. Convergence of sampling in protein simulations. Phys Rev E Stat Nonlin Soft Matter Phys. 2002 Mar;65(3 Pt 1):31910.

174. Lindorff-Larsen K, Ferkinghoff-Borg J. Similarity measures for protein ensembles. PLoS One. 2009;4(1).

175. Frey BJ, Dueck D. Clustering by passing messages between data points. Science. 2007 Feb 16;315(5814):972–6.

176. Agrafiotis DK, Xu H, Levitt M. A self-organizing principle for learning nonlinear manifolds. Proc Natl Acad Sci U S A. 2002;99(25):15869–15872.

177. Kullback S, Leibler RA. On Information and Sufficiency. Ann Math Stat. Institute of Mathematical Statistics; 1951 Mar;22(1):79–86.

178. Tiberti M, Papaleo E, Bengtsen T, Boomsma W, Lindorff-Larsen K. ENCORE: Software for Quantitative Ensemble Comparison. PLoS Comput Biol. 2015;11(10):1–16.

179. Vishveshwara S, Brinda K V, Kannan N. Protein Structure: Insights from graph theory. J Theor Comput Chem. 2002;1(1):1–25.

180. Taylor WR. Protein structure comparison using bipartite graph matching and its application to protein structure classification. Mol Cell Proteomics. 2002;1(4):334–9.

181. Csermely P, Korcsmáros T, Kiss HJM, London G, Nussinov R. Structure and dynamics of molecular networks: A novel paradigm of drug discovery: A comprehensive review. Pharmacol Ther. 2013;138(3):333–408.

182. Brinda K V, Vishveshwara S. A Network Representation of Protein Structures: Implications for Protein Stability. Biophys J. 2005;89:4159–70.

183. Rao F, Caflisch A. The protein folding network. J Mol Biol. 2004;342(1):299–306.

184. Bhattacharyya M, Ghosh A, Hansia P, Vishveshwara S. Allostery and conformational free energy changes in human tryptophanyl-tRNA synthetase from essential dynamics and structure networks. Proteins Struct Funct Bioinforma. 2010;78(3):506–17.

185. Ghosh A, Vishveshwara S. Variations in Clique and Community Patterns in Protein Structures during Allosteric Communication: Investigation of Dynamically Equilibrated Structures of Methionyl tRNA Synthetase Complexes †. Biochemistry. 2008 Nov 4;47(44):11398–407.

186. Del Sol A, O'Meara P. Small-world network approach to identify key residues in protein-protein interaction. Proteins Struct Funct Genet. 2005;58(3):672–82.

187. Watts DJ, Strogatz SH. Collective dynamics of "small-world" networks. Nature. 1998;393(6684):440–2.

188. Vendruscolo M, Dokholyan N V, Paci E, Karplus M. Small-world view of the amino acids that play a key role in protein folding. Phys Rev E Stat Nonlin Soft Matter Phys. 2002;65(6):4.

189. del Sol A, Tsai CJ, Ma B, Nussinov R. The Origin of Allosteric Functional Modulation: Multiple Pre-existing Pathways. Structure. 2009. p. 1042–50.

190. Bhattacharyya M, Bhat CR, Vishveshwara S. An automated approach to network features of protein structure ensembles. Protein Sci. 2013;22(10):1399–416.

191. Seeber M, Felline A, Raimondi F, Muff S, Friedman R, Rao F, et al. Wordom: a user-friendly program for the analysis of molecular structures, trajectories, and free energy surfaces. J Comput Chem. Wiley-Blackwell; 2011 Apr 30;32(6):1183–94.

192. Grant BJ, Rodrigues APC, ElSawy KM, McCammon JA, Caves LSD. Bio3d: an R package for the comparative analysis of protein structures. Bioinformatics. 2006 Nov 1;22(21):2695–6.

193. Vijayabaskar MS, Niranjan V, Vishveshwara S. GraProStr – Graphs of Protein Structures: A Tool for Constructing the Graphs and Generating Graph Parameters for Protein Structures. Open Bioinforma J. 2011;5:53–8.

194. Chakrabarty B, Parekh N. NAPS: Network analysis of protein structures. Nucleic Acids Res. 2016;44(W1):W375–82.

195. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A software Environment for integrated models of biomolecular interaction networks. Genome Res. 2003;

196. Warshel A. Computer modeling of chemical reactions in enzymes and solutions. Wiley; 1997. 236 p.

197. Warshel A, Weiss RM. An Empirical Valence Bond Approach for Comparing Reactions in Solutions and in Enzymes. J Am Chem Soc. 1980;102:6218–26.

198. Leach AR. Molecular modelling : principles and applications. Second Edi. Pearson Education Limited; 2001. 744 p.

199. Niiranen L, Altermark B, Brandsdal BO, Leiros H-KS, Helland R, Smalås AO, et al. Effects of salt on the kinetics and thermodynamic stability of endonuclease I from Vibrio salmonicida and Vibrio cholerae. FEBS J. 2008;275(7):1593–605.

200. Friedhoff P, Franke I, Meiss G, Wende W, Krause KL, Pingoud A. A similar active site for non-specific and specific endonucleases. Nat Struct Biol. 1999;6(2):112–3.

201. Bueren-Calabuig JA, Coderch C, Rico E, Jiménez-Ruiz A, Gago F. Mechanistic insight into the catalytic activity of ββα-metallonucleases from computer simulations: Vibrio vulnificus periplasmic nuclease as a test case. ChemBioChem. 2011;12(17):2615–22.

202. Fedøy AE, Yang N, Martinez A, Leiros HKS, Steen IH. 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. 2007;372(1):130–49.

203. Olufsen M, Papaleo E, Smalås AO, Brandsdal BO. Ion pairs and their role in modulating stability of cold- and warm-active uracil DNA glycosylase. Proteins Struct Funct Genet. 2008;71(3):1219–30.

204. Svingor Á, Kardos J, Hajdú I, Németh A, Závodszky P. A better enzyme to cope with cold: Comparative flexibility studies on psychrotrophic, mesophilic, and thermophilic IPMDHS. J Biol Chem. in Press; 2001;276(30):28121–5.

205. Collins T, Meuwis MA, Gerday C, Feller G. Activity, stability and flexibility in glycosidases adapted to extreme thermal environments. J Mol Biol. 2003;328(2):419–28.

206. Choi YG, Park CJ, Kim HE, Seo YJ, Lee AR, Choi SR, et al. Comparison of backbone dynamics of the type III antifreeze protein and antifreeze-like domain of human sialic acid synthase. J Biomol NMR. 2015;61(2):137–50.

207. Papaleo E, Tiberti M, Invernizzi G. Molecular Dynamics Simulations to Study Structure-Function Relationship in Psychrophilic Enzymes. In: Biotechnology of Extremophiles. Springer International Publishing Switzerland; 2016. p. 675–98.

208. Kazemi M, Himo F, Åqvist J. Enzyme catalysis by entropy without Circe effect. Proc Natl Acad Sci U S A. 2016;113(9):2406–11.

209. Isaksen GV, Hopmann KH, Åqvist J, Brandsdal BO. Computer Simulations Reveal Substrate Specificity of Glycosidic Bond Cleavage in Native and Mutant Human Purine Nucleoside Phosphorylase. Biochemistry. 2016 Apr 12;55(14):2153–62.