UiT
THE ARCTIC
UNIVERSITY
OF NORWAY

Department of Chemistry

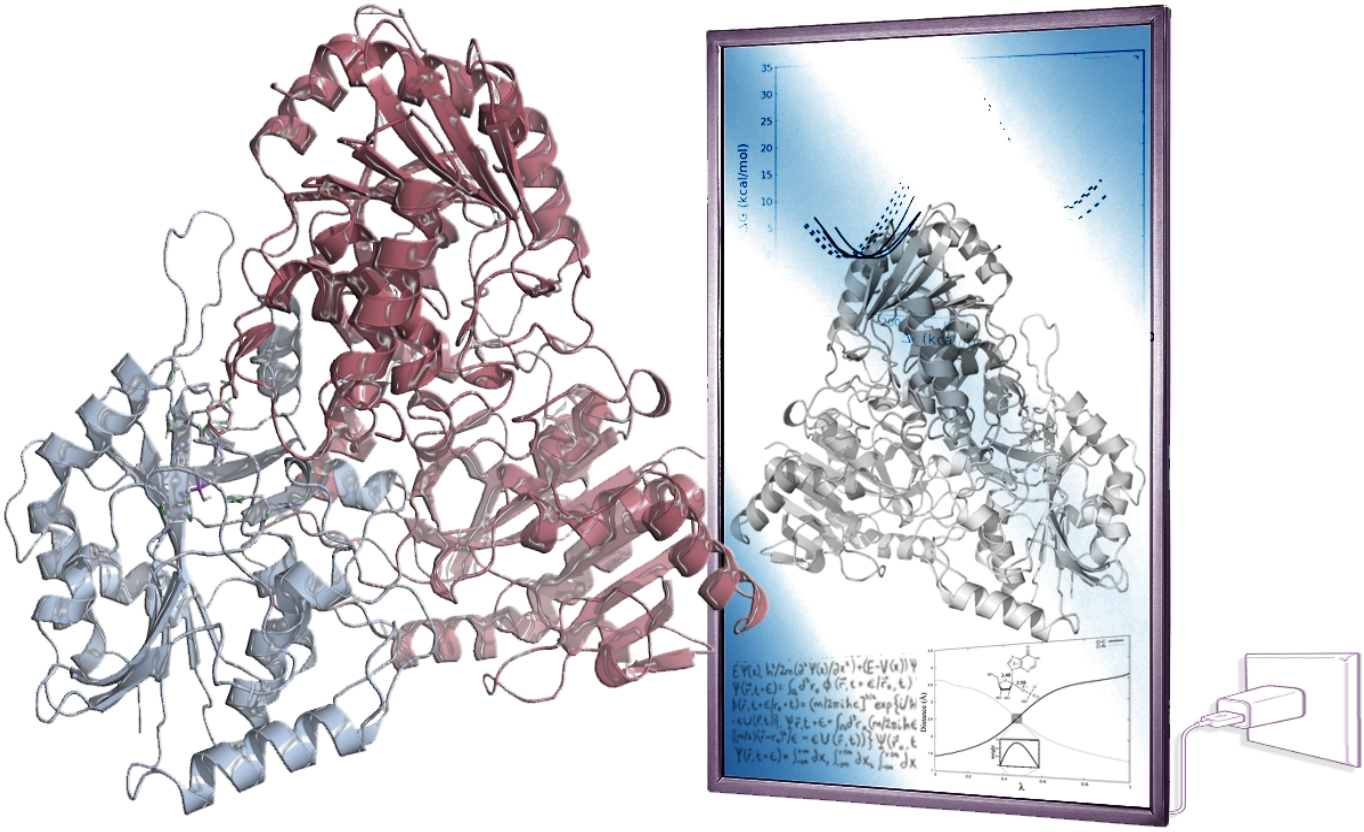
Protein Dynamics Regulates Enzyme Enthalpy-Entropy Balance

Computational Modeling of Enzyme Structure Activity Relationships

—

Geir Villy Isaksen

A dissertation for the degree of Philosophiae Doctor – April 2015



UNIVERSITY OF TROMSØ

UNIVERSITY OF TROMSØ

FACULTY OF SCIENCE AND TECHNOLOGY
DEPARTMENT OF CHEMISTRY

Protein Dynamics Regulates Enzyme Enthalpy-Entropy Balance

Computational Modeling of Enzyme Structure Activity Relationships

Geir Villy Isaksen
A dissertation for the degree of Philosophiae Doctor

April 2015

Abstract

Accurate predictions of free energies and the corresponding enthalpies and entropies from computer simulations are invaluable for understanding enzymatic catalysis and drug actions in terms of actual structure activity relationships. To facilitate the management and handling of the enormous amount of data generated when calculating thermodynamic parameters, we have developed a high-throughput interface (Qgui) for automated free energy and empirical valence bond calculations. We present simulations of cold- and warm-adapted trypsin that quantitatively capture experimental catalytic rates of the two enzymes and further yield high-precision Arrhenius plots, which show the characteristic trends in activation enthalpy and entropy. The relationship between these parameters and the 3D structure is reflected by significantly different internal protein energy changes during the reaction. The origin of this effect is not localized to the active site, but in the outer regions of the protein, where the cold-active enzyme has higher degree of softness. Stiffening the outermost loop regions in cold-adapted trypsin completely changes the enzyme from cold- to warm-adapted in terms of enthalpy and entropy. Several structural mechanisms for softening the protein surface in warm-adapted trypsin are also identified, together with key mutations responsible for this effect. The effect to the thermodynamic activation parameters induced by remote mutations was further investigated utilizing purine nucleoside phosphorylase (PNP). Here the exact reaction mechanism was not known, and it was thus necessary to resolve this first. The presented reaction mechanism reproduce experimental trends for both reaction rates and thermodynamic activation parameters in native and mutated PNP. Our calculations also show that the human PNP substrate specificity for 6-oxopurines over 6-aminopurines originates from significant differences in electrostatic preorganization. Both calculations on trypsin and PNP show that mutations distant to the active site alter the enthalpy-entropy balance without significantly changing the reaction rates. The enthalpy-entropy compensation induced by the mutations originates from altered protein surface softness.

Acknowledgments

The work presented in this thesis was carried out at the Center of Theoretical and Computational Chemistry (CTCC), Department of Chemistry, Faculty of Science and Technology at the University of Tromsø from July 2010 to April 2015.

I would like to thank my supervisor Bjørn Olav Brandsdal for excellent guidance during the work presented in this thesis. Thank you for always being both supportive and brutally honest. Thanks for all the nice conversations, great trips to Uppsala, and for introducing me to O'Connor's and the world of great whiskeys. It has been an honor to be part of establishing the *Brandsdal Group*. I would also like to express my sincere gratitude to Johan Åqvist for great collaborations. Your scientific knowledge and insight is truly inspiring. Thank you for always welcoming me to your lab, the yearly 'julebord' and for great Italian lunches.

My co-authors Tor Arne Heim Andberg and Kathrin Helen Hopmann are sincerely acknowledged for great collaborations and their contributions to the presented works in this thesis. I would also like to thank Christoffer Lind for coming all the way from Uppsala to Tromsø to help out on the FEP development for our Qgui software. NOTUR is sincerely acknowledged for the many millions of CPU hours they have granted me on STALLO. A particular thanks goes to Roy Dragseth, Steinar Traedal-Henden, Espen Tangen and the rest of the crew at the HPC for superb service and support. Thank you for the many times you have fixed a STALLO-related problem on any day and time of the week.

I would like to acknowledge the rest of the *Brandsdal Group* people; Davide-1 (D1/store-Davide), Laura, Ravna and Davide-2 (D2/lille-Davide) for collaborations and for contributing to a great working environment. Thanks to all the people at the CTCC and the Åqvist lab in Uppsala for coffee breaks and nice conversations. Thanks to BioStruct for funding, courses, and for hosting the all time famous annual conference at Jægtvolden.

My special heartfelt gratitude goes to my wife Silje and our two beautiful daughters Othelie and Tora Sofie. Every day with you girls is an adventure. Silje, you deserve a "PhD in patience" for always supporting my workaholic tendencies and my, probably even worse, addiction for fly-fishing after arctic silver.

List of papers

The thesis is based on the following papers and manuscripts, which are referred to in the text by their Roman numerals.

I. **Isaksen GV**, Andberg TAH, Åqvist J & Brandsdal BO (2015) Q-GUI: A high-throughput interface for automated setup and analysis of free energy calculations and empirical valence bond simulations. *J Mol Graph Mod. In print.*

II. **Isaksen GV**, Åqvist J, & Brandsdal BO (2014) Protein surface softness is the origin of enzyme cold-adaptation of trypsin. *PLoS computational biology* 10(8):e1003813.

III. **Isaksen GV**, Åqvist J, & Brandsdal BO (2015) Computational proof for protein surface softness regulating the enthalpy-entropy balance responsible for trypsin cold-Adaptaion *Manuscript*.

IV. **Isaksen GV**, Hopmann, KH., Åqvist, J, & Brandsdal BO (2015) Computer simulations reveals substrate specificity of glycosidic bond cleaveg in native and mutant Human Purine Nucleoside Phosphorylase. *Submitted to Biochemistry*

V. **Isaksen GV**, Åqvist J, & Brandsdal BO (2015) Thermodynamics of glycosidic bond cleavage in human purine nucleoside phosphorylase reavealed with empirical valence bond free energy calculations. *Manuscript*

Other publications

The following papers were not included in the thesis.

i. Karstad, R, **Isaksen, GV**, Brandsdal, B-O, Svendsen, & JS, Svenson, J (2010) Unnatural Amino Acid side Chains as S1, S1', and S2' Popes Yield Cationic Antimicrobial Peptides with Stability toward Chymotryptic Degradation. *J. Med. Chem. 53(15), pp 5558-5566*

ii. Karstad, R, **Isaksen, GV**, Wynendaele, E, Guttormsen, Y, Spiegeleer, BD, Brandsdal, B-O, Svendsen, JS, & Svenson, J (2012) Targeting the S1 and S3 Subsite of Trypsin with Unnatural Cationic Amino Acids Generates Antimicrobial Peptides with Potential for Oral Administration. *J. Med. Chem. 55(14), pp 6294-6305*

iii. Flate, GE, Kottra, G, Stensen, W, **Isaksen, GV**, Karstad, R, Svendsen, JS, Daniel, H, & Svenson, J (2011) In Vitro Characterization of Human Peptide Transported hPEPT1 Interactions and Passive Permeation Studies of Short Cationic Antimicrobial Peptides. *J. Med. Chem. 54(7), pp 2422-2432*

Abbreviations

ADE	Adenine
ADO	Adenosine
AST	Anionic salmon trypsin
BT	Bovine Trypsin
ES	Enzyme-substrate
EVB	Empirical valence bond
FEP	Free energy perturbation
fs	Femtosecond (10^{-15})
GUA	Guanine
GUO	Guanosine
HYP	Hypoxanthine
INO	Inosine
LIE	Linear interaction energy
LRA	Linear response approximation
MD	Molecular dynamics
MM	Molecular mechanics
PNP	Purine nucleoside phosphorylase
QM	Quantum mechanics
RMSF	Root-mean-square fluctuation
RRC	Reference reaction calibration

Contents

1 Introduction

Almost 200 years has passed since the French chemist Anselme Payen discovered the first enzyme, distase (1). Since then the enormous catalytic power utilized by enzymes has remained one of the key secrets in biology. Catalytic power here means the ability to significantly increase the rates that chemical reactions use from reactants to products. Enzymes play fundamental roles in almost all life processes, allowing cells to carry out reactions that otherwise would not occur on biological useful timescales, rendering them literally indispensable for any living organism. Their involvement in transcription and translation of genetic information, signaling control, energy transduction and a great variety of metabolic reactions, underscores why there exists a broad interest in understanding the origin of this catalytic power on a molecular level. However, the complexity of the numerous bonded and non-bonded interactions functioning as the machinery behind enzyme-catalyzed reactions makes the problem of identifying the responsible mechanisms at a molecular level far from intuitive.

One very useful way of narrowing the problem down is to investigate nature's own enzyme design strategy by studying differently adapted enzyme orthologs. A particularly fascinating aspect in this respect regards the mechanisms involved in the adaptive capabilities for life in extreme environments. For example, enzymes from cold and warm adapted organisms typically display very similar three dimensional (3D) structures and regions with highly conserved amino acid sequences, but their operational temperature optima are significantly different. A comprehensive understanding of such enzyme structure – activity relationships is of considerable interest, not only to provide information regarding evolutionary adaptation processes, but potentially also to enable rational design of enzymes. Mutational experiments have been useful for identifying catalytic factors in enzymatic reactions (2), but they cannot uniquely identify the mechanisms of catalysis or environmental adaptation. That is, no current experimental techniques can provide direct correlations between enzyme-substrate (ES) complexes and the detailed contributions to their transition states (3). In principle, computer simulations can provide a unique way of analyzing enzyme reaction energetics and establishing such correlations.

The ability to efficiently calculate free energies reflecting reality is of uttermost importance as the energetics often provides the most important and useful link between structure and function of biomolecular systems. Moreover, binding free energies, solvation energies and activation free energies from thermodynamic and kinetic experiments are directly comparable to results obtained with simulations techniques. It is however difficult to evaluate enzyme – structure activity relationships solely based on the catalytic barriers for different temperature adapted enzyme orthologs as they usually are similar at room temperature (4). It thus becomes critical that thermodynamic activation parameters ($\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$) are reproduced in addition to high precision free energy profiles along the reaction pathway. The only way to do this computationally is to obtain Arrhenius plots for the activation free energy as a function of the temperature. As this involves calculating a significantly large number of free energy profiles over a suitable temperature range, the extensive sampling by molecular dynamics (MD) simulations precludes the use of most standard quantum mechanics (QM) / molecular mechanics (MM), QM/MM approaches. However, the empirical valence bond (EVB) model (5, 6) utilized in the presented work provides a very efficient method for this purpose.

1.1 Objectives of Study

The main objectives of the presented work in this thesis can be summarized as follows:

- Develop a graphical user interface for high-throughput setup and analysis of free energy calculations and EVB simulations in biological systems.
- Investigate the molecular mechanism of trypsin temperature adaptation.
- Investigate the catalytic mechanism of purine nucleoside phosphorylase (PNP).
- Investigate the impact on the enthalpy – entropy balance induced by distant mutations to the active site in trypsin and PNP.

The software development was crucial for the presented work as it enabled comprehensive submission and analysis of high precision free energy calculations that otherwise would be exceedingly time dependent and cumbersome. It was utilized advantageously for exploring enzyme structure – activity relationships through extensive free energy calculations. The possible origin of trypsin temperature adaptation was reviewed, and the enthalpy – entropy balance was explored as a function

of distant mutations/modifications to the active site in both trypsin and human PNP. In the latter case, the exact enzyme reaction mechanism for the glycosidic bond cleavage step was not known and it was thus necessary to resolve this first.

2 Enzyme Catalysis: Quid Agis?

In this chapter some relevant concepts and proposals put forward to rationalize the catalytic effect of enzymes will be introduced to emphasize the motivation for the presented work.

2.1 Defining the Catalytic Effect

Enzymes catalyze the conversion of substrates (S) into corresponding products (P). However, before catalysis takes place a substrate has to diffuse and fit into the active site generating an enzyme-substrate (ES) complex. The ES complex is then converted to an enzyme-product (EP) complex prior to release of the product (P). The process can be illustrated schematically by the following simple generic enzymatic reaction:

$$\mathrm{E+S} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{ES} \xrightarrow{k_{\mathrm{cat}}} \mathrm{EP} \rightarrow \mathrm{E+P} \tag{1}$$

Here k_1, k_{-1} and k_{cat} are the reaction rate constants for the formation and dissociation of the ES complex, and the catalytic step resulting in the EP complex, respectively. It has previously been shown that many enzymes appear to have evolved to optimize the catalytic efficiency (7) given by $k_{\mathrm{cat}}/K_{\mathrm{M}}$ where

$$K_{\mathrm{M}} = \frac{(k_{-1} + k_{\mathrm{cat}})}{k_1} \tag{2}$$

is known as the Michaelis constant, which represents the substrate concentration at which the reaction rate is half of the maximum rate. Either maximizing k_{cat} or minimizing K_{M} can optimize the catalytic efficiency. In the work presented in this thesis k_{cat} has been considered through its relation to the activation free energy from transition state theory:

$$k_{\mathrm{cat}} = \kappa \frac{k_B T}{h} \exp\left(\frac{-\Delta G^{\ddagger}}{k_B T}\right) = \kappa \frac{k_B T}{h} \exp\left(\frac{\Delta S^{\ddagger}}{k_B} - \frac{\Delta H^{\ddagger}}{k_B T}\right) \tag{3}$$

Here κ is the transmission coefficient (generally assumed close to 1), k_B is the Boltzmann constant, h Planck's constant, T the temperature and $\Delta G^{\ddagger}$, $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ are the activation free energy, activation enthalpy and activation entropy, respectively. However, to quantitatively evaluate enzyme catalysis it is necessary to define the catalytic effect relative to something.

The most logical reference to an enzyme-catalyzed reaction is the uncatalyzed reaction in water (see Figure 1). Obviously, the mechanism of the reaction can be

different in water than in the enzyme, and this must be taken into consideration. As previously pointed out by Warshel and co-workers (5, 6), investigating how enzyme environments are able to accelerate reactions can instead be more beneficially compared to a reference reaction in water with the same mechanism. As illustrated in Figure 1, this involves comparing the free energy for the solvent cage relative to that of the same reaction in the enzyme.

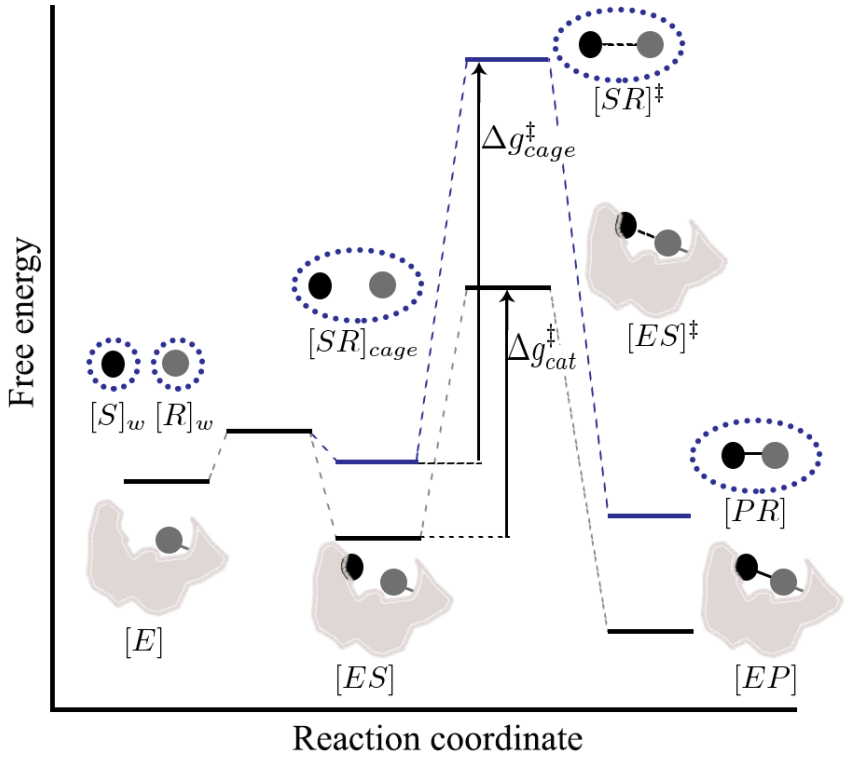


Figure 1: Free energies along a reaction coordinate illustrated for a reaction in solution (blue dotted circles) and in enzyme (gray shape). In solution, the substrate (S) and the reactive group (R) must move from separate solvent shells to become within reacting distance in a single solvent cage. It is the free energy required to form the solvent cage transition state ($\Delta \boldsymbol{g}^{\ddagger}_{\boldsymbol{cage}}$) that is of interest since it is comparable to the activation free energy of the enzyme-substrate complex ($\Delta \boldsymbol{g}^{\ddagger}_{\boldsymbol{cat}}$).

Here the solvent cage refers to the situation of moving the substrate and the reactive group from individual water shells to a reacting distance in the same solvent cage. Thus, the effect of the enzyme to the reaction rate can be observed simply in terms of changing the surrounding environment of the reacting fragment. Relevant k_{cage} values can furthermore be obtained both from experimental information and *ab initio* calculations for appropriate reactions in solutions (5, 6). The task then reduces to finding the true origin for the fact that $\Delta g^{\ddagger}_{\text{cat}}$ is significantly lower than $\Delta g^{\ddagger}_{\text{cage}}$. Thus, catalytic power is in

the context of the work presented in this thesis referred to as the reduction in $\Delta g^{\ddagger}_{\mathrm{cat}}$ relative to an appropriate reference reaction in water, $\Delta g^{\ddagger}_{\mathrm{cage}}$.

2.2 Are Free Energies the Answer to Everything?

As indicated above, free energies are unquestionably a key parameter for understanding enzymatic reactions and also chemical reactions in general. This is particularly true from a computational point of view since energetics from thermodynamic and kinetic experiments are directly comparable to those obtained with simulations techniques. Moreover, in Biochemistry 5th edition by Berg *et al.* (8) under section 8.2 it is stated that:

> *"To fully understand how enzymes operate, we need to consider two thermodynamic properties of the reaction: 1) the free-energy difference between the products and reactants (ΔG_0) and 2) the energy required to initiate the conversion of reactants to products ($\Delta G^{\ddagger}$)"*.

From this statement it is easy to get the impression that the catalytic power of enzymes can be rationalized solely based on ΔG_0 and $\Delta G^{\ddagger}$. Clearly, a key feature of enzymatic reactions is the observed lowering of the activation free energy compared to the uncatalyzed reaction. However, the free-energy difference between the products and the reactants is completely independent of the enzyme. That is, ΔG_0 (S → P) is a state function and therefore unaffected by the reaction pathway and corresponding energetics between its initial and final state. Considering the free energy difference between the products (P) and the reactants (R) can consequently not reveal how enzymes operate. As can be recognized from Eq. 1 and Figure 1, it makes more sense in this context to consider the free energy difference between the EP and ES complex relative to an appropriate solvent cage reaction. Furthermore, realizing that enzymes are able to drastically increase chemical reaction rates by lowering the activation free energy does not explain the origin of the catalytic effect. It has, however, been obvious for decades (9) that enzymes must reduce the catalytic barrier by interacting differently with the ES and ES‡ states compared to the relevant reference reaction. However, as the concept of rationalizing the origin of the catalytic power is an energy issue, resolving it without the ability of dissecting the observed energy into individual contributions becomes very difficult.

Computer simulations provide a unique way of analyzing enzyme reaction energetics at a molecular level. However, analysis of indirect or circumstantial factors such as electrostatics, hydrophobicity and flexibility alone does not suffice for obtaining conclusive evidence. In this respect it is critical that reliable free energy profiles for the reaction pathway reproducing experimental reaction rates can be obtained. That is, it is essential that the observed $\Delta g^{\ddagger}_{\mathrm{cat}} - \Delta g^{\ddagger}_{\mathrm{cage}}$ can be reproduced. If this is accomplished, revealing which energy components are responsible for the observed effect is in principle simple. Along the appropriate reaction coordinate, it must ultimately be possible to translate the corresponding energy components to the enzyme 3D structure. Such free energy calculations have for example previously been applied to identify the general notion that electrostatic effects play an important role in stabilizing the transition state of enzymatic reactions (10, 11).

2.3 Electrostatic Preorganization Explains Enzyme Catalysis

From extensive EVB studies reported in the literature, it has consistently been found that the largest catalytic effect is associated with electrostatic contributions (6, 12-17). As also mentioned above, it is the electrostatic stabilization of the transition state that is improved in the enzyme with respect to the solvent cage. This conception has become more or less the general notion for understanding and explaining the origin of enzyme catalysis. However, even though the importance of electrostatic contributions to enzyme catalysis was proposed as early as 1967 (18), the finding is not as trivial as implied here.

Enzyme electrostatic contribution to catalysis was at first seemingly inconsistent with studies prior to computer modeling. This was for example challenging since experiments with model compounds in solutions were not able to reproduce large electrostatic effects, even with ionized groups covalently linked to the enzyme aligned to stabilize the ionic transition state (19, 20). It has also been argued that protein active sites must have low dielectric constants to enhance the electrostatic effects relative to a reaction in aqueous solution where the dielectric constant is large (21). A problem with the last statement was that ionized groups that were supposed to be the source of electrostatic effects in proteins would not be ionized in low dielectric sites. In fact, it has been shown that the activation barrier will increase in a hypothetical low dielectric protein site due to desolvation effects (22). However, as pointed out in references (3, 23,

24), protein active sites are usually very polar heterogeneous sites. Consequently, such arguments become irrelevant in this respect. Thus, it could be argued that the same stabilization effect as for enzymes can be obtained in other polar solvents that can reorient their dipoles toward the charge distribution of the transition state. The fact that the average electrostatic (q) interaction between the transition state of an enzyme and the surrounding dipoles (μ), $\langle \Delta U_{q\mu}^{p} \rangle$, is not larger than the corresponding interaction in the solvent cage, $\langle \Delta U_{q\mu}^{w} \rangle$, became the big puzzle in explaining the electrostatic stabilization effect (23). However, in solution it costs substantial energy to reorient the solvent dipoles to their transition-state configuration. Generally, it is found that about half of the solvation free energy (ΔG_{sol}) associated with charge-dipole interactions is spent on dipole-dipole repulsion, $\langle \Delta U_{\mu\mu} \rangle$, so that (23)

$$\Delta G_{\mathrm{sol}} \cong \left\langle \Delta U_{q\mu} \right\rangle + \left\langle \Delta U_{\mu\mu} \right\rangle \cong \frac{1}{2} \left\langle \Delta U_{q\mu} \right\rangle \tag{4}$$

The explanation to this, as demonstrated in reference (23), is that enzymes have already paid a significant part of $\langle \Delta U_{\mu\mu} \rangle$ during the folding process. The folding energy is used to compensate for the dipole-dipole repulsion and to align the active site dipoles in a way that allows the solvation of the transition state (ΔG_{sol}) to approach $\langle \Delta U_{q\mu} \rangle$. In other words, enzymes are preorganized with respect to their dipoles, and this is the true origin of the observed electrostatic stabilization of the transition state (5, 23). The dipole-dipole repulsion term here is basically the reorganization energy (5, 25) involved in the process of forming the transition state charges. Thus, the observed electrostatic stabilization is a result of smaller reorganization energies in enzyme-catalyzed reactions with respect to the solvent cage reaction.

The electrostatic preorganization concept is today well established and has existed for decades (23) explaining the origin of enzyme's catalytic power. It has been applied numerous times in the literature to argue for and against popular assumptions and misconceptions aiming at explaining how enzymes are able to significantly increase chemical reaction rates. For further reading beyond the cases mentioned above, see e.g. references (3, 5, 16, 17, 26, 27) and references cited therein. However, the origin of catalytic rate optimization in cold-adapted enzymes, in terms of actual structure-function relationships, still remains one of the most intriguing problems in biology.

2.4 Clues from Nature's Enzyme Design in Cold Adaptation

Organisms adapted to survive in cold environments (psychrophiles) display a remarkable ability to grow and colonize at temperatures close to the freezing point of water. As can be recognized from Eq. 3, a key problem from the viewpoint of chemical kinetics is the exponential decrease in reaction rates upon decreasing the temperature. Decomposing the activation free energy ($\Delta G^{\ddagger}$) into entropic ($-T\Delta S^{\ddagger}$) and enthalpic ($\Delta H^{\ddagger}$) contributions further reveals that it is $\Delta H^{\ddagger}$ giving rise to the temperature dependency (Eq. 3). Lowering the temperature from 37°C to 0°C for a mesophilic enzyme typically results in a 20 – 250 fold reduction of the activity (28). Obviously, life in cold environments requires a vast array of adaptive features at nearly all level of the cell architecture and function, including protein stability and enzyme kinetics.

The fact that different temperature adapted enzyme orthologs usually display similar overall activation free energies around room temperature (29), renders it challenging to reveal the molecular source of such adaptations. Thus, in light of the discussions in the previous sections, the origin of catalytic rate optimization cannot be resolved solely by reproducing $\Delta g^{\ddagger}_{\mathrm{cat}} - \Delta g^{\ddagger}_{\mathrm{cage}}$. However, there exist a seemingly universal characteristic that catalyzed reactions of cold-adapted enzymes display lower enthalpy and a more negative entropy of activation compared to their mesophilic and thermophilic counterparts (28-30). The decrease in $\Delta H^{\ddagger}$ is furthermore believed to be the primary adaptation in psychrophilic enzymes (4, 31, 32) rendering the reaction rates less temperature dependent (Eq. 3). The observed lowering of $\Delta S^{\ddagger}$ at the expense of $\Delta H^{\ddagger}$ has also led to the long-standing assumption that cold-adaption originates from increased flexibility of the active site (30). Surely, increased flexibility in the active site would require more ordering of the substrate and the active site along the reaction coordinate as the reaction barrier is surmounted (i.e., a more negative $\Delta S^{\ddagger}$). Regardless of this, no strong experimental support for this hypothesis seems to exist. For example, X-ray analysis of warm- and cold-active trypsins did not indicate any overall flexibility differences between the two enzymes (33). Moreover, recent computer simulations for different temperature adapted citrate synthases, reproducing the important trends with respect to the thermodynamic activation parameters, showed that the flexibility of the highly conserved active site residues was close to identical (34). Instead, it was indicated that differences in protein softness outside the active site appeared to be correlated with the differences in the thermodynamic activation parameters.

If computer simulations are to successfully aid in resolving the origin of cold adaptation, it is crucial that the characteristic enthalpy-entropy balance can be reproduced in addition to reliable free energy profiles along the reaction pathway. As already stated in the introduction chapter, activation enthalpies and entropies can be obtained from Arrhenius plots, but this is extremely demanding in terms of computational costs and time. This is likely the reason why there to date exist very few examples of Arrhenius plots obtained with simulation techniques for enzyme reactions in the literature. In the work presented in this thesis, enzyme kinetics has been extensively calculated for trypsin (papers **II** and **III**) with a particular focus on the distinctive enthalpy-entropy balance obtained from high precision Arrhenius plots.

3 Computational Approaches

In this chapter the theoretical framework that the presented work is based upon will be introduced.

3.1 Force Fields

The interactions between atoms in a system are in molecular mechanics described by an empirical or semi-empirical derived energy function (U_{pot}), which typically takes the form

$$\begin{aligned} U_{pot} = & \sum_{bonds} \frac{1}{2} k_b (b - b_0)^2 + \sum_{angles} \frac{1}{2} k_\theta (\theta - \theta_0)^2 \\ & + \sum_{impropers} \frac{1}{2} k_\xi (\xi - \xi_0)^2 + \sum_{torsions} \frac{1}{2} k_\phi (1 + \cos(n\phi - \delta)) \\ & + \sum_{nonbonded} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + \sum_{nonbonded} \left(\frac{A_i A_j}{r_{ij}^{12}} - \frac{B_i B_j}{r_{ij}^{6}} \right) \end{aligned} \quad (5)$$

The first 4 terms in Eq. 5 describe the bonded interactions in the system. Here the k's represent the varying force constants for bond lengths (b), angles (θ), improper dihedral angels (ξ) and torsions (ϕ). Bond stretching, angle bending and improper dihedral bending are described with harmonic functions relative to the reference bond length (b_0), angle (θ_0) and improper dihedral angle (ξ_0), respectively. Torsions are described in terms of a series of periodic functions with periodicity n and phase shift δ. The last two terms in Eq. 5 define the non-bonded electrostatic and van der Waals interactions in terms of the interatomic distances (r_{ij}). The electrostatic interaction energy between each atom pair is calculated according to Columb's law, where q_i and q_j are the partial charges of atoms i and j and ε_0 denotes the electric permittivity of vacuum. Finally, the van der Waals interactions between two atoms i and j are defined by the Lennard-Jones potential with corresponding atom type dependent parameters A_i, A_j, B_i and B_j where the r_{ij}^{-12} and r_{ij}^{-6} terms describe the short-range Pauli repulsion and the attractive long-range dispersion, respectively.

The parameters described above and in Eq. 5 are typically determined by calibration against results from experimental methods such as spectroscopy and crystallography and from QM calculations. In the work presented in this thesis, the OPLS-AA (35, 36) force field was applied.

3.2 Molecular Dynamics

Potential energy functions can be utilized to locate energy minima of molecules, but real biological systems are not static entities trapped in a frozen energy minimum. They are constantly changing due to thermal motions. To reflect reality computationally it is necessary that these motions are captured so that the potential energies can be related to experimentally measurable thermodynamic properties. Thus, an ensemble of thermally accessible configurations must be generated and sampled according to the Boltzmann distribution. That is, the probability P of a system being in a state with potential energy U_{pot} at a temperature T must be related as

$$P \propto \exp(-U_{pot} / k_B T) \tag{6}$$

where k_B is the Boltzmann constant. Such configurations can be obtained by moving the atoms as a function of time according to Newton's law of motion, a common method known as molecular dynamics (MD) simulations. The force acting on any atom i ($\mathbf{F}_i$) at any time t is computed from the gradient (∇_i) of the given force field potential energy function (Eq. 5). This is then used for calculating the acceleration ($\boldsymbol{a}_i$) of the atom according to Newton's second law

$$\mathbf{a}_i(t) = \frac{\mathbf{F}_i}{m_i} = -\frac{1}{m_i}\nabla_i U_{pot} \tag{7}$$

where m_i is the mass of the given atom. By using a suitable time-step, Δt, new velocities and coordinates at time $t + \Delta t$ can be approximated from truncated Taylor series expansions. In the presented work, the molecular dynamics package Q (37) was used for generating configurations for the time development of the molecular systems. Here the leap-frog Verlet MD algorithm is utilized for obtaining the velocities ($\boldsymbol{v}_i$) and positions ($\boldsymbol{r}_i$):

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \mathbf{v}_i\left(t + \frac{\Delta t}{2}\right)\Delta t \tag{8}$$

$$\mathbf{v}_i\left(t + \frac{\Delta t}{2}\right) = \mathbf{v}_i\left(t - \frac{\Delta t}{2}\right) + \mathbf{a}_i(t)\Delta t \tag{9}$$

The initial velocities (t = 0) are assigned randomly from the Maxwell-Boltzmann distribution, $P(\mathbf{v}_i)$, at a given temperature

$$P(\mathbf{v}_i) = \sqrt{\frac{m_i}{2\pi k_B T}} \cdot \exp\left(\frac{m_i \mathbf{v}_i^2}{2k_B T}\right) \tag{10}$$

Upon simulating biomolecules, the time step is usually set to 1 or 2 fs to properly sample the fastest vibrations. MD simulations were utilized to generate configurations for the molecular systems in all the included papers, **I-V**.

3.3 Statistical Mechanics

The information generated at the microscopic level, including atom's positions and velocities, from MD simulations would not have been very useful if it could not have been linked to experimentally determined macroscopic properties. Fortunately this connection is made possible through statistical mechanics (38) where a given macroscopic property is formulated in terms of an ensemble average, $\langle A \rangle$:

$$\left\langle A \right\rangle_{\text{ensemble}} = \iint A(\mathbf{p},\mathbf{r}) \rho(\mathbf{p},\mathbf{r}) d\mathbf{p}^N \, d\mathbf{r}^N \tag{11}$$

The desired property A is given as a function of momenta $\mathbf{p}_\text{i}(\text{t}), \ldots, \mathbf{p}_\text{N}(\text{t})$ and positions $\mathbf{r}_\text{i}(\text{t}), \ldots, \mathbf{r}_\text{N}(\text{t})$ of N particles at time t in the abbreviated 6N-dimensional integral above. Every configuration with momenta $\mathbf{p}$ and position $\mathbf{r}$ is weighted by the corresponding probability density ρ. In the canonical ensemble where the number of particles N, volume V and temperature T are constant, the probability density takes the form of the Boltzmann distribution

$$\rho(\mathbf{p},\mathbf{r}) = \frac{\exp\left(-\beta E(\mathbf{p},\mathbf{r})\right)}{\left(h^{3N} N!\right)^{-1} \iint \exp\left(-\beta E(\mathbf{p},\mathbf{r})\right) d\mathbf{p}^N \, d\mathbf{r}^N} \tag{12}$$

where $\beta = 1/k_B T$, h is the Planck's constant, and $N!$ is the quantum correction for indistinguishable particles, which is the general case for nearly all of matter. The denominator in Eq. 12 is known as the classical canonical partition function, Q. The $(h^{3N} N!)^{-1}$ term is a correction factor for agreement with the quantum behavior in the high temperature limit, where quantum systems behave like classic ones. The partition function can be simplified since the total energy, E, is the sum of potential (U) and kinetic (K) energies that are independent and thus separable:

$$Q(T,V,N) = \frac{1}{h^{3N} N!} \left[\int \exp\left(-\beta U(\mathbf{r})\right) d\mathbf{r}^N \right] \left[\int \exp\left(-\beta K(\mathbf{p})\right) d\mathbf{p}^N \right] \tag{13}$$

The first integral in the above equation is the configurational integral Z that only depends on the positional part of the degrees of freedom and the potential energy function. The last integral can be evaluated analytically and has the general solution

$$\int \exp\left(-\beta K(\mathbf{p})\right) d\mathbf{p}^N = \left(\frac{2\pi m}{\beta}\right)^{\frac{3N}{2}} \tag{14}$$

Thus, the total canonical partition function simplifies to

$$Q(T,V,N) = \frac{1}{N!}\left(\frac{2\pi m}{h^2\beta}\right)^{\frac{3N}{2}} \int \exp\left(-\beta U(\mathbf{r})\right) d\mathbf{r}^N = \frac{Z(T,V,N)}{\Lambda(T)^{3N} N!} \tag{15}$$

where $\Lambda(T) = (\beta h^2/2\pi m)^{1/2}$ is the thermal de Broglie wavelength. Thus, the task of evaluating a property A actually amounts to the task of evaluating the configurational partition function Z. Moreover, the fundamental principle in statistical mechanics, the ergodic hypothesis, states that the time averages for a single system equal the ensemble average. In MD simulations the time average of A is determined from

$$\left\langle A \right\rangle_{\text{time}} = \left\langle A \right\rangle_{\text{ensemble}} = \lim_{\tau\to\infty} \frac{1}{\tau} \int_{t=0}^{\infty} A\left(\mathbf{p}(t),\mathbf{r}(t)\right) dt \approx \frac{1}{M} \sum_{t=1}^{M} A\left(\mathbf{p}(t),\mathbf{r}(t)\right) \tag{16}$$

where $A(\mathbf{p}(t),\mathbf{r}(t))$ is the instantaneous value of the property A, t is the simulation time and M is the number of time steps in the simulation. One goal of MD simulations is therefore to generate enough representative conformations such that the equality in Eq. 16 is satisfied. When this is fulfilled, experimentally relevant information regarding structural, dynamic and thermodynamic properties may be calculated using a feasible amount of computer resources.

3.4 Free Energy Perturbation

MD simulations can be utilized to generate an ensemble average of thermally accessible configurations of an equilibrium state. From this average, corresponding potential energies can readily be calculated. More commonly, the interesting and desired information is relative energetics related to changes such as ligand binding or chemical modifications to the enzyme or ligand. However, the time scales involved in for example ligand binding events are extremely long compared to what is typically reached computationally. Instead of waiting for these processes to happen spontaneously, free

energy calculations can be formulated in terms of estimating the relative free energy difference, ΔG, between two related equilibrium states.

The free energy difference between two states ($\phi_1 \rightarrow \phi_2$) can be obtained from the Zwanzig's exponential formula (39):

$$\Delta G = -\beta^{-1} \ln \left\langle \exp\left(-\beta \Delta \epsilon\right) \right\rangle_{\phi_1} \tag{17}$$

$\Delta\varepsilon$ is the difference (ε_2 - ε_1) of the potential energies (e.g. ΔU_{pot} in Eq. 5) for the two states, $\beta = 1/k_B T$ and $\langle \ldots \rangle_{\phi_1}$ denotes averaging over the configurations representative of the initial state. The Zwanzig's formula thus enables free energy calculations from simulation methods by replacing the ensemble average by an average over many sampled conformations. It should be noted that Eq. 17 assumes that the configurational sampling is carried out in the isothermal-isobaric ensemble. That is, applying constant temperature (T) and pressure (P). Using the canonical ensemble (N, V, T) would instead yield the corresponding Helmholtz free energy (ΔF). For Eq. 17 to be useful It is however critical that the thermal accessible regions of the two potentials have a significant degree of overlap. Thus, the configurations sampled on the potential ϵ_1 must have at least a non-vanishing probability of also occurring on ϵ_2 and vice versa. This means that the two sampled states need to be very similar and this is generally not the case. The solution to this rather hopeless situation is the division of the free energy calculation into smaller steps.

In the free energy perturbation (FEP) scheme an alchemical transformation between two states ($\phi_1 \rightarrow \phi_2$) is driven by a linear combinations of the corresponding potential energy functions (ϵ_1 and ϵ_2) via a set of intermediate mapping potentials:

$$\epsilon_m = \lambda_m \epsilon_1 + (1 - \lambda_m)\epsilon_2 \quad ; \lambda_m \in [0,1] \tag{18}$$

In the above equation, ϵ_m denotes the effective potential energy of a particular FEP window where the coupling parameter λ_m is decremented from 1 to 0 in n steps. The resulting series of unphysical hybrids between state 1 and 2 are thus made sufficiently similar for Eq. 17 to be valid. Obviously, such hybrid molecules would never exist in reality. The fact that the free energy is a thermodynamical state function however implies that such intermediate states still can be used with confidence. Finally, the total free energy difference associated with the perturbation is calculated as a sum over the n intermediate states by modifying Eq. 17 accordingly

$$\Delta G_m = \beta^{-1} \sum_{m=0}^{n-1} \ln \left\langle \exp\left[-\beta \left(\epsilon_{m+1} - \epsilon_m \right) \right] \right\rangle_m \tag{19}$$

where $\langle \ldots \rangle_m$ denotes the average on the mapping surface ϵ_m. Combining the FEP method with a thermodynamic cycle (Figure 2) enables calculation of relative binding free energies between two similar compounds, as presented in paper **I**.

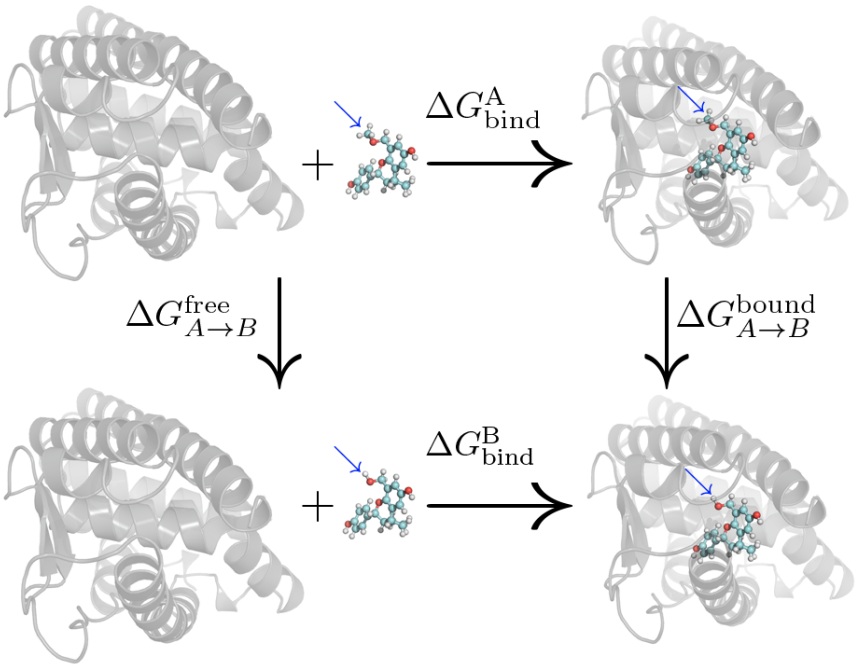


Figure 2: Thermodynamic cycle of ligands A and B binding to a protein. The horizontal paths are measured experimentally, whereas the vertical paths are calculated computationally. The blue arrow indicates the ligand region that is changing.

The FEP calculations, represented in Figure 2 by the vertical lines, are performed with the ligand transformation A → B in solution and bound to the protein. The horizontal lines represent the experimentally obtained binding free energies of A and B to the target protein. Historically FEP has been limited to situations where the differences between molecules A and B are small due to convergence problems. However, recent published works addressing this issue indicate that more complex transformations can be accomplished by introducing improved FEP schemes (40-42).

3.5 Linear Interaction Energy

Upon predicting ligand binding free energies, the FEP method presented in the previous section is often limited to situations where the difference between the two potentials of interest is small. However, several semi-empirical methods have been proposed for the

determination of absolute binding free energies, for example linear response approximation (LRA) (43, 44), molecular mechanics/Poisson-Boltzmann/surface area (MM-PBSA) (45) and the linear interaction energy (LIE) method (46). In the presented work, the LIE method has been implemented and utilized for predicting binding free energies (paper **I**). Performing two separate MD simulations, one with the solvated ligand and one with the solvated protein-ligand complex, allows the binding free energy to be estimated as

$$\Delta G_{bind} = \alpha\left(\left\langle U_{l-s}^{vdw}\right\rangle_{bound} - \left\langle U_{l-s}^{vdw}\right\rangle_{free}\right) + \beta\left(\left\langle U_{l-s}^{el}\right\rangle_{bound} - \left\langle U_{l-s}^{el}\right\rangle_{free}\right) + \gamma \qquad (20)$$

where *el* and *vdW* are electrostatic and van der Waals interactions of the ligand with surroundings (*l-s*), respectively, in water (*free*) and in the protein (*bound*). In the above equation, α and β are theoretically and empirical derived scaling constants for the polar and non-polar components of the free energy of binding, respectively. Initially, β was derived from the linear response approximation (LRA) for electrostatic forces with a value of 0.5 (43, 47). However, minor deviations from the exact LRA scaling of 0.5 have been demonstrated for hydration free energies and these are important to take into account in order to improve the accuracy (48, 49). The non-polar scaling factor, α, has been estimated using a pure empirical approach to a value of 0.18 (46). The final parameter in Eq. 20, γ, is an offset parameter usually necessary upon estimating absolute binding free energies, but not required for calculations of relative binding affinities. The parameter has been found to correlate to the hydrophobicity of the binding site (50) and could be interpreted as reflecting the energetics of water expulsion from the binding site (51).

3.6 Empirical Valence Bond

Being able to calculate free energy profiles along the reaction pathway is essential to gain a quantitative understanding of enzyme catalysis. Chemical reactions involving bond breaking and formation and their corresponding potential energy surfaces are most commonly obtained with *ab initio* QM approaches. Such methods have become quite effective in treating small molecules in gas phase (52), but quickly grow too computationally expensive for larger systems. This has to some extent been addressed by *ab initio* QM/MM methods (53, 54), but despite recent advances it is still not at a stage where these methods can be used in fully quantitative studies of enzyme catalysis

(54-57). However, using a semi-empirical QM/MM approach that is calibrated on the energetics of an appropriate reference reaction can drastically reduce the computational costs. In the presented work, the EVB method (5, 6) has been heavily utilized (papers **I-V**) as it provides probably the most effective way of simulating enzyme reactions, and also for comparing reactions in different environments.

With the EVB method (5, 6) a reactive potential energy surface is obtained by letting the reactant and product diabatic energy functions (resonance structures) interact via an empirical Hamiltonian. The diagonal elements of the Hamiltonian are described by a potential energy function somewhat similar to that in Eq. 5

$$H_{ii} = \epsilon_i = U^i_{bnd} + U^i_{ang} + U^i_{tor} + U^i_{imp} + U^i_{nb,rr} + U^i_{nb,rs} + U^i_{ss} + \alpha^{i\neq 1} \tag{21}$$

where the subscripts bnd, ang, tor, imp and nb are abbreviations for bond, angle, torsion, improper and non-bonded, whereas r and s denotes the reacting fragments and the surroundings, respectively. Here, the use of Morse rather than harmonic potentials allows the changing bonds to be broken and formed. The last term of the Hamiltonian, $\alpha^{i\neq 1}$, represents the intrinsic gas-phase energy of the given resonance structure with all fragments at infinite separation. The off-diagonal matrix element, H_{ij}, represents the quantum mechanical coupling of the states. That is, the adiabatic mixing of the diabatic free energy functions necessary for calculating the reaction free energy profiles. It can be applied as a simple constant or an exponential function

$$H_{ij} = H_{ji} = A_{ij} \exp(-\mu_{ij} \cdot r_{kl}) \tag{22}$$

where A_{ij} and μ_{ij} are fitting parameters for the two states (i and j) and r_{kl} is the distance between a specified pair of atoms (k and l). In the presented work, H_{ij} has been used simply as a constant by setting $\mu_{ij} = 0$. For a 2-state reaction the resulting Hamiltonian thus becomes:

$$\mathbf{H} = \begin{bmatrix} \epsilon_1 & H_{12} \\ H_{12} & \epsilon_2 \end{bmatrix} \tag{23}$$

The adiabatic ground-state energy (E_g) of the system can be obtained by solving the characteristic equation:

$$\left| \mathbf{H} - \mathbf{I} E_g \right| = 0 \tag{24}$$

$$\left\| \begin{bmatrix} \epsilon_1 & H_{12} \\ H_{12} & \epsilon_2 \end{bmatrix} - \begin{bmatrix} E_g & 0 \\ 0 & E_g \end{bmatrix} \right\| = 0 \tag{25}$$

$$E_g = \frac{1}{2}\left(\epsilon_1 + \epsilon_2\right) - \sqrt{\left(\epsilon_1 - \epsilon_2\right)^2 + 4H_{12}} \tag{26}$$

In similarity, the adiabatic ground-state energy and the corresponding eigenvector coefficients ($\boldsymbol{C}_g$) are obtained by solving

$$\mathbf{HC}_g = E_g \mathbf{C}_g \tag{27}$$

that for a 2-state reaction yields the general solution

$$E_g = c_1^2\epsilon_1 + c_2^2\epsilon_2 + 2c_1c_2H_{12} \tag{28}$$

The free energies are evaluated by adiabatically changing the system from one diabatic state to another. This is done as linear combinations of the corresponding potential energy functions (ϵ_1 and ϵ_2) via a set of intermediate mapping potentials, ϵ_m (see Eq. 18). The free energy, ΔG_m, of stepwise decreasing λ from 1 to 0 (Eq. 18) is then evaluated using the FEP procedure described in section 3.4 (Eq. 19). The free energy functional, $\Delta G(\Delta\epsilon_n)$, which corresponds to trajectories moving on the actual adiabatic ground state potential is finally obtained from the FEP-umbrella sampling (FEP/US) method (5, 58)

$$\Delta G(\Delta\epsilon_n) = \sum_{m \supset \Delta\epsilon_n} w_m \left(\Delta G_m - \beta^{-1} \ln \left\langle \exp - \left[\beta \left(E_g(\Delta\epsilon_n) - \epsilon_m(\Delta\epsilon_n) \right) \right] \right\rangle_m \right) / \sum_{m \supset \Delta\epsilon_n} w_m \tag{29}$$

where $\Delta\epsilon_n$ is the discretized reaction coordinate that is defined as the energy gap ($\epsilon_1 - \epsilon_2$). This reaction coordinate can be particular powerful when one is attempting to represent the entire multidimensional conformational space of the solvent by means of a single reaction coordinate (5, 59). The $w_m/\sum w_m$ term in Eq. 29 furthermore ensures that the different contributing vectors to the reaction coordinate interval are weighted proportionally to the total contribution to the respective interval.

A key feature of the EVB method with respect to enzyme catalysis is that it relies on the calibration of empirical parameters ($\alpha^{i\neq1}$ and H_{ij} in Eq. 21 and 22, respectively) for the energetics of a suitable reference reaction (e.g. $\Delta g^{\ddagger}_{\text{cage}}$ in section 2.1). Once calibrated, the same parameters are used to describe the same process in the enzyme, or any other different environment for that matter. Not only does this procedure eliminate many potential force field errors, it allows one to directly compare how the presence of a

different electrostatic environment affects the energetics of the studied reaction (see Figure 1).

3.7 Thermodynamic Activation Parameters

With the EVB method described in the previous section, activation free energies can readily be computed. However, as already discussed in the introduction and section 2.4, it is critical that also the enthalpy-entropy balance can be reproduced upon studying enzyme structure-activity relationships involved in temperature-adaptation, so that

$$\Delta G^{\ddagger} = \Delta H^{\ddagger} - T\Delta S^{\ddagger} \tag{30}$$

The activation enthalpy is in principle the difference in energy between ES‡ and ES (Figure 1) in terms of heat content given as

$$\Delta H^{\ddagger} = \Delta U^{\ddagger} + P\Delta V^{\ddagger} \tag{31}$$

where P is the pressure and V the volume. However, the pressure-volume term is completely negligible and $\Delta H^{\ddagger}$ is determined by the corresponding change in the total internal energy of the system. The latter can be decomposed into contributions from the reacting fragments (i.e., the EVB atoms whose interaction parameters change), *rr*, their interactions with the surrounding, *rs*, and the interactions within the surrounding, *ss*:

$$\Delta H^{\ddagger} = \Delta U^{\ddagger}_{rr} + \Delta U^{\ddagger}_{rs} + \Delta U^{\ddagger}_{ss} \tag{32}$$

The activation entropy, which depending on whether the degree of order is lower ($\Delta S^{\ddagger} > 0$) or higher ($\Delta S^{\ddagger} < 0$) with respect to the ground state, can then in its most simplistic terms express how easy the ES‡ is formed through

$$\Delta S^{\ddagger} = \left(\Delta H^{\ddagger} - \Delta G^{\ddagger}\right)T^{-1} \tag{33}$$

However, as the last term of Eq. 32 pertains to a huge number of interactions within the surrounding protein (p) and solvent (w)

$$\Delta U^{\ddagger}_{ss} = \Delta U^{\ddagger}_{pp} + \Delta U^{\ddagger}_{pw} + \Delta U^{\ddagger}_{ww} \tag{34}$$

the resulting energy is extremely large. Consequently, it becomes practically impossible to obtain a converged value for this quantity directly from MD simulations. Thus, obtaining reliable values for $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ through the seemingly simple relationship of Eq. 32 and 33 becomes very difficult. A solution to this hopeless situation is to calculate $\Delta G^{\ddagger}$ with sufficient high precision as a function of the inversed temperature over a suitable temperature range

$$\Delta G^{\ddagger}(T^{-1}) = \Delta H^{\ddagger} \cdot T^{-1} - \Delta S^{\ddagger} \tag{35}$$

Thus, the activation entropy and enthalpy can be obtained from simple linear regression from the corresponding Arrhenius plot so that

$$\Delta H^{\ddagger} = \frac{\sum_{i=1}^{n} \left(T_i^{-1} - \left\langle T^{-1} \right\rangle \right) \left(\Delta G_i^{\ddagger} \cdot T_i^{-1} - \left\langle \Delta G^{\ddagger} \cdot T^{-1} \right\rangle \right)}{\sum_{i=1}^{n} \left(T_i^{-1} - \left\langle T^{-1} \right\rangle \right)^2} \tag{36}$$

$$\Delta S^{\ddagger} = \left\langle \Delta G^{\ddagger} \cdot T^{-1} \right\rangle - \Delta H^{\ddagger} \left\langle T^{-1} \right\rangle \tag{37}$$

where $\langle \dots \rangle$ denotes average values.

3.8 Reorganization Energies

As discussed in section 2.3, the improved electrostatic stabilization of the transition state in enzyme catalyzed reactions compared to the solvent cage reactions originate from significantly smaller reorganization energies, λ (5, 23). The EVB diabatic free energy profiles represent microscopic equivalents of the Marcus Parabolas in electron-transfer theory (60) and can be used to obtain λ. From the diabatic free energy profiles illustrated in Figure 3, the reorganization energy corresponds to the energy required to force the reactants (Δg_1) to have the same configuration as the products (Δg_2) without allowing any charge transfer. If the diabatic free energy profiles illustrated in Figure 3 have the same curvature, we can for simplicity express the parabola for state 1 as

$$\Delta g_1 = \left(\Delta \epsilon \right)^2 \tag{38}$$

This is completely general because we have not assigned any scale to the axes. The second parabola is then just Δg_1 shifted with the diabatic free energy difference between the two parabolas, ΔG_0, and displaced along the reaction coordinate direction by $\sqrt{\lambda}$:

$$\Delta g_2 = \Delta g_1 - \Delta G_0 = (\Delta \epsilon - \sqrt{\lambda})^2 = (\Delta \epsilon)^2 - 2\Delta \epsilon \sqrt{\lambda} + \lambda \tag{39}$$

At the intersection point (transition state) we furthermore have that

$$\Delta g_1(\Delta \epsilon^{\ddagger}) = \Delta g_2(\Delta \epsilon^{\ddagger}) \tag{40}$$

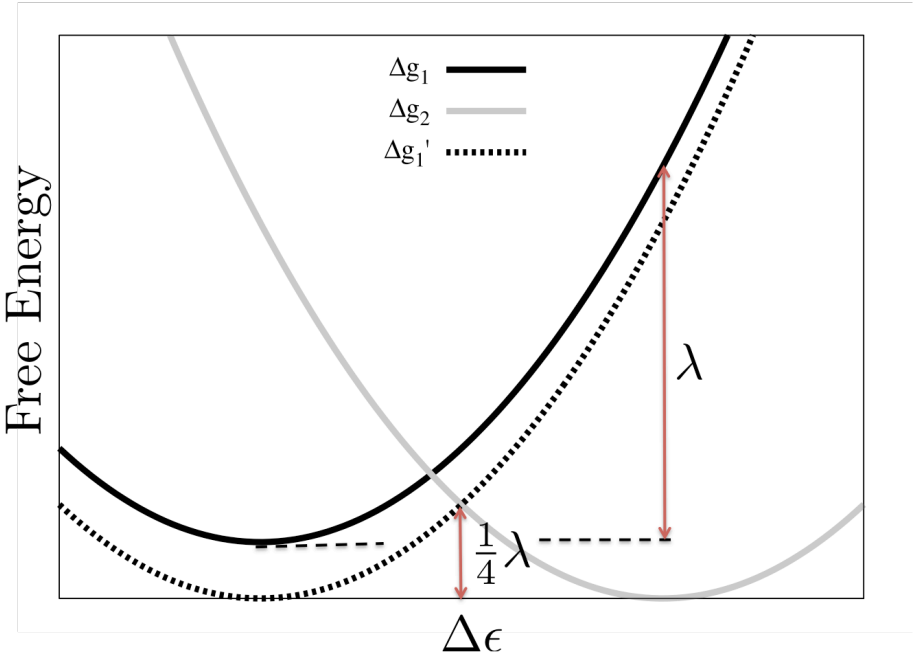


Figure 3: Two methods for obtaining the reorganization energy from the EVB diabatic free energy profiles representing microscopic equivalents of the Marcus parabolas in electron-transfer theory.

Solving Eq. 40 gives the following solution for the transition state with respect to the reaction coordinate:

$$\Delta\epsilon^{\ddagger} = \frac{\Delta G_0 + \lambda}{2\sqrt{\lambda}} \tag{41}$$

Finally, putting this result back into the original expression for Δg_1 (Eq. 38) gives:

$$\Delta g_1(\Delta\epsilon^{\ddagger}) = \frac{\left(\Delta G_0 + \lambda\right)^2}{4\lambda} \tag{42}$$

From Eq. 42 it can immediately be recognized that if the minimum of Δg_1 is shifted to the same height as Δg_2 ($\Delta g_1'$ in Figure 3), ΔG_0 becomes 0 and the reorganization energy is simply obtained as

$$\lambda = 4\Delta g_1(\Delta\epsilon^{\ddagger}) \tag{43}$$

In the work presented in this thesis, two methods have been utilized for estimating the reorganization energy (see Figure 3): The first locates λ simply as the diabatic free energy of Δg_1 at the minimum (product intermediate) of Δg_2. The second methods involves shifting the minima of Δg_1 to the same height as Δg_2 ($\Delta g_1'$ in Figure 3) and taking the resulting height of the intersection, the intrinsic barrier, as λ/4. However, as pointed out by Liu *et al.* (61), the EVB free energy profiles of the diabatic states do not correspond to perfect parabolas and the curvature between the states can be different. The reported reorganization energies in the presented works are therefore computed by

fitting the diabatic free energy profiles up to the intersection point to perfect parabolas by polynomial regression. There may also exist small variations in λ obtained from the two methods presented above, and in such cases it may be useful to take an average of the two values.

4 Development of a GUI for Free Energy Calculations (I)

In this chapter the development of Qgui (paper **I**) and its key features will be presented.

4.1 Introduction

The use of classical MD simulations to calculate free energies (Figure 4) is rapidly becoming a high-throughput technique due to advancements in algorithms together with the tremendous increase in computer power witnessed the last decade. As a consequence it has become not only more efficient to capture larger parts of phase space with respect to time, but also feasible to perform quantitative calculations on highly complex systems such as the ribosome (62). More importantly, the ability to accurately predict free energies renders it possible to characterize the structure and energetics of molecular complexes, and is often the key to understand many biological functions. High accuracy is usually obtained by conducting several replicas of the simulations with different starting conditions to ensure that statistically significant results with adequate sampling of the relevant parts of phase space are achieved. However, most MD software packages suited for these types of calculations, including Q (37) used in the presented work, are command line interfaces. The main challenge with this is management and handling of input files and the enormous amount of data generated. For example, upon studying enzyme structure-activity relationships with the EVB method (papers **I-V**) the overall process can be divided into 3 main stages as illustrated in Figure 4. Stage 1 typically involves manually assignment of atom parameters (charges, bonds, angles, torsions, impropers etc.), which depending on the number of reacting atoms quickly becomes a very tedious and time-consuming task.

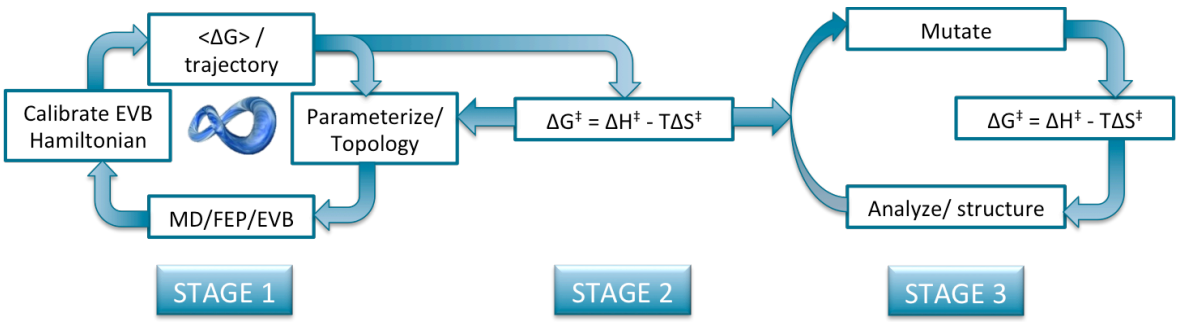


Figure 4: Flowchart illustrating the typical workflow upon doing enzyme structure activity calculations with the empirical valence bond method.

After calibrating the reference reaction(s) (manually adjusting $\alpha^{i\neq1}$ and H_{ij} in Eq. 21 and 22) and successfully reproducing the desired experimental k_{cat} of the enzyme reaction, the next stage typically involves obtaining activation enthalpies and entropies. If stage 2 succeeds, stage 3 is similar, but with mutations selected based on careful analysis of the enzyme energetics and structures obtained in stages 1 and 2. Moreover, stage 2 and each mutation in stage 3 typically involve 500 – 1000 unique EVB simulations each distributed over a suitable temperature range (~6 – 10 temperatures). Clearly, both generating input files and analyzing the resulting free energy profiles, corresponding energy components and relevant structures/trajectories is a challenging task.

In comparison to command line utilities, graphical user interfaces have the potential not only to lower the barrier for new user, but also of making the process of calculating free energies more efficient. In light of the increasing availability of computational power and the need for extensive free energy calculations in for example ligand binding and design, enzyme catalysis and mechanisms of action of biological macromolecules, we have developed a graphical user interface, Qgui, specialized for a wide array of these purposes (Figure 5).

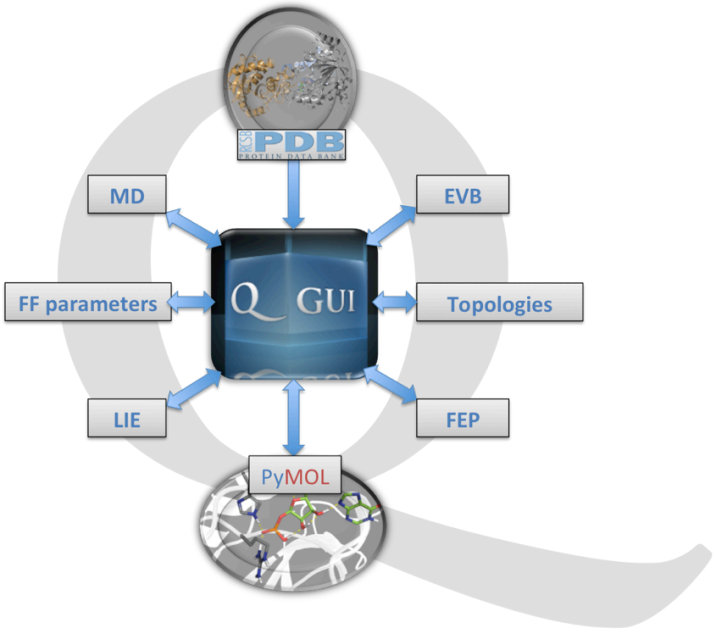


Figure 5: Qgui is a graphical user interface developed for high-throughput setup and analysis of free energy calculations and empirical valence bond simulations in biological systems using the molecular dynamics package Q.

4.2 What is Qgui?

Qgui is a graphical user interface developed for extensive free energy calculations and empirical valence bond simulations using the molecular dynamics package Q (37). It should however be noted that Qgui is not limited to Q, and can in principle be extended to function with other MD software packages. The software has been written in the Python language using object oriented design principles based on the TkInter GUI library. A key feature of the interface is gluing together all the tasks involved in the various setups and analyzes of MD simulations and free energy calculations as illustrated in Figure 5. In addition to communicating with Q, the software also (optionally) communicates with any version of the Open Source PyMOL (Schrödinger, LCC, New York), which is freely available, for visualization and manipulation of 3D structures and the ffld_server utility that comes with Maestro (Schrödinger, LCC, New York) for force field parameter assignment. PDB files can also be imported directly from the RCSB Protein Data Bank (http://www.pdb.org). From a simplistic main window that is always present and functioning as a feedback/logging window, all the available tools are easily accessed from four dropdown menus (Figure 6).

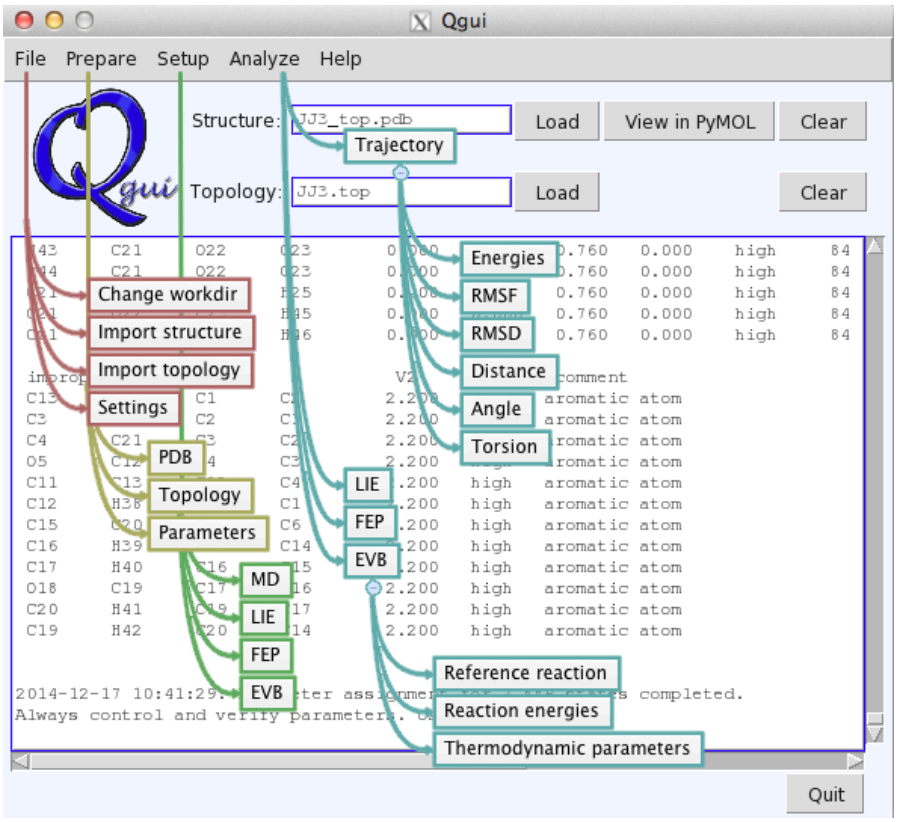


Figure 6: Qgui main window with illustrated overview of dropdown menus and their contents.

4.3 Key Features

Qgui is designed for efficient preparation of necessary input files and analysis of the resulting output files from MD, LIE, FEP and EVB (Figure 5). Much effort has been put into automating typical time consuming tasks such as generation of force field parameter and topology files, preparation of MD equilibration procedures and calibration of EVB reference reactions, to mention a few. The software has particularly aimed at enabling efficient calculations of thermodynamic activation parameters as illustrated in Figure 4. More importantly, Qgui is not a black box. That is, all settings and parameters are visible and editable for the user at any time. In the following, some key features of the software presented in paper **I** will be summarized.

Visualize, click and auto-generate force field parameters

Defining state dependent force field parameters in the context of changing atoms in FEP or bond breaking and formation in EVB is efficiently handled in Qgui. It automatically maps all parameters for each defined state to the corresponding atom numbers in the topology. In addition, syncing with PyMOL allows different states in FEP or EVB to be defined and visualized simply by clicking on the atom(s) of interest. This is particularly convenient because it is both fast and the different states can be visualized together with their parameters (charges, bonds, angles, torsions etc.), making it easy to verify that the prepared simulation is going to be as intended. After defining the desired FEP or EVB states, force field parameters can furthermore be auto-generated and assigned based on template structures generated through PyMOL utilizing the ffld_server utility. In Figure 7, the auto-generation scheme is illustrated for a 2-state EVB, but 3- and 4-state EVB/FEP are also implemented. Parameter and library files can also efficiently be generated independently using the 'Prepare Parameters' tool (Figure 6), which is typically done prior to MD or LIE. Once generated, they can be loaded and organized through the settings window (Figure 6) for future use. Thus, there is no need to specify parameter and library files again when preparing simulations since they are globally defined in the settings. Moreover, there is no limit to how many files that can be loaded.

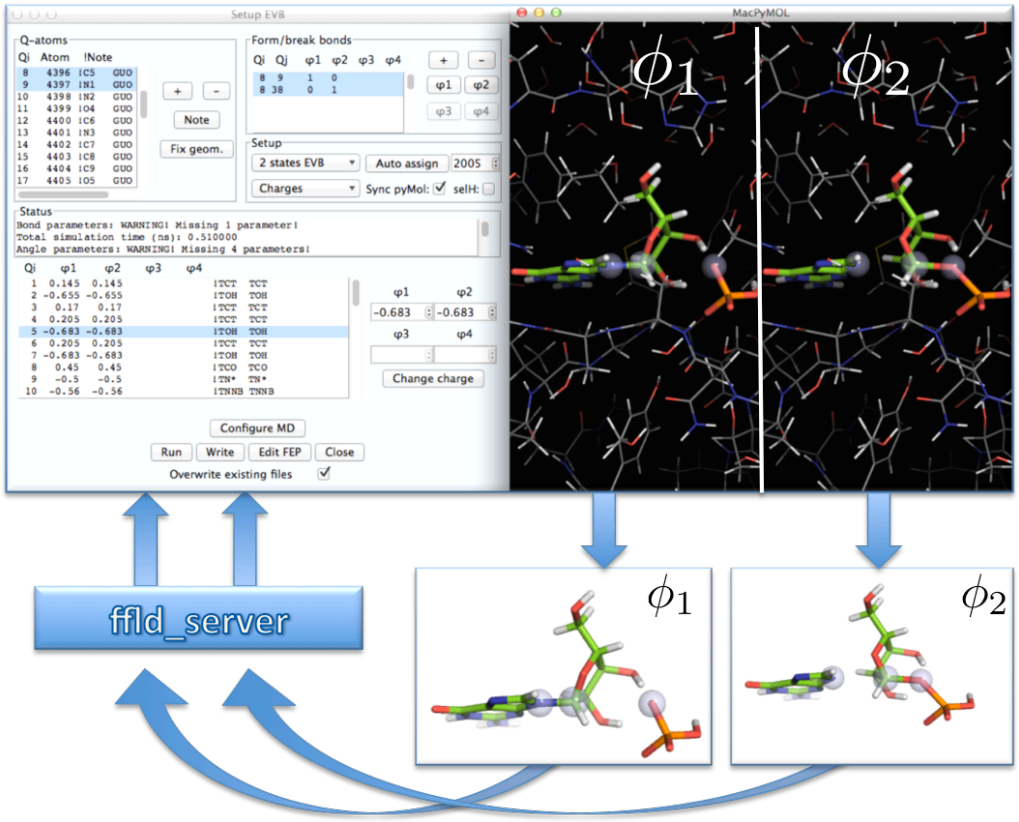


Figure 7: Auto-generation of state dependent force field parameters from PyMOL template structures utilizing the ffld_server illustrated for the Qgui EVB setup using 2 states.

Intuitive fitting and visualization of predictive LIE models

The LIE method is implemented in Qgui and approximates ligand binding free energies as described in section 3.5 through Eq. 20. Normally this involves applying α = 0.18 and assigning β as described by Hansson *et al.* (63). The γ parameter is system dependent and usually fitted so that the calculated absolute binding free energies coincide with those experimentally obtained. We have implemented an automatic scheme for generating predictive LIE models by combining available experimental binding free energies with the MD energy terms. In addition to the γ parameter, α and β can also be optimized, which in some cases can be useful to generate predictive models. The optimal parameter(s) are obtained by mapping the sum of squared errors (SSE) between the calculated and the experimental binding free energies until sufficient information exist to compute the polynomial regression on the form:

$$\mathrm{SSE} = b_2 x^2 + b_1 x + b \qquad ;x \in [\alpha, \beta, \gamma] \tag{44}$$

The optimal parameter value minimizing SSE is then obtained simply by solving

$$\frac{\delta \mathrm{SSE}}{\delta x} = 0 \qquad ; x \in [\alpha, \beta, \gamma] \tag{45}$$

The resulting LIE parameter(s) from this approach will ensure that the SSE is at its minimum. The fitting of LIE binding data for five ligands bound to the estrogen receptor α presented in paper **I** is illustrated in Figure 8 with the Qgui LIE window to the left and the SSE function exemplified to the right.

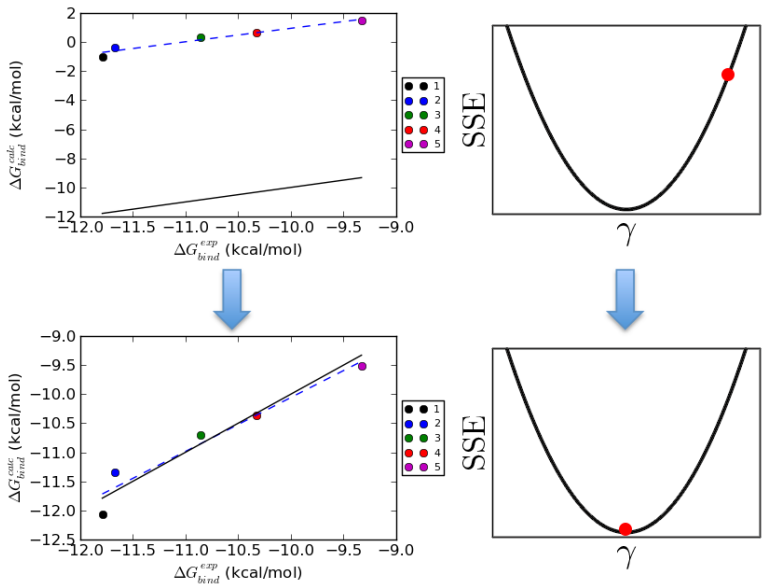


Figure 8: Fitting of the LIE binding data for five ligands bound to the estrogen receptor with the unfitted data points (top) and the data points fitted to experimental values (bottom). The blue dotted line indicates the optimal linear regression line to the calculated binding free energies independently of the experimental values. The solid black line is the experimental binding free energies. The SSE plots (top and bottom right) illustrate how SSE is minimized to optimize the LIE model (not visible in the Qgui window).

The fit LIE parameters tool (Figure 8) allows for visualization of the calculated and experimental data in real time together with the corresponding linear regression lines. The models can be generated by manually or automatically adjusting the parameters. Moreover, visualizing both the linear regression line for the predicted data points independently of the experimental data (blue dotted) together with the experimental line (black solid) makes it easy to validate the predictive power of the model outside the computed region (Figure 8). It can easily be recognized from Figure 8 that the model generated using ligands 1-5 is excellent in the binding free energy region. However, the minor difference in slopes indicate that the predictive power will become weaker when computing ligands outside the binding range used here with more than ~4 kcal/mol.

Automatic calibration of EVB reference reactions

A prerequisite for doing successful reaction dynamics with the EVB method described in section 3.6 is that a free energy surface for a suitable reference reaction is fitted to existing experimental or *ab initio* data. This is typically done by manually adjusting $\alpha^{i\neq1}$ and H_{ij} (Eq. 21 and 22) until the target activation and reaction free energies are reproduced. The resulting parameters often vary between individual runs and it is thus desirable to run several reference reactions and use the more reliable average obtained from these. This is however a tedious and time-consuming task. We have therefore developed and implemented a protocol for automatic reference reaction calibration (RRC) in Qgui (Figure 9). It is possible to calibrate single runs individually or several runs merged together.

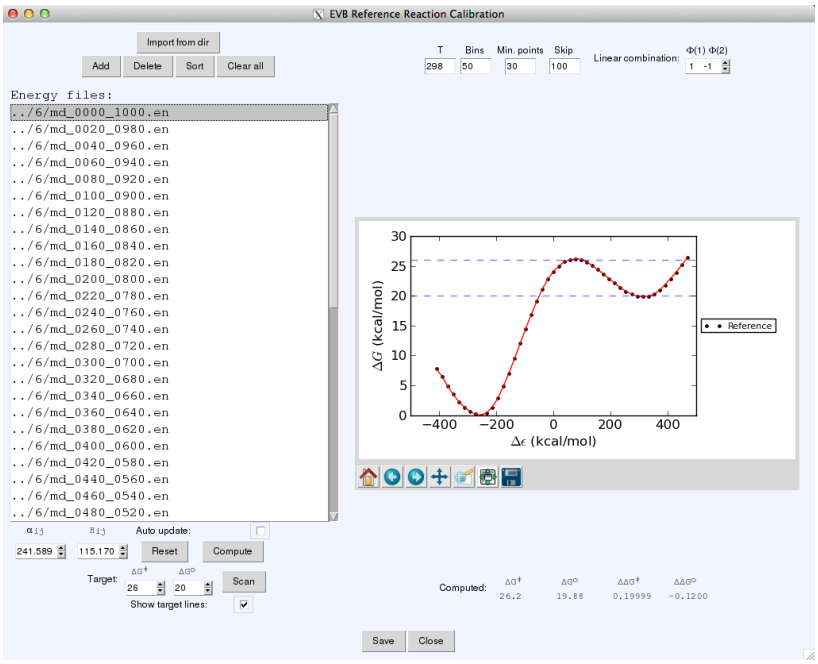


Figure 9: EVB reference reaction calibration (RRC) tool in Qgui.

In summary, the target activation and reaction free energies (blue dotted lines in Figure 9) are efficiently located by taking advantage of the approximate linear relationship between H_{ij} and $\alpha^{i\neq1}$ against $\Delta G^{\ddagger}$ and ΔG_0, respectively. Thus by mapping Eq. 29 iteratively with $\alpha^{i\neq1}$ constant varying H_{ij} and then with H_{ij} constant varying $\alpha^{i\neq1}$, the parameters corresponding to the target values are approached by standard linear regression:

$$\Delta G^{\ddagger}(H_{ij}) = \beta_0 + \beta_1 H_{ij} \tag{46}$$

$$\Delta G_0(\alpha^{i\neq1}) = \beta_0 + \beta_1 \alpha^{i\neq1} \tag{47}$$

Two iterations utilizing the above equations, applying updated parameters in each round, typically brings the reaction free energy profile very close to the target values. If necessary, a fine-tuning stage is automatically launched where the parameters are adjusted in an iterative manner until the target values are within a threshold (ΔΔG = 0.2 by default). The reaction free energy profile is updated in real time during the automatic calibration procedure, which also can be performed by manually adjusting the parameters. In addition, visualizing the reaction free energy profiles is an important tool for revealing potential sampling problems along the reaction coordinate, which typically results in discontinuous points or uneven spacing between the energy gaps.

Thermodynamic activation parameters

Qgui is written with a particular focus on automating the task of obtaining thermodynamic activation parameters ($\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$) from EVB simulations. Specialized setup and analyze tools automatically generating and recognizing directory hierarchies consisting of a range of temperatures with a series of subdirectories have been implemented (Figure 10).

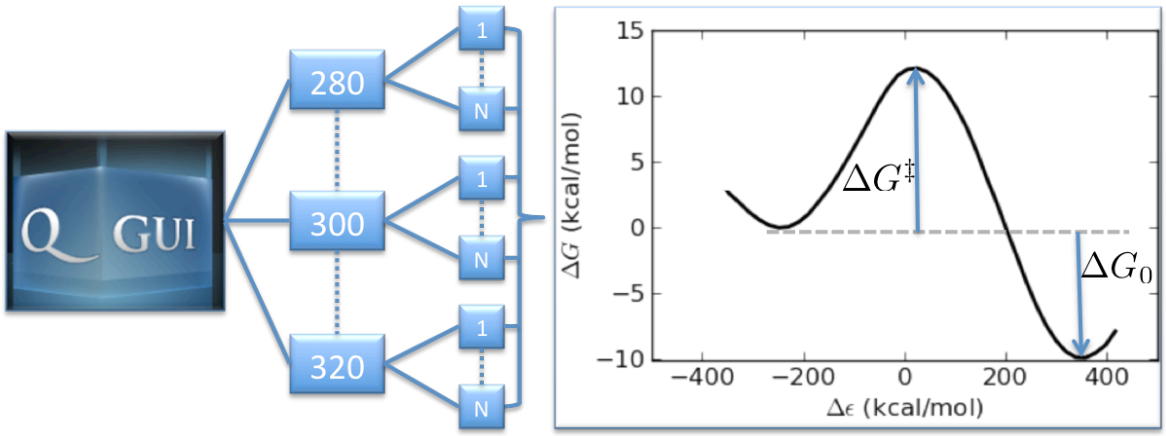


Figure 10: Illustration of how Qgui organizes submission and analysis of EVB simulations over range of temperatures with several runs for each temperature.

The analyze EVB thermodynamic parameters tool (Figure 6) efficiently goes through all the directories computing reaction free energy profiles and locating the transition and product state free energies as illustrated in Figure 10. The thermodynamic activation parameters are then obtained by computing the activation free energy as a function of the inversed temperature as described in section 3.7. The resulting Arrhenius plots together with the corresponding regression statistics can be visualized and modified in Qgui. In addition, we have also implemented tools for obtaining the EVB activation

potential energy terms ($\Delta U^{\ddagger}$) given in Eq. 21. This is particularly useful for efficiently identifying the energy term(s) contributing to the catalytic effect. Moreover, combined with the activation enthalpy it gives a way to obtain the EVB activation potential energy for the interactions within the surroundings (Eq. 34 in section 3.7) as:

$$\Delta U_{ss}^{\ddagger} = \Delta H^{\ddagger} - \Delta U_{rr+rs}^{\ddagger} \tag{48}$$

However, computing $\Delta U^{\ddagger}$ from the corresponding lambda windows' MD energies is not as straightforward as for the Arrhenius plot thermodynamic activation parameters. Locating the energy gaps corresponding to the reactant and transition state is trivial, but every energy gap in the reaction coordinate typically originates from several lambda windows with different weights. For example, the energy gap corresponding to the transition state ($\Delta\epsilon^{\ddagger}$) illustrated in Figure 11 originates from a total of 5 lambda windows (0.44 – 0.52) with different weights ranging from 0.05 – 0.35.

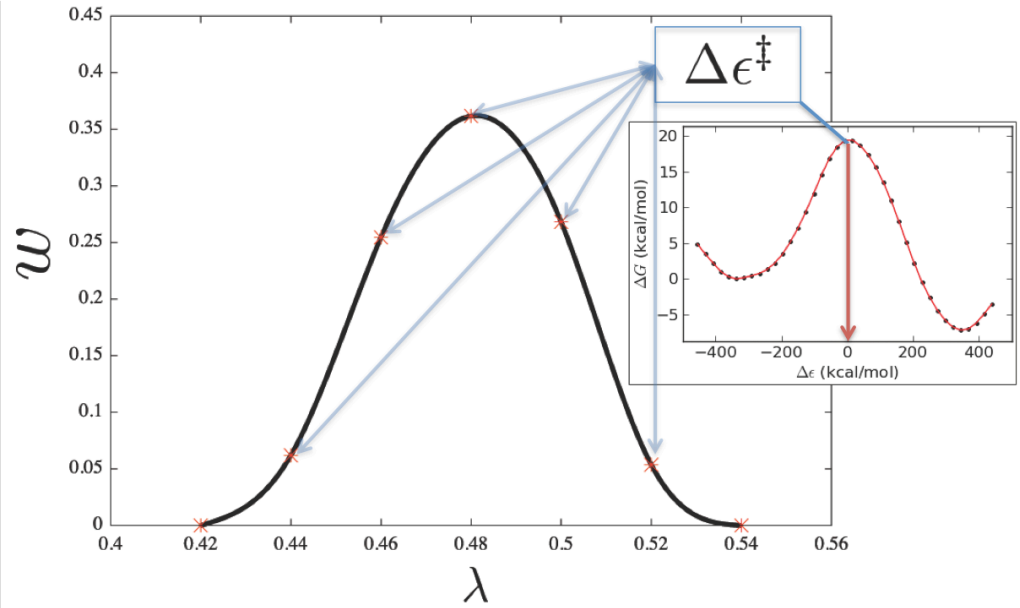


Figure 11: Several lambda windows (λ) contribute with different weights (w) to the transition state energy gap ($\Delta\epsilon^{\ddagger}$).

Qgui computes the exact EVB potential activation energy terms by collecting all lambda windows (λ_m) contributing to the energy gaps corresponding to the transition and reactant state. The energies originating from each lambda value are then scaled by w_i according to the actual state contribution so that

$$U = \sum_i w_i \left(U_{1,i} + U_{2,i} \right) \quad ; i \in \lambda_m \tag{49}$$

where $\sum_i w_i = 1$. From the general solution of the secular equation given in Eq. 28 for a 2-state reaction, the potential activation energy within the reacting region and to the surroundings, $\Delta U_{rr+rs}^{\ddagger}$, is then calculated exactly as

$$\Delta U^{\ddagger}_{rr+rs} = \Delta\left(c_1^2 U_{1,rr+rs}\right)^{\ddagger} + \Delta\left(c_2^2 U_{2,rr+rs}\right)^{\ddagger} + \Delta\left(c_2^2 \alpha_2\right)^{\ddagger} + 2\Delta\left(c_1 c_2 H_{12}\right)^{\ddagger} \quad (50)$$

where all terms are scaled according to Eq. 49. Together with the activation enthalpy, the relationship of Eq. 48 can be used to evaluate the interactions within the surroundings, U_{ss}, which is not obtained directly from the EVB calculations due to slow convergence. The reaction free energy profiles with the corresponding EVB potential activation terms are easily visualized and compared using the EVB reaction energies tool (Figure 6).

4.4 Qgui in the Future

As the field of computational chemistry is continuously advancing and changing for better reflection of the reality, Qgui can obviously not be considered as completed. There are always new features that can and should be implemented and existing ones that could be improved. The development of Qgui has been, and still is, a continuous project where every new molecular system investigated reveals features that would be convenient to implement. Most of the initial code was developed simultaneously with the study presented in paper **II** and many additional features and improvements have been implemented during the works presented in papers **III-V**. The list of potential things to implement and improve is growing almost exponentially. However, some examples include implementing automatic generation of force field parameter and library files for other force fields than OPLS, which is currently the only option for the automatic assignment scheme in Qgui. We are also planning to implement alternatives to the Zwanzig's formula for FEP calculations, such as BAR (64) and WHAM (65). It would furthermore be interesting to investigate the possibility for automatic calibration of EVB reference reactions applying more than 2 states, which is the current limit for the EVB RRC tool.

5 Trypsin Cold-Adaptation (II and III)

In this chapter the key findings for the adaptive mechanisms responsible for trypsin cold-adaption presented in papers **II** and **III** will be summarized. The reader is referred to section 2.4 for a general introduction to the topic of cold-adaptation.

5.1 Arrhenius Plots Reveal Characteristics of Cold-Adaptation

Trypsin belongs to the serine protease class of enzymes that catalyze the cleavage of peptide bonds in proteins and polypeptides. This class of enzymes has been extensively studied for many decades and is in this respect very well suited for generating reliable model systems. As an approach to investigate key aspects of enzyme cold-adaption (section 2.4), we have in the works presented in papers **II** and **III** modeled the rate-limiting step of the peptide bond cleavage reaction, the formation of the tetrahedral intermediate (66-68), in psychrophilic anionic salmon trypsin (AST) and mesophilic bovine trypsin (BT) utilizing the MD/EVB method (Figure 12). The imidazole-catalyzed methanolysis of formamide in water was used as a reference reaction and fitted to activation and reaction free energies of 26 and 20 kcal/mol, respectively (69). The results from the simulations at 300 K are shown in Figure 12 as free energy profiles along the reaction coordinate for the two enzymes together with the reference reaction.

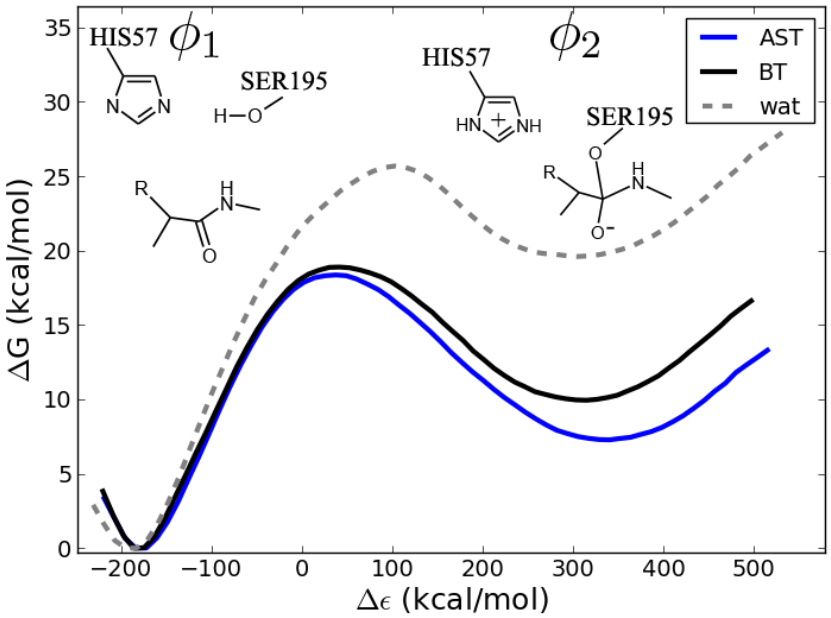


Figure 12: Calculated free energy profiles (300 K) for tetrahedral intermediate formation in the acylation step of anionic salmon trypsin (AST), bovine trypsin (BT) and in the imidazole catalyzed reference reaction in water (wat).

The EVB simulations yield activation energies at 300 K of 18.2 ± 0.2 kcal/mol and 19.0 ± 0.2 kcal/mol for the Cys-Lys-Ala (CKA) substrate in AST and BT, respectively, and are in excellent agreement with the substrate dependent barrier of 15 – 20 kcal/mol (66). This difference in activation free energies corresponds to a 4-fold increase in k_{cat} for AST and is in remarkable good agreement with experiments that shows a 2- to 4-fold increase in k_{cat} depending on the temperature (70). Since the catalytic rates of AST and BT are well reproduced at room temperature, stage 1 illustrated in Figure 4 is completed and it is possible to move on to stage 2 examining the actual temperature dependence.

The thermodynamic activation parameters were obtained by computing 100 – 150 independent reaction free energy profiles at each temperature in a range of eight temperatures from 275 K to 310 K (see section 3.7). From the resulting Arrhenius plots in Figure 13 it can immediately be seen that the cold-adapted trypsin (AST) has a significantly smaller slope compared to the mesophilic counterpart (BT).

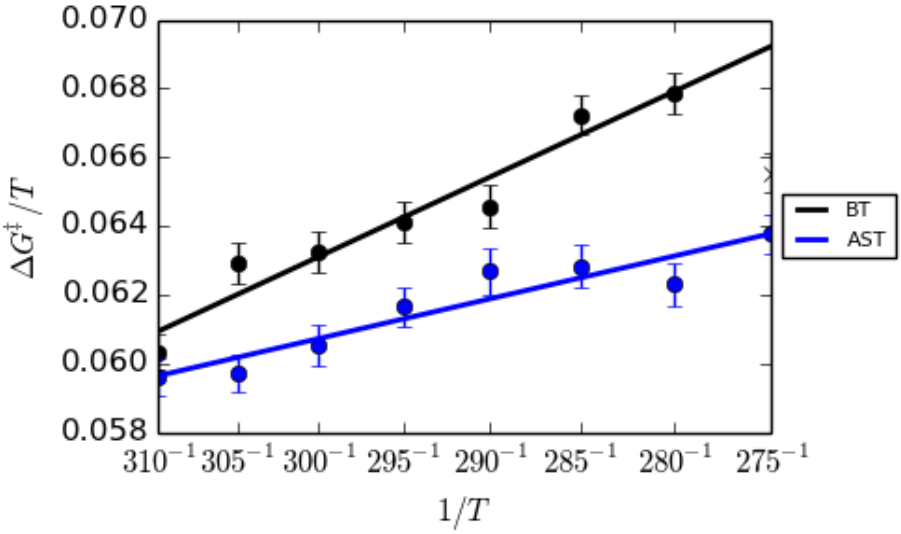


Figure 13: The clear difference in slopes from calculated Arrhenius plots for BT (black) and AST (blue) demonstrate that cold adapted trypsin has a lower activation enthalpy compared to the warm-active ortholog.

The calculated activation parameters at 300 K are $\Delta H^{\ddagger}$ = 20.4 kcal/mol and $T\Delta S^{\ddagger}$ = 1.4 kcal/mol for BT, while the corresponding values for AST are $\Delta H^{\ddagger}$ = 9.9 kcal/mol and $T\Delta S^{\ddagger}$ = -8.3 kcal/mol. This is thus a remarkable example of enthalpy-entropy compensation where the large differences in $\Delta H^{\ddagger}$ are balanced by $T\Delta S^{\ddagger}$ to yield similar activation free energies. It is also noteworthy that computer simulations reproduce both the absolute rates at 300 K and the characteristic balance between activation enthalpy and entropy for mesophilic and psychrophilic enzymes. Thus, with both stages 1 and 2 in Figure 4 successfully completed, the controversial question regarding what the origin of this effect really is remains to be answered.

5.2 Protein Surface Softness Regulates Trypsin Cold-Adaptation

As discussed in section 2.2, resolving the origin of the difference in the thermodynamic activation parameters renders it necessary to dissect the observed energies into individual contributions. The source of the difference in activation enthalpies is from the relationship given in Eq. 32, however, relatively straightforward to identify. Table 1 shows the breakdown of the energetics (Eq. 32) where $\Delta U^{\ddagger}_{rr+rs}$ is calculated as in Eq. 50.

Table 1: Calculated thermodynamic activation parameters for native and mutant bovine (BT) and salmon (AST) trypsin at 300 K.

Enzyme	$\Delta G^{\ddagger}$	$\Delta H^{\ddagger}$	$T\Delta S^{\ddagger}$	$\Delta U^{\ddagger}_{rs+rr}$ [a]	$\Delta U^{\ddagger}_{ss}$ [a]
BT_{native}	19.0 ± 1.4	20.4 ± 1.0	1.4 ± 1.0	14.6 ± 0.7	5.8 ± 1.3
BT_{S150D}	18.6 ± 1.3	14.5 ± 0.9	-4.1 ± 1.0	14.9 ± 1.1	-0.4 ± 1.5
BT_{N97Y}	18.4 ± 1.3	10.6 ± 0.9	-7.8 ± 1.0	15.9 ± 1.1	-5.3 ± 1.5
AST_{native}	18.2 ± 0.8	9.9 ± 0.6	-8.3 ± 0.6	13.1 ± 0.9	-3.2 ± 1.1
AST_{D150S}	18.0 ± 1.1	14.7 ± 0.8	-3.2 ± 0.8	11.5 ± 1.9	3.2 ± 1.2
AST_{Y97N}	18.4 ± 1.1	12.0 ± 0.8	-6.4 ± 0.8	13.9 ± 1.2	-1.9 ± 1.4
AST_{Y97_5}	18.1 ± 0.9	12.2 ± 0.6	-5.9 ± 0.6	12.3 ± 1.1	-0.1 ± 1.3
$AST_{R18\_0.1}$	17.5 ± 1.1	12.6 ± 0.8	-4.9 ± 0.8	12.2 ± 1.7	0.4 ± 1.9
$AST_{R18\_0.5}$	17.5 ± 1.0	15.6 ± 0.7	-1.9 ± 0.7	13.1 ± 2.1	2.5 ± 2.2
AST_{R18_1}	18.0 ± 1.9	21.9 ± 1.3	3.9 ± 1.3	12.8 ± 1.9	9.1 ± 2.3
AST_{R18_5}	17.8 ± 1.7	20.6 ± 1.2	2.8 ± 1.2	12.9 ± 2.4	7.7 ± 2.7
AST_{R18_100}	17.8 ± 1.1	19.5 ± 0.8	1.6 ± 0.8	13.7 ± 1.9	5.8 ± 2.1

[a] Subscripts rr, rs and ss denote, respectively, interactions among atoms in the EVB region, their interactions with the surroundings, and the internal interactions within the surroundings. Error bars denote standard deviations of the mean.

It can immediately be seen from Table 1 that the source of the decreased activation enthalpy in AST is not associated with a more favorable $\Delta U^{\ddagger}_{rr+rs}$, as could have been expected, but rather a significantly lower $\Delta U^{\ddagger}_{ss}$. That is, the contribution from the surroundings (Eq. 34) is predicted to be about 9 kcal/mol more favorable for AST than BT, whereas the internal energy change involving the reacting groups is similar. Thus, at any rate we can conclude that the reduction of the activation enthalpy originates from interactions outside the active site (mainly protein – water and protein – protein). From an evolutionary point of view, this is perhaps not so strange since the active site residues surrounding the substrate are conserved between the two proteins. The fact that $\Delta U^{\ddagger}_{ss}$ is lower in AST compared to BT, furthermore suggest that the surroundings of the active site are effectively softer in the cold-adapted enzyme. This was further demonstrated by

computing the enzyme backbone root-mean-square fluctuations (RMSF) averaged over 100 ns simulation time in the reactant and transition state for AST and BT (paper **II**). In similarity to a previous work on different temperature adapted citrate synthases by Bjelic *et al.* (34), we found that the active site and substrate mobilities were virtually identical for cold- and warm-adapted trypsin. Note that these findings are a direct contrast to the assumption that the more negative activation entropy observed for cold-adapted enzymes could be interpreted in terms of increased flexibility of the active site (32). Instead, it was found that both enzymes have a rather hard core that becomes softer towards the surface (Figure 14) where AST contains regions that are significantly softer compared to BT.

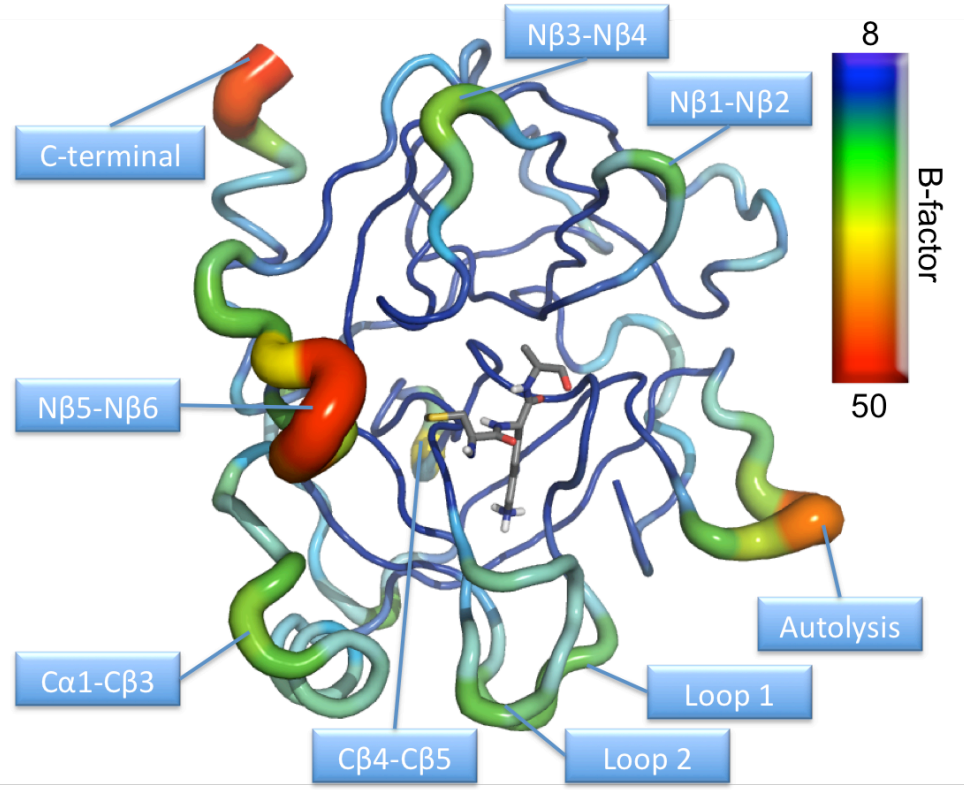


Figure 14: B-factor representation of anionic salmon trypsin averaged over 100 ns MD simulation at the reactant state illustrating that the enzyme core is rather rigid, whereas the surface contains regions with high flexibility.

Interestingly, from the sequence alignment of 7 different temperature adapted trypsins presented in paper **II**, the strictly conserved residues within cold-adapted trypsins are typically located around the flexible surface regions illustrated in Figure 14. In particular, the Nβ5-Nβ6 and the autolysis loops were found to be significantly more flexible in AST compared to BT. The direct effect of the conserved residues 97 and 150 located in the Nβ5-Nβ6 and the autolysis loops, respectively, to the thermodynamic activation parameters were investigated by point mutations. The resulting calculated

activation free energies remain essentially unchanged by the mutations (Table 1), which underscores the general notion that mutations far away from the active site do not significantly affect the catalytic rates (71). As can be seen from Table 1, both mutations in warm-adapted trypsin (BT_{N97Y} and BT_{S150D}) render the entropy-enthalpy balance remarkably similar to that of the cold-adapted enzyme. BT_{N97Y} was predicted with thermodynamic activation parameters virtually identical to those of AST with $\Delta H^{\ddagger}$ = 10.6 kcal/mol and $T\Delta S^{\ddagger}$ = -7.8 kcal/mol. Similar to the cold-adapted trypsin, it is the lowering of the interactions within the surrounding that is the origin of the decrease in the activation enthalpy. Moreover, the backbone RMSF profile averaged over 100 ns in the reactant state reveal a significant increase in mobility of the Nβ5-Nβ6 loop for the mutant compared to the native enzyme as illustrated in Figure 15.

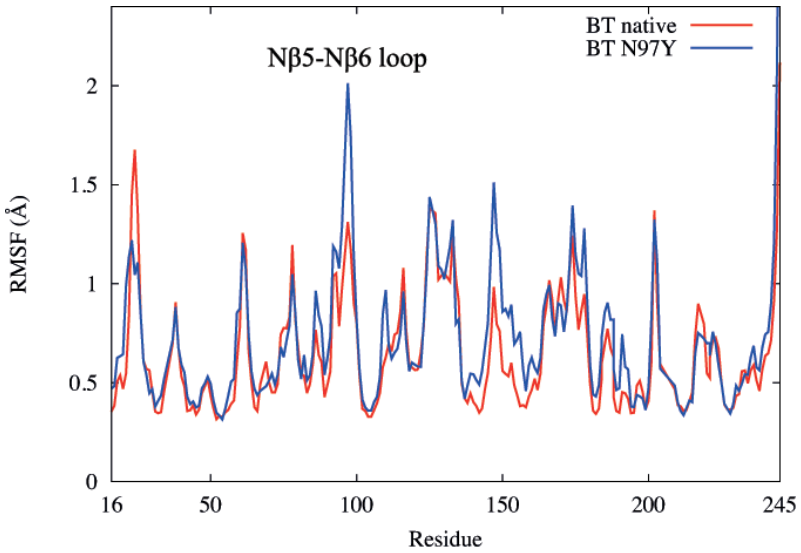


Figure 15: Average backbone RMSF profiles in the reactant state for native bovine trypsin and the N97Y mutant.

A very similar result was also observed in the autolysis loop for BT_{S150D} with predicted $\Delta H^{\ddagger}$ = 14.5 kcal/mol and $T\Delta S^{\ddagger}$ = -4.1 kcal/mol. In both mutations, it is the lowering of $\Delta U_{ss}^{\ddagger}$ that is responsible for the remarkable enthalpy-entropy balancing rendering the warm-adapted trypsin cold-adapted like. The origin to this reduction is a softer protein surface. For the reverse mutations AST_{Y97N} and AST_{D150S}, the thermodynamic activation parameters are rendered more mesophilic like. However, the AST_{Y97N} mutation yields relatively smaller effects on both the activation parameters and $\Delta U_{ss}^{\ddagger}$. This probably just reflects that it easier to disrupt the H-bonding networks rigidifying local parts of the enzyme as in BT_{N97Y}, whereas correlated mutations may be needed to build up the same network, so that a single-point mutation does not suffice.

As an alternative to the complex task of succeeding with correlated mutations, we have in paper **III** calculated the thermodynamic activation parameters upon restraining the mobility of the surface atoms in AST. Restraining the backbone of residues 95 – 99 in Nβ5-Nβ6 loop to their initial positions (AST_{97_5}) resulted in thermodynamic activation parameters and $\Delta U_{ss}^{\ddagger}$ virtually identical to AST_{Y97N}. This clearly demonstrates that the mutational effects predicted for AST_{Y97N} originates from altered surface softness (Table 1). To further test our hypothesis, we performed a series of simulations restraining all protein atoms outside a given radius centered on the AST active site. Seven different radii in the range 18 – 35 Å were examined, and the thermodynamic activation parameters were extracted from each radius. Upon decreasing the radius, both the enthalpy-entropy balance and $\Delta U_{ss}^{\ddagger}$ were gradually rendered more warm-adapted like without altering $\Delta G^{\ddagger}$. Moreover, at 18 Å (AST_{R18_100}) the calculated thermodynamic activation parameters were virtually identical to that of warm-adapted trypsin. However, restraining the mobility of atoms closer to the active site than 18 Å resulted in altered $\Delta G^{\ddagger}$ values. This shows that when the restraints on the enzyme start to affect the interior parts closer to the active site the reaction free energy profile indeed changes, as expected. We also run a series of simulations increasing the force constant from 0 to 100 kcal/mol/Å^2 for the protein atoms outside a sphere of 18 Å as illustrated in Figure 16.

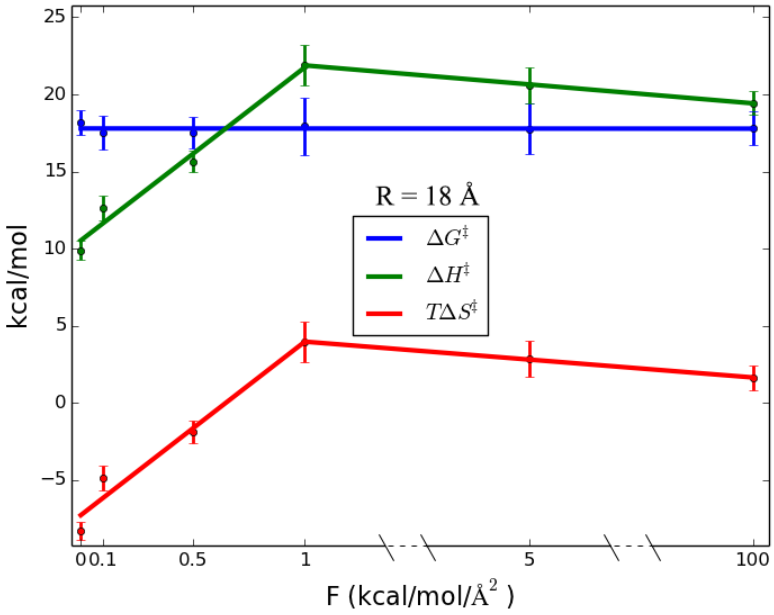


Figure 16: Thermodynamic activation parameters obtained by restraining all protein atoms outside a spherical boundary with radius 18 Å centered around the active site to their initial positions with force constants, F, of 0 - 100 kcal/mol/Å^2.

Restraining the atoms with a force constant of 1 kcal/mol/Å^2 is sufficient to completely render the thermodynamic activation parameters warm-adapted like (Table 1). As can be seen from Figure 16, there is a close to perfect linear correlation upon increasing the force constant from 0 to 1 kcal/mol/Å^2 with respect to transforming the enthalpy-entropy balance from cold- to warm-adapted in AST. What is rather remarkable is that the calculations predict significant changes to $\Delta S^{\ddagger}$ and $\Delta H^{\ddagger}$, but these are again nearly perfect compensating. This is to our knowledge the first time computational evidence is presented showing that protein surface softness controls trypsin cold-adaptation. It should also be noted that presented data in papers **II** and **III** is based on analysis from a tremendous amount of simulations and unique reaction free energy profiles. The data presented in Table 1 alone involves the analysis of approximately 10 000 reaction free energy profiles distributed over 5 μs simulation time, which roughly amounts to 5 million CPU hours.

5.3 How can Point Mutations Soften the Protein Surface?

With the potential of some time being able to successfully design enzymes, it is obviously important that we understand the mechanisms involved in temperature adaption. The works presented in papers **II** and **III** strongly suggest that the surface of cold-adapted enzymes is softer compared to their warm-adapted counterpart. From a mutational point of view, softening the enzyme surface is in light of the above discussion seemingly easy. However, the key question that remains to be answered is how this can be accomplished. In this respect, ultra-high resolution (0.75 – 1.00 Å) crystal structures of both BT (72) and AST (73) provide useful clues since a large numbers of surface bound water molecules are resolved in these structures. The BT_{N97Y} mutation, for example, effectively renders the thermodynamic activation parameters virtually identically to those of AST. In this case, the surface is softened through the mutation of a polar to a less polar residue, which effectively disrupts the water mediated H-bond network as illustrated in Figure 17. The similar result obtained for BT_{S150D} on the other hand originates from a complete change of H-bonding network, which also significantly alters the entire autolysis loop conformation. This mechanism is however achieved by correlated mutations (T21E and V154K) in addition to S150D (Figure 17).

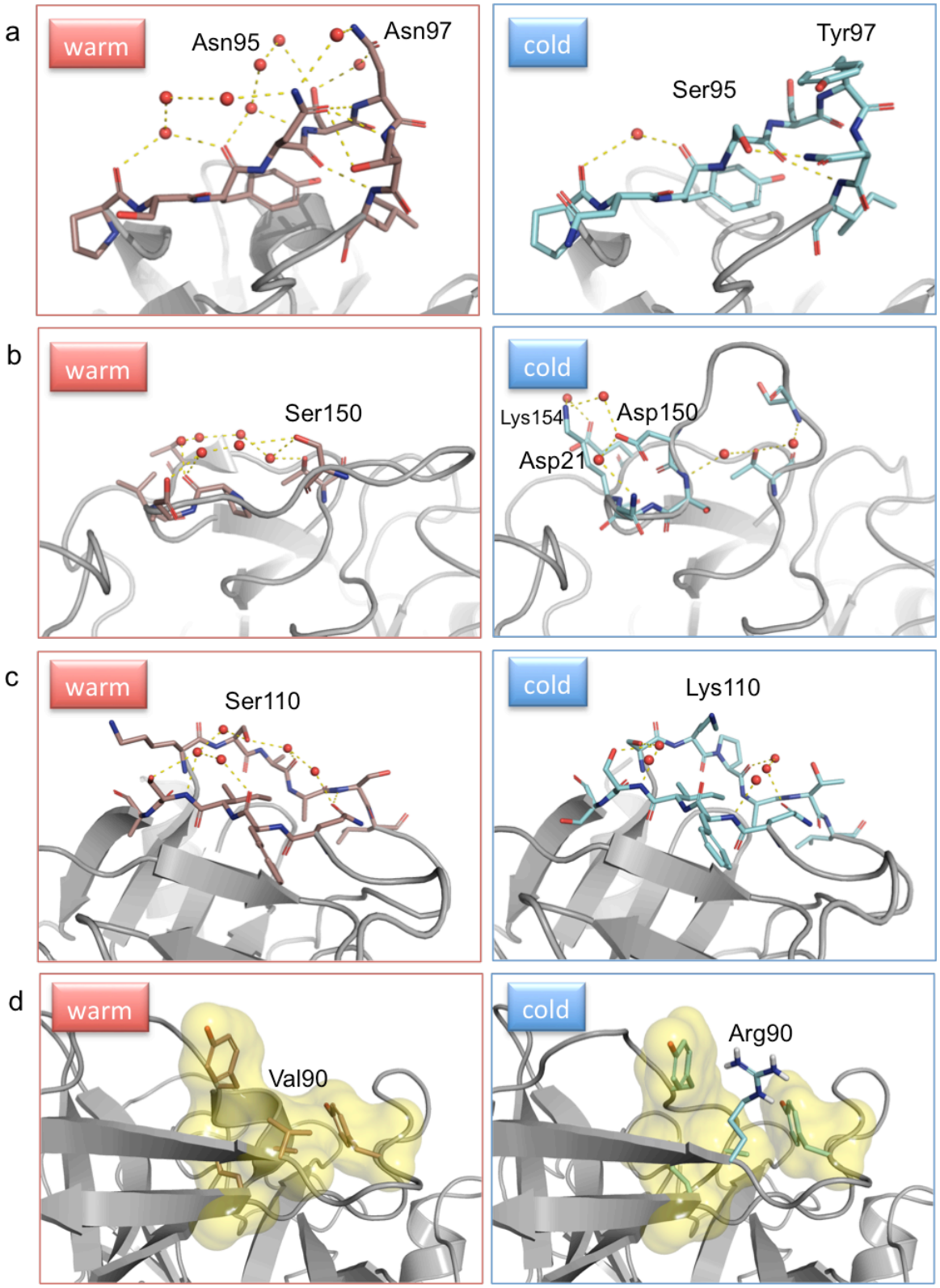


Figure 17: Different mechanisms for increasing the surface softness in cold-adapted trypsin (right) relative to warm-adapted trypsin (left): (a) Disruption of an H-bond network by mutation of polar residue into a less polar (N97Y). (b) Complete change of H-bond network and loop orientation (S150D and correlated mutations). (c) Disruption of an H-bond network by mutation of a polar residue into a charged one (S110D). (d) Destabilization of the packing of hydrophobic surface patches by mutation of a nonpolar residue into a charged one (V90R).

The ultra-high resolution trypsin structures further reveal that a surface H-bonding network can also be disrupted by mutations of a polar residue into a charged one, since the latter one may prefer to interact with bulk solvent. This is exemplified by the mutation S110 to K110 in AST. Finally, one can also identify mutations that destabilize the packing of hydrophobic surface patches by mutation of nonpolar residue into a charged one, as is the case for the mutation of V90 in BT to R90 in AST (Figure 17). Of the examples mentioned above, we have demonstrated the validity of these through the mutations BTS_{150D} and BT_{N97Y} presented in Table 1.

5.4 Concluding Remarks and Future Perspectives

The works presented in papers **II** and **III** clearly reproduce the characteristic balance between activation enthalpy and entropy for cold adapted versus warm-active enzymes. However, what is rather surprising is that it is the protein-water surface that appears to regulate the activation enthalpy-entropy balance. Softening the protein surface of warm-adapted trypsin is easily achieved through single point mutations that efficiently rearrange the thermodynamic activation parameters and $\Delta U_{ss}^{\ddagger}$ to become cold adapted like. The phenomena of changing the balance between $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ due to surface mutations have, in fact, also been experimentally observed for other orthologous enzymes (74). We have shown that modifying the surface softness, either through mutations or by restraining the surface atom mobilities simply by force, apparently change the activation enthalpy of the catalyzed reaction at the expense of the activation entropy without altering activation free energy. Mutations softening the surface are beneficial for adaption to low temperatures as they, not only make the rate more temperature insensitive by reducing $\Delta H^{\ddagger}$, but also counteract the structural rigidity imposed by lowering the temperature. Moreover, given that the active site residues are basically always conserved between highly similar orthologous warm- and cold-adapted enzymes, it seems not so far-fetched to assume that the difference in protein surface properties presented in papers **II** and **III** is the general feature of cold-adapted enzymes.

Our results successfully identifies the same loop regions important for adaption to cold as those reported by previous studies on trypsin thermostability (75). However, experimental characterization is needed to examine whether the reported mutations presented here only change the thermodynamic activation parameters or whether they affect thermostability as well. As indicated above, softening the enzyme surface is

seemingly easy, but the task of stiffening it (e.g. make the enzyme more warm-adapted like) through mutations is not as trivial. A future work could in such respect examine the possibility of regenerating the H-bond network responsible for the surface rigidity in warm-adapted trypsin through correlated mutations in cold-adapted trypsin. More importantly, we need to investigate a wider range of enzymes to verify if the findings for trypsin temperature adaptation apply to enzymes in general. Thus, the grand question remaining is whether or not the protein-water surface softness is globally used in nature's enzyme design strategy as a tool to fine-tune both the enthalpy-entropy balance and the thermostability as the mechanism responsible for temperature adaptation.

6 Human Purine Nucleoside Phosphorylase (IV and V)

In this chapter the reaction mechanism for the glycosidic bond cleavage step in human purine nucleoside phosphorylase, PNP, (paper **IV**) and the thermodynamic activation parameters calculated for native and mutant PNP (paper **V**) will be discussed.

6.1 Introduction

Purine nucleoside phosphorylase (PNP) catalyzes the reversible cleavage of the glycosidic bond of ribo- and deoxyribonucleosides to generate the corresponding purine base and ribose/deoxyribose-1-phoshapte (76). Human PNP (HsPNP) is a homotrimer (low-mm) with the catalytic site located near the subunit-subunit interface as illustrated in Figure 18.

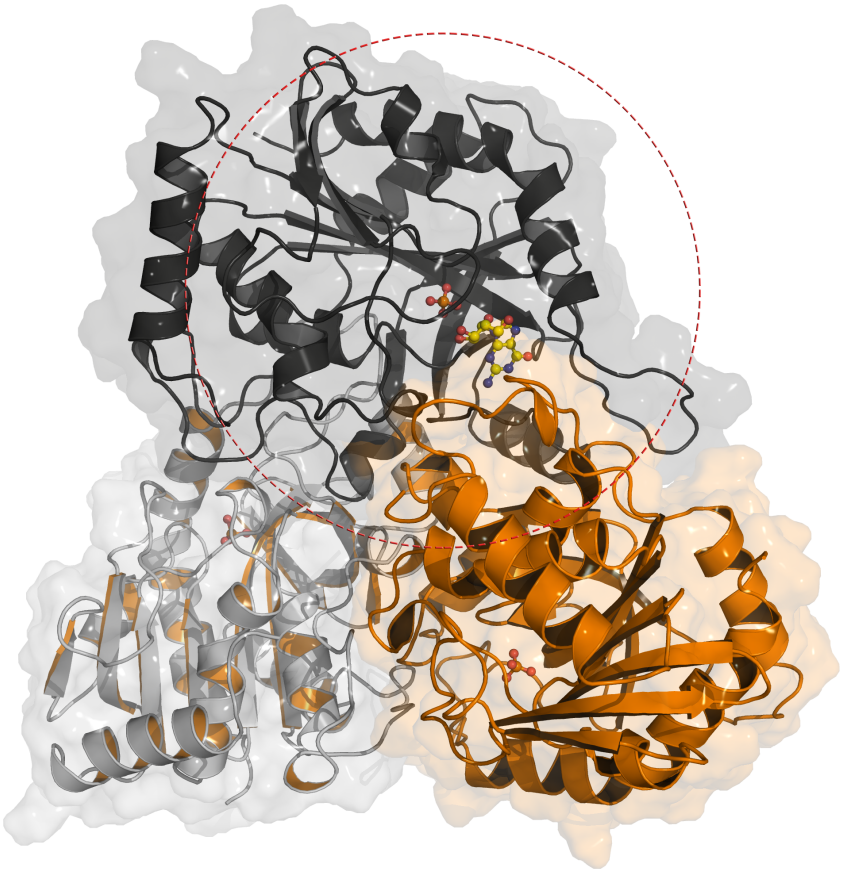


Figure 18: Human purine nucleoside phosphorylase homotrimer with guanosine (yellow) in the active site of the black subunit with the F159* loop from the adjacent subunit (orange) covering it from accessible solvent. The red dashed circle illustrates the 25 Å simulation sphere centered on the active site.

HsPNP is specific for 6-oxopurines, e.g. inosine (INO) and guanosine (GUO), and shows only marginal affinity for 6-aminopurines (e.g. adenosine, ADO). Thus, this class of PNPs is often referred to as "Ino-Guo phosphorylases" (77, 78). Molecular structures of INO, GUO and ADO are illustrated in Figure 19.

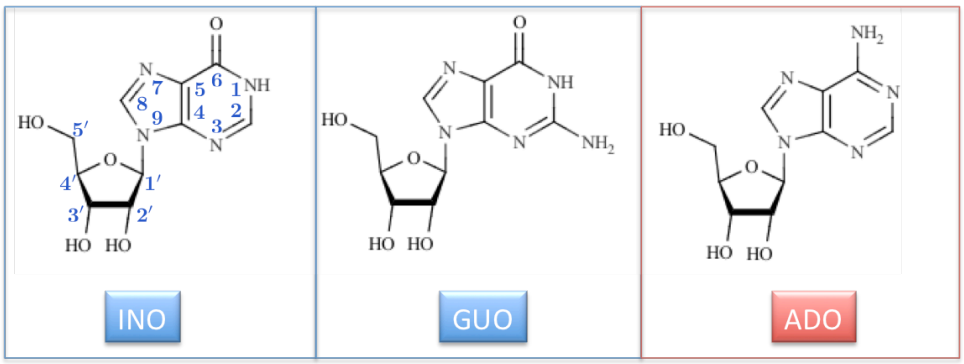


Figure 19: Structures of the included substrates inosine (INO), guanosine (GUO) and adenosine (ADO). The general atom-numbering scheme is illustrated for INO.

Besides potential important applications within the field of medicinal chemistry, a recent publication with experimental mutations of distant residues to the active site altering the thermodynamics of human toward bovine PNP (74) is intriguing in the context of the work presented in papers **II** and **III**.

Despite being studied for decades, the catalytic mechanism for low-mm PNPs is still not adequately elucidated and a number of key questions remain. Firstly, the catalytic role of the phosphate and in particular its protonation state is unclear. Previous works have proposed inconsistent reaction mechanisms in the sense that they employ singly, doubly or even a triply protonated phosphate group (78-80). Secondly, the amino acids stabilizing the transition state and their role are not completely elucidated. Finally, the protonation state of the purine ring during the glycosidic bond cleavage is highly controversial. Even though fluorescence studies have shown that the purine base binds in the anionic form in the reverse reaction (81), indicating that protonation (at N7 or N9) occurs after phosphorolysis, proposed mechanisms employing a positively charged ribonucleoside protonated at N7 still exist (82). However, many suggested mechanisms seem to agree with the fluorescence studies, but with the conflict of whether the negative charge is distributed between N7 and N9 (80, 83) or mainly on O6 (78). A puzzle here is the fact that there is no obvious proton donor to facilitate the protonation of the anionic base after phosphorolysis in the vicinity of neither the base N7 nor N9 in the active site of HsPNP. Clearly, it is of considerable interest to resolve both the HsPNP reaction mechanism and the origin of the observed substrate specificity.

6.2 Computer Simulations Reveal the HsPNP 6-Oxopurine Specificity

In papers **IV** and **V** we have applied extensive EVB simulations to investigate the glycosidic bond cleavage step ($\phi_1 \rightarrow \phi_2$ in Figure 20) in HsPNP. The EVB reference reactions for GUO, INO and ADO were fitted to free energy surfaces obtained at the DFT-level utilizing a singly protonated phosphate nucleophile. The alternative doubly protonated phosphate group resulted in an increased activation barrier and a substantial more endergonic reaction energy. In addition, the reaction was rendered more S_N1 like, compared to S_N2 utilizing HPO_4^{2-}, which generally reflects the weaker nucleophilicity of $H_2PO_4^-$. To our knowledge, the DFT profiles for purine phosphorolysis have not been reported previously.

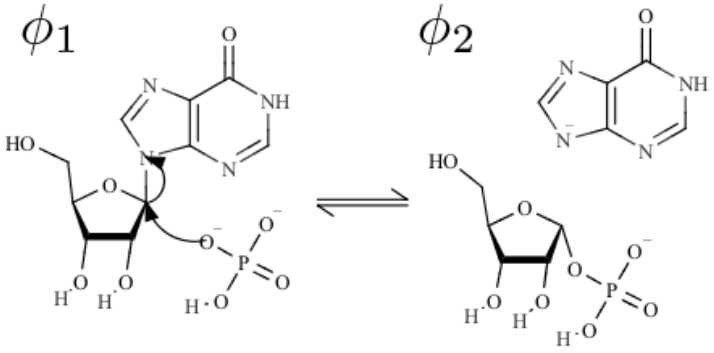


Figure 20: EVB resonance structures applied in the phosphorolytic cleavage of the glycosidic bond of inosine, guanosine and adenosine ($\boldsymbol{\phi_1 \rightarrow \phi_2}$).

The EVB reference reactions were fitted to $\Delta G^{\ddagger}$ of 30.6, 28.5 and 30.6 kcal/mol and ΔG_0 of 9.2, 7.2 and 9.4 kcal/mol for GUO, INO and ADO (Figure 19), respectively. The EVB parameters applied in the enzyme-catalyzed reactions were taken as an average of 10 calibrated reference reactions for each substrate. Trimeric HsPNP was modeled using a 25 Å simulation sphere centered on the active site of the reacting subunit as illustrated in Figure 18. A total of 100 EVB simulations were run for the enzyme reactions, resulting in a simulation time of 51 ns for each substrate. The resulting reaction free energy profiles for the glycosidic bond cleavage step in HsPNP and water are illustrated in Figure 21. The average activation free energies from the EVB simulations in the enzyme-catalyzed reactions are 12.8, 13.0 and 20.2 kcal/mol for GUO, INO and ADO, respectively.

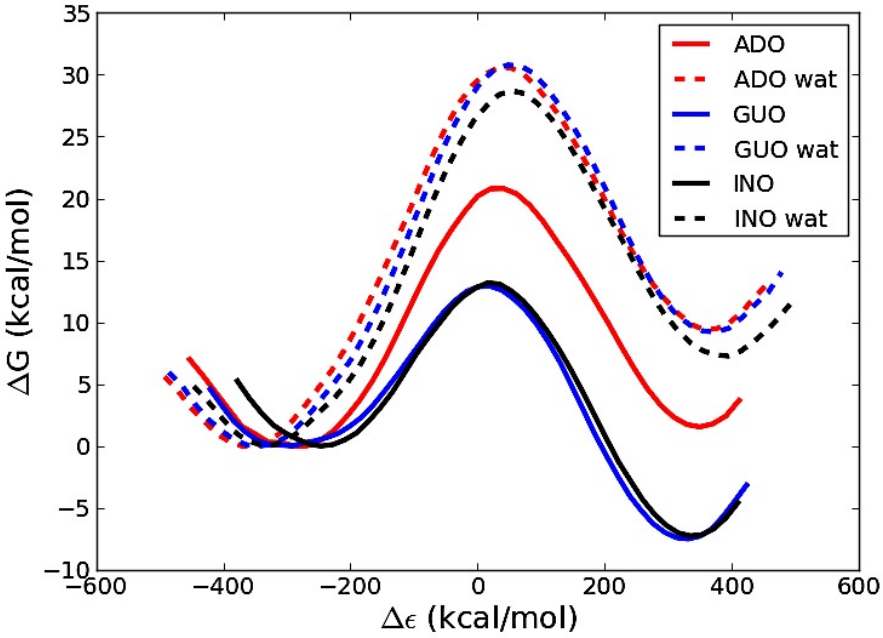


Figure 21: Calculated reaction free energy profiles for the glycosidic bond cleavage step ($\boldsymbol{\phi_1} \rightarrow \boldsymbol{\phi_2}$) for adenosine (red), guanosine (blue) and inosine (black) in HsPNP together with the corresponding reference reactions in water (dashed lines).

It must, however, be emphasized that the rate-limiting step is generally assumed to be the release of purine (74, 84, 85). Consequently, experimentally measured k_{cat} values and corresponding thermodynamic parameters cannot be used directly to evaluate the glycosidic bond cleavage step, but rather as an upper threshold for the activation free energies. Thus, the calculated barriers presented in Table 2 are fully compatible with the experimentally measured k_{cat} values reported by Stoeckler *et al.* (86). More importantly, the calculated $\Delta G^{\ddagger}$ for GUO is in excellent agreement with the experimentally determined $\Delta G^{\ddagger}$ of 14 ± 1 kcal/mol for the glycosidic bond cleavage step obtained from pre-steady-state kinetics (74).

Table 2: Average activation and reaction free energies calculated at 298 K for the glycosidic bond cleavage step together with the experimental activation free energies for the reaction in native and N243D HsPNP for inosine (INO), guanosine (GUO), and adenosine (ADO)

	$\langle \Delta G^{\ddagger} \rangle$	$\langle \Delta G_0 \rangle$	$\Delta G^{\ddagger}_{expt}$	$\langle \Delta G^{\ddagger} \rangle_{N243D}$	$\langle \Delta G_0 \rangle_{N243D}$	$\Delta G^{\ddagger}_{expt,N243D}$
INO	13.0 ± 0.2	-7.0 ± 0.3	15.1 ± 0.1	15.9 ± 0.2	-2.3 ± 0.3	17.0 ± 0.1
GUO	12.8 ± 0.1	-7.1 ± 0.2	15.5 ± 0.1	16.8 ± 0.4	-4.3 ± 0.5	16.8 ± 0.2
ADO	20.2 ± 0.2	0.9 ± 0.5	21.0 ± 0.1	14.7 ± 0.2	-6.9 ± 0.3	16.0 ± 0.3

$\Delta G^{\ddagger}_{expt}$ calculated from k_{cat} (s^{-1}) values reported by Stoeckler *et al.* (86) where substrate release is assumed rate limiting

It is rather remarkable that our simulations are able to reflect the fact that HsPNP display negligible activity for 6-aminopurines (ADO) and not 6-oxopurines (GUO and

INO). Nevertheless, the enzyme lowers both the activation and reaction free energies for all substrates compared to the reference reactions in water as illustrated in Figure 21. The lowering is significantly smaller for ADO, and this clearly demonstrates that HsPNP catalyzes the chemical step with ADO as substrate, but at much lower rates compared to GUO and INO.

Analysis of the transition state structures reveal a rather comprehensive H-bond network stabilizing both the ribonucleoside and the phosphate group (Figure 22). This is also reflected by a significant reduction in the EVB potential activation energy (mainly the electrostatic part) in HsPNP compared to the reaction in water. With the exception of one additional H-bond between E201 and the base C2 amine group for GUO, the two 6-oxopurines display literally identical transition state configurations. This is not surprising considering that both are natural substrates of HsPNP and with similar reaction rates. The EVB simulations furthermore indicate that the low activity for ADO originates from the chemical step where $\Delta G^{\ddagger}$ is $\sim$7 kcal/mol higher compared to INO and GUO. It is therefore likely that the rate-limiting step with ADO as substrate is the glycosidic bond cleavage step and not adenine release. Moreover, the increased reorganization energy of $\sim$20 kcal/mol for ADO clearly demonstrate that HsPNP is significantly more electrostatically preorganized for the 6-oxopurines.

In an attempt to structurally understand the increased barrier for ADO we found that the residues stabilizing the phosphate group illustrated in Figure 22 were critical for the reaction free energies. In particular S33 together with H64 and H86 were found to play key roles upon orienting the phosphate group properly for the nucleophilic attack in the glycosidic bond cleavage step (paper **IV**). In our initial simulations optimizing the EVB starting structures, we observed a few simulations with GUO or INO as substrates where these interactions where randomly disrupted. This resulted in increased activation energies of 4 – 6 kcal/mol. These observations are in agreement with a recent study that entitled S33-H64-H86 the “catalytic triad” and also proposed it to stabilize the phosphate in the active site (87). Interestingly, the phosphate group and the S33-H64-H86 triad does not share identical interactions along the reaction coordinate for ADO as is the situation with INO and GUO.

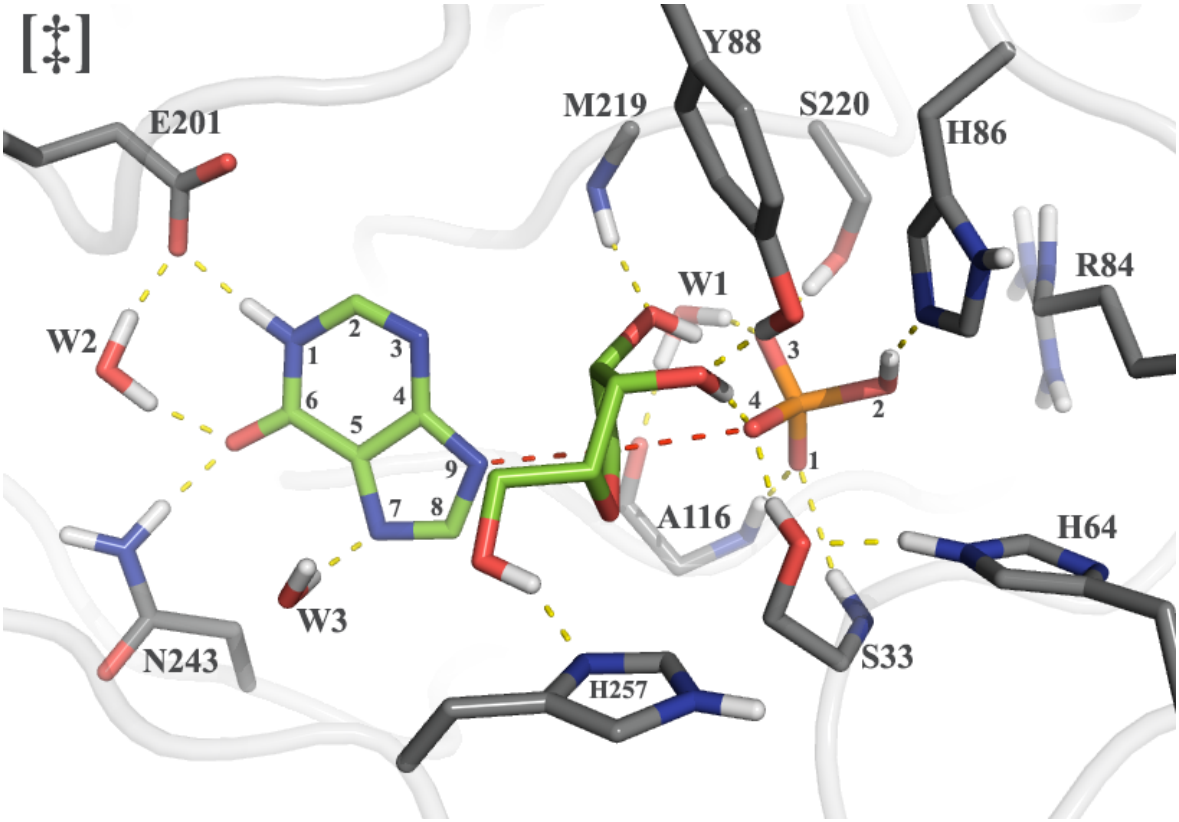


Figure 22: Snapshot of the transition state for inosine in HsPNP as observed from the EVB simulations. The dashed red lines represent the reaction coordinate (Op4-C1'-N9).

Analysis of the ADO transition state structure reveal a different interaction pattern with the base where E201 is shifted away (no H-bond partner to N1) and N243 has flipped making H-bond interactions to both N7 and the C6 amino group. As a result the entire ribonucleoside position is shifted slightly downwards in the active site compared to INO and GUO. Moreover, the zigzag pattern of H-bond interactions between N243, W2 and E201 to the base Figure 22), which previously has been suggested by Tebbe *et al.* (78) as a potential key for substrate recognition, is not present with ADO as substrate. Thus, the low activity for ADO could originate from altered interactions to N243 and E201.

6.3 N243D Mutation Renders HsPNP 6-Aminopurine Active

The fact that the EVB simulations are able to nicely differentiate between the HsPNP specificity for 6-oxo and 6-aminopurines is very encouraging. In light of the above discussion, and to further test the robustness of the presented reaction mechanism for the chemical step, we chose to compute the N243D mutation. This mutation is of notable interest as it is able to eliminate the HsPNP selectivity rendering it active for both 6-oxo and 6-aminopurines, but with slightly higher specificity for ADO compared to GUO and

INO (86). From the results presented in Table 2 it is evident that our model is able to reproduce the experimental trends also for the N243D mutation with all 3 substrates. The average $\Delta G^{\ddagger}$ of 16.8, 15.9 and 14.7 kcal/mol for GUO, INO and ADO, respectively, are in excellent agreement with the experimentally obtained values (Table 2). Thus, the predicted reduction of ~6 kcal/mol in $\Delta G^{\ddagger}$ for ADO perfectly reproduces the experimentally measured ~5000-fold increase in k_{cat} with respect to wild-type HsPNP (86). From the transition state structure of ADO in N243D HsPNP illustrated in Figure 23, it can immediately be seen that the H-bond zigzag pattern discussed above is not present. It was furthermore rather surprising that the pattern remained present for INO and GUO, though bridged by water molecules to E201 instead of N243. This interesting observation thus indicates that the previously proposed H-bond pattern is rather unimportant for the glycosidic bond cleavage step.

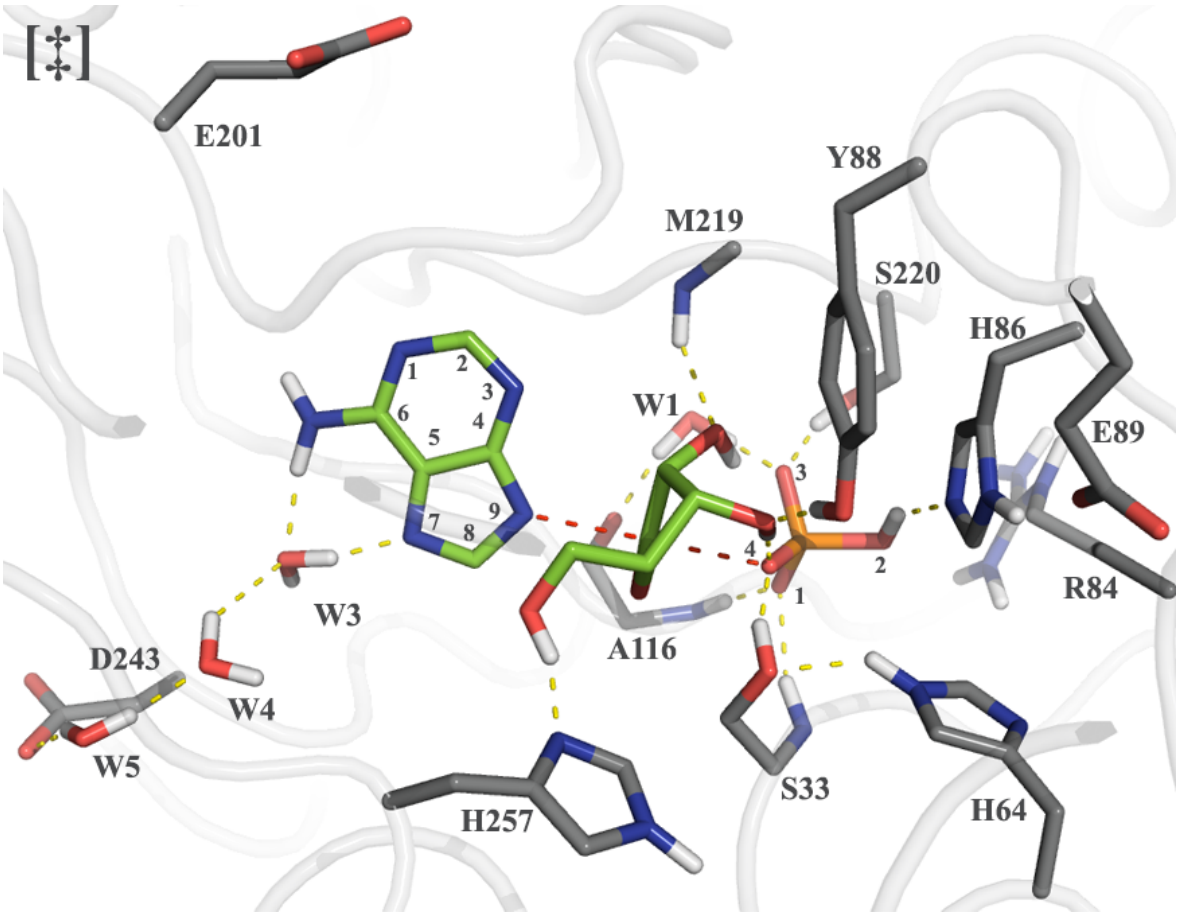


Figure 23: Snapshot of the transition state for adenosine in N243D HsPNP as observed from the EVB simulations. The dashed red lines represent the reaction coordinate (Op4-C1'-N9).

Common for all three substrates is that D243 is observed shifted away from the original position of N243 and a water molecule stabilizes the negative charge on N7. In the case of ADO, the H-bonds between W3 and the base N7 and C6 amino group is

preferable compared to that of N243 since the amino – amino repulsion is eliminated and the water molecule (W3) establishes a significantly better H-bond interaction to N7. It is also notable that the ADO N7 is linked to D243 through a stable chain of 3 water molecules (Figure 23). In the case of GUA and INO, one additional water molecule is observed stabilizing O6, resulting in a chain of 4 water molecules bridging to D243. The key improvement for ADO compared the 6-oxopurines is, however, the average electrostatic interaction energy between the C6 amino group and residue 243, which is significantly improved upon replacing asparagine with glutamic acid. Moreover, the water molecule interacting directly with the base N7 (W3 in Figure 23) has enhanced the electrostatic interaction compared to wild-type HsPNP. This is also reflected by the difference in reorganization energy (Δλ) between native and mutated HsPNP, which was estimated to be ~ -10 kcal/mol. INO and GUO on the other hand were estimated with Δλ of ~10 kcal/mol, which is the opposite of that to ADO. This strongly suggest that the N243D mutation has rendered the enzyme more electrostatic preorganized for 6-amino purines (i.e. ADO) and less for 6-oxopurines (i.e. GUO and INO).

6.4 Nucleobase Protonation is Not Rate-Limiting

It is rather remarkable that the presented EVB model is able not only to reproduce the HsPNP specificity, but also the experimental trends for the N243D mutation. As discussed above, substrate release is indicated as the rate-limiting step in phosphorolysis, but prior to release the anionic product intermediate must be protonated. However, we were not able to locate any suitable residues in the range of the base that could potentially serve as a proton donor. As a consequence, we ran several 5 ns long simulations of ϕ_2 (Figure 20) completely unrestrained in the active site of HsPNP. Interestingly, 2 different stable conformations with the potential of shuttling a proton from ribose-1-phosphate to the base N9 were observed. The calculated DFT barrier for $\phi_2 \rightarrow \phi_3$ illustrated in Figure 24 of 4.9 kcal/mol supports the possibility of this type of proton shuttle mechanism. The average activation free energies in HsPNP from the EVB simulations are 6.7 kcal/mol.

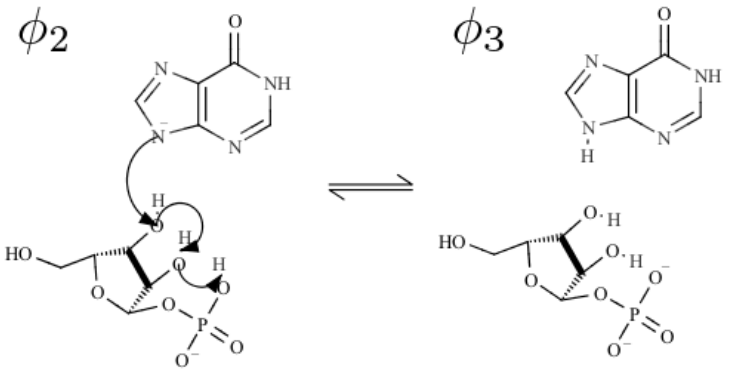


Figure 24: The EVB resonance structures used for modeling the hypothesized proton shuttle to hypoxanthine (glycosidic bond cleave product of inosine).

Other alternatives for potential routes of course exist, and one observed conformation (paper **IV**) suggests that the proton could be linked through a chain of four water molecules. Nevertheless, there probably exist numerous travel routes for the proton, and it is a complex task to suggest the exact mechanism of the nucleobase protonation. The key point in this matter is that the nucleobase protonation has a barrier that is significantly lower compared to the glycosidic bond cleavage. In our example, the activation free energy is only 6.7 kcal/mol for the proton shuttle from the phosphate group to the base N9. Consequently, the glycosidic bond cleavage step would be rate limiting compared to potential protonation steps. To our knowledge, this is also the first time this kind of proton shuttling has been suggested and tested for HsPNP.

6.5 Remote Mutations Altering the Enthalpy-Entropy Balance

In context of the workflow illustrated in Figure 4 (section 4.1), the work presented in paper **IV** focused on stage 1 resolving the glycosidic bond cleavage step. With the encouraging results obtained utilizing the reaction mechanism illustrated in Figure 20, we have in paper **V** advanced to stage 2 and 3 investigating the thermodynamic activation parameters related to the chemical step in HsPNP. In papers **II** and **III** we found that the protein – water interactions in trypsin regulates the enthalpy-entropy balance. Point mutations altering the trypsin surface softness furthermore predict significant changes to $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ that are nearly perfect compensating. The importance of such enthalpy-entropy compensation is however likely easy to overlook since mutations far away from the active site generally do not significantly affect the catalytic rates (71). A recent publication with experimental remote mutations in HsPNP altering

the enthalpy-entropy balance (74) is thus of considerable interest. The double mutation K22E and H104R (HsPNP E:R) with GUO as substrate was reported by Ghanem *et al.* to render the HsPNP thermodynamics similar to that of bovine PNP (BtPNP). As illustrated in Figure 25, residues 22 and 104 are far away from the active site. To capture these, the entire homotrimer was soaked in a spherical droplet of water molecules with radius 53 Å. All simulations were run between 50 and 100 times at 6 different temperatures in the range 283 – 308 K for HsPNP and HsPNP-E:R.

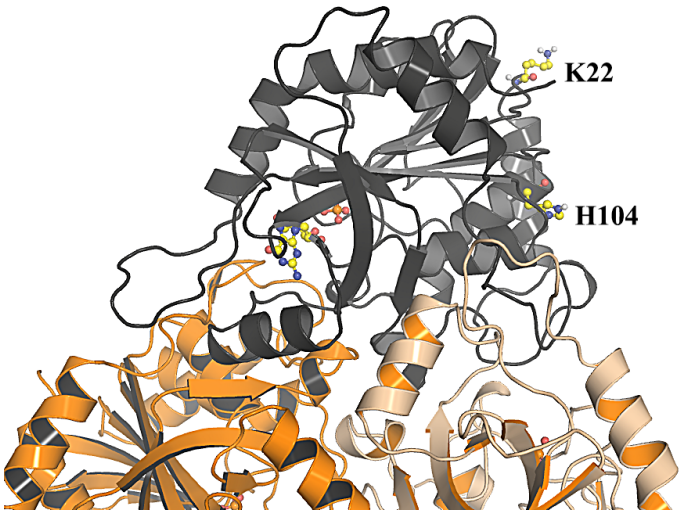


Figure 25: Illustration of the positions of the distant residues K22 and H104 relative to the active site in HsPNP.

The activation free energy upon simulating the entire homotrimer is 13.1 ± 1.2 kcal/mol and in excellent agreement with the experimental $\Delta G^{\ddagger}$ = 14.3 ± 1.0 kcal/mol (Table 3). Moreover, the calculated $\Delta H^{\ddagger}$ and $-T\Delta S^{\ddagger}$ for GUO of 14.7 and -1.6 kcal/mol are in very good agreement with the values obtained by extrapolating stopped-flow fluorescence rates using the Arrhenius equation of 18.5 ± 1 and -4.1 ± 1 kcal/mol, respectively (74). The slight underestimation of the parameters is largely justified by the fact that the calculated activation free energy is ~1.2 kcal/mol lower compared to that obtained from the Arrhenius plot in ref. (74). This could in principle have been corrected for by fitting the EVB reaction free energy surface for the enzyme-catalyzed reaction to coincide with the experimental rates. This is nevertheless irrelevant since it is the relative values that are interesting here. What is however rather remarkable, is the fact that the EVB simulations are able to perfectly reproduce the experimentally observed enthalpy-entropy balance. It should also be noted that the same thermodynamic activation parameters with GUO as substrate was obtained using the smaller 25 Å simulations sphere (paper **V**).

Table 3: Calculated thermodynamic activation parameters at 298 K for the glycosidic bond cleavage step in HsPNP and HsPNP-E:R with guanosine as substrate together with experimental data including BtPNP adapted from ref. (74).

	HsPNP		E:R-PNP		BtPNP	
Parameter	EVB	expt	EVB	expt	EVB	expt
$\Delta H^{\ddagger}$	14.7 ± 0.8	18.6 ± 0.7	18.3 ± 1.0	20.5 ± 0.7	-	16.1 ± 0.7
$-T\Delta S^{\ddagger}$	-1.6 ± 0.8	-4.1 ± 0.2	-3.8 ± 1.0	-6.2 ± 0.2	-	-2.1 ± 0.1
E_a	15.3 ± 0.8	19.1 ± 0.7	18.9 ± 1.0	20.8 ± 0.7	-	16.7 ± 0.7
$\Delta G^{\ddagger}$	13.1 ± 1.2	14.5 ± 1.0	14.4 ± 1.4	14.3 ± 1.0	-	13.9 ± 0.7
$\Delta U^{\ddagger}_{rr+rs}$	19.7 ± 2.0	-	33.9 ± 4.2	-	-	-
$\Delta U^{\ddagger}_{ss}$	-5.0 ± 2.2	-	-15.6 ± 4.3	-	-	-

Our calculations thus indicate that it is possible to extract accurate activation enthalpies and entropies without including the entire enzyme. This can be of great importance upon reducing the high computational costs involved in simulating large enzymes like the homotrimer described here. Nevertheless, in order to capture effects of surface mutations and corresponding changes in external and internal potential activation energies obviously render it necessary to include the entire enzyme.

The computed free energy of activation for HsPNP-E:R is 14.4 ± 1.4 kcal/mol and in good agreement with the experimental $\Delta\Delta G^{\ddagger}$ of ~0 kcal/mol for the mutation (Table 3). This again just underscores the general notion that mutations far away from the active site do not significantly alter the catalytic rates (71). The calculated $\Delta\Delta H^{\ddagger}$ and $-T\Delta\Delta S^{\ddagger}$ of 3.6 kcal/mol and -2.2 kcal/mol, respectively, are furthermore in excellent agreement with the experimentally obtained $\Delta\Delta H^{\ddagger} = 1.9$ kcal/mol and $-T\Delta\Delta S^{\ddagger} =$ -2.1 kcal/mol (Table 3). This type of remarkable enthalpy-entropy compensation induced by mutations distant to the active site is consistent with the results presented for trypsin in section 5 (papers **II** and **III)**. It should however be noted that the observed changes in the thermodynamic activation parameters presented here are relatively small compared to those for trypsin. Thus, it is in this respect rather remarkable that our EVB simulations are able to capture these trends. The internal and external potential activation energies contributing to the activation enthalpy (Eq. 32) are however significantly different between native and mutated HsPNP. From the results presented in Table 3 it can immediately be recognized that $\Delta U^{\ddagger}_{ss}$ is rendered significantly more favorable in HsPNP-E:R compared to the native enzyme. This is reflected by the calculated backbone RMSF values at the transition state (Figure 26), which

demonstrates that the mutations have rendered the protein softer in the mutated regions (residues 22 and 104). The average RMSF for HsPNP and HsPNP-E:R is 0.56 and 0.55, respectively, which is not significantly different. In similarity to the results obtained with trypsin, it is thus smaller regions that become more flexible. However, in contrast to our previous works, the favorable $\Delta U_{ss}^{\ddagger}$ term is here cancelled by $\Delta U_{rr+rs}^{\ddagger}$ that has increased by ~10 kcal/mol in mutated HsPNP. This is not surprising considering the fact that total charge of the surroundings (the enzyme) to the reacting fragment is rendered more negative by the K22E and H104R mutations. Thus, the electrostatic stabilization of the substrate's anionic development in the transition state becomes less favorable, which is also evident from the increased enthalpy of activation.

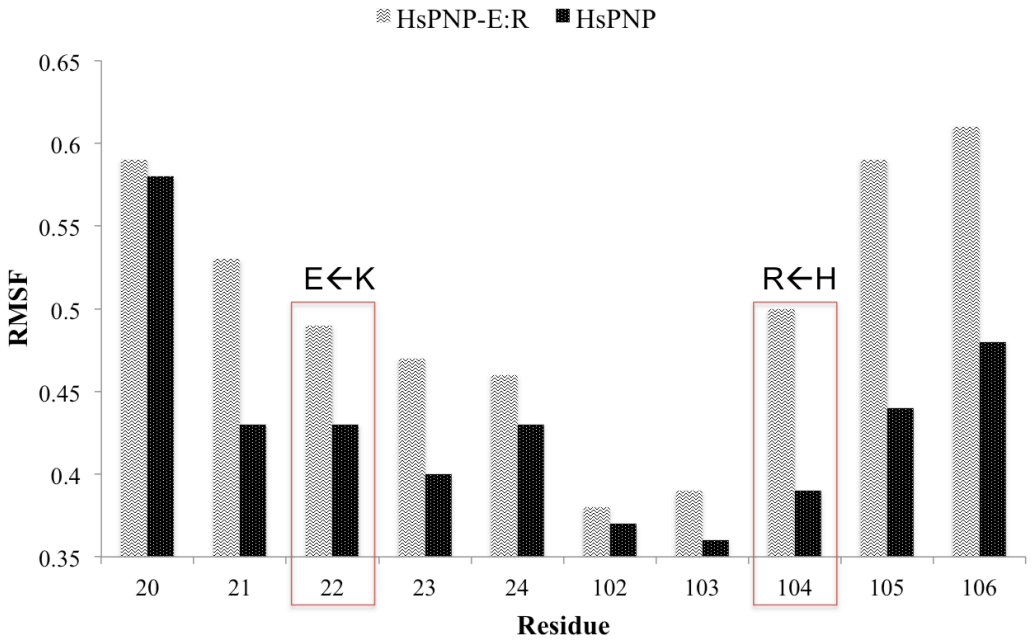


Figure 26: Average backbone fluctuations at the transition state demonstrates that HsPNP-E:R is rendered softer around the mutated (K22E and H104R) regions compared to native HsPNP.

As also recognized by Ghanem *et al.*, the K22E and H104 mutations do not alter the thermodynamics of HsPNP towards BtPNP (74). That is, $\Delta\Delta H^{\ddagger}$ and $-T\Delta\Delta S^{\ddagger}$ for BtPNP is -2.5 kcal/mol and 2.0 kcal/mol, respectively, compared to HsPNP, which is the exact opposite to the enthalpy-entropy compensation obtained with HsPNP-E:R. Based on the experimental data presented in ref. (74) and from our calculations it must be concluded that the these mutations do not alter the thermodynamics of human towards bovine purine nucleoside phosphorylase. Nevertheless, these long-ranged mutations still

serve as a perfect example to how protein surfaces can regulate the enthalpy-entropy balance without significantly altering the free energy of activation.

6.6 Concluding Remarks and Future Aspects

In the works presented in papers **IV** and **V** we have successfully reproduced both the reaction rates and the thermodynamic activation parameters for the glycosidic bond cleavage step in native and mutant human purine nucleoside phosphorylase. Our calculations strongly suggest that HsPNP is significantly more electrostatically preorganized for catalysis of 6-oxopurines compared to 6-aminopurines. This is furthermore reflected by the thermodynamic activation parameters presented in paper **V** where the $\Delta H^{\ddagger}$ for ADO is ~2-fold higher compared to that of INO and GUO. The presented reaction mechanism furthermore reproduces the active site mutation, N243D, rendering HsPNP highly active also for adenosine. Our calculations demonstrate that the mutation significantly improves the electrostatic preorganization for catalysis of ADO compared to the native enzyme.

In paper **V** we computed long-ranged mutations in HsPNP changing the enthalpy-entropy balance without significantly affecting the reaction rates. However, the thermodynamic activation parameters for HsPNP-E:R are only altered ~±2-3 kcal/mol with respect to the native enzyme. Thus, it is rather remarkable that our EVB simulations are able to capture the mutational trends for these marginal changes. Nevertheless, the internal and external potential activation energies showed significant changes. The enzyme surface is rendered softer in the mutated regions, which is also reflected by the favorable $\Delta U_{ss}^{\ddagger}$ calculated for HsPNP-E:R. However, this effect is almost completely cancelled by the unfavorable $\Delta U_{rr+rs}^{\ddagger}$ term, which explains the small change in $\Delta H^{\ddagger}$. Nevertheless, the presented EVB simulations combined with the available experimental data provide valuable insight to how orthologous enzymes regulate the enthalpy-entropy balance through surface mutations.

Clearly it is challenging to derive the exact origin of the observed effect to the thermodynamic parameters for the double mutation K22E and H104R. This is particularly true with only the experimental data presented in ref. (74) available. However, our calculations indicate that the H104R mutation is the most likely source of the improved external activation potential energy, whereas K22E counteracts the effect

by acting unfavorable to $\Delta U^{\ddagger}_{rr+rs}$. Thus, the mutation H104R could in fact render the thermodynamics of human towards bovine PNP when not counteracted by K22E. A future study should therefore do these mutations individually. Moreover, in light of the work presented in papers **II** and **III**, our EVB model could in a future project also be used for studying different temperature adapted PNP orthologs.

References

1. Payen A & Persoz JF (1833) Memoir on diastase, the principal products of its reactions, and their applications to the industrial arts. *Ann Chim Phys* 53:73-92.
2. Leatherbarrow RJ, Fersht AR, & Winter G (1985) Transition-state stabilization in the mechanism of tyrosyl-tRNA synthetase revealed by protein engineering. *Proc Nat Acad Sci U S A* 82(23):7840-7844.
3. Warshel A & Florian J (1998) Computer simulations of enzyme catalysis: Finding out what has been optimized by evolution. *Proc Nat Acad Sci U S A* 95(11):5950-5955.
4. Somero GN (1995) Proteins and temperature. *Ann rev physio* 57:43-68.
5. Warshel A (1991) *Computer modeling of chemical reactions in enzymes and solutions* (Wiley, New York) pp xiv, 236 p.
6. Aqvist J & Warshel A (1993) Simulation of enzyme-reactions using valence-bond force-fields and other hybrid quantum-classical approaches. *Chem Rev* 93(7):2523-2544.
7. Wolfenden R & Snider MJ (2001) The depth of chemical time and the power of enzymes as catalysts. *Acc chem res* 34(12):938-945.
8. Berg J, Tymoczko J, & Stryer L (2002) Free energy is a useful thermodynamic function for understanding enzymes. *Biochemistry*, (W H Freeman, New York), 5 Ed.
9. Pauling L (1946) Molecular architecture and biological reactions. *Chem Eng News Archive* 24(10):1375-1377.
10. Warshel A*, et al.* (2006) Electrostatic basis for enzyme catalysis. *Chem Rev* 106(8):3210-3235.
11. Hansson T, Nordlund P, & Aqvist J (1997) Energetics of nucleophile activation in a protein tyrosine phosphatase. *J Mol Biol* 265(2):118-127.
12. Fuxreiter M & Warshel A (1998) Origin of the catalytic power of acetylcholinesterase: Computer simulation studies. *J Am Chem Soc* 120(1):183-194.
13. Aqvist J & Fothergill M (1996) Computer simulation of the triosephosphate isomerase catalyzed reaction. *J Biol Chem* 271(17):10010-10016.
14. Fothergill M, Goodman MF, Petruska J, & Warshel A (1995) Structure-energy analysis of the role of metal-ions in phosphodiester bond hydrolysis by DNA-polymerase-I. *J Am Chem Soc* 117(47):11619-11627.
15. Hwang JK & Warshel A (1996) How important are quantum mechanical nuclear motions in enzyme catalysis? *J Am Chem Soc* 118(47):11745-11751.
16. Adamczyk AJ, Cao J, Kamerlin SC, & Warshel A (2011) Catalysis by dihydrofolate reductase and other enzymes arises from electrostatic preorganization, not conformational motions. *Proc Nat Acad Sci U S A* 108(34):14115-14120.
17. Pisliakov AV, Cao J, Kamerlin SCL, & Warshel A (2009) Enzyme millisecond conformational dynamics do not catalyze the chemical step. *Proc Nat Acad Sci* 106(41):17359-17364.
18. Vernon CA (1967) The mechanisms of hydrolysis of glycosides and their revelance to enzyme-catalysed reactions. *Proc R Soc London, Ser B* 167(1009):389-401.
19. Dunn BM & Bruice TC (1973) Physical organic models for mechanism of lysozyme action. *Adv Enzymol Ramb* 37:1-60.

20. Bruice TC (1976) Some pertinent aspects of mechanism as determined with small molecules. *Ann rev biochem* 45:331-373.
21. Fife TH, Jaffe SH, & Natarajan R (1991) Intramolecular general-acid and electrostatic catalysis in acetal hydrolysis - hydrolysis of 2-(substituted phenoxy)-6-carboxytetrahydropyrans and 2-alkoxy-6-carboxytetrahydropyrans. *J Am Chem Soc* 113(20):7646-7653.
22. Yadav A, Jackson RM, Holbrook JJ, & Warshel A (1991) Role of solvent reorganization energies in the catalytic activity of enzymes. *J Am Chem Soc* 113(13):4800-4805.
23. Warshel A (1978) Energetics of Enzyme Catalysis. *Proc Nat Acad Sci U S A a* 75(11):5250-5254.
24. Warshel A & Russell ST (1984) Calculations of electrostatic interactions in biological-systems and in solutions. *Q Rev Biophys* 17(3):283-422.
25. Marcus RA (1956) On the theory of oxidation‐reduction reactions involving electron transfer. I. *J Chem Phys* 24(5):966-978.
26. Olsson MH, Parson WW, & Warshel A (2006) Dynamical contributions to enzyme catalysis: critical tests of a popular hypothesis. *Chem Rev* 106(5):1737-1756.
27. Warshel A (1998) Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites. *J biol chem* 273(42):27035-27038.
28. Feller G & Gerday C (2003) Psychrophilic enzymes: hot topics in cold adaptation. *Nat rev. Microbiol* 1(3):200-208.
29. Siddiqui KS & Cavicchioli R (2006) Cold-adapted enzymes. *Ann rev biochem* 75:403-433.
30. Fields PA & Somero GN (1998) Hot spots in cold adaptation: localized increases in conformational flexibility in lactate dehydrogenase A4 orthologs of Antarctic notothenioid fishes. *Proc Nat Acad Sci U S A* 95(19):11476-11481.
31. Low PS, Bada JL, & Somero GN (1973) Temperature adaptation of enzymes: roles of the free energy, the enthalpy, and the entropy of activation. *Proc Nat Acad Sci U S A* 70(2):430-432.
32. Lonhienne T, Gerday C, & Feller G (2000) Psychrophilic enzymes: revisiting the thermodynamic parameters of activation may explain local flexibility. *Bba-Protein Struct M* 1543(1):1-10.
33. Smalas AO, Heimstad ES, Hordvik A, Willassen NP, & Male R (1994) Cold Adaption of Enzymes - Structural Comparison between Salmon and Bovine Trypsins. *Proteins: Struct Funct Genet* 20(2):149-166.
34. Bjelic S, Brandsdal BO, & Aqvist J (2008) Cold adaptation of enzyme reaction rates. *Biochemistry* 47(38):10049-10057.
35. Jorgensen WL, Maxwell DS, & TiradoRives J (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J Am Chem Soc* 118(45):11225-11236.
36. Kaminski GA, Friesner RA, Tirado-Rives J, & Jorgensen WL (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J Phys Chem B* 105(28):6474-6487.
37. Marelius J, Kolmodin K, Feierberg I, & Aqvist J (1998) Q: A molecular dynamics program for free energy calculations and empirical valence bond simulations in biomolecular systems. *J Mol Graph Model* 16(4-6):213-+.
38. Allen MP & Tildesley DJ (1989) *Computer simulation of liquids* (Oxford university press, Oxford England) p 385.

39. Zwanzig RW (1954) High-temperature equation of state by a perturbation method .1. Nonpolar gases. *J. Chem. Phys.* 22:1420-1426.
40. Boukharta L, Gutierrez-de-Teran H, & Aqvist J (2014) Computational prediction of alanine scanning and ligand binding energetics in G-protein coupled receptors. *PLoS comp biol* 10(4):e1003585.
41. Keranen H, Gutierrez-de-Teran H, & Aqvist J (2014) Structural and energetic effects of A2A adenosine receptor mutations on agonist and antagonist binding. *PloS one* 9(10):e108492.
42. Keranen H, Aqvist J, & Gutierrez-de-Teran H (2015) Free energy calculations of A adenosine receptor mutation effects on agonist binding. *Chem commun*.
43. Lee FS, Chu ZT, Bolger MB, & Warshel A (1992) Calculations of antibody-antigen interactions: microscopic and semi-microscopic evaluation of the free energies of binding of phosphorylcholine analogs to McPC603. *Prot eng* 5(3):215-228.
44. Sham YY, Chu ZT, Tao H, & Warshel A (2000) Examining methods for calculations of binding free energies: LRA, LIE, PDLD-LRA, and PDLD/S-LRA calculations of ligands binding to an HIV protease. *Proteins* 39(4):393-407.
45. Kollman PA*, et al.* (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc Chem Res* 33(12):889-897.
46. Aqvist J, Medina C, & Samuelsson JE (1994) A new method for predicting binding affinity in computer-aided drug design. *Prot eng* 7(3):385-391.
47. Marcus RA (1964) Chemical + electrochemical electron-transfer theory. *Annu Rev Phys Chem* 15:155-&.
48. Aqvist J & Hansson T (1996) On the validity of electrostatic linear response in polar solvents. *J Phys Chem-Us* 100(22):9512-9521.
49. Almlof M, Carlsson J, & Aqvist J (2007) Improving the accuracy of the linear interaction energy method for solvation free energies. *J Chem Theo Comp* 3(6):2162-2175.
50. Almlof M, Brandsdal BO, & Aqvist J (2004) Binding affinity prediction with different force fields: examination of the linear interaction energy method. *J Comp Chem* 25(10):1242-1254.
51. Mishra SK, Sund J, Aqvist J, & Koca J (2012) Computational prediction of monosaccharide binding free energies to lectins with linear interaction energy models. *J Comp Chem* 33(29):2340-2350.
52. Pople JA (1999) Quantum chemical models (Nobel lecture). *Angew Chem Int Edit* 38(13-14):1894-1902.
53. Warshel A & Levitt M (1976) Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J Mol Biol* 103(2):227-249.
54. Kamerlin SC, Haranczyk M, & Warshel A (2009) Progress in ab initio QM/MM free-energy simulations of electrostatic energies in proteins: accelerated QM/MM studies of pKa, redox reactions and solvation free energies. *J Phys Chem. B* 113(5):1253-1272.
55. Rosta E, Klahn M, & Warshel A (2006) Towards accurate ab initio QM/MM calculations of free-energy profiles of enzymatic reactions. *J Phys Chem. B* 110(6):2934-2941.
56. Štrajbl M, Hong G, & Warshel A (2002) Ab Initio QM/MM simulation with proper sampling: "first principle" calculations of the free energy of the autodissociation of water in aqueous solution. *J Phys Chem B* 106(51):13333-13343.

57. Rosta E, Haranczyk M, Chu ZT, & Warshel A (2008) Accelerating QM/MM free energy calculations: representing the surroundings by an updated mean charge distribution. *J Phys Chem. B* 112(18):5680-5692.
58. Hwang JK, King G, Creighton S, & Warshel A (1988) Simulation of free-energy relationships and dynamics of sn2 reactions in aqueous-solution. *J Am Chem Soc* 110(16):5297-5311.
59. Warshel A (1982) Dynamics of reactions in polar solvents. Semiclassical trajectory studies of electron-transfer and proton-transfer reactions. *J Phys Chem* 86(12):2218-2224.
60. Marcus RA (1993) Electron transfer reactions in chemistry: theory and experiment (nobel lecture). *Angew Chen Int Ed* 32(8):1111-1121.
61. Liu H & Warshel A (2007) The catalytic effect of dihydrofolate reductase and its mutants is determined by reorganization energies. *Biochemistry* 46(20):6011-6025.
62. Shaw JJ, Trobro S, He SL, Aqvist J, & Green R (2012) A Role for the 2' OH of peptidyl-tRNA substrate in peptide release on the ribosome revealed through RF-mediated rescue. *Chem Biol* 19(8):983-993.
63. Hansson T, Marelius J, & Aqvist J (1998) Ligand binding affinity prediction by linear interaction energy methods. *J Comput Aided Mol Des* 12(1):27-35.
64. Bennet CH (1976) Efficient estimation of free-energy differences from monte-carlo data. *J Comput Phys* 22:245-268.
65. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, & Kollman PA (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J Comp chem* 13(8):1011-1021.
66. Fersht A (1999) *Structure and mechanism in protein science : a guide to enzyme catalysis and protein folding* (W.H. Freeman, New York) pp xxi, 631 p.
67. Kraut J (1977) Serine proteases: structure and mechanism of catalysis. *Ann Rev Biochem* 46:331-358.
68. Warshel A, Naray-Szabo G, Sussman F, & Hwang JK (1989) How do serine proteases really work? *Biochemistry* 28(9):3629-3637.
69. Štrajbl M, Florián J, & Warshel A (2000) Ab initio evaluation of the potential surface for general base- catalyzed methanolysis of formamide: a reference solution reaction for studies of serine proteases. *J Am Chem Soc* 122(22):5354-5366.
70. Outzen H, Berglund GI, Smalas AO, & Willassen NP (1996) Temperature and pH sensitivity of trypsins from Atlantic salmon (Salmo salar) in comparison with bovine and porcine trypsin. *Comp Biochem Physiol B; Biochem Mol Biol* 115(1):33-45.
71. Morley KL & Kazlauskas RJ (2005) Improving enzyme properties: when are closer mutations better? *Trends Biotechnol* 23(5):231-237.
72. Liebschner D, Dauter M, Brzuszkiewicz A, & Dauter Z (2013) On the reproducibility of protein crystal structures: five atomic resolution structures of trypsin. *Acta crystallogr Sect D: Biol Crystallogr* 69(Pt 8):1447-1462.
73. Leiros HK, McSweeney SM, & Smalas AO (2001) Atomic resolution structures of trypsin provide insight into structural radiation damage. *Acta crystallogr Sect D: Biol Crystallogr* 57(Pt 4):488-497.
74. Ghanem M, Li L, Wing C, & Schramm VL (2008) Altered thermodynamics from remote mutations altering human toward bovine purine nucleoside phosphorylase. *Biochemistry* 47(8):2559-2564.

75. Leiros HK, Willassen NP, & Smalas AO (1999) Residue determinants and sequence analysis of cold-adapted trypsins. *Extremophiles* 3(3):205-219.
76. Stoeckler JD, Cambor C, & Parks RE, Jr. (1980) Human erythrocytic purine nucleoside phosphorylase: reaction with sugar-modified nucleoside substrates. *Biochemistry* 19(1):102-107.
77. Wielgus-Kutrowska B, Bzowska A, Tebbe J, Koellner G, & Shugar D (2002) Purine nucleoside phosphorylase from Cellulomonas sp.: physicochemical properties and binding of substrates determined by ligand-dependent enhancement of enzyme intrinsic fluorescence, and by protective effects of ligands on thermal inactivation of the enzyme. *BBA* 1597(2):320-334.
78. Tebbe J*, et al.* (1999) Crystal structure of the purine nucleoside phosphorylase (PNP) from Cellulomonas sp. and its implication for the mechanism of trimeric PNPs. *J Mol Biol* 294(5):1239-1255.
79. Canduri F*, et al.* (2005) New catalytic mechanism for human purine nucleoside phosphorylase. *Biochem Biophys Res Commun* 327(3):646-649.
80. Erion MD, Stoeckler JD, Guida WC, Walter RL, & Ealick SE (1997) Purine nucleoside phosphorylase. 2. Catalytic mechanism. *Biochemistry* 36(39):11735-11748.
81. Porter DJ (1992) Purine nucleoside phosphorylase. Kinetic mechanism of the enzyme from calf spleen. *J Biol Chem* 267(11):7342-7351.
82. Fedorov A*, et al.* (2001) Transition state structure of purine nucleoside phosphorylase and principles of atomic motion in enzymatic catalysis. *Biochemistry* 40(4):853-860.
83. Mao C*, et al.* (1998) Calf spleen purine nucleoside phosphorylase complexed with substrates and substrate analogues. *Biochemistry* 37(20):7135-7146.
84. Bzowska A (2002) Calf spleen purine nucleoside phosphorylase: complex kinetic mechanism, hydrolysis of 7-methylguanosine, and oligomeric state in solution. *BBA* 1596(2):293-317.
85. Ghanem M*, et al.* (2008) Tryptophan-free human PNP reveals catalytic site interactions. *Biochemistry* 47(10):3202-3215.
86. Stoeckler JD*, et al.* (1997) Purine nucleoside phosphorylase. 3. Reversal of purine base specificity by site-directed mutagenesis. *Biochemistry* 36(39):11749-11756.
87. Suarez J*, et al.* (2013) Catalytic site conformations in human PNP by 19F-NMR and crystallography. *Chem Biol* 20(2):212-222.