Institute of Physical and Theoretical Chemistry
Department of Chemistry
Wrocław University of Technology

Modeling of enzymatic inhibition and catalysis within the framework of the theory of intermolecular interactions.

Edyta Dyguda-Kazimierowicz

Supervisor: Professor W. Andrzej Sokalski

A thesis submitted for the degree of Doctor of Philosophy
in the Wrocław University of Technology

· 2009 ·

ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere gratitude to my supervisor, Prof. W. Andrzej Sokalski, for the constructive advice and constant support. His understanding, encouraging, and thorough guidance have been invaluable at each stage of my doctoral research.

I am indebted to the people, who made this thesis possible by sharing their results with me or by providing a guidance. In particular, I would like to acknowledge Prof. J. Andrew McCammon, Dr. Caterina Ghio, and Dr. Jerzy Zoń. My warm thanks also goes to Dr. Adrian Mulholland and Dr. Jolanta Żurek for the generous advice and help on my phosphotriesterase research. I deeply appreciate the valuable collaboration of Prof. Jerzy Leszczyński and Prof. Sławomir Grabowski.

Finally, I offer my regards to all of those who supported me in any way during these years.

This work was partially funded by the L'Oréal Poland-Unesco *For Women in Science* Fellowship and British-Polish Young Scientists Programme. I am also grateful for the financial support from Wrocław University of Technology and Jackson State University.

CONTENTS

Glossary **v**

Preface and Goals **viii**

1 Computational Enzymology **1**

1.1 Enzyme Inhibition 2

1.1.1 Thermodynamics of binding 3

1.1.2 Prediction of binding affinity 5

1.2 Enzymatic Catalysis 8

1.2.1 Origins of the catalytic effects 10

1.2.2 Insights into the enzyme catalysis 12

1.3 Molecular Recognition 15

1.3.1 Physical nature of ligand binding 16

1.3.2 Differential Transition State Stabilization 22

1.3.3 Physical nature of hydrogen bonds 25

2 Enzyme Inhibitory Effects **28**
2.1 Phenylalanine Ammonia-lyase Inhibition 28
2.1.1 PAL characteristics 28
2.1.2 Inhibitors of potato PAL 30
2.1.3 Inhibitors of parsley PAL 32
2.1.4 Docking of parsley PAL inhibitors 35
2.1.5 Interaction energy analysis 40
2.2 Urokinase Inhibition 57
2.2.1 Urokinase characteristics 57
2.2.2 Inhibitors of urokinase 57
2.2.3 Interaction energy analysis 61

3 Enzyme Catalytic Activity **66**
3.1 Kinase-catalyzed Phosphorylation Reaction 66
3.1.1 Hydroxyethylthiazole kinase 66
3.1.2 cAMP-Dependent protein kinase 77
3.2 PTE-catalyzed Hydrolysis Reaction 86
3.2.1 Gas-phase mechanisms of basic hydrolysis of PTE substrates 88
3.2.2 Modeling of the PTE-sarin complex 102
3.2.3 QM modeling of PTE-catalyzed sarin hydrolysis 116
3.3 Intermediate Binding in TIM Catalysis 125
3.3.1 Properties of TIM active site hydrogen bonds 127

4 Summary **130**

List of Figures **136**

List of Tables **140**

Bibliography **142**

Appendix **169**

GLOSSARY

ABNR	Adopted Basis Newton-Raphson
AIM	atoms-in-molecules
AIP	2-aminoindane-2-phosphonic acid
AM1	the Austin model 1
***(S)*-AOPP**	*(S)*-2-aminooxy-3-phenylpropanoic acid
***(R)*-APEP**	*(R)*-1-amino-2-phenylethylphosphonic acid
ATP	adenosine triphosphate
BCP	bond critical point
BSSE	basis set superposition error
CCSD	coupled cluster theory with single and double excitations
CG	conjugate gradient
CoMFA	comparative molecular field analysis
CHELP	charges from electrostatic potential
CP	counterpoise correction
DFT	density functional theory

DTSS	Differential Transition State Stabilization
DFP	*O,O*-diisopropyl phosphorofluoridate
FEP	free energy perturbation
FMO	fragment molecular orbital
HAL	histidine ammonia-lyase
HF	Hartree-Fock method
HIV	human immunodeficiency virus
IC_{50}	half maximal inhibitory concentration
IRC	intrinsic reaction coordinate
KIE	kinetic isotope effects
LIE	linear interaction energy
MC	Monte Carlo
MD	molecular dynamics
MEP	molecular electrostatic potential
MFCC	molecular fractionation with conjugate caps
MIO	5-methylene-3,5-dihydroimidazol-4-one
MM	molecular mechanics
MMPBSA	molecular mechanics Poisson-Boltzmann surface area
MMGBSA	molecular mechanics generalized Born surface area
MP2	Møller-Plesset second-order perturbation theory
NAC	Near Attack Conformation
NMR	nuclear magnetic resonance
ONIOM	our own *n*-layered integrated molecular orbital molecular mechanics
PAL	phenylalanine ammonia-lyase

PCM	polarizable continuum model
PES	potential energy surface
PIEDA	pair interaction energy decomposition analysis
PKA	cAMP-dependent protein kinase
PM3	Parametric Method Number 3
PTE	phosphotriesterase
QM	quantum mechanics
QM/MM	quantum mechanics/molecular mechanics
QSAR	quantitative structure-activity relationship
RMSD	root mean square deviation
RMSF	root mean square fluctuation
SAPT	symmetry-adapted perturbation theory
SAPT(MP)	SAPT with electron correlation described at the Møller-Plesset level of theory
SCF	self-consistent field method
SD	steepest descent
SEE	standard error of estimate
ThiK	4-methyl-5-β-hydroxyethylthiazole kinase
Thz	4-methyl-5-β-hydroxylethylthiazole
TI	thermodynamic integration
TIM	triosephosphate isomerase
uPA	urokinase-type plasminogen activator
ZPE	zero-point energy

PREFACE AND GOALS

Noncovalent interactions, being an essence of molecular recognition between receptor and ligand molecules, determine a great variety of chemical and biological phenomena of contemporary interest [1]. Whether it is enzymatic transformation, signal transduction, molecular motors, or substance transport, the nature and strength of protein-ligand interactions controls an entire repertoire of processes essential for biological systems [2].

Due to possible application as drugs, herbicides or pesticides, enzyme inhibitors have received much attention [3, 4]. With the goal of achieving high binding affinity and selectivity, structure-based inhibitor design usually aims at ensuring substantial complementarity between a ligand and an enzyme binding site [5]. To meet the challenge of quantitative assessment of a protein-ligand fit, the methods allowing for reliable yet rapid ligand affinity prediction have been sought [6, 7]. However, this is done mainly by empirical force field based methods [2] which prevents deeper analysis of the interactions involved. When similar ligands are considered, the relative interaction energy emerges as the decisive factor in the discrimination of particular guest molecules, whereas entropic and solvation contributions could often be neglected [8]. Consequently, studies based on the comparison of interaction energies have proven their utility in the prediction of binding affinity [9–15].

The exceptional enhancement of reaction rates arising from enzymes action has been a matter of the utmost interest for over a century. Both fundamental and practical importance of unravelling the principal driving force in enzymatic catalysis has fuelled a longstanding debate about the origin of enzyme proficiency [16]. While several proposals have recently been put forward including desolvation, ground-state destabilization, and dynamic effects [17], it has been argued that what really contributes to the activation barrier lowering is the preferential transition

state binding relative to reactants [18–20]. In other words, the impact of an enzyme environment consists in its improved complementarity toward the transition state which is bound more tightly than the substrates. Thus, enzymatic catalysis exploits molecular recognition at the highest level of refinement [21], and, as such, could be investigated in terms of the theory of intermolecular interactions.

The understanding of enzymes catalysis and inhibition holds the key to the rational design of man-made catalysts as well as to the development of novel and more potent inhibitors. That the successful design of artificial enzymes is within the computational and conceptual limits, has recently been demonstrated by David Baker and coworkers [22, 23]. While impressive results were achieved in the cited contributions, there is still much room for improvement, as obtained rate enhancement only slightly exceeds the one provided by catalytic antibodies [24, 25].

The research described in this thesis was aimed at **application of the analysis of intermolecular interactions to the study of enzymatic inhibition and catalytic activity.** This is one of the first such undertakings in the literature. The availability of theoretically rigorous yet computationally tractable hybrid variation-perturbation scheme of interaction energy partitioning [26] allowed for both **the study of the nature of receptor-ligand interactions and for the systematic derivation of simplified models of inhibitory and catalytic activity by the gradual neglect of the active site residues and stabilization energy components of minor importance.**

In the case of enzymatic inhibition (Chapter 2) the attained goals included **obtaining reliable models of inhibitory activity as well as determination of the binding contribution arising from particular receptor site residues.** The results were validated by comparison with experimental inhibitor potency to establish the practical yet sufficiently accurate **level of approximation that could be employed in the prediction of binding affinity.** In contrast to QSAR methodology, inhibitory activity models obtained herein were derived solely from the first principles of quantum mechanics, as no empirical quantities were employed. Three enzymes were examined including phenylalanine ammonia-lyase (PAL) from *Solanum tuberosum* (6 inhibitors [27, 28]), PAL from *Petroselinum crispum* (11 inhibitors [29]; see section 2.1), and urokinase-type plasminogen activator (uPA) (5 inhibitors [30]; section 2.2). The structures of enzyme-inhibitor complexes of *Solanum tuberosum* PAL and uPA were prepared by Dr. Jolanta Grembecka and Renata Grzywa, respectively. The majority of the inhibitors studied herein were synthesized and studied at the Wrocław University of Technology by Dr. Jerzy Zoń (parsley PAL inhibitors [29]) and Dr. Marcin Sieńczyk (uPA inhibitors [30]). **The docking of 11 parsley PAL inhibitors**

followed by description of their mode of binding constituted an additional goal fulfilled here.

Enzymatic activity was the subject of Chapter 3. Differential Transition State Stabilization (DTSS) analysis [31, 32] applied to the phosphoryl transfer reaction catalyzed by 4-methyl-5-β-hydroxyethylthiazole kinase (ThiK) (section 3.1.1) and cAMP-dependent protein kinase (PKA) (section 3.1.2) allowed for **assessment of the catalytic contribution of active site residues and their possible role in a molecular mechanism of catalysis.** By examining the physical nature of interactions taking place in an enzyme active site, the reasonable **level of approximation was established that could further be employed in prediction of the influence of a mutation.** In particular, the dominant electrostatic contribution justified application of catalytic fields representing the optimal electrostatic characteristics of a molecular environment acting as a catalyst. Due to the lack of structural data concerning ThiK complexes with substrate, transition state and product, additional step performed here featured **modeling of ThiK-catalyzed reaction.** Analogous geometries for PKA-catalyzed process were obtained courtesy of Prof. J. Andrew McCammon [33]. DTSS analysis of PKA reaction was performed in collaboration with Dr. Paweł Szarek [34].

Due to the importance of phosphotriesterase (PTE) in biodegradation of toxic organophosphorus compounds and remarkable flexibility of PTE allowing for tailoring enzyme properties [35], **PTE catalytic mechanism** constituted another goal attempted here (section 3.2). Once fully understood, the molecular mechanism of PTE-catalyzed reaction can be of tremendous help in the rational control and design of a desired enzyme activity. In addition to the modeling of gas phase alkaline hydrolysis of several PTE substrates, this preliminary study provided a description of the substrate and reaction intermediate binding. Possible catalytic role of a zinc ion and a particular aspartate residue was also suggested on the basis of the catalytic fields approach.

Finally, a controversial hypothesis of Zhang and Houk [36] regarding possible covalent or partially covalent binding of enzymatic transition state and/or intermediate inspired detailed analysis of the nature of binding occurring in the active site of triosephosphate isomerase (TIM). The structure of TIM-reaction intermediate complex was kindly provided by Dr. Caterina Ghio [37], allowing for **the study of possible covalent characteristics of intermediate binding** (section 3.3).

CHAPTER 1

Computational Enzymology

Enzymes are complex proteins that act as catalysts by accelerating a great variety of chemical processes in comparison with the related reactions in solution. Most of chemical reactions occurring within living cells would be to slow to sustain life in the absence of enzymes. The crucial role of these amazing and sophisticated molecules is undeniable since a malfunction of even a single enzyme in a living organism can lead to a lethal illness. Although the effects of enzymes action are well described, the way they are achieved in most cases remains unclear.

For complete understanding of the factors that govern not only the binding of ligands in protein cavities, but most of all facilitate the subsequent reaction, it is essential to analyze the interactions within the enzyme active site. Although noncovalent intermolecular forces play a key role in a great diversity of molecular phenomena including the enzyme catalytic activity or the activity of enzyme inhibitors, their physical nature is not as thoroughly established as in the case of a well-understood chemical bond [38]. Theoretical methods of analysis are usually based on an extensive approximation of interactions employing empirical intermolecular potentials and, thus, do not allow for rigorous representation of physically well defined interaction energy contributions. However, accurate energy partitioning based on rigorous *ab initio* calculations was limited until recently by the size of investigated models [39], owing to the high cost of calculations which could scale (in the case of SAPT(MP)) as N^7 (N is the number of orbitals) [40]. Fortunately, recent developments in hybrid variation-perturbation interaction energy partitioning scheme [26] allowed for the investigation of considerably larger molecular complexes of the size of an inhibitor bound within an enzyme active site [13, 14, 31, 41]. Since application of sophisticated but computationally de-

manding tools for many biomolecular structures and related problems is still hardly feasible, it is tempting to employ some reasonable approximations. To do that, however, one should first compare the accuracy of results obtained at different levels of simplification. The energy decomposition scheme mentioned above defines a hierarchy of gradually simplified theoretical models and so provides an opportunity for testing hypotheses regarding the importance of particular interaction energy contribution and establishing the most appropriate level of theory for sufficiently correct description of a given phenomenon.

The following introduction provides general theory regarding enzyme inhibition and catalysis as well as molecular recognition. Special emphasis is given to the contribution of computational chemistry and molecular modeling techniques to the atomic-level understanding of basic principles underlying enzymes action. While more or less accurate data obtained via experimental methods refer to realistic systems, their interpretation is rarely straightforward as careful dissection of factors contributing to the final outcome is required in most cases. Noticeably, there is no single technique, neither experimental nor computational, that would be capable of providing full and precise knowledge concerning both structural and kinetic features of enzymes action [42]. Therefore, the multidisciplinary approach combining experimental and computational enzymology is essential for gaining a comprehensive and detailed insight into the enzyme-related phenomena.

1.1 Enzyme Inhibition

Inhibitors interact specifically with enzymes decreasing or even blocking their catalytic activity. The way it happens is by altering enzyme ability to bind substrate and/or hampering enzymatic activity [43]. Most therapeutic drugs act as enzyme inhibitors—the examples range from long known aspirin and penicillin to recently developed drugs exploited in cancer chemotherapy or treating HIV infection [44]. Many naturally occurring toxins (e.g., snake venom peptides) as well as chemical warfare agents (e.g., sarin) are also enzyme inhibitors [45]. Finally, some artificial inhibitors have been used as insecticides (e.g., parathion) or herbicides (e.g., glyphosate [46]). A variety of applications resulted in a great interest in the discovery and design of enzyme inhibitors. Since common experimental approaches involving trial and error screening of large libraries of drug-like compounds are time-consuming and expensive, rational inhibitor design taking advantage of structure-based information has recently become a standard method [47]. However, despite the vigorous development of numerous computational techniques for predicting the receptor-ligand structure and/or affinity, unsatisfactory results

have been obtained implying that some important aspects of specific ligand binding are still missing [48].

1.1.1 Thermodynamics of binding

The affinity between a ligand and a protein is commonly described in terms of the inhibition constant, i.e., the quantity expressing the equilibrium of enzyme-inhibitor complex dissociation (square brackets denote molar concentrations):

$$\mathrm{EI} \rightleftharpoons \mathrm{E} + \mathrm{I} \tag{1.1}$$

$$K_i = \frac{[E] \cdot [I]}{[EI]} \tag{1.2}$$

Since K_i corresponds to such a concentration of an inhibitor, $[I]$, at which the concentration of an enzyme-inhibitor complex, $[EI]$, equals the concentration of free enzyme, $[E]$, the smaller is the value of an inhibition constant, the larger inhibitory activity is observed. In the case of inhibitors considered as potential drugs, K_i value in the nanomolar range is regarded as sufficient for further development [44]. One of the most potent noncovalent protein-ligand interactions occurs in the avidin-biotin complex, as the corresponding inhibition constant equals 10^{-15} M [49].

The following formula relates the equilibrium constant defined in Eq. 1.2 and the standard Gibbs free energy change:

$$\Delta G^\circ = -RT \ln K_i \tag{1.3}$$

In the above equation, R and T refer to the gas constant and the absolute temperature, respectively. The Gibbs free energy of binding results from enthalpic (ΔH°) and entropic ($T\Delta S^\circ$) contributions:

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ \tag{1.4}$$

Therefore, the propensity of two monomers to associate with each other is driven by two opposing fundamental effects—a tendency to lower the energy of the system (e.g., by formation of noncovalent bonds associated with negative ΔH) compensated by a tendency of thermal motions to disrupt the dimer structure (positive ΔS). When the binding occurs in a solution, the overall description of the association process gets more complicated, as solvent molecules interacting with both monomers have to be replaced upon complexation. Accordingly, ΔH reflects the strength of the binding between interacting partners relative to the analogous interactions with solvent molecules and ΔS includes both the change in solvation

entropy (the increase in entropy of solvent molecules released upon complex formation) and the decrease in conformational entropy accompanying the partial loss of dimer conformational freedom.

Complete knowledge of the factors underlying molecular recognition requires both enthalpic and entropic contributions to be assessed. A common way of binding affinity measurement is via indirect methods employing the changes in a physical property recording the influence of ligand concentration on the enzyme kinetics (e.g., absorption or fluorescence of one of the reaction partners) [2]. The binding enthalpy could be further obtained using van't Hoff plots of temperature dependence of the binding affinity. Direct experimental determination of ΔH and ΔS can be accomplished by microcalorimetric measurements (e.g., isothermal titration calorimetry [50]), however, such results are still scarce [8]. While of utmost interest for the detailed understanding of binding principles, experimental methods providing thermodynamic binding parameters are impractical for routine testing of vast libraries of compounds. To allow for rapid *in silico* determination of putative binding characteristics of a hypothetical ligand, theoretical methods of predicting molecular recognition have been developed [51] (see the next section).

Binding affinity (or inhibitory activity) is used in reference to both inhibition constant, K_i and binding free enthalpy, ΔG (as these two quantities are related via Eq. 1.3). In certain cases, half maximal inhibitory concentration (IC_{50}) quantity is also employed as an experimental measure of the inhibitory activity. By definition, IC_{50} corresponds to the inhibitor concentration decreasing an enzyme activity by half. In contrast to K_i, IC_{50} values are dependent on conditions under which they are measured including the concentration of an enzyme. In the case of a competitive inhibition (i.e., inhibition occurring via the ligand binding within an enzyme active site), the concentration of an enzyme substrate (if present) also affects the IC_{50} value. In such a case, IC_{50} quantity could be related to inhibition constant by the equation [52]:

$$K_i = \frac{IC_{50}}{1 + \frac{[S]}{K_M}} \tag{1.5}$$

where $[S]$ and K_M denote, respectively, the substrate concentration and the Michaelis constant (see Eq. 1.7 in section 1.2). Irrespectively of the type of inhibition, IC_{50} values for a set of inhibitors measured under identical conditions provide a description of relative binding affinity analogous to that given by inhibition constant.

1.1.2 Prediction of binding affinity

Theoretical approaches employed in binding affinity prediction encompass two general categories of methods differing in the requirement for the known three-dimensional structure of a receptor. In the absence of an enzyme structure, the prediction of binding affinity is performed under the assumption that ligands with similar physico-chemical properties trigger similar biological response [53]. A correlation between biological properties and certain physico-chemical and structural ligand parameters employed in the quantitative structure-activity relationship (QSAR) methods allows for quantitative predictions regarding the binding affinity [54]. In particular, the comparative molecular field analysis (CoMFA) method, taking advantage of the three-dimensional structure of ligand molecules, has gained considerable popularity due to improved precision of forecasts [55]. However, the quality of the CoMFA results depends heavily on the accurate superposition of the ligands, which usually cannot be verified due to the absence of receptor structure [53]. Other limitations arise from the use of force field parameters as well as the inherent inability of providing the outcome interpretable in terms of physico-chemical factors underlying ligand-protein interactions, i.e., the results yielded by CoMFA have only statistical meaning [56].

Assessment of ligand binding thermodynamics through the approaches relying on the knowledge of a target protein structure fall into the following categories [57]:

1. molecular simulation based methods
2. empirical/force field/additivity techniques
3. knowledge based approaches
4. hybrid protocols

Molecular simulation based methods The approaches belonging to this category aim at the direct calculation of relative binding free energies via rigorous free energy perturbation (FEP) or thermodynamic integration (TI) protocols derived from statistical mechanics [58, 59]. The binding free energy is calculated from the ensemble average of an energy function describing a given system. Configurational space is sampled during molecular dynamics (MD) or Monte Carlo (MC) simulation, usually with explicit consideration of solvent. In principle, only the differences in binding affinities are predicted, however, calculation of the absolute receptor-ligand binding free energy is also possible [60]. Due to the exact treatment of essentially all aspects of protein-ligand binding, the FEP and TI methods have

provided valuable insight into individual contributions to free energy on an atomic level. However, a basic requirement here is the availability of reliable force field parameters, as the system under study is described at the molecular mechanics level. Since the structures of most inhibitor candidates usually lack valid parameters, an additional effort that needs to be undertaken encompasses the development and testing of the missing force field parameters. Finally, conventional molecular mechanics description of the system is not capable of capturing some fundamental phenomena of quantum nature (e.g., polarization). While first applications of semiempirical quantum mechanics based FEP simulation have already been reported [61], complexity and low throughput of these methods severely limits their application.

Additivity based approaches The basic assumption behind these methods is the partitioning of the free energy of ligand-receptor binding into a sum of individual contributions [62]. In particular, the free-energy contributions are no longer associated with the ensemble average values, but derived from model structures/calculations. The force field scoring functions (e.g., AutoDock [63]) approximate the binding free energy by a sum of van der Waals, electrostatic and other contributions. For a series of related compounds, in which the entropic term can be assumed to remain constant, considering the enthalpic contribution only (i.e., electrostatic and van der Waals interactions), appears to be sufficient to obtain correlation with experimental data [2]. Nevertheless, most scoring functions depend on some sort of empirical estimates of binding entropy and solvation. The empirical scoring functions (e.g., FlexX [64], Gold [65, 66]) rely on determination of the individual contributions of the separate terms by means of multiple linear regression, partial least-squares regression or a neural network approach, employing a training set of protein-ligand complexes with known structures and binding affinities [2]. Apparently, the predictive ability of such models depends significantly on the compilation of the training set. In contrast to the FEP/TI methods capable of providing near-quantitative description of binding trends, scoring functions usually provide qualitative results only.

Knowledge based approaches Prediction of the binding affinity via knowledge base approaches stems from the statistical analysis of a large number of protein-ligand complexes deposited in the structural databases [67]. According to the concept of the "inverse Boltzmann law" [68], interatomic interactions analyzed in terms of their frequency distributions derived from experimental structures are converted into the knowledge based potentials (e.g., BLEEP [69]). The performance of these approaches is comparable to that of empirical scoring functions.

Hybrid approaches Due to the intrinsic limitations of the methods discussed above, i.e., vast computational expense of FEP/TI protocols and insufficient accuracy of scoring functions, another approach has been tested employing a combination of molecular simulation and additivity approximation. Hybrid approaches aiming at a quick but reasonably precise binding affinity estimation include linear interaction energy (LIE) [70], MMPBSA and MMGBSA [71]. Binding energetics within these methods is predicted using the configurations of a protein, a ligand, and a protein-ligand complex obtained from the molecular dynamics in the associated and dissociated states (in FEP/TI methods the gradual conversion between these states is also considered). Despite a potential of yielding accurate results [72], hybrid approaches also suffer from the necessity of using molecular mechanics level of theory.

While the entropy term remains an important part of absolute free energy of binding, various entropic contributions have been reported depending on a somewhat arbitrary ways of entropy partitioning into different components [73–75]. When applied to a series of related ligands, inhibitory activity analysis could be performed on the basis of relative interaction energies only [2, 8]. Apparently, the differences in electronic binding energy within a set of similar inhibitors become a dominant contribution to the relative binding affinity. As a consequence, strong relationship concerning interaction energy, ΔE, and $\ln K_i$ (or $\log K_i$) is observed [9–15]. Noticeably, electronic interaction energy (at the semiempirical level of theory) has been shown to provide a better correlation with experimental binding affinity (within a set of ligands interacting with a common protein target) than the full parametrized scoring function including, in addition to this electronic interaction energy, a variety of terms estimating, for instance, the binding entropy and solvation [76].

Large scale full *ab initio* computation of the interaction energy could, in principle, be performed with the fragment-based methods, e.g., molecular fractionation with conjugate caps (MFCC) [77] or fragment molecular orbital (FMO) [78] approaches. The examples include MFCC-based binding energy calculation within the complexes of streptavidin-biotin [79], β-trypsin/benzamidine [80], and human immunodeficiency virus (HIV)-1 protease inhibitors [81] or FMO determination of the interactions involving human estrogen receptor and its ligands [15, 82]. Apparent advantages of these approaches encompass the potential of employing accurate *ab initio* methods and the ability of analysing global interaction patterns, as binding of all the enzyme residue-ligand pairs is considered. Nevertheless, little knowledge is gained in terms of the physical nature of interactions. While pair interaction energy decomposition analysis (PIEDA) scheme [83, 84] developed recently within the framework of FMO approach is capable, in general, of providing such an information, the Kitaura and Morokuma energy decomposition

method [85] implemented in PIEDA does not introduce a well defined hierarchy of the interaction energy terms [86], and, thus, systematic derivation of approximate models cannot be attained. Moreover, components of the interaction energy derived from Kitaura and Morokuma scheme are contaminated by basis set superposition error (BSSE). On the other hand, variation-perturbation partitioning [26] employed in this work is free of the BSSE and defines a set of models allowing for the rational construction of approximate, first principles based models of inhibitory activity [13, 14].

An example of a simple, yet instructive model is the electrostatic lock-and-key analogy proposed by Náray-Szabó [11]. The approximation of enzyme active site as an electrostatic lock enabled simple but effective way to compare a series of inhibitors with respect to the fit between the lock (a charge pattern of active site residues) and the key (a molecular electrostatic potential of inhibitors).

1.2 Enzymatic Catalysis

The phenomenological description of a general enzymatic process takes advantage of the kinetic scheme proposed by Michaelis and Menten [87], whereby an enzyme-catalyzed reaction proceeds with the formation of an enzyme-substrate complex, ES, followed by the actual reaction, i.e., product formation:

$$\mathrm{E} + \mathrm{S} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{ES} \xrightarrow{k_{cat}} \mathrm{E} + \mathrm{P} \tag{1.6}$$

The ES complex formation is described by the apparent dissociation constant, so called Michaelis constant:

$$K_M = \frac{k_{-1} + k_{cat}}{k_1} \tag{1.7}$$

By definition, K_M value corresponds to the substrate concentration, at which half of the available enzyme active sites is occupied or, in other words, the rate of enzymatic reaction, V, is equal to the half of a limiting value ($V = \frac{1}{2}V_{max}$) [43, 88]. It appears, that enzymes have evolved to exhibit K_M values within the range describing the physiological concentration of their substrates [1].

Under the conditions of high (saturating) substrate concentration, the initial reaction rate, V_0, depends on the rate constant and enzyme concentration:

$$V_0 = V_{max} = k_{cat}[E] \tag{1.8}$$

The catalytic constant k_{cat}, i.e., the first-order rate constant for the chemical transformation of the ES complex to the products, is often called the enzyme turnover

number, as it represents the maximum number of substrate molecules converted to products per unit time per active site.

At low (subsaturating) substrate concentration (i.e., under conditions close to physiological), V_0 is given by the equation:

$$V_0 = \frac{k_{cat}}{K_M}[E][S] \tag{1.9}$$

In this case, k_{cat}/K_M is an apparent second-order rate constant. From the Eq. 1.7, the upper limit of the k_{cat}/K_M ratio can be derived, as it cannot exceed the k_1 rate constant, i.e., the rate of diffusion-controlled enzyme-substrate encounter. Accordingly, enzymes exhibiting k_{cat}/K_M value in the range of 10^8 to 10^9 $s^{-1} \cdot M^{-1}$ (e.g., triosephosphate isomerase) are referred to as "diffusion-limited" [88].

The value of k_{cat}/K_M is often employed as a measure of the catalytic efficiency of enzymes. However, most enzymes operate with k_{cat}/K_M value of about 10^7 $s^{-1} \cdot M^{-1}$ [89]. Due to relatively narrow range of K_M values, the respective turnover numbers are also similar (k_{cat} in the range of 10^2 to 10^3 s^{-1} [90]). As the spontaneous rates of uncatalyzed biological reactions in a neutral solution vary considerably, the comparison of enzyme efficiencies should account also for the rate of the corresponding uncatalyzed reaction, k_{uncat}. Thus, the appropriate indication of the catalytic power of enzymes is the dimensionless k_{cat}/k_{uncat} ratio, i.e., the rate enhancement representing the factor by which an enzyme's affinity toward the reaction transition state is greater relative to the ground state [89]. The k_{cat}/k_{uncat} ratio reflects the actual differences in enzymes ability to accelerate a given reaction, as the rate enhancements they provide ranges from 10^7 to 10^{19}-fold [90].

The more suitable description of enzymes power under physiological conditions is provided by $(k_{cat}/K_M)/k_{uncat}$ (M^{-1}), i.e., the catalytic proficiency. The reciprocal of this quantity corresponds to the dissociation constant of the enzyme-transition state complex and, therefore, it can be employed as a benchmark for the design of potential transition-state analogue inhibitors [89, 90]. Accordingly, these values are within the range of 10^{-8} to 10^{-27} M (or -11 to -38 kcal $\cdot$ mol^{-1} in terms of binding free energy) [36], setting an upper limit on the dissociation constant of the enzyme-transition state complex.

1.2.1 Origins of the catalytic effects

Within the framework of the transition state theory, the rate constant for a reaction is determined by the Eyring equation [91]:

$$k(T) = \gamma(T)\frac{k_B T}{h}\exp\left(-\frac{\Delta G^{\ddagger}(T)}{RT}\right) \tag{1.10}$$

where $\Delta G^{\ddagger}$ denotes the free energy of activation (i.e., the difference in free energy between the transition state, $X^{\ddagger}$, and the ground state, X) and γ is the transmission coefficient. The reaction rate can thus be affected by the perturbations of the activation free energy and/or the transmission coefficient. Apparently, an exponential dependency of the rate constant on the activation free energy results in a large contribution of the latter to the changes in a reaction rate.

The transmission coefficient is a measure of the dynamic catalytic effects including recrossing, quantum mechanical tunneling and deviations of the equilibrium distribution in phase space [21]. As these phenomena are interesting in terms of the rigorous, quantitative description of the enzyme-catalyzed reaction, they are rarely accounted for due to their minor contribution and prohibitively large computational cost. One exception applies to quantum tunneling, which has been confirmed both experimentally and theoretically to occur in the enzymatic reactions, especially those involving the proton transfer [92]. It remains an open question, however, if the enzymes are capable of contributing to the intensity of tunneling or it is an inherent feature of the catalyzed reaction [93].

Probably the first explanation of the origins of enzyme power—the lock-and-key analogy—was proposed by Fisher in 1894 [94]. As stated by Polanyi in reference to the catalytic processes occurring at the surfaces [95], the reaction rate enhancement might result from the transition state being adsorbed more strongly compared with the reactants. The connection between the enzymatic catalysis and the concept of transition state stabilization was then made by Pauling [96], twelve years before the first experimental structure of the protein came up [97]. In terms of the thermodynamic quantities, this effect can be expressed as [20]:

$$\Delta G^{\ddagger}_{uncat} - \Delta G^{\ddagger}_{cat} = \Delta G^{ES}_{bind} - \Delta G^{TS}_{bind} \tag{1.11}$$

For the left-hand side of the above equation to be positive, the transition state of the enzyme-catalyzed reaction needs to be stabilized with respect to the corresponding ground state. Preferential transition state stabilization as a source of the enzymatic activity has been put forward in a number of contributions [18–20]. In particular, it has been argued that this effect is electrostatic in na-

ture [19, 21, 98–100]. According to the proposition of Warshel and coworkers [19, 101], the electrostatic effects play a major role in the enzyme catalysis due to the favourable electric field within an enzyme active site that is optimal to accommodate the transition state charge distribution. Unlike the charge fluctuations in a water environment, fixed charge distribution of the enzyme catalytic site is already preorganized to match the electrostatic properties of the transition state.

Another way of thinking of the factors underlying enzymes power refers to the concept of ground state destabilization [102]. While the overall transition state characteristics seem to be common for the catalyzed and uncatalyzed reactions, bringing together fully separated and solvated reactants to the spatial arrangement analogous to that featured by an enzyme-substrate complex might require additional energy. The corresponding ΔG^{ES}_{bind} quantity could then include, in addition to the contribution resulting from an enzyme-substrate interaction ($\Delta G^{ES'}_{bind}$), the energy change due to the conformational change of a substrate upon binding within an enzyme active site, $\Delta G^{ES'}_{R}$ [20]:

$$\Delta G^{ES}_{bind} = \Delta G^{ES'}_{R} + \Delta G^{ES'}_{bind} \tag{1.12}$$

where ES' denotes the hypothetical ES-like structure of reactants in solution (a reactive conformation). Substituting Eq. 1.12 into Eq. 1.11 results in the following relationship:

$$\Delta G^{\ddagger}_{uncat} - \Delta G^{\ddagger}_{cat} = \Delta G^{ES'}_{R} + \Delta G^{ES'}_{bind} - \Delta G^{TS}_{bind} \tag{1.13}$$

Due to positive value of $\Delta G^{ES'}_{R}$, lowering of the activation energy could be obtained without a requirement for the higher enzyme affinity toward the transition state [20]. Instead of focusing on the enzyme preorganization to accommodate the transition state, the discussion is then focused on the enzyme-driven substrate preorganization. Such a ground state destabilization could arise from entropic, strain and desolvation effects [102]. According to the NAC proposal of Bruice et al. [103], enzymes act by populating the ground state conformations essentially resembling the transition state. However, it has been argued that ground state destabilization is simply a consequence of geometrical and electrostatic complementarity of an enzyme active site to the transition state [20]. It is then because of the high affinity to the transition state, that the reactive substrate conformations are favoured. Finally, the amount of ground state destabilization cannot be large as the increase in the k_{cat} value is accompanied by the increase in K_M, leaving the catalytic efficiency, k_{cat}/K_M, unaffected [104].

1.2.2 Insights into the enzyme catalysis

The difficulty in experimental dissecting different catalytic contributions makes the problem of explaining the origin of enzymatic catalysis even more complicated. The kinetic measurements and structural studies supply either general information about mechanism of an enzyme or concern only ground-state properties of inactive complexes and so the structural changes along the reaction path remain unknown. It is then clear that numerous questions regarding enzymatic catalysis cannot be uniquely addressed by current experimental approaches [42]. On the other hand, the use of computer simulations turned out to be an effective way of exploring mechanisms of enzyme action [105–108]. A detailed description of enzyme catalyzed reaction at the atomic level is not, however, a trivial task as the size of models limits the accuracy of employed methods. In view of the points listed above, the combination of theoretical and experimental information seems to be much more successful in revealing the way of enzymes action. Along with the increase in speed and availability of computers it is possible to address a growing number of catalysis aspects. Taking things altogether, the elucidation of enzyme catalytic power is therefore presumably one of the most important and challenging problems in current computational chemistry and molecular modeling field.

Basic prerequisite for the understanding of the principles of enzymes efficiency is their structure. The experimental methods most commonly employed include X-ray crystallography and NMR spectroscopy. Despite obvious advances resulting from the wealth of experimental structures deposited in public databases [109], certain limitations are inherent for these techniques. The conventional X-ray crystal structures provide a static information only. While this deficiency could be compensated for by means of time-resolved crystallography [110], this method has not yet been widely applied. In a crystal structure determined at 2.0 Å resolution (commonly accepted as a structure of sufficient quality), the average errors in atomic positions exceed 0.3 Å [8]. Finally, the positions of hydrogen atoms cannot usually be resolved and, thus, uncertainty remains with respect to the ionization state of titratable groups. The widespread use of neutron protein crystallography providing the exact positions of hydrogen atoms is prohibited by the requirements regarding the size and quality of crystals as well as long data-acquisition times [111]. NMR structures are generally less accurate compared to X-ray geometries, however, they represent structures in solution (under physiological instead of artificial conditions imposed by crystal lattice).

A powerful technique providing insight into the enzyme function is protein engineering [88]. In particular, site-directed mutagenesis has enabled enzymologists not only to identify residues responsible for the substrate specificity and catalysis, but also to quantify a given residue's contribution to the substrate and/or transi-

tion state binding [112]. Remarkably, mutagenesis study allows for the strength of a particular hydrogen bond to substrate and transition state to be measured with an accuracy of ~ 0.1 kcal $\cdot$ mol^{-1} [113], while the best experimental estimate [114] of a water dimer association energy is equal to -5.44 ± 0.7 kcal $\cdot$ mol^{-1} [115]. Nevertheless, the results of site-directed mutagenesis might be difficult to interpret [101, 116], as mutations could affect both the enzyme structure and mechanism [117]. By combination of structural and mutational data with computer simulations, this problem can be surmounted illuminating the actual factors underlying the catalytic events [100]. Important contribution to interpretation of the mutagenesis effects can be made with the use of Differential Transition State Stabilization (DTSS) approach (see section 1.3.2). In addition to quantifying the catalytic contributions of particular active site residues, DTSS method allows the physical nature of these effects to be assessed in terms of the intermolecular interaction theory [31].

Experimental determination of enzymatic transition states, vital to understanding of the principles of catalysis and the design of transition state analogs [118, 119], constitutes a real challenge due to elusive nature of these transient species. Important insight into the structural features of transition states can be gained from the kinetic isotope effects (KIE) [120]. Assuming that the rate-determining step involves the formation and/or cleavage of a bond, the rate constant of the reaction will be affected by isotopic labelling. Remarkably, isotopic substitution provides the most subtle perturbation, as it arises from the different number of neutrons which (within the Born-Oppenheimer approximation) affects only the energetics, leaving the potential energy surface unchanged [121]. The primary kinetic isotope effects result from the breakage of a bond to the isotopically substituted atom. Owing to the increased atomic mass, the frequency of a vibration associated with the reaction coordinate is decreased (or, viewed quantum mechanically, the corresponding zero-point energy is lower). As a consequence, the activation energy barrier is elevated and the reaction involving a heavier isotope becomes decelerated. For instance, the cleavage of a C-D bond is 7 times slower compared to that of C-H bond (the actual KIE might be within a range of 2–15 due to possible quantum tunneling or compensation by bending motions in the transition state [88]). Smaller isotope effects are expected upon the replacement resulting in a smaller relative change in the atomic mass (e.g., ^{14}N by ^{15}N and ^{16}O by ^{18}O substitutions). Nevertheless, the magnitude of KIE gives a clue about the relevance of bond formation or breakage to the proposed reaction mechanism. In the case of alkaline and phosphotriesterase-catalyzed hydrolysis of paraoxon (see section 3.2), analysis of the primary and secondary ^{18}O kinetic isotope effects confirmed the expected associative mechanism and shown the transition state similarity when passing from solution to the enzyme-catalyzed process [122]. Of the two alternative mechanisms

proposed for the PAL-catalyzed deamination of *(S)*-phenylalanine (i.e., elimination (E_1cB) [123] or Friedel-Crafts type [124, 125] reaction), secondary hydrogen kinetic isotope effects arising from tritium ortho-substitution of a substrate ring have been interpreted as an indication of Friedel-Crafts-like mechanism [126]. The latter has recently been confirmed by means of kinetic measurements and computer simulation [127].

The power of theoretical methods of computational enzymology [105–108] consists in their ability of providing a uniquely detailed insight into the enzyme reactions. Different reaction mechanisms can be tested along with the hypotheses explaining the sources of catalytic effects. However, due to inherent limitations of particular computational approaches, careful examination of the results should involve their comparison with available experimental data. Studying of an enzymatic reaction is particularly demanding, as the requirement for an accurate quantum mechanical description of a chemical reaction contradicts the approximations that have to be invoked for a cost-effective treatment of a large molecular system including the solvated enzyme-ligand complex. While molecular mechanics (MM) methods featuring a classical, force field based description of the potential energy surface, are fast enough to allow for the MD study of a dynamical behaviour of biomolecules [128], they cannot be applied to the study of any process involving bond making or breaking [129]. On the other hand, reliable description of a reaction demands versatility and precision of high-level quantum mechanics (QM) methods (i.e., *ab initio* molecular orbital or density functional approaches), and these are hardly feasible for the systems composed of more than 200 atoms [42]. Hybrid quantum mechanics/molecular mechanics (QM/MM) approaches attempt to overcome these shortcomings by dividing the modelled system into two regions: a small part where chemical transformation takes place (the QM region) and the remaining environment treated classically [130, 131]. Since the first application in 1976 [132], QM/MM method has become the state-of-the-art computational technique for simulations involving chemical process occurring in large systems. Depending on the desired chemical accuracy and the available computational resources, various QM levels are applied within the QM/MM scheme—from semiempirical methods (e.g., AM1 [133], PM3 [134]), to HF and DFT. Semiempirical level of theory, while suitable for larger systems and fast enough to allow for MD or MC sampling, is generally not accurate enough to give quantitatively meaningful results. Greater precision of DFT methods comes with a considerably increased computational cost over semiempirical methods. Despite the facts, that crucial physical interactions (e.g., dispersion) are mostly not properly represented by the DFT theory and possible errors in the calculations cannot be systematically improved, DFT is usually the method of choice due to advantageous cost/accuracy ratio [130]. For the predictions to be quantitatively reliable, high level *ab ini-*

tio methods are required, such as MP2 or CCSD. However, these methods are prohibitively expensive owing, essentially, to the delocalized character of molecular orbitals [129]. Recent advancement in linear-scaling local versions of high level electron-correlation methods allowed for chemically accurate (i.e., featuring the $\sim$ 1 kcal $\cdot$ mol^{-1} error) QM/MM calculations of enzyme-catalyzed reaction barriers [135]. Nevertheless, due to arbitrary character of the procedure of wavefunction localization, the errors reported in Ref. [135] could be underestimated.

While crucial for modeling of the enzymatic reactions, QM/MM methods provide little insight into the physical nature of the preferential transition state stabilization. The most feasible approach encompassing calculation of the enzyme-transition state/enzyme-reactant binding energies by means of electrostatic and van der Waals contributions [33] provides only approximate results. Full *ab initio* assessment of the catalytic effects by means of the fragment-based methods [136] is not capable of capturing the fine details regarding the nature of interactions. As discussed in the final paragraphs of section 1.1.2, this problem could be surmounted with the use of variation-perturbation scheme of interaction energy decomposition [26]. In particular, such an approach applied to the study of ribonuclease A [137] and chorismate mutase [31] allowed for the assessment of the interaction energy terms playing the dominant role in terms of the differential transition state stabilization.

1.3 Molecular Recognition

Given a fundamental importance of the molecular recognition resulting from the strength and specificity of ligand binding, the process of complex association has remained an active area of theoretical research. Since the experimental measurements provide Gibbs free energy of binding, quantification of interactions requires both the enthalpic and entropic contributions to be calculated. On the other hand, when similar set of ligands is considered, they are assumed to exhibit comparable characteristics in terms of desolvation, protein (ligand) reorganization energy and the conformational entropy loss. In fact, these effects are not regarded as negligible, only that they are mostly relatively constant across the series. Thus, computationally-demanding study of association free enthalpy may be replaced by the more affordable analysis of a binding energy that constitutes the most characteristic contribution to the observed binding affinity.

1.3.1 Physical nature of ligand binding

All the forces between either charged or polar and even nonpolar atoms and molecules arise in essence from electrostatics [138, 139]. In general, the two classes of interactions can be introduced as a result of both long- and short-range effects. Among the former are the electrostatic, induction and dispersion interactions, where the energy at large distances is proportional to some inverse power of the intermolecular separation. Either attractive or repulsive electrostatic term originates basically from the interaction of static charge distributions of isolated molecules and is exclusively pairwise additive. The electron density distortion of a given molecule caused by the electric field of its neighbours (polarization) and the resulting stabilizing interactions of induced multipole moments is responsible for induction effects. Finally, the dispersion contribution, especially important in the case of interactions between nonpolar molecules (where neither electrostatic nor induction effects are dominant) is associated with the attractive interaction of instantaneous charge fluctuations arising from correlated electron movements. Both induction and dispersion terms are non-additive and so are the remaining short-range effects. When the separation of the two molecules is small, the significant overlap of monomers' molecular wavefunctions takes place causing the repulsive effect characterized by energy diminishing exponentially with the intermolecular distance—exchange repulsion interaction, which results from the Pauli exclusion principle. The importance of revealing the physical nature of interactions in terms of the perturbation theory arises from the fact, that since essentially the same types of interactions govern all the range of molecular complexes (e.g., hydrogen bonded, van der Waals, donor-acceptor, and dihydrogen bonded complexes), the reason for observed diversity is basically the different contribution of particular interaction energy components. Thus, elucidation of some more general and fundamental principles related to the role of specific components is undoubtedly of special interest.

There are two general approaches allowing for the evaluation of binding energy. Probably the most intuitive one is to subtract the energies of isolated monomers from the total energy of a dimer (the supermolecular method [86]), assuming the same internal coordinates of monomers and dimer are employed:

$$\Delta E_{int} = E_{AB} - E_A - E_B \tag{1.14}$$

The apparent advantage of such a definition is its simplicity and possible application in the cases of strongly interacting species. Computationally expensive yet accurate methods as well as simplified algorithms can be employed increasing the range of applications. The energy from Eq. 1.14 represents the overall interaction effect, but little, if any, knowledge about particular energy contributions can

be gained. Furthermore, as ΔE_{int} is orders of magnitude smaller than monomer energies, the required precision makes calculations of large systems extremely costly. Basic requirement in any application of the supermolecular method consists of treating all system configurations in identical, size-consistent manner. However, this rule is not followed if basis sets of different size are applied for each of the interacting molecules - which results in so called basis set superposition error (BSSE). For instance, if one intends to reveal the interaction energy of a dimer by first calculating its total energy, then obtaining the energies of the two molecules forming dimer separately (in their monomer basis sets) and finally subtracting this two values from the first calculated energy, one will find that such an interaction energy will be an overestimate of the true value. This happens because each monomer within the dimer has a larger number of basis functions available as it takes advantage of another monomers basis functions. According to the variational principle, the overall dimer energy lowering with respect to the isolated species is then not only on account of favourable intermolecular interactions but also due to the the more flexible description of the wavefunction. An obvious way of correcting for this non-physical error, i.e., the use extremely large basis sets, is hardly practical for most of the systems investigated. Another approach commonly employed when dealing with BSSE is to calculate monomers energies in the basis set of a dimer (counterpoise correction of Boys and Bernardi [140]).

In revealing the nature of interactions in a nonempirical manner, the accurate contribution of particular components to the binding energy is most important, whereas the total energy obtained within supermolecular method is insufficient for understanding of the underlying physical forces. The interaction energy partitioning is enabled with an use of perturbational approach, where the intermolecular interaction is introduced as a perturbation into the $A + B$ supermolecule Hamiltonian and subsequent corrections to the initial energy of an isolated molecules unperturbed state can be attributed a particular physical meaning.

The unperturbated Hamiltonian operator of the two non-interacting atoms or molecules (treated as exactly solvable system) can be written as the sum of the Hamiltonians of isolated systems:

$$\hat{H}^{(0)} = \hat{H}_A + \hat{H}_B \tag{1.15}$$

According to the Rayleigh-Schroedinger Perturbation Theory (RSPT), the relative weakness of intermolecular interactions allows for treating this effect as a perturbation and hence expressing the final state of interacting components A and B as the perturbated Hamiltonian:

$$\hat{H} = \hat{H}^{(0)} + \hat{H}^{(1)} \tag{1.16}$$

where perturbation operator $\hat{H}^{(1)} = \hat{V}$ represents their interaction. After introducing the polarization approximation, the unperturbated state is assumed to be a simple product function of monomers wavefunctions (subscript "0" at eigenfunctions denotes the electronic ground state):

$$\psi_0^{(0)} = \psi_{A,0}\psi_{B,0} \tag{1.17}$$

Here $\psi_0^{(0)}$ designates an eigenfunction of unperturbated Hamiltonian, $\hat{H}^{(0)}$:

$$\hat{H}^{(0)}\psi_0^{(0)} = E_0^{(0)}\psi_0^{(0)} \tag{1.18}$$

Since

$$\hat{H}_A\psi_{A,0} = E_{A,0}\psi_{A,0} \tag{1.19}$$

$$\hat{H}_B\psi_{B,0} = E_{B,0}\psi_{B,0} \tag{1.20}$$

one gets the unperturbated energy called zeroth-order energy as the sum of monomer energies:

$$E_0^{(0)} = E_{A,0} + E_{B,0} \tag{1.21}$$

As mentioned above, the overall effect of interaction is expressed as an infinite sum of successive physically meaningful higher-order corrections to the energy:

$$\Delta E_{int} = \sum_{n=1}^{\infty} E_0^{(n)} \tag{1.22}$$

Accordingly, the first-order energy being the expectation value of the perturbation operator for the unperturbated wavefunction gives directly electrostatic energy:

$$\Delta E_0^{(1)} = \Delta E_{el}^{(1)} = \langle \psi_0^{(0)} | \hat{V} \psi_0^{(0)} \rangle \tag{1.23}$$

and hence describes Coulomb's interaction of the two unperturbated (fixed) charge distributions. If one expresses the electrostatic interaction of each particles from molecule A and B as the infinite sum of the interacting multipole moments (multipole expansion), the first-order electrostatic energy turns out to be the sum of multipole and penetration terms:

$$\Delta E_{el}^{(1)} = \Delta E_{mtp} + \Delta E_{pen} \tag{1.24}$$

The equality between electrostatic and multipole components would take place if the charge distributions of both molecules were completely separated, that is no penetration was observed.

The second order correction consists of two terms, i.e., induction and dispersion:

$$\Delta E_0^{(2)} = \Delta E_{ind}^{(2)} + \Delta E_{disp}^{(2)} = \sum_{n_A} \sum_{n_B}{}' \frac{|\langle \psi_{A,n_A} \psi_{B,n_B} | \hat{V} \psi_{A,0} \psi_{B,0} \rangle|^2}{(E_{A,0} - E_{A,n_A}) + (E_{B,0} - E_{B,n_B})} \tag{1.25}$$

Here the subscripts n_A and n_B denote the states of particular molecules A and B, that form the total n-th state of the system:

$$\psi_n^{(0)} = \psi_{A,n_A} \psi_{B,n_B} \tag{1.26}$$

and the ground state $n = 0$ (the one that is the state of interest in this equation and is expressed by both n_A and n_B equal zero) is omitted in the above summation. A contribution to the $\Delta E_0^{(2)}$ from states, where solely either molecule A or molecule B are excited (either n_A or n_B equals zero) is then interpreted as the interaction between the molecule A multipoles induced by the electric field of molecule B with its static multipoles, and vice versa. This type of interaction energy is simply the induction energy. Likewise, the remaining component of second-order energy (the one corresponding to the excitement of both molecules) expresses, from a classical point of view, the effect of two induced multipoles interactions—the dispersion energy. Another interpretation of this phenomenon is that one molecule charge distribution fluctuations cause the appearance of temporary multipole moments in this molecule. The electric field from this multipoles induces multipole moments in the second molecule and these two attract one another. Finally, the time-averaged effect of such an induced multipoles interaction is referred as dispersion energy.

The key assumption in foregoing approach is the polarization approximation, neglecting the possible electron exchange. This simplification is reasonable in the case of long intermolecular distances. However, if molecules are not well separated, the overlap between their wavefunctions causes the repulsion that here is ignored: neither interaction energy term in the perturbation theory employing the polarization approximation accounts for this effect. To consider electron exchange (resulting from the electrons indistinguishability and Pauli exclusion principle), the wavefunction for the combined system should be antisymmetrized product of the two molecules wavefunctions, that is its sign should revert when the two electrons are exchanged. Symmetry-adapted perturbation theory SAPT has been developed to handle this issue and deal with intermediate and short range of intermolecular distances [141]. Similarly to polarization perturbation treatment, the zeroth-order energy stands for the sum of isolated molecules and interaction energy is represented by subsequent corrections. Unlike, however, previously demonstrated approach, the corrections of each order consist of more terms than precedent: interaction energy components listed above are supplemented by

exchange repulsion terms. For instance, the first-order interaction energy (equivalent to the Heitler-London energy) contains the electrostatic contribution as well as the exchange component:

$$\Delta E^{(1)} = \Delta E^{(1)}_{pol} + \Delta E^{(1)}_{exch} = \Delta E^{(1)}_{el} + \Delta E^{(1)}_{exch} \tag{1.27}$$

Making allowance for the additional exchange term prevents the non-physical effect of infinite interaction energy lowering when the two neutral molecules approach each other. symmetry-adapted perturbation theory (SAPT) methods enable probably the most accurate yet very expensive interaction energy calculations, their application is therefore limited by the size of investigated systems.

While no unique way of energy partitioning exists and a multitude of formulations can be encountered, the variation-perturbation scheme [26] utilized herein constitutes a reasonable compromise between accuracy and computational cost. The latter has been achieved as a result of combination of both variational and perturbational theories: the first-order electrostatic interaction energy component is defined as in perturbation approach, whereas the next terms in a hierarchy of accuracy, i.e., Heitler-London and SCF energies, are evaluated within variational formulation. The remaining components can be obtained from consecutive differences introduced as follows. The SCF interaction energy, determined as in a supermolecule method and calculated in a dimer basis set to correct for BSSE, constitutes a starting point for the energy partitioning. After subtracting the first-order Heitler-London term, $\Delta E^{(1)}$, representing an interaction of monomers with frozen electronic density distributions, the remaining delocalization component expresses an effect of interaction on the electronic distribution relaxation of the monomer and comprises several higher-order SAPT corrections including induction:

$$\Delta E_{SCF} = \Delta E^{(1)} + \Delta E^{(R)}_{DEL} \tag{1.28}$$

$\Delta E^{(1)}$ energy is obtained after the first SCF interaction cycle, starting from the Gram-Schmidt orthogonalized monomer eigenvectors calculated in a dimer basis set. According to the SAPT formulation of the first-order interaction energy, the Heitler-London term consists of electrostatic and exchange contributions:

$$\Delta E^{(1)} = \Delta E^{(1)}_{EL} + \Delta E^{(1)}_{EX} \tag{1.29}$$

Following the perturbation theory, first-order electrostatic energy is given by the equation:

$$\begin{aligned}\Delta E_{EL}^{(1)} &= \sum_a \sum_b Z_a Z_b R_{ab}^{-1} + \sum_r \sum_s \sum_t \sum_u D_{rs}^A D_{tu}^B \langle rs|tu\rangle \\ &- \sum_r \sum_s \sum_b D_{rs}^B \langle r|Z_b R_{1b}^{-1}|s\rangle \\ &- \sum_t \sum_u \sum_a D_{tu}^A \langle t|Z_a R_{1a}^{-1}|u\rangle \end{aligned} \tag{1.30}$$

where r, s, t and u indices refer to dimer atomic orbitals; a, b correspond to atoms of A and B molecules; Z_a, Z_b stand for nuclear charges; R is the separation of two particles (electron-electron, nucleus-nucleus or electron-nucleus) and D_{rs}^A and D_{tu}^B designate monomer electron density matrices obtained in a dimer basis set.

Since the calculation of electrostatic energy contribution is less computationally expensive compared to the second component of first-order interaction energy, it is advantageous to obtain the exchange term as a plain difference:

$$\Delta E_{EX}^{(1)} = \Delta E^{(1)} - \Delta E_{EL}^{(1)} \tag{1.31}$$

While SCF method entirely excludes dispersion, an accurate description of all types of weak interactions requires the correlated *ab initio* methods to be used. To account for electron correlation effects, $\Delta E_{CORR}^{(2)}$ term (including intramolecular correlation and intermolecular dispersion components) can be obtained from the MP2 (or CCSD) interaction energy:

$$\Delta E_{MP2} = \Delta E_{SCF} + \Delta E_{CORR}^{(2)} \tag{1.32}$$

In summary, decomposition of the MP2 (or CCSD) interaction energy reveals the following terms:

$$\Delta E_{MP2} = \Delta E_{EL}^{(1)} + \Delta E_{EX}^{(1)} + \Delta E_{DEL}^{(R)} + \Delta E_{CORR}^{(2)} \tag{1.33}$$

As one can easily note, a well-defined hierarchy of successive interaction energy approximations, varying from the most expensive MP2 (or CCSD) method to the electrostatic energy representation (the more simplified the theory, the less computationally demanding calculation; see Figure 1.1), demonstrates the utility of this decomposition scheme:

$$\Delta E_{EL}^{(1)} < \Delta E^{(1)} < \Delta E_{SCF} < \Delta E_{MP2} \tag{1.34}$$

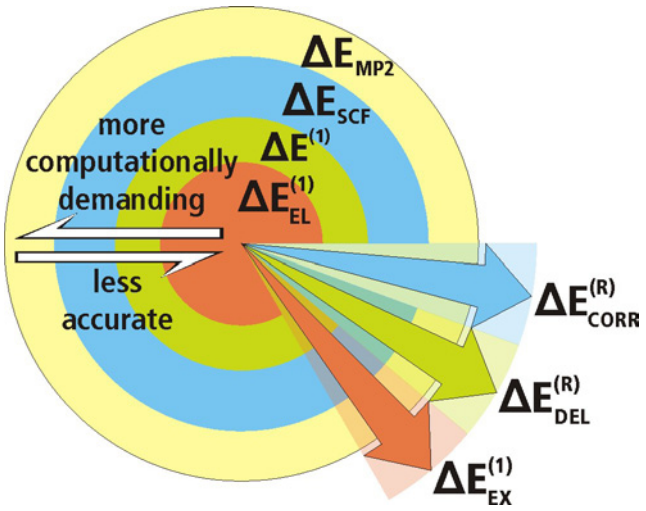


Figure 1.1: Variation-perturbation scheme of interaction energy decomposition. *Circles* denote the consecutive levels of theory, while *arrows* designate the corresponding correction terms.

It performs exceptionally well when derivation and testing of simplified yet reasonable models by a subsequent elimination of less important and more time-consuming contributions is required. Moreover, owing to the full counterpoise correction in each of the interaction energy components, the basis set dependency has been significantly reduced. Finally, additional advantage of a presented treatment comes from the implementation of a direct SCF technique, so that the storage of integrals during calculations can be avoided and the efficient study of relatively large models is possible. Currently molecular systems containing up to circa 300 atoms could be investigated, opening the possibility to analyze the physical nature of interactions in enzyme active sites.

Variation-perturbation energy decomposition scheme was successfully employed in the investigation of a number of diverse phenomena including the influence of mutations within an enzyme active site [41], the contribution of particular active site residues to catalytic effects [31, 137] (see the next section) and the relation between inhibitory activity and the interaction energy [13, 14].

1.3.2 Differential Transition State Stabilization

The key concept of DTSS method [31, 32] follows the hypothesis of Pauling [96] and states that the catalytic (or inhibitory) activity of a given environment arising

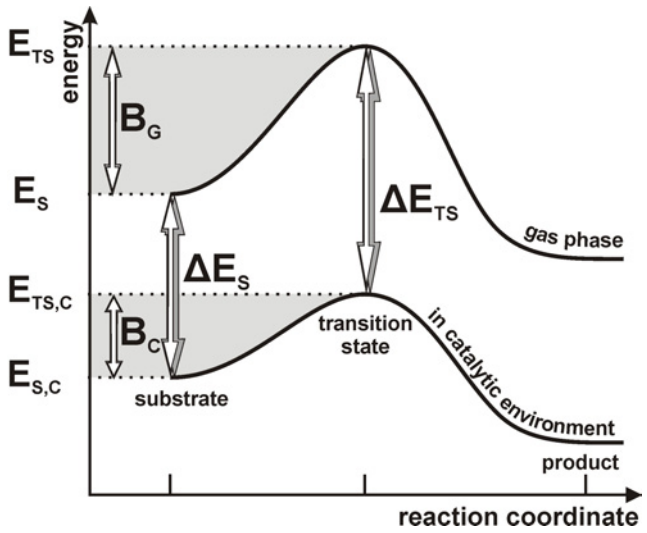


Figure 1.2: Energy diagram illustrating the DTSS approach [31, 32]. B_C and B_G denote the activation energy barriers for the catalyzed and uncatalyzed (gas phase) processes, respectively. E_{TS} (E_S) or $E_{TS,C}$ ($E_{S,C}$) are the transition state (reactant) energies in the gas phase or in the presence of a catalyst.

from lowering of the activation energy barrier, $\Delta = B_C - B_G$ (see Figure 1.2), can be associated with the difference in a strength of transition state, ΔE_{TS}, versus substrate binding, ΔE_S, with the catalytic environment:

$$\Delta_{DTSS} = \Delta E_{TS} - \Delta E_S \tag{1.35}$$

Thus, the stronger a given environment interacts with the transition state compared to the corresponding interaction with reactants, the lower is the resulting activation energy barrier and the faster the reaction proceeds.

Derivation of the catalytic effects from the interactions involving the catalyst (e.g., enzyme) and the substrate/transition state species allows for assessing the physical nature of catalytic activity. In particular, decomposition of binding energies according to variation-perturbation scheme provides the following contributions to overall DTSS effect:

$$\Delta_{MP2} = \Delta_{EL}^{(1)} + \Delta_{EX}^{(1)} + \Delta_{DEL}^{(R)} + \Delta_{CORR}^{(2)} \tag{1.36}$$

As discussed in section 1.3.1, consecutive interaction energy terms are related via a clear hierarchy of approximations offering the more accurate description at the expense of a higher computational cost. Such a relationship refers also to

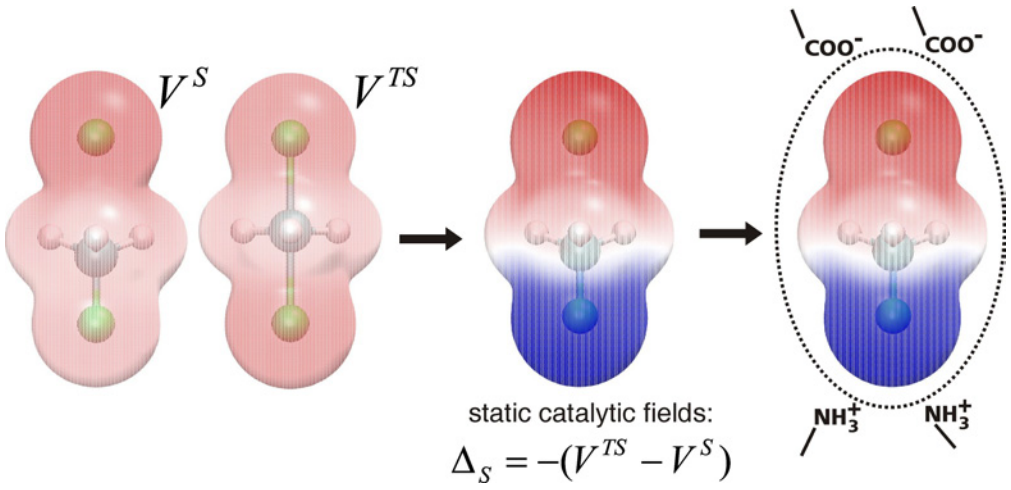


Figure 1.3: Derivation and interpretation of the catalytic fields [31, 32] for a model reaction (i.e., symmetric nucleophilic substitution of methyl fluoride).

DTSS quantities:

$$\Delta_{EL}^{(1)} < \Delta^{(1)} < \Delta_{SCF} < \Delta_{MP2} \tag{1.37}$$

allowing for systematic, first principles based derivation of approximate models of catalytic activity.

Whenever the electrostatic term is dominant, even more simplified approach can be applied to estimate the catalytic contribution for a given reaction from partial atomic charges of a molecular environment, q_i, and the difference in molecular electrostatic potential (MEP) of a transition state and a substrate:

$$\Delta_{EL,Vq} = \sum_i q_i (V_i^{TS} - V_i^S) \tag{1.38}$$

Since the difference in MEP generated by transition state and reactants expresses the catalytic influence of a positive unit charge, simply considering the so-called static catalytic field [31, 32], allows the most favourable distribution of charges within a hypothetical catalytic environment to be predicted:

$$\Delta_S = -(V^{TS} - V^S) \tag{1.39}$$

By inverting the sign of the transition state and substrate MEP difference, one can derive the optimal electrostatic characteristics of a complementary molecular environment acting as a catalyst (Figure 1.3). Notably, only the knowledge of a transition state and a reactant structure is assumed, i.e., no prior knowl-

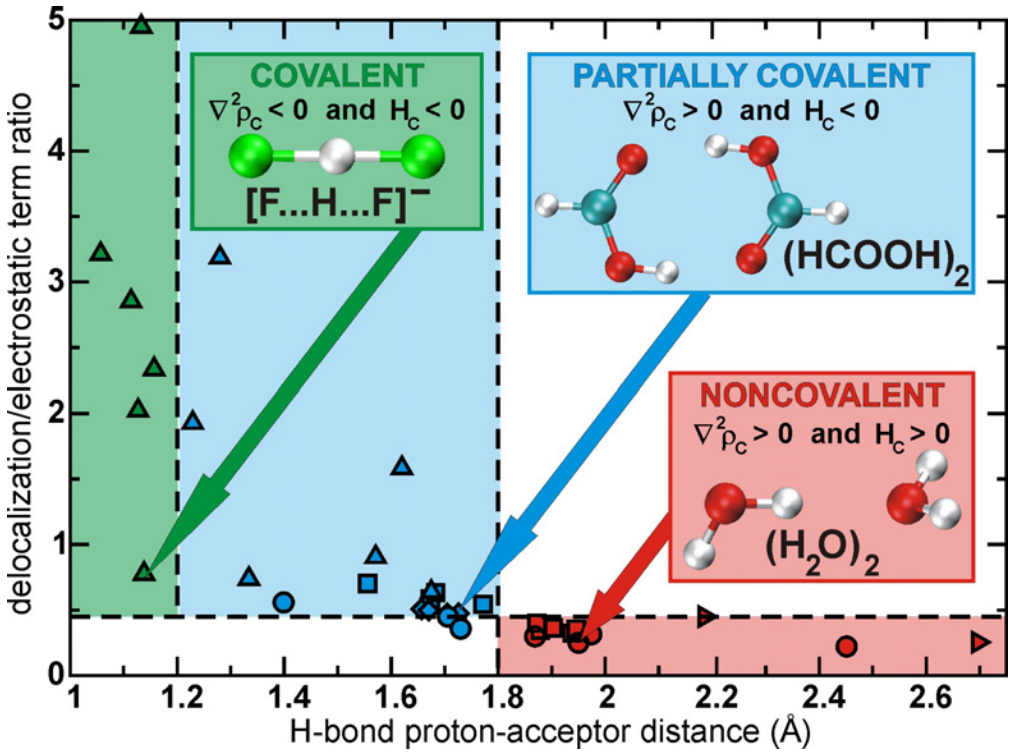


Figure 1.4: Classification of the hydrogen bonding interactions within 34 hydrogen bonded complexes studied in Ref. [142].

edge of the catalytic environment is required. As it will be shown in Chapter 3, comparison of the catalytic fields with the actual arrangement of charged enzyme active site residues aids determination of catalytically important residues along with the prediction of the impact of a mutation on an enzyme catalytic activity.

1.3.3 Physical nature of hydrogen bonds

The conventional hydrogen bond is said to be electrostatic in nature [143]. However, some classes of hydrogen bonds recognized within the last years do not conform with this statement [144, 145]. It seems that hydrogen bonds could occur anywhere between the two extremes associated with covalent and ionic bonds [146]. In particular, very strong hydrogen bonds appear to involve some degree of covalency. The contribution described in Ref. [142] has aimed at providing the classification of a variety of hydrogen bonded complexes in terms of their possible covalent nature.

Straightforward criteria for assessment of the degree of a hydrogen bond covalency are defined within the topological atoms-in-molecules (AIM) theory of Bader [147]. The nature of a specific interaction is described by a number of properties related to the electron density, ρ_C, of the corresponding bond critical point (BCP). The latter is characterized by the zeroth gradient of ρ_C. Depending on the sign of the Laplacian of ρ_C, the covalent bonds can be recognized as they are associated with the concentration of electron density in an interatomic region (i.e., $\nabla^2\rho_C < 0$). Consequently, the positive $\nabla^2\rho_C$ value indicates depletion of the electron density which implies the interaction of closed-shell species (i.e., ionic, hydrogen bonding, and van der Waals interactions). Further characteristics of an interaction in such a case is made on the basis of the sign of electronic energy density, H_C, in a given BCP. Local energy density is given by the equation:

$$H_C = G_C + V_C \tag{1.40}$$

Local one-electron kinetic energy density, G_C, and local potential energy density, V_C, are related to the Laplacian of electron density in the following way:

$$\frac{1}{4}\nabla^2\rho_C = 2G_C + V_C \tag{1.41}$$

The sign of Laplacian is then an indication whether the negative V_C or the positive G_C is in excess of the virial ratio equal to 2. If $\nabla^2\rho_C < 0$, the potential energy dominates. In the case of a bond with any degree of covalency, the condition that $|V_C| > G_C$ implies that $H_C < 0$. Accordingly, partially covalent bonds are associated with $G_C < |V_C| < 2G_C$ (i.e., positive $\nabla^2\rho_C$) [148]. Finally, $H_C > 0$ indicates the noncovalent interaction.

The possible covalent nature of the interaction can also be determined by means of the theory of intermolecular interactions, e.g., with the variation-perturbation scheme of interaction energy partitioning. In particular, $\Delta E^{(R)}_{DEL}$ term may be related to covalent effects, whereas $\Delta E^{(1)}_{EL}$ represents ionic interactions [149]. Analysis of 34 hydrogen bonded complexes resulted in the classification presented in Figure 1.4 [142], which demonstrates close relationship between hydrogen bond lengths, AIM properties and $\Delta E^{(R)}_{DEL}/\Delta E^{(1)}_{EL}$ ratio. Remarkably, covalent and partially covalent bonds (e.g., charge assisted hydrogen bonds) are characterized by $\Delta E^{(R)}_{DEL}/\Delta E^{(1)}_{EL}$ value exceeding ~ 0.45. Further distinction is based on the hydrogen bond proton-acceptor distance, i.e., covalent hydrogen bonds within the studied set exhibit $R_{XH\ldots Y} < 1.2$ Å (e.g., $(FHF)^-$). Noncovalently bound complexes ($\Delta E^{(R)}_{DEL}/\Delta E^{(1)}_{EL} > \sim 0.45$) feature $R_{XH\ldots Y} > 1.8$ Å.

It should be emphasized that this classification [142] is in line with topological characteristics of BCPs associated with particular hydrogen bonded complexes. While the dataset investigated herein was necessarily limited, it will be interesting to compare these conclusions with properties of other hydrogen bonding interactions, especially of biological relevance. In particular, assessing the possible covalent nature of hydrogen bonds occurring in the enzyme active sites might help to address the hypothesis about covalent catalysis performed by enzymes with remarkable proficiencies [36].

CHAPTER 2

Enzyme Inhibitory Effects

2.1 Phenylalanine Ammonia-lyase Inhibition

2.1.1 PAL characteristics

Phenylalanine ammonia-lyase (PAL, E.C. 4.3.1.5) is a crucial enzyme of general phenylpropanoid pathway operating in plant and fungi. Non-oxidative deamination of amino acid substrate *(S)*-phenylalanine followed by the formation of *(E)*-cinnamic acid (Figure 2.1) redirects the carbon flow from primary metabolism to the synthesis of phenylpropanoid compounds [150] (e.g., lignins, flavonoids, coumarins). Phenylpropanoid products play multiple roles including mechanical support, signaling, and protection against biotic and abiotic stress [151]. PAL, lying at a branching point between the primary metabolism and natural products biosynthesis, has been extensively studied as a possible control site of desirable product accumulation and potential target for herbicides [152]. The reverse PAL reaction encompassing the formation of L-phenylalanine has been applied in an industrial process of aspartam production [153]. An employment of PAL in the treatment of a genetic disorder phenylketonuria has also been considered [154].

While the experimental structure of PAL has recently been solved for a yeast and plant enzyme [155, 156], little is known about the actual catalytic mechanism of PAL [125]. The PAL-catalyzed reaction is particularly challenging as it involves abstraction of a non-acidic β proton located in the vicinity of an acidic α ammonium group which should remain protonated to allow for its departure. An active site prosthetic group of 5-methylene-3,5-dihydroimidazol-4-one (MIO),

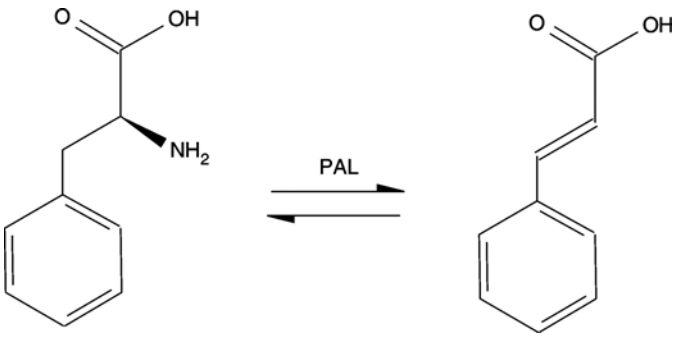


Figure 2.1: PAL-catalyzed reaction.

formed autocatalitycally by cyclization and dehydration of residues from a highly conserved Ala–Ser–Gly triad [157], is suggested to perform an electrophilic attack at the phenylalanine aromatic ring in a Friedel-Crafts type mechanism [124–127]. The following activation of the β proton allows for its abstraction by an enzymatic base (presumably Tyr351 [158]). After the proton transfer step, ammonia elimination occurs. Other PAL active site residues anticipated to contribute to substrate binding and/or catalysis include Arg354, Glu348 (interaction with the substrate carboxylic moiety) and Tyr110 (binding of the eliminated ammonia molecule).

Among the first specific PAL inhibitors was a derivative of *(S)*-phenylalanine with α amino group replaced by the aminooxy moiety, i.e., *(S)*-2-aminooxy-3-phenylpropanoic acid (*(S)*-AOPP) [159] (Figure 2.2). Phosphonic acid analogues of substrate, such as *(R)*-1-amino-2-phenylethylphosphonic acid (*(R)*-APEP), have also been proved to serve as moderately potent inhibitors of PAL [160, 161]. The most potent *in vivo* inhibitor of PAL is 2-aminoindane-2-phosphonic acid (AIP), a conformationally constrained derivative of phenylalanine [162].

Both *(S)*-AOPP and AIP inhibit PAL in a time-dependent manner and are assumed to act as reaction intermediate analogues [163]. Combined experimental and computational analysis of AIP revealed that since its geometry in solution presumably differs from that in PAL active site, the necessary conformational transition occurring upon docking might explain the observed slow binding [164].

Despite the abundant experimental data regarding structure-activity relationships of PAL inhibitors [165–167], there exists no experimental or theoretical evidence explaining the molecular basis of PAL inhibition. The only theoretical analysis performed for AIP-inhibited PAL has suggested that the inhibitor molecule is more likely to bind in a conformation with axial phosphonic group [164]. The following steps were undertaken to gain insight into the mode of PAL inhibitors binding. Detailed and systematic analysis of binding energy was performed to de-

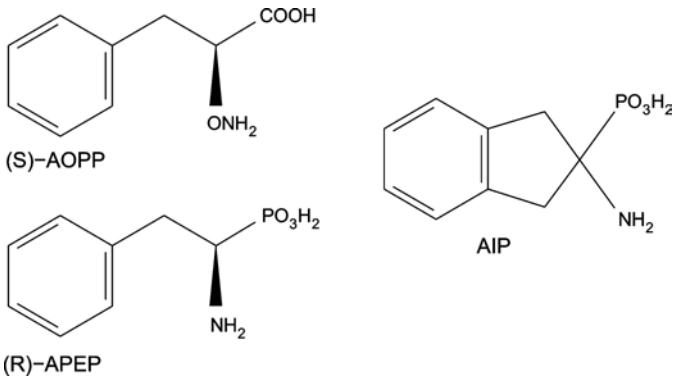


Figure 2.2: *(S)*-phenylalanine analogs known as PAL inhibitors [159–162].

scribe the physical nature of receptor-ligand interactions. Finally, the approximate models of inhibition were derived based on the comparison of predicted and actual biological activity of particular inhibitors.

2.1.2 Inhibitors of potato PAL

Due to availability of experimental binding affinities for a set of inhibitors of PAL from potato (*Solanum tuberosum*) [27, 28], this particular enzyme was chosen for further investigation. Except for *(S)*-AOPP, the remaining five compounds considered herein are aminophosphonic derivatives of phenylalanine differing in terms of Cα configuration and/or substituents at phosphorus atom (Figure 2.3).

Technical details Since no experimental PAL structure had been available by the time this analysis was performed [168], active site homology model based on the X-ray structure of a closely related histidine ammonia-lyase (HAL) was employed, as prepared by Dr. Jolanta Grembecka. The details of modeling the enzyme-inhibitor structures are provided in Ref. [168].

The generated active site model consisted of six highly conserved amino acid residues surrounding the variable part of inhibitor structures: Tyr35^{A}, Asn187^{A}, Gln275^{B}, Tyr278^{B}, Arg^{+}281^{B}, and Asn311^{A} (the superscripts indicate the monomer of PAL homotetramer). To reduce computational effort, asparagine and glutamine residues were represented by acetamide molecule, arginine was modelled by methylguanidinium cation, whereas tyrosine residues were truncated at Cβ atom and mimicked by *p*-cresol (Figure 2.4).

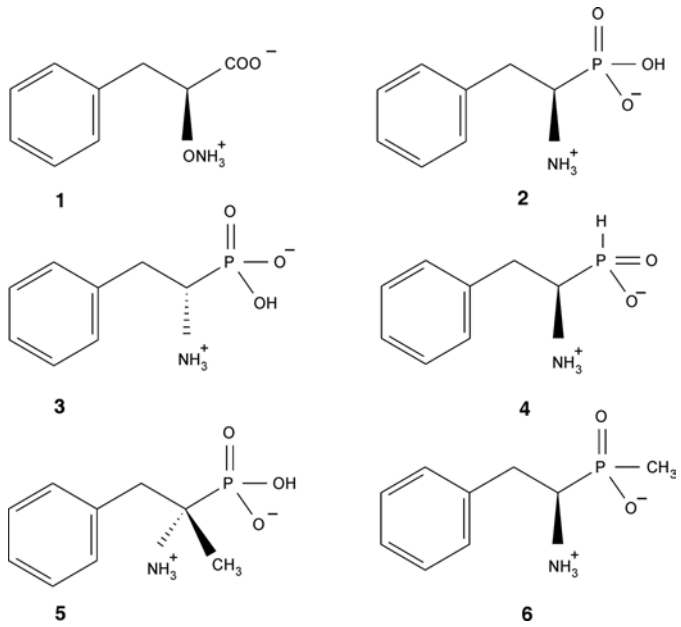


Figure 2.3: Structures of potato PAL inhibitors in the order of decreasing inhibitory activity [27, 28].

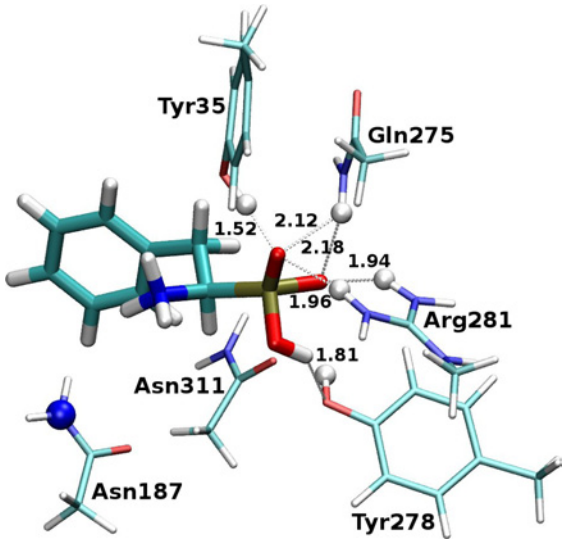


Figure 2.4: Potato PAL active site model in complex with inhibitor **2**. The shortest atomic contacts [Å] are marked as *dashed lines*. *Ball representation* is used to distinguish 6 atoms with atomic point charges employed in an approximate electrostatic model of inhibitory activity.

Calculation of receptor-ligand interaction energy was performed in a pairwise fashion with the total binding energy expressed as a sum of two-body contributions revealed for each inhibitor-amino acid residue dimer. Variation-perturbation scheme of interaction energy decomposition[26] was applied as implemented [169] in the modified version of Gamess–US [170]. Consecutive components of interaction energy were computed using 6-31G(d) basis set and full counterpoise correction [140].

Correlation between experimental inhibitory activity and theoretical interaction energy was determined using regression analysis and standard least square method of line fitting. The strength of correlation was assessed by means of two statistical parameters: coefficient of determination and standard error of estimate (r^2 and SEE, respectively). Coefficient of determination, corresponding to the square of the Pearson's correlation coefficient, is a measure of the proportion of variation in a dependent variable explained by an independent one. Standard error of estimate, describing the uncertainty of a predicted dependent variable, is given by the equation:

$$SEE = \sqrt{\frac{\sum (Y - Y')^2}{N - 2}} \tag{2.1}$$

where Y, and Y' are the actual and predicted values, respectively; while N denotes the number of (X, Y) pairs. The reason for $N - 2$ expression in the denominator is that two parameters (the slope and the intercept) were estimated in order to obtain the sum of squares.

2.1.3 Inhibitors of parsley PAL

Since new experimental data have recently been provided by Dr. Jerzy Zoń [29] for a larger set of eleven inhibitors with known biological activity toward PAL from parsley (*Petroselinum crispum*), more accurate investigation of PAL inhibition has been undertaken in this study, augmented by an availability of the X-ray structure of parsley PAL [156]. The following compounds were considered: *(S)*-AOPP, AIP, 5- and 4-substituted AIP derivatives (6 structures), ring-substituted 1-aminobenzylphosphonic acids (2 structures), and an analogue of AIP with indane moiety replaced by benzocyclobutane (Figure 2.5). Except for enantiopure *(S)*-AOPP and achiral AIP, all inhibition constant values refer to a racemic mixture. Of the two possible enantiomers of phenylalanine derivative, higher inhibitory potency is exhibited by the enantiomer with Cα atom configuration corresponding to that of *(S)*-phenylalanine (i.e., R isomer) [29]. It was assumed that relative binding affinity of racemic mixtures does not change compared to the corresponding quantities for R isomer. Accordingly, only R isomers were studied in what follows.

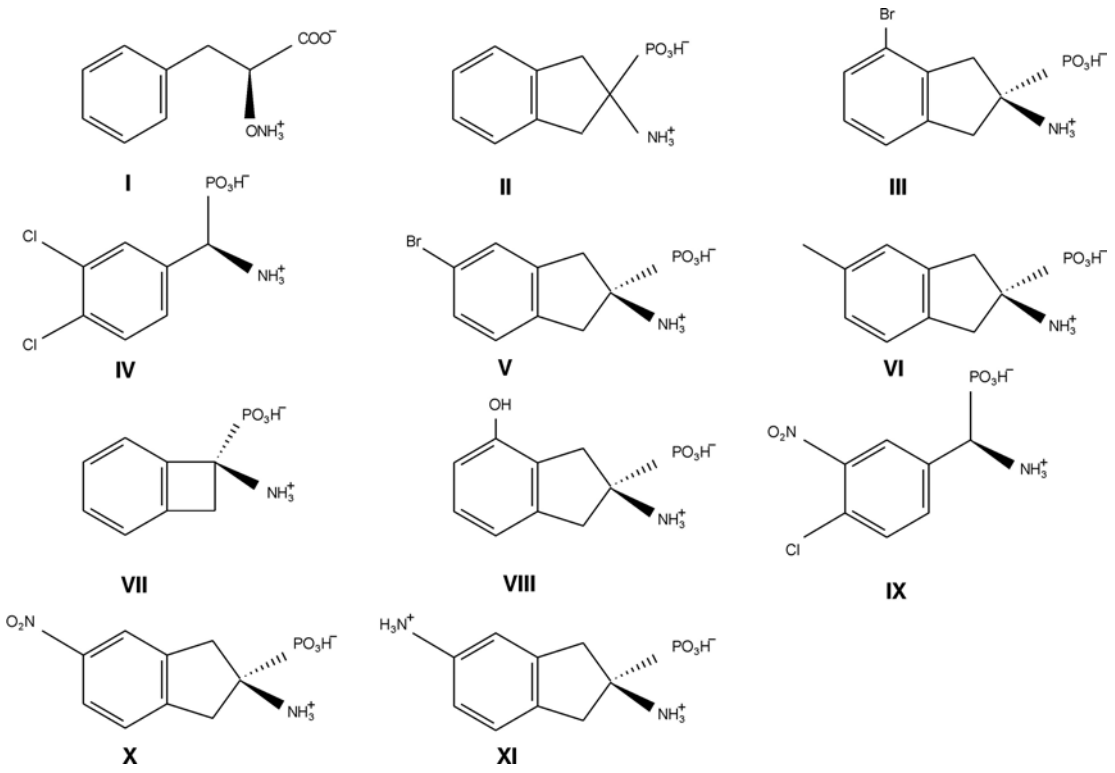


Figure 2.5: Structures of parsley PAL inhibitors in the order of decreasing inhibitory activity [29].

It is noteworthy, that in contrast to potato PAL inhibitors, the compounds considered herein differ mainly in the structure and/or substituents of a cyclic part of a molecule, while the aminophosphonic moiety is retained.

Technical details Crystal structure of parsley PAL [156] was taken from the Protein Data Bank (1.70 Å resolution, PDB accession code 1W27). The crystallographic unit contains two equivalent PAL subunits. Apart from the chemical modification of MIO prosthetic group due to solvent composition throughout the crystallization process, no ligands are present in the X-ray structure. MIO-bound 1,4-dithiothreitol was then removed and the catalytically functional PAL homotetramer was generated based on the symmetry information provided with the PDB file. Protonation of titratable residues was determined based on the Optimal Hydrogen Bonding method [171] available via What If program [172]. In particular, histidine residue present in the active site, His396, was modelled as neutral with proton attached to δ nitrogen atom. Possible flipped positions of ring

nitrogen and carbon atoms were identified in the case of histidine residues His153, His306, and His600. Asp489 was assumed to be protonated at carboxylic OD2 atom due to its close contact to Glu429 side chain (Asp489/OD2...Glu429/OE2 distance of 2.42 Å). The remaining aspartate and glutamate residues were negatively charged (−1), whereas lysine and arginine residues were assigned a positive charge (+1). The missing hydrogen atoms were added with Charmm procedure Hbuild and optimized throughout a series of 500 steps of steepest descent minimization followed by conjugate gradient optimization until the RMS gradient of 0.01 kcal $\cdot$ mol$^{-1}\cdot$Å^{-1} was satisfied. Charmm package [173] and all–hydrogen parameter set from Charmm27 forcefield [174] were employed in the simulation.

The structures of the inhibitors were optimized at the HF/6-31G(d) level of theory with Gaussian 03 program [175]. In the case of AIP or AIP-derivatives (compounds **II**, **III**, **V**, **VI**, **VIII**, **X**, and **XI**; see Figure 2.5) two conformations were taken into account encompassing two possible arrangements of the phosphonic group, i.e. axial and equatorial (**AC** and **EC**, respectively). Except for inhibitor **I** modelled in *S*-isomeric form, the remaining chiral compounds were taken as *R* isomers.

Docking of the optimized structures was performed using Gold (Genetic Optimization for Ligand Docking) software, version 4.0.1 [65, 66]. The receptor site was defined as a 10 Å radius sphere centered on a certain point with coordinates determined by the position of a ligand crystallized with yeast PAL [155] (PDB accession code 1T6J). Against the binding site model, 10 ligand binding poses were generated for each inhibitor using Gold Fitness function and default genetic search parameters. Docking solutions were manually inspected to determine the representative binding poses. In the case of considerably different binding modes, the top solutions (according to fitness function score) representing each ligand pose were selected. Poorly populated clusters with low fitness score were neglected. As the binding mode of most compounds was well defined (i.e., all docking solutions were similar or identical), alternative arrangement was considered only in three cases (inhibitors **II-EC**, **III-AC**, and **III-EC**). However, subsequent interaction energy calculation revealed that binding poses with lower fitness score were also characterized by lower interaction energy than the respective top-ranked solutions. Therefore, only top docking solutions were subjected to further analysis.

The following 21 amino acid residues were selected to represent PAL active site during the interaction energy calculation: Gly115^{B}, Phe116^{B}, Leu134^{B}, Phe137^{B}, Leu138^{B}, MIO^{B}, Leu206^{B}, Leu256^{B}, Asn260^{B}, Gln348^{A}, Tyr351^{A}, Arg^{+}354^{A}, Asn384^{B}, Pro385^{B}, His396^{B}, Phe400^{B}, Lys^{+}456^{C}, Ile460^{C}, Glu^{-}484^{B}, Asn487^{B}, and Gln488^{B} (the superscripts indicate the monomer of PAL homotetramer). Unless the main chain of a given residue appeared to be involved in the ligand bind-

ing (i.e., for Gly115, Phe116, Asn384, Pro385), only side chains were included in the model of PAL binding cavity. In particular, Arg354 was represented by methylguanidinium cation, Phe137, Tyr351, His396, Phe400, Lys456, and Gln488 were truncated at Cγ atoms, while the remaining residues were present as respective Cβ-capped side chains (Figure 2.6). All broken bonds were saturated with hydrogen atoms. Except for the hydroxyl group of Tyr351 with hydrogen atom rotated to optimize the possible hydrogen bonding to ligand, each inhibitor-enzyme complex involved the same geometry of PAL active site.

Except for Tyr35, the remaining 5 amino acid residues present in a previously employed representation of potato PAL active site (Figure 2.4) have their counterparts in current parsley PAL model of binding pocket. In particular, parsley PAL Asn260, Gln348, Tyr351, Arg354, and Asn384 correspond to potato PAL Asn187, Gln275, Tyr278, Arg281, and Asn311 residues, respectively. Tyr35 of potato PAL, an equivalent of parsley PAL Tyr110, was excluded from the receptor site model of the latter due to its little participation in ligand binding (Tyr110–inhibitor distance exceeding 14Å). Since preparation of parsley PAL model involved homology modeling [168], it is unclear whether such a position of Tyr35 relative to bound inhibitor molecule is characteristic of parsley PAL specifically, or if it indicates a mistake in a model.

The subsequent interaction energy calculation and analysis of correlation with experimental ligand activity was performed according to the protocol described for the analogous project concerning potato PAL inhibitors (see *Technical details* paragraph in 2.1.2 subsection).

2.1.4 Docking of parsley PAL inhibitors

Docking results suggest that parsley PAL inhibitors studied herein (Figure 2.5) bind in two possible arrangements differing in a position of a cyclic moiety (Figure 2.7–*A*). Except for inhibitor **III**, AIP-derivatives with axial phosphonic group (i.e., **AC** conformation) form a cluster of very similar docking solutions characterized by the ring portion of a molecule directed toward the more polar part of a binding pocket composed of residues Gly115, Gln348, Pro385, and His396 (Figure 2.7–*B*). **EC** conformers of the same compounds bind with their cyclic part buried within more hydrophobic region of an active site (i.e., residues Phe116, Leu134, Phe137, Leu138, and Leu206; see Figure 2.7–*C*). This particular cavity appears to be more spacious, as it accommodates several docking poses including **AC** conformation of inhibitor **III** (shown in Figure 2.7–*B*) as well as inhibitors **IV**, **IX** (Figure 2.7–*D*) and **VII** (Figure 2.7–*E*). It seems that whenever an inhibitor molecule does not fit into the smaller size binding pocket (e.g., due to its

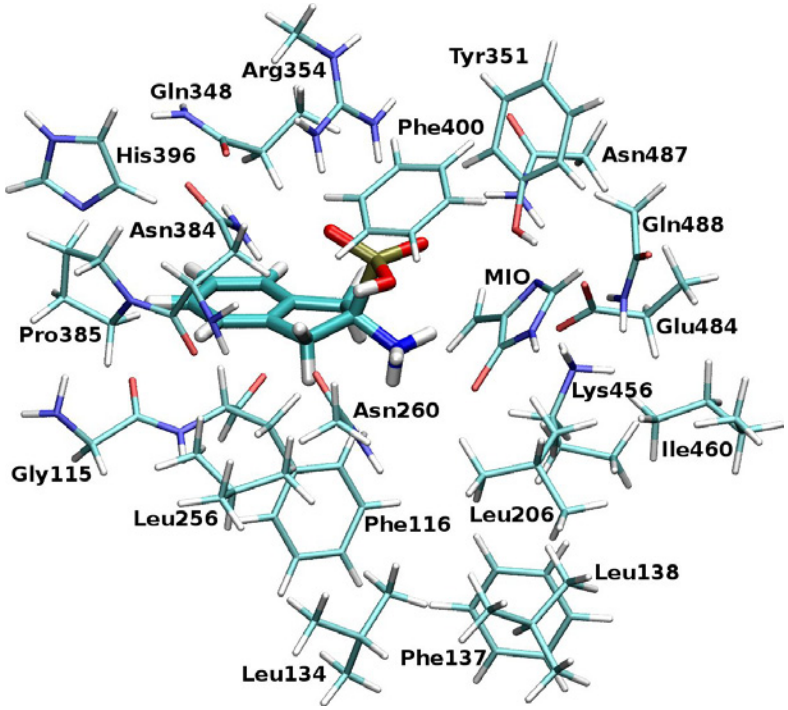


Figure 2.6: Parsley PAL active site model in complex with inhibitor **II-AC**.

conformation and/or the size of substituents), it can still be incorporated within the larger one. In addition to AIP derivatives in **AC** conformation, inhibitor **I** is the only other compound exhibiting complementarity with the smaller binding cavity (Figure 2.7–*F*). On the other hand, inhibitor **III-AC** occupies the larger binding pocket as the size and position of its bromine substituent probably prevents the mode of binding observed for the remaining **AC** conformers of AIP derivatives.

Independently on the binding mode, all inhibitors are anchored by salt bridge involving their phosphonic group and arginine (Arg354) residue. The position of positively charged amino moiety is more variable as it does not participate in any ionic or hydrogen bonding interactions. The only exception is the hydrogen bond occurring between aminooxy group of inhibitor **I** and Tyr351 hydroxyl group (Figure 2.8–*A*) which might explain the highest binding affinity of this particular compound. Compared to only slightly worse inhibitor **II**, the orientation of charged groups of compound **I** is optimal for simultaneous interaction with Arg354, Tyr351, and Glu484. The position of amino group of inhibitor **II-AC** (and the remaining compounds) prohibits any direct interactions with amino acid residues of PAL active site (Figure 2.8–*B*).

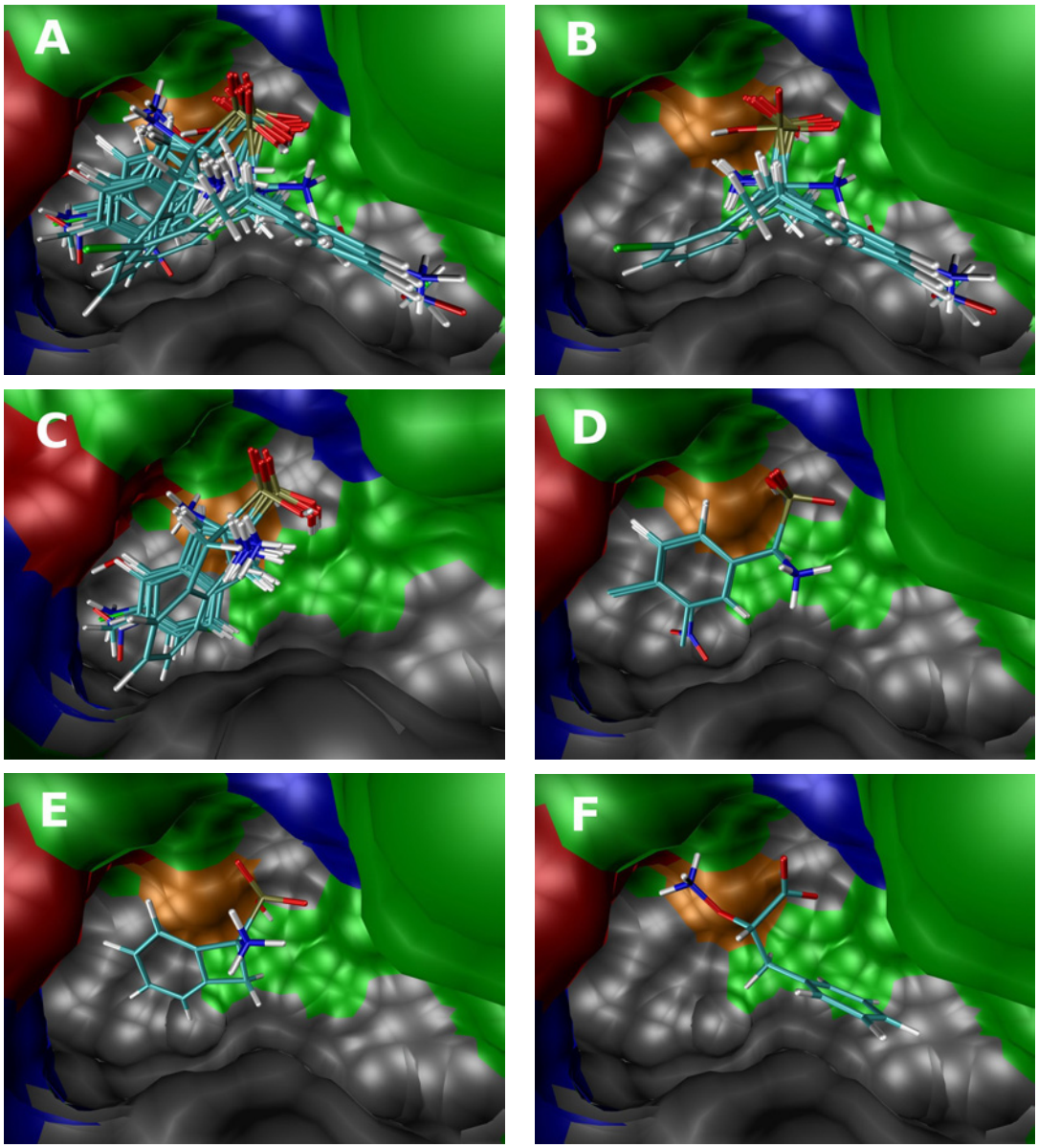


Figure 2.7: Binding poses of parsley PAL inhibitors. The surface representing PAL active site is coloured according to the type of particular amino acid residues: *blue/red colour* denotes positively/negatively charged residues; neutral polar and hydrophobic residues are shown in *green and grey*, respectively. MIO prosthetic group is given in *orange colour*. *A.* Superimposed geometries of all docked inhibitors. *B.* Inhibitors **II**, **III**, **V**, **VI**, **VIII**, **X** and **XI** in **AC** conformation. *C.* Inhibitors **II**, **III**, **V**, **VI**, **VIII**, **X** and **XI** in **EC** conformation. *D.* Inhibitors **IV** and **IX**. *E.* Inhibitor **VII**. *F.* Inhibitor **I**.

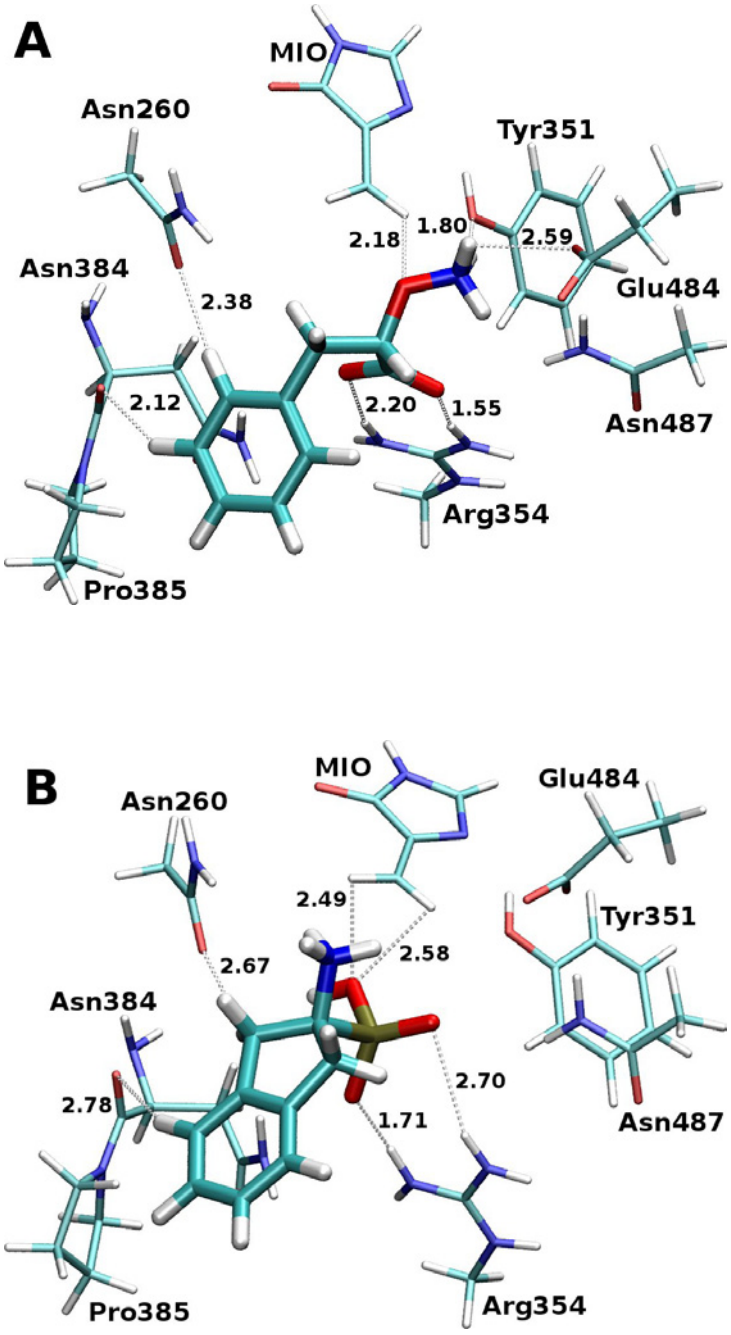


Figure 2.8: Selected amino acid residues of parsley PAL active site in complex with inhibitors **I** and **II-AC** (*A* and *B* part, respectively). The shortest atomic contacts [Å] are marked as *dashed lines*.

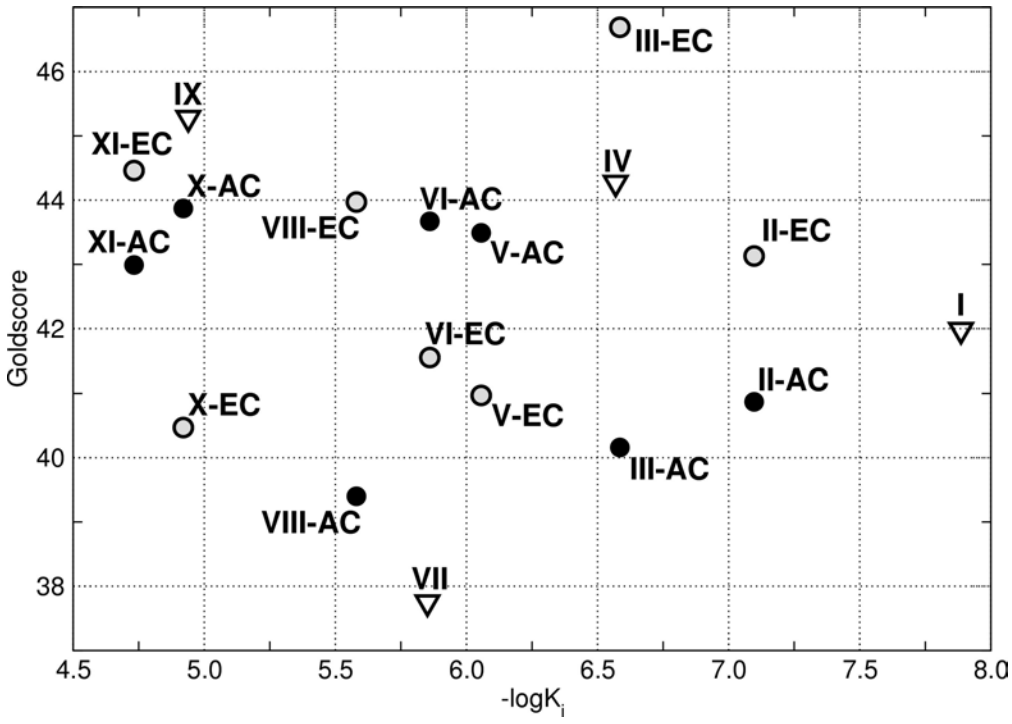


Figure 2.9: Goldscore function against the experimental inhibitory activity of parsley PAL inhibitors. AIP derivatives in **AC** and **EC** conformation are denoted by *full and grey-filled circles*, respectively; the remaining compounds are marked by *triangles*. Numbering of inhibitors corresponds to Figure 2.5.

Gold Fitness scoring function employed in ranking the docking solutions for respective inhibitors has been optimized for the prediction of ligand binding geometry and so it cannot be directly compared with binding affinity. It has been proposed [176] that removal of the intramolecular terms from Gold Fitness function could remediate this issue, so the resulting Goldscore function can be associated with biological activity. However, the consecutive analysis of possible relationships between Goldscore function and inhibitory potency of particular compounds revealed that this is not a case here, as no correlation is observed concerning the scoring function and experimental data (Figure 2.9). In particular, the most potent inhibitor **I** is associated with intermediate Goldscore, while much weaker inhibitors **IX** and **XI** rank among the top solutions. Likewise, considering Goldscore or Gold Fitness functions does not allow to determine the preferred docking conformation of AIP derivatives (i.e., unless one assumes that no common docking conformation describes the binding of this group of compounds).

Table 2.1: Total binding energy [$\mathrm{kcal} \cdot \mathrm{mol}^{-1}$] of potato PAL inhibitors calculated as a sum of two-body (i.e., inhibitor-residue) interactions at various levels of theory.

Inhibitor	$\Delta E_{EL}^{(1)}$	$\Delta E^{(1)}$	ΔE_{SCF}	ΔE_{MP2}
1	−154.91	−49.69	−95.76	−100.47
2	−129.76	−35.46	−79.55	−83.57
3	−129.77	−33.43	−75.54	−78.51
4	−131.09	−19.94	−69.65	−72.39
5	−127.90	−20.80	−63.91	−69.57
6	−105.52	−25.42	−61.21	−66.80

2.1.5 Interaction energy analysis

Potato PAL inhibitors The total binding energy [168, 177] was taken from a pairwise analysis of the interactions between the 6 PAL active site residues and particular inhibitor molecules (**A** model). With the goal of deriving some justified approximations in mind, the decrease in correlation with experimental data was monitored. This resulted in a selection of 4 active site residues (i.e., Tyr35, Asn187, Tyr278, and Arg281) that served as a limited size model of binding pocket (**B** model). The most complete description of the binding energy obtained at the MP2 level is very closely mimicked in both models by its SCF counterpart, indicating minor influence of correlation effects (Table 2.1). Qualitatively correct estimations of relative binding energies are also gained from the $\Delta E^{(1)}$ term, especially for the full model (**A** model; see Table 2.1 and Figure 2.10).

Remarkable correlation with experiment (i.e., with the value of $-\log K_i$) was revealed for subsequent interaction energy approximations including the first-order electrostatic term (Figure 2.10 and Table 2.2). This finding justifies an assumption regarding constant contribution of entropic and solvation effects to the free enthalpy of binding. Noticeably, the higher the level of theory applied to the description of interactions, the more pronounced the relationship of resulting binding energies with experimental data [27, 28]. The corresponding coefficients of determination and standard errors of estimate (Table 2.2) indicate negligible change at the MP2 and SCF levels when passing from model **A** to **B**. Although the analogous values for $\Delta E^{(1)}$ and $\Delta E_{EL}^{(1)}$ are affected, neglecting delocalization and exchange terms still produces reasonable agreement. Overall, **B** model provides a suffi-

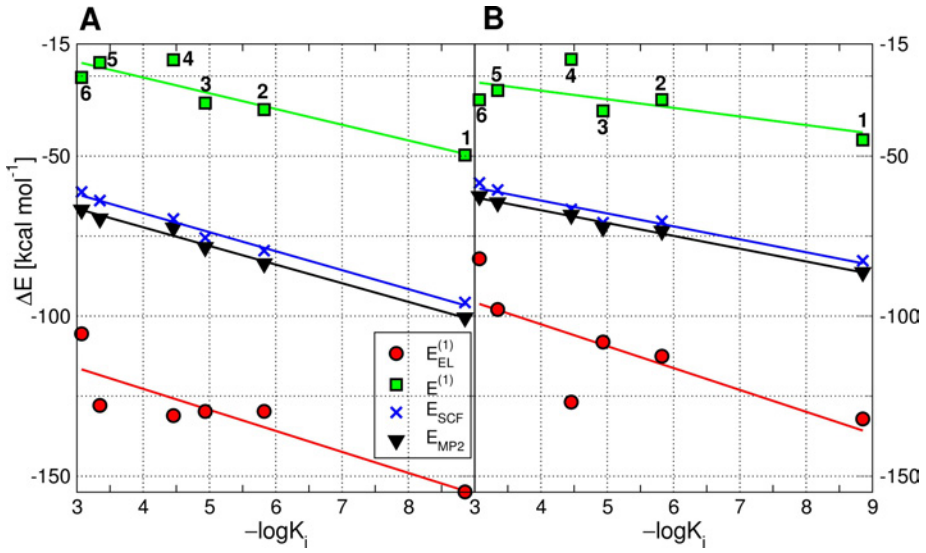


Figure 2.10: Total interaction energy at various levels of theory as a function of experimental inhibitory activity [27, 28] expressed as a logarithm of inhibition constant. *A.* Full model of potato PAL active site (i.e., 6 residues; **A** model). *B.* Potato PAL active site represented by 4 residues (**B** model). Numbering of inhibitors corresponds to Figure 2.3.

ciently accurate description of the experimentally observed characteristics of PAL inhibitors binding. The above conclusions are further confirmed by an excellent accuracy of the predicted inhibition constants based on the equations derived from correlation analysis of the MP2 interaction energy (Table 2.3).

As noted above, an adequately correct description of this particular set of PAL inhibitors binding is provided by the PAL active site model encompassing 4 out of the initial number of 6 amino acid residues, especially when the SCF (or higher) level of theory is employed.

Another question might arise, namely which PAL active site residues are crucial for the observed specificity of binding? Surprisingly, despite Arg281 exhibiting the largest strength of inhibitors binding (Figure 2.11), Tyr278 is more responsible for the observed differences in ligands stabilization (Table 2.4). This conclusion applies not only to interaction energy at SCF and MP2 levels of theory, but even more significant loss of correlation is found for electrostatic term. On the other hand, elimination of Arg281 does not alter the latter. Finally, omitting the two residues assumed to be of a minor importance while constructing **B** model (i.e., Gln275 and Asn311) does not result in the loss of correlation.

Table 2.2: Performance of interaction energy-based models of inhibitory activity. Regression equations and statistical parameters correspond to the fit shown in Figure 2.10. Standard error of estimate, SEE, is given in $\mathrm{kcal} \cdot \mathrm{mol}^{-1}$.

Model		Regression equation	r^2	SEE
ΔE_{MP2}	**A**[a]	$y = -5.8205 \cdot x - 48.9782$	0.99	1.48
	B[b]	$y = -4.0042 \cdot x - 50.9016$	0.99	0.99
ΔE_{SCF}	**A**	$y = -5.9282 \cdot x - 44.1450$	0.99	1.44
	B	$y = -4.0404 \cdot x - 47.7255$	0.96	1.99
$\Delta E^{(1)}$	**A**	$y = -4.9132 \cdot x - 5.8265$	0.85	4.88
	B	$y = -2.6834 \cdot x - 18.8540$	0.47	6.67
$\Delta E^{(1)}_{EL}$	**A**	$y = -6.5627 \cdot x - 96.4802$	0.78	8.17
	B	$y = -6.8373 \cdot x - 75.2016$	0.61	12.93

[a]Full PAL active site model (6 amino acid residues)
[b]Limited size PAL active site model (4 amino acid residues)

Considering the potential for practical applications, what is probably more important than the analysis of physical nature of interactions itself is the first principles based derivation of approximate yet reasonable models of binding affinity prediction. The use of empirical scoring functions only, although often successful, does not necessarily prove the validity of a computational model, that is the agreement with experiment might result from, for example, an accidental cancellation of errors. On the other hand, nonempirical approach can also lead to simple models relating the binding energy to quantities that may be rapidly evaluated (e.g., electrostatic potential of inhibitor in some selected positions). In such a case, however, the limit of accuracy has already been established by the highest level of theory and the following systematic neglect of the least significant contributions is well-founded. To address this idea a simplified model was investigated employing electrostatic potential generated by inhibitors and atomic point charges at selected binding site residues. Strikingly, significant correlation was found described by r^2 of 0.92 and 0.90 for Mulliken and potential-derived CHELP [178] charges. By consecutive elimination of components contributing the least to total electrostatic energy, six PAL active site atoms (Figure 2.4; atoms shown in ball representation) were selected to mimic the electrostatic properties of the entire PAL binding

Table 2.3: Comparison of experimental [27, 28] and predicted inhibition constant values [μM]. Numbering of inhibitors corresponds to Figure 2.3.

Inhibitor	$K_{i,exp}$	$K_{i,\Delta E_{MP2}}$	
		A[a]	**B**[b]
1	0.0014	0.0014	0.0014
2	1.5	1.1	2.4
3	11.6	8.4	4.6
4	35.0	95.0	42.0
5	450.0	290.0	390.0
6	850.0	870.0	1230.0

[a]Full potato PAL active site model (6 am ino acid residues)
[b]Limited size potato PAL active site model (4 amino aci d residues)

site. Coefficients of determination evaluated for an interaction between these six atoms point charges and molecular electrostatic potential of particular inhibitors were only slightly diminished (0.69 and 0.83 for Mulliken and CHELP charges, respectively). Finally, atomic point charges were arbitrarily assigned for the above listed set of six atoms. In particular, charges of $+0.4$ and -1.0 described hydrogen and nitrogen atoms (total charge of a system, $+1$, reflects the summary charge of a full model). Electrostatic interaction energy was then evaluated according to the formula:

$$\begin{aligned} \Delta E_{EL,Vq} &= 0.4 \cdot V_{Tyr35:HH} - V_{Asn187:ND2} \\ &\quad +0.4 \cdot V_{Gln275:HE22} + 0.4 \cdot V_{Tyr278:HH} \\ &\quad +0.4 \cdot (V_{Arg281:HH11} + V_{Arg281:HH21}) \end{aligned} \tag{2.2}$$

The results concerning $\Delta E_{EL,Vq}$ are in remarkable agreement with experimental inhibitory activity ($r^2 = 0.78$). Further improvement could be gained by determination of optimal (instead of arbitrary) charges via least squares fit of approximate electrostatic energy derived from equation 2.2 to the reference ΔE_{MP2} values. Such an analysis demonstrates a stepwise generation of simplified models, upon which knowledge of the factors neglected plays as important a role as the final results

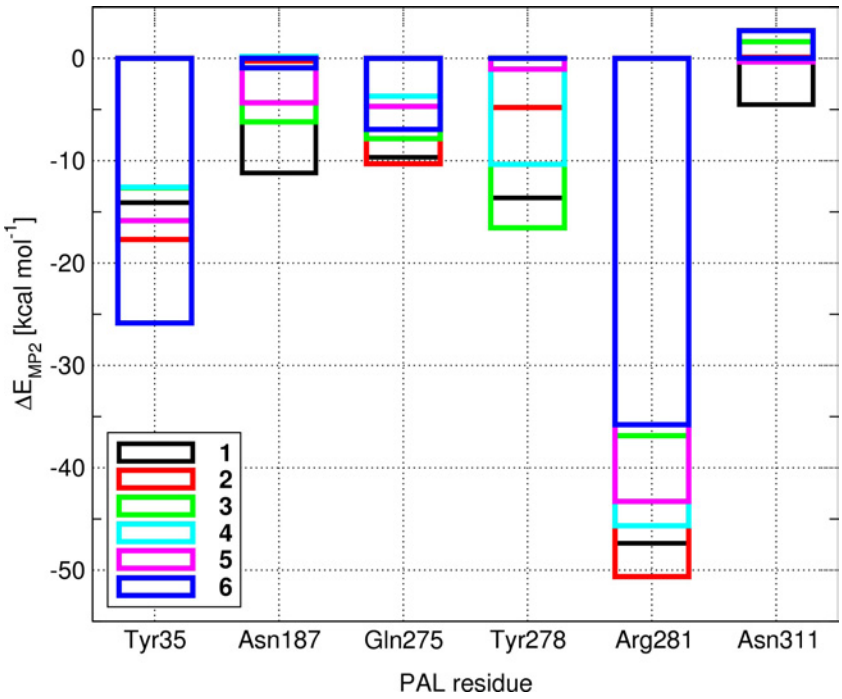


Figure 2.11: Binding contribution of particular potato PAL active site residues. *The colour* of each bar corresponds to an inhibitor according to the legend in *left bottom corner.*

themselves. In particular, binding affinity of a novel inhibitor could be estimated based on its MEP values in six points (Eq. 2.2; assuming $\Delta E_{EL,Vq}$ is in units of kcal $\cdot$ mol^{-1}) in a following way:

$$\log K_i = \frac{\Delta E_{EL,Vq} + 105.058}{10.203} \tag{2.3}$$

Inspired by Náray-Szabó electrostatic lock-and-key analogy [11], successfully applied to leucine aminopeptidase inhibitors [179], the complementarity between active site charge pattern constituting an electrostatic "lock", and MEP generated by inhibitor molecule (a "key") was examined resulting in a simple yet reasonable estimate of inhibitory activity. Symbolic representation of the fit between electrostatic lock of active site charge pattern (i.e., arbitrarily assigned set of atomic charges on selected atoms in short contact with inhibitor) and a MEP key of a given inhibitor is proposed in Figure 2.13. Noticeably, electrostatic keys of inhibitors are compatible with the PAL charge configuration constituting a lock. The best coincidence is observed for the most potent PAL inhibitor (inhibitor **1**),

Table 2.4: Coefficients of determination obtained in the absence of individual potato PAL active site residues. The latter are ranked according to the decreasing significance (i.e., the lower the r^2 value upon excluding a given residue, the greater importance of the latter).

Residue	$r^2/\Delta E_{EL}^{(1)}$	$r^2/\Delta E^{(1)}$	$r^2/\Delta E_{SCF}$	$r^2/\Delta E_{MP2}$
Tyr278	0.37	0.88	0.71	0.66
Arg281	0.76	0.53	0.74	0.74
Asn187	0.19	0.85	0.86	0.92
Tyr35	0.71	0.88	0.90	0.92
Asn311	0.71	0.66	0.94	0.96
Gln275	0.66	0.67	0.98	0.98

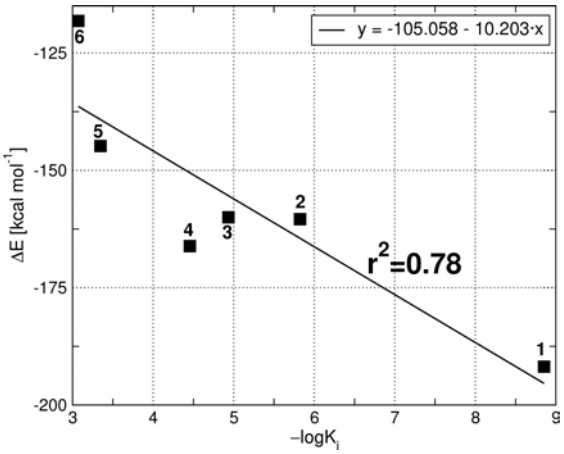


Figure 2.12: Electrostatic binding energy (according to Eq. 2.2) as a function of experimental [27, 28] inhibitory activity. *Solid line* shows the regression fit. Numbering of inhibitors corresponds to Figure 2.3.

while non-perfect fit of the remaining inhibitors is consistent with their lower biological activity. In general, differences in PAL specificity toward analyzed ligands can be explained by diversity in the MEP exhibited by particular compounds and its complementarity with receptor charges.

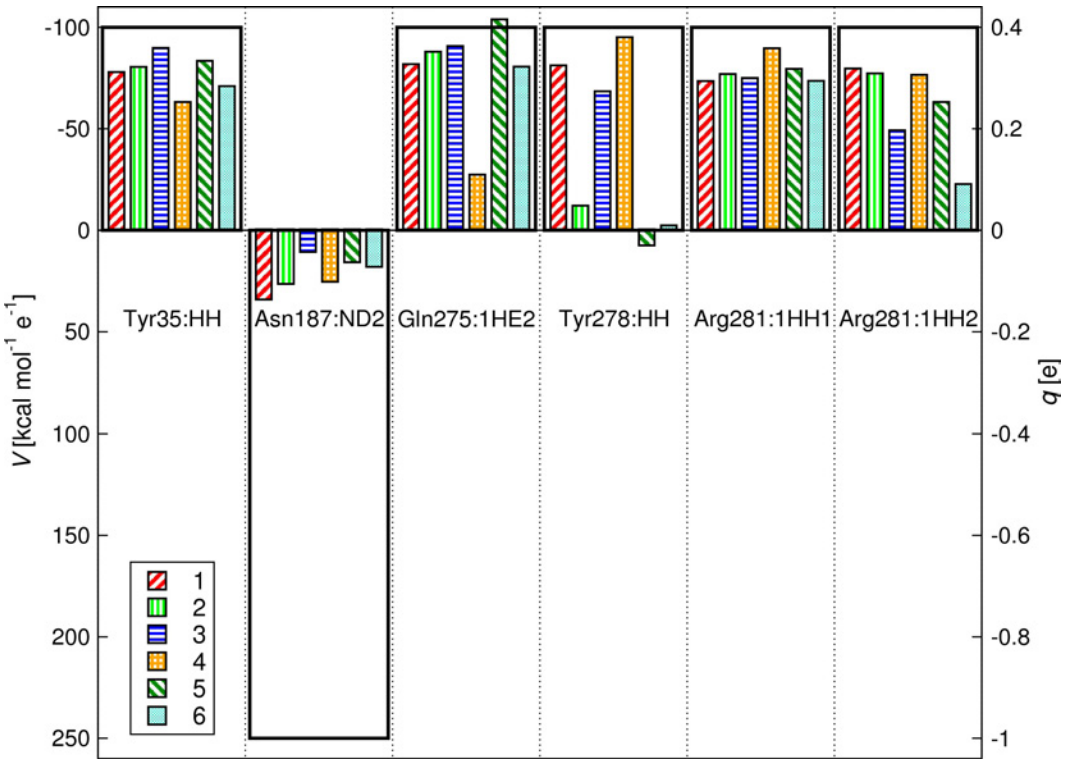


Figure 2.13: Diagram of the electrostatic lock-and-key model. *Thin bars* represent molecular electrostatic potential of a given inhibitor (a "key") in the positions of selected potato PAL atoms (*ball representation* in Figure 2.4) described by atomic point charges (*thick black bars* representing an electrostatic "lock"). Numbering of inhibitors corresponds to Figure 2.3.

Parsley PAL inhibitors Total binding energy between the residues representing parsley PAL active site (Figure 2.6) and respective parsley PAL inhibitors (Figure 2.5) was revealed in a pairwise manner at several levels of theory up to the MP2 interaction energy. Since the docking results were inconclusive with respect to the preferred binding conformation of AIP and its derivatives (see section 2.1.4), interaction energy analysis was performed for both axial and equatorial conformations (Table 2.5). As evidenced by negative values of total binding energy at the MP2 level of theory, all inhibitors considered here bind within PAL active site irrespective of their conformations. However, except for the results at the lowest level of theory (i.e., first-order electrostatic energy), AIP derivatives with axial phosphonic group are clearly favored. While the MP2 interaction energy difference between **AC** and **EC** conformers could be considered as insignificant in the case of inhibitors **II** and **III** (-3.7 and -1.6 kcal $\cdot$ mol^{-1}, respectively), it

Table 2.5: Total binding energy [$\mathrm{kcal \cdot mol^{-1}}$] of parsley PAL inhibitors calculated as a sum of two-body (i.e., inhibitor-residue) interactions at various levels of theory.

Inhibitor		$\Delta E_{EL}^{(1)}$	$\Delta E^{(1)}$	ΔE_{SCF}	ΔE_{MP2}
I		−123.36	−20.97	−67.12	−91.99
II	**AC**	−76.94	−0.82	−29.21	−53.54
	EC	−69.84	12.03	−23.04	−49.86
III	**AC**	−58.70	10.24	−21.43	−47.04
	EC	−95.76	30.32	−8.86	−45.44
IV		−100.05	25.10	−16.30	−45.52
V	**AC**	−83.86	18.83	−13.03	−42.69
	EC	−104.87	61.34	13.72	−26.01
VI	**AC**	−80.87	13.18	−17.62	−47.18
	EC	−85.97	51.38	8.47	−26.89
VII		−68.20	3.26	−23.58	−44.68
VIII	**AC**	−87.75	17.94	−16.30	−42.11
	EC	−68.25	39.31	0.58	−30.89
IX		−94.91	28.63	−10.26	−42.48
X	**AC**	−70.18	18.52	−10.34	−42.90
	EC	−104.70	82.09	32.15	−8.98
XI	**AC**	−38.53	23.65	−10.24	−35.72
	EC	−46.92	77.44	14.36	−18.42

is much more pronounced for the remaining AIP derivatives (e.g., it amounts to -33.9 $\mathrm{kcal \cdot mol^{-1}}$ in the case of inhibitor **X**). The energetic preference for axial binding conformation is consistent with an observation that only axial placement of phosphonic group allows for specific interaction with PAL active site [164].

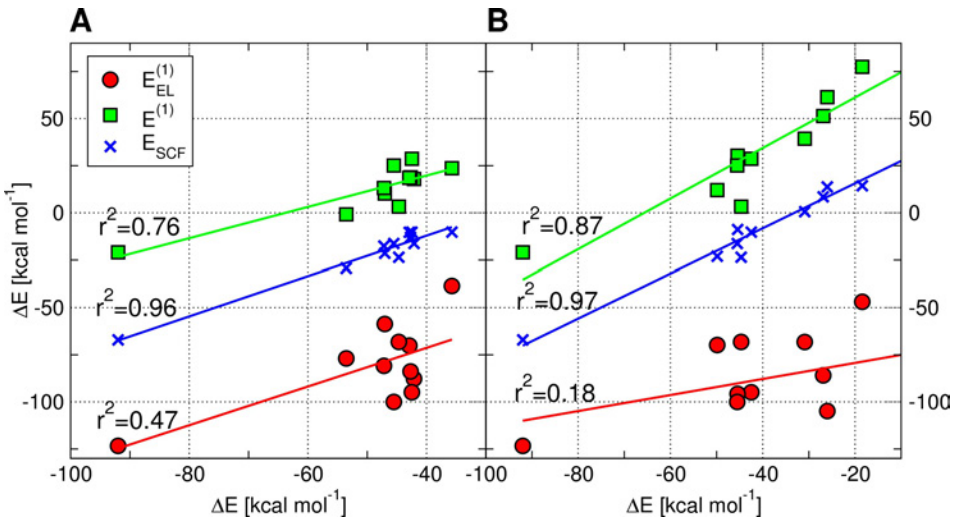


Figure 2.14: Total interaction energy at various levels of theory versus reference MP2 binding energy. The coefficients of determination corresponding to respective regression lines are given as *labels*. *A*. The entire set of inhibitors with AIP derivatives in **AC** conformation. *B*. The entire set of inhibitors with AIP derivatives in **EC** conformation.

The quality of interaction energy approximation obtained by omitting the consecutive binding energy corrections was assessed by examining correlation between the reference MP2 results and a given level of approximation (Figure 2.14). Despite the unquestionable contribution of electron correlation effects, the overall trend in SCF binding energy closely follows the changes at the MP2 level of theory. Analogous finding applies to delocalization term, the neglect of which shifts first-order interaction energy toward positive values for all but the two most potent inhibitors. Unexpectedly poor quality of $\Delta E^{(1)}_{EL}$ approximation could possibly be caused by non-optimized geometries of enzyme-inhibitor complexes. Of the 21 PAL active site residues, 11 are charged or polar. It is noteworthy, that within the set containing **EC** conformations of AIP derivatives (Figure 2.14–*B*), the correlation between electrostatic and MP2 interaction energy is even lower than in the case of **AC** conformers (Figure 2.14–*A*). As described in 2.1.4 section, all **EC** conformations of AIP derivatives bind with their ring portion buried in the hydrophobic region of PAL active site (Figure 2.7–*C*). When no apparent (e.g., ionic or hydrogen bonding) interactions are present, the task of predicting an optimal ligand positioning is even more challenging. As a consequence, non-optimal ligand pose will increase the overall noise due to structure deficiency.

The ultimate validation of docking results can only be performed by comparison to experimental data. Since no structure of the PAL-inhibitor complex is available, the reliability of theoretical results can indirectly be assessed in reference to known biological activity of ligands. The relationship between binding energy and inhibition constants is plotted in Figure 2.15 for the two sets of results differing in a conformation of AIP and its analogs. The corresponding statistical measures of the fit between theoretical and experimental data are provided in Table 2.6. Except for the $\Delta E_{EL}^{(1)}$ energy, reasonable correlation with experiment was found for interaction energy calculated at subsequent levels of theory. In particular, coefficient of determination corresponding to the ΔE_{MP2} energy is equal to 0.63 and 0.69 for the sets including axial and equatorial AIP conformers, respectively. Neglecting the electron correlation and delocalization contributions does not affect the observed correlation as neither coefficients of determination nor standard errors of estimate are changed in a way indicating the weakening of a mutual relationship. Poor performance of the electrostatic energy based model might stem from the considerable involvement of nonpolar residues and/or the insufficient quality of receptor-ligand geometries.

While the overall fit concerning **AC** and **EC** series binding energy versus inhibitory activity is similar in terms of the respective r^2 values, standard errors of estimate are larger in the case of a set including AIP derivatives in **EC** conformation. This observation seems to support a postulated lower accuracy of the corresponding structures due to more difficult docking in a binding pocket region occupied by this particular conformers. Taken together with the less favorable binding of the **EC** series, these results confirm that AIP derivatives are more likely to bind in an axial conformation. Accordingly, further analysis is focused on the set of 11 parsley PAL inhibitors with AIP analogs in **AC** conformation, exclusively.

Compared to an excellent agreement with experiment obtained in a previous study concerning potato PAL inhibitors, the quality of results presented here might be viewed as inadequate. It should be emphasized, however, that the current analysis encompasses almost twice as many inhibitors, as previously, and their binding manner includes two possible arrangements within PAL active site. Consequently, the latter is represented by 21 amino acid residues compared to 6 residues employed in the case of potato PAL inhibition. While enlarged model of binding pocket seems to provide a better description, it also increases the noise arising from any inaccuracies in the structures, which underwent no post-docking treatment. With these facts in mind, the correlation between *ab initio* binding energy and experimental activity observed here can be considered as remarkable. When limited to 7 closely related compounds (i.e., AIP and its analogs), even stronger relationship with experiment might be established. In particular, r^2 and SEE parameters amount to 0.77 and 2.91, 0.84 and 2.98, or 0.79 and 4.03 kcal $\cdot$ mol^{-1} for the ΔE_{MP2}, ΔE_{SCF}

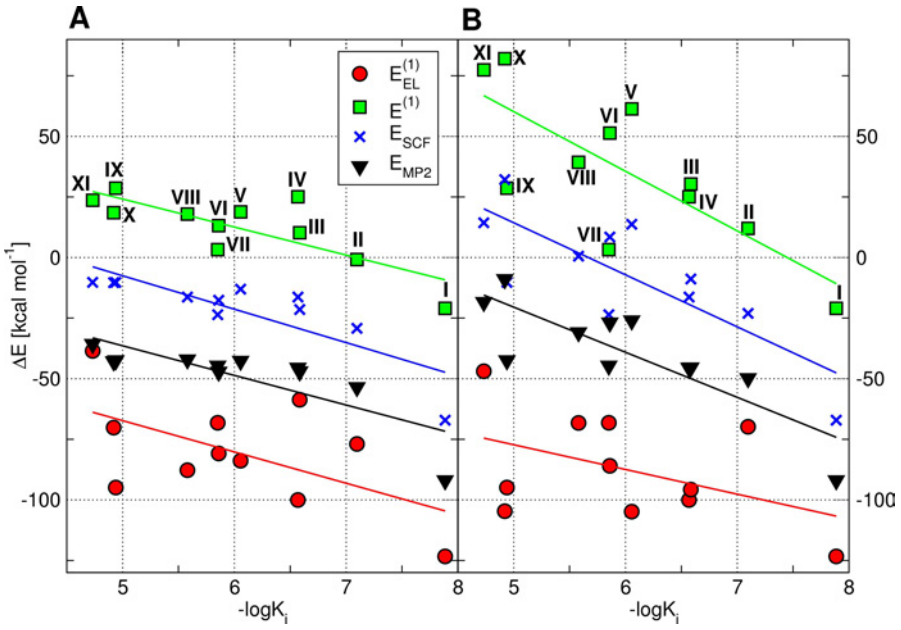


Figure 2.15: Total interaction energy at various levels of theory as a function of inhibitory activity [29]. *A.* The entire set of inhibitors with AIP derivatives in **AC** conformation. *B.* The entire set of inhibitors with AIP derivatives in **EC** conformation. Numbering of inhibitors corresponds to Figure 2.5.

or $\Delta E^{(1)}$ energy, respectively. An improvement in coefficients of determination is therefore accompanied by significant decrease in standard errors of estimate (see the results for a full set of inhibitors—Table 2.6). Contrary to expectations, analogous parameters for $\Delta E_{EL}^{(1)}$ energy (i.e., 0.15 and 17.45 kcal $\cdot$ mol^{-1}) show weaker correlation than when calculated for all inhibitors. However, the same analysis performed without inhibitor **III** (as the only AIP derivative occupying more hydrophobic region of a binding site) results in a retained correlation at the MP2, SCF and Hitler-London levels of theory, while statistical measures associated with the first-order electrostatics are improved (0.33 and 16.36 kcal $\cdot$ mol^{-1} for r^2 and SEE, respectively). This finding suggests that incorrect electrostatic description might be limited to ligands populating the highly nonpolar part of PAL active site. Indeed, examining the correlation without inhibitors bound in the hydrophobic region (i.e., compounds **III**, **IV**, **VII**, and **IX**) produces little variation in statistical parameters corresponding to the ΔE_{MP2}, ΔE_{SCF}, and $\Delta E^{(1)}$ energy based models but considerable improvement is achieved for electrostatic description, at least concerning the coefficient of determination (r^2 and SEE equal to 0.66 and 16.14 kcal $\cdot$ mol^{-1}, respectively).

Table 2.6: Performance of the interaction energy-based models of inhibitory activity. Regression equations and statistical parameters correspond to the fit shown in Figure 2.15. Standard errors of estimate, SEE, are given in $\mathrm{kcal} \cdot \mathrm{mol}^{-1}$.

Model		Regression equation	r^2	SEE	q^2	SEE_{cv}
ΔE_{MP2}	**AC**	$y = -12.2276 \cdot x + 24.7344$	0.63	9.56	0.65	14.93
	EC	$y = -18.6102 \cdot x + 72.5917$	0.69	12.71	0.69	17.54
ΔE_{SCF}	**AC**	$y = -13.7960 \cdot x + 61.4688$	0.68	9.70	0.69	14.83
	EC	$y = -21.4691 \cdot x + 121.7004$	0.63	16.87	0.62	22.06
$\Delta E^{(1)}$	**AC**	$y = -11.5528 \cdot x + 81.9004$	0.63	9.21	0.62	12.02
	EC	$y = -24.6126 \cdot x + 183.2912$	0.59	21.19	0.58	25.41
$\Delta E^{(1)}_{EL}$	**AC**	$y = -12.9157 \cdot x - 2.7218$	0.32	19.45	0.32	25.52
	EC	$y = -10.2011 \cdot x - 26.2528$	0.20	20.75	0.21	27.09

Establishing the correlation with experiment, apart from validation purposes, was also aimed at exploring the predictive ability of interaction energy based models employing the first principles instead of a empirical parametrization. When explicit validation set is not available, one can assess the significance of the derived correlation equations by cross-validation approach, e.g. with the leave-one-out method. Analysis of the cross-validated r^2 and SEE values (i.e., q^2 and SEE_{cv}) provides an estimate of the predictive performance of a given model. In particular, q^2 value exceeding 0.5 indicates an acceptable predictive potential, while q^2 greater than 0.6 is considered quite significant [180]. However, due to the limited size of a data set employed here, care should be taken when using the statistical estimates for purposes other than assessment of a relative performance of approximate models derived from the reference one (i.e., the MP2 interaction energy).

Cross-validated measures of the strength of correlation equations provided in Table 2.6 show the statistical significance of relationships involving interaction energy calculated at the MP2, SCF and first-order levels of theory. Accordingly, these models could be employed in binding affinity prediction. Electrostatic energy based model exhibits insufficient quality for any prognostic applications. Comparison of the experimental inhibition constants [29] with those calculated using the regression equations derived for the MP2 and SCF results is presented in Figure 2.16.

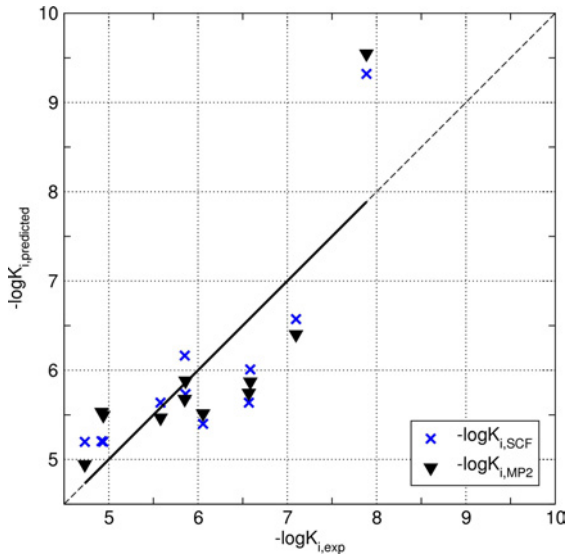


Figure 2.16: Predicted versus experimental [29] inhibitory activity of parsley PAL inhibitors. *Solid line* shows the regression fit.

The qualitative agreement between theoretical and experimental inhibitory activity supports the possible application of such models in binding affinity prediction.

In practice, successful application of a given model to the prediction of ligand affinity demands both its reliability and low computational cost. While extremely useful in gaining insight into physical nature of interactions and providing the molecular basis of inhibition phenomena, a procedure employing *ab initio* interaction energy is of no use when it comes to virtual screening of a large number of compounds. However, understanding the intrinsic factors underlying binding strength and specificity allows for systematic derivation of approximate models. Such models could be generated by neglecting the computationally expensive corrections to binding energy (as shown above) and/or by limiting the receptor site representation. Additional advantage of the latter approach is that the quality of a model might actually improve as amino acid residues of minor importance introduce unnecessary errors arising from imperfect geometries.

The aspects considered during the assessment of a residue significance included its contribution to ligand binding (Figure 2.17) and changes in r^2 values upon excluding a given residue (Table 2.7). According to the expectations, Arg354 residue involved in an ionic pair with phosphonic group of all inhibitors contributes the major part of total binding energy. The only exception is due to positively charged inhibitor **XI** being stabilized mainly by Glu484 residue bearing a negative

Table 2.7: Coefficients of determination obtained in the absence of individual parsley PAL active site residues. The latter are ranked according to the decreasing significance (i.e., the lower the r^2 value upon excluding a given residue, the greater importance of the latter).

Residue	$r^2/\Delta E_{EL}^{(1)}$	$r^2/\Delta E^{(1)}$	$r^2/\Delta E_{SCF}$	$r^2/\Delta E_{MP2}$
Arg354	0.01	0.12	0.10	0.07
Lys456	0.22	0.27	0.32	0.26
Glu484	0.19	0.38	0.62	0.52
Phe116	0.28	0.53	0.62	0.59
Asn384	0.28	0.55	0.64	0.59
Asn487	0.34	0.61	0.64	0.59
Leu134	0.34	0.66	0.67	0.61
Gln348	0.31	0.56	0.64	0.61
Gln488	0.32	0.62	0.66	0.61
Leu138	0.34	0.62	0.67	0.62
Leu206	0.35	0.59	0.66	0.62
Phe137	0.31	0.64	0.69	0.64
MIO203	0.35	0.67	0.66	0.64
Leu256	0.30	0.61	0.69	0.64
Pro385	0.35	0.58	0.67	0.64
Phe400	0.37	0.58	0.67	0.64
Ile460	0.32	0.66	0.69	0.64
Gly115	0.32	0.62	0.71	0.66
Asn260	0.36	0.67	0.72	0.66
His396	0.31	0.61	0.71	0.66
Tyr351	0.36	0.72	0.72	0.67

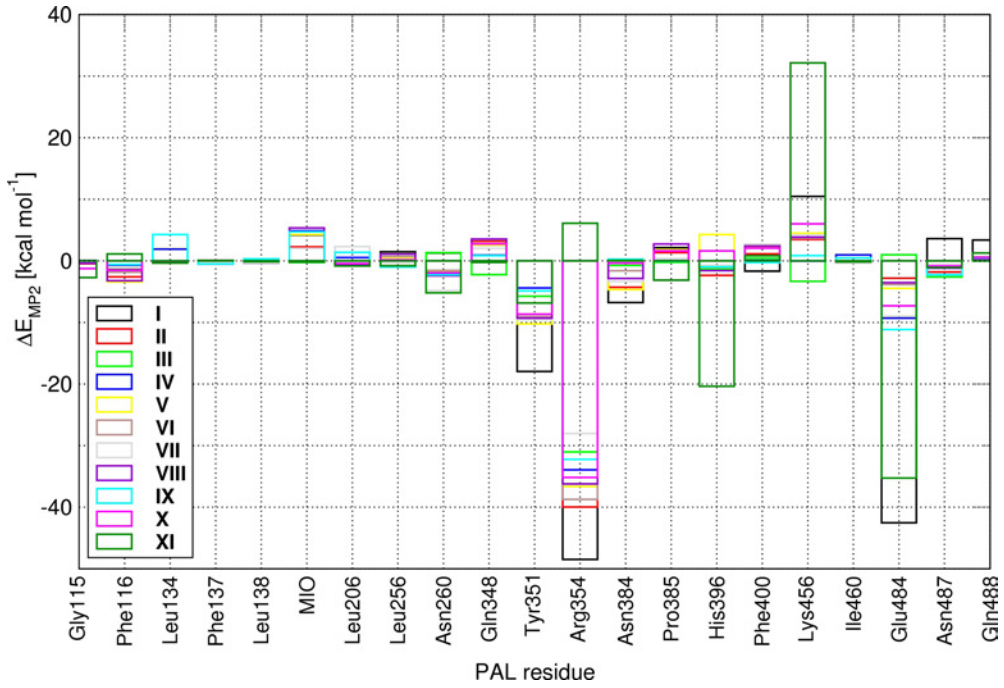


Figure 2.17: Binding contribution of particular parsley PAL active site residues. *The colour* of each bar corresponds to an inhibitor according to the legend in *left bottom corner*.

charge. The significance of Arg354 is also evidenced by the loss of correlation when this particular residue is neglected. Contribution of the remaining charged residues, Glu484 and Lys456, is less substantial and much more dependent on an actual inhibitor. In terms of importance for retaining correlation, these two residues are also significant. Comparison of the changes in coefficients of determination provided in Table 2.7 with the initial values for a full model (Table 2.6) reveals that for 10 residues their neglect actually improves the correlation. This finding might provide clues to determination of residues with incorrect geometry. It could also guide the selection of residues allowed to be flexible during the docking procedure.

Since majority of PAL active site components has been shown to contribute little to the overall correlation with experiment, several selection schemes could be employed while determining the most optimal receptor site representation. As an example, neglecting 10 nonpolar residues results in a minor decrease in the strength of correlation (Table 2.8). Noticeably, including only 3 charged residues still gives an acceptable agreement with experiment. Significant improvement in electrostatic description achieved in the last case has suggested a possibility of building simple inhibitory activity model employing fixed atomic point charges

Table 2.8: Performance of the inhibitory activity models encompassing the limited size representation of parsley PAL active site.

Model		$\Delta E_{EL}^{(1)}$	$\Delta E^{(1)}$	ΔE_{SCF}	ΔE_{MP2}
21 PAL residues	r^2	0.32	0.63	0.68	0.63
	SEE	19.45	9.21	9.70	9.56
11 PAL residues (charged or polar)	r^2	0.42	0.51	0.64	0.61
	SEE	16.84	8.88	9.44	10.01
3 PAL residues (Arg354, Lys456, Glu484)	r^2	0.50	0.41	0.51	0.51
	SEE	17.23	13.61	14.26	14.00

as a representation of Arg354, Lys456, and Glu484. In particular, two +0.5 point charges, a single +1.0 charge, and two −0.5 charges were associated with the positions of hydrogen or oxygen atoms of the respective guanidinium, amino, and carboxylic moieties (Figure 2.18). Molecular electrostatic potential of particular inhibitors generated at the positions of these 5 point charges was then employed in calculation of the approximate electrostatic binding energy. The resulting regression equation (see Figure 2.19) exhibits the same quality of correlation as the full electrostatic description employing 3 charged PAL residues (r^2 and SEE equal to 0.51 and 17.57 kcal $\cdot$ mol^{-1}, respectively). Noteworthily, such a model allows for a rapid prediction of the inhibitory activity using the following relationships (assuming that the interaction energy is in the units of kcal $\cdot$ mol^{-1}):

$$\log K_i = \frac{\Delta E_{EL,Vq} - 49.437}{17.360} \tag{2.4}$$

$$\begin{aligned} \Delta E_{EL,Vq} \;=\; & 0.5 \cdot (V_{Arg354:HH11} + V_{Arg354:HH21}) \\ & +1.0 \cdot V_{Lys456:HZ2} \\ & -0.5 \cdot (V_{Glu484:OE1} + V_{Glu484:OE2}) \end{aligned} \tag{2.5}$$

Previously described potato PAL results regarding the approximate models of inhibitory activity implied that sufficient accuracy is retained while using limited active site model composed of 4 residues (**B** model; Table 2.2). Herein, 3 residues are required to represent parsley PAL receptor site (Table 2.8). Only one residue

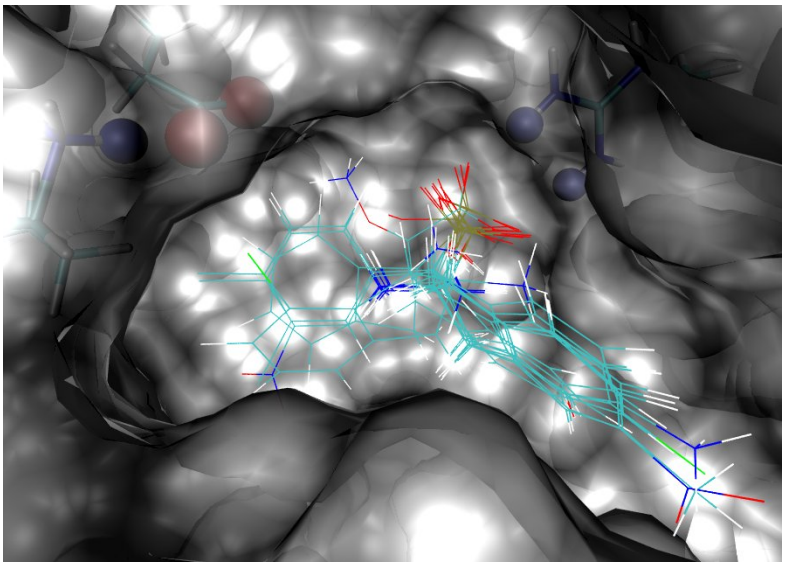


Figure 2.18: Parsley PAL active site atoms (shown as *blue and red balls*) selected to mimic the electrostatic properties of a binding pocket.

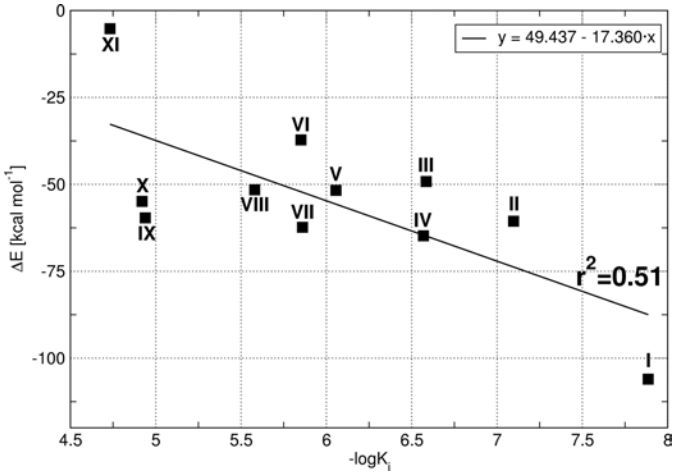


Figure 2.19: Electrostatic binding energy (according to Eq. 2.5) as a function of experimental [29] inhibitory activity. *Solid line* shows the regression fit. Numbering of inhibitors corresponds to Figure 2.5.

(i.e., potato PAL Arg281 or parsley PAL Arg354) is common in these two limited size representations of a binding pocket. Such a difference concerning simplified inhibitory activity models might result from various sets of inhibitors analyzed in the case of potato and parsley PAL enzymes (Figures 2.3 and 2.5, respectively) and/or the size and quality of active site models (see Figures 2.4 and 2.6 for potato and parsley PAL representations of a binding cavity). The findings obtained in the case of parsley PAL appear to be more general, as the enzyme model was based on experimental structure and larger series of inhibitors was accounted for. While direct comparison of these two sets of results is impossible (due to different size of data samples and various procedures employed during preparation of enzyme-inhibitor structures), it seems likely that the predictive ability of approximate models of inhibitory activity is reasonable provided that one is interested in derivatives of compounds included in the set analyzed, while building a given model. In other words, what can be done with inhibitory activity models is interpolating rather that extrapolating.

2.2 Urokinase Inhibition

2.2.1 Urokinase characteristics

Urokinase-type plasminogen activator (urokinase, uPA; E.C. 3.4.21.73) is a trypsin-like serine protease involved in plasminogen activation. Plasmin, an active form of plasminogen, is responsible for numerous proteolytic events encompassing degradation of extracellular matrix proteins, thereby triggering the process of cell migration. Accordingly, urokinase has been implicated in the pathogenesis of a variety of diseases that require tissue remodeling and/or its abnormal destruction including wound healing [181, 182], atherosclerosis [183, 184], multiple sclerosis [185] and metastatic processes in cancer [186–188]. The latter connotation is especially promising as uPA plays an important role also in angiogenesis (i.e., the growth of new blood vessels), a process exploited in tumor progression and metastasis [189]. Hence, urokinase has become a target of potential importance for specific and non-cytotoxic cancer therapy [190].

2.2.2 Inhibitors of urokinase

Trypsin-like serine proteases are responsible for the cleavage of peptide bonds following a positively charged amino acid residue. Accordingly, the S1 binding site of uPA active site [191] contains an aspartic acid (Asp189) residue recognizing

the key arginine of plasminogen. In contrast to other trypsin-like serine proteases possessing a hydrophobic alanine residue adjacent to Asp189 (e.g., thrombin), the binding cavity of urokinase contains a hydrophilic serine (Ser190), a feature that might be used in design of selective uPA inhibitors. As a consequence of uPA binding preferences, urokinase inhibitors encompass mainly arginine mimetics such as guanidines and amidines [192]. While α-aminoalkylphosphonate diphenyl esters (serving as phosphonic analogues of amino acids) are known as specific, essentially irreversible inhibitors of serine proteases [193], only a few papers reported application of these phosphonic compounds as uPA inhibitors [194–196]. Additionally, the mode of urokinase-phosphonic inhibitors interaction had not been established experimentally. Availability of experimental human uPA-directed inhibitory activity of several α-aminophosphonates [30, 195, 197] resulted in an attempt to explain the molecular basis of urokinase inhibition by means of molecular docking followed by *ab initio* analysis of the physical nature of interactions taking place in uPA active site [198]. It is worth mentioning that theoretical results presented herewith were compared to experimental inhibitory activities obtained consistently under identical conditions [30].

Technical details The structures of inhibitors considered herein are presented in Figure 2.20. Experimental ligand activity data apply to the racemic mixtures of particular compounds [30]. Despite relatively minor structural changes, experimental IC_{50} values span the range of five orders of magnitude. The docking part of this project was performed by Renata Grzywa (for more details see Ref. [198]) and employed crystal structure of urokinase [199] (1.65 Å resolution; PDB accession code 1C5Y). Noticeably, the final mode of receptor-ligand binding [198] is in agreement with the general knowledge regarding inhibition of serine proteases by α-aminoalkylphosphonate diphenyl esters [200]. In particular, the distance between inhibitor phosphorus atom and Ser195 alcohol group is consistent with the irreversible inhibition via enzyme phosphonylation (i.e., covalent modification of Ser197 from catalytic triad). Guanidino or amidino moieties were found to occupy S1 pocket and form hydrogen bonds with Asp189, Ser190, and Gly219 residues (Figure 2.21. Analogous theoretical results published recently [201] further confirm the correct mode of phosphonic inhibitors binding as obtained from the docking procedure applied here [198].

Urokinase binding pocket employed in nonempirical analysis of interaction energy consisted of 9 amino acid residues: $Asp^{-}189$, Ser190, Gln192, Val213, Gly216, Gly219, Cys220, Ala221, and Pro225. To reduce the size of a system, selected molecular fragments most distant from an inhibitor molecule were neglected (see Figure 2.21 for details). In particular, the main chain amino group of Asp189 was

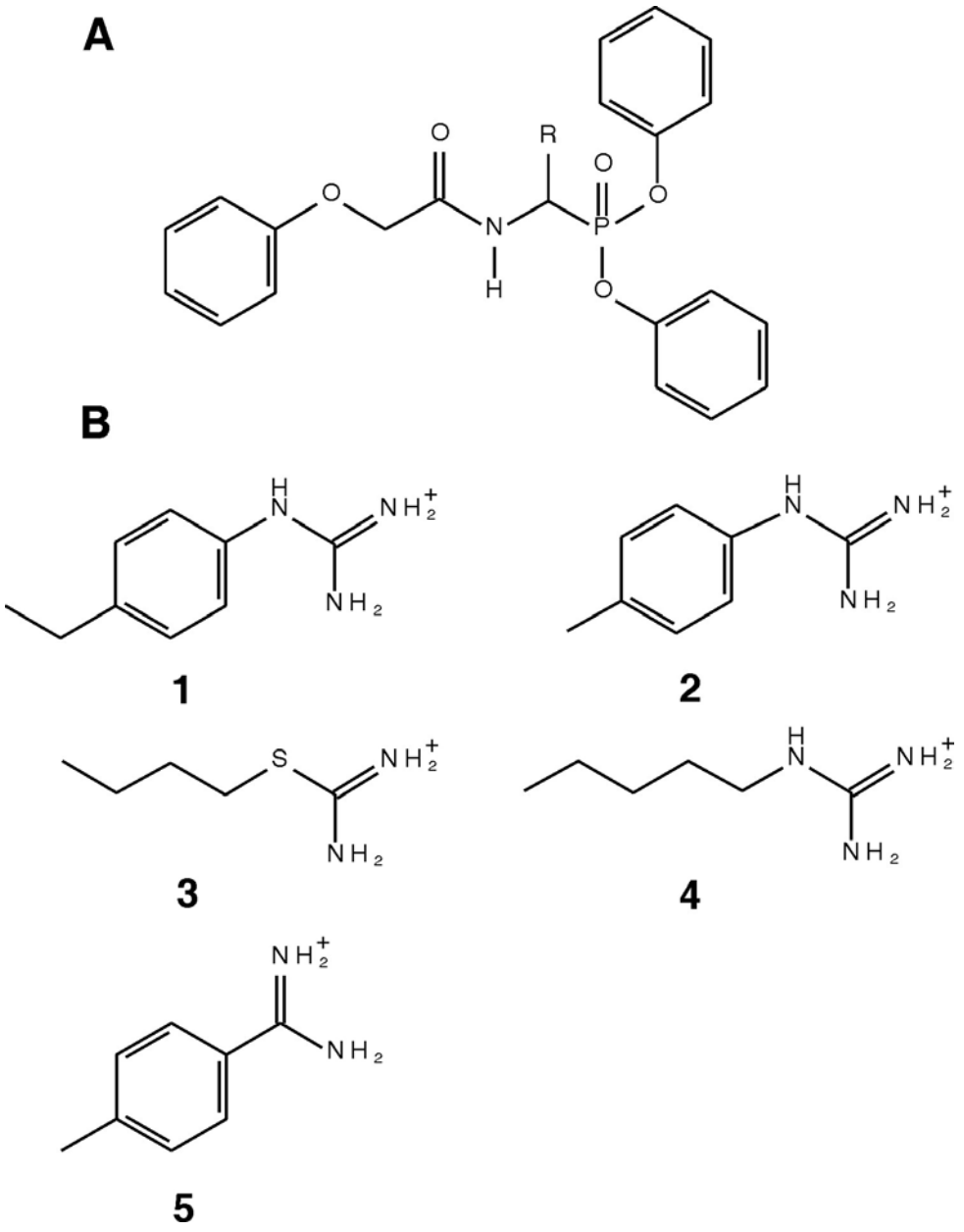


Figure 2.20: Structures of α-aminoalkylphosphonate diphenyl esters known as urokinase inhibitors in the order of decreasing inhibitory activity [30]. *A.* General scheme with a variable part marked as R group. *B.* The detailed structures of R substituents belonging to particular inhibitors.

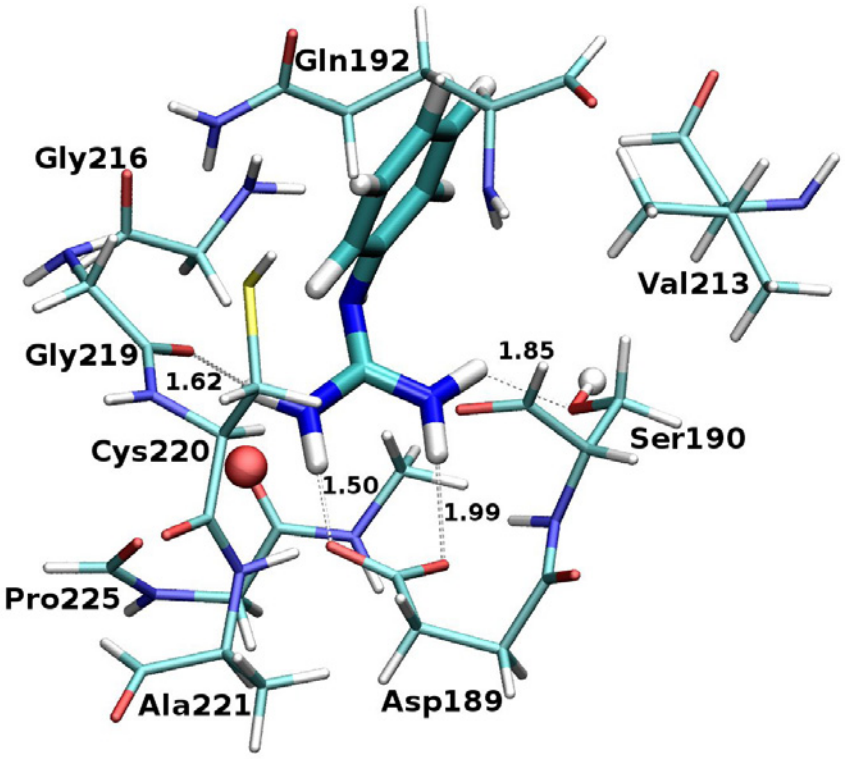


Figure 2.21: Urokinase active site model in complex with inhibitor **2**. The shortest atomic contacts [Å] are marked as *dashed lines*. *Ball representation* is used to distinguish 2 atoms with atomic point charges employed in an approximate electrostatic model of inhibitory activity.

replaced by a hydrogen atom. Since Pro225 residue interacts with an inhibitor primarily via its main chain atoms, only the latter were considered along with a main chain carbonyl group of the preceding residue (i.e., Lys224) as well as an amino group of the succeeding Gly226 residue. All broken bonds were saturated with hydrogen atoms. The inhibitor structures were also simplified to include only the variable part of α-aminophosphonates considered in this study (i.e., the substituents presented in Figure 2.20–*B*). All inhibitor molecules carried a positive charge (+1), whereas uPA residues were modelled as neutral species (except for negatively charged aspartate $Asp^{-}189$). Stabilization energy was then evaluated in a pairwise manner with total binding energy of particular inhibitors being the sum of two-body interactions. The detailed protocol followed a description provided in *Technical details* paragraph of subsection 2.1.3.

Table 2.9: Total binding energy [kcal·mol^{-1}] of uPA inhibitors calculated as a sum of two-body (i.e., inhibitor-residue) interactions at various levels of theory.

Inhibitor	$\Delta E_{EL}^{(1)}$	$\Delta E^{(1)}$	ΔE_{SCF}	ΔE_{MP2}
1	−215.53	−41.06	−110.25	−135.83
2	−174.01	−42.25	−97.91	−122.60
3	−197.24	−45.69	−109.88	−133.22
4	−165.80	−58.30	−103.22	−124.03
5	−122.63	−73.63	−100.55	−119.59

2.2.3 Interaction energy analysis

Interaction energy analysis was performed for the uPA-inhibitor complexes derived from docking calculations. Due to the significant size of an active site model composed of 9 residues (Figure 2.21), binding energy was evaluated as a sum of pairwise interactions between respective inhibitor molecules and amino acid residues. Interaction energy was further partitioned allowing for the magnitude of particular binding energy terms to be examined (Table 2.9). In addition to the study of a physical nature of the total observed interaction, this analysis also aimed at the determination of uPA residues responsible for inhibitor specificity. Accordingly, the smaller active site model (referred to as **B** in contrast to the full 9-residues model denoted by **A**) was derived by stepwise neglect of 7 residues with minor contribution to overall ligand specificity. The results for active site **A** and **B** models are presented in Figure 2.22.

Remarkable correlation with experiment [30] was found for the first-order electrostatic energy (coefficient of determination for the relationship with experimental inhibitory activity is equal to 0.88, Table 2.10). The subsequent stabilization energy terms exhibit rather poor agreement with experimental data probably due to artificially shortened distances in force field-optimized complexes (see the last paragraph of this section). However, interaction energy at the MP2 level of theory is still in a qualitative agreement with experiment suggesting that the overall model of inhibitors binding is reasonable. Apparently, the cancellation of higher corrections to binding energy results in a good performance of the electrostatic term. Exchange and delocalization components of stabilization energy are much more sensitive to the actual distance between interacting species and, thus, any errors in the latter are much more pronounced.

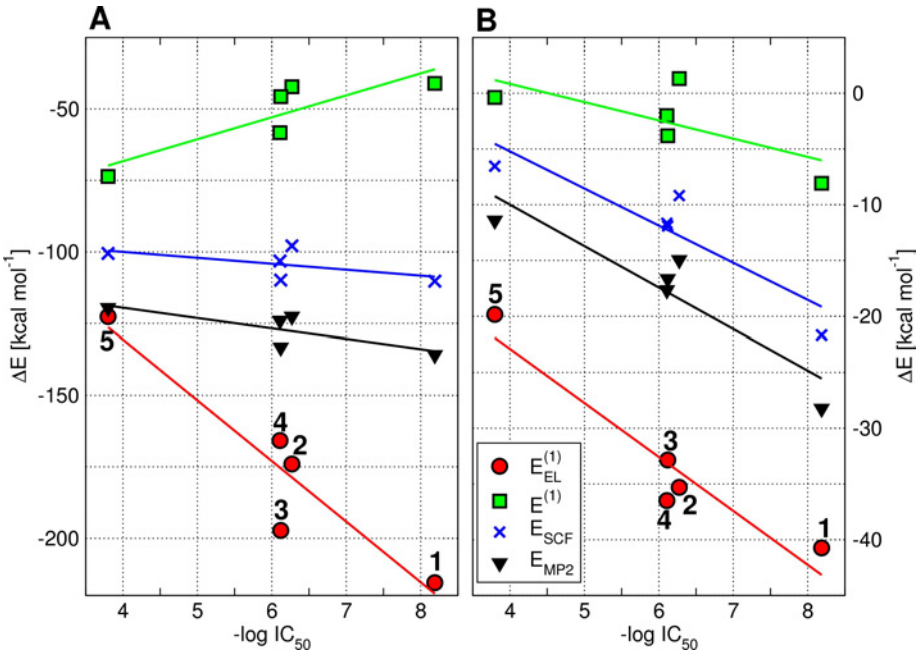


Figure 2.22: Total interaction energy at various levels of theory as a function of inhibitory activity [27, 28]. *A*. Full model of uPA active site (i.e., 9 residues; **A** model). *B*. uPA active site represented by 2 residues (**B** model). Numbering of inhibitors corresponds to Figure 2.20.

Since several uPA residues were found to interact with all inhibitors to a similar extent, limited-size model of uPA active site (**B** model) was constructed including only Ser190 and Pro225. The quality of such an approximation was validated in terms of its capability to retain the initial correlation with experimental inhibitory activity (Figure 2.22–*B* and Table 2.10). Noticeably, the first-order electrostatic energy is able to reproduce experimental binding affinity with r^2 and SEE of 0.91 and 2.81 kcal$\cdot$mol^{-1}, respectively. As regards the stabilization energy described by the higher levels of theory, **B** model provides better correlation with experiment, especially concerning the ΔE_{SCF} and ΔE_{MP2} terms. Lower r^2 value associated with the **B** model Heitler-London energy actually indicates an improvement, since the regression equation referring to the $\Delta E^{(1)}$ term features an incorrect, positive value of a slope in the case of **A** model. An observed increase in the strength of correlation upon reducing the size of an active site representation also suggests that the full model might suffer from imperfect geometries of inhibitor-residue complexes.

Table 2.10: Performance of the interaction energy-based models of inhibitory activity. Regression equations and statistical parameters correspond to the fit shown in Figure 2.22. Standard error of estimate, SEE, is given in $\mathrm{kcal} \cdot \mathrm{mol}^{-1}$.

Model		Regression equation	r^2	SEE
ΔE_{MP2}	**A**[a]	$y{=}-3.5924 \cdot x - 105.1530$	0.63	4.99
	B[b]	$y{=}-3.7114 \cdot x + 4.8452$	0.84	2.92
ΔE_{SCF}	**A**	$y{=}-2.0663 \cdot x - 91.7633$	0.34	5.20
	B	$y{=}-3.3270 \cdot x + 8.0899$	0.82	2.82
$\Delta E^{(1)}$	**A**	$y{=}7.6775 \cdot x - 98.9936$	0.75	7.97
	B	$y{=}-1.6388 \cdot x + 7.3970$	0.50	2.95
$\Delta E^{(1)}_{EL}$	**A**	$y{=}-21.1891 \cdot x - 45.8529$	0.88	14.29
	B	$y{=}-4.8366 \cdot x - 3.5550$	0.91	2.81

[a]Full uPA active site model (9 amino acid residues)
[b]Limited size uPA active site model (2 amino acid residues)

Due to the formation of an ionic pair with positively charged guanidinylated or amidinylated inhibitor moieties, the strongest interaction is observed with the Asp189 residue. However, the strength of binding is similar in all these cases and Asp189 exclusion from **B** model does not affect the agreement with experimental data. Therefore, Asp189 residue is responsible rather for the overall positioning and strong binding of inhibitor molecules, while it appears to contribute little to substrate specificity, at least as long as the compounds analogous to those considered here are taken into account.

Sufficient performance of a limited size **B** model composed of as few as 2 residues along with the correct description provided by the $\Delta E^{(1)}_{EL}$ interaction energy suggested that even more approximate electrostatic model could be derived. In analogy to the simple electrostatic models obtained for potato and parsley PAL inhibitors, atomic point charges of -0.4 and 0.1 were arbitrarily assigned to Pro225 carbonyl oxygen and Ser190 hydroxyl hydrogen atoms (ball representation in Figure 2.21). Electrostatic interaction energy was then calculated based on the molecular electrostatic potential of inhibitors in the positions of the two selected atoms:

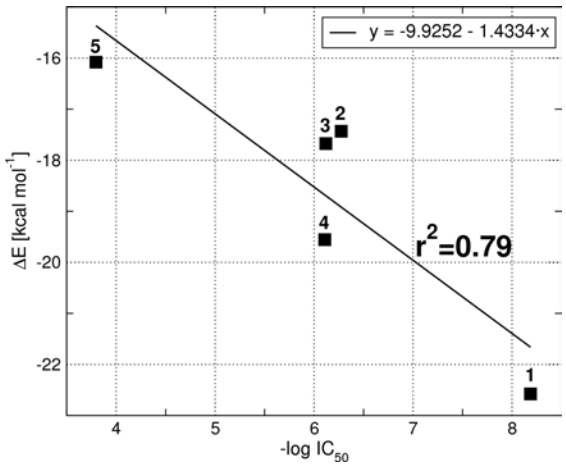


Figure 2.23: Electrostatic binding energy (according to Eq. 2.6) as a function of experimental [30] inhibitory activity. *Solid line* shows the regression fit. Numbering of inhibitors corresponds to Figure 2.20.

$$\Delta E_{EL,Vq} = -0.4 \cdot V_{Pro225:O} + 0.1 \cdot V_{Ser195:HG1} \tag{2.6}$$

As presented in Figure 2.23, $\Delta E_{EL,Vq}$ values are in a good agreement with the experimental ligand activity. The corresponding r^2 and SEE values equal to 0.79 and 1.34 kcal $\cdot$ mol^{-1}, respectively, indicate usefulness of a model in prediction of the inhibitory activity according to the formula (applying to the $\Delta E_{EL,Vq}$ energy expressed in the units of kcal $\cdot$ mol^{-1}):

$$\log IC_{50} = \frac{\Delta E_{EL,Vq} + 9.925}{1.433} \tag{2.7}$$

The loss of correlation with experimental data for the first-order Heitler-London energy suggested the improper (i.e., too short) distances between the interacting monomers. Nonempirical analysis of the interaction energy terms along the selected contacts was performed by Mikołaj Feliks [198]. It appears that the minimum energy distances from force field-optimized complexes are about 0.5 Å shorter than the corresponding *ab initio* values (see sample results in Figure 2.24). As a result of artificially high exchange repulsion term, the first order interaction energy is overestimated and does not correlate with the experimentally determined inhibitory activity. It can be seen from Figure 2.24, that the exchange component of binding energy is the most distance-dependent term. While the difference in the MP2 interaction energies for force field and *ab initio* optimized structures

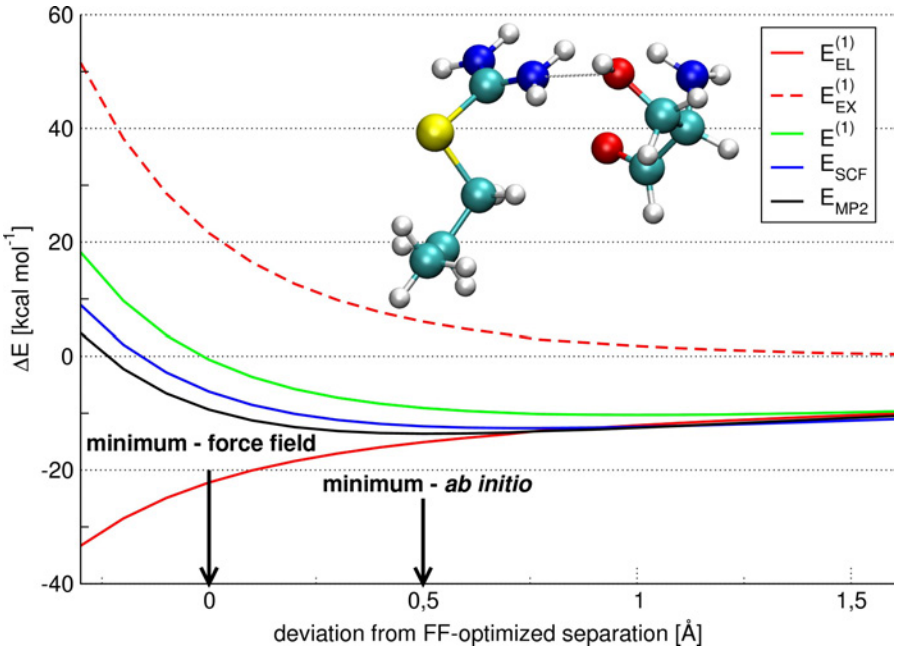


Figure 2.24: Binding energy terms of the interaction between inhibitor **3** and Ser190 residue [198]. The distance between monomers was sampled by 0.1 Å in both directions. Zero value corresponds to the distance found in a final docked complex (i.e., force field optimized structure). Optimal force field and *ab initio* separations are denoted by *arrows*.

is equal to 4.2 kcal $\cdot$ mol^{-1}, the corresponding value for the exchange energy is 15.6 kcal $\cdot$ mol^{-1}. This observation can be employed to explain the sudden drop in correlation with experiment when passing from the electrostatic to the Heitler-London interaction energy.

CHAPTER 3

Enzyme Catalytic Activity

3.1 Kinase-catalyzed Phosphorylation Reaction

The great interest in the catalytic properties of kinases originates from their participation in multiple metabolic and regulatory pathways. Despite facilitating essentially the same phosphoryl transfer reaction, kinases exhibit remarkable diversity in their structure and substrate specificity [202]. The latter implies either that the same (or similar) reaction can be accomplished in several ways or that analogous catalytic mechanisms are achieved with the involvement of different active site residues that together exhibit some common characteristics. The most extensively studied kinases are protein kinases transferring a terminal phosphate group from adenosine triphosphate (ATP) to protein substrates [203]. On the other hand, small molecule kinases, such as those representing the ribokinase superfamily [204], are much less frequently investigated and little is known about the molecular basis of their catalytic activity.

3.1.1 Hydroxyethylthiazole kinase

4-Methyl-5-β-hydroxyethylthiazole kinase (Thz kinase, ThiK, E.C. 2.7.1.50) is a member of the ribokinase family of sugar kinases that phosphorylate substrates containing a hydroxymethyl group [116]. ThiK-catalyzed phosphorylation of the thiazole moiety constitutes a salvage branch in the biosynthetic pathway of thiamine. Despite significant similarity to other ribokinase-like kinases (e.g., ribokinase [205], adenosine kinase [206], pyridoxal kinase [207]), ThiK lacks the highly

conserved aspartate residue (replaced by Cys198) otherwise believed to serve as a catalytic base during the phosphorylation process. Surprisingly, a cysteine to aspartate mutation results in a ninefold increase in ThiK enzymatic activity [116]. Other features of the ThiK active site, including the presence of magnesium ion(s), a positive helix dipole and an oxyanion hole, as well as the overall positioning of ligands within existing X-ray structures, are consistent with an S_N2-like mechanism comprising a direct nucleophilic attack of a substrate hydroxyl group oxygen on the γ-phosphorus of ATP [208]. Nevertheless, this hypothesis has been tested neither experimentally nor via theoretical modeling. Among the questions that remained unanswered is the actual number of magnesium ions required for catalysis as well as the role of Cys198 and Arg121 residues. The guanidinium moiety of the latter provides an additional electrostatic stabilization of β-phosphate—the interaction specific to Thz kinases.

The purpose of this study [209] was to test the suggested pathway of ThiK-catalyzed phosphorylation of 4-methyl-5-β-hydroxylethylthiazole (Thz) and to determine the possible involvement in catalysis of specific ThiK active site components. By rigorous *ab initio* calculations of the interaction energy, the concept of differential transition state stabilization was evaluated. Furthermore, application of interaction energy decomposition enabled determination of both the physical nature of binding and the residues that are crucial for catalytic activity. As a result, approximate, yet well-founded models can be constructed that aid in prediction of the influence of enzyme mutations on enzymatic activity. Finally, catalytic fields methodology was employed in the development of qualitative characteristics of a catalyst optimal for a given reaction.

Technical details The geometries of reactants, transition state and products were determined using the two-layered ONIOM(B3LYP/6-31G(d):PM3) approach [134, 210–212] as implemented in the Gaussian 03 program [175]. The starting structure of reactants was extracted from previous molecular mechanics and semiempirical simulations [213]. The model reaction system was then simplified to include highly conserved residues (see Figure 3.1) suggested to be important for catalysis (i.e., Asp94, Arg121, Glu126, and Cys198) as well as two magnesium ions together with their solvation spheres (water molecules Wat82, Wat112, Wat126, Wat567, and Wat575). The magnesium ion denoted as Mg1 interacts with oxygen atoms from β- and γ-phosphates and is coordinated to Wat112, Wat126, and Wat575. The two remaining water molecules, Wat82 and Wat567, are bound to Mg2 interacting with α- and β-phosphates. Both Asp94 and Glu126 participate in water-mediated coordination of Mg1.

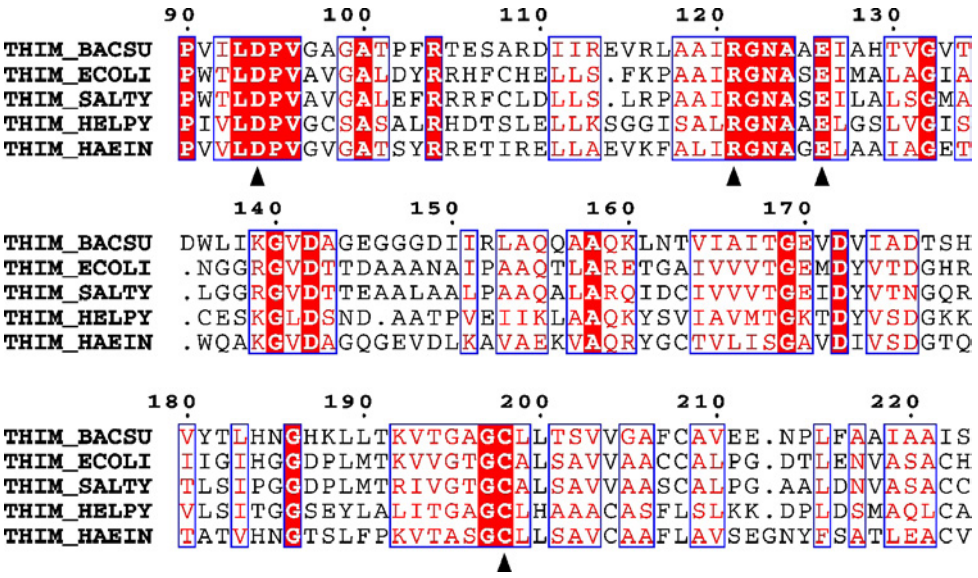


Figure 3.1: Multiple sequence alignment of hydroxyethylthiazole kinases (Thz kinases, ThiK). *Blue boxes* indicate significant similarity (in terms of amino acid physical-chemical properties). *Red background* denotes identity. *Triangles* mark residues serving as a model active site in this study. Thz kinase sequences from the following species are shown: *Bacillus subtilis* (THIM_BACSU), *Escherichia coli* (THIM_ECOLI), *Salmonella typhimurium* (THIM_SALTY), *Helicobacter pylori* (THIM_HELPY),*Haemophilus influenzae* (THIM_HAEIN).

The multiple sequence alignment of Thz kinases from several bacterial species shown in Figure 3.1 indicates the ThiK residues used in this analysis as an active site model. The first sequence in this alignment (denoted as THIM_BACSU according to Swiss-Prot database [214] nomenclature) belongs to *Bacillus subtilis* ThiK, which is the subject of this study. As can be seen from Figure 3.1, the four amino acid residues considered here are among the most highly conserved residues within the Thz kinase family. Sequence alignment was performed with CLUSTAL W [215] and printed using ESPript [216] tool.

To limit the size of a system, arginine and cysteine residues were represented by methylguanidinium cation and methanethiol, whereas an acetate molecule was used to mimic aspartate and glutamate side chains. All broken bonds were saturated with hydrogen atoms. The only neutral residue was then Cys198 (see explanation below), while carboxylic moieties as well as guanidinium group were assumed to be either negatively ($Asp^{-}94$, $Glu^{-}126$) or positively charged ($Arg^{+}121$). The protonation state of aspartate and glutamate was chosen as the most prob-

able due to their involvement in coordination of positively charged magnesium ion [116]. Similarly, the arginine residue has been attributed a role in stabilization of negatively charged phosphates [116] and thus is likely to be positively charged. ATP and Thz were modelled by methyl triphosphate and ethanol, respectively, and only these molecules were included in the region described by higher level of theory. As suggested recently, the triphosphate moiety of ATP in aqueous solution is fully deprotonated under neutral conditions [217]. Due to the influence of the protein environment, there are numerous possible protonation states; however, the charge of -4 is generally accepted as the most likely, especially when the triphosphate tail is coordinated to two magnesium ions [33, 218]. Therefore, the methyl triphosphate considered herein was assumed to bear a charge of -4. The other substrate—ethanol—is unlikely to be deprotonated in the initial stages of reaction [116] and it was thus modelled as a neutral species.

Since a reliable verification of transition state relevance would not be possible in the case of a model containing fixed entities, all atoms in the system were allowed to move during optimization and the final complexes turned out to be stable. The change in substrate structure associated with the optimization procedure was measured by the root mean square deviation (RMSD) of its heavy atoms calculated with respect to the starting complex. Relatively high RMSD values (i.e., 2.3 Å) could partially be explained by a possible errors in the reference structure itself (due to its molecular mechanics origin [213]). RMSD corresponding to the displacement of heavy atoms from the reacting molecules (including magnesium ions) is significantly lower (i.e., 0.9 Å) indicating greater stability of the model core as opposed to increased mobility of the surrounding amino acid residues. Comparison with any experimental structure could not be performed due to the lack of data regarding ThiK-substrate complexes; crystal structures of Thz kinase are available only for Thz alone or ATP/Thz-phosphate [116] (Protein Data Bank access codes of 1C3Q and 1ESQ, respectively). Moreover, the structure containing ATP and phosphorylated Thz corresponds to the Cys198Ser mutant. Nonetheless, the stability of final quantum mechanically derived structures can be used to support their reliability. The nature of particular stationary points found along the reaction coordinate as well as thermodynamic properties (zero point energy correction, enthalpies, and Gibbs free energies) were determined by vibrational frequency analysis.

Based on the structures of reactant and transition state complexes with ThiK active site residues (magnesium ions were also considered as a part of active site environment), the energies of interaction between those residues and the reacting molecules (i.e., ATP and Thz) were then evaluated in a pairwise manner and further partitioned according to the variation-perturbation decomposition scheme [26]. Interaction energy values presented in what follows were calculated

with 6-31G(d) basis set using the modified version of Gamess–US program [170]. Dimer-centered basis set was employed to account for BSSE by means of counterpoise correction scheme [140]. The HF/6-31G(d) electron density and electrostatic potentials required for catalytic fields were obtained with the Gaussian 03 program [175].

To assess the importance of ThiK residues not included in an active site model considered herein, the reactant and transition state structures were inserted into the starting molecular mechanics model of ThiK-substrates complex. Any possible bad contacts were removed by short minimization performed with Charmm program [173] and all-hydrogen parameter set from Charmm27 forcefield [174]. Optimization was followed by evaluation of pairwise interaction energy between each residue located within 10 Å cutoff around reactant/transition state and the latter. Binding energy was expressed as a sum of electrostatic and van der Waals terms, as defined within Charmm27 force field. Since no additional parametrization was performed for transition state structure, it was described by the same set of atomic point charges as reactants. Apart from residues already included in quantum mechanical model of active site, three amino acid residues (i.e., Arg104, Asn123, and Gly197) were found to contribute more than 1 $\mathrm{kcal \cdot mol^{-1}}$ to differential transition state stabilization. Due to the average distance between Arg104 and reactant/transition state being equal to 6 Å as well as rather insignificant contribution to DTSS ($+1.3\ \mathrm{kcal \cdot mol^{-1}}$), this residue is probably of minor importance for the study of ThiK-catalyzed reaction. Similar characteristics in terms of preferential transition state binding is exhibited by Gly197 ($+1.4\ \mathrm{kcal \cdot mol^{-1}}$), however, the close distance (i.e., 1.8 Å) between its main chain amino group and γ-phosphate oxygen might indicate that more thorough study should probably account for the presence of this residue. Finally, much higher contribution to DTSS ($+8.00\ \mathrm{kcal \cdot mol^{-1}}$) is provided by Asn123 interacting via its side chain amino group with β-phosphate oxygen atom (minimum distance of 2.5 Å in the case of ThiK-reactants complex). The magnitude of differential transition state destabilization arising from Asn123 residue results from the significant rearrangement of its side chain when going from reactants to transition state complex. The corresponding minimum distance is decreased from 2.5 Å to 2.0 Å. It is then unclear whether this outcome should be considered as a conclusive result, since it might originate from the lack of transition state structure reparametrization. Due to uncertain role of Asp123, further calculations should probably take this residue into account. Finally, it should be clearly stated that Charmm calculation was performed only to estimate the possible contributions to DTSS arising from ThiK residues neglected while building the ONIOM active site model, whereas the latter was utilized for the nonempirical analysis of interaction energy. As it has been shown in a parallel study of urokinase inhibition [198], force field optimization tends to underestimate

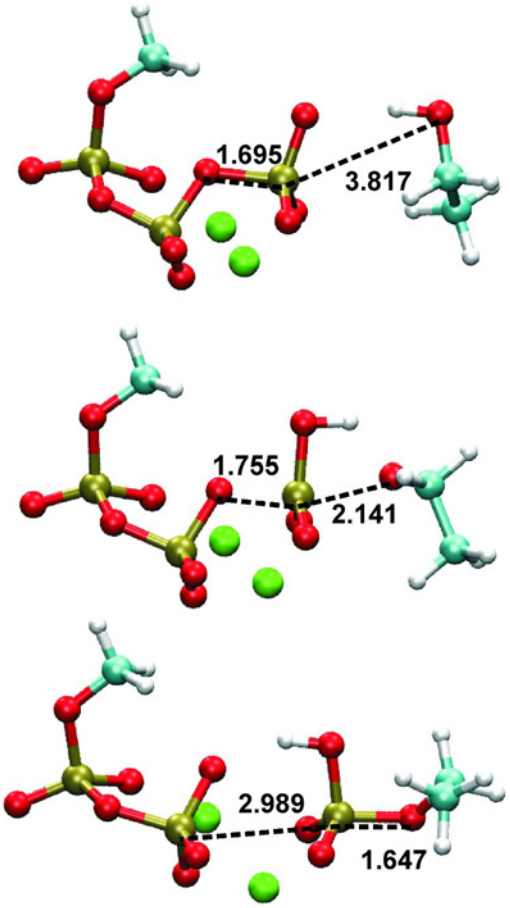


Figure 3.2: B3LYP/6-31G(d):PM3 optimized geometries of reactants *(top)*, transition state *(middle part of drawing)* and products *(bottom)*. Only the reacting molecules (i.e., methyl triphosphate and ethanol representing ATP and Thz) and the two magnesium ions are shown. The crucial distances are denoted as *black dashed lines*.

the distances between interacting species and, thus, binding energy decomposition carried out herein was based on the quantum mechanically-derived structures.

Phosphoryl transfer reaction The structures of the reactants, transition state and products for ThiK-catalyzed process are presented in Figure 3.2. In this study, only the most probable reaction course, which consists of an S_N2-like nuclephilic substitution at the phosphorus atom [116], was modelled. When considering the cleavage of a phosphoester bond, the range of mechanisms, that can

Table 3.1: Relative energies [$\mathrm{kcal \cdot mol^{-1}}$] of the transition state and product with respect to reactant structures as derived from B3LYP/6-31G(d):PM3 calculations. ΔE corresponds to the sum of electronic and zero-point energies, while ΔH and ΔG stand for the enthalpy and Gibbs free energy, respectively.

	ΔE	ΔH	ΔG
Reactant	0.0	0.0	0.0
Transition state	40.2	38.7	44.4
Product	9.2	8.1	10.6

occur, is bound by two limiting cases, i.e., the mechanism is either fully dissociative or fully associative [219]. The dissociative pathway, analogous to the S_N1 reaction in organic chemistry, proceeds via a highly reactive trigonal metaphosphate intermediate. The corresponding distances between the oxygen atoms from a leaving/entering group and the phosphorus atom is greater than the respective sum of van der Waals radii (i.e., 3.3 Å). The fully associative mechanism involves a pentavalent bipyramidal phosphorane intermediate with the length of the axial phosphorus-oxygen bonds (i.e., 1.73 Å) indicating their covalent character [219]. Between these two extremes are located intermediate cases described by partially associative S_N2-like mechanism and the occurrence of a transition state with axial bonds length between 1.73 and 3.3 Å.

As mentioned above, kinase-catalyzed phosphorylation reactions often imply the involvement of an aspartate residue acting as an alcohol activating base. Its role consists of deprotonation of an attacking nucleophile prior to the actual formation of a phosphorus-oxygen bond. In the case of Thz kinase, only the cysteine residue (Cys198) is present in the vicinity of an attacking hydroxyl group. Moreover, site-directed mutagenesis studies on ThiK have suggested that the cysteine residue is not likely to be deprotonated during the phosphoryl transfer and, thus, it is not directly involved in catalysis (cysteine to alanine or serine mutants are only slightly less active) [116]. Presumably, the γ-phosphate oxygen atoms could also act as a catalytic base increasing nucleophilicity of a substrate hydroxyl group and aiding in proton transfer during the reaction. The results of our simulation seem to confirm this suggestion—one of the γ-phosphate oxygen atoms is hydrogen bonded to the Thz hydroxyl moiety and its distance to the hydroxyl proton is equal to 2.05 Å.

Since no intermediates were found along the considered reaction coordinate, and a single transition state encompasses trigonal bipyramidal geometry of a transferred phosphate, the overall phosphorylation pathway is consistent with an S_N2-like mechanism. According to Pauling formula as proposed by Mildvan [219], the fractional bond number n providing a quantitative measure of associativity can be calculated based on the length of an axial bond to entering group in a model considered, $D(n)$, and the single phosphorus-oxygen bond distance, $D(1)$:

$$D(n) = D(1) - 0.60 \log(n) \tag{3.1}$$

Since $D(n)$ and $D(1)$ are equal to 2.14 and 1.73 Å, respectively, the discussed mechanism is 80% dissociative. A minor increase in the distance to the leaving group oxygen (1.76 compared to 1.70 Å) suggests the late departure of a phosphate donor. Noticeably, the expected proton transfer to the γ-phosphate oxygen atom occurs prior to transition state formation. However, the rather large activation energy barrier of $\sim$ 40 kcal $\cdot$ mol^{-1} (see Table 3.1) indicates that either a more realistic catalytic environment should be considered (i.e., the ThiK active site model consisting of a greater number of residues) or an alternative reaction pathway associated with a lower energy barrier may exist.

Differential transition state stabilization The next step in this study consisted of assessing the involvement in catalysis of each active site residue present in the model system. The putative catalytic role can be analyzed within the DTSS methodology allowing the most favourable (i.e., those increasing the reaction rate) residues to be recognized. In general, the greater the strength of interaction of a given residue with the transition state with respect to the substrates, the lower the activation energy barrier, resulting in a higher rate enhancement. Thus, such a residue can be considered as catalytically active, since its presence promotes catalysis.

Table 3.2 gives the results of interaction energy analysis expressed in terms of the DTSS energy. The ordering of residues reflects their decreasing contribution to differential transition state stabilization. These results are also presented in Figures 3.3 and 3.4. The most catalytically effective residues (i.e., those that bind the transition state stronger than the reactants) are Glu126, Cys198, and Mg2. The presence of Mg1 appears to have a moderately destabilizing effect on the transition state. This result is rather unexpected, since Mg1 is generally believed to increase the rate of catalyzed phosphoryl transfer, whereas an inhibitory influence has been attributed to Mg2 (as reported in the case of cAMP-dependent protein kinase (PKA) [220]). However, similar results (i.e., greater contribution of

Table 3.2: Differential transition state stabilization energy [$\mathrm{kcal} \cdot \mathrm{mol}^{-1}$] at various levels of theory.

Residue	$\Delta_{EL}^{(1)}$	$\Delta^{(1)}$	Δ_{SCF}	Δ_{MP2}
Glu126	−5.95	−5.94	−5.88	−5.44
Cys198	−5.13	1.39	−3.95	−5.30
Mg2	−3.30	−5.18	−4.53	−3.90
Wat112	−3.61	−2.25	−2.97	−3.02
Wat567	−0.35	−0.91	−0.85	−0.81
Wat82	0.21	−0.37	−0.28	−0.20
Wat575	−4.29	1.58	0.76	−0.12
Asp94	−0.16	−0.14	−0.30	0.19
Mg1	7.34	5.34	4.49	3.26
Wat126	7.19	3.86	4.32	4.12
Arg121	7.89	10.67	12.86	12.52
sum	−0.16	8.04	3.67	1.30
$r^{2\,a}$	0.77	0.80	0.98	1.00

[a] Coefficient of determination referring to the relationship between results at a particular level of theory and the MP2 results.

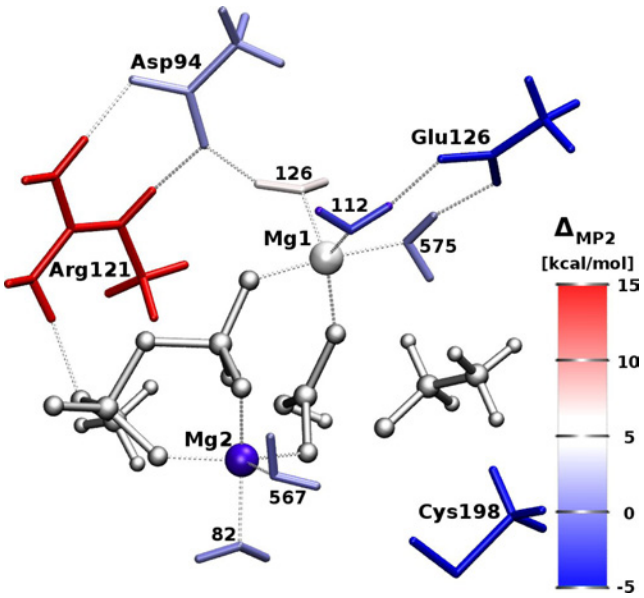


Figure 3.3: Structure of the ThiK active site model in complex with transition state (*ball-and-stick representation*). Active site residues *(stick representation)* are coloured according to their contribution to differential transition state stabilization (see energy scale on *the right hand side*).

Mg2 to transition state stabilization) were obtained in an independent QM/MM simulation of PKA-catalyzed phosphoryl transfer [33]. Surprisingly, Arg121 seems to disfavor catalysis by stabilizing the substrates instead of the transition state and, due to its significant influence, the overall differential effect is slightly positive (total $\Delta_{MP2} = 1.30\ \text{kcal} \cdot \text{mol}^{-1}$), indicating the actual lack of preferential transition state stabilization. This observation could suggest that it might be necessary to include the more complete active site representation in further calculation.

Considering the accuracy of consecutive levels of theory applied to the description of interactions, analysis of correlation with the most precise results at hte MP2 level demonstrates very good agreement between corresponding interaction energy approximations (Table 3.2). In particular, electron correlation effects seem to be of minor importance in the case analyzed (the coefficients of determination of the SCF results with respect to those at the MP2 level is 0.98). Moreover, neglecting delocalization and exchange effect is also justified—the corresponding r^2 values are 0.80 and 0.77, respectively. As can be seen in Figure 3.4, subsequent corrections to binding energy tend to cancel each other to a remarkable degree. Nonetheless, systematic improvement is obtained by application of higher level of theory in description of interactions.

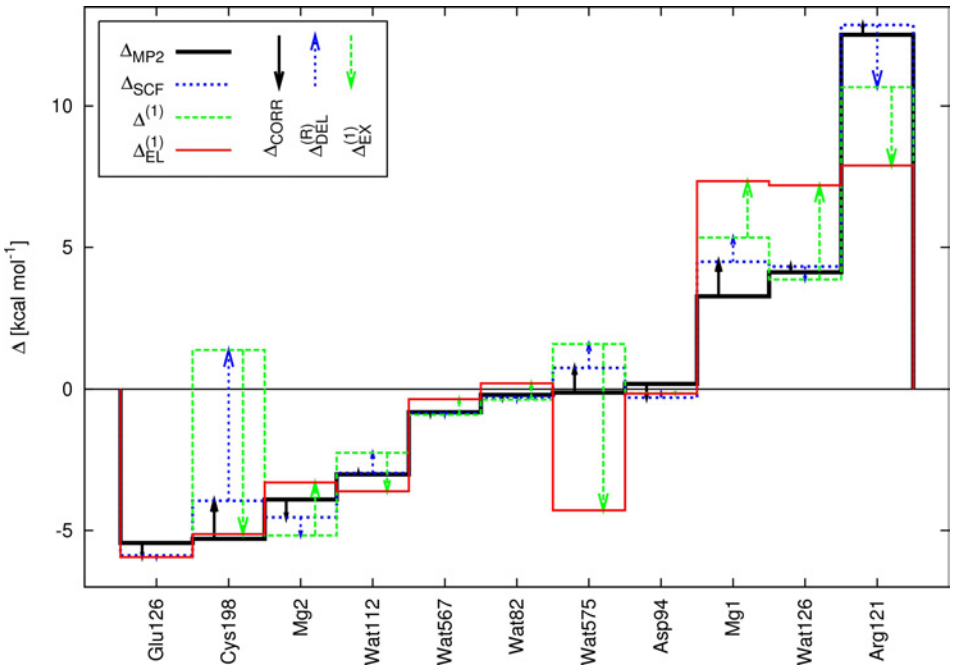


Figure 3.4: The components of differential transition state stabilization energy [kcal·mol^{-1}]. For each residue, the transition state stabilization energy (relative to substrates) is given at subsequent levels of theory. *Vertical arrows* show the mutual cancellation of correlation, delocalization, and exchange corrections to the MP2 interaction energy.

Optimal catalytic environment Catalytic fields represent a hypothetical optimal molecular environment with electrostatic properties adjusted to enhance the rate of a given reaction (i.e., acting as a catalyst; see section 1.3.2). Figure 3.5 illustrates the catalytic field for the phosphoryl transfer reaction catalyzed by ThiK. The corresponding arrangement of ThiK active site residues is also presented. Examination of differential electrostatic potential reveals a pronounced change in charge distribution when going from reactants to the transition state. This indicates a legitimate need for this type of analysis to be performed as a part of the catalyst design process. Moreover, the evident agreement between electrostatic interaction energy and the reference MP2 results ($r^2 = 0.77$, see Table 3.2) further justifies the approximations applied in catalytic field derivation.

As anticipated by the catalytic field, the presence of a negative charge in the red-coloured region would favor catalysis. One of these areas is already occupied by the negatively charged Asp94, while the other is located close to the neutral Cys198 residue. However, the experimental substitution of cysteine for the neg-

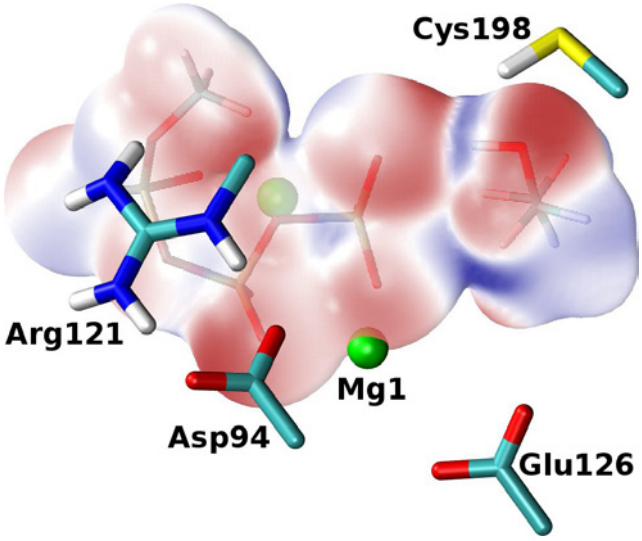


Figure 3.5: Catalytic fields for the ThiK-catalyzed reaction superimposed with the positions of active site residues in a model system. The electronic isodensity surface of 0.01 a.u. is coloured according to the differential electrostatic potential of transition state and reactants calculated at the HF/6-31G(d) level of theory. The sign of the differential potential is inverted to reflect the electrostatic properties of a complementary molecular environment. *Red (blue)* colour denotes regions where a negative (positive) charge would be optimal for catalytic activity.

atively charged aspartate residue results in increased catalytic activity [116], in a perfect agreement with the catalytic field predictions.

3.1.2 cAMP-Dependent protein kinase

Due to their participation in protein phosphorylation and the resulting involvement in cell signaling as well as metabolic regulation, protein kinases have received much attention from the scientific community [203]. Among the most extensively studied is cAMP-dependent protein kinase (PKA) with its catalytic subunit structure and mechanism serving as a prototype for the entire family of eukaryotic protein kinases [221]. PKA-catalyzed process encompasses the transfer of γ-phosphate from ATP to the specific serine residue located on the substrate peptide. Despite numerous experimental [222] and theoretical [33, 218, 223–231] studies, no agreement has been reached regarding all details of PKA catalytic machinery. In particular, the involvement of several highly conserved active site residues (e.g.,

Asp166, Lys168) along with the functional relevance of magnesium ions has remained unclear.

While divalent metal ions are generally thought of as facilitating the phosphorylation processes [217] the mechanism of their participation appears to be complex and, possibly, varying between different metal sites. X-ray crystallographic data display the presence of two magnesium ions within PKA active site. It has been known, that at least one magnesium ion is essential for PKA action. This metal ion occupies a high-affinity metal binding site and chelates the β- and γ-phosphates of ATP as well as Asp184 residues. At high magnesium concentration also the low-affinity site is filled with the second $Mg2^+$ bridging α- and γ-phosphates and Asp184, Asn171 residues. As shown experimentally, PKA activity decreases with an increasing occupation of the secondary metal site [232]. Accordingly, this particular magnesium ion is termed inhibitory. However, at the limited ATP concentration, the reaction rate is actually accelerated owing to increased ATP binding affinity [203].

Another unresolved issue is the PKA phosphorylation mechanism itself, i.e., the presence and identity of a residue serving as a catalytic base. While many studies suggest that no general-base catalyst is required for the phosphoryl transfer process [224, 225, 228, 229], others contradict this hypothesis proposing the involvement of Asp166 residue in deprotonation of a hydroxyl group of serine residue prior to its phosphorylation [218, 230, 233, 234]. Moreover, direct Asp166 participation in the general-base catalysis has been supported by the latest high-level computational results [33, 223, 226, 231], while the earlier proposals neglecting Asp166 involvement have been put forward based on semiempirical QM (or QM/MM) models [224, 225, 228]. Whether this particular residue contributes any more effects to PKA catalytic rate enhancement has yet to be determined.

Since the function of the remaining conserved residues has not been unequivocally ascertained, our work [34] aimed at revealing the actual role of PKA active site components in the observed catalytic activity. In particular, the activation barrier lowering resulting from the presence of a given residue was investigated based on the PKA mechanism and structures demonstrated in Ref. [33]. As suggested by classical calculations of the interactions occurring in enzyme-substrate and enzyme-transition state complexes, the PKA active site provides stabilization of the transition state [33]. In what follows, nonempirical decomposition of interaction energy was applied providing a comprehensive evaluation of the physical nature of catalysis along with the possible involvement of individual active site residues. The overall picture of an enzyme catalytic mechanism could aid prediction of the impact of enzyme mutations on the enzymatic activity.

Technical details The structures of PKA-bound substrate and transition state were obtained from previous density functional theory QM/MM simulation [33]. The model complexes used for interaction energy calculation were then simplified to include PKA residues in a close vicinity of the reaction site (Figure 3.6). As the adenosine part of ATP does not change its position and geometry throughout the phosphoryl transfer, ATP molecule was represented by the methyl triphosphate only. Similarly, the substrate peptide was limited to serine residue. Whenever justified by sufficient distance to the substrate/transition state complex, the backbone of selected residues was also removed leaving the respective side chains capped with Cα atoms. Such a treatment was applied to Val44, Lys^{+}72, Gln84, Glu^{-}91, Lys168, Asn171, and Phe187 residues. All broken bonds were saturated with hydrogen atoms optimized at the HF/6-31G(d) level of theory. The positively charged lysine and negatively charged aspartate and glutamate residues were considered. The model included also several conserved water molecules [235] (Wat410, Wat412, Wat447, Wat459, Wat476, Wat477, Wat597, Wat635) as observed in 1ATP PDB structure, and two water molecules from QM/MM calculation [33] (WatSOL1 and WatSOL2) without any crystallographic counterparts. Magnesium ion occupying the high-affinity Mg1 site is surrounded by β- and γ-phosphates of ATP, two oxygen atoms from Asp184 carboxylate group, and two water molecules (Wat447, Wat477). Magnesium ion from the secondary metal site, Mg2, is coordinated by the α- and γ-phosphates, the bridging oxygen located between β- and γ-phosphates, one oxygen atom of Asp184, the side chain carbonyl oxygen of Asn171, and water molecule Wat635.

The HF/6-31G(d) electron density and electrostatic potentials required for catalytic fields were obtained with the Gaussian 03 program [175]. Calculation of binding energy and its decomposition according to the variation-perturbation procedure [26] was performed in collaboration with Dr. Paweł Szarek [34].

Differential transition state stabilization The catalytic contribution of a given residue can be expressed as its ability to lower the activation energy barrier by stronger interaction with transition state in comparison to reactants, i.e., differential transition state stabilization. Accordingly, the presence of such a catalytically active residue results in the reaction rate enhancement and promotes catalysis. Considering partitioning of the system into transition state/substrates and a remaining environment, magnesium ions were arbitrarily assigned to the latter. Table 3.3 and Figure 3.7 provide the catalytic activity of individual active site components in terms of their DTSS energy values calculated at different levels of theory (as provided by Dr. Paweł Szarek [34]). The residues are arranged according to decreasing contribution to activation barrier lowering as indicated by the increasing values of the DTSS energy. Noticeably, the entire PKA active site

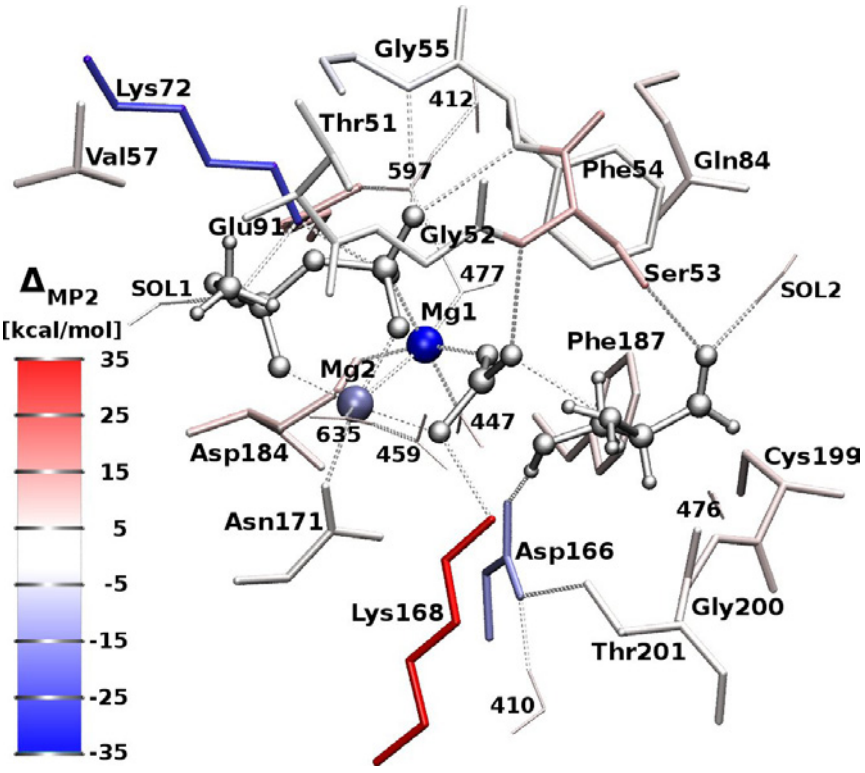


Figure 3.6: Structure of the PKA active site model in complex with transition state (*ball-and-stick representation*). Active site residues *(stick representation)* are coloured according to their contribution to differential transition state stabilization (see energy scale on *the left hand side*). Except for the transition state and water molecules, only heavy atoms are shown.

Table 3.3: Differential transition state stabilization energy [$\mathrm{kcal} \cdot \mathrm{mol}^{-1}$] [34].

Residue	$\Delta_{EL}^{(1)}$	$\Delta^{(1)}$	Δ_{SCF}	Δ_{MP2}
Mg2	−28.57	−29.42	−37.21	−32.36
Lys72	−21.83	−16.38	−22.50	−22.71
Mg1	−13.88	−13.12	−16.70	−15.15
Asp166	−19.80	−11.01	−16.02	−13.26
Gly55	−3.64	−1.62	−4.16	−4.09
Thr51	−2.14	−1.82	−2.72	−2.42
Gly52	−1.05	−0.19	−2.66	−2.21
WatSOL1	−2.42	−0.39	−1.95	−1.81
Phe54	−0.91	1.80	−2.10	−1.68
Asn171	−0.69	0.92	−0.58	−1.35
Thr201	−1.82	−2.20	−1.22	−1.13
Wat477	−1.34	−0.36	−1.03	−1.09
Val57	−0.08	0.23	−0.53	−0.51
Gln84	−0.02	−0.02	−0.04	−0.02
Wat412	0.17	0.17	0.14	0.11
Gly200	−0.32	−0.62	−0.24	0.16
Wat410	0.25	0.25	0.29	0.27
Wat476	0.46	0.47	0.53	0.49
Wat635	2.63	0.93	0.87	0.63
Cys199	0.69	0.68	0.87	0.70
WatSOL2	1.11	0.40	0.92	0.73
Wat597	1.03	1.15	0.87	0.78
Wat459	1.09	0.99	1.05	0.81
Phe187	0.61	0.30	1.08	0.97
Wat447	1.54	2.83	2.66	2.21
Asp184	7.04	6.98	4.24	3.06
Ser53	5.17	2.56	4.57	4.72
Glu91	5.90	5.90	5.48	4.96
Lys168	28.21	21.03	27.74	27.83
sum	−42.60	−29.57	−58.34	−51.39
$r^{2\,a}$	0.97	0.97	0.99	1.00

[a]Coefficient of determination referring to the relationship between results at a particular level of theory and the MP2 results.

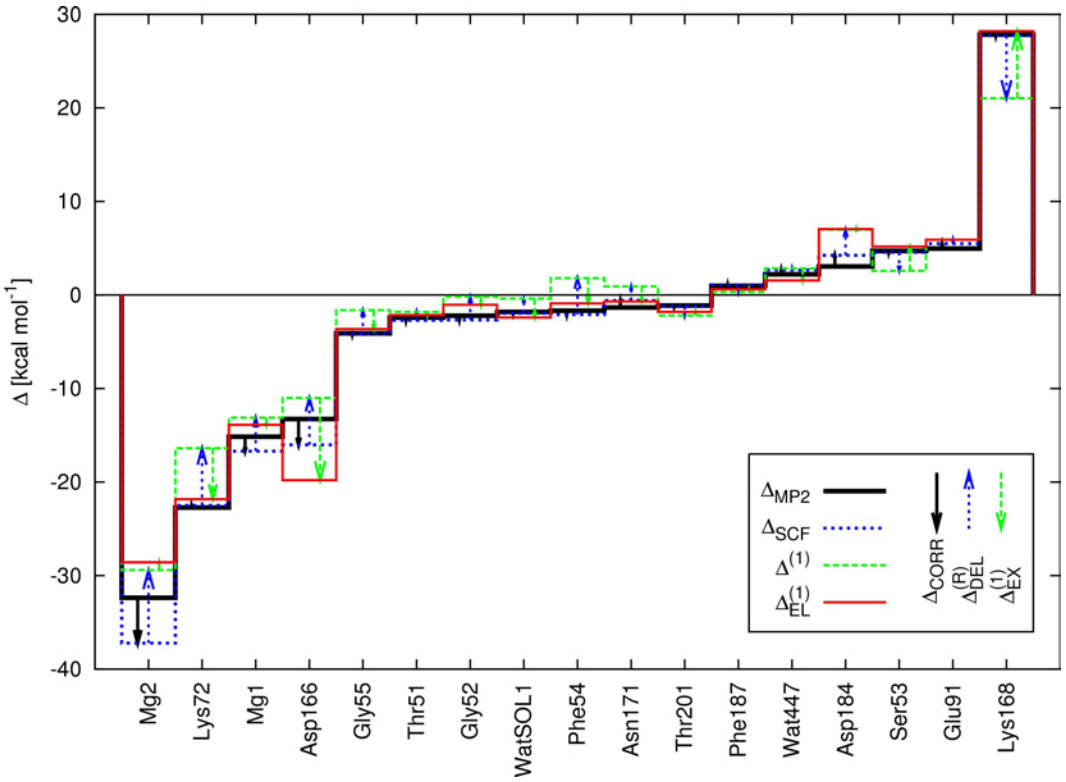


Figure 3.7: The components of differential transition state stabilization energy [kcal $\cdot$ mol^{-1}] [34]. For each residue, the transition state stabilization energy (relative to substrates) is given at subsequent levels of theory. *Vertical arrows* show the mutual cancellation of correlation, delocalization, and exchange corrections to the MP2 interaction energy.

provides as much as -51.39 kcal $\cdot$ mol^{-1} of total differential transition state stabilization (Table 3.3).

Apart from magnesium ions, the most pronounced catalytic effects seem to be exhibited by absolutely conserved Lys72 and Asp166 residues ($\Delta_{MP2} = -22.71$ and -13.26 kcal $\cdot$ mol^{-1}, respectively; see Table 3.3). Lys72 interacting with α- and β-phosphates (Figure 3.6) has been postulated to facilitate the phosphorylation process without affecting the binding of ATP [203]. Owing to 844-fold decrease in k_{cat} and essentially retained ATP binding properties, alanine-substituted mutants seem to be deficient in ATP transition state binding [236]. In excellent agreement with these experimental results, Lys72 appears to participate in a catalytic step by outstanding preferential stabilization of the transition state (Figure 3.7). Another strictly conserved active site residue, Glu91, is located behind Lys72 (relative

to ATP-serine complex). Instead of being involved in direct contact with any of the substrates, it forms a conserved salt bridge with Lys72 residue (Figure 3.6). According to DTSS results, the presence of Glu91 results in a moderate transition state destabilization characterized by $\Delta_{MP2} = 4.96$ kcal $\cdot$ mol^{-1} (Table 3.3). Presumably, the function of Glu91 consists in a proper positioning of the flexible Lys72 side chain to maximize its favorable interaction with ATP phosphate tail. Relatively large catalytic advantage resulting from the presence of Lys72 cancels destabilizing influence of Glu91 residue.

As confirmed by the results presented in Ref. [33], Asp166 residue serves as a catalytic base that accepts substrate peptide proton during the phosphorylation process. However, the proton transfer step occurs after the transition state which still encompasses deprotonated Asp166 carboxylate and neutral hydroxyl moiety of substrate serine. Therefore, to reveal its possible role in the DTSS effects, this particular residue was treated as a part of a catalytic environment. The significant negative value of Asp166 DTSS energy ($\Delta_{MP2} = -13.26$ kcal $\cdot$ mol^{-1}, Table 3.3) indicates that its additional function might indeed consist in preferential stabilization of the transition state.

Both magnesium ions are among the most catalytically effective components of PKA active site (Figure 3.7). Their contribution to DTSS effects is equal to -32.36 (Mg2) and -15.15 kcal $\cdot$ mol^{-1} (Mg1; Table 3.3). Interestingly, magnesium ion occupying the so-called inhibitory site, i.e., Mg2, is capable of greater transition state stabilization than Mg1 located at a high-affinity metal binding site. Nonetheless, these results are consistent with preliminary interaction energy analysis from Ref. [33] as well as the results obtained for Thz kinase [209]. In contrast to magnesium ions themselves, their ligands appear to contribute little if any to overall transition state stabilization. The most pronounced inhibitory effects result from the presence of Asp184 and Wat447 belonging to Mg1 coordination sphere. These particular residues destabilize the transition state by 3.06 and 2.21 kcal $\cdot$ mol^{-1} (Table 3.3). Noteworthily, mutagenesis results show that Asp184 residue is essential for the catalytic function of PKA as yeast cells with Asp184Ala mutation are inviable [236]. Presumably, Asp184 role encompasses the chelation of Mg1 and the lack of the latter is the main reason for impaired PKA action. Considering the overall impact of individual magnesium-ligands complexes, the sum of DTSS energy values associated with a given magnesium ion and its coordination sphere (i.e., Asp184, Wat447, and Wat477 or Asn171, and Wat635 for Mg1 and Mg2, respectively) indicates similar influence in terms of transition state stabilization to that exhibited by magnesium ions treated separately. In particular, the DTSS energy of Mg1 and Mg2 ions in complexes with their ligands is equal to -10.98 and -33.09 kcal $\cdot$ mol^{-1}, respectively. It seems that catalytic benefits resulting

from the presence of magnesium ions make it worthwhile to employ residues that anchor a magnesium ion even if they exhibit moderate inhibitory effects.

The glycine-rich loop encompassing residues 47–57 constitutes a highly conserved motif within a kinase catalytic core. Owing to its extended "U" shape spatially aligned with ATP triphosphate tail, the glycine-rich loop is capable of tightly enfolding the nucleotide by means of both hydrogen bonding and hydrophobic interactions. Accordingly, its primary function appears to involve the positioning of γ-phosphate of ATP for the subsequent phosphoryl transfer step [222]. Since mutagenesis studies have revealed no major changes in PKA catalytic competency upon the mutation within this particular region [237, 238], it is not clear whether PKA benefits any transition state stabilization due to the presence of glycine-rich loop. Out of the six glycine-rich loop residues present in our model (see Figure 3.6), only Ser53 exhibits a certain degree of transition state destabilization ($\Delta_{MP2} = 4.72$ kcal $\cdot$ mol^{-1}). This particular residue interacts via its backbone amide hydrogen with γ-phosphate. As shown experimentally [237], it is the presence of a backbone interaction that matters most: side-chain modifications do not significantly alter the catalytic features of PKA. Presumably, the role of Ser53 is essentially structural as it involves the optimal positioning of a terminal phosphate group. The catalytic activity of the remaining glycine-rich loop residues is moderate with the most prominent transition state stabilization coming from Gly55 residue ($\Delta_{MP2} = -4.09$ kcal$\cdot$mol^{-1}). It is noteworthy, however, that the collective influence of Thr51, Gly52, Ser53, Phe54, Gly55, and Val57 residues encompasses Δ_{MP2} of -6.20 kcal$\cdot$mol^{-1}, indicating relatively large contribution to the lowering of activation energy barrier.

Lys168 residue, being an important component of a catalytic loop, has been suggested to support both the phosphoryl transfer process and peptide binding [203]. Apparently, this residue is not directly involved in the phosphorylation reaction [33], yet the presence of hydrogen bonding interactions between its side chain and both the γ-phosphate as well as the hydroxyl group from substrate serine implies its possible contribution to the catalytic process. Surprisingly enough, the most obvious role consisting in transition state stabilization provided by Lys168 side chain charged oppositely to the reactants is not an issue here, as Lys168 has been found to strongly destabilize the transition state (Table 3.3 and Figure 3.7). This result is consistent with interaction energy analysis presented in Ref. [33] and further supports the conclusion stated therein (based on the results of molecular dynamics simulation) that the primary role of Lys168 is to keep the reactants in a geometry of the catalytically competent complex.

The remaining components of PKA active site do not seem to provide an appreciable degree of transition state stabilization or destabilization. Except for water

molecule designated as WatSOL1 and Thr201 residue, the absolute contribution of the other residues to the lowering of activation energy barrier does not exceed 1 kcal $\cdot$ mol^{-1}. WatSOL1 molecule located in the proximity of Lys72 is hydrogen bonded to α-phosphate. Similarly to the Lys72 residue, WatSOL1 seems to promote catalysis providing $\Delta_{MP2} = -1.81$ kcal$\cdot$mol^{-1} of transition state stabilization. Since no counterpart of WatSOL1 has been found in PKA X-ray structure, this result should probably be considered as a suggestion that the presence of a water molecule in this particular position would be favorable for catalysis. Thr201 anchors both Asp166 and Lys168 via hydrogen bonds to their side chains. While it has been suggested that such a location of Thr201 residue might imply its involvement in a proton shuttle between Asp166 and Lys168 [222], it has not yet been confirmed. According to DTSS analysis, Thr201 role encompasses additional activation barrier lowering of -1.13 kcal $\cdot$ mol^{-1}.

In addition to total DTSS contributions to PKA catalytic activity, this analysis aimed at determining the physical nature of active site interactions. Therefore, binding energy was partitioned into the components with clear physical sense as defined within the variation-perturbation decomposition scheme [26]. Noticeably, such a description of the interactions provides also a convenient way to assess the most suitable level of theory. The overall quality of consecutive levels of theory applied in analysis of the interactions can be judged by their comparison to the most accurate description provided by the MP2 calculations. The results of such a comparison encompassing the respective coefficients of determination are presented in Table 3.3. Only minor decrease in correlation accompanies neglect of subsequent interaction energy corrections. In particular, electron correlation effects appear to be insignificant as the r^2 value of the SCF results is still equal to 0.99. Similarly, r^2 values of 0.97 and 0.97 indicate minor contribution of delocalization and exchange components. As it can be seen from Figure 3.7, these two terms tend to cancel each other, resulting in a relatively correct description provided by the electrostatic energy. Apparently, electrostatic effects play a dominant role in the system under study supporting the general hypothesis about an electrostatic origin of the enzymatic catalysis [19, 98].

Optimal catalytic environment Catalytic fields for PKA-catalyzed phosphoryl transfer reaction are presented in Figure 3.8. In excellent agreement with the predicted optimal arrangement of charged moieties are the Asp166 and Lys72 residues as well as both magnesium ions. Similar to the DTSS results, Lys168 seems to disfavor catalysis as it is present in a region where the negative charge is expected to be optimal. Nevertheless, Lys168 (or its equivalent arginine) is a highly conserved residue, indicating that catalytic machinery of kinases employs its flexible, positively charged side chain. The possible structural role of Lys168

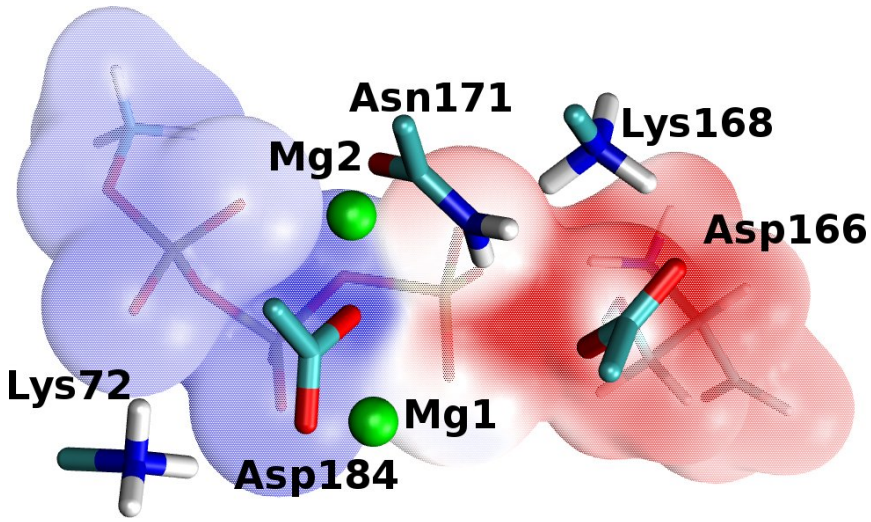


Figure 3.8: Catalytic fields for the PKA-catalyzed reaction superimposed with the positions of active site residues in a model system. The electronic isodensity surface of 0.01 a.u. is coloured according to the differential electrostatic potential of transition state and reactants calculated at the HF/6-31G(d) level of theory. The sign of the differential potential is inverted to reflect the electrostatic properties of a complementary molecular environment. *Red (blue)* colour denotes regions where a negative (positive) charge would be optimal for catalytic activity.

has already been discussed [33]. Another inconsistency between catalytic field predictions and the electrostatic characteristics of PKA active site appears to involve Asp184 residue. Since this particular residue anchors the mandatory magnesium ion, it does not probably affect the catalytic activity as the negative charge of its side chain is screened by the magnesium ion.

3.2 Phosphotriesterase-catalyzed Hydrolysis Reaction

High toxicity of organophosphorus compounds, directed not only toward crop protection but also rising threat of being used as chemical warfare agents [239], resulted in an urgent need to develop safe and efficient decontamination methods. Consequently, there is a growing interest in the fundamental mechanisms of organophosphate hydrolysis as reflected by a number of contributions devoted to the computational investigation of phosphotriesters degradation [240–242].

The non-harmful transformation of such hazardous chemicals could also be attained by enzymatic biodegradation [243, 244] and the factors determining selection of leading enzyme candidates include both the efficiency of biocatalysis and the ability to react with a broad spectrum of substrates. From this perspective, bacterial phosphotriesterase, PTE [245], constitutes a prime target, as it has been shown to cleave a variety of phosphorus-ester bonds [246, 247] and to approach the limits of diffusion in the case of its best known substrate [248]. Moreover, much experimental evidence (e.g., directed evolution of mutant enzymes) has shown that PTE substrate specificity as well as enantioselectivity can be modified, leading to engineered enzymes capable of enhanced decontamination of toxic agents [35]. Such approaches, however, require a great amount of time-consuming laboratory work as well as dealing with hazardous materials that one could possibly avoid by application of theoretical methods allowing the relationship between mutations and catalytic activity to be predicted. Noticeably, despite many experimental (as referenced below) and theoretical efforts [249–259], the actual catalytic mechanism of PTE has remained unexplained. Since a detailed understanding of the driving force behind catalysis constitutes a crucial prerequisite in computer-aided catalyst design, the overall objective of this study has been to gain insight into the catalytic properties of PTE that could further be utilized in the rational, knowledge-based control of enzyme catalytic activity.

Phosphotriesterase exhibits wide substrate specificity encompassing the hydrolysis of phosphorus-oxygen (e.g., *O,O*-diethyl *p*-nitrophenyl phosphate (paraoxon), *O,O*-diethyl *p*-nitrophenyl thiophosphate (parathion)), phosphorus-fluorine (e.g., *O,O*-diisopropyl phosphorofluoridate (DFP), *O*-isopropyl methyl phosphonofluoridate (sarin)), phosphorus-sulphur (e.g., *O,O*-diethyl *S*-2-ethylthioethyl phosphorothioate (demeton-S)) and phosphorus-cyanide (e.g., *O*-ethyl *N,N*-dimethyl phosphoramidocyanidate (tabun)) bonds [247, 260] (Table 3.4 and Figure 3.9). Phosphomonoesters as well as phosphodiesters are not cleaved by PTE at a reasonable rate [261]. Interestingly, a natural PTE substrate has not been revealed, whereas limits of diffusion are reached in the hydrolysis of paraoxon, the best substrate identified to date [248]. Taking into account the chemical diversity of PTE substrates, it is essential to determine some common features of their hydrolysis that allow all of them to be accommodated in the PTE active site and subsequently subjected to catalysis.

Experimental data for PTE chemical and kinetic mechanism suggest that phosphotriester hydrolysis occurs upon inversion of configuration at the phosphorus atom, consistent with an in-line displacement mechanism [262] (Figure 3.9). The pH-dependence of the activity suggests an involvement of one ionizable residue with a pK_avalue of 6.1 [263]. On the basis of the available X-ray structures and theoretical studies it was proposed that this ionizable residue is the hydroxyl group

Table 3.4: The kinetic parameters for the hydrolysis of selected PTE substrates.

	DFP	sarin	tabun	paraoxon	parathion	demeton-S
k_2 [a]	1	25.8	7.49	$2.7 \cdot 10^{-2}$	$3.6 \cdot 10^{-4}$	
E		9.1	10.1	13.5	16.6	
k_{cat} [b]	465	56	76.7	3170	630	1.3

[a] Second-order rate constant, k_2 [$M^{-1}s^{-1}$] and activation energy, E [$kcal \cdot mol^{-1}$] refer to the reaction with an aqueous base [267].

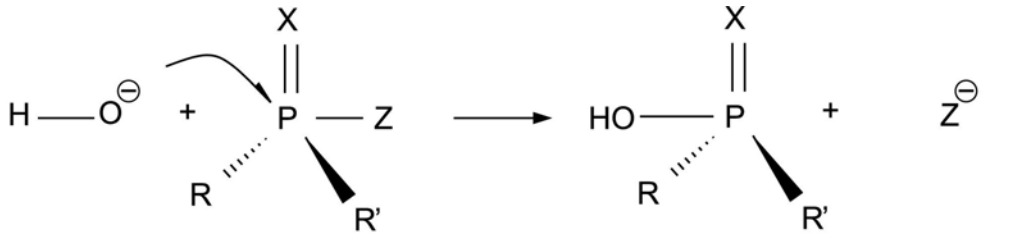
[b] Turnover number, k_{cat} [s^{-1}] applies to the PTE-catalyzed process [247, 260].
Since the particular experiments were not carried out under identical conditions, caution should be used when comparing these numbers.

that bridges the two zinc ions within the PTE active site [249, 264]. Alternatively, recent crystallographic analysis [265] might indicate that the hydroxide in question is another solvent molecule terminally coordinated to the more buried zinc ion. This suggestion has been contradicted by the latest experimental data [259] consistent with the initially proposed mechanism whereby hydrolysis is initiated by the nucleophilic attack of a bridging hydroxide on the phosphorus atom followed by the phosphoester bond cleavage and the expulsion of a leaving group in the anionic form. In addition, the reaction might be facilitated by the zinc ion-substrate interaction resulting in the polarization of a phosphoryl oxygen bond and the increase in phosphorus atom electrophilicity [266]. Comparison of kinetic isotope effects for alkaline and enzymatic hydrolysis of a P-O bond (paraoxon) [122] suggests that the PTE-catalyzed process is slightly less associative than the one in solution. Nevertheless, it can be interpreted as consistent with an S_N2-like concerted associative mechanism in both enzymatic and nonenzymatic phosphotriesters hydrolysis. The latter finding along with the apparent involvement of a hydroxide as a nucleophile, was used to rationalize the following approach assuming that the knowledge of the gas phase mechanism for alkaline hydrolysis would aid further study of the reaction occurring in the PTE active site.

3.2.1 Gas-phase mechanisms of basic hydrolysis of PTE substrates

While the hydrolysis of anionic phosphomonoesters is likely to proceed via a dissociative, unimolecular mechanism that assumes the presence of a metaphosphate

Figure 3.9: Scheme of alkaline hydrolysis of PTE substrates.



DFP:	X=O	Z=F	R=R'=*i*-C_3H_7O	
sarin:	X=O	Z=F	R=*i*-C_3H_7O	R'=CH_3
tabun:	X=O	Z=CN	R=$(CH_3)_2N$	R'=C_2H_5O
paraoxon:	X=S	Z=*p*-$NO_2C_6H_4O$	R=R'=C_2H_5O	
parathion:	X=O	Z=*p*-$NO_2C_6H_4O$	R=R'=C_2H_5O	
demeton-S:	X=O	Z=$SC_2H_4SC_2H_5$	R=R'=C_2H_5O	

intermediate [122], phosphotriester hydrolysis seems to involve some kind of an associative pathway. As to the latter, two limiting scenarios are possible [267, 268], i.e., fully associative mechanism comprising an addition-elimination reaction associated with the presence of a pentacoordinate phosphorane intermediate or a direct-displacement mechanism accompanied by a single S_N2-like transition state. Thus, the addition-elimination pathway is a two-step process in which the attack of a nucleophile results in the formation of an intermediate that further decomposes into products. On the other hand, direct-displacement mechanism involves only one step in which the expulsion of a leaving group occurs at the same time the substituting nucleophile is entering. Current theoretical studies have suggested that depending on the actual trisubstituted phosphoric acid derivative both these pathways are possible. In particular, basic hydrolysis of several phosphofluoridate compounds (i.e., DFP, sarin, soman) was shown to proceed according to an addition-elimination pathway involving trigonal bipyramidal intermediates [240, 242], whereas the hydrolysis of paraoxon is probably a single-step process consistent with a direct-displacement pathway [242]. Since no theoretical analysis has yet been performed for a hydrolysis of a number of PTE substrates and, as it will be argued in what follows, alternative reaction pathways have remained to be explored, these calculations [269] aimed at systematic *ab initio* investigation of the possible gas phase mechanisms of alkaline hydrolysis of a variety of phosphoester bonds in compounds known as PTE substrates (Table 3.4 and Figure 3.9).

Since analysis of the PTE catalytic mechanism presented in subsequent sections is focused on the enzyme-catalyzed hydrolysis of sarin, the gas phase mechanism of the basic hydrolysis of the latter will be presented here in details. For the remaining compounds (Figure 3.9), only the summary of the results will be provided as modeling of the gas phase reaction mechanisms is beyond the scope of this thesis and the relevant details are described elsewhere [269].

Technical details The gas phase reaction profiles were initially studied at the Hartree-Fock (HF) level with the 6-31+G(d) basis set. The nature of stationary points was verified by vibrational analysis. For all the first-order saddle points, intrinsic reaction coordinate (IRC) calculations [270] were performed revealing the geometries of the local minima associated with a given transition state. Since the Hartree-Fock theory neglects the electron correlation effects, it is also not able to accurately predict the barrier heights. Theoretical studies of closely related compounds have shown that results of second-order Møller-Plesset (MP2) calculation of energy with the use of HF-optimized geometries are comparable to the analogous outcome obtained using structures optimized at the MP2 level of theory [242]. Such an approach was also tested herein for a model compound, *O,O*-dimethyl phoshorofluoridate and these calculations [269] confirmed a relative insensitivity of the HF geometries to the electron correlation effects. Thus, the energies reported herein refer to the MP2/6-311++G(d,p) model chemistry applied to HF-optimized structures (denoted as the MP2/6-311++G(d,p)//HF/6-31+G(d) level of theory).

Thermodynamic properties (enthalpies and Gibbs free energies) were determined from vibrational frequencies computed at the fully optimized structures of stationary points along a reaction coordinate. Unless stated otherwise, energy values reported herein include zero-point vibrational energy (ZPE). Considering that the mechanism of an enzyme-catalyzed reaction resembles the analogous gas phase process rather than the one taking place in a solution [271–274], research presented herein has focused primarily on the study of possible gas phase mechanisms of the alkaline hydrolysis of PTE substrates. However, to relate the results to available experimental quantities referring to the analogous processes in aqueous solution (see Table 3.4), the influence of aqueous solvation was accounted for by polarizable continuum model (PCM) [275, 276] applied in the single point MP2/6-311++G(d,p) calculations employing gas phase HF/6-31+G(d) geometries.

All calculations were performed using Gaussian 03 program [175].

Multistep mechanism of sarin hydrolysis Two possible reaction pathways starting from a common structure of a prereactive complex of substrates (INT1)

Figure 3.10: HF/6-31+G(d) geometries of the stationary points along a reaction coordinate for the alkaline hydrolysis of sarin (**A** path). Selected interatomic distances [Å] are indicated. The analogous distances associated with structures from the **B** path are given in parentheses.

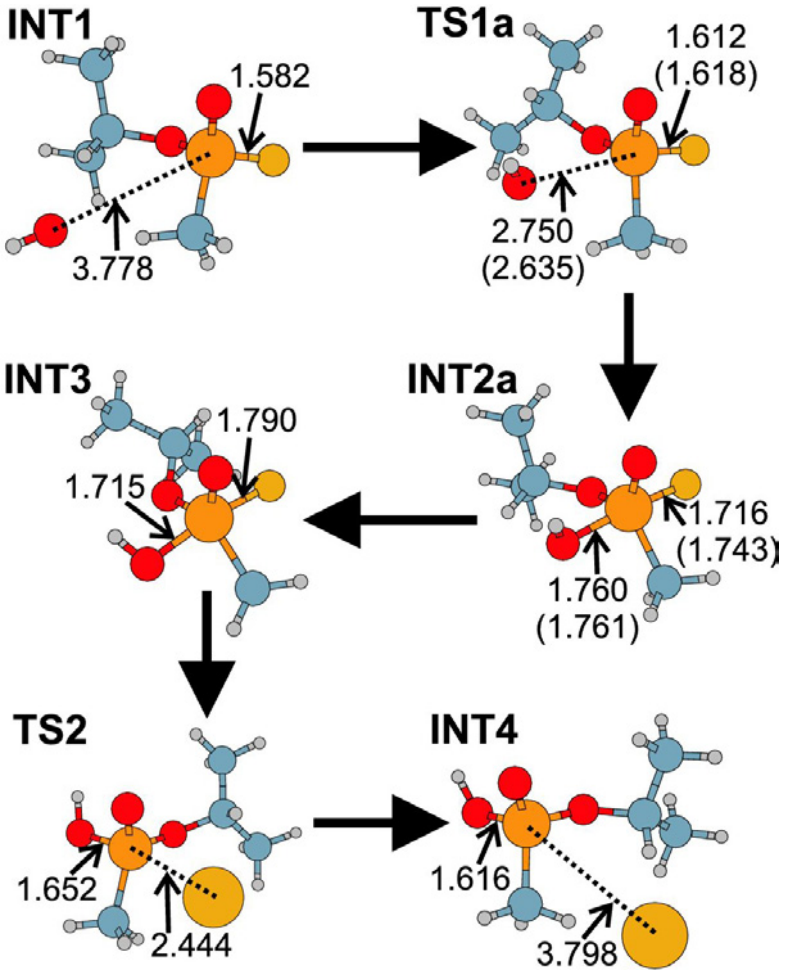


Figure 3.11: Energy profile for the first step of alkaline hydrolysis of sarin, i.e., formation of pentavalent intermediate. Potential energy changes with respect to the separately optimized reactant structure, INT1, were generated by the HF/6-31+G(d) IRC calculations starting with TS1 geometries corresponding to **A** and **B** paths (*black* and *blue dotted lines*, respectively). Reaction coordinate "0" corresponds to the TS1 structure; its negative and positive values indicate the directions toward reactants (INT1) and pentavalent intermediates (INT2a or INT2b), respectively. Variation in the interatomic distances is given for a phosphorus-hydroxide oxygen as well as phosphorus-fluorine atom separation. **A** and **B** pathways differ in the arrangement of a nucleophilic hydroxide, that can be positioned with its proton pointing in the direction of a phosphoryl oxygen atom (**A** path) or in an opposite way (**B** path).

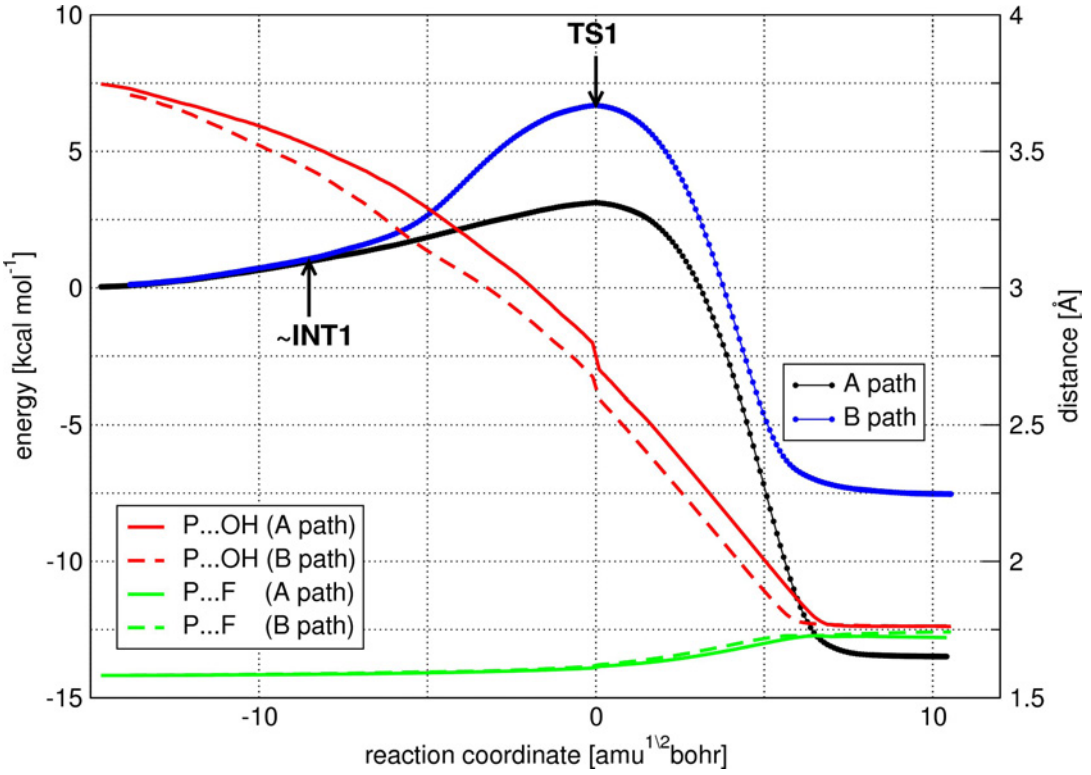


Figure 3.12: Energy profile for the last step of alkaline hydrolysis of sarin, i.e., decomposition of pentavalent intermediate. Potential energy changes with respect to the separately optimized intermediate structure, INT3, were generated by the HF/6-31+G(d) IRC calculations starting with TS2 geometry). Reaction coordinate "0" corresponds to the TS2 structure; its negative and positive values indicate the directions toward pentavalent intermediate (INT3) and products (INT4), respectively. Variation in the interatomic distances is given for a phosphorus-hydroxide oxygen as well as phosphorus-fluorine atom separation.

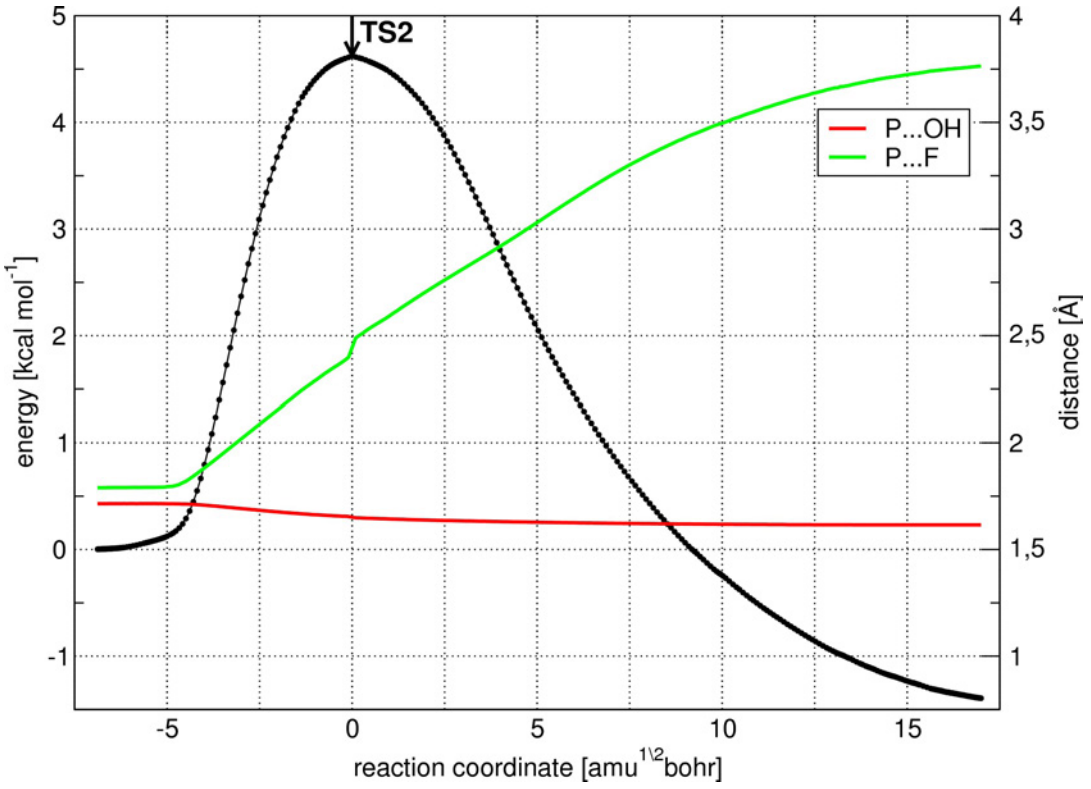
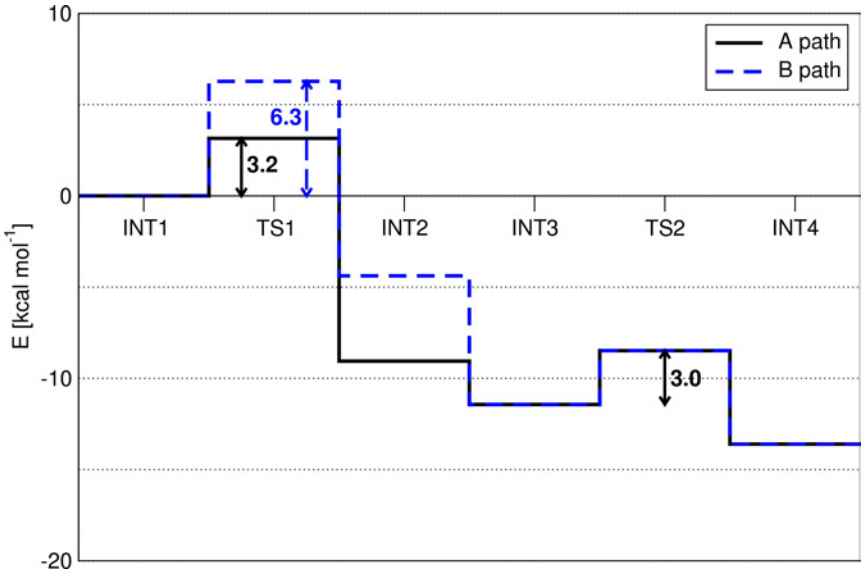


Figure 3.13: Potential energy surface along a reaction coordinate for the hydrolysis of sarin. MP2/6-311++G(d,p) relative energies (normalized with respect to the reactants geometry, INT1; ZPE included) were evaluated for structures fully optimized at the HF/6-31+G(d) level of theory. Energy barriers for the chemical steps encompassing the formation of the first intermediate and decomposition of the final one are indicated.



were determined at the HF/6-31+G(d) level of theory. The difference between these two reaction coordinates consists in the orientation of an attacking hydroxide ion; in a lower energy pathway, its hydrogen atom points in the same direction as the phosphoryl oxygen (path designated as **A**), whereas the other one involves hydroxide positioned in an opposite way (**B** path). In what follows, the "a" and "b" suffixes will be used to distinguish between the structures associated with **A** and **B** paths (e.g., TSa and TSb). Structures related to **A** path are presented in Figure 3.10. For the transition state geometries located along each of those pathways, HF/6-31+G(d) IRC simulation was employed to verify the connectivity of particular reaction pathway components. The results for two steps of sarin hydrolysis (i.e., formation and decomposition of the pentavalent intermediate) are plotted in Figures 3.11 and 3.12. Additionally, the MP2/6-311++G(d,p)//HF/6-31+G(d) energetics of both paths is compared in Figure 3.13.

The molecular scenario of the alkaline hydrolysis of sarin is consistent with the addition-elimination mechanism. The consecutive steps of this process can be summarized as follows. First, a stable complex of reactants (INT1) is formed

Table 3.5: Relative energies ($\text{kcal} \cdot \text{mol}^{-1}$) for the multistep hydrolysis of sarin as determined at the MP2/6-311++G(d,p)//HF/6-31+G(d) level of theory.

	ΔE^a	ΔG^b
INT1→TS1	3.2 (6.3)^c	4.4 (7.6)
INT1→INT2	-9.1 (-4.4)	-7.4 (-3.3)
INT2→INT3	-2.4 (-7.1)	-2.7 (-6.7)
INT3→TS2	3.0	3.0
INT3→INT4	-2.2	-3.3
INT1→INT4	-13.6	-13.3

[a] ΔE corresponds to the sum of electronic and zero-point energies.
[b] ΔG stands for the Gibbs free energy.
[c] The results for the **B** path are given in parentheses.

that subsequently proceeds through the two possible transition states for nucleophile attack (TS1a or TS1b) to the two pentacoordinate intermediates (INT2a and INT2b, respectively). TS1a with the hydroxide pointing "upward" (i.e., in the same direction as a phosphoryl oxygen atom) is associated with a lower energy barrier ($\Delta E = 3.2$ compared to 6.3 $\text{kcal} \cdot \text{mol}^{-1}$; see Table 3.5). Since the intermediate decomposition requires an isopropyl group of sarin to be oriented toward a fluorine atom, another trigonal bipyramidal intermediate (INT3; differing from INT2 in the conformation of the isopropyl substituent) undergoes transformation to final products via transition state TS2. The energy barrier in this case is equal to $\Delta E = 3.0$ $\text{kcal} \cdot \text{mol}^{-1}$ Analogous results concerning hydrolysis of a model compound, *O,O*-dimethyl phoshorofluoridate [269] suggest that energy barriers associated with the conformational transitions are insignificant compared to the chemical steps involving formation and decomposition of the pentavalent intermediate. Accordingly, only these steps were analyzed in the case of sarin. For both **A** and **B** paths of sarin hydrolysis, intermediate formation step seems to be associated with a higher energy barrier compared to the fluoride expulsion (see Table 3.5). Since the loss of fluoride occurs only via **A** path, which also favors the initial formation of an intermediate, the overall gas phase sarin hydrolysis appears to involve only the **A** path structures, i.e., those with a hydroxide stabilized by an additional interaction with a phosphoryl oxygen atom.

Table 3.6: Multistep hydrolysis: the activation energy barriers (in units of $\text{kcal} \cdot \text{mol}^{-1}$) as determined at the MP2/6-311++G(d,p)//HF/6-31+G(d) level of theory.

		DFP	sarin	tabun
intermediate formation	ΔE_a [a]	3.3 (5.0) [b]	3.2 (6.3)	5.7 (8.6)
	ΔG_a [c]	5.3 (7.6)	4.4 (7.6)	6.2 (9.4)
	ΔE_a^{PCM} [d]	3.4 (0.6)	3.1 (3.1)	4.6 (4.9)
intermediate decomposition	ΔE_a	2.3 (1.9)	3.0	0.4
	ΔG_a	2.7 (1.9)	3.0	1.0
	ΔE_a^{PCM}	0.5 (1.9)	-2.3	0.1

[a] ΔE_a corresponds to the sum of electronic and zero-point energies.
[b] The results for the **B** path are given in parentheses.
[c] ΔG_a stands for the Gibbs free energy.
[d] ΔE_a^{PCM} stands for the electronic energy in solution.

The results discussed herein are consistent with the mechanism of sarin hydrolysis proposed by Ornstein et al. [242]. Nonetheless, close inspection of structures and energetics of the corresponding potential energy surfaces supports the notion that only higher energy reaction pathway was proposed therein. While the results of Cramer et al. regarding alkaline sarin hydrolysis [240] encompass the reaction coordinate characterized by a more favorable hydroxide orientation, no attempts have been made to investigate an alternative reaction pathway. As regards the available experimental data, the highest Gibbs free energy barrier along the sarin hydrolysis pathway, i.e., 4.4 $\text{kcal} \cdot \text{mol}^{-1}$ (see Table 3.5) does not seem to compare favorably with an experimental value of 9.1 $\text{kcal} \cdot \text{mol}^{-1}$ (Table 3.4). Again, this theoretical analysis have been carried out for gas phase structures, while experimental data apply to an aqueous solution.

Alkaline hydrolysis of P-F, P-CN, P-O, and P-S bonds Basic hydrolysis of P-F and P-CN bonds in DFP and tabun, respectively, appears to follow multistep addition-elimination mechanism presented for sarin. More detailed discussion of these results is given in Ref. [269]. The energy barriers corresponding to the chemical steps of formation and decomposition of the pentavalent intermediate are provided in Table 3.6. In general, the rate-controlling step in the case of all

Table 3.7: Single-step hydrolysis: the activation energy barriers (in units of $\mathrm{kcal} \cdot \mathrm{mol}^{-1}$) as determined at the MP2/6-311++G(d,p)//HF/6-31+G(d) level of theory.

	paraoxon	parathion	demeton-S
$\Delta E_a{}^a$	7.4 (8.7)^b	8.6 (9.7)	4.8 (6.4)
$\Delta G_a{}^c$	7.7 (9.3)	8.8 (10.3)	6.8 (8.7)
$\Delta E_a^{PCM\,d}$	6.2 (4.9)	7.1 (5.3)	9.1 (7.2)

[a] ΔE_a corresponds to the sum of electronic and zero-point energies.
[b] The results for the **B** path are given in parentheses.
[c] ΔG_a stands for the Gibbs free energy.
[d] ΔE_a^{PCM} stands for the electronic energy in solution.

multistep mechanisms discussed here is associated with the intermediate formation. Experimental activation energy for the hydrolysis of tabun by an aqueous base is higher by 1.0 $\mathrm{kcal} \cdot \mathrm{mol}^{-1}$ compared to that of sarin (Table 3.4). Noticeably, reasonable resemblance occurs also for Gibbs free energy barriers associated with a rate-limiting step of sarin and tabun (Table 3.6).

In contrast to the complex addition-elimination reaction pathways for decomposition of P-F and P-CN bonds, base-catalyzed hydrolysis of P-O (paraoxon, parathion) and P-S bonds (demeton-S) proceeds according to the direct-displacement mechanism. In particular, no pentacoordinate intermediate is present and a single transition state is further decomposed by a leaving group departure. As previously, **A** and **B** paths differing in the position of a nucleophilic hydroxide are possible and the favored one, i.e., **A** path, involves the transition state structure with a hydroxide proton stabilized by phosphoryl oxygen (sulfur) atom. The activation energy barriers for these mechanisms are compared in Table 3.7. Noticeably, the results regarding paraoxon hydrolysis are consistent with experimental data suggesting associative bimolecular mechanism of alkaline hydrolysis that involves a single S_N2-like transition state structure [122]. Relative activation energy barriers for parathion and paraoxon decomposition are in a remarkable agreement with the chemical kinetics data. Assuming the 5.6-fold rate acceleration when comparing parathion and paraoxon hydrolysis [277] (the experimental data from Table 3.4 are not used herein as they were not gathered under common temperature), it corresponds to a decrease in activation free energy of about 1.1 $\mathrm{kcal} \cdot \mathrm{mol}^{-1}$ (at 40 ℃). Moreover, comparison of turnover numbers for a PTE-catalyzed reaction results

in similar values. Thus, the difference in the activation free energy barrier for parathion and paraoxon (*i.e.*, $\Delta\Delta G = 1.1$ kcal $\cdot$ mol^{-1}; see Table 3.7) resulting from this model is in an excellent agreement with the experimental results related to the enzymatic reaction.

Alkaline hydrolysis versus PTE-catalyzed process Because of the lack of experimental data for gas phase alkaline hydrolysis of the compounds studied herein, a concluding validation of the corresponding theoretical results cannot be performed. To compare the results to available experimental quantities describing basic hydrolysis in aqueous solution (see Table 3.4), self-consistent reaction field model [275, 276] of bulk water has been applied. Solvation-induced effects on the potential energy surface for particular reactions are given in Ref. [269]. Activation energy barriers in solution are compared in Tables 3.6 and 3.7. In general, the height of energy barriers is decreased upon solvation with the most pronounced differences occurring for the **B** paths. As a consequence, the reaction pathways involving a hydroxide proton positioned in an opposite way relative to the phosphoryl oxygen atom are associated with lower energy barriers than those employing an alternative hydroxide arrangement. Apparently, solvent is better suited for stabilization of the attacking hydroxyl group than is the phosphoryl oxygen alone. Since analogous solvent effects can be observed regarding the energy barriers for intermediate decomposition, it is presumably the intermediate formation step that controls the overall reaction rate, similar to the gas phase results. Additionally, on account of the strong polarity of the compounds studied herein, the reactant complex may no longer be associated with a minimum on the potential energy surface and the reaction barrier for the intermediate formation (or the overall reaction barrier in the case of single-step mechanisms) might correspond to the amount of energy required for separated reactants to reach the transition state geometry [242]. This could possibly explain an apparent discrepancy between experimental data and the results discussed herein. Another likely reason for the observed inconsistency is the error introduced by the use of the implicit solvation method itself [278]. As it has been shown for sarin and soman [279], polarizable continuum models does not seem to be capable of reproducing the experimental data. It is possibly due to the greater resemblance between the gas phase and the enzymatic transition state[271–274] that an exact agreement has been obtained for paraoxon and parathion gas phase results and experimental turnover numbers for the PTE-catalyzed process [277].

Since including the solvent effects by means of quantum mechanical continuum model does not seem to yield a reliable outcome, a comparison to experimental data obtained for processes occurring in an aqueous solution is also inconclusive.

As it can be seen from Table 3.4, the ordering of compounds in terms of their susceptibility to hydrolysis undergoes a major change upon PTE action. The most pronounced example encompasses paraoxon, which, despite being relatively insensitive to hydrolysis, is the best PTE substrate. Nonetheless, the recent theoretical analyses of PTE-catalyzed hydrolysis of paraoxon [250, 251, 258, 259] have confirmed the mechanism proposed on the basis of the experimental kinetic isotope effects [122]. In particular, no protonation of a leaving *p*-nitrophenolate is required, and the bridging hydroxide is a sufficiently strong nucleophile to initiate the reaction. It is worth emphasizing, however, that the reaction coordinate observed for an enzyme-catalyzed process might differ from that in a gas phase or an aqueous solution. In the case of PTE-catalyzed cleavage of paraoxon, the single-step process proposed herein is likely to be modified, as additional intermediates have been observed along a reaction pathway [250]. Still, the general characteristics of these two mechanisms have remained unaltered, and therefore the knowledge of the most probable events following the attachment of a hydroxide is of great importance for the successful modeling of a reaction taking place in an enzyme active site.

Catalytic fields for sarin hydrolysis Catalytic fields [31, 32] represent the optimal electrostatic characteristics of a hypothetical catalyst. The only knowledge required while generating the catalytic fields is that of the reactant and transition state geometries. Differential electrostatic potential around the INT1 and TS1a (or TS1b) structures, i.e., catalytic fields for the first step of sarin hydrolysis are presented in Figure 3.14. Depending on the way, a nucleophilic hydroxide approaches the sarin molecule, different arrangement of charges is required for enhancing the reaction. In particular, **A** path is favored by the presence of two negative charges on the "top" and "bottom" sides of the approaching hydroxide (see *side view* in Figure 3.14). Positive charge in the vicinity of a phosphoryl oxygen is optimal for the **B** path. As it will be discussed in what follows, this difference in comparison with the arrangement of charges present in the PTE active site might provide clues regarding the preferred pathway of enzyme-catalyzed sarin hydrolysis.

Since the exact geometry of reactant and/or transition state is not always available, it might be helpful to establish the sensitivity of catalytic fields to the accuracy of reactant/transition state geometry. Accordingly, catalytic fields for sarin hydrolysis were also generated on the basis of the "true" transition state and the approximate reactant structure. The latter was derived from the respective IRC simulations by selecting the geometry corresponding to 100 steps (~ 8.5 amu$^{1/2}$bohr), i.e., half way along the reaction coordinate leading toward INT1 structure (see the point labeled by ∼INT1 in Figure 3.11). It can be seen from Figure 3.15, that such approximate catalytic fields are in good agreement with the exact ones

Figure 3.14: Catalytic fields for intermediate formation step of alkaline sarin hydrolysis (**A** and **B** paths). The structures of reactants are shown in *ball-and-stick representation*. The electronic isodensity surface of 0.01 a.u. is coloured according to the differential electrostatic potential of transition state and reactants calculated at the HF/6-31+G(d) level of theory. The sign of the differential potential is inverted to reflect the electrostatic properties of a complementary molecular environment. *Red (blue)* colour denotes regions where a negative (positive) charge would be optimal for catalytic activity.

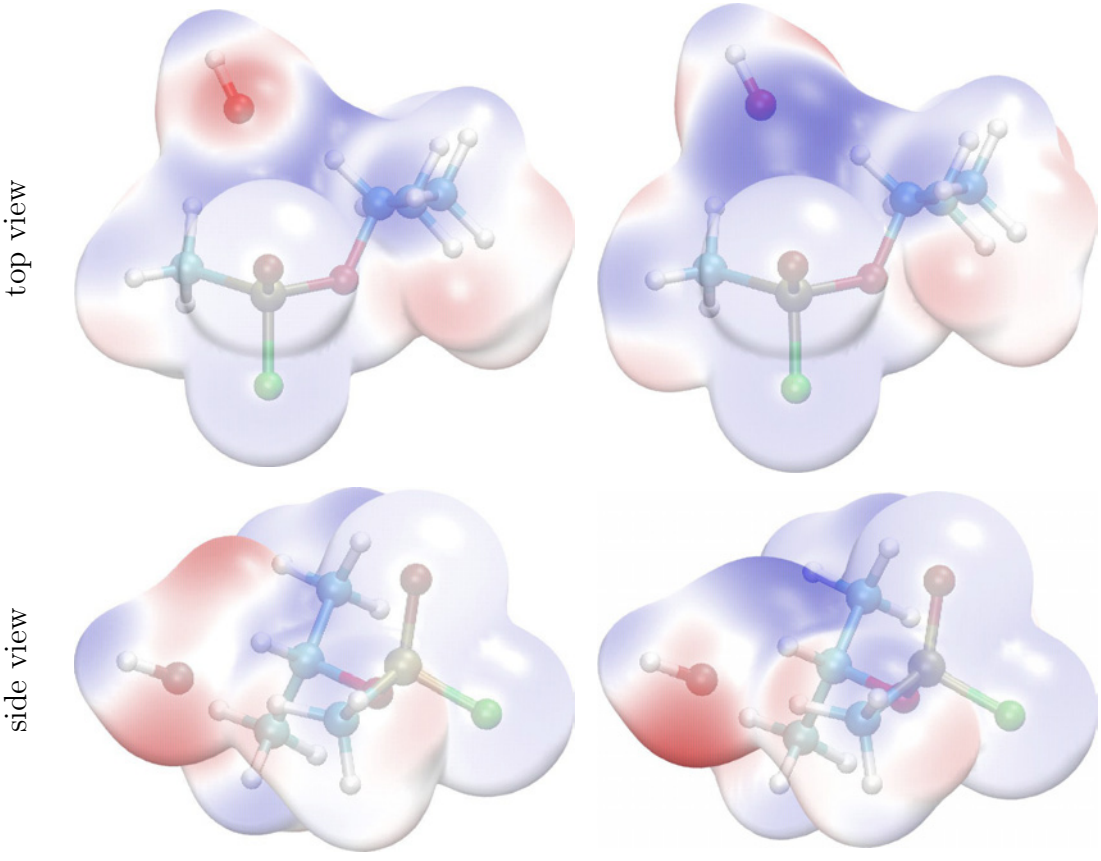
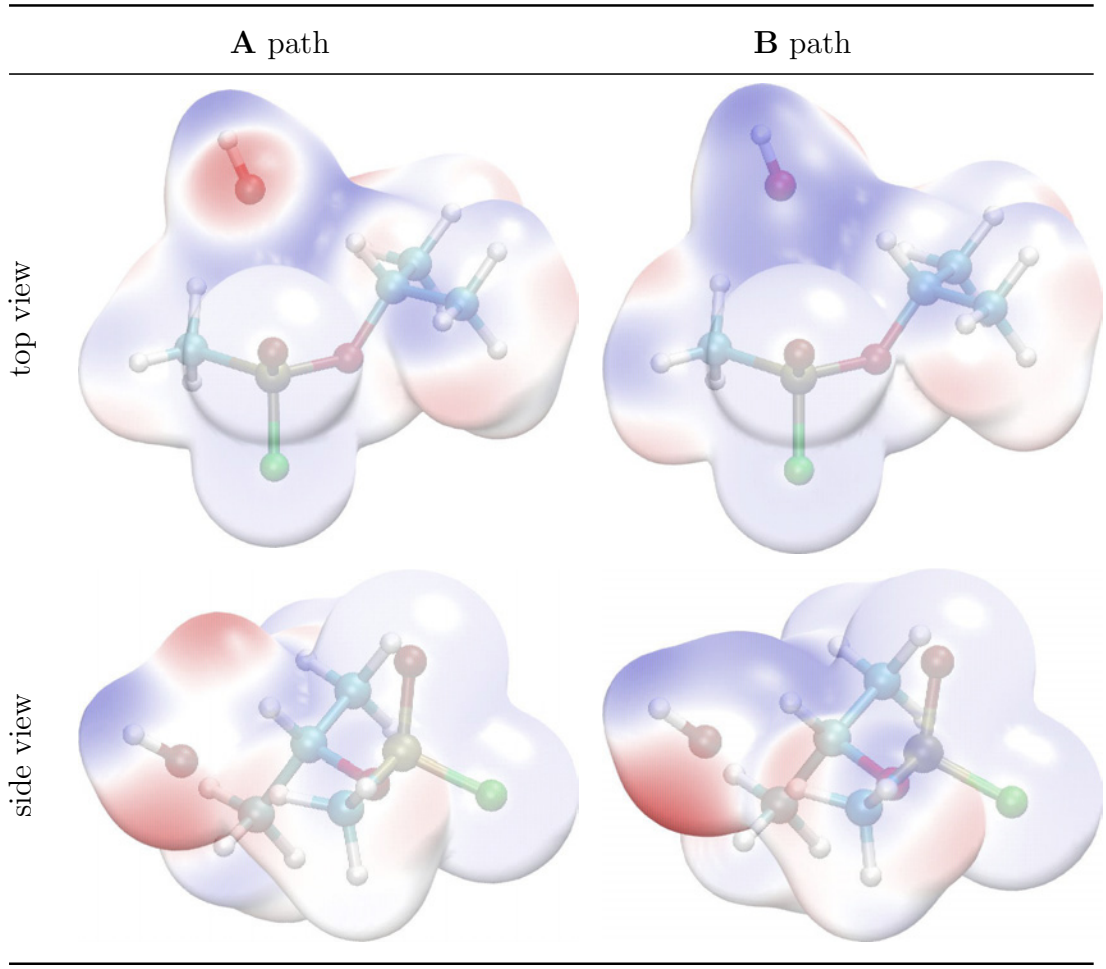


Figure 3.15: Approximate catalytic fields for intermediate formation step of alkaline sarin hydrolysis (**A** and **B** paths). The approximate structures of reactants (shown in *ball-and-stick representation*) were taken from IRC simulation as geometries obtained half way between transition state and actual reactants (see the point labeled by ∼INT1 in Figure 3.11). The electronic isodensity surface of 0.01 a.u. is coloured according to the differential electrostatic potential of transition state and reactants calculated at the HF/6-31+G(d) level of theory. The sign of the differential potential is inverted to reflect the electrostatic properties of a complementary molecular environment. *Red (blue)* colour denotes regions where a negative (positive) charge would be optimal for catalytic activity.



(Figure 3.14). This preliminary results suggest that qualitative features predicted by catalytic fields are relatively insensitive to the accuracy of initial geometries.

3.2.2 Modeling of the PTE-sarin complex

Well-founded model of an enzyme-substrate complex constitutes an essential prerequisite to the study of a mechanism of enzyme-catalyzed reaction [280]. Due to the considerable size of a protein-ligand system, preparation and testing of the initial model was performed with the utilization of molecular mechanics methods implemented in Charmm package [173] along with the all-hydrogen parameter set from Charmm27 force field [174].

Technical details The first step encompassed the development of sarin parameters necessitated by the lack of respective force field parametrization for this molecule. Based on the similarity to already parametrized molecules, initial set of parameters was assigned to isopropyl moiety. However, no analogous parameters were found for the fluoro and methyl group attached to phosphorus atom. Molecular mechanics description of a molecule requires both the intra- and intermolecular terms to be determined. The former includes equilibrium values for geometry parameters as well as the force constants for bond stretching, angle bending, and dihedral rotation. Since the normal modes of molecule vibrations originate from the coupled motion of multiple atoms (e.g., bond stretching and angle bending), their separation is not possible and, thus, the force constants associated with the potential energy curve for these particular deformations cannot be computed directly. Accordingly, the most reasonable approach to the force constants evaluation is the search for parameters with which the experimental and/or quantum mechanically-derived vibrational spectrum will be sufficiently reproduced. Taking advantage of the availability of automated iterative procedure for force constants fitting (Automated Frequency Matching Method [281]), the testing set of initial parameters was refined until the agreement with *ab initio* vibrational frequencies was improved.

The intermolecular parameter optimization involves the electrostatic and van der Waals interactions. General Charmm strategy with respect to intermolecular terms parametrization is to reproduce structural and energetic aspects of a given molecule interactions with water. The van der Waals constants depend heavily on atomic properties and, thus, they are relatively insensitive to changes in molecular environment. Consequently, the existing parameters from Charmm27 force field were utilized with no further attempt to modify them. Of special importance is the electrostatic term resulting from partial atomic charges assigned to particular

atoms. Noticeably, even the slight variation in point charges has a great impact on the overall interaction energy. Since partial atomic charges influence also the intramolecular terms (e.g., the dihedral rotational profiles), extreme care must be taken in their adjustment. The approach followed here involved the simultaneous manual modification of the initial set of electrostatic potential derived CHELP charges from quantum mechanical calculations to reproduce the energetics of interactions with water as well as the rotational profiles for selected dihedral angles.

Throughout all the steps of sarin parametrization, *ab initio* results obtained with Gaussian 03 program [175] were used as a reference. In particular, geometry optimization, vibrational frequency analysis and rotational profiles for selected dihedral angles were performed at the B3LYP/6-31G(d) level of theory. The same model chemistry was applied in the analysis of sarin-water molecule interactions. Minimum energy structures, CHELP charges and the corresponding interaction energies were evaluated with the counterpoise correction [140] for BSSE.

Several available X-ray structures of PTE are remarkably similar—the RMSD of Cα atoms does not exceed 0.2 Å. Factors considered during the selection of a model structure included its quality, the absence/presence of ligands (e.g., inhibitor molecules) within an active site and the correspondence to native enzyme. The latter resulted in the rejection of both cadmium and manganese-containing structures (two zinc ions are present in native PTE) as well as phosphotriesterase mutants. Similarly, apo forms of an enzyme were not considered. The remaining three structures encompass PTE complexes with the following inhibitors: diethyl 4-methylbenzylphosphonate (2.1 Å resolution of crystal structure), triethyl phosphate (1.9 Å), and diisopropyl methyl phosphonate (1.9 Å). Due to relatively high resolution and close resemblance of a ligand to the sarin molecule, the latter structure [282] was chosen (PDB accession code 1EZ2). Noticeably, the selected enzyme-substrate analog complex exhibits strong coordination of ligand phosphoryl oxygen to the more exposed zinc ion, as supported by the shortest zinc-oxygen distance of 2.5Å.

The mode of substrate analog binding within an active site of 1EZ2 phosphotriesterase structure is consistent with the proposed catalytic mechanism. The only difference between diisopropyl methyl phosphonate and sarin consists in substitution of the sarin fluoro leaving group by isopropyl moiety. As evidenced experimentally, the highest is the rate of R_P-sarin hydrolysis [283]. This finding can be rationalized by examination of the active site geometry. The latter is usually described as being composed of three pockets: the large one (formed by His257, Leu271, and Met317), the small one (residues Ile106, Leu303, Phe306, His57, and Gly60) and the leaving group pocket (Trp131, Phe132, Leu271, Phe306, Tyr309). Possibly, productive enzyme-substrate complex requires the leaving group (fluoro

substituent in the case of sarin) to be positioned in a leaving group cavity. In a similar manner, the size of a bulky substituent, isopropyl moiety, is probably the best suited for a large pocket. Finally, the small pocket should be occupied by the smallest substituent at phosphorus atom, i.e., methyl group. In the case of diisopropyl methyl phosphonate, the small, large, and leaving group pockets bind isopropyl, methyl and isopropyl groups, respectively. Simply replacing the isopropyl moiety occupying the leaving group with a fluoro group would give S_P-sarin isomer. To perform simulation with the more active sarin enantiomer, substrate molecule was spatially aligned with diisopropyl methyl phosphonate by superimposing the sarin methyl group onto the small pocket isopropyl moiety, so that the fluoro group remained positioned in a leaving group pocket and a desired configuration at phosphorus atom was retained.

From two identical monomers present in an asymmetric unit of X-ray structure, only A chain was included in further simulation (the active site is formed entirely from residues donated by the single subunit of a homodimer). The protein chain is composed of 329 amino acid residues. In addition, two zinc ions bridged by catalytic hydroxide and carbamylated lysine are present in an active site. Crystallographic water molecules (total of 138 molecules) accompanying the protein A chain were also included. Unless patched (see below), all aspartate and glutamate residues (including the carboxylic moiety of carbamylated lysine side chain) were negatively charged (-1), while lysine and arginine residues were assigned a positive charge of $+1$. The pK_avalues of particular titratable residues were verified independently using Propka method [284] and Uhbd program [285]. Protonation of histidines was determined based on the Optimal Hydrogen Bonding Network method [171] accessible via What If program [172]. With the exception of a positively charged $His^{+}254$, all the remaining histidine residues were modelled as neutral with proton attached either to δ (His55, His57, His123, and His230) or to ϵ-nitrogen atom (His201 and His257). The latter approach resulted also in the identification of possibly flipped positions of nitrogen and oxygen atoms forming the amide moiety of Gln212 residue (similar size of these atoms makes them indistinguishable by X-ray spectroscopy). Finally, both zinc ions were simulated as divalent cations Zn^{2+}. All the missing hydrogen atoms were added with Charmm procedure Hbuild and their positions were further optimized using a series of 500 steps of steepest descent (SD) minimization followed by conjugate gradient (CG) optimization repeated until the RMS gradient of 0.01 $kcal \cdot mol^{-1} \cdot Å^{-1}$ was satisfied.

To reduce the simulation effort, the whole system was truncated to a 25 Å sphere centered on the more exposed zinc ion, Zn2. Since the charged residues on protein surface might provide artificially large contribution to the energy, all solvent-exposed residues capable of forming salt bridges were patched with residues

Table 3.8: Harmonic force constants [$\text{kcal} \cdot \text{mol}^{-1}\text{Å}^{-2}$] acting within particular buffer regions.

buffer region	protein backbone	Cβ atoms	Cγ atoms	the rest
21–22 Å	0.27	0.24	0.23	0.22
22–23 Å	0.85	0.78	0.73	0.69
23–24 Å	1.44	1.31	1.22	1.18
24–25 Å	1.70	1.55	1.45	1.40

of the same geometry but neutral charge. Final net charge of the system was equal to -4.

Next step involved solvation of the system. The center of separately equilibrated 30 Å cubic box of $8 \cdot 10^3$ water molecules was superimposed with Zn2 position. Water molecules located beyond the 25 Å sphere and those with oxygen atom closer than 2.6 Å to any heavy atom of protein framework were removed. As a result, total number of 717 water molecules was added to the model. To optimize the positions of newly included solvent molecules, the latter were subjected to 1500 steps of SD minimization followed by CG until the gradient tolerance of 0.01 $\text{kcal} \cdot \text{mol}^{-1} \cdot \text{Å}^{-1}$ was obtained.

Final system used for simulation of enzyme-ligand complex included 313 amino acid residues, two zinc ions, hydroxyl and sarin residues as well as 824 water molecules (7316 atoms in total).

The actual MD run was preceded by the equilibration of all solvent molecules (including the crystallographic ones). All non-water atoms were kept fixed at their initial positions. Stochastic boundary potential (i.e., gradually increasing harmonic forces acting beyond the radius of 19 Å) was applied to the surface shell of water molecules to maintain a resemblance to bulk solvent and to prevent water from diffusion. After the initial optimization encompassing 600 steps of SD and 1500 steps of Adopted Basis Newton-Raphson (ABNR) minimization algorithms, molecular dynamics of solvent was performed. First, the system was heated from 10 to 300 K over 1 ps (30 K per 0.1 ps) followed by 25 ps of equilibration. Both stages involved 0.5 fs timestep. Nonbonded interactions were calculated using 20 Å cutoff distance. Finally, analogous optimization procedure (600 steps of SD followed by the ABNR minimization exited with gradient tolerance satisfied) was carried out.

In what follows, special treatment was required to avoid artificial results originating from truncation of protein framework—the overall shape of the sphere was maintained with an aid of harmonic constraints. Atoms within an inner reaction zone were free to move while the outer buffer zone beyond 21 Å was further divided into four buffer regions differing in the magnitude of harmonic potential acting on all atoms except for water molecules and hydrogen atoms (see Table 3.8 for details). Starting from the minimized final snapshot from water equilibration dynamics, the whole system was optimized during 500 steps of SD followed by 1250 steps of Powell minimization algorithms. No atoms were kept frozen. Water behaviour was controlled with stochastic boundary potential and the above listed harmonic constraints within the particular buffer regions were applied. Subsequently, molecular dynamics was carried out in two stages. The first one involved heating of the system from 10 to 300 K in three 10 ps time intervals (100 K per 10 ps). The following equilibration run lasted 50 ps. The common timestep of 1 fs was utilized throughout the simulation. In addition to the stochastic boundary potential and harmonic constraints, the behaviour of all heavy atoms beyond 21 Å was affected by a friction force with friction coefficients of 62 and 250 ps^{-1} (water oxygen and non-water heavy atoms, respectively). As previously, 20 Å cutoff for nonbonded interactions was used. Positively charged His254 (HSP) was initially selected because of its location between two negatively charged aspartate residues (i.e., Asp301 and Asp253). Due to unclear protonation state of this particular residue, the overall procedure was also performed for neutral form protonated either at Nδ or Nϵ (HSD and HSE, respectively). Throughout the course of these short test simulations, the protonated form of His254 turned out to be optimal, as both setups containing neutral forms resulted in higher RMSD values of not only the histidine under question, but also the residues encompassing PTE active site (Figure 3.16).

The production phase of MD simulation consisted of three independent 5 ns simulations differing in the values of seed for random number generator. The final frames of MD simulations were further optimized within two series of 250 SD steps and 1500 steps of ABNR.

Sarin parametrization Newly developed sarin parameters along with the corresponding Charmm27 topology file are given in the Appendix. The general performance of a final set of partial atomic charges (see Figure 3.17) is summarized in Table 3.9. An agreement between the reference *ab initio* interaction energy and the corresponding force field-derived value is very good. The empirical length of individual sarin-water molecule hydrogen bonds is slightly decreased, however, the observed values are still reasonable [286].

Figure 3.16: RMSD of His254 residue (*left-hand side*) and active site (*right-hand side*) heavy atoms along the trajectories corresponding to different protonation state of His254.

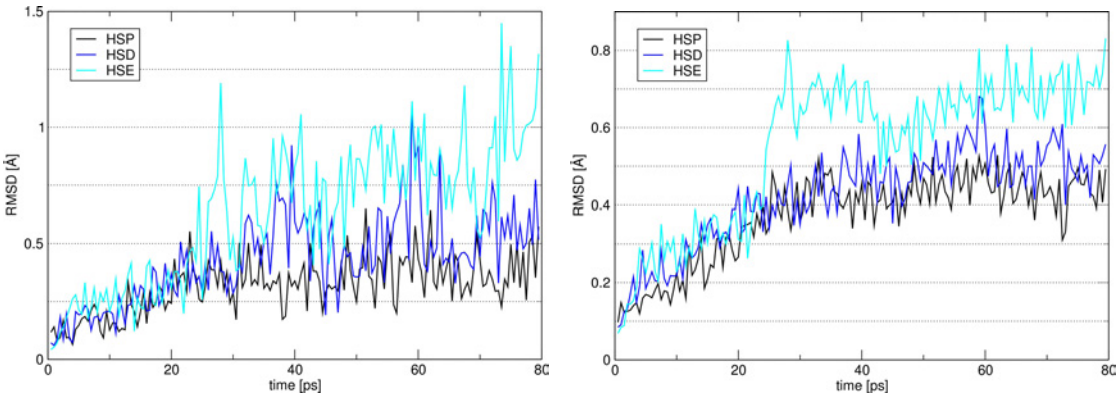
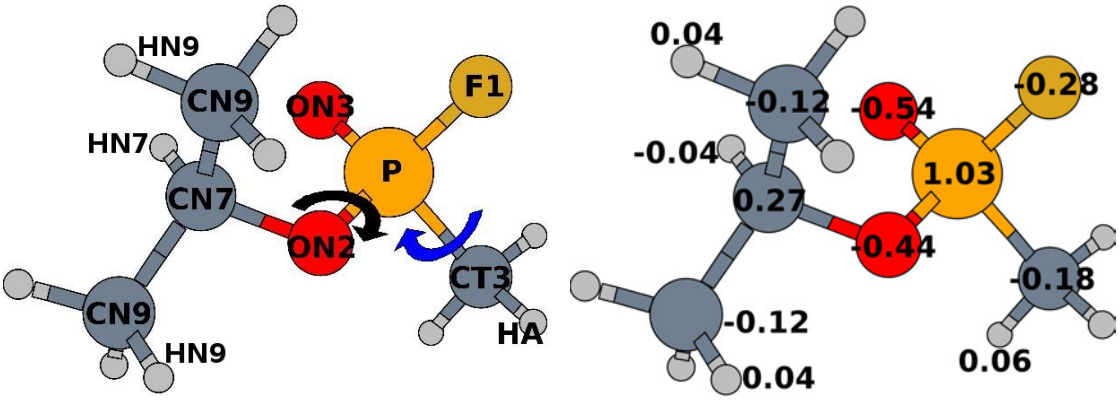
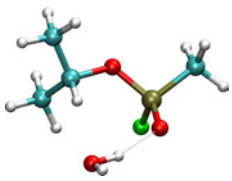
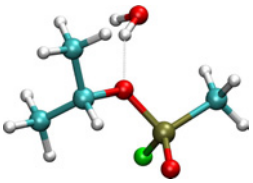
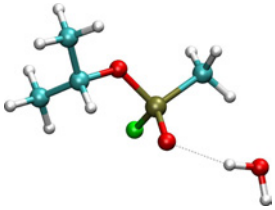


Figure 3.17: Sarin molecule—assignment of atom types (according to Charmm27 force field) and partial atomic charges. *Black and blue arrows* show the dihedral angles discussed in what follows.



Initial set of intramolecular parameters was obtained either directly from the *ab intio* reference geometry (i.e., equilibrium values for bonds and angles) or by similarity to the already parametrized molecules (i.e., force constants). The latter was verified in terms of an agreement between resulting vibrational spectrum and QM-generated normal modes. In particular, the starting RMSD between force field and *ab initio* vibrational frequencies was equal to 127 cm^{-1}. After the iterative fitting procedure was applied to selected force constants [281], the discussed RMSD value was further reduced to 112 cm^{-1} (Figure 3.18). Although the improvement is not

Table 3.9: *Ab initio* and force field-optimized sarin-water molecule complexes. Full geometry optimization and interaction energy calculation were performed at the B3LYP/6-31G(d) level of theory. For CHARMM27 calculation, the final set of partial atomic charges presented in Figure 3.17 was employed. Since no novel atom type was introduced during the parametrization procedure (see Figure 3.17 for details), already existing van der Waals parameters were also utilized without any further modification. ΔE [kcal·mol^{-1}] and R [Å] correspond to the interaction energy and the shortest minimum energy distance, respectively.

	ab initio	FF	*ab initio*	FF	*ab initio*	FF
ΔE	-7.06	-7.07	-3.72	-3.82	-6.64	-6.83
R	1.916	1.753	2.100	1.974	1.971	1.759
geometry						

very impressive, these results can also indicate that the initial set of force constants was already sufficiently close to the optimal values. Nonetheless, reasonable agreement between empirical and QM vibrational spectra confirms the validity of tested bonded parameters.

Comparison of force field-optimized sarin geometry with the QM-derived structure given is in Figure 3.19. Noticeably, the reference geometry is very well reproduced within the empirical approach.

Another test of force field performance involves the correct estimation of potential energy surface associated with conformational changes. In particular, two dihedral angles (marked by arrows in Figure 3.17) were considered due to their uniqueness within the available CHARMM27 torsional parameters. Figure 3.20 shows the energy profile corresponding to rotation of phosphorus atom-attached methyl group (*i.e.*, rotation about the P-C bond). As indicated by a perfect match between dashed black line (total CHARMM energy) and black dots (QM-generated values), remarkable agreement with the reference *ab initio* energy profile was obtained. Additional analysis of the magnitude of individual empirical energy terms reveals the torsional potential energy as a main contribution to the overall poten-

Figure 3.18: Vibrational frequency analysis of sarin molecule—performance of force field parameters *(squares and triangles)* with respect to the B3LYP/6-31G(d) calculation *(circles)*.

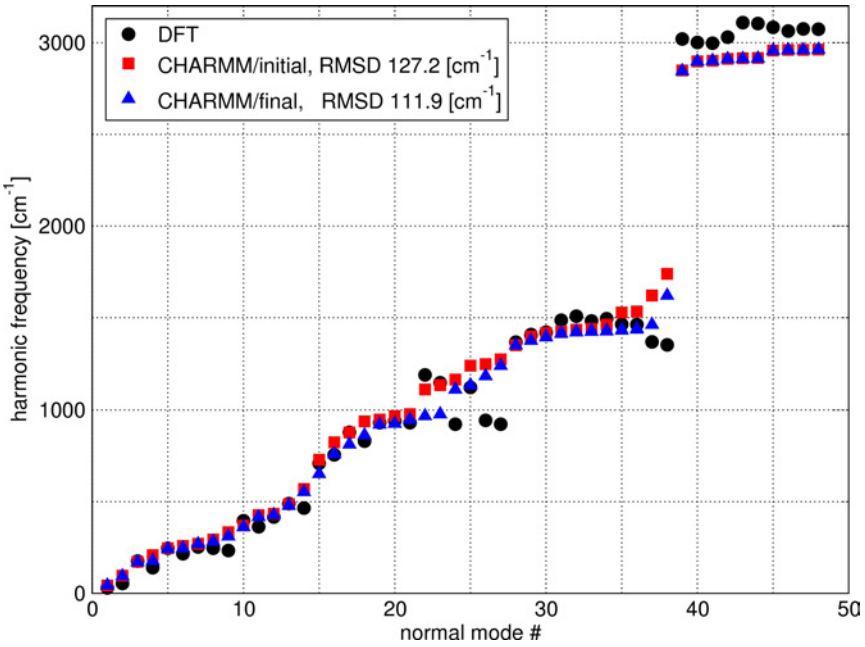


Figure 3.19: Comparison of the B3LYP/6-31G(d) and force field-optimized sarin geometries. *In blue* given is the empirical geometry; RMSD between the two structures is equal to 0.123 Å.

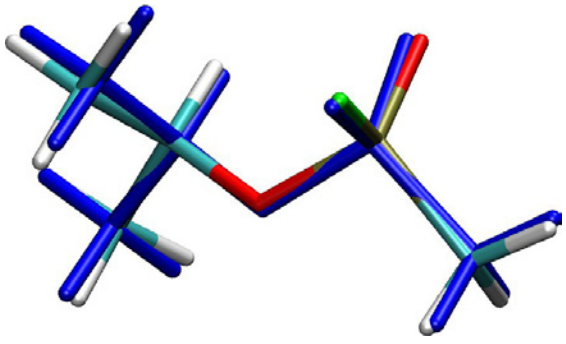
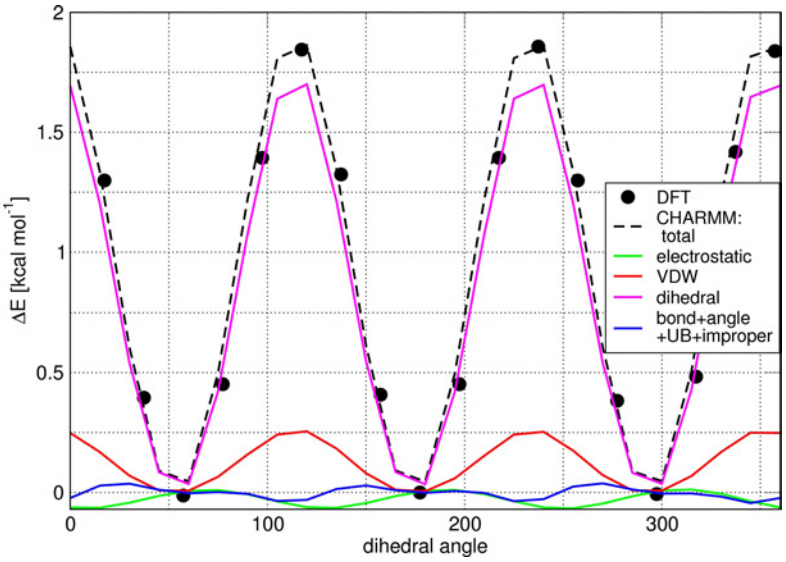


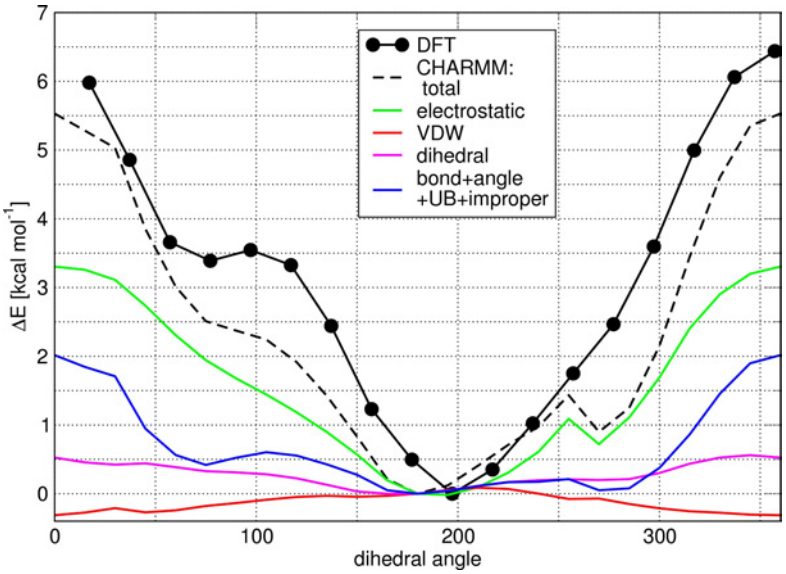
Figure 3.20: Potential energy as a function of the F1-P-CT3-HA dihedral angle. DFT energy profile was obtained using the B3LYP/6-31G(d) model chemistry.



tial energy curve. Only a fraction of rotational barrier arises from van der Waals interactions, whereas electrostatic term is of minor importance in this case.

This is not a case, however, for the torsional profile of another dihedral angle considered here, *i.e.*, the rotation about P-O bond (Figure 3.21). The match between total Charmm energy and the QM-derived one is no longer as close as previously. The empirical equilibrium dihedral angle is shifted by 10 degrees and the height of barrier is underestimated with respect to the *ab initio* value. Finally, empirical energy profile exhibits additional minimum for 260 degrees, not present in the reference potential energy curve. Contrarily to the P-C bond rotation, total Charmm energy is dominated by electrostatic term. Thus, the discussed torsional profile is very sensitive to assignment of partial atomic charges and great care must have been taken while performing the latter procedure. In fact, the final set of partial atomic charges constitutes a compromise between correct interaction energies with water and reasonable behaviour of the P-O bond rotational profile. Remarkably, imperfect match between empirical and QM energy profiles for P-O bond torsion does not imply its uselessness, since the artificial minimum is very shallow and it is not likely to influence the overall behaviour of sarin molecule during molecular mechanics simulation.

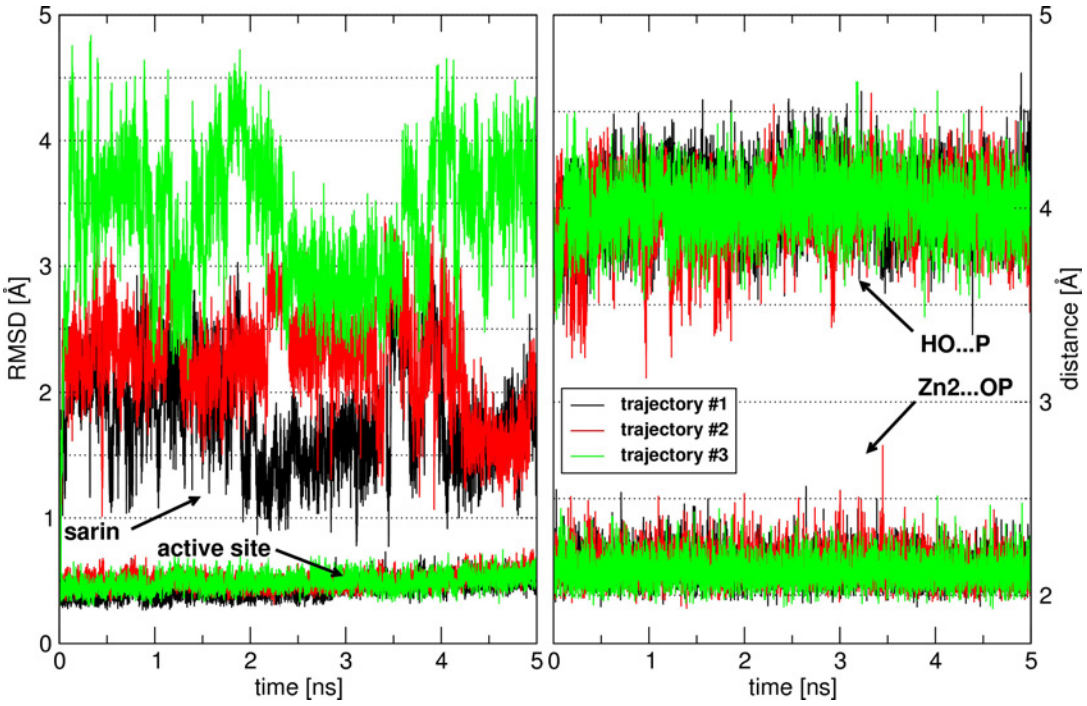
Figure 3.21: Potential energy as a function of the CT3-P-ON2-CN7 dihedral angle. DFT energy profile was obtained using the B3LYP/6-31G(d) model chemistry.



Molecular dynamics simulation of PTE-sarin complex The dynamic properties of PTE-sarin complex were analyzed based on three independent 5 ns MD simulations. The stability of energy components and structural parameters (e.g., RMSD of protein geometry from crystal structure, Figure 3.22) ensured that the system under study is in equilibrium. As evidenced by RMSD values of active site heavy atoms oscillating about 0.5 Å (Figure 3.22), active site geometry is well preserved. Presumably, high degree of active cleft organization results from the presence of two zinc ions that hold ligands in relatively stable positions. Additional stabilization of the bridging hydroxyl group is achieved by hydrogen bond interaction with Asp301 residue (Figure 3.23). Analogous results obtained for all three trajectories also support these conclusions.

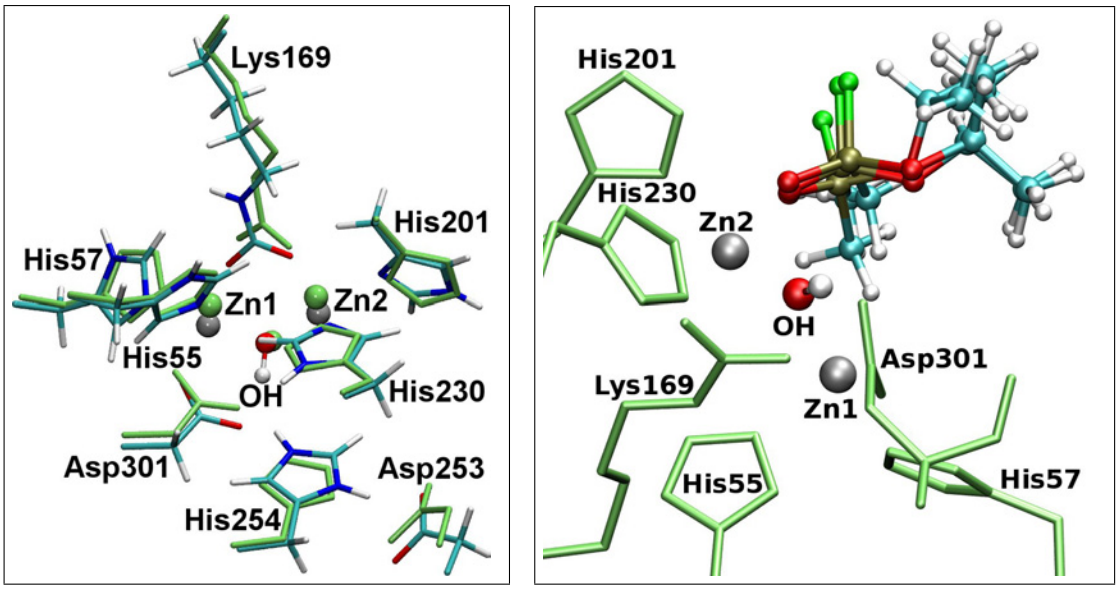
In addition to the dynamic behaviour of active site residues during simulation, their deviation from initial X-ray structure should also be analyzed. Figure 3.23 shows the arrangement of PTE active site from final MD snapshot compared to initial framework (only the first trajectory is considered, since the two remaining ones exhibit essentially identical characteristics). The RMSD of active site heavy atoms with respect to X-ray structure is equal to 0.33 Å (mean value for the three trajectories). In contrast with this low value remains the corresponding displacement of sarin molecule—depending on the trajectory, values of RMSD in the range of 2.97–3.43 Å are observed. It should be kept in mind, however,

Figure 3.22: RMSD of the PTE active site (i.e., His55, His57, carbamylated $Lys^{-}169$, His201, His230, $Asp^{-}301$, OH^{-}, $Zn^{2+}1$, and $Zn^{2+}2$) and sarin heavy atoms (*left-hand side*) and variation in selected distances (*right-hand side*) along the three independent MD trajectories.



that initial sarin positioning was very crude, as it involved only spatial alignment of separately optimized sarin molecule with an inhibitor present in X-ray structure. It can be seen from Figures 3.22 and 3.23 that relatively high RMSD values observed for sarin molecule 3.22 result mainly from the degrees of freedom associated with an isopropyl group, whereas the phosphoryl oxygen is firmly attached to the more exposed zinc ion, Zn2. In fact, this interaction provides a kind of anchoring for sarin molecule, since only hydrophobic interactions are possible for the remaining part of sarin residue and protein cavities that accommodate particular substituent at phosphorus atom (especially the large and leaving group pocket) provide relatively much space for PTE substrate. Another interesting observation concerning the final sarin placement relative to the initial one is that sarin methyl group becomes buried in the small group pocket, entirely in agreement with initial assumptions behind the starting mode of sarin positioning.

Figure 3.23: *Left-hand side:* deviation of active site residues in the optimized final MD snapshot relative to the X-ray structure (shown in *green colour*). *On the right:* comparison of the sarin positioning (*ball-and-stick representation*) after the three independent MD simulations.



Of utmost importance for further simulation of PTE-catalyzed process is the distance between hydroxyl oxygen and phosphorus atom corresponding to the reaction coordinate. Due to the stable hydroxyl position and strong zinc-phosphoryl oxygen interaction, HO...P separation is relatively constant and its average value (i.e., 4.00 Å) is reasonable for a reactive complex to be formed (Figure 3.22).

Recently published results of MD simulation performed for phosphotriesterase-soman complex [257] have implied that the active site geometry is flexible in a sense that it accommodates several zinc ions coordination patterns. The latter was not observed for any of the MD trajectories reported here as no changes in zinc ions coordination occurred (Figure 3.24). While this finding is in disagreement with the results presented in Ref. [257] it might be argued that the difference results from the substitution of a bridging hydroxide by a water molecule. It has been shown both theoretically [287] and experimentally [264] that the bridging solvent molecule is most likely to be a hydroxide and, thus, its replacement by water might disrupt the active site framework.

Figure 3.24: Coordination pattern of zinc ions shown as zinc-ligand distances along the three independent MD trajectories.

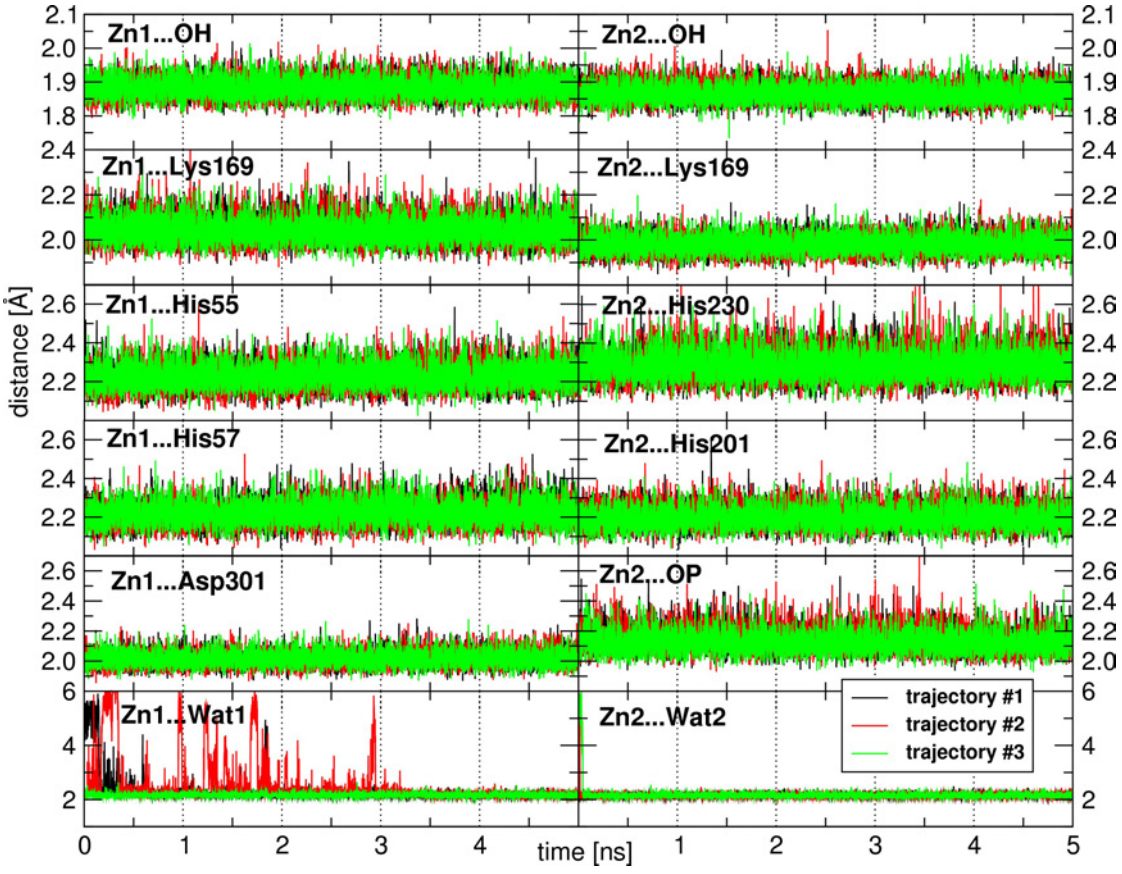
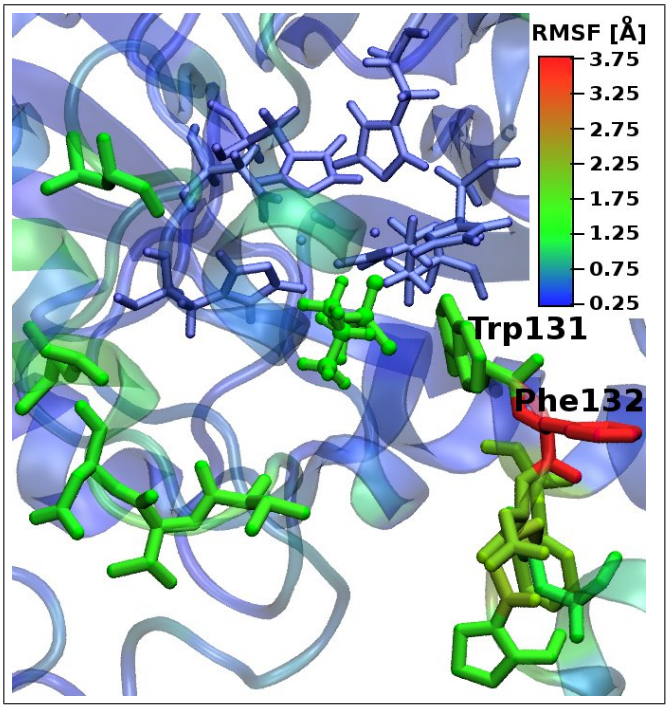


Figure 3.25: Mobility of PTE residues in terms of their RMSF values (i.e., RMSD averaged over the trajectory). *Blue* and *red colours* indicate low and high flexibility, respectively. Active site residues are shown as *thin sticks*; *thicker sticks* (or *ball-and-stick representation* in the case of sarin) denote residues with RMSF exceeding 1.0 Å.



Due to the presence of a water molecule in the vicinity of each zinc ion (Figure 3.24), both zinc ions acquire octahedral coordination pattern. Since PTE crystal structures show the presence of at most five zinc ligands, this outcome might result from the the deficiencies in force field description of zinc complexes by means of nonbonded method [256], as applied herein. Since this simulation aimed at revealing the details of sarin binding and providing a starting geometry for further QM/MM and QM models, this aspect was not pursued.

Exploration of flexibility of a protein-substrate complex during MD simulation shows the increased mobility of several other protein residues. As presented in Figure 3.25, these residues are located almost exclusively within the protein regions assigned to the leaving group pocket, which is consistent with the role of the latter: motions of residues forming leaving group cavity (Trp131, Phe132) might facilitate the release of a fluoride (or another leaving group).

3.2.3 QM modeling of PTE-catalyzed sarin hydrolysis

In the first attempt at reaction pathway modeling, umbrella sampling simulation coupled with QM/MM description of the system was applied to reveal the potential of mean force, i.e., the free energy surface along a specific reaction coordinate. To allow for the sufficient sampling of a conformational space, the QM region (i.e., sarin molecule, both zinc ions and a bridging hydroxide as well as the side chains of two zinc ligands: Asp301 and carbamylated Lys169; a total of 38 atoms including link atoms) was described at the semiempirical level of theory. Both AM1 [133] and PM3 [134] semiempirical methods (as available in Charmm program [173]) were tested. Initial optimization and MD simulation of the enzyme-reactant complex gave the most reliable outcome when conducted with the AM1 method and weak restraints to control the zinc-histidine coordination. In the following umbrella sampling simulation, two definitions of a reaction coordinate were considered, i.e., the reaction coordinate employing oxygen-phosphorus atom distance (the bond being formed) or the more complex one corresponding to the difference between oxygen-phosphorus and phosphorus-fluorine distances (the bond being broken). Unfortunately, none of this approaches yielded a meaningful result as the former failed to describe any bond breaking, while the latter resulted in an unrealistic free energy profile and only P-F bond partially broken.

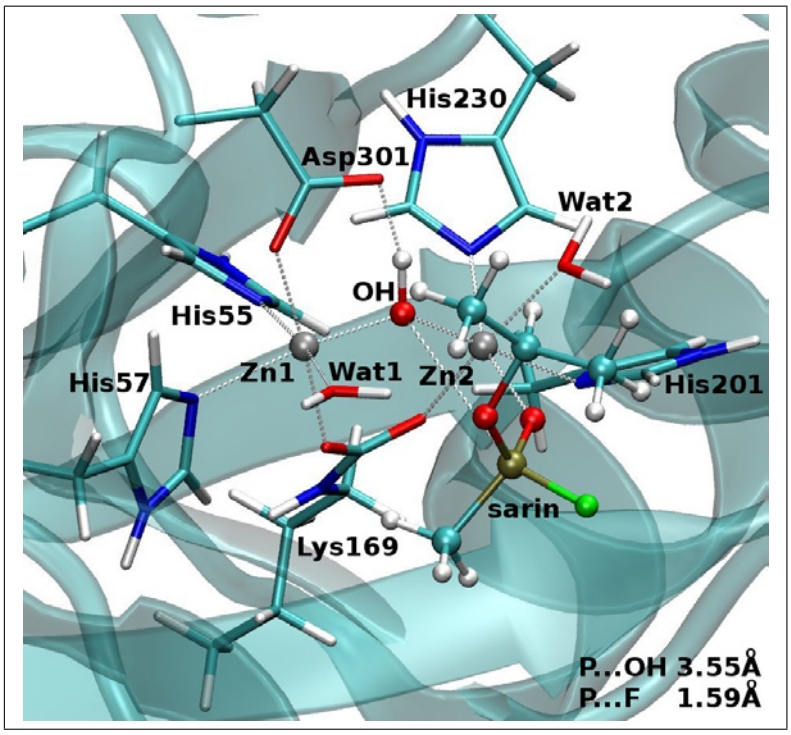
Subsequently, two alternative approaches were employed differing in the size of a system under study. The most obvious way of improvement upon the previously observed flaws of semiempirical description would be to model the QM region with a more reliable *ab initio* level of theory. However, computational demands of the latter limit its applicability to optimization of the stationary points along a reaction pathway. Even with the QM region restricted to reactants, any optimization procedure lasts much longer due to tedious multi-step reoptimization of enzyme environment. Hence, more accessible quantum-chemical description of a limited active site model was also employed as a complementary way to provide quick insight into the reaction scenario. In what follows, preliminary results of QM/MM and QM models of PTE-substrate and PTE-intermediate complexes will be compared.

Technical details Molecular mechanics model of PTE-sarin complex (section 3.2.2) served as a starting point for quantum mechanical modeling of the enzyme-catalyzed reaction pathway. The MM part was treated in a way analogous to the one employed in an entirely force field based simulation. In particular, harmonic constraints were applied within a buffer region to avoid non-physical effects originating from truncation of a protein framework, while water behaviour was controlled by stochastic boundary potential (see section 3.2.2 for

details). Prior to QM/MM optimization, the entire system was reoptimized until RMS gradient of 10^{-5} kcal·mol^{-1}·Å^{-1} was reached. QM region selected for the following calculations encompassed the reactants, i.e., sarin molecule and hydroxyl ion. Initial optimization employed the HF/3-21G model chemistry and consisted of two stages. First, thorough optimization of the enzyme environment was performed with QM atoms kept frozen. The resulting geometry was reoptimized without any constraints (except for those applied at the boundaries of the system investigated). Both stages were carried out until convergence criterion (i.e., RMS gradient of 10^{-5} kcal·mol^{-1}·Å^{-1}) was satisfied. The resulting geometry was then minimized with QM region described at the B3LYP/6-31G(d) level of theory and using the same convergence criterion. To model the enzyme-product complex, approximate structure obtained from the initial umbrella sampling simulation was subjected to the analogous procedure as that employed for preparation of an enzyme-reactant complex. The phosphorus-fluoride distance of 2.5 Å in a starting structure decreased significantly in the course of optimization arriving at the final value of 1.76 Å. Thus, the final optimized geometry turned out to represent the pentavalent intermediate. All QM/MM calculations were performed with Charmm [173] interfaced with Gamess(US) [170].

Initial geometry of QM model was prepared using the HF/3-21G//Charmm27 structure of PTE-sarin complex. Enzyme active site consisted of the two zinc ions along with their entire coordination spheres (i.e., His55, His57, Lys 169, His201, His230, and Asp301 residues). Amino acid residues were truncated so that only respective side chains were kept. Due to uncertainty in the number of zinc ligands, two cases were considered differing in the presence of water molecules bound to zinc ions (as found in the final MM and QM/MM geometries). Including these water molecules (Wat1 and Wat2; see Figure 3.26) resulted in pentacoordinated zinc ions, as both Wat1 and Wat2 molecules were expelled from the first coordination shell of Zn1 and Zn2 ions. This outcome appears to support zinc preference to become pentacoordinated as suggested by X-ray data. However, water molecule initially interacting with Zn1 (i.e., Wat1) was found to be hydrogen bonded to both Asp301 and sarin ester oxygen atom. To test the possible influence of this water molecule on the reaction coordinate, two model systems were considered consisting of a total of 74 or 77 atoms (differing by the absence or presence of aspartate- and sarin-bound water molecule). Stationary points along the reaction coordinate were calculated using the B3LYP/6-31G(d) model chemistry. Final optimizations involved no constraints as they were followed by vibrational frequency calculation to verify the nature of a given stationary point and to obtain thermodynamic quantities (e.g., enthalpy and Gibbs free energy).

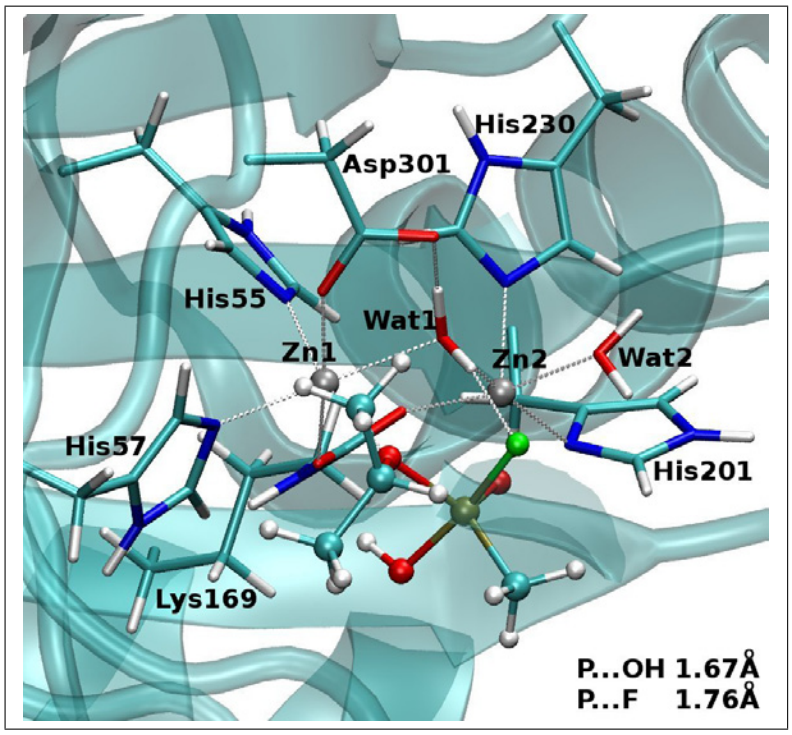
Figure 3.26: B3LYP/6-31G(d)//Charmm27 geometry of the PTE-substrate complex.



QM/MM geometry The structure of an enzyme-substrate complex optimized using the B3LYP/6-31G(d)//Charmm27 model chemistry is given in Figure 3.26. While RMSD of active site heavy atoms with respect to starting MM model is equal to 0.15 Å, the displacement of sarin is more pronounced as it is associated with RMSD of 0.65 Å. Nonetheless, the key sarin-active site contacts considered as crucial for catalysis are retained. In particular, the distance associated with a reaction coordinate, i.e., HO...P separation, is decreased arriving at the final value of 3.55 Å. The octahedral coordination of both zinc ions, resulting from the presence of an additional water molecule at each metal ion, remains the same as in MM structure. Since this region of PTE active site was treated by force field parameters, no improvement upon using this particular QM/MM scheme could be gained in terms of description of zinc coordination pattern.

As another stationary point on the potential energy surface, the geometry of PTE-intermediate complex was determined. Surprisingly, this particular structure was obtained using the probable PTE-product complex as a starting point. In the course of optimization, an approximate geometry of product evolved to-

Figure 3.27: B3LYP/6-31G(d)//Charmm27 geometry of the PTE-intermediate complex.



ward the pentacoordinate intermediate interacting with binuclear zinc center by means of both phosphoryl and ester oxygen atoms (Figure 3.27). Similarly to PTE-substrate complex, octahedral coordination of both zinc ions is retained also in a discussed geometry. Interestingly, water molecule initially coordinated to the more buried zinc ion was found to serve as a bridging ligand held firmly by hydrogen bonding interactions to Asp301 carboxylate and the leaving fluoride. This outcome might suggest the reaction mechanism encompassing a nucleophilic attack of a bridging hydroxide accompanied by substitution of the latter with a water molecule bound to more buried metal ion, Zn1. It is unclear, however, whether this events are concerted or if one precedes another. Moreover, the structure of PTE-intermediate complex is higher in energy relative to the PTE-substrate geometry by about 40 $\mathrm{kcal} \cdot \mathrm{mol}^{-1}$, rendering these results inconclusive even without the knowledge of the transition state(s) separating the reactant and intermediate state.

Table 3.10: Selected distances [Å] in the reactant, transition state and intermediate structures of the PTE-catalyzed sarin hydrolysis as derived from the B3LYP/6-31G(d) calculations. The models composed of 74 and 77 atoms differ in the absence/presence of additional water molecule. QM/MM models correspond to the B3LYP/6-31G(d)//Charmm27 structures given in Figures 3.26 and 3.27.

Model		HO...P	P...F	PO...Zn2	Zn1...Zn2
Reactant	74 atoms	3.02	1.60	2.21	3.52
	77 atoms	2.95	1.60	2.19	3.48
	QM/MM	3.55	1.59	1.95	3.38
Transition state	74 atoms				
	77 atoms	2.17	1.63	2.25	4.56
	QM/MM				
Intermediate	74 atoms	1.81	1.70	1.89	4.17
	77 atoms	2.01	1.64	2.18	4.60
	QM/MM	1.67	1.76	1.85	3.71

QM geometry Enzyme-substrate and enzyme-intermediate complexes were obtained at the B3LYP/6-31G(d) level of theory using small representation of PTE active site including both zinc ions along with the respective coordination shells (see geometries provided in Figure 3.28). PTE-reactant structure bears close resemblance to QM/MM results as all crucial contacts are maintained. It is remarkable that despite of the two water molecules no longer interacting with zinc ions, QM active site model is in good agreement with the parallel simulation employing MM description of an enzyme environment. However, HO...P distance equal to 3.02 Å (Table 3.10) is much shorter than those predicted in MM or QM/MM simulations ($\sim$ 4 and 3.55 Å, respectively). Including aspartate- and sarin-bound water molecule (model referred to as "77 atoms") does not affect the key active site distances (Table 3.10).

The complexes of PTE active site-bound pentacoordinate intermediate resulting from the nucleophilic attack of a hydroxide on sarin molecule differ qualitatively depending on the presence of the additional water molecule (Figure 3.28). If the latter is not included (model referred to as "74 atoms"), more exposed zinc

Figure 3.28: Structures of the PTE active site in complex with substrate, transition state and pentacoordinate intermediate *(ball-and-stick representation)* as derived from the B3LYP/6-31G(d) calculations. The models composed of 74 and 77 atoms differ in the absence/presence of additional water molecule.

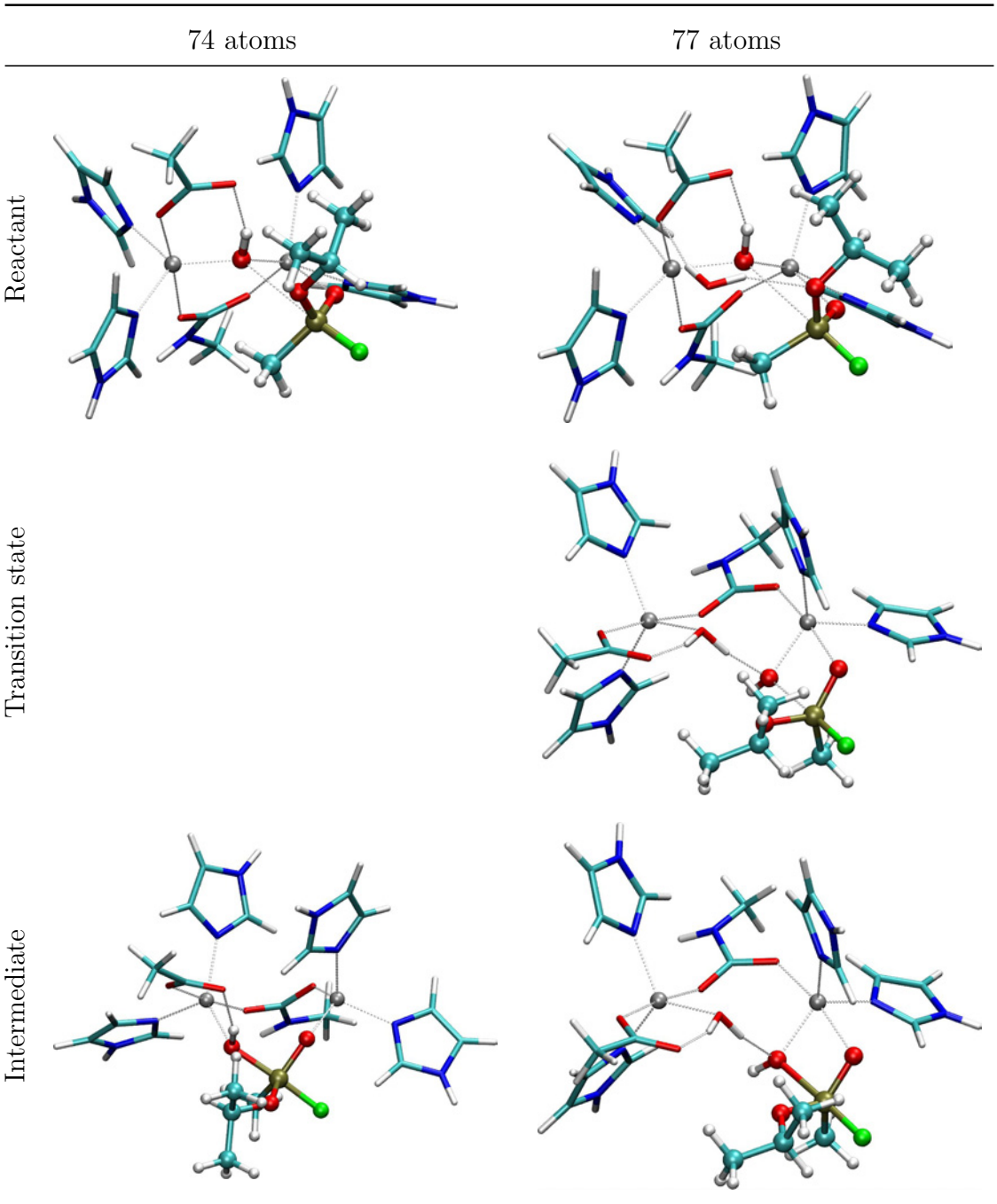


Table 3.11: B3LYP/6-31G(d) relative energies [kcal · mol^{-1}] of the transition state and intermediate with respect to reactant structures as derived from B3LYP/6-31G(d) calculations. The models composed of 74 and 77 atoms differ in the absence/presence of additional water molecule. ΔE refers to electronic energy; ΔE_{ZPE} denotes the sum of electronic and zero-point energies, while ΔH and ΔG stand for the enthalpy and Gibbs free energy, respectively.

Model		ΔE	ΔE_{ZPE}	ΔH	ΔG
Reactant	74 atoms	0.0	0.0	0.0	0.0
	77 atoms	0.0	0.0	0.0	0.0
Transition state	74 atoms				
	77 atoms	4.7	4.8	3.6	5.3
Intermediate	74 atoms	15.4	16.0	15.7	16.6
	77 atoms	4.6	5.1	4.3	5.4

ion becomes tetracoordinated as the oxygen atom from a former hydroxyl group is no longer capable of bridging both Zn^{2+} ions. The coordination number of more buried zinc ion remains the same due to maintained contact with nucleophilic oxygen atom. Accordingly, pentavalent intermediate binds via simultaneous interaction with both Zn^{2+} ions, in a manner similar to that observed in the case of QM/MM PTE-intermediate structure (Figure 3.27). If present, water molecule appears to influence the reaction scenario as it serves as a new ligand to more buried zinc ion, Zn1, while pentavalent intermediate is bidentally coordinated to Zn1 (Figure 3.28). Consequently, both zinc ions maintain their trigonal bipyramidal coordination pattern. Compared to the corresponding reactant geometry, significant increase in the separation of zinc ions can be observed (Zn1...Zn2 distance of 4.60 in contrast with 3.48 Å).

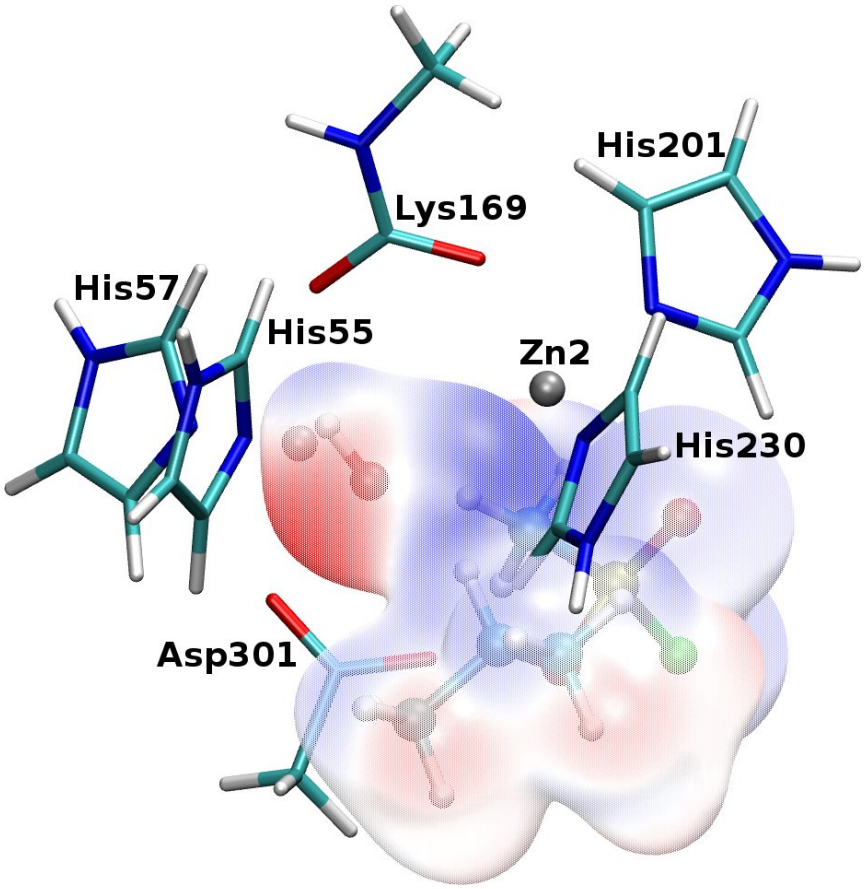
Some of the recently reported studies of the PTE-catalyzed hydrolysis of paraoxon [250, 259] has shown no changes in hydroxide-zinc ions interactions upon formation of a pentacoordinate intermediate. However, this is not the only likely scenario, as Gao et al. [258] proposed a single-step mechanism, whereby the paraoxon hydrolysis proceeds via the transition state involving a hydroxide interacting with one zinc ion only. Noticeably, the resulting enzyme-product complex includes additional Zn1-bound water molecule, while hydrogen bonding pat-

tern involving this water molecule, phosphorus-bound hydroxyl group, and Asp301 side chain [258] is very similar to that found herein for the enzyme-intermediate structure (Figure 3.28).

In the case of "77 atoms" model, transition state structure immediately preceding the pentavalent intermediate was successfully determined (Figure 3.28). However, the entire reaction pathway leading from reactant to this intermediate seems to involve other transition state(s), as it is unlikely that rearrangement in zinc coordination required to convert substrate to transition state occurs in a single step. The energy of this particular transition state geometry is slightly higher than that of subsequent intermediate only when electronic energy difference is considered (Table 3.11). Nonetheless, the intermediate energy relative to reactants is about 3 times higher in the case of "74 atoms" model (Table 3.11), which seems to imply that the catalytic mechanism involving water molecule is favored. Another advantage of such a reaction scenario involves straightforward restitution of the active site, i.e., Zn1-bound water molecule additionally hydrogen bonded to aspartate could be deprotonated by the latter and become a bridging hydroxide serving as a nucleophile in another reaction cycle. While the remaining stationary points along a reaction coordinate are yet to be found, these preliminary results are consistent with the generally accepted catalytic mechanism of PTE, whereby the nucleophilic attack of a bridging hydroxide triggers the hydrolysis of Zn2-bound substrate molecule.

Catalytic field for sarin hydrolysis Catalytic fields for the first step of alkaline hydrolysis of sarin (i.e., intermediate formation) were superimposed with the structure of PTE active site (Figure 3.29). Interestingly, only the catalytic field associated with **B** path seems to fit into the enzyme active site, as optimal electrostatic environment for **A** path implies the presence of negative charges on both sides of the attacking hydroxide (see *side view* in Figure 3.14). Negative and positive charges on the opposite sides of a nucleophilic hydroxide proposed by **B** path catalytic fields coincide with the positions of Asp301 residue and Zn2 zinc ion. In addition to likely preference for **B** path, the agreement between electrostatic pattern of PTE active site and the one predicted by catalytic field suggests that the corresponding active site constituents (i.e., Asp301 and Zn2) are catalytically important as they are involved in transition state stabilization.

Figure 3.29: Catalytic fields for the alkaline sarin hydrolysis (**B** path; see Figure 3.14) superimposed with the positions of PTE active site residues. The electronic isodensity surface of 0.01 a.u. is coloured according to the differential electrostatic potential of transition state and reactants calculated at the HF/6-31+G(d) level of theory. The sign of the differential potential is inverted to reflect the electrostatic properties of a complementary molecular environment. *Red (blue)* colour denotes regions where a negative (positive) charge would be optimal for catalytic activity.



3.3 Intermediate Binding in Triosephosphate Isomerase Catalysis

Triosephosphate isomerase (TIM, E.C. 5.3.1.1) is a highly efficient glycolytic enzyme performing the conversion of dihydroxyacetone phosphate to glyceraldehyde-3-phosphate [288]. Due to the considerable reaction rate enhancement of 10^9 times relative to that in solution [289], TIM has served as an example of the extreme power of an enzyme catalyst [290]. Accordingly, a hypothesis about covalent or partially covalent catalysis in enzymes with remarkable proficiencies [36] included TIM as a possible case where the binding of intermediate and/or transition state may exhibit some degree of covalency.

The first and the last steps of TIM-catalyzed reaction involve formation of an enediolate, EDT1, via proton abstraction by a catalytic base, Glu^-165 residue, and conversion of another enediolate, EDT2 (differing by the position of a hydroxyl group), to product by back donation of proton from Glu165 residue. Despite a number of studies, uncertainty remains as to the actual mechanism of the middle step [37, 291, 292], which might encompass either the intramolecular proton transfer (see reaction course marked by *black arrow* in Figure 3.30) or yet another series of proton donation and abstraction steps involving an electrophilic TIM residue (Glu165 or His95).

While general acid-base catalysis exploited by TIM seems to fulfill the covalency requirement put forward by Zhang and Houk [36], it would be interesting to check whether there is any degree of covalency in binding of the reaction intermediate and/or transition state. The transition state associated with intramolecular second proton transfer is particularly well-suited for this purpose, as it does not engage any enzyme residue directly. However, such a reaction course is less likely than the scenario assuming His95-assisted EDT1 to EDT2 conversion [37]. Since the structure of TIM-EDT1 complex [37] was kindly provided by Dr. Caterina Ghio, this preliminary study of the covalent nature of interactions taking place in an enzyme active site was focused on the binding of the first intermediate of TIM-catalyzed reaction. Selected hydrogen bonds were analyzed in terms of their geometry, energetics, and topological characteristics of electron density.

Technical details The structure of TIM-EDT1 complex [37], courtesy of Dr. C. Ghio, was employed in analysis of the possible covalent aspects of enzyme-intermediate binding. Based on the distance to EDT1, 6 molecular fragments were selected that make hydrogen bonds to intermediate (see Figure 3.30). In particular, imidazole, propionic acid, ethylamine and acetamide represented His95,

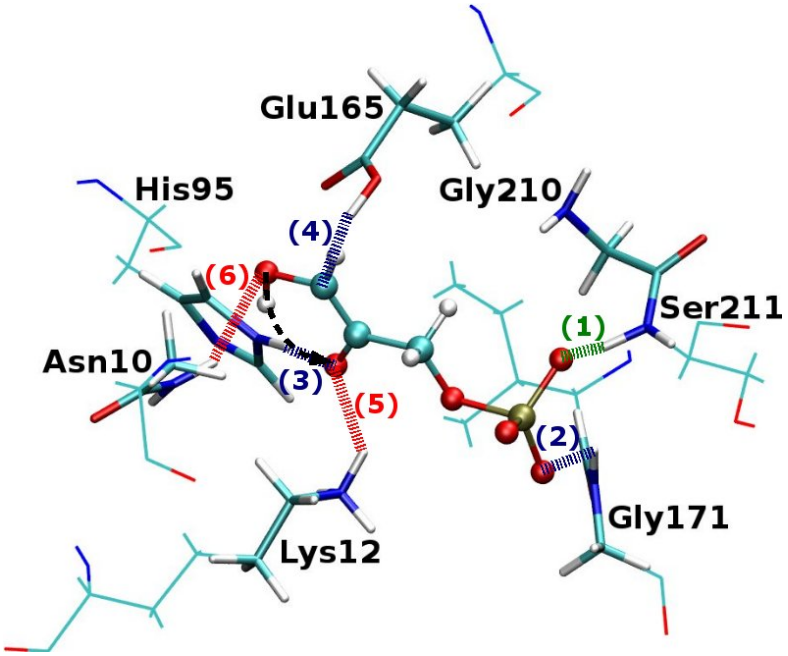


Figure 3.30: Structure of the TIM active site model in complex with EDT1 intermediate (*ball-and-stick representation*). Molecular fragments shown in *stick representation* were included in analysis of the nature of hydrogen bonds marked by *dashed lines*. The *colour scheme* applied in the latter is consistent with the classification of hydrogen bonding interactions given in Figures 1.4 and 3.31.

Glu165, Lys^+12, and Asn10 residues, respectively. Two hydrogen bonds involving the EDT1 phosphate tail are formed with Ser211 and Gly171 main chain amino groups. To mimic these interactions, Ser211 amino group was treated simultaneously with the preceding Gly210 residue, whereas Gly171 Cα and amino groups were capped with carbonyl moiety of the neighbouring Ile170. Except for positively charged Lys^+12 residue, all the remaining molecular fragments are neutral. EDT1 intermediate bears a charge of -3.

Interaction energy was calculated with 6-31G(d) basis set in a pairwise manner, i.e., within 6 dimers of EDT1 and respective residues/fragments. Variation-perturbation decomposition of binding energy [26] was carried out as implemented in the modified version of Gamess–US program [170]. The AIM theory of Bader [147] was employed in topological analysis of the electron density distribution based on the MP2/6-31G(d) wavefunction. Location and characterization of the critical points was performed with EXT94B module of Aimpac code [293].

3.3.1 Properties of TIM active site hydrogen bonds

Table 3.12: Properties of TIM-EDT1 intermediate hydrogen bonding interactions. Topological parameters and interaction energies are given in a.u. and kcal · mol^{-1}, respectively.

	Residue-EDT1 dimer					
Property	1[a] (Ser211)	2 (Gly171)	3 (His95)	4 (Glu165)	5 (Lys12)	6 (Asn10)
$R_{XH\cdots Y}$[b]	1.189	1.375	1.520	1.746	2.023	2.273
ρ_C[c]	0.166	0.118	0.073	0.052	0.026	0.015
$\nabla^2\rho_C$[d]	0.079	0.622	0.218	0.100	0.070	0.051
H_C[e]	−0.138	−0.019	−0.010	−0.011	−0.002	0.0002
ratio[f]	0.70	0.46	0.51	0.69	0.08	0.30
ΔE_{MP2}	−18.49	46.77	−29.08	−11.21	−235.31	−23.55
$\Delta E^{(2)}_{CORR}$	−2.01	−9.71	−3.91	−5.36	−3.16	−0.37
ΔE_{SCF}	−16.48	56.49	−25.17	−5.85	−232.15	−23.18
$\Delta E^{(R)}_{DEL}$	−73.70	−60.76	−28.21	−19.17	−18.01	−6.55
$\Delta E^{(1)}$	57.23	117.24	3.04	13.32	−214.14	−16.64
$\Delta E^{(1)}_{EX}$	162.45	248.63	58.13	41.14	12.67	5.51
$\Delta E^{(1)}_{EL}$	−105.22	−131.38	−55.09	−27.82	−226.81	−22.15

[a]Numbering of complexes according to designation introduced in Figure 3.30
[b]Hydrogen bond proton-acceptor distance in Å
[c]Electron density at the BCP
[d]Laplacian of the electron density at the BCP
[e]Local electron energy density at the BCP
[f]$\Delta E^{(R)}_{DEL}/\Delta E^{(1)}_{EL}$

As described in section 1.3.3, the degree of covalency of hydrogen bonding interaction can be assessed based on the values of Laplacian of the electron density, $\nabla^2\rho_C$, and local electron energy density, H_C, in a given proton-acceptor bond critical point (BCP). Covalent bonds, implying the concentration of electronic density in an interatomic region, exhibit negative values of both $\nabla^2\rho_C$ and H_C.

Positive Laplacian coupled with $H_C < 0$ is presumed to indicate a partially covalent nature of interaction [148]. This classification appears to be reflected in the ratio of delocalization to electrostatic components of interaction energy, i.e., $\Delta E^{(R)}_{DEL}/\Delta E^{(1)}_{EL}$ [142]. In particular, covalent or partially covalent hydrogen bonds ($H_C < 0$) are accompanied by the value of $\Delta E^{(R)}_{DEL}/\Delta E^{(1)}_{EL}$ greater than ~ 0.45.

According to the properties of 6 hydrogen bonds studied herein (Table 3.12), 4 of them seem to be partially covalent, since the $\Delta E^{(R)}_{DEL}/\Delta E^{(1)}_{EL}$ ratio exceeds 0.45. Remarkably, this finding is confirmed by the negative H_C values. While the shortest hydrogen bond present in Ser211-EDT1 complex ($R_{NH...O} = 1.189$ Å) falls into the region associated with covalent bonds (Figure 3.31), its possible covalency is not confirmed by negative $\nabla^2\rho_C$. However, considerable value of electronic density in the corresponding BCP, ρ_C= 0.166 a.u., is characteristic of covalent bonds. It is then unclear, whether this inconsistency results from some intrinsic features of the system under study or it is due to application of relatively small basis set without diffuse functions.

A distinctive feature of Lys12-EDT1 complex is the interaction of two oppositely charged monomers, one of them bearing a charge as large as -3. This is reflected in the large value of interaction energy, a major part of which encompasses first-order electrostatic term (Table 3.12). Despite the hydrogen bond proton-acceptor distance ($R_{NH...O} = 2.023$ Å) ruling out any degree of covalency, H_C value is negative. Again, this might result from unique features of the Lys12-EDT1 complex or be simply an artifact arising from basis set inadequacy.

Comparison of the topological and energetic results describing hydrogen bonds considered herein (Table 3.12) reveals lack of correlation between increasing length of a hydrogen bond and decreasing binding strength. While the former is closely followed by lowering of the electronic density in a given BCP, the ordering of complexes based on the interaction energy is inconsistent with their intermolecular separation or ρ_C value. Based on the positive value of binding energy, Gly171-EDT1 complex, suggested to be partially covalent (Figure 3.31) does not seem to occur at all. Repulsive interaction observed here results from too short intermonomer distance—nitrogen and oxygen atoms belonging to different monomers are separated by $R_{O...N} = 1.802$ Å. Significant contribution of delocalization effect is then artificial and the two monomers appear to be too close to each other, probably as a result of molecular mechanics origin of the structure [37]. Finally, apparent inconsistency concerning the binding strength versus the proton-acceptor distance and BCP characteristics could result from the arbitrary division of an active site into molecular fragments interacting separately with EDT1 intermediate. Still, this preliminary results seem to confirm that TIM-EDT1 interaction involves hydrogen bonds of partially covalent nature.

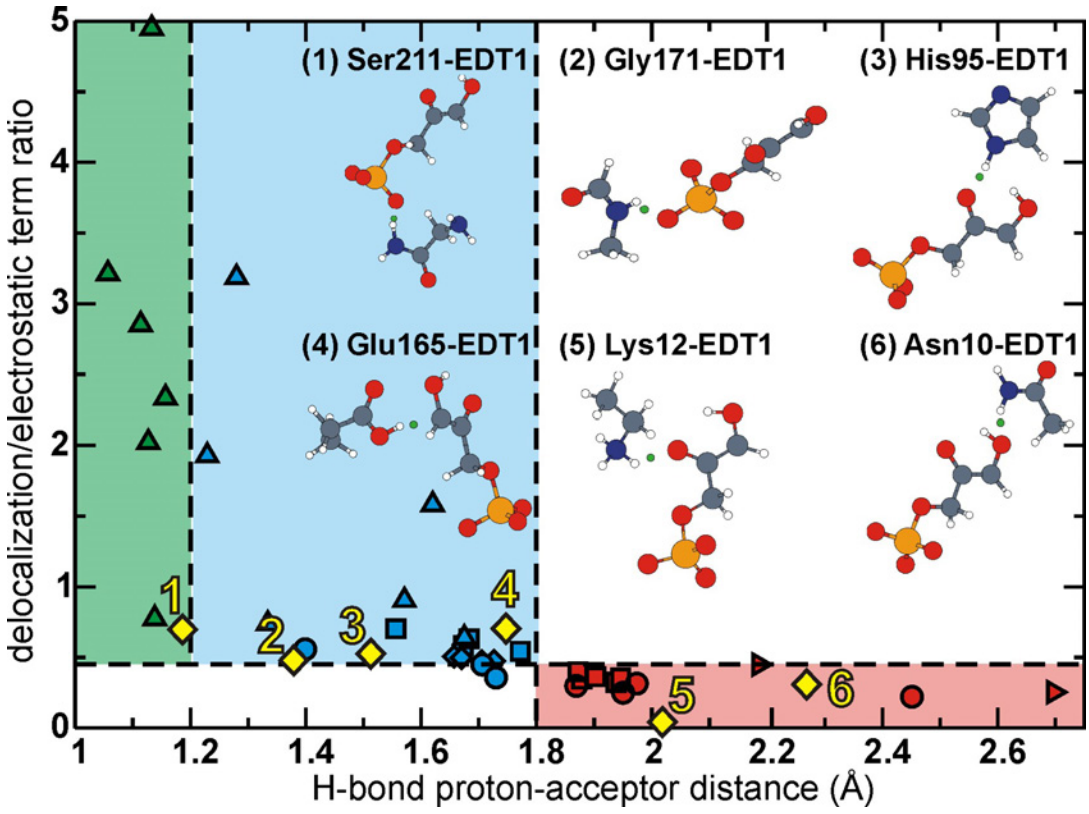


Figure 3.31: Classification of hydrogen bonding interactions including 6 points corresponding to the intermediate binding in TIM active site (marked as *yellow diamonds*). The numbering of these points corresponds to structures of dimers inserted into the plot. The location of proton-acceptor bond critical points within each dimer is shown as *green ball*. The remaining points in the plot refer to 34 hydrogen bonded complexes studied in Ref. [142].

CHAPTER 4

Summary

This work was intended to provide insight into the fundamental phenomena responsible for enzyme-ligand interactions. As one of the first such contributions in the literature, it was focused on the nature of intermolecular forces that govern the recognition of ligands within a binding pocket of an enzyme or enhance a particular reaction performed by an enzyme. The main tool to study the intermolecular interactions was the nonempirical variation-perturbation scheme of binding energy partitioning. Catalytic activity of enzymes was analyzed in the framework of Differential Transition State Stabilization (DTSS) methodology, which provides the clear connection between the binding energy and the catalytic function. In addition to these methods, a variety of computational chemistry techniques was applied in preparation of the structures of enzyme-ligand complexes (e.g., docking, molecular dynamics, QM and QM/MM calculations). The key conclusions of this work are presented herein.

Enzyme inhibition

A detailed, novel in the existing literature, nonempirical analysis of the physical nature of interactions taking place in an enzyme binding pocket was performed for potato phenylalanine ammonia-lyase (PAL), parsley PAL and human urokinase-type plasminogen activator (uPA) inhibitors (Chapter 2).

General conclusions The following conclusions regarding the intermolecular interactions-based study of enzyme inhibitory effects can be drawn upon these results:

- Binding energy within three sets of enzyme-inhibitor complexes shows **correlation with experimental inhibitory activity** (i.e., $-\log K_i$ or $-\log IC_{50}$). Quantitative agreement obtained for potato PAL inhibitors features r^2 and standard error of estimate (SEE) values of 0.99 and $1.48\,\mathrm{kcal}\cdot\mathrm{mol}^{-1}$ (at the MP2 level of theory). Weaker correlation found in the case of parsley PAL inhibitors might result from considering a larger and more variable set of inhibitors bound in two different arrangements as well as a lack of post-docking refinement of receptor-ligand complexes. Similarly, structural flaws due to force field based optimization of uPA-inhibitor complexes could explain worse performance of inhibitory activity models obtained in this case. Nonetheless, **these results indicate that solvation and entropic contributions are relatively constant across the three series considered herein and valid binding affinity prediction could be made based on the analysis of interaction energy only.**

- In general, **the higher the level of theory applied to the description of intermolecular interactions, the greater the degree of correlation with experimental binding affinity.** This observation does not seem to apply to the cases, where inaccurate structures of enzyme-inhibitor complexes were employed (e.g., uPA complexes), as derived from force field calculations. Even in such a case, fortunate cancellation of errors in short range terms might result in a reasonable description of the inhibitory activity despite an application of a more approximate electrostatic model. Additionally, it appears that limiting the size of an active site representation could be a remedy to some structural flaws of receptor-ligand complexes, as the noise introduced by residues contributing little to the overall interaction is probably larger than their actual involvement in inhibitor binding.

- *Ab initio* results indicate that some force field optimized uPA-inhibitor structures contained shortened intermolecular contacts due to the deficiencies of empirical parametrization (section 2.2). This resulted in a poor correlation of the ΔE_{SCF} and $\Delta E^{(1)}$ levels of theory containing exaggerated exchange and delocalization components. **As these terms are of opposite sign, they tend to cancel each other to a significant degree explaining a remarkably reasonable correlation of the electrostatic term $\Delta E_{EL}^{(1)}$ with experimental results.**

- **Monitoring of the drop in correlation when omitting a given receptor site residue(s) in the inhibitory activity model allows for determination of residues important for ligand specificity.** As a result, the minimal active site representation can be constructed leading to the more computationally tractable models. For example, accounting for the two of a total of nine uPA active site residues results in a remarkable agreement with experimental data.

- Systematic analysis of both the physical nature of interactions and binding contribution of receptor site residues allows for rational derivation of **simple yet sufficiently accurate models of inhibitory activity** related, for example, to the molecular electrostatic potential of inhibitors in several selected contact points. **Such models, derived from the first principles of quantum mechanics (in contrast to QSAR approach), provide rapid and straightforward estimation of binding affinity without using any empirical parametrization.**

Parsley PAL inhibitors The results of the docking study of parsley PAL inhibitors (section 2.1.4) can be summarized as follows:

- The binding of 11 parsley PAL inhibitors studied herein is directed by an ionic pair involving the phosphonate (carboxylate) group of inhibitors and the side chain of Arg354 residue. Depending on the size and flexibility of a remaining part of a ligand molecule, two different arrangements of an inhibitor aromatic ring are possible. Complementarity with a more polar region of parsley PAL active site (i.e., residues Gly115, Gln348, Pro385, and His396) appears to increase the binding affinity. Of the two possible conformers of 2-aminoindane-2-phosphonic acid derivatives, the one with axial phosphonic group seems to be favored as it binds within this particular region of a receptor site, resulting in a stronger enzyme-inhibitor interaction.

Enzymatic catalysis

DTSS methodology supplemented by catalytic fields approach was employed in the investigation of phosphorylation process catalyzed by 4-methyl-5-β-hydroxyethylthiazole kinase (ThiK) and cAMP-dependent protein kinase (PKA). With the aim of further application of an analogous approach to the study of organophosphate hydrolysis catalyzed by phosphotriesterase (PTE), the attempts at modeling

of PTE catalytic mechanism were also undertaken here. To address the Zhang and Houk [36] hypothesis concerning covalent catalysis possibly exploited by proficient enzymes, the last part of this thesis focused on the nature of a reaction intermediate binding within triosephosphate isomerase (TIM) active site.

General conclusions The main findings regarding the intermolecular interactions-based study of enzyme catalytic effects (section 3.1) are the following:

- Due to the minor role of electron correlation and mutual cancellation of delocalization and exchange effects, **interaction energy within ThiK and PKA active sites is mainly electrostatic in nature.** This outcome further confirms the hypothesis about electrostatics playing a major role in stabilization of the enzymatic transition states [21, 98, 99]. Accordingly, computationally inexpensive electrostatic interaction energy could be employed in the assessment of a possible catalytic contribution arising from the presence of a given residue.

- **DTSS analysis reveals the role of active site residues in enzyme catalytic activity.** Of particular importance for differential transition state stabilization are Glu126 and Cys198 residues as well as Mg2 magnesium ion of ThiK. PKA active site components with the most significant contribution to the lowering of activation energy barrier include both magnesium ions as well as Lys72 and Asp166 residues. Despite moderate inhibitory effects due to the presence of PKA Ser53 residue, the total DTSS energy of residues building the glycine-rich loop is negative confirming its involvement in transition state stabilization. Noticeably, **identity of enzyme residues exhibiting outstanding catalytic or inhibitory effects matches their sequence and structure conservation.**

- Destabilization of the transition state resulting from the presence of a particular residue suggests a structural role of the latter (e.g., Lys168 residue of PKA presumably involved in an optimal spatial alignment of the reactants or magnesium ligands engaged in metal ion binding). Another likely explanation involves the pairwise fashion of interaction energy calculation that entirely neglects any possible many-body contributions.

- **Analysis of the electronic charge redistribution along a reaction pathway allows the ideal catalytic environment to be derived from the first principles of quantum mechanics.** The comparison between spatial arrangement of enzyme active site components and the catalytic

fields representing the electrostatic characteristics of an optimal catalyst provides a qualitative measure of complementarity of both charge distributions achieved in an evolutionary way. More importantly, **such an analysis could be employed in the prediction of an influence of mutation allowing for the rational catalyst design.** Accordingly, catalytic fields for ThiK-catalyzed reaction correctly predict increased catalytic activity upon Cys198Asp mutation.

ThiK-catalyzed reaction The results concerning ThiK catalytic mechanism lead to the following conclusions:

- ThiK-catalyzed phosphoryl transfer reaction appears to proceed according to an S_N2-like mechanism. The latter involves a nucleophilic attack of the substrate hydroxyl group on the phosphorus atom followed by a single transition state with the hydroxyl proton already transferred to the γ-phosphate oxygen atom. No direct involvement of Cys198 residue was confirmed.

PTE-catalyzed hydrolysis of organophosphates The results involving gas phase alkaline hydrolysis of PTE substrates as well as preliminary findings regarding PTE catalytic mechanism (section 3.2) can be summarized in a following way:

- While all base-catalyzed hydrolysis reactions studied herein appear to follow an associative mechanism, the cleavage of P-O and P-S bonds occurs according to a one-step direct-displacement mechanism involving the presence of a single S_N2-like transition state, whereas the hydrolysis of P-F and P-CN bonds is consistent with an addition-elimination scenario employing several trigonal bipyramidal intermediates. Two alternative reaction pathways are possible for each of these mechanisms that differ in a position of the attacking hydroxide relative to the phosphoryl oxygen atom. Apparently, the most energetically favorable reaction coordinate involves the hydroxide proton being stabilized by a phosphoryl oxygen. As suggested by catalytic fields, the mechanism involving opposite placement of hydroxide appears to be exploited inside the PTE active site. The rate-limiting step of multistep mechanisms appears to be associated with an intermediate formation.
- Reliable model of PTE-sarin complex was developed that is consistent with the current hypotheses regarding substrate binding and the following catalytic events.

- PTE active site is well organized by the stabilizing interactions between zinc ions and their ligands. Bridging hydroxide ion is hydrogen bonded to Asp301 carboxylate group, which is crucial for proper positioning of hydroxide with respect to phosphorus atom of substrate.

- Although isopropyl group of sarin molecule is able to acquire a variety of positions within large group binding pocket of PTE, sarin itself is anchored by strong interaction between more exposed zinc ion and phosphoryl oxygen atom. Small group binding pocket accommodates methyl substituent, whereas fluorine atom is directed toward the leaving group portion of PTE.

- In agreement with both MD results and available X-ray data, QM/MM and QM models of PTE-sarin complex support the generally accepted catalytic mechanism of PTE. The latter might involve additional water molecule that becomes zinc ligand in the course of a reaction and, possibly, aids the active site restitution.

- The comparison between catalytic fields derived from gas phase mechanism of sarin hydrolysis and the arrangement of PTE active site residues suggests that Asp301 residue and Zn2 zinc ion are catalytically important, as their formal charge is in agreement with the electrostatic properties of an optimal catalytic environment predicted by the catalytic fields.

The nature of intermediate binding in TIM catalysis The binding of enediolate (EDT1) intermediate of TIM-catalyzed reaction involves 6 hydrogen bonds. *Ab initio* analysis of the physical nature of these hydrogen bonds (section 3.3) leads to the following results:

- Topological and energetic properties of the hydrogen bonds anchoring EDT1 in the TIM active site suggest that 4 hydrogen bonds are partially covalent in nature. This finding is supported by the value of $\Delta E^{(R)}_{DEL}/\Delta E^{(1)}_{EL}$ ratio greater than ~ 0.45 as well as the negative value of electron energy density, H_C, in a given proton-acceptor bond critical point. While more comprehensive investigation encompassing other enzymes as well as transition state binding is required to confirm the hypothesis of Zhang and Houk [36], it appears that at least partially covalent interactions with reaction intermediates and/or transition states might account for the catalytic advantage of the most powerful enzymes.

LIST OF FIGURES

1.1 Variation-perturbation scheme of interaction energy decomposition . 22

1.2 Energy diagram illustrating the DTSS approach 23

1.3 Derivation and interpretation of the catalytic fields 24

1.4 Classification of the hydrogen bonding interactions 25

2.1 PAL-catalyzed reaction. 29

2.2 *(S)*-phenylalanine analogs known as PAL inhibitors 30

2.3 Structures of potato PAL inhibitors 31

2.4 Potato PAL active site model in complex with inhibitor **2** 31

2.5 Structures of parsley PAL inhibitors 33

2.6 Parsley PAL active site model in complex with inhibitor **II-AC** . . 36

2.7 Binding poses of parsley PAL inhibitors 37

2.8 Selected parsley PAL residues in complex with inhibitors **I** and **II-AC** 38

2.9 Goldscore function against the experimental inhibitory activity of parsley PAL inhibitors 39

2.10 Total interaction energy of potato PAL inhibitors 41

2.11 Binding contribution of particular potato PAL active site residues . 44

2.12 Electrostatic binding energy as a function of experimental inhibitory activity (potato PAL inhibitors) 45

2.13 Diagram of the electrostatic lock-and-key model 46

2.14 Total interaction energy at various levels of theory versus reference MP2 binding energy 48

2.15 Total interaction energy of parsley PAL inhibitors 50

2.16 Predicted versus experimental inhibitory activity of parsley PAL inhibitors 52

2.17 Binding contribution of particular parsley PAL active site residues . 54

2.18 Parsley PAL active site atoms selected to mimic the electrostatic properties of a binding pocket 56

2.19 Electrostatic binding energy as a function of experimental inhibitory activity (parsley PAL inhibitors) 56

2.20 Structures of α-aminoalkylphosphonate diphenyl esters known as urokinase inhibitors 59

2.21 Urokinase active site model in complex with inhibitor **2** 60

2.22 Total interaction energy of uPA inhibitors 62

2.23 Electrostatic binding energy as a function of experimental inhibitory activity (uPA inhibitors) 64

2.24 Binding energy terms of the interaction between uPA inhibitor **3** and Ser190 residue 65

3.1 Multiple sequence alignment of hydroxyethylthiazole kinases 68

3.2 B3LYP/6-31G(d):PM3 geometries of reactants, transition state and products 71

3.3 Structure of the ThiK active site model in complex with transition state 75

3.4 The components of differential transition state stabilization energy (ThiK-catalyzed reaction) 76

3.5 Catalytic fields for the ThiK-catalyzed reaction 77

3.6 Structure of the PKA active site model in complex with transition state 80

3.7 The components of differential transition state stabilization energy (PKA-catalyzed reaction) 82

3.8 Catalytic fields for the PKA-catalyzed reaction 86

3.9 Scheme of alkaline hydrolysis of PTE substrates 89

3.10 HF/6-31+G(d) geometries of the stationary points along a reaction coordinate for the alkaline hydrolysis of sarin 91

3.11 Energy profile for the first step of alkaline hydrolysis of sarin . . . 92

3.12 Energy profile for the last step of alkaline hydrolysis of sarin 93

3.13 Potential energy surface along a reaction coordinate for the hydrolysis of sarin 94

3.14 Catalytic fields for intermediate formation step of alkaline sarin hydrolysis 100

3.15 Approximate catalytic fields for intermediate formation step of alkaline sarin hydrolysis 101

3.16 RMSD of His254 residue and heavy atoms of PTE active site . . . 107

3.17 Assignment of sarin atom types and partial atomic charges 107

3.18 Vibrational frequency analysis of sarin molecule 109

3.19 Comparison of QM and force field-optimized sarin geometries 109

3.20 Potential energy as a function of F1-P-CT3-HA dihedral angle . . . 110

3.21 Potential energy as a function of CT3-P-ON2-CN7 dihedral angle . 111

3.22 RMSD of the PTE active site and sarin heavy atoms, and variation in selected distances 112

3.23 Deviation of active site residues in the optimized final MD snapshot relative to X-ray structure 113

3.24 Coordination pattern of zinc ions 114

3.25 Mobility of PTE residues 115

3.26 B3LYP/6-31G(d)//Charmm27 geometry of the PTE-substrate complex 118

3.27 B3LYP/6-31G(d)//Charmm27 geometry of the PTE-intermediate complex 119

3.28 B3LYP/6-31G(d) structures of the PTE active site in complex with substrate, transition state and pentacoordinate intermediate 121

3.29 Catalytic fields for the alkaline sarin hydrolysis superimposed with the PTE active site 124

3.30 Structure of the TIM active site model in complex with EDT1 intermediate 126

3.31 Classification of hydrogen bonding interactions including 6 points corresponding to the intermediate binding in TIM active site 129

LIST OF TABLES

2.1 Total binding energy of potato PAL inhibitors 40

2.2 Performance of the interaction energy-based models inhibitory activity (potato PAL inhibitors) 42

2.3 Comparison of experimental and predicted inhibition constant values 43

2.4 Coefficients of determination obtained in the absence of individual potato PAL residues 45

2.5 Total binding energy of parsley PAL inhibitors 47

2.6 Performance of the interaction energy-based models of inhibitory activity (parsley PAL inhibitors) 51

2.7 Coefficients of determination obtained in the absence of individual parsley PAL residues 53

2.8 Performance of the inhibitory activity models (limited size representation of parsley PAL active site) 55

2.9 Total binding energy of uPA inhibitors 61

2.10 Performance of the interaction energy-based models of inhibitory activity (uPA inhibitors) 63

3.1 B3LYP/6-31G(d):PM3 relative energies of the transition state and product (ThiK-catalyzed reaction) 72

3.2 Differential transition state stabilization energy (ThiK-catalyzed reaction) 74

3.3 Differential transition state stabilization energy (PKA-catalyzed reaction) 81

3.4 The kinetic parameters for the hydrolysis of PTE substrates 88

3.5 Relative energies for the multistep hydrolysis of sarin 95

3.6 Multistep hydrolysis: the activation energy barriers 96

3.7 Single-step hydrolysis: the activation energy barriers 97

3.8 Harmonic force constants acting within particular buffer regions . . 105

3.9 *Ab initio* and force field-optimized sarin-water molecule complexes . 108

3.10 Selected distances in the reactant, transition state and intermediate structures of the PTE-catalyzed sarin hydrolysis 120

3.11 B3LYP/6-31G(d) relative energies of the transition state and intermediate (PTE-catalyzed reaction) 122

3.12 Properties of TIM-EDT1 intermediate interactions 127

BIBLIOGRAPHY

[1] Houk, K. N.; Leach, A. G.; Kim, S. P.; Zhang, X. Binding affinities of host-guest, protein-ligand, and protein-transition-state complexes. *Angew. Chem. Int. Ed.*, **2003**, *42*(40), 4872–4897.

[2] Gohlke, H.; Klebe, G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew. Chem. Int. Ed.*, **2002**, *41*(15), 2644–2676.

[3] Guerrero, A.; Rosell, G. Biorational approaches for insect control by enzymatic inhibition. *Curr. Med. Chem.*, **2005**, *12*(4), 461–469.

[4] Smith, A. J. T.; Zhang, X.; Leach, A. G.; Houk, K. N. Beyond picomolar affinities: Quantitative aspects of noncovalent and covalent binding of drugs to proteins. *J. Med. Chem.*, **2009**, *52*(2), 225–233.

[5] Congreve, M.; Murray, C. W.; Blundell, T. L. Structural biology and drug discovery. *Drug Discovery Today*, **2005**, *10*(13), 895–907.

[6] de Azevedo Jr., W. F.; Dias, R. Computational methods for calculation of ligand-binding affinity. *Curr. Drug Targets*, **2008**, *9*(12), 1031–1039.

[7] Lazaridis, T. Binding affinity and specificity from computational studies. *Curr. Org. Chem.*, **2002**, *6*(14), 1319–1332.

[8] Böhm, H.-J.; Klebe, G. What can we learn from molecular recognition in protein-ligand complexes for the design of new drugs? *Angew. Chem. Int. Ed.*, **1996**, *35*(22), 2588–2614.

[9] Holloway, M. K. A priori prediction of ligand affinity by energy minimization. *Perspect. Drug DiscoVery Des.*, **1998**, *9–11*, 63–84.

[10] Morreale, A.; Maseras, F.; Iriepa, I.; Gálvez, E. Ligand-receptor interaction at the neural nicotinic acetylcholine binding site: A theoretical model. *J. Mol. Graphics Modell.*, **2002**, *21*(2), 111–118.

[11] Náray-Szabó, G. Quantum chemical calculation of the enzyme-ligand interaction energy for trypsin inhibition by benzamidines. *J. Am. Chem. Soc.*, **1984**, *106*(16), 4584–4589.

[12] Raha, K.; van der Vaart, A. J.; Riley, K. E.; Peters, M. B.; Westerhoff, L. M.; Kim, H.; Merz, Jr., K. M. Pairwise decomposition of residue interaction energies using semiempirical quantum mechanical methods in studies of protein-ligand interaction. *J. Am. Chem. Soc.*, **2005**, *127*(18), 6583–6594.

[13] Grembecka, J.; Kędzierski, P.; Sokalski, W. A. Non-empirical analysis of the nature of the inhibitor-active-site interactions in leucine aminopeptidase. *Chem. Phys. Lett.*, **1999**, *313*(1–2), 385–392.

[14] Grembecka, J.; Sokalski, W. A.; Kafarski, P. Quantum chemical analysis of the interactions of transition state analogs with leucine aminopeptidase. *Int. J. Quantum Chem.*, **2001**, *84*(2), 302–310.

[15] Fukuzawa, K.; Kitaura, K.; Uebayasi, M.; Nakata, K.; Kaminuma, T.; Nakano, T. Ab initio quantum mechanical study of the binding energies of human estrogen receptor with its ligands: An application of fragment molecular orbital method. *J. Comp. Chem.*, **2004**, *26*(1), 1–10.

[16] Borman, S. Much ado about enzyme mechanisms. *Chem. Eng. News*, **2004**, *82*(8), 35–39.

[17] Gao, J.; Ma, S.; Major, D. T.; Nam, K.; Pu, J.; Truhlar, D. G. Mechanisms and free energies of enzymatic reactions. *Chem. Rev.*, **2006**, *106*(8), 3188–3209.

[18] Schramm, V. L. Enzymatic transition state theory and transition state analogue design. *J. Biol. Chem.*, **2007**, *282*(39), 28297–28300.

[19] Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H. M. Electrostatic basis for enzyme catalysis. *Chem. Rev.*, **2006**, *106*(8), 3210–3235.

[20] Martí, S.; Maite Roca, J. A.; Moliner, V.; Silla, E.; Tuñón, I.; Bertrán, J. Theoretical insights in enzyme catalysis. *Chem. Soc. Rev.*, **2004**, *33*(2), 98–107.

[21] Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G. How enzymes work: Analysis by modern rate theory and computer simulations. *Science*, **2004**, *303*(5655), 186–195.

[22] Jiang, L.; Althoff, E. A.; Clemente, F. R.; Doyle, L.; Röthlisberger, D.; Zanghellini, A.; Gallaher, J. L.; Betker, J. L.; Tanaka, F.; Barbas, III, C. F.; Hilvert, D.; Houk, K. N.; Stoddard, B. L.; Baker, D. De novo computational design of retro-aldol enzymes. *Science*, **2008**, *319*(5868), 1387–1391.

[23] Röthlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O.; Albeck, S.; Houk, K. N.; Tawfik, D. S.; Baker, D. Kemp elimination catalysts by computational enzyme design. *Nature*, **2008**, *453*(7192), 190–195.

[24] Ghirlanda, G. Computational biochemistry: Old enzymes, new tricks. *Nature*, **2008**, *453*(7192), 164–166.

[25] Martí, S.; Andrés, J.; Moliner, V.; Silla, E.; Tuñón, I.; Bertrán, J. Computational design of biological catalysts. *Chem. Soc. Rev.*, **2008**, *37*(12), 2634–2643.

[26] Sokalski, W. A.; Roszak, S.; Pecul, K. An efficient procedure for decomposition of the scf interaction energy into components with reduced basis set dependence. *Chem. Phys. Lett.*, **1988**, *153*(2,3), 153–159.

[27] Langer, B.; Langer, M.; Rétey, J. Methylidene-imidazolone (mio) from histidine and phenylalanine ammonia-lyase. *Adv. Protein Chem.*, **2001**, *58*, 175–214.

[28] Maier, L.; Diel, P. J. Synthesis, physical and biological properties of the phosphorus analogues of phenylalanine and related compounds. *Phosphorus Sulfur*, **1994**, *90*(1–4), 259–279.

[29] Amrhein, N.; Miziak, P.; Zoń, J. to be published.

[30] Sieńczyk, M. *Nowe inhibitory urokinazowego aktywatora plazminogenu o potencjalnych właściwościach antyangiogennych i przeciwnowotworowych*. PhD thesis, Wrocław University of Technology, 2006.

[31] Szefczyk, B.; Mulholland, A. J.; Ranaghan, K. E.; Sokalski, W. A. Differential transition-state stabilization in enzyme catalysis: Quantum chemical analysis of interactions in the chorismate mutase reaction and prediction of the optimal catalytic field. *J. Am. Chem. Soc.*, **2004**, *126*(49), 16148–16159.

[32] Sokalski, W. A. The physical nature of catalytic activity due to the molecular environment in terms of intermolecular interaction theory: derivation of simplified models. *J. Mol. Catalysis*, **1985**, *30*(3), 395–410.

[33] Cheng, Y.; Zhang, Y.; McCammon, J. A. How does the cAMP-dependent protein kinase catalyze the phosphorylation reaction: An ab initio QM/MM study. *J. Am. Chem. Soc.*, **2005**, *127*(5), 1553–1562.

[34] Szarek, P.; Dyguda-Kazimierowicz, E.; Tachibana, A.; Sokalski, W. A. Physical nature of intermolecular interactions within cAMP-dependent protein kinase active site: Differential transition state stabilization in phosphoryl transfer reaction. *J. Phys. Chem. B*, **2008**, *112*(37), 11819–11826.

[35] Hill, C. M.; Li, W.-S.; Thoden, J. B.; Holden, H. M.; Raushel, F. M. Enhanced degradation of chemical warfare agents through molecular engineering of the phosphotriesterase active site. *J. Am. Chem. Soc.*, **2003**, *125*(30), 8990–8991.

[36] Zhang, X.; Houk, K. N. Why enzymes are proficient catalysts: Beyond the pauling paradigm. *Acc. Chem. Res.*, **2005**, *38*(5), 379–385.

[37] Alagona, G.; Campanile, S.; Ghio, C.; Molin, D. Is the bias introduced in a FEP calculation by parameterizing a QM reaction acceptable? Comparison with Car-Parrinello MD/AMBER results for the second proton transfer in triosephosphate isomerase (TIM). *J. Mol. Struct. (THEOCHEM)*, **2005**, *729*(1–2 Spec. Iss.), 131–139.

[38] Szalewicz, K.; Jeziorski, B. in *Molecular Interactions: From van der Waals to Strongly Bound Complexes*, Scheiner, S., Ed., (John Wiley & Sons, New York). pp 3–43. 1997.

[39] Alagona, G.; Ghio, C. Ab initio investigation of the methylimidazole-indole complexes as models of the histidine-tryptophan pair. *J. Phys. Chem. A*, **1998**, *102*(30), 6152–6160.

[40] Bukowski, R.; Cencek, W.; Jankowski, P.; Jeziorski, B.; Jeziorska, M.; Kucharski, S. A.; Misquitta, A. J.; Moszynski, R.; Patkowski, K.; Rybak, S.;

Szalewicz, K.; Williams, H. L.; Wormer, P. E. *SAPT2002: An Ab Initio Program for Many-Body Symmetry-Adapted Perturbation Theory Calculations of Intermolecular Interaction Energies.* 2004.

[41] Sokalski, W. A.; Kędzierski, P.; Grembecka, J. Ab initio study of the physical nature of interactions between enzyme active site fragments in vacuo. *Phys. Chem. Chem. Phys.*, **2001**, *3*(5), 657–663.

[42] Náray-Szabó, G. Enzyme mechanisms: interplay of theory and experiment. *J. Mol. Struct. (THEOCHEM)*, **2000**, *500*(1–3), 157–167.

[43] Berg, J. M.; Tymoczko, J. L.; Stryer, L. *Biochemistry.* W. H. Freeman and Company, 2002.

[44] Bommarius, A. S.; Riebel, B. R. *Biocatalysis.* Wiley-VCH, 2004.

[45] Bencsura, A.; Enyedy, I.; Kovach, I. M. Origins and diversity of the aging reaction in phosphonate adducts of serine hydrolase enzymes: What characteristics of the active site do they probe? *Biochemistry*, **1995**, *34*(28), 8989–8999.

[46] Franz, J. E.; Mao, M. K.; Sikorski, J. A. *Glyphosate: A Unique Global Herbicide.* American Chemical Society, Washington, DC, 1997.

[47] Scapin, G. Structural biology and drug discovery. *Curr. Pharm. Des.*, **2006**, *12*(17), 2087–2097.

[48] Kim, R.; Skolnick, J. Assessment of programs for ligand binding affinity prediction. *J. Comput. Chem.*, **2008**, *29*(8), 1316–1331.

[49] Green, N. M. Avidin. *Adv. Protein Chem.*, **1975**, *29*, 85–133.

[50] Wiseman, T.; Williston, S.; Brandts, J. F.; Lin, L.-N. Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Analyt. Biochem.*, **1989**, *179*(1), 131–137.

[51] J. A, M. Theory of biomolecular recognition. *Curr. Opin. Struct. Biol.*, **1998**, *8*(2), 245–249.

[52] Cheng, Y.-C.; Prusoff, W. H. Relationship between the inhibition constant (KI) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. *Biochem. Pharmacol.*, **1973**, *22*(23), 3099–3108.

[53] Lill, M. A. Multi-dimensional qsar in drug discovery. *Drug Discov. Today*, **2007**, *12*(23–24), 1013–1017.

[54] Du, Q.-S.; Huang, R.-B.; Chou, K.-C. Recent advances in QSAR and their applications in predicting the activities of chemical molecules, peptides and proteins for drug design. *Curr. Protein Pept. Sci.*, **2008**, *9*(3), 248–259.

[55] Martin, Y. C. in *3D QSAR in Drug Design*, Kubinyi, H.; Folkers, G.; Martin, Y. C., Eds., (Springer, Dordrecht, The Netherlands). volume 3 of *Three-Dimensional Quantitative Structure Activity Relationships*. pp 3–23. 1998.

[56] Folkers, G.; Merz, A.; Rognan, D. in *3D QSAR in Drug Design*, Kubinyi, H., Ed., (Springer, Dordrecht, The Netherlands). volume 1 of *Three-Dimensional Quantitative Structure Activity Relationships*. pp 583–618. 1994.

[57] Jain, T.; Jayaram, B. An all atom energy based computational protocol for predicting binding affinities of protein-ligand complexes. *FEBS Lett.*, **2005**, *579*(29), 6659–6666.

[58] Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. The statistical-thermodynamic basis for computation of binding affinities: A critical review. *Biophys. J.*, **1997**, *72*(3), 1047–1069.

[59] Brandsdal, B. O.; Österberg, F.; Almlöf, M.; Feierberg, I.; Luzhkov, V. B.; Åqvist, J. Free energy calculations and ligand binding. *Adv. Protein Chem.*, **2003**, *66*, 123–158.

[60] Woo, H.-J.; Roux, B. Calculation of absolute protein-ligand binding free energy from computer simulations. *Proc. Nat. Acad. Sci. USA*, **2005**, *102*(19), 6825–6830.

[61] Reddy, M. R.; Erion, M. D. Relative binding affinities of fructose-1,6-bisphosphatase inhibitors calculated using a quantum mechanics-based free energy perturbation method. *J. Am. Chem. Soc.*, **2007**, *129*(30), 9296–9297.

[62] Dill, K. A. Additivity principles in biochemistry. *J. Biol. Chem.*, **1997**, *272*(2), 701–704.

[63] Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function. *J. Comp. Chem.*, **1998**, *19*(14), 1639–1662.

[64] Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.*, **1996**, *261*(3), 470–489.

[65] Jones, G.; Willett, P.; Glen, R. C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.*, **1995**, *245*(1), 43–53.

[66] Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.*, **1997**, *267*(3), 727–748.

[67] Gohlke, H.; Klebe, G. Statistical potentials and scoring functions applied to protein-ligand binding. *Curr. Opin. Struct. Biol.*, **2001**, *11*(2), 231–235.

[68] Sippl, M. J. Knowledge-based potentials for proteins. *Curr. Opin. Struct. Biol.*, **1995**, *5*(2), 229–235.

[69] Mitchell, J. B. O.; Laskowski, R. A.; Alex, A.; Forster, M. J.; Thornton, J. M. BLEEP - Potential of mean force describing protein-ligand interactions: II. calculation of binding energies and comparison with experimental data. *J. Comp. Chem.*, **1999**, *20*(11), 1177–1185.

[70] Åqvist, J.; Luzhkov, V. B.; Brandsdal, B. O. Ligand binding affinities from md simulations. *Acc. Chem. Res.*, **2002**, *35*(6), 358–365.

[71] Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, III, T. E. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.*, **2000**, *33*(12), 889–897.

[72] Wittayanarakul, K.; Hannongbua, S.; Feig, M. Accurate prediction of protonation state as a prerequisite for reliable MM-PB(GB)SA binding free energy calculations of HIV-1 protease inhibitors. *J. Comp. Chem.*, **2008**, *29*(5), 673–685.

[73] Carlsson, J.; Åqvist, J. Calculations of solute and solvent entropies from molecular dynamics simulations. *Phys. Chem. Chem. Phys.*, **2006**, *8*(46), 5385–5395.

[74] Chang, C. E. A.; Chen, W.; Gilson, M. K. Ligand configurational entropy and protein binding. *Proc. Nat. Acad. Sci. USA*, **2007**, *104*(5), 1534–1539.

[75] Irudayam, S. J.; Henchman, R. H. Entropic cost of protein-ligand binding and its dependence on the entropy in solution. *J. Phys. Chem. B*, **2009**, *113*(17), 5871–5884.

[76] Raha, K.; Merz, Jr., K. M. Large-scale validation of a quantum mechanics based scoring function: Predicting the binding affinity and the binding mode of a diverse set of protein-ligand complexes. *J. Med. Chem.*, **2005**, *48*(14), 4558–4575.

[77] Zhang, D. W.; Zhang, J. Z. H. Molecular fractionation with conjugate caps for full quantum mechanical calculation of proteinmolecule interaction energy. *J. Chem. Phys.*, **2003**, *119*(7), 3599–3605.

[78] Kitauraa, K.; Ikeoa, E.; Asadaa, T.; Nakanob, T.; Uebayasi, M. Fragment molecular orbital method: an approximate computational method for large molecules. *Chem. Phys. Lett.*, **1999**, *313*(3–4), 701–706.

[79] Zhang, D. W.; Xiang, Y.; Zhang, J. Z. H. New advance in computational chemistry: Full quantum mechanical ab initio computation of streptavidin-biotin interaction energy. *J. Phys. Chem. B*, **2003**, *107*(44), 12039–12041.

[80] Zhang, D. W.; Xiang, Y.; Gao, A. M.; Zhang, J. Z. H. Quantum mechanical map for protein-ligand binding with application to beta-trypsin/benzamidine complex. *J. Chem. Phys.*, **2004**, *120*(3), 1145–1148.

[81] Zhang, D. W.; Zhang, J. Z. H. Full quantum mechanical study of binding of HIV-1 protease drugs. *Int. J. Quant. Chem.*, **2005**, *103*(3), 246–257.

[82] Fukuzawa, K.; Mochizuki, Y.; Tanaka, S.; Kitaura, K.; Nakano, T. Molecular interactions between estrogen receptor and its ligand studied by the ab initio fragment molecular orbital method. *J. Phys. Chem. B*, **2006**, *110*(32), 16102–16110.

[83] Fedorov, D. G.; Kitaura, K. Pair interaction energy decomposition analysis. *J. Comp. Chem.*, **2007**, *28*(1), 222–237.

[84] Fedorov, D. G.; Kitaura, K. Extending the power of quantum chemistry to large systems with the fragment molecular orbital method. *J. Phys. Chem. A*, **2007**, *111*(30), 6904–6914.

[85] Kitaura, K.; Morokuma, K. A new energy decomposition scheme for molecular interactions within the Hartree-Fock approximation. *Int. J. Quant. Chem.*, **1975**, *10*(2), 325–340.

[86] Moszyński, R. in *Molecular Materials with Specific Interactions – Modeling and Design*, Sokalski, W. A., Ed., (Springer, Dordrecht, The Netherlands). volume 4 of *Challenges and Advances in Computational Chemistry and Physics*. chapter 1, pp 1–152. 2007.

[87] Michaelis, L.; Menten, M. L. The kenetics of the inversion effect. *Biochem. Z.*, **1913**, *49*, 333–369.

[88] Fersht, A. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*. W. H. Freeman and Company, New York, 1999.

[89] Miller, B. G.; Wolfenden, R. Catalytic proficiency: The unusual case of OMP decarboxylase. *Annu. Rev. Biochem.*, **2002**, *71*, 847–885.

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

[91] Eyring, H. The activated complex and the absolute rate of chemical reactions. *Chem. Rev.*, **1935**, *17*(1), 65–77.

[92] Kohen, A.; Klinman, J. P. Enzyme catalysis: Beyond classical paradigms. *Acc. Chem. Res.*, **1998**, *31*(7), 397–404.

[93] Ball, P. Enzymes: By chance, or by design? *Nature*, **2004**, *431*(707), 396–397.

[94] Fisher, E. Einfluss der Configuration auf die Wirkung der Enzyme. *Ber. Deutsch. Chem. Ges.*, **1894**, *27*(3), 2985–2993.

[95] Polanyi, M. Adsorption catalysis. *Z. Elektrochem.*, **1921**, *27*, 142–150.

[96] Pauling, L. Molecular architecture and biological reactions. *Chem. Eng. News*, **1946**, *24*(10), 1375–1377.

[97] Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C. A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature*, **1958**, *181*(4610), 662–666.

[98] Náray-Szabó, G.; Fuxreiter, M.; Warshel, A. in *Computational Approaches to Biochemical Reactivity*, Mezey, P. G.; Náray-Szabó, G.; Warshel, A., Eds., (Springer, Netherlands). volume 19 of *Understanding Chemical Reactivity*. pp 237–293. 1997.

[99] Villa, J.; Warshel, A. Energetics and dynamics of enzymatic reactions. *J. Phys. Chem. B*, **2001**, *105*(33), 7887–7907.

[100] Hwang, J. K.; Warshel, A. Semiquantitative calculations of catalytic free energies in genetically modified enzymes. *Biochemistry*, **1987**, *26*(10), 2669–2673.

[101] Warshel, A. Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites. *J. Biol. Chem.*, **1998**, *273*(42), 27035–27038.

[102] Jencks, W. P. Destabilization is as important as binding. *Philos. T. R. Soc. A*, **1993**, *345*(1674), 3–10.

[103] Bruice, T. C. A view at the millennium: The efficiency of enzymatic catalysis. *Acc. Chem. Rev.*, **2002**, *35*(3), 139–148.

[104] Warshel, A.; Florián, J. Computer simulations of enzyme catalysis: Finding out what has been optimized by evolution. *Proc. Nat. Acad. Sci. USA*, **1998**, *95*(11), 5950–5955.

[105] Bruice, T. C.; Kahn, K. Computational enzymology. *Curr. Opin. Chem. Biol.*, **2000**, *4*(5), 540–544.

[106] Ramos, M. J.; Fernandes, P. A. Computational enzymatic catalysis. *Acc. Chem. Res.*, **2008**, *41*(6), 689–698.

[107] Mulholland, A. J. Computational enzymology: Modelling the mechanisms of biological catalysts. *Bioch. Soc. Trans.*, **2008**, *36*(1), 22–26.

[108] van der Kamp, M. W.; Mulholland, A. J. Computational enzymology: insight into biological catalysts from modelling. *Nat. Prod. Rep.*, **2008**, *25*(6), 1001–1014.

[109] Protein Data Bank. http://www.pdb.org.

[110] Moffat, K. Time-resolved biochemical crystallography: A mechanistic perspective. *Chem. Rev.*, **2001**, *101*(6), 1569–1581.

[111] Myles, D. A. Neutron protein crystallography: current status and a brighter future. *Curr. Opin. Struct. Biol.*, **2006**, *16*(5), 630–637.

[112] Wells, T. N. C.; Fersht, A. R. Hydrogen bonding in enzymatic catalysis analysed by protein engineering. *Nature*, **1985**, *316*(6029), 656–657.

[113] Wells, T. N. C.; Fersht, A. R. Use of binding energy in catalysis analyzed by mutagenesis of the tyrosyl-tRNA synthetase. *Biochemistry*, **1986**, *25*(8), 1881–1886.

[114] Wormer, P. E. S.; van der Avoird, A. Intermolecular potentials, internal motions, and spectra of van der Waals and hydrogen-bonded complexes. *Chem. Rev.*, **2000**, *100*(11), 4109–4144.

[115] Curtiss, L. A.; Frurip, D. J.; Blander, M. Studies of molecular association in H_2O and D_2O vapors by measurement of thermal conductivity. *J. Chem. Phys.*, **1979**, *71*(6), 2703–2711.

[116] Campobasso, N.; Mathews, I. I.; Begley, T. P.; Ealick, S. E. Crystal structure of 4-methyl-5-beta-hydroxyethylthiazole kinase from Bacillus subtilis at 1.5 A resolution. *Biochemistry*, **2000**, *39*(27), 7868–7877.

[117] Nickbarg, E. B.; Davenport, R. C.; Petsko, G. A.; Knowles, J. R. Triosephosphate isomerase: removal of a putatively electrophilic histidine residue results in a subtle change in catalytic mechanism. *Biochemistry*, **1988**, *27*(16), 5948–5960.

[118] Mader, M. M.; Bartlett, P. A. Binding energy and catalysis: The implications for transition-state analogs and catalytic antibodies. *Chem. Rev.*, **1997**, *97*(5), 1281–1301.

[119] Schramm, V. L. Enzymatic transition state theory and transition state analogue design. *J. Biol. Chem.*, **2007**, *282*(39), 28297–28300.

[120] Cleland, W. W. The use of isotope effects to determine enzyme mechanisms. *Arch. Biochem. Biophys.*, **2005**, *433*(1), 2–12.

[121] Williams, I. H. Interplay of theory and experiment in the determination of transition-state structure. *Chem. Soc. Rev.*, **1993**, *22*(4), 277–283.

[122] Caldwell, S. R.; Raushel, F. M.; Weiss, P. M.; Cleland, W. W. Transition-state structures for enzymic and alkaline phosphotriester hydrolysis. *Biochemistry*, **1991**, *30*(30), 7444–7450.

[123] Hanson, K. R.; Havir, E. A. L-phenylalanine ammonia-lyase. IV. Evidence that the prosthetic group contains a dehydroalanyl residue and mechanism of action. *Arch. Biochem. Biophys.*, **1970**, *141*(1), 1–17.

[124] Schuster, B.; Rétey, J. The mechanism of action of phenylalanine ammonia-lyase: the role of prosthetic dehydroalanine. *Proc. Nat. Acad. Sci. USA*, **1995**, *92*(18), 8433–8437.

[125] Poppe, L.; Rétey, J. Friedel-Crafts-type mechanism for the enzymatic elimination of ammonia from histidine and phenylalanine. *Angew. Chem. Int. Ed.*, **2005**, *44*(24), 3668–3688.

[126] Lewandowicz, A.; Jemielity, J.; Kanńska, M.; Zoń, J.; Paneth, P. Tritium secondary kinetic isotope effect on phenylalanine ammonia-lyase-catalyzed reaction. *Arch. Biochem. Biophys.*, **1999**, *370*(2), 216–221.

[127] Bartsch, S.; Bornscheuer, U. T. A single residue influences the reaction mechanism of ammonia lyases and mutases. *Angew. Chem. Int. Ed.*, **2009**, *48*(18), 3362–3365.

[128] Klepeis, J. L.; Lindorff-Larsen, K.; Dror, R. O.; Shaw, D. E. Long-timescale molecular dynamics simulations of protein structure and function. *Curr. Opin. Struct. Biol.*, **2009**, *19*(2), 120–127.

[129] Cramer, C. J. *Essentials of Compuational Chemistry.* John Wiley & Sons, 2004.

[130] Senn, H. M.; Thiel, W. QM/MM methods for biomolecular systems. *Angew. Chem. Int. Ed.*, **2009**, *48*(7), 1198–1229.

[131] Field, M. J.; Bash, P. A.; Karplus, M. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J. Comp. Chem.*, **1990**, *11*(6), 700–733.

[132] Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.*, **1976**, *103*(2), 227–249.

[133] Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.*, **1985**, *107*(13), 3902–3909.

[134] Stewart, J. J. P. Optimization of parameters for semiempirical methods i. method. *J. Comput. Chem.*, **1989**, *10*(2), 209–220.

[135] Claeyssens, F.; Harvey, J. N.; Manby, F. R.; Mata, R. A.; Mulholland, A. J.; Ranaghan, K. E.; Schütz, M.; Thiel, S.; Thiel, W.; Werner, H.-J. High-accuracy computation of reaction barriers in enzymes. *Angew. Chem. Int. Ed.*, **2006**, *45*(41), 6856–6859.

[136] Ishida, T.; Fedorov, D. G.; Kitaura, K. All electron quantum chemical calculation of the entire enzyme system confirms a collective catalytic device in the chorismate mutase reaction. *J. Phys. Chem. B*, **2006**, *110*(3), 1457–1463.

[137] Kędzierski, P.; Sokalski, W. A.; Krauss, M. Nonempirical analysis of nature of catalytic effects in ribonuclease A active site. *J. Comp. Chem.*, **2000**, *21*(6), 432–445.

[138] Piela, L. *Ideas of quantum chemistry.* Elsevier, 2006.

[139] Stone, A. *The Theory of Intermolecular Forces.* Oxford University Press, 1996.

[140] Boys, S. F.; Bernardi, F. The calculation of small molecular interactions by the differences of separate total energies. some procedures with reduced errors. *Mol. Phys.*, **1970**, *19*(4), 553–566.

[141] Jeziorski, B.; Kołłos, W. Symmatry forcing in perturbation theory of weak intermolecular interactions. *Int. J. Quant. Chem.*, **1977**, *12*(Suppl. 1), 91–117.

[142] Grabowski, S. J.; Sokalski, W. A.; Dyguda, E.; Leszczyński, J. Quantitative classification of covalent and noncovalent H-bonds. *J. Phys. Chem. B*, **2006**, *110*(13), 6444–6446.

[143] Minkin, V. I. Glossary of terms used in theoretical organic chemistry. *Pure Appl. Chem.*, **1999**, *71*(10), 1919–1981.

[144] Mautner, M. M.-N. The ionic hydrogen bond. *Chem. Rev.*, **2005**, *105*(1), 213–284.

[145] Sobczyk, L.; Grabowski, S. J.; Krygowski, T. M. Interrelation between H-bond and π-electron delocalization. *Chem. Rev.*, **2005**, *105*(10), 3513–3560.

[146] Rozas, I. On the nature of hydrogen bonds: an overview on computational studies and a word about patterns. *Phys. Chem. Chem. Phys.*, **2007**, *9*(22), 2782–2790.

[147] Bader, R. F. W. *Atoms in Molecules: A Quantum Theory (The International Series of Monographs on Chemistry, No 22).* Oxford University Press, USA, 1994.

[148] Cremer, D.; Kraka, E. Chemical bonds without bonding electron density - does the difference electron-density analysis suffice for a description of the chemical bond? *Angew. Chem. Int. Ed.*, **1984**, *23*(8), 627–628.

[149] Grabowski, S. J.; Sokalski, W. A.; Leszczynski, J. How short can the H...H intermolecular contact be? New findings that reveal the covalent nature of extremely strong interactions. *J. Phys. Chem. A*, **2005**, *109*(19), 4331–4341.

[150] Hanson, K. R.; Havir, E. A. in *Secondary Plant Metabolites*, Conn, E. E., Ed., (Academic Press, New York). volume 7 of *The Biochemistry of Plants*. pp 577–625. 1981.

[151] Ferrer, J.-L.; Austin, M. B.; Stewart, C.; Noel, J. P. Structure and function of enzymes involved in the biosynthesis of phenylpropanoids. *Plant Physiol. Biochem.*, **2008**, *46*(3), 356–370.

[152] MacDonald, M. J.; D'Cunha, G. B. A modern view of phenylalanine ammonia lyase. *Biochem. Cell Biol.*, **2007**, *85*(3), 273–282.

[153] Finkelman, M. A. J.; Yang, H.-H. Method for the production of phenylalanine ammonia-lyase by fermentation, 1985. United States Patent 4584273.

[154] Harding, C. O. Progress toward cell-directed therapy for phenylketonuria. *Clin. Genet.*, **2008**, *74*(2), 97–104.

[155] Calabrese, J. C.; Jordan, D. B.; Boodhoo, A.; Sariaslani, S.; Vannelli, T. Crystal structure of phenylalanine ammonia lyase: multiple helix dipoles implicated in catalysis. *Biochemistry*, **2004**, *43*(36), 11403–11416.

[156] Ritter, H.; Schulz, G. E. Structural basis for the entrance into the phenylpropanoid metabolism catalyzed by phenylalanine ammonia-lyase. *Plant Cell.*, **2004**, *16*(12), 3426–3436.

[157] Röther, D.; Merkel, D.; Rétey, J. Spectroscopic evidence for a 4-methylidene imidazol-5-one in histidine and phenylalanine ammonia-lyases. *Angew. Chem. Int. Ed.*, **2000**, *39*(14), 2462–2464.

[158] Röther, D.; Poppe, L.; Morlock, G.; Viergutz, S.; Rétey, J. An active site homology model of phenylalanine ammonia-lyase from Petroselinum crispum. *Eur. J. Biochem.*, **2002**, *269*(12), 3065–3075.

[159] Amrhein, N.; Gödeke, K.-H. α-aminooxy-β-phenylpropionic acid – a potent inhibitor of L-phenylalanine ammonia-lyase in vitro and in vivo. *Plant Sci. Lett.*, **1977**, *8*(4), 313–317.

[160] Janas, K. M.; Filipiak, A.; Kowalik, J.; Mastalerz, P.; Knypl, J. S. 1-amino-2-phenylethylsulphonic acid: an inhibitor of L-phenylalanine ammonia-lyase in vitro. *Acta Biochim. Polon.*, **1985**, *32*(2), 131–143.

[161] Janas, K. M.; Olechnowicz, D. 1-amino-3-phenylpropylphosphonic acid, the inhibitor of L-phenylalanine ammonia-lyase activity of higher plants. *Acta Biochim. Polon.*, **1994**, *41*(2), 191–193.

[162] Zoń, J.; Amrhein, N. Inhibitors of phenylalanine ammonia-lyase: 2-aminoindan-2-phosphonic acid and related compounds. *Liebigs Ann. Chem.*, **1992**, *1992*(6), 625–628.

[163] Appert, C.; Zoń, J.; Amrhein, N. Kinetic analysis of the inhibition of phenylalanine ammonia-lyase by 2-aminoindan-2-phosphonic acid and other phenylalanine analogues. *Phytochemistry*, **2003**, *62*(3), 415–422.

[164] Zoń, J.; Szefczyk, B.; Gancarz, W. S.-D. R.; Kucharska-Zoń, M.; Latajka, R.; Amrhein, N.; Miziak, P.; Szczepanik, W. Experimental and ab initio calculated structures of 2-aminoindane-2-phosphonic acid, a potent inhibitor of phenylalanine ammonia-lyase, and theoretical studies of its binding to the model enzyme structure. *New J. Chem.*, **2004**, *28*(8), 1048–1055.

[165] Zoń, J.; Miziak, P.; Amrhein, N.; Gancarz, R. Inhibitors of phenylalanine ammonia-lyase (PAL): synthesis and biological evaluation of 5-substituted 2-aminoindane-2-phosphonic acids. *Chem. Biodivers.*, **2005**, *2*(9), 1187–1194.

[166] Miziak, P.; Zoń, J.; Amrhein, N.; Gancarz, R. Inhibitors of phenylalanine ammonia-lyase: Substituted derivatives of 2-aminoindane-2-phosphonic acid and 1-aminobenzylphosphonic acid. *Phytochemistry*, **2007**, *68*(4), 407–415.

[167] Zoń, J.; Amrhein, N.; Gancarz, R. Inhibitors of phenylalanine ammonia-lyase: 1-aminobenzylphosphonic acids substituted in the benzene ring. *Phytochemistry*, **2002**, *59*(1), 9–21.

[168] Dyguda, E.; Grembecka, J.; Sokalski, W. A.; Leszczyński, J. Origins of the activity of PAL and LAP enzyme inhibitors: Toward ab initio binding affinity prediction. *J. Am. Chem. Soc.*, **2005**, *127*(6), 1658–1659.

[169] Góra, R. W. EDS Package v. 2.1.2, 1998–2003.

[170] Schmidt, M. S.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. General atomic and molecular electronic structure system. *J. Comput. Chem.*, **1993**, *14*(11), 1347–1363.

[171] Hooft, R. W. W.; Sander, C.; Vriend, G. Positioning hydrogen atoms by optimizing hydrogen-bond networks in protein structures. *Proteins*, **1996**, *26*(4), 363–376.

[172] Vriend, G. WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph.*, **1990**, *8*(1), 52–56.

[173] Brooks, B. R.; Bruccoleri, R. D.; Olafson, B. O.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, **1983**, *4*(2), 187–217.

[174] MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, **1998**, *102*(18), 3586–3616.

[175] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03, Revision B.04*. Gaussian, Inc., Pittsburgh PA, 2003.

[176] Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved protein-ligand docking using GOLD. *Proteins*, **2003**, *52*(4), 609–623.

[177] Berlicki, Ł.; Grembecka, J.; Dyguda-Kazimierowicz, E.; Kafarski, P.; Sokalski, W. A. in *Molecular Materials with Specific Interactions – Modeling and Design*, Sokalski, W. A., Ed., (Springer, Dordrecht, The Netherlands). volume 4 of *Challenges and Advances in Computational Chemistry and Physics*. chapter 8, pp 365–398. 2007.

[178] Chirlian, L. E.; Francl, M. M. Atomic charges derived from electrostatic potentials - a detailed study. *J. Comp. Chem*, **1987**, *8*(6), 894–905.

[179] Grembecka, J.; Kędzierski, P.; Sokalski, W. A.; Leszczyński, J. Electrostatic models of inhibitory activity. *Int. J. Quantum Chem.*, **2001**, *83*(3–4), 180–192.

[180] Famini, G. R.; Wilson, L. Y. in *Reviews in Computational Chemistry*, Lipkowitz, K. B.; Boyd, D. B., Eds., (John Wiley & Sons, Hoboken, New Jersey). volume 18. pp 211–255. 2003.

[181] Kataoka, K.; Asai, T.; Taneda, M.; Ueshima, S.; Matsuo, O.; Kuroda, R.; Kawabata, A.; Carmeliet, P. Roles of urokinase type plasminogen activator in a brain stab wound. *Brain Res.*, **2000**, *887*(1), 187–190.

[182] Solberg, H.; Ploug, M.; Hoyer-Hansen, G.; Nielsen, B. S.; Lund, L. R. The murine receptor for urokinase-type plasminogen activator is primarily expressed in tissues actively undergoing remodeling. *J. Histochem. Cytochem.*, **2001**, *49*(2), 237–246.

[183] Giglio, M. F. D.; Björnheden, T.; Nygren, H.; Risberg, B. Urokinase plasminogen activator colocalizes with CD25+ cells in atherosclerotic vessels. *J. Vasc. Res.*, **1998**, *35*(5), 318–324.

[184] Plekhanova, O.; Parfyonova, Y.; Bibilashvily, R.; Domogatskii, S.; Stepanova, V.; Gulba, D. C.; Agrotis, A.; Bobik, A.; Tkachuk, V. Urokinase plasminogen activator augments cell proliferation and neointima formation in injured arteries via proteolytic mechanisms. *Atherosclerosis*, **2001**, *159*(2), 297–306.

[185] Gveric, D.; Hanemaaijer, R.; Newcombe, J.; van Lent, N. A.; Sier, C. F.; Cuzner, M. L. Plasminogen activators in multiple sclerosis lesions: implications for the inflammatory response and axonal damage. *Brain*, **2001**, *124*(10), 1978–1988.

[186] Almholt, K.; Lund, L. R.; Rygaard, J.; Nielsen, B. S.; Danø, K.; Rømer, J.; Johnsen, M. Reduced metastasis of transgenic mammary cancer in urokinase-deficient mice. *Int. J. Cancer*, **2005**, *113*(4), 525–532.

[187] Andreasen, P. A.; Egelund, R.; Petersen, H. H. The plasminogen activation system in tumor growth, invasion, and metastasis. *Cell. Mol. Life Sci.*, **2000**, *57*(1), 25–40.

[188] Schweinitz, A.; Steinmetzer, T.; Banke, I. J.; Arlt, M. J.; Stürzebecher, A.; Schuster, O.; Geissler, A.; Giersiefen, H.; Zeslawska, E.; Jacob, U.; Krüger, A.; Stürzebecher, J. Design of novel and selective inhibitors of urokinase-type plasminogen activator with improved pharmacokinetic properties for use as antimetastatic agents. *J. Biol. Chem.*, **2004**, *279*(32), 33613–33622.

[189] Rabbani, S. A. Metalloproteases and urokinase in angiogenesis and tumor progression. *In Vivo*, **1998**, *12*(1), 135–142.

[190] Nozaki, S.; Endo, Y.; Nakahara, H.; Yoshizawa, K.; Ohara, T.; Yamamoto, E. Targeting urokinase-type plasminogen activator and its receptor for cancer therapy. *Anticancer Drugs*, **2006**, *17*(10), 1109–1117.

[191] Nienaber, V.; Wang, J.; Davidson, D.; Henkin, J. Re-engineering of human urokinase provides a system for structure-based drug design at high resolution and reveals a novel structural subsite. *J. Biol. Chem.*, **2000**, *275*(10), 7239–7248.

[192] Rockway, T. W.; Giranda, V. L. Inhibitors of the proteolytic activity of urokinase type plasminogen activator. *Curr. Pharm. Des.*, **2003**, *9*(19), 1483–1498.

[193] Oleksyszyn, J.; Powers, J. C. Irreversible inhibition of serine proteases by peptide derivatives of (alpha-aminoalkyl)phosphonate diphenyl esters. *Biochemistry*, **1991**, *30*(2), 485–493.

[194] Fastreza, J.; Jespersa, L.; Lisonb, D.; Renardb, M.; Sonveaux, E. Synthesis of new phosphonate inhibitors of serine proteases. *Tetrahedron Lett.*, **1989**, *30*(49), 6861–6864.

[195] Joossens, J.; der Veken, P. V.; Lambeir, A.-M.; Augustyns, K.; Haemers, A. Development of irreversible diphenyl phosphonate inhibitors for urokinase plasminogen activator. *J. Med. Chem.*, **2004**, *47*(10), 2411–2413.

[196] Joossens, J.; der Veken, P. V.; Surpateanu, G.; Lambeir, A.-M.; El-Sayed, I.; Ali, O. M.; Augustyns, K.; Haemers, A. Diphenyl phosphonate inhibitors for the urokinase-type plasminogen activator: Optimization of the P4 position. *J. Med. Chem.*, **2006**, *49*(19), 5785–5793.

[197] Oleksyszyn, J.; Boduszek, B.; Kam, C. M.; Powers, J. C. Novel amidine-containing peptidyl phosphonates as irreversible inhibitors for blood coagulation and related serine proteases. *J. Med. Chem.*, **1994**, *37*(2), 226–231.

[198] Grzywa, R.; Dyguda-Kazimierowicz, E.; Sieńczyk, M.; Feliks, M.; Sokalski, W. A.; Oleksyszyn, J. The molecular basis of urokinase inhibition: from the nonempirical analysis of intermolecular interactions to the prediction of binding affinity. *J. Mol. Model.*, **2007**, *13*(6–7), 677–683.

[199] Katz, B. A.; Mackman, R.; Luong, C.; Radika, K.; Martelli, A.; Sprengeler, P. A.; Wang, J.; Chan, H.; Wong, L. Structural basis for selectivity of a small molecule, S1-binding, submicromolar inhibitor of urokinase-type plasminogen activator. *Chem. Biol.*, **2000**, *7*(4), 299–312.

[200] Oleksyszyn, J.; Powers, J. C. in *Methods in Enzymology*, Barrett, A. J., Ed., (Academic Press, San Diego). volume 244 of *Proteolytic Enzymes: Serine and Cysteine Peptidases*. chapter 30, pp 423–441. 1994.

[201] Joossens, J.; Ali, O. M.; El-Sayed, I.; Surpateanu, G.; der Veken, P. V.; Lambeir, A.-M.; Setyono-Han, B.; Foekens, J. A.; Schneider, A.; Schmalix, W.; Haemers, A.; Augustyns, K. Small, potent, and selective diaryl phosphonate inhibitors for urokinase-type plasminogen activator with in vivo antimetastatic properties. *J. Med. Chem.*, **2007**, *50*(26), 6638–6646.

[202] Cheek, S.; Zhang, H.; Grishin, N. V. Sequence and structure classification of kinases. *J. Mol. Biol.*, **2002**, *320*(4), 855–881.

[203] Adams, J. A. Kinetic and catalytic mechanisms of protein kinases. *Chem. Rev.*, **2001**, *101*(8), 2271–2290.

[204] Bork, P.; Sander, C.; Valencia, A. Convergent evolution of similar enzymatic function on different protein folds: the hexokinase, ribokinase, and galactokinase families of sugar kinases. *Protein Sci.*, **1993**, *2*(1), 31–40.

[205] Sigrell, J. A.; Cameron, A. D.; Jones, T. A.; Mowbray, S. L. Structure of Escherichia coli ribokinase in complex with ribose and dinucleotide determined to 1.8 A resolution: insights into a new family of kinase structures. *Structure*, **1998**, *6*(2), 183–193.

[206] Mathews, I. I.; Erion, M. D.; Ealick, S. E. Structure of human adenosine kinase at 1.5 A resolution. *Biochemistry*, **1998**, *37*(45), 15607–15620.

[207] Li, M. H.; Kwok, F.; Chang, W. R.; Lau, C. K.; Zhang, J. P.; Lo, S. C.; Jiang, T.; Liang, D. C. Crystal structure of brain pyridoxal kinase, a novel member of the ribokinase superfamily. *J. Biol. Chem.*, **2002**, *277*(48), 46385–46390.

[208] Matte, A.; Tari, L. W.; Delbaere, L. T. J. How do kinases transfer phosphoryl groups? *Structure*, **1998**, *6*(4), 413–419.

[209] Dyguda-Kazimierowicz, E.; Sokalski, W. A.; Leszczyński, J. Non-empirical study of the phosphorylation reaction catalyzed by 4-methyl-5-beta-hydroxyethylthiazole kinase: Relevance of the theory of intermolecular interactions. *J. Mol. Model.*, **2007**, *13*(6–7), 839–849.

[210] Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.*, **1993**, *98*(7), 5648–5652.

[211] Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, **1988**, *37*(2), 785–789.

[212] Maseras, F.; Morokuma, K. IMOMM: A new integrated ab initio + molecular mechanics geometry optimization scheme of equilibrium structures and transition states. *J. Comput. Chem.*, **1995**, *16*(9), 1170–1179.

[213] Dyguda, E.; Szefczyk, B.; Sokalski, W. A. The mechanism of phosphoryl transfer reaction and the role of active site residues on the basis of ribokinase-like kinases. *Int. J. Mol. Sci.*, **2004**, *5*(4–7), 141–153.

[214] Bairoch, A.; Apweiler, R.; Wu, C. H.; Barker, W. C.; Boeckmann, B.; Ferro, S.; Gasteiger, E.; Huang, H.; Lopez, R.; Magrane, M.; Martin, M. J.; Natale, D. A.; O'Donovan, C.; Redaschi, N.; Yeh, L. S. The Universal Protein Resource (UniProt). *Nucleic Acids Res.*, **2005**, *33*(Database issue), D154–D159.

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

[216] Gouet, P.; Courcelle, E.; Stuart, D. I.; Métoz, F. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics*, **1999**, *15*(4), 305–308.

[217] Akola, J.; Jones, R. O. ATP hydrolysis in water — a density functional study. *J. Phys. Chem. B*, **2003**, *107*(42), 11774–11783.

[218] Díaz, N.; Field, M. J. Insights into the phosphoryl-transfer mechanism of cAMP-dependent protein kinase from quantum chemical calculations and molecular dynamics simulations. *J. Am. Chem. Soc.*, **2004**, *126*(2), 529–542.

[219] Mildvan, A. S. Mechanisms of signaling and related enzymes. *Proteins*, **1997**, *29*(4), 401–416.

[220] Herberg, F. W.; Doyle, M. L.; Cox, S.; Taylor, S. S. Dissection of the nucleotide and metal-phosphate binding sites in cAMP-dependent protein kinase. *Biochemistry*, **1999**, *38*(19), 6352–6360.

[221] Taylor, S. S.; Bubis, J.; Toner-Webb, J.; Saraswat, L. D.; First, E. A.; Buechler, J. A.; Knighton, D. R.; Sowadski, J. cAMP-Dependent protein kinase: Prototype for a family of enzymes. *FASEB J.*, **1988**, *2*(11), 2677–2685.

[222] Johnson, D. A.; Akamine, P.; Radzio-Andzelm, E.; Madhusudan, M.; Taylor, S. S. Dynamics of cAMP-dependent protein kinase. *Chem. Rev.*, **2001**, *101*(8), 2243–2270. and references therein.

[223] Cheng, Y.; Zhang, Y.; McCammon, J. A. How does activation loop phosphorylation modulate catalytic activity in the cAMP-dependent protein kinase: A theoretical study. *Protein Sci.*, **2006**, *15*(4), 672–683.

[224] Hart, J. C.; Hillier, I. H.; Burton, N. A.; Sheppard, D. W. An alternative role for the conserved Asp residue in phosphoryl transfer reactions. *J. Am. Chem. Soc.*, **1998**, *120*(51), 13535–13536.

[225] Hart, J. C.; Sheppard, D. W.; Hillier, I. H.; Burton, N. A. What is the mechanism of phosphoryl transfer in protein kinases? A hybrid quantum mechanical/molecular mechanical study. *Chem. Commun.*, **1999**, (1), 79–80.

[226] Henkelman, G.; LaBute, M. X.; Tung, C.-S.; Fenimore, P. W.; McMahon, B. H. Conformational dependence of a protein kinase phosphate transfer reaction. *Proc. Natl. Acad. Sci. U.S.A.*, **2005**, *102*(43), 15347–15351.

[227] Hirano, Y.; Hata, M.; Hoshino, T.; Tsuda, M. Quantum chemical study on the catalytic mechanism of the C-subunit of cAMP-dependent protein kinase. *J. Phys. Chem. B*, **2002**, *106*(22), 5788–5792.

[228] Hutter, M. C.; Helms, V. Influence of key residues on the reaction mechanism of the cAMP-dependent protein kinase. *Proein Sci.*, **1999**, *8*(12), 2728–2733.

[229] Hutter, M. C.; Helms, V. Mechanism of phosphoryl transfer in kinases investigated by semiempirical calculations. *Int. J. Quantum Chem.*, **2003**, *95*(4–5), 479–486.

[230] Valiev, M.; Kawai, R.; Adams, J. A.; Weare, J. H. The role of the putative catalytic base in the phosphoryl transfer reaction in a protein kinase: First-principles calculations. *J. Am. Chem. Soc.*, **2003**, *125*(33), 9926–9927.

[231] Valiev, M.; Yang, J.; Adams, J. A.; Taylor, S. S.; Weare, J. H. Phosphorylation reaction in cAPK protein kinase-free energy quantum mechanical/molecular mechanics simulations. *J. Phys. Chem. B*, **2007**, *111*(47), 13455–13464.

[232] Armstrong, R. N.; Kondo, H.; Granot, J.; Kaiser, E. T.; Mildvan, A. S. Magnetic resonance and kinetic studies of the manganese(II) ion and substrate complexes of the catalytic subunit of adenosine 3',5'-monophosphate dependent protein kinase from bovine heart. *Biochemistry*, **1979**, *18*(7), 1230–1238.

[233] Cook, A.; Lowe, E. D.; Chrysina, E. D.; Skamnaki, V. T.; Oikonomakos, N. G.; Johnson, L. N. Structural studies on phospho-CDK2/cyclin A bound to nitrate, a transition state analogue: Implications for the protein kinase mechanism. *Biochemistry*, **2002**, *41*(23), 7301–7311.

[234] Yoon, M.-Y.; Cook, P. F. Chemical mechanism of the adenosine cyclic 3',5'-monophosphate dependent protein kinase from pH studies. *Biochemistry*, **1987**, *26*(13), 4118–4125.

[235] Shaltiel, S.; Cox, S.; Taylor, S. S. Conserved water molecules contribute to the extensive network of interactions at the active site of protein kinase A. *Proc. Natl. Acad. Sci. U.S.A.*, **1998**, *95*(2), 484–491.

[236] Gibbs, C. S.; Zoller, M. J. Rational scanning mutagenesis of a protein kinase identifies functional regions involved in catalysis and substrate interactions. *J. Biol. Chem.*, **1991**, *266*(14), 8923–8931.

[237] Aimes, R. T.; Hemmer, W.; Taylor, S. S. Serine-53 at the tip of the glycine-rich loop of cAMP-dependent protein kinase: Role in catalysis, P-site specificity, and interaction with inhibitors. *Biochemistry*, **2000**, *39*(28), 8325–8332.

[238] Hemmer, W.; McGlone, M.; Tsigelny, I.; Taylor, S. S. Role of the glycine triad in the ATP-binding site of cAMP-dependent protein kinase. *J. Biol. Chem.*, **1997**, *272*(27), 16946–16954.

[239] Talmage, S. S.; Watson, A. P.; Hauschild, V.; Munro, N. B.; King, J. Chemical warfare agent degradation and decontamination. *Curr. Org. Chem.*, **2008**, *11*(3), 285–298.

[240] Šeckute, J.; Menke, J. L.; Emnett, R. J.; Patterson, E. V.; Cramer, C. J. Ab initio molecular orbital and density functional studies on the solvolysis of sarin and O,S-dimethyl methylphosphonothiolate, a VX-like compound. *J. Org. Chem.*, **2005**, *70*(22), 8649–8660.

[241] Xiong, Y.; Zhan, C.-G. Reaction pathways and free energy barriers for alkaline hydrolysis of insecticide 2-trimethylammonioethyl methylphosphonofluoridate and related organophosphorus compounds: Electrostatic and steric effects. *J. Org. Chem.*, **2004**, *69*(24), 8451–8458.

[242] Zheng, F.; Zhan, C.-G.; Ornstein, R. L. Theoretical studies of reaction pathways and energy barriers for alkaline hydrolysis of phosphotriesterase substrates paraoxon and related toxic phosphofluoridate nerve agents. *J. Chem. Soc. Perkin. Trans.*, **2001**, *2*(12), 2355–2363.

[243] Raushel, F. M. Bacterial detoxification of organophosphate nerve agents. *Curr. Opin. Microbiol.*, **2002**, *5*(3), 288–295.

[244] Yair, S.; Ofer, B.; Arik, E.; Shai, S.; Yossi, R.; Tzvika, D.; Amir, K. Organophosphate degrading microorganisms and enzymes as biocatalysts in environmental and personal decontamination applications. *Crit. Rev. Biotechnol.*, **2008**, *28*(4), 265–275.

[245] Dumas, D. P.; Caldwell, S. R.; Wild, J. R.; Raushel, F. M. Purification and properties of the phosphotriesterase from Pseudomonas diminuta. *J. Biol. Chem.*, **1989**, *264*(33), 19659–19665.

[246] Dumas, D. P.; Durst, H. D.; Landis, W. G.; Raushel, F. M.; Wild, J. R. Inactivation of organophosphorus nerve agents by the phosphotriesterase from Pseudomonas diminuta. *Arch. Biochem. Biophys.*, **1990**, *277*(1), 155–159.

[247] Lai, K.; Stolowich, N. J.; Wild, J. R. Characterization of P-S bond hydrolysis in organophosphorothioate pesticides by organophosphorus hydrolase. *Arch. Biochem. Biophys.*, **1995**, *318*(1), 59–64.

[248] Caldwell, S. R.; Newcomb, J. R.; Schlecht, K. A.; Raushel, F. M. Limits of diffusion in the hydrolysis of substrates by the phosphotriesterase from Pseudomonas diminuta. *Biochemistry*, **1991**, *30*(30), 7438–7444.

[249] C.-G.Zhan; De Souza, O. N.; Rittenhouse, R.; Ornstein, R. L. Determination of two structural forms of catalytic bridging ligand in zinc - phosphotriesterase by molecular dynamics simulation and quantum chemical calculation. *J. Am. Chem. Soc.*, **1999**, *121*(32), 7279–7282.

[250] Chen, S.-L.; Fang, W.-H.; Himo, F. Theoretical study of the phosphotriesterase reaction mechanism. *J. Phys. Chem. B*, **2007**, *111*(6), 1253–1255.

[251] Chen, S.-L.; Fang, W.-H.; Himo, F. Technical aspects of quantum chemical modeling of enzymatic reactions: The case of phosphotriesterase. *Theor. Chem. Account*, **2008**, *120*(4–6), 515–522.

[252] Koca, J.; Zhan, C.-G.; Rittenhouse, R. C.; Ornstein, R. L. Mobility of the active site bound paraoxon and sarin in zinc-phosphotriesterase by molecular dynamics simulation and quantum chemical calculation. *J. Am. Chem. Soc.*, **2001**, *123*(5), 817–826.

[253] Koca, J.; Zhan, C.-G.; Rittenhouse, R. C.; Ornstein, R. L. Coordination number of zinc ions in the phosphotriesterase active site by molecular dynamics and quantum mechanics. *J. Comput. Chem.*, **2003**, *24*(3), 368–378.

[254] Krauss, M. Ab initio structure of the active site of phosphotriesterase. *J. Chem. Inf. Model.*, **2001**, *41*(1), 8–17.

[255] Krauss, M.; Olsen, L.; Antony, J.; Hemmingsen, L. Coordination geometries of Zn(II) and Cd(II) in phosphotriesterase: Influence of water molecules in the active site. *J. Phys. Chem. B*, **2002**, *106*(36), 9446–9453.

[256] Pang, Y.-P. Successful molecular dynamics simulation of two zinc complexes bridged by a hydroxide in phosphotriesterase using the cationic dummy atom method. *Proteins*, **2001**, *45*(3), 183–189.

[257] Soares, T. A.; Osman, M. A.; Straatsma, T. P. Molecular dynamics of organophosphorous hydrolases bound to the nerve agent soman. *J. Chem. Theory Comput.*, **2007**, *3*(4), 1569–1579.

[258] Wong, K.-Y.; Gao, J. The reaction mechanism of paraoxon hydrolysis by phosphotriesterase from combined QM/MM simulations. *Biochemistry*, **2007**, *46*(46), 13352–13369.

[259] Kim, J.; Tsai, P.-C.; Chen, S.-L.; Himo, F.; Almo, S. C.; Raushel, F. M. Structure of diethyl phosphate bound to the binuclear metal center of phosphotriesterase. *Biochemistry*, **2008**, *47*(36), 9497–9504.

[260] Raveh, L.; Segall, Y.; Leader, H.; Rothschild, N.; Levanon, D.; Henis, Y.; Ashani, Y. Protection against tabun toxicity in mice by prophylaxis with an enzyme hydrolyzing organophosphate esters. *Biochem. Pharmacol.*, **1992**, *44*(2), 397–400.

[261] Shim, H.; Hong, S.-B.; Raushel, F. M. Hydrolysis of phosphodiesters through transformation of the bacterial phosphotriesterase. *J. Biol. Chem.*, **1998**, *273*(28), 17445–17450.

[262] Lewis, V. E.; Donarski, W. J.; Wild, J. R.; Raushel, F. M. Mechanism and stereochemical course at phosphorus of the reaction catalyzed by a bacterial phosphotriesterase. *Biochemistry*, **1988**, *27*(5), 1591–1597.

[263] Dumas, D. P.; Raushel, F. M. Chemical and kinetic evidence for an essential histidine in the phosphotriesterase from Pseudomonas diminuta. *J. Biol. Chem.*, **1990**, *265*(35), 21498–21503.

[264] Samples, C. R.; Howard, T.; Raushel, F. M.; DeRose, V. J. Protonation of the binuclear metal center within the active site of phosphotriesterase. *Biochemistry*, **2005**, *44*(33), 11005–11013.

[265] Jackson, C. J.; Foo, J.-L.; Kim, H. K.; Carr, P. D.; Liu, J.-W.; Salem, G.; Ollis, D. L. In crystallo capture of a Michaelis complex and product-binding modes of a bacterial phosphotriesterase. *J. Mol. Biol.*, **2008**, *375*(5), 1189–1196.

[266] Hong, S.-B.; Raushel, F. M. Metal-substrate interactions facilitate the catalytic activity of the bacterial phosphotriesterase. *Biochemistry*, **1996**, *35*(33), 10904–10912.

[267] Cox Jr, J. R.; Ramsay, O. B. Mechanisms of nucleophilic substitution in phosphate esters. *Chem. Rev.*, **1964**, *64*(4), 317–352.

[268] Knowles, J. R. Enzyme-catalyzed phosphoryl transfer reactions. *Annu. Rev. Biochem.*, **1980**, *49*, 877–919.

[269] Dyguda-Kazimierowicz, E.; Sokalski, W. A.; Leszczyński, J. Gas-phase mechanisms of degradation of hazardous organophosphorus compounds: Do they follow a common pattern of alkaline hydrolysis reaction as in phosphotriesterase? *J. Phys. Chem. B*, **2008**, *112*(32), 9982–9991.

[270] Gonzalez, C.; Schlegel, H. B. An improved algorithm for reaction path following. *J. Chem. Phys.*, **1989**, *90*(4), 2154–2161.

[271] Dewar, M. J. S.; Storch, D. M. Alternative view of enzyme reactions. *Proc. Nat. Acad. Sci. USA*, **1985**, *82*(8), 2225–2229.

[272] Lightstone, F. C.; Zheng, Y.-J.; Maulitz, A. H.; Bruice, T. C. Non-enzymatic and enzymatic hydrolysis of alkyl halides: A haloalkane dehalogenation enzyme evolved to stabilize the gas-phase transition state of an SN2 displacement reaction. *Proc. Nat. Acad. Sci. USA*, **1997**, *94*(16), 8417–8420.

[273] Rucker, V. C.; Byers, L. D. An assessment of desolvation on rates of acetyl transfer: Insights into enzyme catalysis. *J. Am. Chem. Soc.*, **2000**, *122*(35), 8365–8369.

[274] Devi-Kesavan, L. S.; Gao, J. Combined QM/MM study of the mechanism and kinetic isotope effect of the nucleophilic substitution reaction in haloalkane dehalogenase. *J. Am. Chem. Soc.*, **2003**, *125*(6), 1532–1540.

[275] Cancés, E.; Mennucci, B.; Tomasi, J. A new integral equation formalism for the polarizable continuum model: Theoretical background and applications to isotropic and anisotropic dielectrics. *J. Chem. Phys.*, **1997**, *107*(8), 3032–3041.

[276] Tomasi, J.; Mennucci, B.; Cammi, R. Quantum mechanical continuum solvation models. *Chem. Rev.*, **2005**, *105*(8), 2999–3094.

[277] Faust, S. D.; Gomaa, H. M. Chemical hydrolysis of some organic phosphorus and carbamate pesticides in aquatic environments. *Environ. Lett.*, **1972**, *3*(3), 171–201.

[278] Foresman, J. B.; Keith, T. A.; Wiberg, K. B.; Snoonian, J.; Frisch, M. J. Solvent effects. 5. Influence of cavity shape, truncation of electrostatics, and electron correlation on ab initio reaction field calculations. *J. Phys. Chem.*, **1996**, *100*(40), 16098–16104.

[279] Kaczmarek, A.; Gorb, L.; Sadlej, A. J.; Leszczyński, J. Sarin and soman: Structure and properties. *Struct. Chem.*, **2004**, *15*(5), 517–525.

[280] Klähn, M.; Braun-Sand, S.; Rosta, E.; Warshel, A. On possible pitfalls in ab initio quantum mechanics/molecular mechanics minimization approaches for studies of enzymatic reactions. *J. Phys. Chem. B*, **2005**, *109*(32), 15645–15650.

[281] Vaiana, A. C.; Cournia, Z.; Costescu, I. B.; Smith, J. C. AFMM: A molecular mechanics force field vibrational parametrization program. *Comp. Phys. Commun.*, **2005**, *167*(1), 34–42.

[282] Benning, M. M.; Hong, S.-B.; Raushel, F. M.; Holden, H. M. The binding of substrate analogs to phosphotriesterase. *J. Biol. Chem.*, **2000**, *275*(39), 30556–30560.

[283] Hong, S.-B.; Raushel, F. M. Stereochemical constraints on the substrate specificity of phosphotriesterase. *Biochemistry*, **1999**, *38*(4), 1159–1165.

[284] Li, H.; Robertson, A. D.; Jensen, J. Very fast empirical prediction and rationalization of protein pKa values. *Proteins*, **2005**, *61*(4), 704–721.

[285] Davis, M. E.; Madura, J. D.; Luty, B. A.; McCammon, J. A. Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian dynamics program. *Comp. Phys. Commun.*, **1991**, *62*(2–3), 187–197.

[286] Feng, M. H.; Philippopoulos, M.; MacKerell Jr., A. D.; Lim, C. Structural characterization of the phosphotyrosine binding region of a high-affinity SH2 domain-phosphopeptide complex by molecular dynamics simulation and chemical shift calculations. *J. Am. Chem. Soc.*, **1996**, *118*(45), 11265–11277.

[287] Zhan, C.-G.; De Souza, O. N.; Rittenhouse, R.; Ornstein, R. L. Determination of two structural forms of catalytic bridging ligand in zinc—phosphotriesterase by molecular dynamics simulation and quantum chemical calculation. *J. Am. Chem. Soc.*, **1999**, *121*(32), 7279–7282.

[288] Rieder, S. V.; Rose, I. A. The mechanism of the triosephosphate isomerase reaction. *J. Biol. Chem.*, **1959**, *234*(5), 1007–1010.

[289] Hall, A.; Knowles, J. R. Uncatalyzed rates of enolization of dihydroxyacetone phosphate and of glyceraldehyde 3-phosphate in neutral aqueous solution. Quantitative assessment of the effectiveness of an enzyme catalyst. *Biochemistry*, **1975**, *14*(19), 4348–4352.

[290] Knowles, J. R. Enzyme catalysis: not different, just better. *Nature*, **1991**, *350*(6314), 121–124.

[291] Alagona, G.; Ghio, C.; Kollman, P. A. The intramolecular mechanism for the second proton transfer in triosephosphate isomerase (TIM): a QM/FE approach. *J. Comput. Chem.*, **2003**, *24*(1), 46–56.

[292] Cui, Q.; Karplus, M. Quantum mechanical/molecular mechanical studies of the triosephosphate isomerase-catalyzed reaction: Verification of methodology and analysis of reaction mechanisms. *J. Phys. Chem. B*, **2002**, *106*(7), 1768–1798.

[293] Bader, R. F. W. AIMPAC. http://www.chemistry.mcmaster.ca/aimpac.

CHARMM27 topology and parameters for sarin

Topology file

```
RESI SAR          0.00   !
GROUP                    !           H21
ATOM P1   P       1.030  !            |
ATOM O1   ON2    -0.440  !      H23--C2--H22    H31
ATOM O2   ON3    -0.540  !            |        /
ATOM FP   F1     -0.280  !      H11--C1----C3--H32
!                        !            |        \
!                        !      HP1  O1         H33
ATOM C1   CN7     0.270  !       |    |
ATOM H11  HN7    -0.040  !  HP2--CP--P1--FP
!                              |   ||
GROUP                    !      HP3  O2
ATOM C2   CN9    -0.120
ATOM H21  HN9     0.040
ATOM H22  HN9     0.040
ATOM H23  HN9     0.040
GROUP
ATOM C3   CN9    -0.120
ATOM H31  HN9     0.040
ATOM H32  HN9     0.040
ATOM H33  HN9     0.040
GROUP
ATOM CP  CT3   -0.18
```

```
ATOM HP1 HA     0.06
ATOM HP2 HA     0.06
ATOM HP3 HA     0.06
BOND P1   O1     P1   O2    P1   FP    P1  CP  O1  C1
BOND C1   H11    C1   C2    C1   C3
BOND C2   H21    C2   H22   C2   H23
BOND C3   H31    C3   H32   C3   H33
BOND CP   HP1    CP   HP2   CP   HP3
```

Additional parameters

```
BONDS
P       F1      233.0  1.600
P       CT3     236.0  1.800
ANGLES
CN9  CN7  CN9   58.35   113.60  11.16   2.561
ON2  CN7  CN9   115.0    109.7
HA   CT3   P    42.7    110.0
CT3  P    ON3   98.9    118.0
CT3  P    ON2   37.0    103.0
F1   P    ON3   50.0    112.0
F1   P    ON2   50.0    103.0
F1   P    CT3   47.1    102.0
DIHEDRALS
CN9  CN7  ON2  P     0.40    1  180.0
CN9  CN7  ON2  P     0.30    2    0.0
CN9  CN7  ON2  P     0.10    3    0.0
CN9  CN7  CN9  HN9   0.195   3    0.0
ON2  CN7  CN9  HN9   0.195   3    0.0
HA   CT3  P    ON3   0.10    3    0.0
HA   CT3  P    ON2   0.10    3    0.0
F1   P    ON2  CN7   0.10    1  270.0
F1   P    CT3  HA    0.10    3    0.0
CT3  P    ON2  CN7   0.10    1    0.0
```