

Charles University in Prague
Faculty of Science

Ph. D. study program: Modelling of Chemical Properties of Nano- and Biostructures



M.Sc. Adam Pecina

Quantum Chemical Approach for *In Silico* Drug Design

Doctoral Thesis

Supervisor:

prof. Ing. Pavel Hobza DrSc., FRSC, dr. h. c.

Institute of Organic Chemistry and Biochemistry,
Academy of Sciences of the Czech Republic, v. v. i.

Prague, 2016

Univerzita Karlova v Praze
Přírodovědecká fakulta

Studijní obor: Modelování chemických vlastností nano- a biostruktur



Mgr. Adam Pecina

Kvantově chemické pojetí návrhu léčiv

Disertační práce

Školitel:

prof. Ing. Pavel Hobza DrSc., FRSC, dr. h. c.

Ústav organické chemie a biochemie,
Akademie věd České republiky, v.v.i.

Praha, 2016

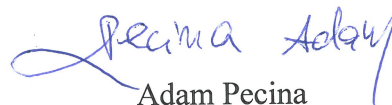
Declaration of Authorship:

I hereby certify that the thesis entitled „Quantum chemical approach for *in silico* drug design“ is entirely my own work. All direct or indirect sources used are acknowledged as references and this work was not previously presented to another examination board.

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracoval samostatně a že jsem uvedl všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

January 2016, Prague


Adam Pecina

Acknowledgement

It is a genuine pleasure to express my deep sense of gratitude to my supervisor Prof. Pavel Hobza for his kind guidance, unflagging enthusiasm and priceless scientific approach. He has literally become a paragon of virtue not only in science for me. I am very much thankful to my colleagues, namely Robo S. and Michal K., Jindra F. and Martin L., Honza Ř. and Kevin R., Tom D. and Jerry H., Susanta H., Ella N., Jiří P. and many others, for their everyday help, constructive criticism and real-life friendship. They all made my stay in „Canon“ one I will always remember.

I also thank all my collaborators whom I had a pleasure to work with, namely Jiří Brynda, Pavlína Řezáčová, Bohumír Grüner, Aleš Růžička, Václav Šícha, Drahomír Hnyk, Jiří Dostál, Pavel Máder, Petr Cígler, René Meier and Carsten Baldauf.

Let me also use this opportunity to thank my loving parents Vlasta and Petr, my sister Adéla and my wife Veronika for their constant encouragement, support and love that were imperative to my success.

Abstract

Computational approaches have become an established and valuable component of pharmaceutical research. Computer-aided drug design aims to reduce the time and cost of the drug development and also to bring deeper insight into the inhibitor binding to its target. The complexity of biological systems together with a need of proper description of non-covalent interactions involved in molecular recognition challenges the accuracy of commonly used molecular mechanical methods (MM). There is on the other side a growing interest of utilizing quantum mechanical (QM) methods in several stages of drug design thanks to increased computational resources.

This doctoral thesis's topic is the QM-based methodology for the reliable treatment of intermolecular interactions. It consists of eight original publications divided into three topics and an accompanying text that aims to emphasize selected outcomes of the work. Firstly, the nature of nonclassical non-covalent interactions - so called σ -hole bonding - is studied by high-level QM methods. The strength and origin of halogen-, chalcogen- and pnictogen bonded model systems in extended datasets are accurately explored by coupled cluster QM method (CCSD(T)/CBS) and symmetry adapted perturbation theory (SAPT). The second part is devoted to three pharmaceutically important protein targets, *i.e.* HIV-1 protease, secreted aspartic protease and carbonic anhydrase, and shows benefits of corrected DFT and semiempirical quantum mechanical (SQM) methods used in protein-ligand complexes involving proton-transfer phenomena, metal ions and unusual compounds such as boranes. A hybrid QM/MM approach unveils here the features of the structure that are not accessible to the crystallographic experiment and explains fundamental differences in the binding modes of inhibitors. Finally, SQM-based scoring function that describes quantitatively all types of non-covalent protein-ligand interactions is simplified for virtual screening of compound libraries. The reliability of this physics-based SQM/COSMO filter is tested on four unrelated difficult-to-handle protein-ligand systems. In this last part of the thesis it is shown how the SQM/COSMO filter outperforms eight standardly used scoring functions and thus may become an effective tool for accurate medium-throughput refinement in later stages of virtual screening.

Abstrakt

Výpočetní metody jsou nedílnou součástí moderního farmaceutického výzkumu. Počítačový návrh léčiv si klade za cíl snížit čas a náklady spojené s vývojem léčiva a také detailněji porozumět vazbě inhibitoru k danému biologickému cíli. Kvůli komplikovanosti biologických systémů a potřebě správného popisu nekovalentních interakcí nutných k molekulárnímu rozpoznávání je přesnost běžně používaných molekulově mechanických (MM) metod na hraně spolehlivosti. Na druhou stranu zde vzrůstá tendence používání kvantově mechanických (QM) metod v různých fázích vývoje léčiv díky rostoucím výpočetním možnostem.

Tato disertační práce se zabývá aplikací kvantově mechanických metod pro věrný popis mezimolekulových komplexů a jejich interakcí. Tato práce zahrnuje osm původních publikací rozdělených do tří témat a doprovodný text, jenž si klade za cíl zdůraznit některé závěry plynoucí z této práce. V první řadě je vysoce přesnými kvantově mechanickými metodami studována povaha neklasických nekovalentních interakcí, tzv. vazebné interakce pomocí sigma díry. Síla a původ halogenové, chalkogenové a pniktogenové vazby v modelových systémech z rozšířených databází molekul jsou zkoumány přesnou metodou vázaných klastrů (CCSD(T)/CBS) a symetricky adaptovanou poruchovou teorií (SAPT). Druhá část se věnuje třem farmaceuticky důležitým proteinům, a to HIV-1 protease, sekretované aspartátové protease a karbanhydrase, a ukazuje výhody aplikace opravených DFT a semiempirických (SQM) metod na protein-ligandové komplexy spojené s přenosy protonu, s ionty kovů a s neobvyklými molekulami jakými jsou borany. Strukturální vlastnosti, jež jsou experimentálně (krystalograficky) nedosažitelné, a zásadní vazebné rozdíly inhibitorů jsou zde odhaleny hybridním QM/MM přístupem. Následně je SQM skórovací funkce, jež kvantitativně správně popisuje všechny typy nekovalentních protein-ligandových interakcí, adaptována pro virtuální prohledávání databází sloučenin (tzv. „virtual screening“). Spolehlivost tohoto fyzikálního „SQM/COSMO“ filtru je testována na čtyřech nepříbuzných netriviálních protein-ligandových systémech. V této poslední části mé disertační práce je ukázáno, jak tento „SQM/COSMO“ filtr předčí osm standardně používaných skórovacích funkcí a jak tedy může být efektivním nástrojem pro zpřesňování v pozdějších fázích virtuálního prohledávání.

Contents

List of Abbreviations	ii
List of Figures	iv
Chapter 1	
Introduction	1
Chapter 2	
Methods	7
2.1 Supermolecular Interaction Energy	7
2.1.1 Coupled Cluster Theory	10
2.1.2 Density Functional Theory	11
2.1.3 Semiempirical Quantum Mechanical Methods	12
2.2 Intermolecular Perturbation Theory	14
2.2.1 DFT-SAPT	15
2.3 Hybrid QM/MM Approach	16
2.4. Solvation Models	18
Chapter 3	
Projects	20
3.1 Nature of σ -hole Bonding	20
3.2 Protein-Ligand Binding	34
3.2.1 Protonation of HIV-1 Protease/Inhibitor complex	36
3.2.2 Secreted Aspartic Protease of Candida Parapsilosis	41
3.2.3 Carborane-based Inhibitors of Carbonic Anhydrases	49
3.3 The SQM/COSMO filter	60
Chapter 4	
Summary and final remarks	69
Bibliography	73
List of Publications	86
Declaration of co-authorship	88
Appendices	89

List of Abbreviations

3D	three-dimensional
AIDS	Acquired Immune Deficiency Syndrome
AIM	Atoms-In-Molecules
BSSE	Basis Set Superposition Error
CA	Carbonic Anhydrase
CADD	Computer Aided Drug Design
CBS	Complete Basis Set
CC	Coupled Cluster
COSMO	COnductor-like Screening MOdel
CP	CounterPoise correction
DFT	Density Functional Theory
DFT-SAPT	Density Function Theory Symmetry Adapted Perturbation Theory
ESP	ElectroStatic Potential
GAFF	General Amber Force Field
GB	Generalized Born
GGA	Generalized Gradient Approximation
HIV	Human Immunodeficiency Virus
HF	Hartree-Fock
LDA	Local-Density Approximation

LRA	Linear Response Approximation
MD	Molecular Dynamics
MM	Molecular Mechanics
MP2	Moller-Plesset perturbative method to second order
NMR	Nuclear Magnetic Resonance
ONIOM	Our own N-layered Integrated molecular Orbital and Molecular mechanics
PDB	Protein Data Bank
PR	Protease
QM	Quantum Mechanical (quantum mechanics)
QM/MM	Quantum Mechanics/Molecular Mechanics
QSAR	Quantitative Structure-Activity Relationships
RESP	Restrained fit to the ElectroStatic Potential
RMSD	Root-Mean-Square Deviations
SAPT	Symmetry-Adapted Perturbation Theory
Saps	Secreted aspartic proteases
SQM	Semiempirical Quantum Mechanical
TD-DFT	Time-Dependent DFT
vHTS	virtual High Throughput Screening

List of Figures

3.1	Halogen bond of bromobenzene...acetone complex	20
3.2	The typical σ -hole	22
3.3	The dependence of σ -hole characteristics.	25
3.4	ESP and dipole moments of 2D and 3D aromatic sulphur-bound structures	27
3.5	Binding motifs of 12-Ph- <i>closo</i> -1-SB ₁₁ H ₁₁ crystal model	28
3.6	DFT-SAPT decomposition for all binding motifs	29
3.7	ESP of <i>closo</i> -1,2-P ₂ B ₁₀ H ₁₀ and <i>closo</i> -1,2-As ₂ B ₁₀ H ₁₀	30
3.8	Active sites of HIV-1 PR/DRV and HIV-1 PR/KI2 complexes	37
3.9	Proton-transfer barriers	39
3.10	The complex of Sapp2p and pepstatin A	42
3.11	Interactions of Sapp2p and pepstatin A	43
3.12	The difference of binding of pepstatin A in Sapp1p and Sapp2p	45
3.13	Virtual glycine scan of Sapp1p/pepstatin A and Sapp2p/pepstatin A	46
3.14	Carborane-based inhibitors of CAII	50
3.15	The active site of CAII	51
3.16	Rotamer preferences of <i>closo</i> -carborane-based inhibitor	53
3.17	Details of binding of <i>closo</i> - and <i>nido</i> - inhibitors in CAII	54
3.18	Virtual glycine scan of CAII/1a and CAII/7a	56
3.19	Comparison of binding of 1a in CAII and CAIX	59
3.20	Virtual glycine scan of CAII/1a and CAIX/1a	59
3.21	Lower boundary of normalized scores	65
3.22	The performance of 9 scoring functions for 4 protein-ligand systems	66

Chapter 1

Introduction

Drug discovery is the process through which potential new medicines are identified. Bringing a drug to the market is still very demanding task which nowadays costs more than 1 billion USD and takes over 10 years.¹ The whole process consists of many stages: from an identification of drug candidates (or so called leads) by chemical synthesis, purchase, curation and biological screening; through an optimization process increasing lead affinity, selectivity, efficacy and metabolic stability; to complex toxicity studies in two animal species and three phases of human clinical trials. The long path to a drug is not only very expensive but it also carries an extremely high risk of failure. The use of various *in silico* techniques can help avert those failures in all mentioned pre-clinical phases and it becomes more and more popular thanks to enormous advances in software and hardware computational power. It is obvious that computer-aided drug design (CADD) can also significantly minimize time and cost requirements of drug development.²

The very first step is an identification of a biomolecular target involved in a particular pathway associated with a studied disease. The target is mostly an enzyme, transmembrane receptor, ion channel or a piece of nucleic acid. Then regulators of the target are identified by relevant biological assays, such as for enzyme inhibition or modulation of an intracellular process. The most active compounds (so called hits) arise from large libraries, often small organic molecules representing the largest class of marketed drugs.³ In the last few decades rational approaches are often used in this stage.

“Ligand-based design” is used when the target has known substrates or inhibitors. Then compound libraries are screened by pharmacophore models using similarities in structure and/or properties.⁴

On the other hand, if a crystal structure of the target is known, the knowledge of the binding site is used in the “structure-based design”. A virtual high throughput screening (vHTS) is an *in silico* equivalent of biological screening of compound libraries that uses a three-dimensional structure of the target molecule. Here, docking algorithms predict geometries of complexes constituted by the target molecule and library compounds and

scoring functions predict the binding affinities. Another approach is “*de novo* design” based on local optimization, where novel structures are built up in the binding site from small molecular fragments or single atoms in stepwise manner.⁵ Computational tools are not only important for the identification of hits but also for a selection of modifications that would improve the potency and other properties of lead compounds and also for a bringing of deeper insight into the mechanism of their action.

The majority of drugs act through a competitive inhibition to their biological targets. The most common case is a reversible non-covalent binding of a ligand to the active site of a protein, thus preventing a native substrate from entering the site. A basic view of the non-covalent binding offers the lock-and-key model introduced by Emil Fischer in 1894⁶ which was later on superseded by more adequate concept of the induced fit.⁷ Non-covalent interactions (known also as Van der Waals forces) play a crucial role in the stability, structure and functionality of biomolecules and also in the fields of supramolecular chemistry and nano-materials.⁸ They are weak but multiple forces, acting at distances from units of Å to several nanometers, that are not only fundamental for the existence of liquid states of matter and molecular clusters but also essential for nothing small than life itself.⁹ The most important representatives are hydrogen bonds and electrostatic forces, but they are comprised also of nonspecific stacking interactions and highly specific interactions such as sigma-hole or dihydrogen bonding. They are usually complicated and combine multipole electrostatic interactions, polarization, dispersion and also charge-transfer interactions. The importance of non-covalent interactions for biomolecules has been known and studied for a long time.¹⁰

The free energy of binding (ΔG_b) between the protein and the ligand, which is related to the dissociation equilibrium constant (K_i) or IC_{50} value of the protein-ligand complex (Eq. 1.1), is expected to be proportional to the ligand potency.

$$\Delta G_b = RT \ln K_i \quad (\text{Eq. 1.1})$$

, where R is the gas constant and T is the absolute temperature.

In the case of known weakly-bound competitive inhibition, IC_{50} value is related to K_i by the Cheng and Prusoff equation¹¹ (Eq. 1.2), whereas in the case of tight-binding inhibitors, the enzyme concentration $[E]$ must be considered (Eq. 1.3)

$$K_i = IC_{50}/(1+[S]/K_M), \quad (Eq. 1.2)$$

$$K_i = (IC_{50}-[E]/2)/(1+[S]/K_M) \quad . \quad (Eq. 1.3)$$

, where [S] is the concentration of the substrate and K_M is the Michaelis constant.

Binding affinities can be determined experimentally by modern biophysical methods like isothermal titration calorimetry (ITC)¹² or surface plasmon resonance (SPR).¹³

Binding event can range from exothermic or entropy driven spontaneous process, depending on the interplay of an enthalpic (ΔH) and an entropic term ($-T\Delta S$) in a negative binding free energy (Eq. 1.4).

$$\Delta G_b = \Delta H - T\Delta S \quad (Eq. 1.4)$$

There are many phenomena that contribute to the total binding free energy of a protein-ligand complex, among these the most important are hydrogen bonds, dispersion and charge transfer interactions, halogen bonding, the deformation and desolvation energies, conformational entropy, and vibrational/rotational entropy. In order to correctly describe ΔG_b computationally, the methods must be able to reliably describe all these contributions as accurately as possible and deal with a sufficient large parts of the system (thousands of atoms) within a reasonable time.

A computational arsenal for estimating the free energy of protein-ligand binding varies from statistics-based methods (reviewed in Ref.14) to physical chemistry-based approaches.

Molecular mechanics (MM) methods based on classical-physics approximations are the most suitable for solving large protein molecules. This well established area was pioneered by Nobel laureates Martin Karplus, Michael Levitt and Arieh Warshel. Aside from routinely used molecular dynamic simulations (MD) and other extensions¹⁵, various methods for affinity predictions exist, ranging from pathway methods (such as free energy perturbation, or thermodynamic integration), through linear response approximation (LRA)-based methods (such as a linear interaction energy method¹⁶ to widely used MM-GB/PBSA methods¹⁷ for example coupled with MD sampling. The first class of perturbative methods is rather used for lead optimization than for virtual screening, because it requires extensive ensemble sampling for obtaining converged free energy values.¹⁸ The second class of methods is system dependent and relies on the quality of the calibration test for determining of adjustable parameters, however the setup of

separation of the binding into Van der Waals and electrostatic parts allows the method to obtain absolute binding free energies.¹⁶ The latter class of methods is able to predict reliably the relative binding affinities¹⁹ but it is dependent on the quality of used implicit solvent models and MM forcefields and also on the converged sampling.²⁰ Besides free energy calculations of protein-ligand binding, all-atom molecular dynamics simulations are used today to study motions of macromolecules and processes by which drugs bind to receptors. Coarse-grained simulations extend the range of problems that can be studied by achieving longer, more biologically relevant timescales²¹ Several approaches, for example metadynamics²² or accelerated MD²³, aim to accelerate sampling of protein conformational states.

In contrast of widely used molecular mechanics, quantum mechanical (QM) methods are by the definition able to reliably describe non-covalent interactions and bond breakage/formation. Quantum mechanics offers proper description of quantum effects such as proton and charge transfer, many body effects, polarization or σ -hole bonding and covers the vast of organic and inorganic chemical space without a need of ligand-dependent parameters. Unfortunately, the exact solutions of the Schrödinger equation is limited to very small systems and on other hand non-covalent interactions generally involve hundreds of atoms and are inherently very complex. The proper treatment thus must find the best balance between computational feasibility and accuracy. However computational cost usually escalates with increased levels of theory, QM have been featured among CADD methods more often with tremendously increase of computer power in last decades

Details of the implementations of QM approach in CADD are well reviewed in the literature²⁴⁻²⁸, so below I only give a brief overview.

There are various efforts toward the improvement of biomolecular force fields, such as AMOEBA²⁹ or QMPFF polarizable forcefields³⁰, using validation against quantum mechanics (QM) data.³¹

“Ligand-based design” standardly contains the use of quantum mechanically derived descriptors in quantitative structure-activity relationships (QSAR), for example atom-centered partial charges, characteristics from atoms-in-molecules (AIM) or other topological indeces.^{32,33} These efforts continue with more classes of QM-derived descriptors for probing protein–ligand complexes such as molecular electrostatic potential (ESP) maps, frontier orbital analysis, density of states, local hardness and softness or Fukui indices.³⁴⁻³⁶

QM methods in the „structure-based design“ are often used in a refinement process in X-ray or NMR structure determination³⁷⁻⁴², nevertheless their main strength lays in an accurate prediction of binding affinities. This is the ‘the holy grail’ of drug design and there is no doubt that the applications of QM will rise among docking and scoring approaches. It is known that the prediction of bound geometry of ligands to a given protein active site is reasonably accurate (with RMSD between X-ray and docked pose below 2Å), however limitations of commonly used scoring functions have been exposed.⁴³ Still docking results could be improved by re-parameterization of scoring functions or via direct inclusion of some QM-based information to model non-covalent interactions more correctly.⁴³⁻⁴⁵ Some full QM or QM/MM-based docking approaches for example use QM-derived charge models⁴⁶⁻⁴⁸ or include polarization effects⁴⁹ to improve geometry predictions. Despite their high computational cost, QM methods can also improve the quality of prediction of docking poses.⁵⁰⁻⁵² On the other hand, knowledge-based, empirical or force field-based scoring functions give poor results in ranking different ligands according to their affinity. However reliable QM estimation of the free energy of protein-ligand binding is limited by the size of the system, it can be solved by the use of hybrid QM/MM approaches (reviewed in 27, 52-55) or various fragmentation schemes⁵⁶⁻⁵⁸, DFT-D3 on truncated protein-ligand complexes⁵⁹ and linear-scaling or efficient parallelization of corrected semiempirical quantum mechanical (SQM) methods.⁶⁰⁻⁶³

Semiempirical QM-based scoring function (QMScore) was firstly introduced by Kenneth Merz group, by using AM1 method augmented with empirical dispersion and combined with Poisson-Boltzmann implicit solvent model.⁶⁴ Authors showed a superior performance of the QMScore over other scoring functions in the case of metalloprotein-ligand binding⁵⁰ but further corrections, especially for hydrogen bonding and dispersion, were needed.^{65,66}

To this end, Pavel Hobza’s group has taken more systematic approach. Firstly, based on comparison with high-level QM calculations on small model systems of non-covalent interactions, the PM6 SQM method (which is valid throughout chemical space⁶⁷ and does not require parameterization for each new system) have been carefully selected and parameterized to describe dispersion as well as hydrogen and halogen bonding reliably and accurately.⁶⁸⁻⁷⁰ Similarly, several implicit solvent models *e.g.* MM-based (PB or GB⁷¹) and QM-based (COSMO⁷² or SMD⁷³), have been carefully compared.⁷⁴ These methods are therefore used within the SQM-based scoring function⁷⁵, in which the binding free energy

is approximated by the total score expressed by Equation 1.5.⁷⁶ Particular terms describe the gas-phase interaction energy (ΔE_{int}), the change of solvation free energy upon complex formation ($\Delta\Delta G_{solv}$), the change of conformational “free” energy ($\Delta G'_{conf}{}^w$) and the change of entropy upon ligand binding ($-T\Delta S$).

$$Score = \Delta E_{int} + \Delta\Delta G_{solv} + \Delta G'_{conf}{}^w(P) + \Delta G'_{conf}{}^w(L) - T\Delta S_{int} \quad (Eq. 1.5)$$

Its generality has been demonstrated in various non-covalent protein-ligand complexes⁷⁷⁻⁸⁰ and moreover it was extended to treat covalent inhibitor binding.⁸¹

The aim of this thesis is to show the ability of some applications of QM-based approaches to contribute hand-in-hand with experiments to the CADD. The thesis consists of 8 original papers published in international peer-reviewed Journals (attached in Appendices) and an accompanying text that aims to emphasize outcomes of individual papers linking them into the complex work. It is organized as follows: All computational methods essential for our work are summarized in Chapter 2, while the following chapter covers individual projects. The first part of Chapter 3 explores accurately the strength and origin of the stabilization for σ -hole bonded model systems by high level QM methods, going from halogen- through chalcogen- to pnicogen- bonding. The second part is devoted to various protein-ligand complexes and shows the capability of QM methods to unveil the features of the structure which are not accessible to the crystallographic experiments. The last part introduces an effective SQM-based tool for virtual screening that was tested together with standardly used scoring functions on different protein-ligand systems. Chapter 4 summarizes the work with some final remarks.

Chapter 2

Methods

There are a wide variety of computational methods that can be used to treat intermolecular complexes. If all kinds of non-covalent interactions are to be reliably calculated, a detailed description of the electron distribution should be used. This section summarizes QM-based methods with the different relative accuracy/computational cost performance that determines their use throughout our work. It comprises particularly from highly accurate CCSD(T)/CBS calculations used in model systems as benchmark data, through suitable SAPT perturbative schemes for a decomposition of the interaction energy, to fast corrected DFT calculations and SQM methods used in protein-ligand studies.

It is clear that all these QM-based methods notwithstanding their merit in calculations of non-covalent interactions are limited by the size of the studied system. On the other hand molecular mechanics approach can easily represent even very large biomolecules using approximated all-atom representations and force field description of the potential energy of the studied system. Therefore to speed up the calculations especially in the studies of protein-ligand complexes, generally comprising thousands of atoms, a combination of QM and MM methods coupled with various implicit solvent models is used in a hybrid scheme.

2.1 Supermolecular Interaction Energy

In general, the interaction energy is caused by an interaction between the objects being considered. In the supermolecular approach, the total many-body interaction energy is defined as a difference between the energies of the complex and its isolated subsystems (Eq. 2.1). This is applicable to any type of molecular clusters.

$$\Delta E_{\text{int}} = E(A_1, A_2, A_3, \dots, A_N) - \sum_{i=1}^N E(A_i) \quad (\text{Eq.2.1})$$

The interaction energy for non-covalently bound binary system (A ... B) is showed in Eq.2.2.

$$\Delta E_{\text{int}}(A \cdots B) = E(A \cdots B) - [E(A) + E(B)] \quad (\text{Eq.2.2})$$

,where $\Delta E_{\text{int}}(A \cdots B)$ is the interaction energy for the complex, $E(A \cdots B)$ stands for the total electronic energy of the complex and $E(A)$, $E(B)$ are electronic energies of the monomers.

The final interaction energy is much smaller than total electronic energies from which it is derived. Basis set superposition error (BSSE) is the main disadvantage of the supermolecular approach and it is due to unequal description of supersystems and subsystems. The supersystem uses functions of both subsystems (contrary to subsystems which use only own functions) and its energy is due to variation principle too negative. The BSSE could be eliminated *a posteriori* - by a counterpoise (CP) correction scheme of Boys and Bernardi⁸², where the BSSE is calculated by re-performing all calculations with mixed basis sets using dummy atoms, and *a priori* by using the chemical Hamiltonian approach introduced by Mayer.⁸³

In the QM, the interaction energy can be expressed as the sum of the Hartree-Fock (HF) interaction energy and correlation interaction energy (Eq. 2.3).

$$\Delta E_{\text{int}} = \Delta E^{\text{HF}} + \Delta E^{\text{corr}} \quad (\text{Eq.2.3})$$

Therefore, in order to obtain interaction energies, all the electronic energies should be calculated with the highest accuracy by using sufficiently large basis set and by covering the major part of correlation energy. Incompleteness of basis sets can be solved by an extrapolation to the complete basis set limit (CBS). Different speed of the convergence of HF and correlated interaction energies cause that both terms can be extrapolated separately. Several extrapolation schemes are well-documented, for instance Helgaker's^{84,85} or Truhlar's⁸⁶ schemes in Eq. 2.4 and Eq. 2.5, respectively:

$$E_X^{HF} = E_{CBS}^{HF} + A \exp(-\alpha X) \quad \text{and} \quad E_X^{corr} = E_{CBS}^{corr} + BX^{-3} \quad (\text{Eq.2.4})$$

$$E_X^{HF} = E_{CBS}^{HF} + BX^{-\alpha} \quad \text{and} \quad E_X^{corr} = E_{CBS}^{corr} + BX^{-\beta} \quad (\text{Eq.2.5})$$

,where E_x and E_{CBS} stand for energies for the basis set with the largest angular momentum X and for the CBS respectively; A is a pre-exponential factor, B is a pre-power factor; α and β are parameters fitted in original works. The two point extrapolation form is preferable, using Dunning's augmented or non-augmented basis sets which have been constructed to converge systematically into the CBS limit.

Kim *et al.* developed quite different kind of extrapolation for interaction energies using a least biased scheme.⁸⁷ The method uses the fact that both BSSE-corrected and – uncorrected interaction energies give the same CBS limit. This asymptotic value based on extrapolation can thus be considered as pseudo-interpolation in terms of energies because the CBS energy is between BSSE-corrected and BSSE-uncorrected values. Thus it is possible to use data from different basis sets and the CBS value is obtained without any predetermined parameter.⁸⁸

It is well known that the role of both terms in Eq. 2.3 is different for different type of non-covalent interactions, for instance ΔE^{HF} is more important in hydrogen bonding, whereas ΔE^{corr} is essential for stacking interactions formed mainly by dispersion (pure correlation) effect. Generally, it is of vital importance to estimate the correlation energy as accurate as possible, however it is very demanding task.

A coupled-cluster technique with a complete basis set description (CCSD(T)/CBS) is widely accepted as the “golden standard” for the accurate calculation of interaction energies for non-covalent complexes. The application of this method is very limited because of the high computational cost, so much research over the past decades has been concerned with the development of other methods capable of accurate determining of interaction energies for larger biological structures. Standard QM methods such as MP2, MP3, CCSD, or DFT fail to describe various types of non-covalent systems with comparable accuracy. Therefore some approximate methods must be used. These methods have been usually parameterised towards non-covalent interactions, that requires a sufficient amount of accurate benchmark data, such as Truhlar's database NCIE53,⁸⁹

Grimme's GMTKN30⁹⁰ or several datasets produced in Pavel Hobza's group accessible online on www.begdb.com.⁹¹⁻⁹⁵

2.1.1 Coupled Cluster Theory

Coupled cluster (CC) theory as a very accurate method for calculation of the correlation energy in atoms and molecules was introduced by Čížek, Paldus and Barlett.⁹⁶⁻⁹⁸ The wave function is constructed from a reference Slater determinant via an exponential formula of an operator expanded into clusters of excitation operators. CC methods are systematically improvable and also classified by inclusion of a higher number of excitations allowed in the definition of the cluster operator. The abbreviations usually starts with CC letters, followed by S, D, T and Q for allowed single, double, triple and quadruple excitations. Terms calculated non-iteratively using perturbation theory are indicated by round brackets.

As already mentioned, accurate interaction energies for non-covalent complexes are generated by the CCSD(T) method calculated with a sufficiently large basis set known as the „golden standard“ in quantum chemistry, which covers single- and double electron excitations iteratively and triple excitations perturbatively in the fourth order. Further, one fifth-order term is here also included.

The benchmark CCSD(T)/CBS interaction energy^{88,99,100} is defined as follows (see Eq. 2.6):

$$\Delta E_{CBS}^{CCSD(T)} = \Delta E_{CBS}^{HF} + \Delta E_{CBS}^{MP2corr} + \left(\Delta E^{CCSD(T)} - \Delta E^{MP2} \right)_{basis.set}^{smaller} \quad (Eq.2.6)$$

,where ΔE^{HF} is the Hartree-Fock interaction energy and $\Delta E^{MP2corr}$ is the correlation interaction energy calculated at MP2 level, both extrapolated to the CBS limit; and $(\Delta E^{CCSD(T)} - \Delta E^{MP2})$ is so called the $\Delta\Delta E^{CCSD(T)}$ correction term calculated as a difference between interaction energies at the CCSD(T) and MP2 level often calculated in one (small) basis set only. The accuracy of this multi-level approach mainly depends on the size of the $\Delta\Delta E^{CCSD(T)}$ correction term and the quality of the small basis set, e.g the interaction energy for dispersion-dominated non-covalent complexes is the error about 3-5%.^{101,102}

The CCSD(T)/CBS procedure provides interaction energies with chemical accuracy (with error less than 1 kcal/mol). However it has the best accuracy/cost ratio, the scaling of

the method is N^7 (where N is a total number of orbitals), so its use is still very limited. It is widely used in a generation of reference datasets, nevertheless the biggest systems calculated up to now by CCSD(T)/CBS have about 70 atoms.^{103,104}

2.1.2 Density Functional Theory

Density functional theory (DFT) offers an alternative point of view on the electronic structure of atoms and molecules. Energy of the molecule is a function of spatially dependent electron density, defined as a functional. Kohn-Sham DFT¹⁰⁵ is now the most used *ab initio* method for electronic structure calculations in condensed matter physics and quantum chemistry, reasonably providing accurate properties of various molecules and solids. The main drawback of commonly used (LDA, GGA or hybrid) density functionals is inability to describe ubiquitous attractive long-range electron correlations. Much research is thus focused on the development of approximate DFT approaches that are able to model very important dispersion interactions (for example meta-hybrid functionals, special correlation or orbital-based DFT methods and DFT/MM-based hybrid methods, reviewed in Ref. 106-108. The most promising approaches in point of view of computational speed and robustness add empirical dispersion correction of the form $-C_6 R^{-6}$ to existing functionals. It must be noted that existing functionals account for some medium-ranged dispersion effects. Therefore atom-atom pairwise empirical corrections (see Eq.2.7) differ not only in methods for derivation of C_6 coefficients but also in formulas of damping functions that are important to avoid electron correlation double-counting effects for small intermolecular distances.

$$E_{disp}[\rho] = -\sum_{i \leq j} \sum_{n \geq 6} C_n^{ij}(\rho) R_{ij}^{-n} f_{damp}^n(R_{ij}, \rho) \quad (Eq.2.7)$$

,where the first sum is over all atom pairs in the system, C_n^{ij} stands for dispersion dependent averaged n^{th} -order dispersion coefficient (orders $n=6, 8, 10, \dots$) for atom pair ij , R_{ij} is their internuclear distance and f_{damp} is the dispersion-dependent damping function.

In 2004 Grimme *et al.* introduced the very first simplified version of empirical dispersion correction for DFT.¹⁰⁹ This DFT-D1 approach includes only one term ($n=6$)

from the expansion in Eq.2.7 and completely system independent dispersion coefficients C_6^{ij} and the damping function with two empirical parameters. Despite this simplification, extensive fitting procedure of a new global scaling parameter (s_6) and other empirical parameters helps with satisfying agreement in description of various weakly bound complexes. In the second generation of empirical dispersion correction, DFT-D2, Grimme *et al.*¹¹⁰ reparameterized B97 functional and brought much less empiricism into the derivation of C_6^{ij} coefficients. Recently introduced geometry dependent DFT-D3 approach¹¹¹ has been completely revised and offers up to now the best performance for different types of non-covalent complexes. In contrary of previous approaches, DFT-D3 is less empirical, C_6 terms are no longer scaled, the higher C_n ($n=8$) terms are used and all parameters, e.g. cutoff radii and dispersion coefficients) are computed from first principles by TD-DFT method. No atom connectivity information is thus required. The method employs the damping function from Chai and Head-Gordon¹¹² which includes the most important order-dependent scaling factor firstly introduced by Jurečka *et al.*¹¹³ Furthermore, the method is robust and very fast and applicable to all elements of the Periodic Table, achieving the error of binding energies with respect to CCSD(T) values mostly only about 10%.¹¹⁴

Similarly, in 2007 Jurečka *et al.* presented the pair-wise empirical dispersion correction for DFT method.¹¹³ In contrast of Grimme's first generation of dispersion correction, Jurečka's approach (DFT-D) uses the damping function of Fermi type, where C_6 coefficients are not scaled. As mentioned above, this damping function includes order-dependent scaling factor of the cutoff radii and thus adapts the correction at small and medium distances to a specific form of the chosen density functional. This type of damping function has been adopted in DFT-D3 method.

2.1.3 Semiempirical Quantum Mechanical Methods

Semiempirical quantum mechanical (SQM) methods introduce many approximations to HF formalism by using only valence electrons explicitly, omitting some integrals and introducing empirical parameters. The fact that SQM methods cover quantum effects due to the quantum mechanical base makes them more preferred over fully empirical MM methods. Other advantage is their application to any system without a need of any input

parameters. SQM methods are fast by definition so they can be in principle applied on extended systems, but are not suitable for calculation of non-covalent interactions. The reason is that the reasonable description of some components of the interaction energy like hydrogen bonding or dispersion is either incorrect or not included at all. In order to solve this problem reparameterization or other empirical corrections are needed. Several attempts have been made to improve description of dispersive contributions or hydrogen bonding, e.g. PM3-D¹¹⁵, OMxD¹¹⁶, AM1-FS1¹¹⁷ or PM3-PIF¹¹⁸ and PM3-PDDG¹¹⁹ however none of these methods is accurate enough to describe different non-covalent complexes. Another very promising approach similar to the traditional SQM methods is self-consistent charges density functional tight binding (SCC-DFTB).¹²⁰ Here, all the parameters are derived from full DFT calculations, which makes the method more robust and often also more accurate.

Stewart *et. al* introduced the NDDO-based PM6 method in 2007.⁶⁷ The PM6 is based around the earlier AM1 formalism, but differs in a method of parameter optimization (for 70 elements) and in a modification of core-core interaction term. Although the method has brought substantial improvements over its predecessors, it still lacks the ability to describe van der Waals systems.

Recently, several corrections for non-covalent interactions have been developed in our laboratory. The first generation of the hydrogen-bond correction is the function of the distance, angle and partial charges of hydrogen bonded atoms. Together with a parameterization of the Jurečka's dispersion correction¹¹³ the resulting PM6-DH method⁶⁸ achieves a good accuracy in small model systems.

The second generation of the corrections (-DH2)¹²¹ aims to improve the previous version by avoiding double-counting of the dispersion energy already described in PM6 and by fixing of discontinuities of the potential in hydrogen bond correction. It results in a higher accuracy especially for hydrogen-bonded systems.

In the third generation (-DH+)¹²², the dispersion correction is the same as in the -DH2 and other modifications of the hydrogen bond correction involved donor-acceptor and partial charges issues. The biggest disadvantage of the latter two methods is a lack of smooth first derivatives of potential energy surface that makes them inapplicable in geometry optimizations of some systems, moreover the PM6-DH+ method systematically underestimates hydrogen bonding interactions in charged systems.

Hence the final version of corrections (-D3H4) has been proposed, which solves all issues encountered in previous generations.⁷⁰ It adopts the Grimme's dispersion correction

term¹¹¹, improving the robustness of the method and completely redesigns the hydrogen bond correction, *e.g* by simplifying of its form and by scaling for charged systems. Yielding not only a smooth potential energy surface but also its derivatives, the new corrections enable geometry optimizations and molecular dynamics and also naturally describe proton transfer along hydrogen bond.

The PM6 method cannot describe halogen bonding properly, because it uses only subminimal basis sets and suffers from the lack of repulsion. To remedy this problem, a simple repulsive correction for halogen bonding (-X) was recently presented.⁶⁹

The resulting PM6-D3H4X method has become the most accurate SQM method for the description of biomolecular systems reaching the chemical accuracy (error of 1 kcal/mol) and so outperforming by its speed DFT-D or MP2 methods.⁷⁰ Its performance is not even overcome by a newly developed PM7 method, which adopted a much of above mentioned conception.¹²³ Moreover PM6-D3H4X combined with a linear scaling algorithm, such as the localized orbital method MOZYME¹²⁴, makes now possible calculations on entire proteins with several thousands atoms routinely.

2.2 Intermolecular Perturbation Theory

Apart from the supermolecular approach, intermolecular interactions bearing dipole-dipole interactions, hydrogen bonding and London forces are most naturally accounted by Rayleigh–Schrödinger perturbation theory.¹²⁵ In this theory, the unperturbed Hamiltonian is defined as the sum of monomers Hamiltonians and the perturbation consists of all interactions between monomers. In London’s method the interaction energy between two monomers is represented by its multiple expansion and the convergence depends on the intermonomer distance. Thus the method is valid only for monomers with fully localised electrons.¹²⁶⁻¹²⁹

This convergence drawback has been overcome by symmetry-adapted perturbation theories (SAPT) introducing intermolecular symmetry projections into appropriate places of energy and wavefunction expansions.¹³⁰⁻¹³² The symmetrized Rayleigh–Schrödinger scheme implemented for many-body system leads to feasible equations and gives reliable results.¹³¹ The SAPT interaction energy is the sum of physically meaningful terms – electrostatic, induction, exchange-repulsion and dispersion energy contributions - and it is

by definition free of the BSSE. These all make the SAPT to be a good option to calculate intermolecular interaction energies and moreover to interpret the nature of binding.

2.2.1 Density Function Theory based Symmetry Adapted Perturbation Theory (DFT-SAPT)

The substantial improvement of the original SAPT is introduced by the combination of DFT method and the perturbation theory in DFT-SAPT method.¹³³⁻¹⁴⁰ This approach accelerates the calculations by one order of magnitude and allows for the treatment of extended complexes (up to 40 atoms). In DFT-SAPT monomers are described in terms of Kohn-Sham orbitals and orbital energies as well as of TD-DFT response functions, whereas intermolecular interactions are solved as the perturbation. The intramolecular treatment needs some corrections, because it is conducted by DFT and so suffers from inaccurate energies of virtual orbitals. This is solved in advance by the gradient-controlled shift procedure¹³⁷, which uses the difference between the exact vertical ionisation potential (IP) and the HOMO energy.

The total interaction energy in the DFT-SAPT is given as the sum of the first- (E^1) and second-order (E^2) perturbation energy terms and a δHF energy terms. (Eq.2.8). The former two terms represent: polarization (E^1_{Pol}), induction (E^2_{Ind}) and dispersion (E^2_{Disp}) together with the exchange-repulsion terms (E^1_{Ex} , E^2_{Ex-Ind} and $E^2_{Ex-Disp}$) and δHF term represents higher than second-order terms covered by the Hartree-Fock approach.

$$E_{int} = E^1_{Pol} + E^1_{Ex} + E^2_{Ind} + E^2_{Ex-Ind} + E^2_{Disp} + E^2_{Ex-Disp} + \delta HF \quad (Eq.2.8)$$

DFT-SAPT decomposition of the interaction energy helps to qualitatively understand non-covalent bonding. The first-order polarization energy (E^1_{Pol}) comes from unperturbed interactions between two charge distributions and is indeed equal to the electrostatic interaction. The second-order induction term (E^2_{Ind}) arises from the polarization of one charge distribution by the electric moment of the other one and it is derived from coupled-perturbed Kohn-Sham equations. It should be mentioned here that the induction energy contains not only the classical induction term but also the charge-transfer energy from the electron donor to electron acceptor. The dispersion energy (E^2_{Disp}) is computationally the most demanding term derived from frequency-dependent propagators obtained from

TD-DFT. The exchange interaction energy (E_{Ex}^I) stems from Pauli or anti-symmetry principle and is proportional to the different overlap between monomer orbitals. It is strongly repulsive, short range and responsible for the volume of the molecule. The second-order E_{Ex-Ind}^2 and $E_{Ex-Disp}^2$ terms are approximated by scaling their counterparts. In principle the method would be exact for all energetic contributions of the interaction energy (asymptotically for exchange terms) if the DFT description of the monomers was exact. However, it has been shown that DFT-SAPT using the localized and asymptotically corrected LPBE0AC exchange-correlation functional and at least aug-cc-pVDZ basis set provides satisfyingly accurate results of various non-covalent systems.^{136-138,141}

Additionally, DFT-SAPT can be significantly accelerated by density fitting^{138,140} that lowers down the scaling of the method from N^6 to N^5 . Even more acceleration is achieved by hybrid a DFT-SAPT approach introduced by Hesselmann.¹⁴² The most computationally demanding dispersion terms are modelled here by adapted Grimme's empirical correction with the adjusted damping function. All used parameters were fitted toward CCSD(T)/CBS data for S22 dataset to achieve a good accuracy of the correction for various non-covalent interactions.

2.3 Hybrid QM/MM Approach

Quantum mechanics/molecular mechanics (QM/MM) approach is an embedding scheme that combines the strengths of both calculations: the accuracy of QM and the speed of MM. This hybrid scheme was firstly introduced by Nobel laureates Warshel and Levitt in 1976.¹⁴³ The QM/MM speeds up the calculations significantly and thus allows to study protein-ligand interactions reliably, when the the active site of the protein, representing the most important part of the system, is treated quantum mechanically and a remaining part of the system is treated classically by MM. Water environment is often approximated by a combination of implicit solvent models with important structure water molecules treated explicitly. In practice, the hybrid scheme is not restricted to two layered QM/MM case but it can also combine more than two levels of theory, e.g. QM together with more modest-cost QM method, QM/SQM, QM/SQM/MM *etc.* Hybrid schemes also differ in the way of solving boundaries between layers, which depends on the nature of studied systems. Total energy of the studied (E^S) system is within QM/MM defined by an additive scheme showed in Eq. 2.9:

$$E^S = E_{QM}^I + E_{MM}^O + E_{coupling}^{I-O} \quad (Eq.2.9)$$

,where E_{QM}^I is the energy of the inner part of the system (e.g. the active site of the protein) calculated by quantum mechanics, E_{MM}^O is the energy of the rest of the system (the outer part) solved by molecular mechanics and the $E_{coupling}^{I-O}$ is a coupling term between both parts. The interaction between QM and MM parts of the studied system is in general described by a mechanical and/or electrostatic embedding. In the mechanical embedding, change of the geometry of both parts is mutually dependent. Within the electrostatic embedding the potential of the outer part affects the inner part and on the contrary the charge redistribution of the inner part influences the outer part.

In the protein-ligand systems, boundaries between both parts are mostly defined by cutting across covalent bonds of the amino acid chains. The saturation of dangling bonds of the inner part is in this case usually provided by using hydrogen atoms as link atoms. It is thus more practical to avoid the calculations of the outer part without the inner part. This is allowed by a subtractive ONIOM (our own n-layered integrated molecular orbital and molecular mechanics) approach.¹⁴⁴ The ONIOM potential (Eq. 2.10), allowing to calculate the interaction between two layers at the low level of theory, is expressed as:

$$E^{ONIOM} = E_{QM}^{Model} + E_{MM}^{Real} - E_{MM}^{Model} \quad (Eq.2.10)$$

,where E_{QM}^{Model} is the energy of the inner part of the system containing link atoms (model system) estimated at QM level, E_{MM}^{Model} is the energy of the same model system calculated by MM and E_{MM}^{Real} is the energy of the full system calculated at MM level. However the border region is artificial in the sense of local introducing of link atoms or by neglecting the short-range interactions with the outer part of the system. The effect of the border region is not reflected in the final energy because it is subtracted in the terms of model system in Eq.2.10.

2.4. Solvation Models

Most chemical processes take place in different solvents, thus to cover the environment into calculations is an inevitable task. Very important especially for thermodynamic considerations is an estimation of the solvation free energy, which is the net energy change upon transferring the molecule from the gas phase into the solvent with which it equilibrates. The solvent can be modeled by several different approaches, such as by explicit solvent molecules, implicit solvent models and a hybrid model combining continuum with explicit solvent.

In the first case the solvent molecules are treated explicitly and it is the most realistic description of solvation, because all specific interactions with a solute are covered. The use of this method is however very limited in QM because it considerably increases the computational requirements. In practise the solvent is often treated at lower level of theory in QM/MM approach, e.g. replacing their actual electron distribution with partial charges, thus only calculating their electrostatic influence on a solute.

In implicit solvent models, the solvent is approximated by homogeneously polarisable continuum characterised by the dielectric constant. This constant is responsible for defining the level of polarisability of the solvent. Implicit models use cavities to exclude the solvent and into which a solute can be inserted. When the solute charge distribution meets continuum dielectric field at the surface of the cavity, the polarization on the solute is changed. The response of solute charge distribution can be modeled by the reaction potential in QM or by partial atomic charges in MM. This approach often gives good results of equilibrium solvation energetics and it is used for estimation of pKs or redox potentials. The implicit solvent models evaluate the solvation free energy (see Eq. 2.11) as a sum of its changes due to the mutual polarization of solute and solvent, differences in solute-solvent dispersion and repulsion interactions and lastly the energy required to create a cavity - the cavitation free energy.¹⁴⁵

$$\Delta G_{solv} = \Delta G_{pol} + \Delta G_{disp} + \Delta G_{rep} + \Delta G_{cav} \quad (Eq.2.11)$$

However electrostatic interactions of small molecules can be estimated by exact Poisson-Boltzmann equations, its use for bigger systems is too much expensive. Therefore some simplifications have arisen.

The *Generalized Born* (GB) implicit solvent model⁷¹ is an approximation to the exact linear Poisson-Boltzmann equations, where the electric field is approximated by Coloumb field model. A solute is modeled as a set of spheres whose internal dielectric constant differs from the external solvent. Continuous charge density is thus replaced by a set of atom-centered partial charges. It introduces errors, for example poorly described local charge distribution around atoms with lone electron pairs. The accuracy of the method depends on a level of the computation of meaningful partial charges. The main advantage of the pair-wise GB model is that the result is analytic and so the forces can be evaluated quite rapidly. It is therefore often used in big molecules like proteins where the solvation evaluation represents a large portion of the overall computation time.

Conductor-like screening model (COSMO)⁷² solves the non-homogeneous Poisson equation by employing a scaled-conductor approximation. The cavity is here defined by series of atom-centered spheres, with usually a bit larger radii than standard Van der Waals radii, augmented with some auxiliary spheres if necessary. The surface of this cavity based on not overlapping spheres is partitioned into segments, *e.g.* triangles, and each of this segment is assigned a adjustable point charge. These surface-point charges are determined in the SCF procedure, from the charge density and corresponding potential of the solute, and the electrostatic equations assuming that the solvent is perfect conductor, *i.e.*, vanishing potential on the cavity surface. The dielectric constant of the real continuum then defines scaling factor for computed energies. The model is thus more accurate for solvents with a higher permittivity like water. Moreover energy derivatives can also be calculated, so geometry optimization or harmonic frequencies are available within this model. It is often used in combination with DFT or SQM methods for a reliable description of solvation effects in protein-ligand complexes.

Chapter 3

Projects

3.1 Nature of σ -hole Bonding

The following section presents the first topic of this thesis, particularly studied in three publications that are attached in Appendices A, B and C. It focuses on nonclassical non-covalent bonding. The bonds are referred to as halogen, chalcogen and pnictogen bonds or in general σ -hole bonds.

A typical σ -hole bond (Figure 3.1) occurs between a Lewis acid and a Lewis base where the Lewis acid is an halogen, chalcogen or pnictogen atom and the Lewis base is an electron donor, *i.e.* a chemical group having a lone electron pair or aromatic π -electrons. However the most electronegative elements such as halogen atoms that are usually considered to be negative when they are covalently bonded to other atoms, are expected to interact only with positive sites, their enigmatic interactions with negative sites had been reported since 1950's.¹⁴⁶ Their origin was that time referred to charge transfer¹⁴⁷ and the concept of sigma-hole bonding^{148,149} had began to be used later on with the knowledge of their electrostatic origin.¹⁵⁰

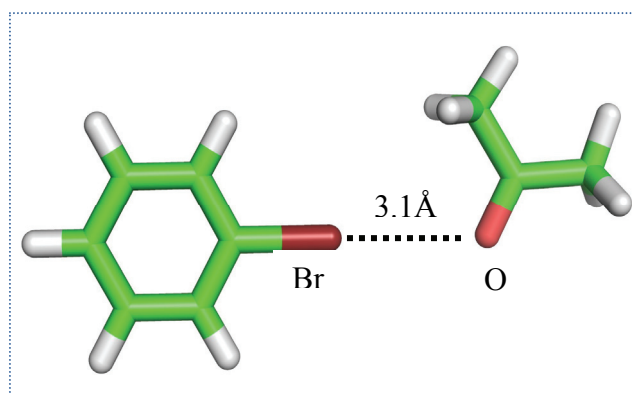


Fig. 3.1: Halogen bond of bromobenzene...acetone complex

The σ -hole is an area of positive electrostatic potential (ESP) that originates in an unequal occupation of valence orbitals on the top of the electron acceptor (*i.e.* Lewis acid). Such a σ -hole thus facilitates the electrostatic interaction with the negative sites. The typical σ -hole is depicted in Figure 3.2. The σ -hole is characterized by its magnitude and size.¹⁵¹ The magnitude of the σ -hole is defined as the value of the most positive or the least negative ESP localized at the halogen boundary defined as a surface of 0.001 e/bohr³ electron density¹⁵² and its size is the spatial extent of the positive region.¹⁵³ It seems to be a key concept for σ -hole bonding, although it concerns only of the two interacting partners. The evidence of $R-X\cdots Y$ halogen bond occurrence, where $R-X$ is a halogen bond donor, X is a halogen atom with electro-poor σ -hole area, R is a group covalently bound to halogen and Y is halogen-bond acceptor, *i.e.* electron donor (O, N, S, P, etc.) can be determined by experiments or/and theoretical studies. The characterization of an interaction as a halogen bond can be done by the satisfaction of its typical features.¹⁵⁴ The interatomic distance tends to be less than the sum of the van der Waals radii and the angle $R-X\cdots Y$ tends to be close to 180°, *i.e.* the halogen bond is strictly directional. Apart from a typical halogen bond where Cl, Br or I covalently bound to an electronegative atom or carbon is in a contact with an electron donor, the case of halogen atom in contact with another halogen exists and is referred to a dihalogen bond. It should be mentioned here that fluorine does not create halogen bonds unless it is bound to a very electronegative entities, such as cyano group or another fluorine atom.

The halogen bond strength decreases with an increased electronegativity of a halogen atom and it can be also tuned by adding more electron-withdrawing substituents in R position.^{153,155} The existence of positive σ -hole elegantly explains the stabilization of a halogen bonding which can reach several kcal/mol. High level QM calculations at the CCSD(T)/CBS level have recently revealed comparable stabilization energies like for a strong $R-H\cdots Y$ hydrogen bond, particularly 5.8 kcal/mol in a iodobenzene...trimethylamine complex from the X40 dataset.¹⁵⁶ However, there are evidences of much stronger halogen bonding stabilization energies, e.g. 17.1 kcal/mol for FI...NH₃¹⁵⁷ or up to 15 kcal/mol for diiodine...1,4-diazabicyclo[2.2.2]octane (DABCO) complex.¹⁵⁸ The question is where these stabilization energies come from.

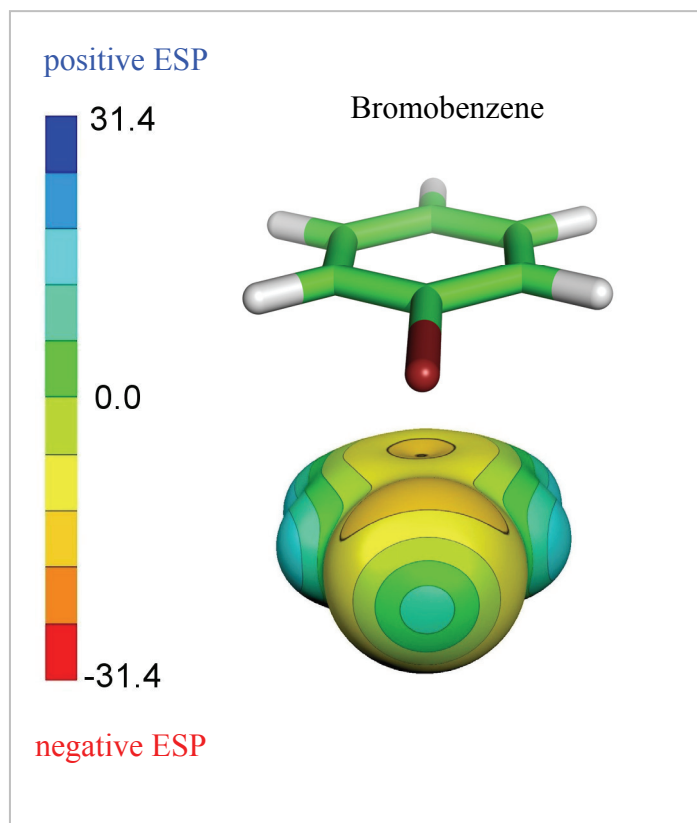


Fig. 3.2: The typical σ -hole, the area of positive ESP created on the top of bromine atom in bromobenzene molecule.

According to IUPAC the definition of halogen bond is following: “the forces involved in the formation of the halogen bond are primarily electrostatic, but polarization, charge transfer, and dispersion contributions all play an important role.”¹⁵⁴ Nevertheless a recent paper of K. E. Riley *et al.*¹⁵⁹ has shown that the clear definition of halogen bonding is maybe still unclear. Authors investigated 10 different halogen bonded complexes by a DFT-SAPT decomposition and it was shown that the electrostatic term slightly dominated in two cases only, whereas in eight cases the dispersion term was dominant. This is in contradiction with the IUPAC definition saying that “the forces involved in the formation of the halogen bond are primarily electrostatic.”¹⁵⁴. To shed light on the nature of halogen bonding more comprehensively, we have extended the dataset of halogen bound complexes significantly and the DFT-SAPT decomposition of the stabilization energies was performed consistently.

Our dataset consisted of 128 halogen-bonded or dihalogen complexes of different size and origin, thus we aimed to cover representatives from weak and moderate complexes

formed by standard electron donors (*e.g.* water, ammonia, formaldehyde, dimethyl ether or trimethylammonia), through standard halogen donors (*e.g.* halobenzenes or substituted halobenzenes), up to strong halogen-bonded complexes with a significant charge transfer.

We made up the dataset from different sources. For the first part of our dataset, the complex geometries were determined at DFT-D3 level and the benchmark CCSD(T)/CBS stabilization energies were known. It particularly consisted of 18 complexes from the X40 dataset (complexes of CH₃Cl, CH₃Br, CH₃I, CF₃Cl, CF₃Br and CF₃I with OCH₂, complexes of chlorobenzene, bromobenzene and iodobenzene with OC₃H₆, NC₃H₉ and SHCH₃, complexes of H₃CBr, H₃Cl, F₃CBr and F₃Cl with benzene)¹⁵⁶, 46 complexes from the XB51 dataset (HCN, NH₃ and HCP in complexes with ICF₃, BrF, ClF, INC₄H₂O₂, BrO₂C₄H₂N, BrC₆H₅, IC₆H₅ and Br₂, FI and H₃Cl in complexes with FC₂H, FCH₃, NCH, NH₃, OCH₂, OPH₃, PCH, NC₅H₅ and H₃Cl...LiH)¹⁵⁷, 11 complexes from the papers previously published by Hobza *et al* (benzene with F₂, Cl₂, and Br₂, I₂...I₂, Br₂...Br₂, Cl₂...Cl₂, F₂...F₂, Br₂ with trimethylbenzene and hexamethylbenzene and twice CH₂BrOH...CH₂BrOH complex with Br-O halogen bonding and with Br-Br dihalogen bond)¹⁶⁰⁻¹⁶² and 13 complexes (C₂H₃Cl, C₂HCl, C₂H₃Br, C₂HBr, C₂H₃I and C₂HI in complexes with H₂CO, H₂O and NH₃) from Ref.163. For the second part of halogen-bonded complexes, the stabilization energies and complex geometries were calculated at MP2 and DFT-D3 level. It consisted of 8 complexes of crystal motifs which were taken from the Cambridge Structure Database (CH₃CN and CO in complex with BrF, ClF, BrF₃ and ClF₃)¹⁶⁴, 15 complexes (ICN, IBr, ICl and I₂ in complexes with NC₅H₅, complexes of I₂, C₄F₉I and C₆F₅I with OSC₂H₆, NC₆H₁₅ and NC₇H₁₃ and complexes of I₂ and C₄F₉I with OPC₃H₉) from Ref. ¹⁶⁵ and finally 17 structures of organic crystals (I₂...1,4-diazabicyclo[2.2.2]octane, I₂...1,3-dithiole-2-thione-4-carboxylic acid, C₆Cl₆...C₆Cl₆, C₆Br₆...C₆Br₆, C₄N₃H₄Br...C₇F₄O₂HBr, C₆F₄I₂...I₂F₄C₆, C₇F₄O₂HBr...NBrC₄N₃H₂, three different orientations of 1,2-TFIB...TMO, 2-mercapto-1-methylimidazole...1,2-TFIB, 4,4'-bipyridine ...1,2-TFIB, (3,4,5-trichlorophenol)₂ and four different orientations of 1,2-TFIB...1,2-TFIB)) taken from Refs. 166-171. Structures of all investigated complexes were taken from the original references without any additional optimization. They are shown in Figure S1-S7 of Appendix A.

We aimed to combine approaches to monomers and complexes in order to provide novel insight into halogen bonding. In the first step we paid an attention to the characterization of isolated halogen donors. The σ -holes of all halogenated subsystems

were described in terms of size and magnitude. The energy minimization prior ESP and ESP calculations were done at the PBE0/aug-cc-pVDZ level with the pseudopotentials on bromine and iodine atoms. In the second step, halogen-bonded complexes were studied by DFT-SAPT decomposition of their total stabilization energy. We used pseudopotentials for bromine and iodine atoms to correctly describe relativistic effects of inner-core electrons. A gradient-controlled shift procedure was carried out by using PBE0/aug-cc-pVDZ and PBE0/aug-cc-pVTZ calculations. The DFT part was treated using the localized and asymptotically corrected LPBE0AC exchange-correlation functional with the density fitting and the aug-cc-pVDZ. It is known that this combination of the functional and the basis set provides reasonably good results for all SAPT energy terms, except of dispersion which is underestimated with a smaller basis set. We thus estimated DFT-SAPT/CBS for 18 complexes from the X40 dataset¹⁵⁶ using two-point extrapolation methods with aug-cc-pVDZ and aug-cc-pVTZ basis sets. The obtained scaling factor for aug-cc-pVDZ dispersion energy was then used for halogen-bonded complexes in which higher level of theory is too demanding. For even more extended complexes we used a hybrid DFT-SAPT approach using Hesselmann empirical dispersion which was scaled in the same manner.¹⁴² For most of the complexes we also calculated BSSE corrected interaction energies at the DFT-D3 (B97-D3/def2-QZVP) level.

It was shown that the all studied subsystems (F₂, Cl₂, ClF, ClF₃, F₃CCl, C₂H₃Cl, C₂HCl, chlorobenzene, C₆Cl₆, C₆H₂OHCl₃, Br₂, BrF, BrF₃, H₃CBr, F₃CBr, C₂H₃Br, C₂HBr, bromobenzene, C₆Br₆, BrC₄H₂NO₂, CH₂BrOH, C₇F₄O₂HBr, I₂, IF, ICl, IBr, ICN, H₃Cl, F₃Cl, C₂H₃I, C₂HI, iodobenzene, C₆F₅I, C₄F₉I, INC₄H₂O₂, HO₂C₇F₄I, TFIB) possess a positive σ -hole (with the exception of slightly negative σ -hole of H₃CCl). It was also proved that the magnitude and size of the σ -hole correlate well (R=0.86) and they increase with the atomic number of the halogen atom and with the presence of electron-withdrawing fluorine atoms. While the magnitude anticorrelates with the LUMO energy, *i.e.* strong electron acceptors have more positive σ -holes. In the case of the dihalogen bonding, the magnitude increases with the decreasing atomic number of the second halogen. All of these trends agree with previously presented dependence. We also tried to relate the properties of monomers with properties of complexes. Figure 3.3 shows that the stabilization energy surprisingly correlates with the magnitude of the σ -holes only weakly (with correlation coefficient R being 0.52). However when we selected the most stable complex of particular halogenated monomers, the correlation increased to R=0.77.

The magnitude of the σ -hole therefore informs about the ability of a monomer to create the halogen bond rather than about the strength of the halogen-bonded complex.

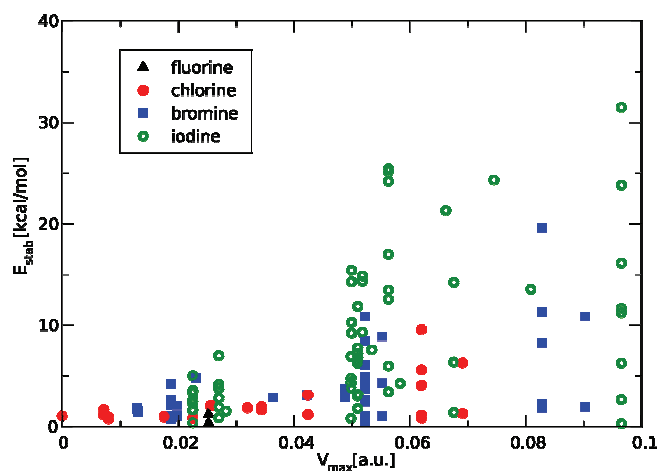


Fig. 3.3: The dependence of the stabilization energy E_{stab} on the magnitude of the σ -hole V_{max}

The results of DFT-SAPT decomposition showed that all studied halogen-bonded complexes can be split into two different classes according to the strength of their total stabilization energy. The first class comprising 38 complexes was characterized by the strong stabilization energies (larger than 7 kcal/mol), relatively small distances between the halogen and the electron donor (even below 2.4 Å) and significant difference between this distance and the sum of the respective vdW radii (up to 1.2 Å). This contraction is connected with the important induction energy, which was here in 21 cases more important than the dispersion energy. The large induction cannot originate in the classical permanent dipole – induced dipole induction energy but rather reflects the importance of charge-transfer contribution that is confirmed by the negative values of the LUMO of these electron acceptors. The polarization (electrostatic) energy was almost systematically dominant for all complexes (with only one exception where the dispersion energy term was larger). The second class of 90 standard halogen-bonded complexes had weaker stabilization energy between 0.3 and 7 kcal/mol. Their contraction of the vdW distances were much smaller (mostly less than 0.5 Å) except for dihalogen-bonded cases. In 48 complexes, the dispersion energy was mostly dominant, followed by the polarization and the induction energies. For the rest of 42 complexes the polarization energy was dominant, followed by the dispersion.

While the electrostatic term is in halogen and hydrogen bonds more or less comparable, the contribution of the dispersion energy to the stability differs a lot. In the halogen bonds dispersion contribution is much larger because there are two heavy atoms in contact, in

contrast of the case of the hydrogen bond where the light hydrogen and electron donors are in contact. To demonstrate the importance of this contact atom pairs we estimated for 14 complexes of the X40 dataset and 8 extended organic crystal complexes the contribution to the total dispersion energy coming from this pair by means of empirical dispersion term.¹⁴² The dominant role of this pair was shown, because the dispersion energy of the contact atom pair equals 40% on average of the total dispersion energy.

To summarize, we have shown that within the whole set of 128 halogen-bonded complexes is the most dominant contribution to the total stabilization energy the polarization (electrostatic) energy in 62% of complexes, whereas in remaining 38% of cases the dominant term is the dispersion energy. Both contributions are thus with the same importance responsible for a characterization of the halogen bonding, where the electrostatic interaction is responsible for stabilization and directionality of the bond and dispersion energy is responsible for its high stabilization.

It was already mentioned that the existence of the σ -hole is not restricted only to halogen atoms, but also for atoms of Group IV-VI and related non-covalent interactions are known as pnicogen and chalcogen bonds.¹⁷² The typical chalcogen bond is formed between a chalcogen atom (S, Se, Te) and particular negative site.¹⁷³⁻¹⁷⁶ Chalcogens are because of their high electronegativity negatively charged in organic structures. The ESP around the chalcogen atom is however in the same manner as in halogens strongly anisotropic and the areas of positive σ -holes are formed. The size and magnitude of the σ -hole can be tuned by adding electron-withdrawing substituents, as it is shown for the case of thioformaldehyde (CH_2S) and thiocarbonyl flourid (CF_2S) in the Figure 3.4. While the σ -hole localized at the top of the divalent sulphur atom in CH_2S is just less negative than the surrounding ESP, in the case of CF_2S it is already positive. In contrast, there have been recently synthesized and crystallized thiaborane structures,^{177,178} where the sulphur atom is bound to five boron atoms. The ESP in the Figure 3.4 shows that a sulphur atom is in this case positively charged with the less positive area at the top of the atom and five highly positive σ -holes on its sides.

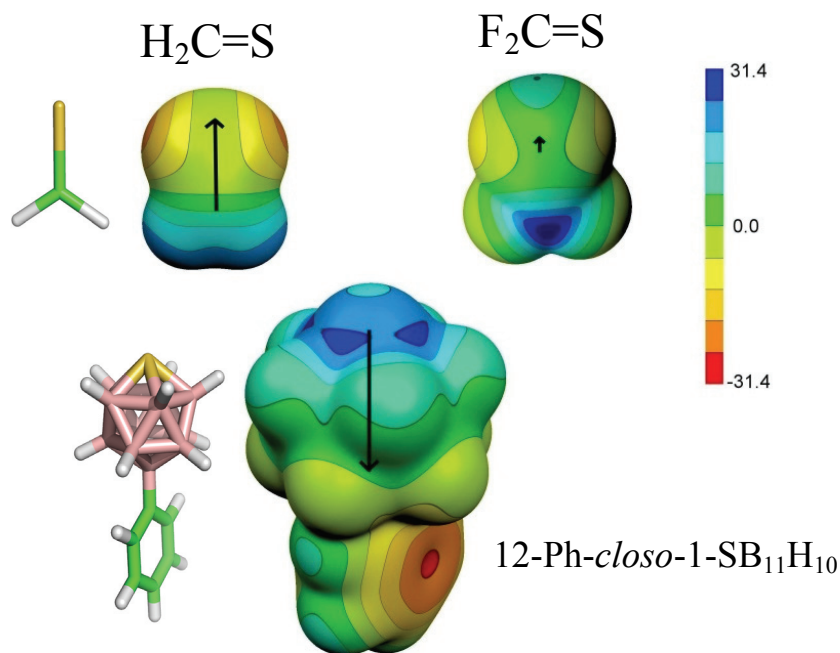


Fig. 3.4: Comparison of calculated ESPs and dipole moments (arrows) for a sulphur-bound 2D organic structures and 3D inorganic boron hydrid.

To reveal an ability of these structures to form chalcogen bonds and its characteristics such as energetic quantification and the directionality we performed quantum mechanical analysis of all binding motifs in inorganic crystals of thiaboranes with an *exo*-substituted chlorine atom (12-Cl-*closo*-1-SB₁₁H₁₀), iodine atom (12-I-*closo*-1-SB₁₁H₁₀) and phenyl ring (12-Ph-*closo*-1-SB₁₁H₁₀).

The crystal model was obtained by a cutoff within 5 Å around of the central molecules and all hydrogens were optimized at DFT-D/BLYP/SVP level. Interaction energies were obtained for all binding motifs by DFT-D3/TPSS/TZVPP method with the pseudopotential for iodine. The total interaction energies were then decomposed by a hybrid DFT-SAPT approach using the empirical dispersion¹⁴² at LPBE0AC/aug-cc-pVDZ level with a gradient-controlled shift procedure. As a benchmark data we calculated BSSE corrected interaction energies at the CCSD(T)/CBS level, using extrapolation from aug-cc-pVDZ and aug-cc-pVTZ basis sets.⁹⁹ The CCSD(T) correction term was obtained by modified 6-31G* basis set with changed exponents of polarization functions.¹⁷⁹

The ESP on 0.001 a.u. and dipole moments computed at HF/cc-pVDZ level revealed that the magnitude of σ -holes is higher than in majority of halogen-bonded systems (for example $V_{s,max}=26.7$ kcal/mol for 12-Ph-*closo*-1-SB₁₁H₁₀ is comparable to multisubstituted bromobenzene).¹⁵⁵ When going to 12-Cl-*closo*-1-SB₁₁H₁₀, the magnitude of σ -holes is

even bigger. It shows that σ -holes of the 3D aromatic cages can be also tuned by adding electron-withdrawing groups.

From the results of interaction energies of all binding motifs (Figure 3.5) it was found that the B-S... π chalcogen bond is the strongest binding motif. It should be stressed here that the B-S... π angle is about 155° and so the chalcogen bond is not linear for thiaboranes as it was predicted.

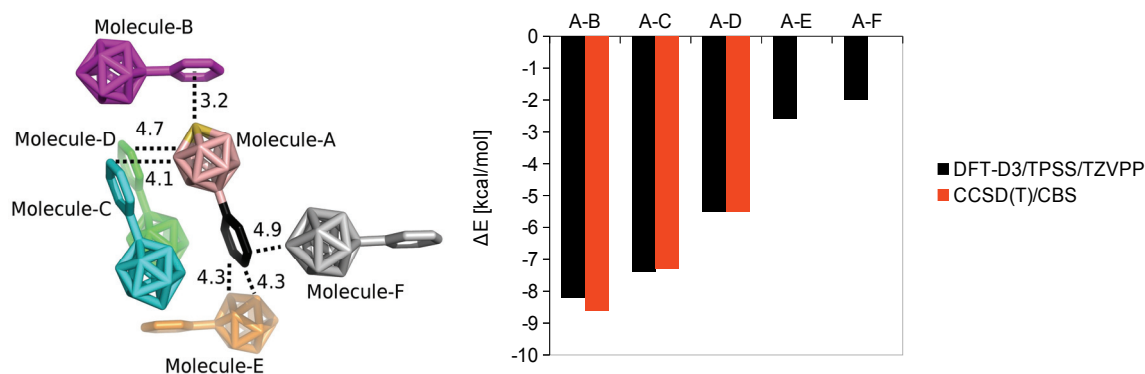


Fig. 3.5: All binding motifs of 12-Ph-*closo*-1-SB₁₁H₁₀ crystal model, where A...B motif is characterized as the B-S... π chalcogen bond (on the left) with the estimated interaction energies of all these motifs at DFT-D3 and CCSD(T)/CBS level (on the right).

The interaction energies of A-C and A-D stacking motifs of 12-Ph-*closo*-1-SB₁₁H₁₀ are weaker than the chalcogen bonding by about 0.8 and 2.7 kcal/mol, respectively. The bonds between molecule-A and molecule-E and F are much distant and thus the head-to-tail bonds are significantly less stable. Passing to the chlorine and iodine *exo*-substitutions it was also shown that the B-S... π chalcogen bond is much stronger than the B-S...X one.

The nature of stabilization of all motifs was elucidated by the DFT-SAPT decomposition (Figure 3.6). It was shown that the dominant contribution to the total stabilization energy for B-S... π chalcogen bond is the dispersion energy, followed by the electrostatic term. Charge transfer causes that the induction energy is systematically larger for chalcogen-bonded motifs. It should be mentioned that the stabilization energy of the B-S... π chalcogen bond exceeding 8 kcal/mol is considerably stronger than those in their organic counterparts and in the known halogen bond.¹⁵⁶

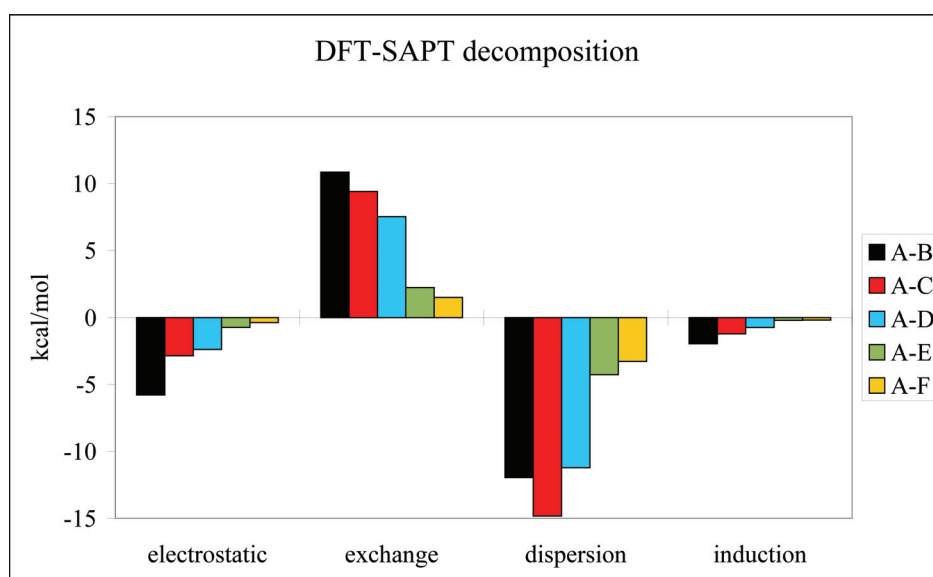


Fig. 3.6: The decomposition of the stabilization energy of all binding motifs of 12-Ph-*closo*-1-SB₁₁H₁₀ estimated by the DFT-SAPT method.

The CCSD(T) benchmark data agreed well with the DFT-D3 calculations, however the total DFT-SAPT interaction energies were slightly more negative. Moreover it was shown that the motif A-C became more stable than chalcogen bond in this case. Based on the comparison with benchmark calculations it indicates an artefact of a hybrid DFT-SAPT method. Therefore to elucidate the complete picture of the σ -hole bonding of heteroboranes, a systematic computational study on bigger dataset is needed. We thus applied high-level QM methods on to the extended dataset made of experimentally known neutral icosahedral and square-antiprismatic *closo*-heteroboranes in which carbon, chalcogen and also pnictogen atoms are incorporated in the 3D cages, whereas halogens are considered as *exo*-substituents of dicarbaboranes.

Our dataset included 12 heteroborane molecules (*closo*-1-SB₁₁H₁₁, 12-F-*closo*-1-SB₁₁H₁₀, 12-Cl-*closo*-1-SB₁₁H₁₀, 12-Br-*closo*-1-SB₁₁H₁₀, *closo*-1-SeB₁₁H₁₁, *closo*-1-SB₉H₉, *closo*-1,2-P₂B₁₀H₁₀, *closo*-1,2-As₂B₁₀H₁₀, *closo*-2,1-PCB₈H₉, *closo*-6,1-PCB₈H₉, 12-Br-*closo*-1,2-C₂B₁₀H₁₁ and 1-Br-*closo*-1,2-C₂B₁₀H₁₁). We studied non-covalent complexes of selected boron clusters with five σ -hole acceptors (benzene, trimethylamine, dimethyl ether, acetone and formamide). The minimum of the complexes was found by geometry optimization with various fixed angles of σ -hole bonds with a step of 5° at DFT-D3:TPSS/TZVPP level together with the confirmation via estimated vibrational frequencies. The structures of studied systems are depicted in Figure 2 of Appendix C.

As previously, the BSSE corrected CCSD(T)/CBS interaction energies were calculated as the benchmark data and the decomposition was performed by the DFT-SAPT method at LPBE0AC/aug-cc-pVDZ with density fitting. To scale the underestimated dispersion energy at aug-cc-pVDZ level¹³⁸ we used the scaling factor estimated on a model system (*closo*-1-SB₉H₉...formamide complex) by DFT-SAPT two-point extrapolation to CBS.

The magnitude of σ -holes was estimated by means of ESP calculation on isolated molecules at HF/cc-pVDZ level. The results showed that the σ -holes of chalcogens and pnicogens are both more positive in heteroboranes than in organic molecules. They are highly positive areas formed on already positively charged chalcogen and pnicogen atoms. In the case of two pnicogens incorporated into the borane cage, the area of the most positive ESP is placed between these atoms as shown in Figure 3.7.

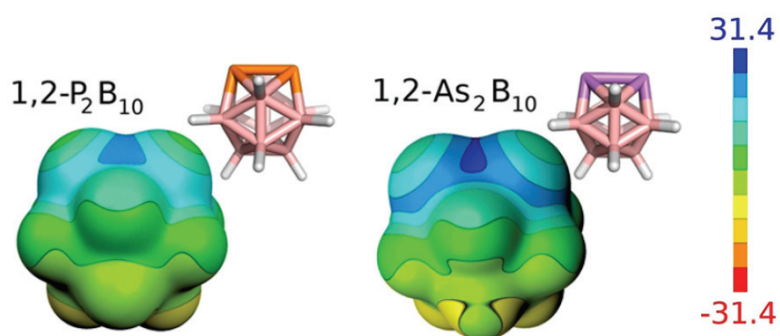


Fig. 3.7: ESP calculated on *closo*-1,2-P₂B₁₀H₁₀ and *closo*-1,2-As₂B₁₀H₁₀ shows the σ -holes position in the valley between pnicogen atoms.

It is also evident that the magnitude of the σ -holes can be tuned by *i*) changing the atomic number, *e.g.* going from S to Se and from P to As the magnitude increased by about 1.4 and 4.3 kcal/mol respectively, *ii*) changing the skeleton of the borane cage, *e.g.* the magnitude of sulphur σ -holes increased by about 5.9 kcal/mol by going from 10-vertex to 12-vertex cage and *iii*) changing the chemical environment, *e.g.* in the case of *closo*-2,1-PCB₈H₉ is the magnitude of the σ -hole higher about 5 kcal/mol than in *closo*-6,1-PCB₈H₉ and also by the *exo*-substitutions of halogen atoms instead of hydrogens in *para* position to the heteroatom increase the magnitude by about 1.8 kcal/mol on average. In the case of halogenated dicarbaboranes the magnitude of σ -hole depends on the position of halogen-bound vertex, *e.g.* when the halogen is bound to a carbon atom of the cage, the σ -hole is highly positive. We also aimed to deduce the characteristics of the chalcogen, pnicogen

and halogen bonding in heteroboranes from the analysis of their complexes (illustrative interactions of these complexes are shown in Figure 3 of Appendix C).

The chalcogen bonding was represented by 1-SB₁₁H₁₁ cage interacting with benzene, trimethylamine, dimethyl ether, acetone and formamide partners. The complex 1-SB₁₁H₁₁...benzene represented the simplified model system of previously studied chalcogen bond in 12-Ph-*closo*-1-SB₁₁H₁₀ in the crystal. The model was approved by the finding of the minima of the angle between both monomers that agreed with experimental value and also by the similarly strong stabilization energy and its DFT-SAPT decomposition. The interaction energies varied among the σ -hole acceptors and DFT-D3 results agreed well with CCSD(T)/CBS reference data. The benzene and trimethylamine molecules were the best σ -hole acceptors, while dimethyl ether was the weakest one. Data showed that the optimal angle of chalcogen bonds of 1-SB₁₁H₁₁ ranges between 130° and 165° which agree with the positions of σ -holes. In all 1-SB₁₁H₁₁ complexes the dispersion energy played the major role in stabilization, followed by polarization term which was dominant only in the case of formamide. Small induction energy contributions show that charge transfer does not contribute here. To model the modulation of chalcogen bonding we selected three other cages, *i.e.* *closo*-1-SeB₁₁H₁₁, 12-Cl-*closo*-1-SB₁₁H₁₀ and *closo*-1-SB₉H₉, and analyzed their interactions with all partners. Results revealed that the S to Se substitution has the biggest impact on the modulation with ΔE being about 1.3 kcal/mol more negative on average and simultaneously all contributions to the total interaction energy become more negative. The changing the skeleton neither brings any significant change of total interaction energies (SB₁₁H₁₁ cage complexes lower down the ΔE by about 0.2 kcal/mol in contrast to SB₉H₉ cage) nor the *exo*-substitution of chlorine atom (12-Cl-*closo*-1-SB₁₁H₁₀ complexes have ΔE comparable to 1-SB₁₁H₁₁).

Pnicogen bonding was represented by 1,2-P₂B₁₀H₁₀ and *closo*-1,2-As₂B₁₀H₁₀ complexes with bond acceptors. The interaction energies of the 1,2-P₂B₁₀H₁₀ are of the similar strength as in the 1-SB₁₁H₁₁ (pnictogen bonds were about 0.5 kcal/mol less stable). The difference can be seen for the interactions with benzene, where the SB₁₁H₁₁ complex was more stable by about 1.4 kcal/mol. This can be caused by the worse accessibility of the σ -hole located in the valley between two pnicogens. It was shown that the P-to-As substitution has smaller impact on modulation than in the case of chalcogen bonding, because the interaction energy lowered down by 0.7 kcal/mol on average. DFT-SAPT decomposition showed that the dispersion energy plays an important role in the pnicogen

bonding, followed by polarization with not negligible induction term. In some cases polarization is even comparable to dispersion.

As it was mentioned before, the halogen bonding is observable only in the case when the halogen is *exo*-substituted to the carbon atom of the heteroborane cage. Contrarily, when it is bound to the boron atom, its σ -hole is just less negative than the negatively charged surrounding. Which means that 12-Br-*closo*-1,2-C₂B₁₀H₁₁ does not form halogen bonding, with only one exception (in the complex with trimethylamine is the halogen bond shorter than the sum of van der Waals radii). DFT-SAPT showed that this weak interaction is enabled mainly by dispersion energy. In the case of 1-Br-*closo*-1,2-C₂B₁₀H₁₁ the strong halogen bonds are formed. The best halogen bond acceptor was trimethylamine, where the stabilization of the complex came from very large polarization term followed by induction energy.

We have demonstrated that the pnictogen and chalcogen atoms incorporated in heteroborane cages are positively charged entities carrying even more positive σ -holes. The same is true for halogen atoms bonded to the carbon atom of dicarbaboranes. These molecules can thus form very strong halogen, pnictogen and chalcogen bonds that are stronger in heteroboranes than in other neutral σ -hole bonded organic complexes.

At the end of this section I would like to stress some conclusions that have arisen from our studies. We have shown that the only way how to elucidate the complete picture of σ -hole bonding is to relate the properties of monomers, *i.e.* σ -holes, with the properties of complexes. It was demonstrated that the fact that strength of σ -hole bonding in isolated complexes is proportional to its magnitude of the σ -hole on the atom may not be so straightforward and many other effects can come to play. Therefore only the high level quantum mechanical methods can answer the question of the nature of such a bonding.

The analysis of contributions to the total stabilization energies of the extended dataset of halogenated complexes calculated by DFT-SAPT method has revealed the concert action of polarization and dispersion energies to the stabilization of halogen bonding. Positive σ -hole and the negative electron donor interact by the electrostatic energy, which is responsible not only to the stability but also for the high directionality of the bond while dispersion energy is responsible for its high stability. The question of the adequacy of the recent IUPAC definition¹⁵⁴ of halogen bonding has thus arisen.

Halogen, chalcogen and pnictogen bonds are found in organic compounds standardly, but have so far never been observed in inorganic boron hydrides. We have thus tackled

experimentally-known inherently electron-deficient heteroboranes in order to examine their ability to form σ -hole bonding. Firstly we studied the inorganic crystal of thiaborane in which 2D and 3D aromatics are connected. We have shown the existence of five highly positive σ -holes on the positively charged pentacoordinated sulphur atom and consequently the ability of this structure to form B-S... π chalcogen bonds which are considerably stronger than these in their organic counterparts and in known halogen bonds. In order to gain a deeper insight into the nature of these nonclassical σ -hole-based non-covalent interactions, we have applied a detailed QM study to the majority of experimentally known *closo*-heteroboranes, where chalcogens and pnicogens are incorporated in the borane cage, together with *exo*-substituted halogens. As opposed to the classical electronegativity concept, we have shown that all these heteroatoms are centers of positive charges and so form very strong σ -holes bonds. DFT-SAPT decompositions of their total stabilization energies have revealed that chalcogen and pnicogen bonds come from dominating dispersion and electrostatic energy, followed by induction showing the not negligible role of charge transfer. Moreover we have also shown and quantified several ways of modulation of σ -hole bonding which can be utilized in its applying in crystal engineering and drug design.

The importance of halogen bonds for rational molecular design is well known and halogen substitutions present a promising way for the improvement of the activity of drugs. The role of halogen bonding in molecular recognition, crystal engineering^{180,181} and drug-target interactions, is now being extensively investigated.^{79,182-186} Although the chalcogen bond is not so well researched compared to halogen bonds, they play an important role in crystal engineering^{174,187} as well as in drug design.^{188,189} It was also shown by an analysis of Protein Data Bank that they also influence protein structures.^{175,190} Recently, pnicogen bonds have been used as new supramolecular linkers.^{191,192}

Heteroboranes have already been used in nanotechnology and medicinal chemistry, mainly because of their specific properties like their hydrofobicity, 3D shape, aromaticity, stability and ability to form dihydrogen bonds.¹⁹³⁻¹⁹⁶ We have just shown the ability of heteroboranes to form all types of σ -hole bonding and it can thus be utilized in the design of heteroborane-based protein ligands, such as enzyme inhibitors or receptor antagonists/agonists. One of the many focuses can aim to the B-S... π chalcogen bonding between heteroborane-based ligand and phenylalanine aminoacid of the active site of the protein target.

3.2 Protein-Ligand Binding

This section of the Chapter 3 is devoted to the second topic of this thesis that has been covered by four original publications attached in Appendices D, E, F and G. It focuses on a detailed quantum mechanical analysis of protein-ligand interactions of three medicinally important targets.

The known atomistic structure of the target molecule, mostly the protein structure, is the main prerequisite of the structure-based drug design. Protein structure is the 3D arrangements of atoms in protein molecule and it can be determined by several techniques, *i.e.* X-ray crystallography, NMR spectroscopy or cryo-electron microscopy. The data of the structures are freely accessible via Protein Data Bank (PDB, <http://www.rcsb.org>)¹⁹⁷ Up to now (January 2016) more than 106.500 protein structures have been deposited and its number is still growing. It should be mentioned that the resulting X-ray structure comes from an average electron density of all molecules within the crystal and not every atom is possible to observe, such as weakly scattering hydrogen atoms. Some groups of atoms can be indistinguishable from each other such as an amide C(=O)-NH₂ or imidazol group. It is also possible that some atoms appear in the X-ray structures multiple times, *e.g.* multiple conformations of the protein sidechains. The dynamical features of atoms are presented in atomic displacement parameters, often so called temperature B-factors. The model can be several times refined until the correlation between the diffraction data and model is maximized and this agreement is measured by R factor, *i.e.* the resolution of the crystal. Despite mentioned approximations, X-ray crystallography is now used routinely to determine the interaction of drug molecule and its protein target.¹⁹⁸

In the protein-ligand complex, multiple non-covalent interactions of different kind play the role in molecular recognition. The complex represents the balance between attractive and repulsive interactions and the role of the structure-based drug design is to identify and optimize these interactions between the ligand and the protein. Experimentally determined binding affinity gives very little insight into the relationship of the geometrical features of the ligand and its interaction with the protein. Whereas computational methods provide access to the detailed decomposition of the interaction between the ligand and the receptor. Theoretically achieved insights can often be validated by experiments and vice versa. QM based methods are in general able to identify all kinds of non-covalent interactions and help to understand the energetic contribution of the particular interaction to the binding

free energy. Understanding the nature of stabilization of protein-ligand complexes helps in the design of more active ligands.

The treatment of protein-ligand complexes with thousands of atoms requires the use of methods which are fast and still accurate. The corrected DFT with density fitting and SQM methods with linear scaling algorithm perfectly match these requirements. The calculations can be speeded up even more using a QM/MM approach. We have used our in house ONIOM-like¹⁴⁴ subtractive QM/MM approach with link atoms and mechanical embedding combined with GB⁷¹ or COSMO⁷² implicit solvent models. The QM part comprising up to 500 atoms is usually solved by corrected DFT methods^{111,113} accelerated by density fitting¹⁹⁹ and the corrected semiempirical PM6 method⁶⁷ with linear scaling MOZYME algorithm¹²⁴ is used for even bigger QM parts covering thousands of atoms standardly. The MM part is usually treated using AMBER ff03 forcefield²⁰⁰ for the protein and GAFF parameters²⁰¹ and RESP charges at HF/6-31G* level²⁰² for the ligand. In some cases the convergence of optimizations can be speeded up by the keeping the outer part of the protein frozen. We have applied this calculation setup at different levels of theory to all studied protein-ligand complexes, however we tackled different tasks within each study.

To provide the meaningful energies, calculations must be performed on a reasonable 3D structure. When the starting point is an experimentally determined X-ray structure of the protein-ligand complex, needs of single conformation of each residue, addition of missing residues and hydrogens, correct protonation variants of all residues with respect to the surroundings, pK_a and optimal pH of the particular target, should be fulfilled. We use for hydrogen addition for ligands USCF Chimera program²⁰³ and for proteins the Reduce and LEaP modules of the AMBER simulation package.²⁰⁴ The positions of added hydrogens are every time relaxed by optimization followed by molecular dynamics-based simulated annealing using the Berendsen thermostat²⁰⁵ in the SANDER module of the AMBER package.²⁰⁴ All these steps of the preparation process should be carried as carefully as possible, because a correct/incorrect computational model influences the results of interaction energy calculations. Prior any energy calculations all complexes should be fully optimized using QM/MM setup.

We have applied reliable and accurate QM methods to unveil the features of the structure which are not accessible to the crystallographic experiment, even at high quality. Most notably, these are *i*) the determination of the protonation states and the

identification of the most stable conformers/tautomers *ii*) the dissection of the energy contributions of the individual amino acids toward the total interaction and *iii*) the characterization of the nature of different binding of similar/different ligands. The prior knowledge is important for building the reliable computational model for interaction energy calculations and the latter two findings are useful for an understanding and selectivity of the ligand binding to the particular protein target as well as for a further rational design of more potent/selective inhibitors.

3.2.1 Protonation of HIV-1 Protease/Inhibitor complex

HIV-1 protease (PR) is a retroviral aspartyl protease and is one of the enzymes of HIV retrovirus that causes AIDS. It has an essential role in the maturation process of infectious virion and thus it is one of the most studied pharmaceutical targets. It has two catalytic aspartates (Asp25/Asp25') in the active site of its C₂-symmetrical dimeric structure. It is known that these coplanar aspartates are close to each other and connected by one proton via double-well low-barrier hydrogen bond.²⁰⁶ However, when the complex is formed with inhibitor featuring hydroxyl isostere this aspartic dyad becomes monoprotonated²⁰⁷ or less frequently diprotonated in the case of statine-based inhibitors.^{208,209}

We studied two model systems of HIV-1 PR/inhibitor complex, particularly with nonpeptidic inhibitor darunavir (DRV, PDB code 3QOZ)²¹⁰ and phenylnorstatine-based peptidomimetic KI2 inhibitor (PDB code 1NH0)²⁰⁹. In the complex of HIV-1 PR with DRV, there were two orientations of the inhibitor and four possible variants of the monoprotonated catalytic dyad (Figure 3.8). The atomic resolution of the HIV-1 PR/KI2 crystal (R=1.03Å) revealed two conformations of P2 benzyloxycarbonyl group of the ligand, where the B conformation offers alternative possibility of hydrogen bonding of the KI2 hydroxyl group – intermolecularly to Asp25' or intramolecularly to O01 of KI2 (for details see Figure 3.8). However the protonation of the active site is defined here, another carboxyl interaction takes place between the KI2 inhibitor and Asp30, where both of the partners can be protonated or just connected by one hydrogen bond which can be localized on one or other partner or can be mobile in the sense of low-barrier hydrogen bond.

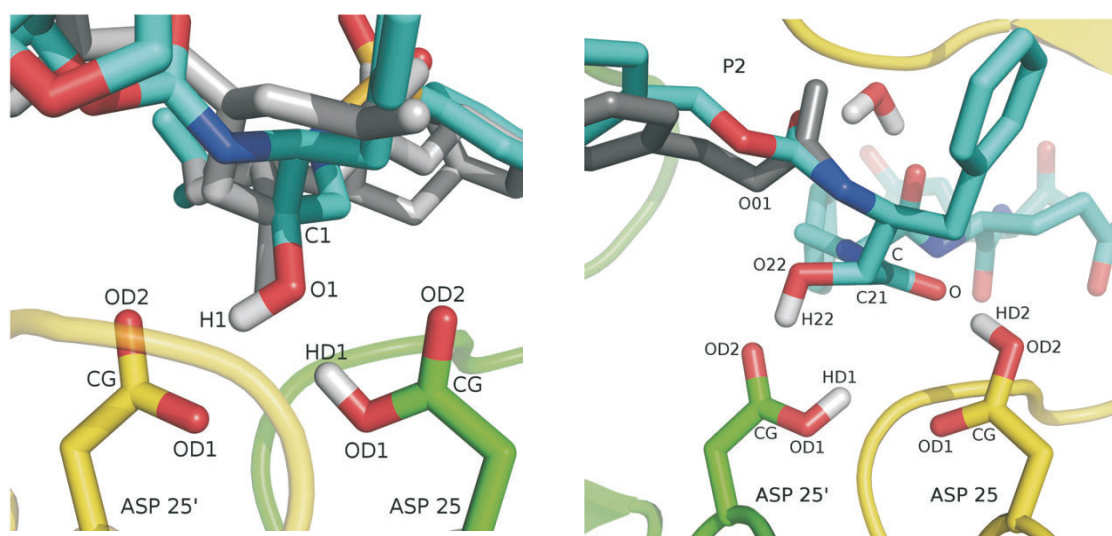


Fig. 3.8: On the left: the active site of HIV-1PR/DRV complex with monoprotonated aspartic dyad, where the first of four possible protonation variants is depicted (H1 of the ligand in the orientation A interacts with OD2 of Asp25' whereas OD1 of Asp25 is protonated). On the right: KI2 inhibitor bound to the diprotonated Asp25/25' dyad of HIV-1 PR. Protease monomers in green and yellow, x-ray pose orientations/conformations in blue and grey, respectively.

We aimed to shed light on the probability of all protonation variants of both crystal complexes with regard to all orientations and conformations of the ligands. All 16 protonation, orientation and conformation variants are summarized in Table 1 of Appendix D. We thus made the respective molecular models and explored them by QM-based methods. All models were optimized by the QM/MM approach and sorted by the relative energies of their QM parts. The semiempirical PM6-DH2//PM6-D method¹²¹ allowed us to explore the influence of the extension of QM parts up to 10 Å from the ligand. The results were checked by RI-DFT-D¹¹³ QM/MM calculations at TPSS/TZVP//B-LYP/SVP level on the QM part defined by 3Å surroundings (~450 atoms) of the inhibitor. The important flap water molecule was kept in the active site explicitly, whereas the bulk was modeled by GB and COSMO implicit solvent model, respectively.

The molecular models were prepared using a special protocol developed for the comparison of the stabilities of differently protonated structures. We started with only three complexes with chosen protonation variant, representing unique structures of the DRV in one orientation (because of the dimeric structure of HIV PR the second orientation forms identical protonation variants) and of the KI2 in both conformations. And only after addition and relaxation of hydrogen atoms in these complexes, all other protonation variants were made and added protons were again optimized. Comparable

computational models of all 16 protonation variants having relaxed hydrogen positions but differing only in the studied protonations were thus the starting points for the QM/MM optimizations.

The single-point energies of the optimized 3Å-QM parts of all protonation, orientation and conformation variants were calculated. In DFT-D QM/MM calculations on HIV-1 PR/DRV it was shown that the same symmetry-related protonation variants were proved to be the most stable ones for both orientations of the darunavir. The most stable protonation variant for the A orientation with proton localized on the lower oxygen OD1 of Asp25' and hydroxyl group of the ligand O1-H1 heading to OD2 of Asp25 is symmetry related and so energy-related to the case of B orientation where proton is localized on the lower oxygen OD1 of Asp25 with the O1-H1 hydroxyl group of DRV heading to OD2 upper oxygen of Asp25' (see Figure 3.8). The second most stable protonation variant differing within 3 kcal/mol was the case when both orientations were exchanged contrary the previous case. All other protonation variants were 4.3-13 kcal/mol less stable. In the SQM/MM optimization, half of the structures resulted in a proton transfer in the active site, transforming protonation variants to the most stable one which was the same one as in DFT-D case. Transformed variants became trapped in the local minima, differing from the ideal structure of the most stable variant and so remained 5.4-7.1 kcal/mol less stable. Consequently we have examined if these structures reaches global minima when we define bigger QM part, *i.e.* allowing more relaxation in more distant surroundings of the active site. We thus increased the size of QM part stepwise up to approximately 1700 atoms and optimize it by the corrected PM6 method. Despite quantitative differences and taking proton transfer into account the most stable variants were found consistently in 3, 6, 8 and 10Å regions surrounding the ligand. However the allowing large parts of the protein to move did not solve the problem of trapping the unstable variants in the local minima of higher energy.

In the case of studying monoprotonated variant of Asp30/Glu-P2' pair of HIV-1 PR/KI2 complex, DFT-D QM/MM optimizations resulted in proton transfer towards Asp30 residue, whereas SQM optimizations resulted in the proton localized between two oxygens of these residues with typical O-H distances 1.2 and 1.3 Å. This can be explained by the shape of proton transfer curves in model systems (see further). The diprotonated variant was energetically less stable and moreover had a bigger RMSD of optimized structures with respect to the crystal positions of Asp30 and Glu-P2'. Therefore we concluded that

monoprotonated variant is more probable. We have also studied the relative stabilities of A and B conformations of the ligand in the HIV-1 PR/KI2 complex. Both methods consistently showed the preference of the A conformation that was more stable about 10 kcal/mol in the DFT case and 10-20kcal/mol in the PM6-DH2. The results qualitatively agrees with the higher occupancy of the A conformation observed in the crystal structure.²⁰⁹ QM/MM optimizations also revealed the third structural detail of the hydrogen bond formed by the hydroxyl of the B conformation of the ligand. In the intermolecular case no significant movement was observed for involved atoms, whereas in the intramolecular case O01 oxygen of Glu-P2' moved in the direction of its position in the A conformation and increased the distance to the O2 oxygen (see Figure 3.8). We thus concluded that the intramolecular case would not be stable.

The enlarging of the size of the QM region did not influence the results. The sizes of the 6 and 8 Å surroundings of the ligand were energetically consistent with DFT and PM6-DH2 results on the small region. However in the regions bigger than 10 Å, the unrelated structural changes occurred far from the active site that affected also the relative stabilities. We therefore recommend an optimal size of the QM region for HIV PR studies of 8 Å of surroundings (~1300 atoms).

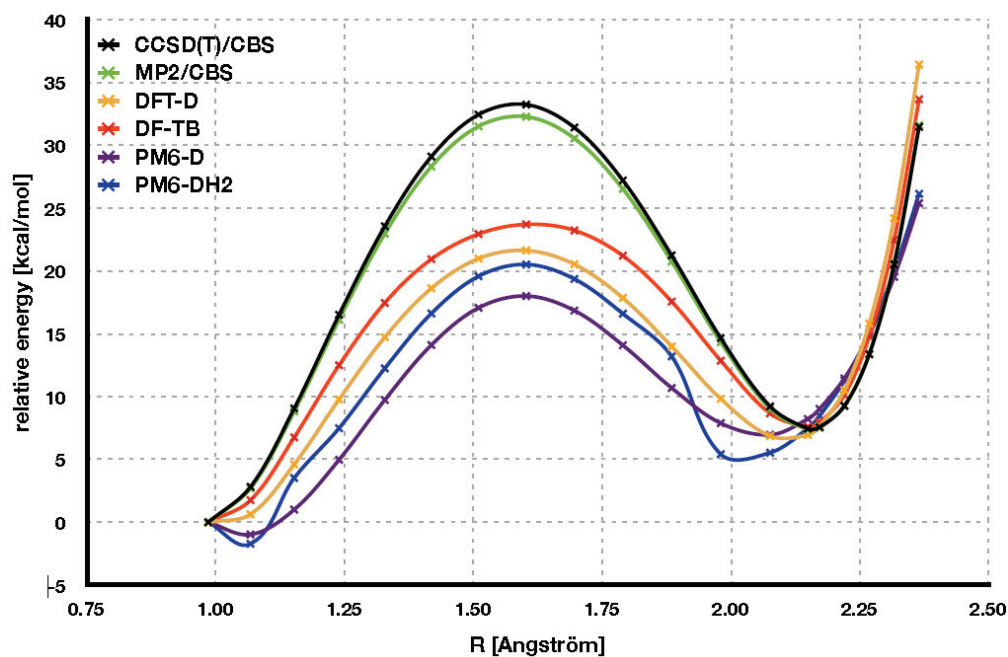


Fig. 3.9: The graph of proton-transfer barriers calculated in several QM-based methods.

Proton transfer has occurred at both the SQM and DFT levels, which underlines the requirement for a QM approach. We thus checked this phenomena by studying proton transfer barrier heights on a small model of the monoprotonated carboxyle pair using DFT, SQM and high-level MP2 and CCSD(T) methods (Figure 3.9). The results not only confirmed a well-known tendency of DFT GGA functionals to underestimate the reaction barriers²¹¹ but also showed an even greater underestimation at the PM6 level and shifting of the minimum towards intermediate positions of the proton between two oxygens.

To conclude, we have presented hybrid QM/MM calculations on a biomolecular system where the protonation phenomena play a pivotal role. The need of a QM-based approach to describe correctly molecular systems in which proton transfer can occur is evident.²¹² We have thus introduced the novel computational protocol using corrected PM6-DH2 and DFT-D method which is useful not only for a determining the most probable protonation states but also for assessing stabilities of various isomers (conformers/tautomers) in biomolecular complexes.

This general methodology was here applied on two complexes of HIV-1 protease. We have shown that in HIV-1 PR/inhibitor complexes with two orientations of the ligand, the symmetry-related pairs of the protonation variants are also energy-related. We have thus confirmed that, if using relaxed QM region, the only one orientation of the inhibitor is sufficient to correctly describe energetics in the active site of HIV-1 PR complexes. We have identified the most stable protonation variant of HIV-1PR/DRV complex that agrees with suggested location of the proton in an atomic resolution of the DRV with HIV-1 PR mutant.²¹³ These findings are the experimental verification of our computational approach. We have also revealed that the Asp30/Glu-P2' carboxylate pair is monoprotonated on the Asp30 and that the acceptor of the hydrogen bond from the hydroxyl group of the KI2 is most probably the OD2 oxygen of the Asp25'. In agreement with the experiment is also our finding that the major A conformation of the KI2 is in the complex with HIV-1 PR more stable than the conformation B.

We can also conclude that the corrected PM6 QM/MM calculations using QM part extended up to 8 Å (~1300 atoms) gives the same qualitative picture as DFT-D QM/MM calculations that are limited by a size of the QM region (up to 500 atoms). The bigger QM parts increases the risk of unrelated distant structural changes that can affect the energetics of the active site. Comparing with the high-level QM methods on the small model of carboxylate pair has shown that proton transfer barriers are underestimated by DFT using

GGA functional and even more by the PM6-D and PM6-DH2 method. These findings pointed to the need for better corrections or even more reparametrizations of PM6 method which would also describe proton transfer. It has been consequently shown that in the DH2 formalism, a proton transfer along a hydrogen bond exhibits a discontinuous potential energy surface and the requirement of better correction was fulfilled by introducing a new generation of corrections in PM6-D3H4.⁷⁰

3.2.2 Secreted Aspartic Protease of *Candida Parapsilosis*

Candida parapsilosis is a fungal species that causes a wide variety of hospital-acquired infections and sepsis in immuno-compromised patients and thus presents a serious problem, particularly in neonatal intensive care units.²¹⁴ *C. parapsilosis* has been isolated most frequently from the human hands but also from nonhuman sources like domestic animals, insects and soil.²¹⁵ *Candida* species secrete hydrolytic enzymes, namely aspartic proteases, lipases and phospholipases which facilitate penetration of the pathogens through host tissues. Secreted aspartic proteases (Saps) of pathogenic *Candida* thus represent possible targets for drug design.

Two *C. parapsilosis* isoenzymes, Sapp1p and Sapp2p, in complexes with the classical aspartic protease inhibitor pepstatin A have been recently crystallized^{216,217} In the cooperation with Dostál group the Sapp2p/pepstatin A structure has been determined at the atomic resolution of 0.825 Å and quantum mechanical calculations have been employed to understand the differences in pepstatin A binding to Sapp1p and Sapp2p on an accurate quantitative basis. We specifically aimed here to unveil the features of the structure which were not accessible to the crystallographic experiment, *i.e.*, *i*) the disproval of the presence of the third proton in the active site and *ii*) the analysis of energy contributions of all important aminoacids in the active site.

Sapp2p protein structure comprises large substrate-binding cleft located between two topologically similar N- and C- terminal domains. Each domain contains a conserved sequence of aminoacids which get closer to each other and form the catalytic aspartate dyad (Asp 32, Asp 211). The Sapp2p cleft is covered by the active site flap which plays an important role in substrate binding by adopting a close conformation. The active site is

further lined by four entrance loops. In our protein-ligand complex the active site is occupied by the substrate-mimicking inhibitor pepstatin A. (Figure 3.10)

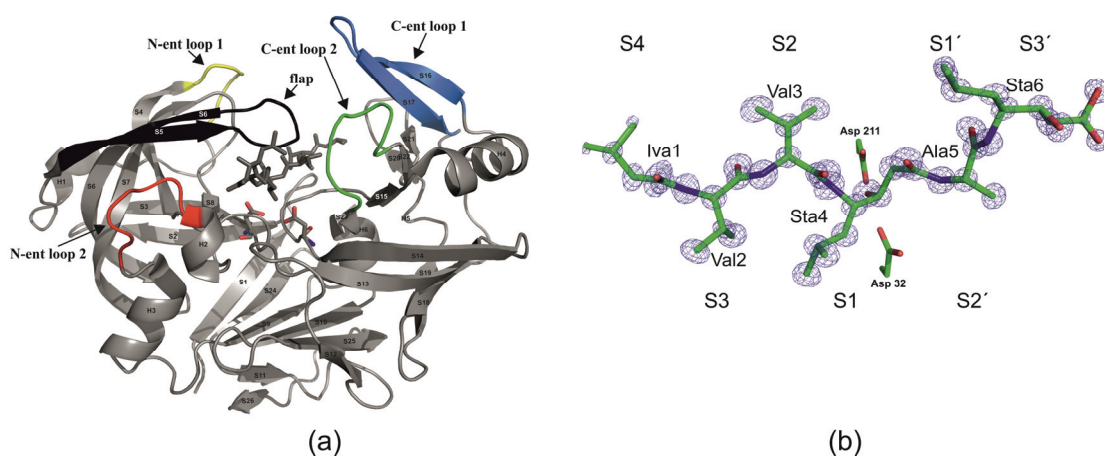


Fig. 3.10: (a) Sapp2p/pepstatin A complex, where protein is shown as ribbon and pepstatin A and catalytic aspartates as sticks. (b) Pepstatin A binding pose in Sapp2p with depicted electron density map in contours. Further names of peptidomimetic pepstatin A residues (Iva1-Val2-Val3-Sta4-Ala5-Sta6), corresponding substrate binding subsites (S4-S3') and catalytic aspartates in sticks are indicated.

The isoenzym Sapp1p is the closest sequence and structural homolog of Sapp2p. The sequence homology is over 80% and the RMSD of their structural superposition of 330 aligned C α atoms amounts to 1.25 Å. Despite a high structural similarity (both isoenzymes have two pairs of cysteins with similar S-S bridge topology and one serine residue within topologically similar loops with low sequence homology), there are noticeable differences. The major differences are located at the loops that line the entrance to the binding cleft. The entrance loops of Sapp2p, differing from Sapp1p case in a deletion of eight aminoacids in one loop and insertion in another one, are in direct contact with C-terminal residue of pepstatin A and their conformations thus significantly affect the shape, size and character of the binding site.

The atomic resolution of the Sapp2p/pepstatin A ($R=0.83\text{\AA}$) allowed to find numerous hydrogen atoms in the difference density maps and thus revealed hydrogen bonding networks. One of these hydrogen networks leading to the catalytic aspartate residue is critical for positioning the carboxyl group of Asp32 to one plane with the carboxyl group of the second catalytic residue Asp211 (for details see Figure 2 in Appendix E). Protonation states of the catalytic aspartates can be assigned by measuring of the bond

lengths which are in such a high quality crystal on subpicometer scale. The protonation of the O δ 2 of Asp211 has thus been revealed by the comparison of the C-O interatomic distances. Moreover also hydroxyl hydrogen of the inhibitor Sta4 residue is visible in the difference electron density map. Contrarily, the C-O distances in the case of the second catalytic aspartate Asp32 were very similar to each other because the crystal structure reflects the superposition of two states. The first one is where both oxygens are deprotonated and the second minor one in which the lower oxygen (O δ 1) is protonated. The occupancy for these two states was estimated to be 60% and 40%, respectively, based on the C-O δ 1 distance. The suggested protonation of Asp32 O δ 1 can be achieved *via* the presence of the proton shared between both catalytic aspartates or by a transient shift of the Sta4 hydroxyl hydrogen toward Asp32 O δ 1. The possible protonation variants of catalytic aspartates are depicted in Figure 3.11.

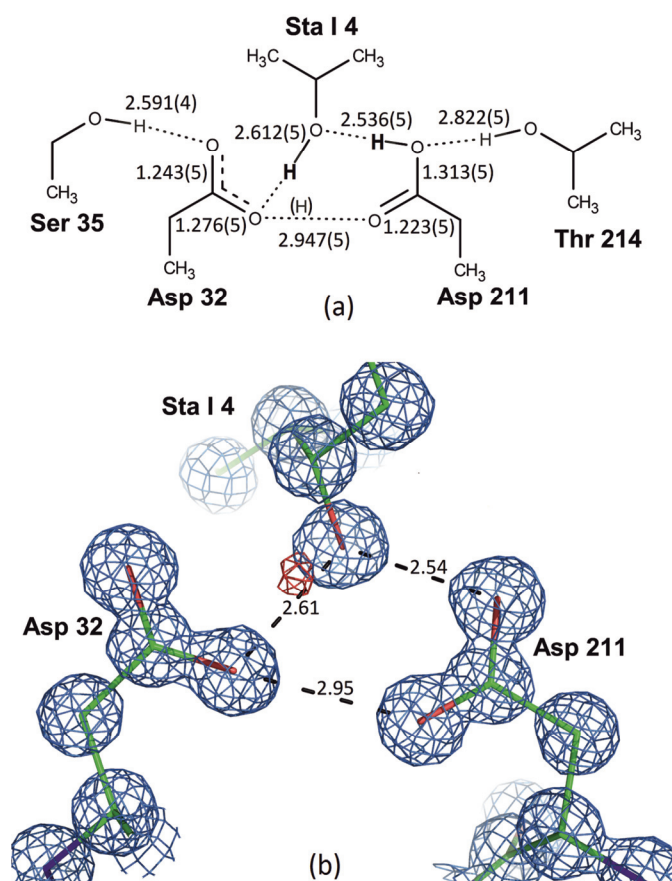


Fig. 3.11: (a) Schematic diagram of polar interactions in the active site of Sapp2p/pepstatin A complex. Clearly assigned hydrogens are in bold, hypothetical hydrogen atom is in parentheses. Distances are in Å with standard deviations in parentheses. (b) The active site in detail in stick

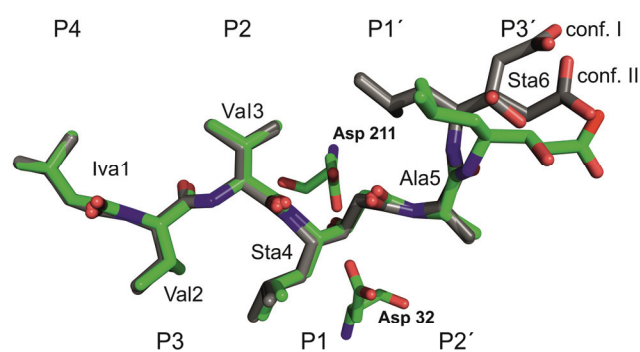
representation with contoured electron density map in blue color and different electron map in red color. Hydrogen bonds are shown as dotted lines with distances in Å.

We used two crystal structures, Sapp2p/pepstatin A (PDB code 4Y9W, $R=0.83\text{Å}$)²¹⁸ and Sapp1p/pepstatin A (PDB code 3FV3, $R=1.85\text{Å}$)²¹⁷ as starting points for our molecular modelling purposes. For the latter complex, two conformations of the ligand were considered. The computational models were prepared using our usual procedure of adding hydrogens (taking into an account protonation states of all residues according to the experimental pH, individual protonations of histidines and ligands) and their relaxation followed by MD-based simulated annealing. Similarly we also modelled aminoacids that were missing or not-well defined in electronic maps. The active site was protonated according to the crystallographic findings of the Sapp2p/pepstatin A complex. We used both variants, *i.e.* monoprotonated Asp211 on the O δ 2 oxygen and Asp32 either deprotonated or monoprotonated on the O δ 1 oxygen. All four model complexes, including two protonation variants of Sapp2p/pepstatin A and two conformations (I and II) of the pepstatin in the Sapp1p, were optimized using the QM/MM scheme. The QM part consisted of 8 Å surroundings of the ligands (approximately 1,560 atoms in total), whereas only atoms in 6 Å surrounding (~1200 atoms) were allowed to move. QM part was solved by PM6-D3H4 method⁷⁰ coupled with the COSMO implicit solvent model⁷² using the linear scaling algorithm MOZYME¹²⁴. The MM part was for speeding up of the calculations kept frozen.

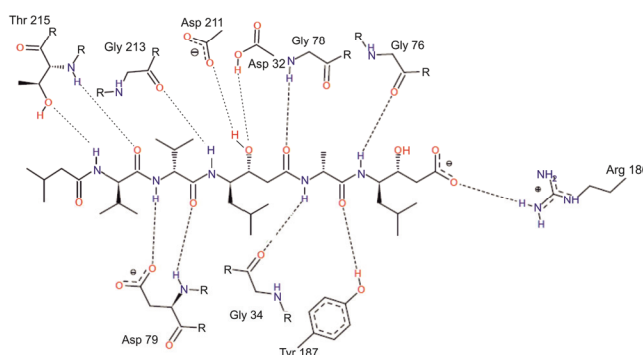
To dissect the energy contributions of the all important amino acids in the active site we applied a “virtual glycine scanning”²¹⁹, which was inspired by “computational alanine scanning”²²⁰ *i.e.* the interacting amino acids in the active site were substituted with glycine. The energy contributions of their side chains ($\Delta\Delta G'_{int}$) were calculated as the difference between the original $\Delta G'_{int}$ with the wildtype amino acid and the new $\Delta G'_{int}$ with the mutated glycine residue. Interaction ‘free’ energies ($\Delta G'_{int}$) of all the studied systems were determined on the whole optimized structures as single-point energies at the PM6-D3H4/COSMO level.

First of all, we aimed to shed light on possible protonation variants of catalytic aspartates. We have found that a shared proton cannot be accommodated in the Sapp2p active site for steric reasons, *i.e.* because of the repulsive interaction with the Asp32 the hydroxyl hydrogen of the inhibitor Sta4 residue moved towards a backbone of Gly213

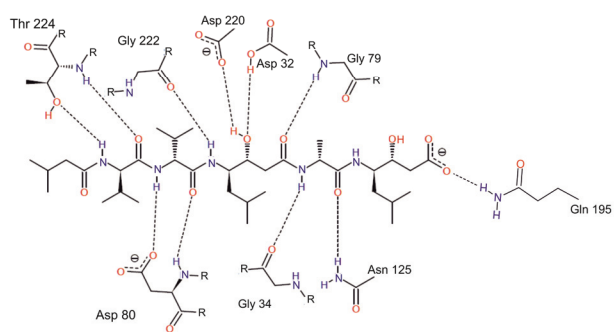
and thus did not fit to electronic maps. Moreover, the shared proton arrangement has far less favorable interaction energy (by about 10 kcal/mol) with the statin inhibitor than the model with only two protons in the active site. We have therefore suggested that the protonated state of Asp32 may indicate a transient shift of the statin hydroxyl proton toward O δ 1 of Asp 32.



(a)



(b)



(c)

Fig. 3.12: (a) Aligned structures of pepstatin A bound to both isoenzymes, *i.e.* Sapp2p (carbon atoms in green color) and Sapp1p in I a II conformations (carbon atoms in grey color). (b) and (c) Schematic representation of hydrogen bonding of pepstatin A with aminoacid residues of Sapp2p and Sapp1p, respectively.

As mentioned before, pepstatin A is a peptidomimetic inhibitor containing six amino acid residues in positions P4-P3' (Iva1-Val2-Val3-Sta4-Ala5-Sta6). It is bound to Sapp2p in an extended conformation, occupying the S4-S1 substrate binding pockets of the enzyme active site. The pepstatin A is bound into the Sapp1p in very similar conformation but differs in P3' position of Sta6 residue. Moreover two conformations (denoted I and II) of the Sta6 residue were observed in the crystal. The binding of the pepstatin A in Sapp2p and Sapp1p thus differs structurally, especially in three changes in hydrogen bonding of the P2' and P3' inhibitor moieties (see Figure 3.12).

Polar interactions between pepstatin A and both isoenzymes are mediated mainly by direct hydrogen bonds supplied by the pepstatin A backbone and water networks formed by P4 and P3' residues. Analogous hydrogen bonds are formed between the inhibitor and Sapp2p and Sapp1p isoenzymes except of some changes caused by a different sequence of aminoacids. In addition, pepstatin A makes numerous van der Waals interactions with the aminoacids in the active site. We have thus performed the “virtual glycine scanning” to quantify all structural differences of the binding of pepstatin A in Sapp1p and Sapp2p isoenzymes. The overall results are shown in the Figure 3.13.

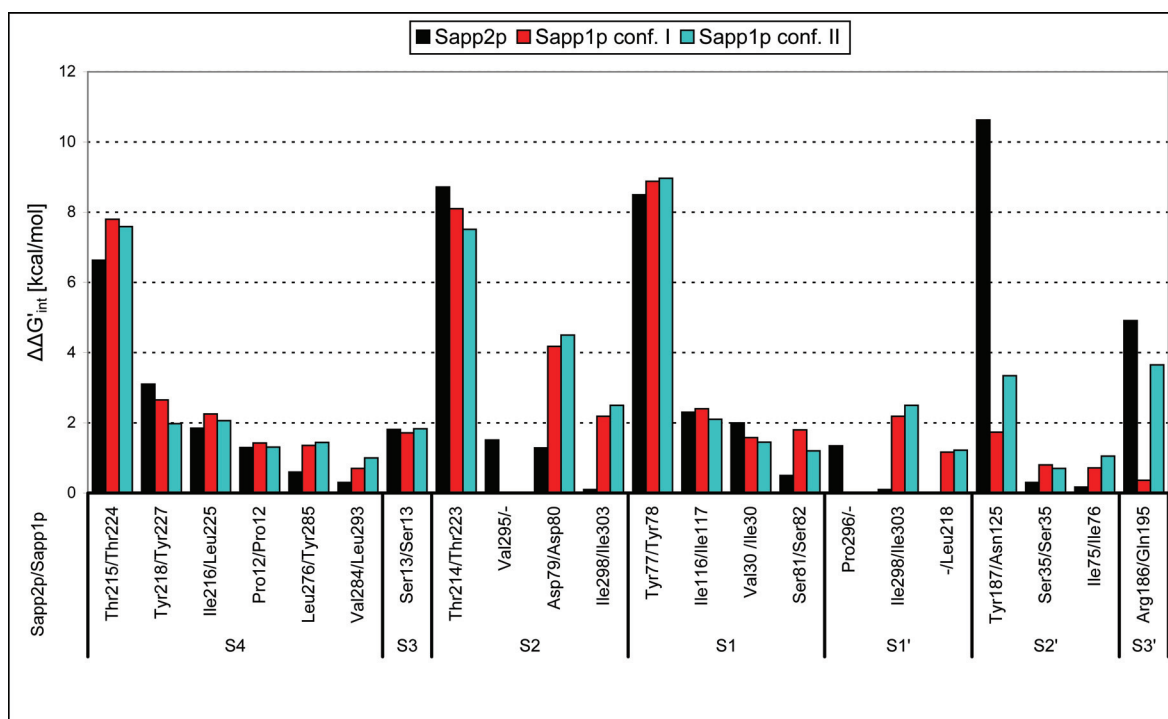


Fig. 3.13: The changes in free energy of interaction ($\Delta\Delta G'_{int}$) in kcal/mol upon mutation of a given amino acid residue to glycine to study the roles of individual amino acid side chains of Sapp2p and Sapp1p active sites in binding with pepstatin A.

As seen in Figure 3.13, individual substrate binding subsites of Sapp2p and Sapp1p (S4-S3') possess amino acid residues that can be divided into four categories: *i*) identical residue, *ii*) similar residue, *iii*) different residue and *iv*) residue that does not form corresponding pairs due to different tracing of the protein backbone. We should keep in mind, that the energy contributions evaluated by the glycine scan inherently contain the effect of hydrogen bonding mediated by the residue side chain and thus the role of glycine residues or residues interacting via their backbone (e.g. Thr 215/Thr 224 in the S4 subsite and Asp 79/Asp 80 in the S2 as shown in Fig.3.12) cannot be evaluated. It was however shown, that these residues have similar conformations in both isoforms and thus the contributions of these residues are likely very similar. The only exception in this category is the Gly76...Sta6 hydrogen bond, which is present in Sapp2p and absent in Sapp1p.

The evaluation of individual subsites has shown that in the S4 subsite all amino acid pairs feature similar energy contributions within 1kcal/mol except of the strongest interaction (around 7 kcal/mol) of Thr215/Thr224, which was mediated by a combination of aliphatic...aliphatic interactions in the S4 subsite and hydrogen bonding in the S3 subsite. This subsite is relatively exposed to the solvent and the only interaction that is energetically identical for both isoenzymes was mediated by Ser13 residue. The similar contributions were observed also for conserved residues in the S2 subsite, where the strongest one was mediated by Thr214/Thr223 pair due to a combination of aliphatic...aliphatic dispersion interactions here and hydrogen bonding in the S1 pocket. The largest differences in Sapp2p and Sapp1p contributions of pepstatin A binding in the S2 subsite have been assigned to two following pairs. The Ile298/Ile303 pair has in the S2 pocket very weak (0.5 kcal/mol) methyl...methyl dispersion interactions for both isoenzymes, whereas it is enhanced by an additional interaction of the ligand in S1' for the Sapp1p case. The second difference is interestingly mediated by the conserved Asp79/Asp80 pair with an identical interaction pattern (two hydrogen bonds and a van der Waals interaction). The binding of pepstatin A to Sapp2p and Sapp1p has however revealed a difference of 3-4 kcal/mol in energy. We have shown that this difference can be ascribed to the long-range electrostatic influence of Lys49 and Lys80, located approximately 7 Å from the charged Asp79 side chain only in the case of Sapp2p. The analysis of the S1 subsite has revealed similar contributions for both isoenzymes with the strongest contribution of Tyr77/Tyr78 pair (almost 9 kcal/mol) due to main-chain/main-chain hydrogen bonding combined with CH... π interactions. The energy

changes in the last three binding pockets (S1', S2' and S3') are caused by the totally different binding moieties of the pepstatin A in Sapp2p and Sapp1p (see Fig. 3.12). In the S1' subsite, Sapp2p/pepstatin A features only one interaction in the case of Pro296 (nonpolar aliphatic...aliphatic type). However Sapp2p does not have a counterpart, this loss is compensated by Leu218 and even improved by a favorable interaction with Ile303. The most significant contribution to the binding of pepstatin A in Sapp2p has been revealed from the quantification of Tyr187. The reason of the strongest interaction ($\Delta\Delta G'_{int}=10.6$ kcal/mol) among all the calculated contributions is a very short hydrogen bond (O...O distance of 2.6 Å) between the phenolic hydroxyl of Tyr187 and the backbone carbonyl of the Ala in P2'. Moreover, the Tyr187 CZ...OH bond length of 1.337 Å suggests that the proton is shared between the two oxygen atoms. Contrarily, in the Sapp1p/pepstatin A the interaction with the carbonyl of the Ala in P2' is mediated by a medium-strong hydrogen bond with Asn125. The P3' terminal carboxylate of pepstatin A Sta6 is exposed to the solvent and thus forms hydrogen bonds with water molecules. Moreover it contributes to the total binding by about 5 kcal/mol coming from a salt bridge with Arg186 in Sapp2p, which is in conformation II in Sapp1p functionally replaced by a charge-assisted hydrogen bond with Gln195 (by about 3.6 kcal/mol).

To conclude, we successfully applied QM approach to quantitatively describe the strength of non-covalent interactions, including quantum effects such as proton transfer, in another protein-ligand system. Using the SQM approximation, we were able to include over 1,000 atoms in the QM part and thus capture the long-range effects, such as electrostatic interactions. We shed light on protonation variants of the catalytic aspartic dyad and suggested both oxygens of Asp32 to be deprotonated. We used a virtual glycine scan, using a fast and reliable SQM method PM6-D3H4X, to study the roles of individual amino acid side chains in binding of pepstatin A in two *C. parapsilosis* isoenzymes. Although the interactions of pepstatin A with Sapp2p and Sapp1p are mostly similar, we noted several mutually compensating differences for the binding of pepstatin A to Sapp2p and Sapp1p. Our conclusions are in line with the similar values of the measured inhibition constants of 0.4 and 0.3 nM, for the binding of pepstatin A to Sapp2p and Sapp1p, respectively.

3.2.3 Carborane-based Inhibitors of Carbonic Anhydrases

Carbonic anhydrases (CAs), *i.e.* monomeric zinc metalloenzymes catalyzing the reversible reaction of carbon dioxide hydration and bicarbonate dehydration, play important roles in many physiological processes, such as maintaining the acid-base balance and facilitating the transport of carbon dioxide and protons out of tissues. The α -CA family found in humans consists of 12 catalytic CA isoforms. They can be localized in cytosol, mitochondria, secretory or membranes of various tissues. While the CAII, ubiquitous isoform essential for the maintenance of general acid-base balance, is one of the most studied isoenzymes with a wealth of structural and biochemical data²²¹, CAIX isoenzyme is expressed selectively in a range of hypoxic tumors and is a validated diagnostic and therapeutic target.²²²⁻²²⁴ Clinical regulation of the activity of human CAs is in general a reliable therapeutic method for a number of human diseases, such as high blood pressure, glaucoma, hyperthyrosis, hypoglycemia, osteoporosis, neurological disorders and cancer.²²⁵ A selective inhibition of these ubiquitous enzymes is thus a very important aspect of drug design.

The CAs inhibitors can be classified into three main classes followingly: metal ion binders (sulfonamides, sulfamides, sulfamates, dithiocarbamates, thiols, and hydroxamates), compounds that anchor the zinc-coordinated water molecule/hydroxide ion (phenols, carboxylates, polyamines, esters, and sulfocoumarins) and coumarins and related compounds that bind further away from the metal ion.²²⁶ The most important classes are inorganic anions and sulfonamides, containing *i*) weakly acidic $\text{SO}_2\text{-NH}_2$ head group approaching the zinc ion and *ii*) the tail of the inhibitor molecule which can be substituted by specific functional groups to provide further interactions with the amino acids of CAs.²²¹ The strength of the inhibitor binding comes from the interaction of the head group to the metal ion, whereas selectivity against different isoforms comes from various interaction patches of the active site (hydrophobic pocket and hydrophilic faces), where the inhibitors bind via van der Waals and polar interactions.²²⁷

Novel carborane-based sulfamide inhibitors of CAII and CAIX have been recently designed and shown to inhibit the enzymes in submicromolar range.¹⁹⁶ Boranes are inherently electrodeficient polyhedral boron clusters that exhibit an astonishing variety of stable three-dimensional structures. Their building blocks are triangles of boron atoms held together by delocalized electron-deficient three-center two-electron bonding with

an extensive electron delocalization.²²⁸ *Closo*-carboranes are symmetrical 12-vertex icosahedron heteroboranes in which one or more {BH} vertex is replaced with {CH} vertex, whereas removing BH- vertex leads to open-cage *nido*-carboranes optionally possessing a B-H-B hydrogen bridge.²²⁹ The properties which make carboranes biologically active compounds include their resistance to catabolism, non-toxicity, high thermal and chemical stability, hydrophobicity, shape and 3D aromaticity.²³⁰⁻²³⁴ Recently, carboranes have been successfully applied as hydrophobic pharmacophores, *e.g.* estrogen receptor agonists/antagonists²³⁵, boron-containing antifolates²³⁶, HIV protease inhibitors^{195,237}, vitamin D ligands^{238,239} among others.^{232,240}

The nature of the non-covalent binding of carboranes to biomolecules range from B-H...H-C dihydrogen bonds²⁴¹⁻²⁴⁴ via B-H...Na⁺ bridges^{245,246} to B₂H... π and C-H... π hydrogen bonds.²⁴⁷ For example, the leading role of van der Waals and electrostatic interactions were found in the case of dihydrofolate reductase carborane inhibitors²³⁶, whereas for a carborane ligand of the vitamin D receptor, a “hydrophobic interaction“ was postulated.²³⁹

We have studied the nature of binding of two carborane-based sulfamide inhibitors bearing *closo*- and *nido*-carborane cages to well-known CAII isoenzyme (Figure 3.14). We used crystal structures of CAII with 1-methylenesulfamide-1,2-dicarba-*closo*-dodecaborane (**1a**) and 7-methylenesulfamide-(7,8-*nido*-dicarbaundecaborate) (**7a**) at resolutions of 1.35 Å and 1.55 Å, respectively (PDB codes 4MDG, 4MDM).

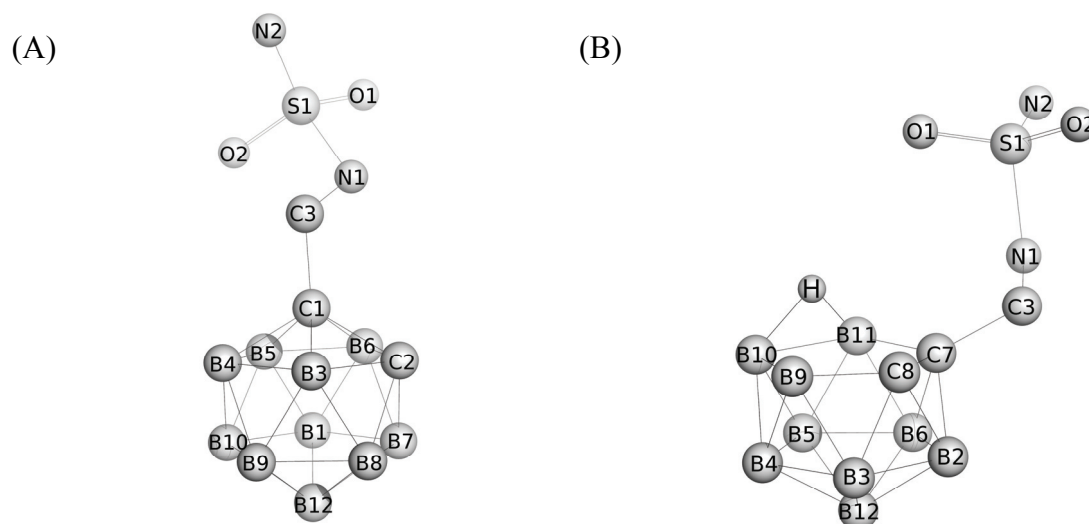


Fig. 3.14: (A) 1-methylenesulfamide-1,2-dicarba-*closo*-dodecaborane (**1a**) structure (B) 7-methylene sulfamide-(7,8-*nido*-dicarbaundecaborate) (**7a**) structure. Hydrogens are omitted for clarity with the exception of a B-H-B bridge in **7a**.

These high-resolution structures clearly revealed a binding mode of both inhibitors in the active site, which is well conserved in sequence among various isoenzymes. It has a conical shape and contains a Zn^{2+} ion coordinated by three histidine residues (His94, His96, His119), which are held in a distorted tetrahedral geometry. The sulfamide head group of the inhibitors is coordinated to the zinc ion and the carborane cluster fills the binding pocket (Figure 3.15).

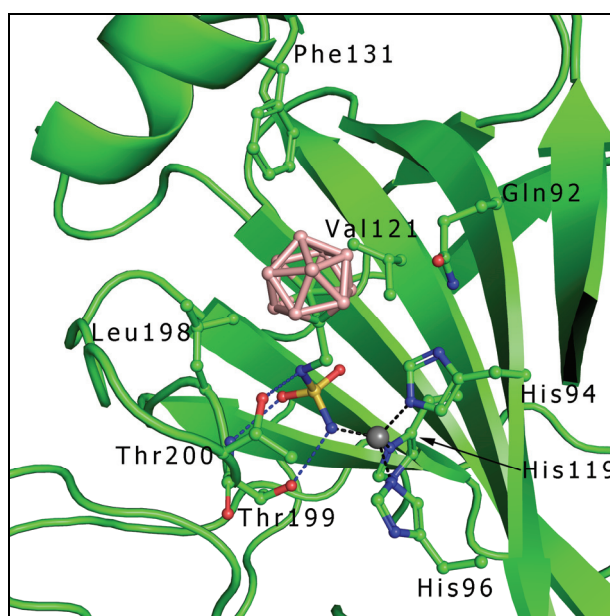


Fig. 3.15: The binding motif of the compound **1a** in CAII. Protein is shown in cartoon representation; residues involved in interactions with the Zn^{2+} ion (gray sphere) and **1a** are shown in stick representation. Polar interactions are represented by blue dashed lines; Zn^{2+} ion coordination is shown as black dashed lines.

We aimed to theoretically explain: *i*) the binding of two various carborane-based inhibitors to the CAII protein. Specifically, it was not clear which physical forces drive the binding of the carborane cages, *e.g.* hydrophobicity of the carborane cage, dispersion interactions, an effect of the cage on the pKa of the sulfamide moiety or the formation of dihydrogen bonds. *ii*) the stability of all possible rotamers and enantiomers of studied inhibitors and finally we aimed to quantify *iii*) contributions of all important aminoacids in the active site of CAII and CAIX isoenzymes.

To answer these questions properly and to gain deeper insight into the nature of the interactions, we performed ONIOM-like QM/MM calculations and virtual glycine scanning procedure.

All computational models were treated following our standard procedure of the structure preparation. Protonation of the enzyme was done to reflect the predominant state at pH 7, with special care of manual protonation of histidines based on visual inspection of their surroundings. The sulfamide moiety binds to the Zn^{2+} of CA in a deprotonated NH^- form and was thus modeled accordingly.²²¹ The *closo*-carborane-based inhibitor (**1a**) has five possible rotational isomers (rotamers), differing in the positions of the carbon atom (C2) in the lower pentagon of the cage, while *nido*-carborane-based inhibitor (**7a**) has two possible positions of the carbon (C8) atom (2 enantiomers) combined with two positions of the B-H-B bridge (B9-H-B10 or B10-H-B11) (see Figure 3.14).

The positions of added hydrogen atoms were relaxed in vacuo using AMBER forcefield, followed by annealing (10 ps) from 600 to 0 K. Complexes of all isomers in the complex with CAII were fully optimized using QM/MM approach, where the QM part consisted of 480 atoms (4 Å surroundings of the ligands) which is around the current limit for used DFT-D calculations (TPSS/TZVP//BLYP/SVP combination of basis sets and functionals was used here for all single point energy calculations and optimizations, respectively). The rest of the protein was solved by MM. Surrounding solvent was approximated by GB model, except of one explicitly treated structural water molecule bridging three protein amino acids and the inhibitors.

The first step was to identify the most stable isomers of the **1a** and **7a** inhibitors. This was done on the basis of the QM/MM energies of the optimized structures in GB solvent together with the analysis of isolated compounds using DFT-D TPSS/TZVP in vacuum.

In the case of **1a** compound, we studied the rotamer preferences obtained by a rigid scan of the N1-C3-C1-C2 dihedral. QM/MM energies then revealed how these preferences are influenced by the protein surroundings. The results are shown in Figure 3.16.

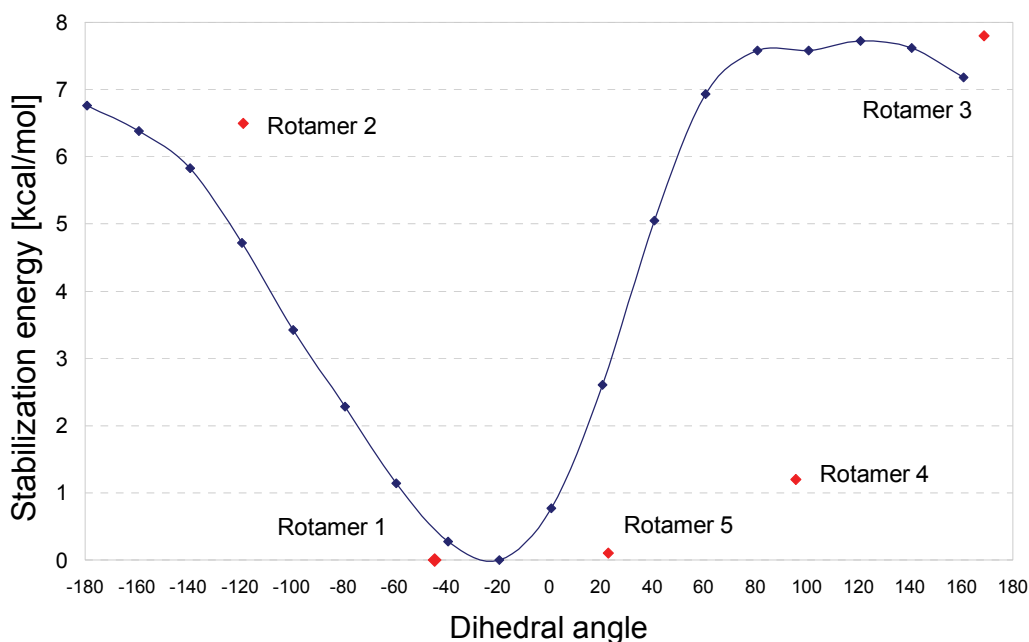


Fig. 3.16: The rotamer preferences of **1a** in the complex with CAII. The rotational profile of isolated inhibitor is shown by blue line, whereas relative stabilization energies of five rotamers in the complex with CAII are shown by red points.

The rigid scan showed that the carborane moiety of isolated **1a** preferred dihedral angle around -20° within the possible range of 80° to $+21^\circ$ with the energy difference up to 2 kcal/mol, whereas the rotational barrier of full 360° rotation was 8 kcal/mol high. The highest stability of the rotamer 1 was caused by weak hydrogen bond between the C2-H vertex and the oxygen of the sulfamide head group. Contrarily, an electrostatic repulsion between the B-H group and the oxygen was the reason of unfavorable energies of the less stable rotamers. The relative QM/MM energies revealed that the well around the minimum in the complex geometry is broader, *i.e.* allowing wider rotation, however the computed minimum agreed with experimentally determined position of the carbon atoms of **1a** structural data (a dihedral angle of -44°). The barrier for the 360° rotation remained about 8 kcal/mol high.

In the case of **7a** compound, we studied the preference of the position of the C8 atom (two enantiomers) and the position of the hydrogen bond bridge, there were thus four isomers of *nido*-carborane-based inhibitor. The relative energies of both enantiomers with both positions of the B-H-B bridge differed only by about 1.5 kcal/mol for isolated systems and both of them should thus be considered. Calculations in the complex geometry with CAII showed that the ρ isomer with the bridged hydrogen between B10 and B11 (Figure

3.14) was by about 3 kcal/mol more stable than others. Moreover the position of C8 atom was in agreement with crystallographic observations. Because of the low energetic barrier we thus assumed that other isomers can be also found in the complex, however they are less populated.

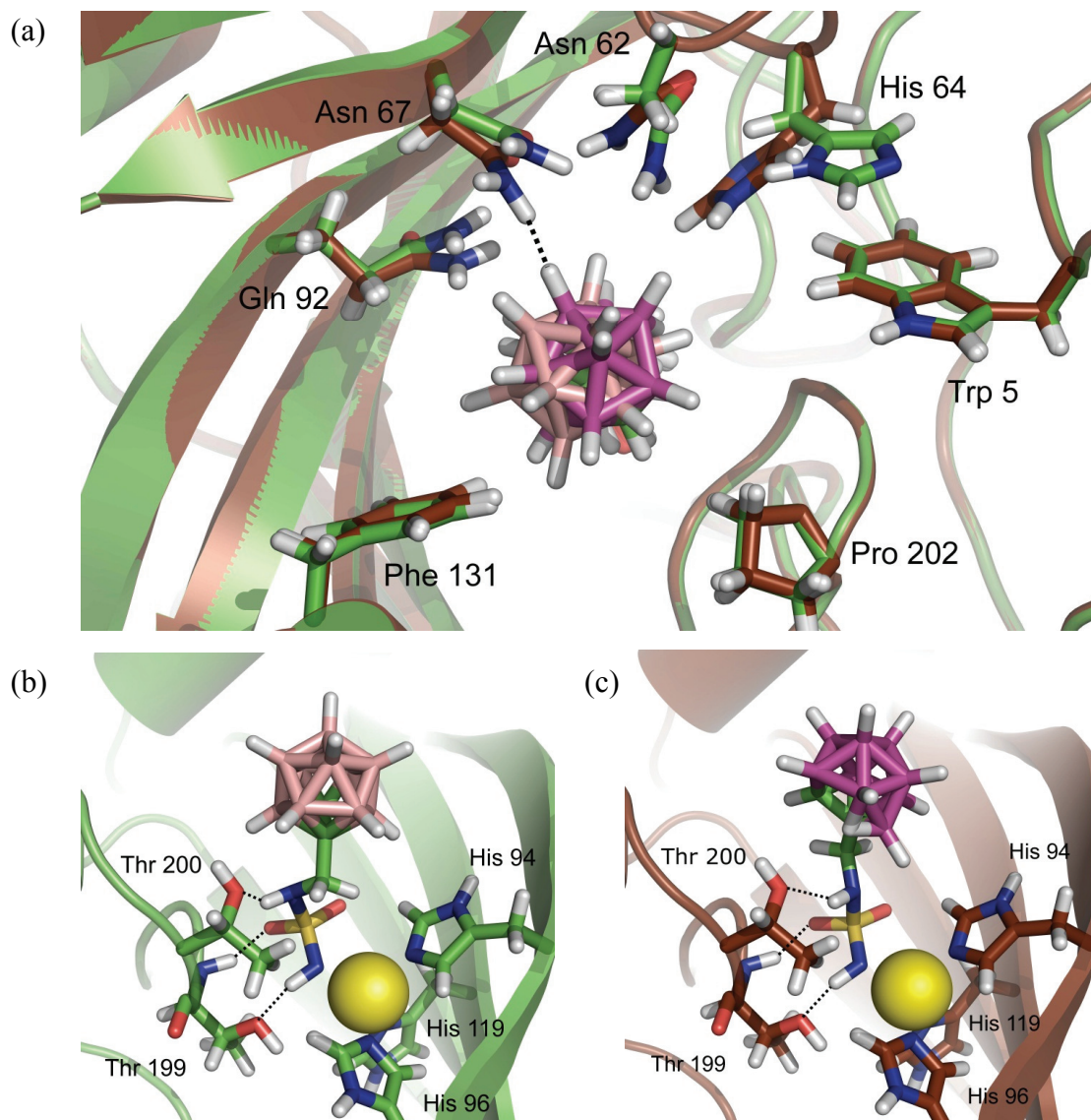


Fig. 3.17: Superposition of the QM/MM optimized complexes of CAII with **1a** and **7a** inhibitors, stressing in (a) the interactions of carborane cages (pink for **1a** and magenta for **7a**) with CAII amino acid sidechains (CAII/**1a** complex in green and CAII/**7a** complex in brown) and the different interactions of sulfamide moiety of **1a** (b) and **7a** (c) where Zn²⁺ is visualized as a yellow sphere.

The binding of **1a** and **7a** to the CAII protein slightly differ in the orientation of amino acid sidechains interacting with the carborane cages and also in the interaction of sulfamide linker (Figure 3.17). We fragmented the ligands into two parts (the carborane

cage and the sulfamide head group) and capped them by hydrogen atoms and calculated their interaction ‘free’ energies ($\Delta G'_{int}$) with the CAII active site (QM part). Because of the fact that the sulfamide head group interacted directly with the Zn^{2+} cation, we applied more accurate COSMO salvation instead of GB model used during the QM/MM optimization. An accurate calculation of the desolvation free energy of the bare cation is, however, a very difficult task.²⁴⁸ To decrease the error of the calculated $\Delta G'_{int}$ we incorporated a single explicit water molecule as the first solvation shell of Zn^{2+} to screen its charge, following Equation 3.1.²²¹



, where RSO_2NH^- stands for the deprotonated sulfamide form of **1a** or **7a**.

Interaction ‘free’ energies ($\Delta G'_{int}$) of ligands and its fragments as well as the experimental binding affinities are summarized in Table 3.1.

	$\Delta G'_{int}$	ΔE_{int}	<i>Disp</i>	ΔG^o_b
1a <i>closo</i> -carborane based inhibitor	-37.7	-184.8	-42.0	-8.4±0.1
7a <i>nido</i> -carborane based inhibitor	-36.8	-246.6	-38.3	-7.0±0.2
1a sulfonamide head group	-26.5	-162.4	-17.1	
7a sulfonamide head group	-26.1	-165.3	-16.8	
1a <i>closo</i> cage	-11.2	-22.4	-24.9	
7a <i>nido</i> cage	-10.7	-81.3	-21.5	

Table 3.1: The decomposition of the DFT-D (TPSS/TZVP) interaction energy between ligand or fragments of ligands and the QM part of CAII. The interaction “free” energy ($\Delta G'_{int}$) calculated in the COSMO solvent model, the gas-phase interaction energy (ΔE_{int}), the dispersion energy (*Disp*) contribution to the interaction energy, experimental binding free energy ΔG^o_b , all energies are in kcal/mol.

Results revealed a good agreement of calculated and experimental relative binding free energies between **1a** and **7a** ($\Delta \Delta G'_{int}$ of -0.9 kcal/mol as compared to $\Delta \Delta G^o_b$ of -1.4 kcal/mol). The $\Delta G'_{int}$ of sulfonamide moiety was significantly stronger than of the carborane cages by about 15.3 kcal/mol on average, which is in agreement with studies revealing its energetic importance.²²¹ However the sulfonamide head groups of **1a** and **7a**

interact with the same strength, the difference between ΔE_{int} of *closo* and *nido* carborane cages was significant. The *closo* cage interacts by 58.9 kcal/mol less strongly than the *nido* cage and the dispersion energy itself played a major role in its binding. In contrast, the dispersion energy of the *nido* cage contributed only about 26.5% of the total ΔE_{int} . The driving force of the *nido* cage hence seemed to be of an electrostatic character. The resulting interaction “free” energy of the *closo* and *nido* cages was however comparable, because of the high desolvation penalty of charged *nido*-carborane cage.

To quantify the roles of individual amino acid sidechains in the active site, we performed the virtual glycine scan. The $\Delta\Delta G'_{int}$ upon single amino-acid mutation into glycine for the most stable **1a** rotamer and the most stable **7a** isomer is shown in Figure 3.18.

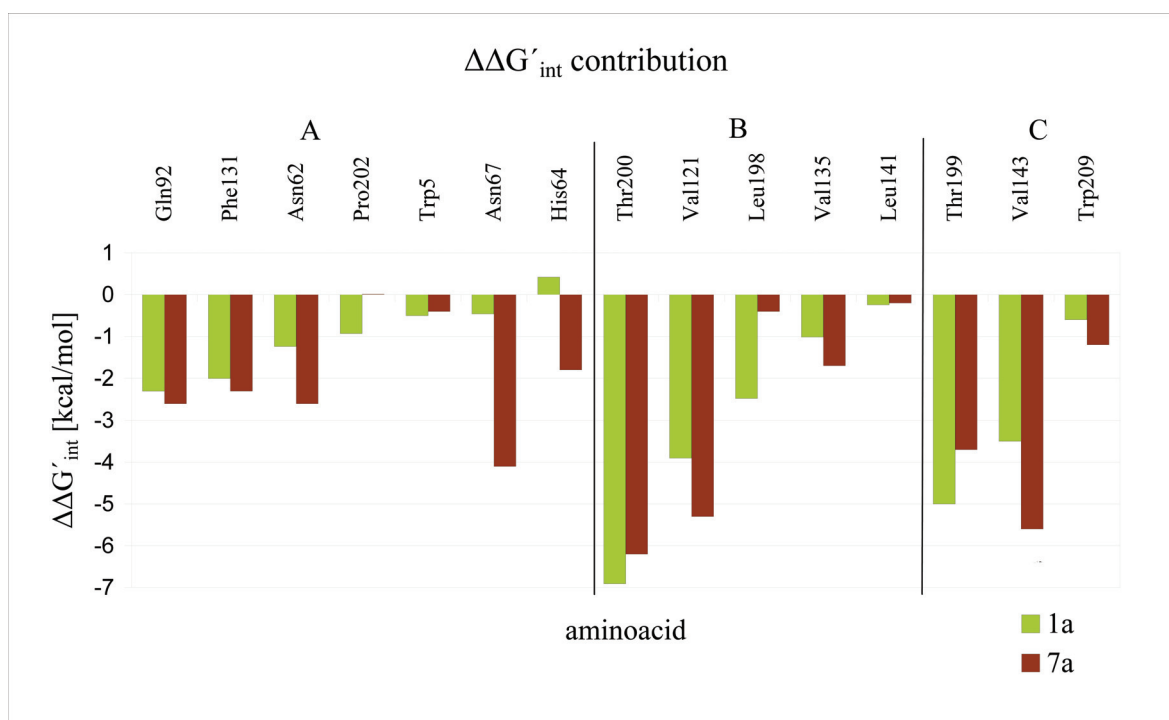


Fig. 3.18: The contribution of single amino acids to the interaction “free” energy $\Delta\Delta G'_{int}$ as obtained from a “virtual glycine scan”. **A)** The first 7 amino acids (from Gln92 to His64) interact only with the carborane cage, **B)** the next 5 (from Thr200 to Leu141) have interactions with both, the cage and the sulfamide head, respectively and **C)** the last 3 only with the sulfamide head (from Thr199 to Trp209).

The binding mode of the sulfonamide moiety differed among inhibitors, however the interaction ‘free’ energies were similar. The most significant difference was revealed in

the interaction with Leu198, where the CAII/**1a** had a more favorable interaction by about 2 kcal/mol due to the presence of a weak CH...N hydrogen bond as opposed to van der Waals interactions only for **7a**.

The *closo* cage of compound **1a**, whose binding was driven mainly by dispersion energy revealed the strongest energy contribution via 2.0 Å short dihydrogen bond with Gln92 and 2.2 Å short interaction with Phe131, whereas the $\Delta\Delta G'_{int}$ did not exceed -2.5 kcal/mol. In general, the dihydrogen bonds of the *closo* cage were weak and rather long (at the margin of the range of H...H distances). *Closo* carborane B-H⁻ vertexes interacted with non-polar C-H of Phe131, Pro202 and Asn62. The second group of aminoacids (Figure 3.18 B) mediated besides dihydrogen bonding with the *closo* cage also strong interactions with the sulfamide head group.

The virtual glycine scan of rotamer 4 explained the reason of its higher stabilization in the complex with CAII protein than in isolation (Figure 3.16). In contrast of the most stable rotamer 1 it had stronger interactions with Thr200, Val121 and Gln92 by about 3.1 kcal/mol in total. Both rotamers differed in the interaction with Thr200, *i.e.* a repulsion between the B-H vertex of rotamer 1 and the Oxygen atom of Thr200 (B-H^{δ-}...O^{δ-}) was replaced by a weak hydrogen bond (C-H^{δ+}...O^{δ-}) of rotamer 4.

The interactions of *nido* carborane cage differed significantly from *closo* cage. It interacted with the protein mainly via electrostatic interactions and formed short and strong dihydrogen bonds mainly with the polar hydrogens of NH₂ groups. The biggest contribution to the total binding ($\Delta\Delta G'_{int} = -4.1$ kcal/mol) was mediated by a short dihydrogen bond with Asn67 with the H...H distance of 1.7 Å. It should be mentioned that Asn67 had no other contacts and thus the calculated interaction can be directly assigned to the single dihydrogen bond. The neighboring Asn62 had also more attractive interaction than with *closo*-cage by about -1.2 kcal/mol. The second biggest difference concerned interaction with flexible His64 which interacted with *nido* by single dihydrogen bond in the distance of 2.0 Å, whereas in the case of **1a** is far away from the inhibitor.

This very first QM/MM study of the two novel carborane-sulfamide inhibitors of CAII has thus unveiled a detailed atomistic and energetic understanding of the nature of inhibitor binding. Whereas the neutral *closo*-carborane cage, bearing boron-bound hydrogens with slightly negative charge, were bound mainly via dispersion interactions and formed only very weak dihydrogen bonds; negatively charged *nido*-carborane, bearing more negative boron-bound hydrogen atoms, interacted with the protein mainly via electrostatic

interactions and formed very strong and short dihydrogen bonds. This knowledge can be utilized in tuning of the binding affinity of carborane-containing ligands in rational drug design.

We showed that various carborane clusters act as CA inhibitors which means that modifying these clusters with an appropriately attached sulfamide group and other substituents can lead to compounds with selectivity toward the cancer-specific CAIX isoenzyme. Recently, it was shown that **1a** compound exhibits inhibitory property to the CAIX isoenzyme with K_i values in submicromolar range.¹⁹⁶ Because of the lack of structural data, we modeled the binding of compound **1a** into the CAIX active site using QM/MM methods. The complex of CAIX/**1a** was modeled by aligning of the crystal structure of human CAII in complex with **1a** determined at 1.0 Å resolution (PDB code 4Q78)²⁴⁹ with the existing crystal structure of the CAIX catalytic domain (PDB code 3IAI).²⁵⁰ The substrate binding sites of CAII and CAIX differ in a shape of the active site cavity caused by variations of six amino acids, *i.e.* Asn67 of CAII is replaced by Gln in CAIX, Ile91 by Leu, Trp123 by Leu, Phe131 by Val, Val135 by Leu, and Leu204 by Ala. The preparation of the structure was performed as described before for all studied protein-ligand systems. The positions of the added hydrogen atoms, the inhibitor, and 15 amino acids surrounding the ligand were relaxed in a GB implicit solvent model⁷¹ using the ff03 AMBER forcefield followed by 10 ps annealing from 150 to 0K using Berendsen thermostat²⁰⁵ in SANDER module of AMBER package.²⁰⁴ We fully optimized the complex using a QM/MM approach, where the QM part was described at the DFT-D TPSS/TZVP//BLYP/SVP level of theory, comprising 8 amino acids and the ligand. The rest of the protein was solved as the MM part and the surrounding solvent was approximated by GB implicit solvent model except of one structural water molecule bridging the inhibitor and amino acid residues.

It was shown that the position of the inhibitor slightly differs in CAIX and CAII. (Figure 3.19). The position of sulfamide head group remained unchanged, whereas the carborane cage shifted by 2.1 Å. and interacts with the opposite site of the active site.

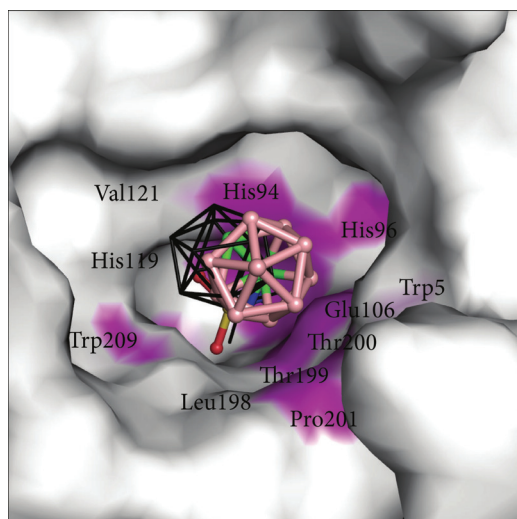


Fig. 3.19: Different binding poses of **1a** inhibitor in CAII (black lines) and CAIX (in pink). Atoms making contacts with **1a** are highlighted in magenta.

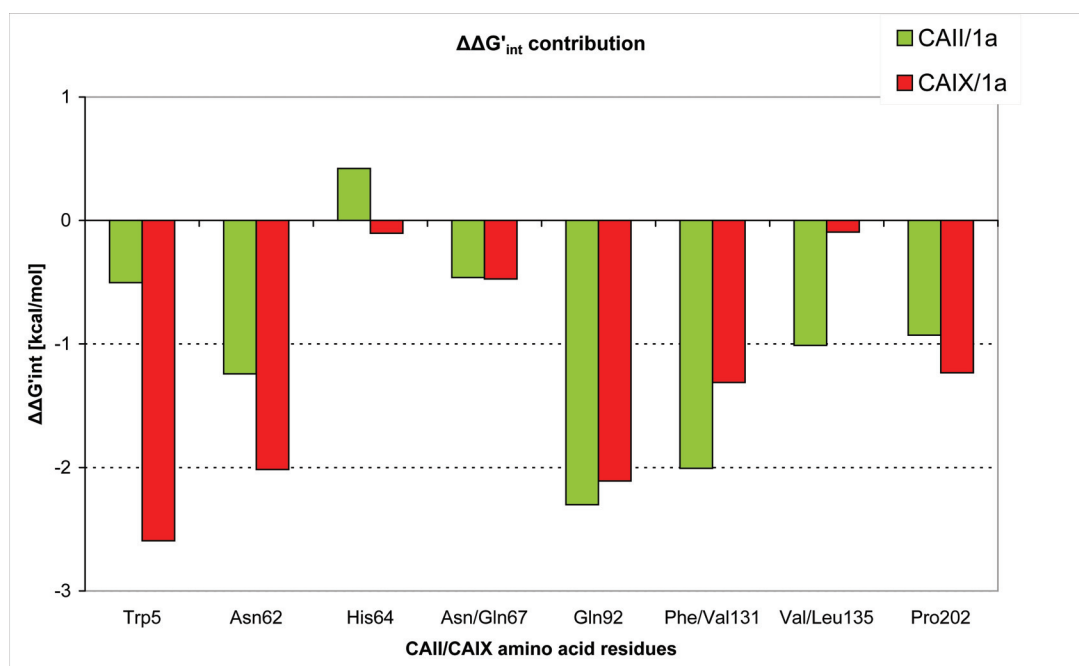


Fig. 3.20: Results of virtual glycine scan showing contributions of individual amino acid residues to the energy of binding of **1a** to CAII and CAIX, respectively.

We performed the virtual glycine scan procedure to reveal the differences in the binding of **1a** compound in CAII and CAIX. From the results (Figure 3.20) it is obvious that the largest difference (2.6 kcal/mol) originated from the closer position of Trp5 in the case of CAIX. It interacted with the inhibitor via several dihydrogen bonds with the shortest one with the distance of 2.3 Å. Another favourable differences were caused by Asn62 and

Pro202 residues. These contributions were however cancelled out by differences in binding of Phe/Val131 and Val/Leu135 which were lower in CAIX by 0.7 and 0.9 kcal/mol, respectively. The fact that all favourable energy changes in CAIX/**1a** were slightly larger than the unfavourable changes is in qualitative agreement with the experimental K_i values (380 ± 111 nM for CAIX/**1a** and 100 ± 141 nM for CAII).¹⁹⁶

To conclude, we applied quantum chemistry to study the non-covalent interactions of carborane-based inhibitors in two isoenzymes of carbonic anhydrase. We showed the benefits of QM-based approach in the study involving metal ions and unusual compounds such as boranes and we used a quantum mechanics/molecular mechanics (QM/MM) methodology to quantitatively describe the binding and to explain fundamental differences in the binding modes of *closo*- and *nido*-cages. We also reported structural and computational analysis applicable to structure-based design of carborane compounds with selectivity towards the cancer-specific CAIX isoenzyme. We successfully introduced the very first model of 1-methylenesulfamide-1,2-dicarba-*closo*-dodecaborane interactions with CAIX.

3.3 The SQM/COSMO filter

In the previous section we have shown that if the crystal structure is not known, QM-based methods can be combined with slower approaches, such as ligand building or MD sampling, to obtain the complex geometry. In a virtual screening, the complex geometry is obtained by using various ultrafast classical scoring functions in docking of ligand structures. One of the most important tasks of docking and scoring is to predict a correct binding mode of known active compounds. This is a challenge where quite a large success has already been achieved. Docking algorithms are very efficient now and correct binding mode is among predicted poses in about 80%. However, to reliably identify a binding pose remains a difficult challenge.^{251,252} In this section we introduce a physics-based filter composed of the semiempirical quantum mechanical description of protein-ligand interaction and solvation that has been adapted for the virtual screening of compound libraries. The original paper is attached in Appendix H.

All methods in vHTS aim to predict the affinity of compounds to a receptor by computing the interaction energy of a predicted pose of the ligand within the binding pocket of the receptor. From a theorist's perspective, the judgment of the binding energy is a special challenge that needs to be tackled *i)* in order to predict a productive binding pose and *ii)* in order to compare the affinity of different ligands. Current methods to rank docking poses²⁵³ are:

- Empirically-derived energy functions describing the individual contributions of various physics-inspired terms, such as hydrogen bonds, ionic interactions, hydrophobic interactions, entropy terms. The parameters necessary are usually optimized to reproduce training sets.
- Knowledge-based potentials are generated by statistical evaluation of large data sets. These potentials do not model specific interaction types, they rather incorporate all interaction types that were present in the data set in the parameters and/or the functional form.
- Force-field derived objective functions adapt functional form of and parameters for electrostatics and van der Waals interactions from empirical force fields like Amber.

All these methods entail a high degree of empiricism and may be unable to go beyond the protein-ligand interactions present in the test set or described by the MM. The ultimate solution of these issues is the application of QM-based methods in the drug design^{24,254} which are able to describe unusual ligands²¹⁹ or non-covalent interactions of quantum origin, such as halogen bonding^{79,80} or covalent ligand-receptor binding.⁸¹ With the ever increasing power of computational infrastructure accompanied by recent developments in the QM methods, algorithms and setups (such as linear scaling or efficient parallelization of SQM^{61-63,76}, hybrid QM/MM^{27,52,54,55,61,62} or fragmentation techniques^{58,59,62,63,255}, the calculation of ligand-receptor complexes of thousands of atoms with methods of electronic structure theory has become feasible.

Recently, we have introduced an advanced SQM-based scoring function and applied it to describe the binding of inhibitors to several protein targets.⁷⁵⁻⁸¹ The scoring function

represents a thought decomposition of drug binding into distinct contributions that can be calculated (Eq. 3.2):

$$Score = \Delta E_{int} + \Delta \Delta G_{solv} + \Delta G'_{conf}{}^w - T\Delta S \quad (Eq. 3.2)$$

The terms are the gas-phase interaction energy (ΔE_{int}), the solvation/desolvation free energy change upon the binding ($\Delta \Delta G_{solv}$), the change of conformational free energy ($\Delta G'_{conf}{}^w$) upon the binding, and the change of entropy ($-T\Delta S$).

ΔE_{int} describes the undamped interaction between ligand and receptor in the gas phase that is governed by electrostatics, dispersion, polarization, and charge-transfer contributions. In addition, inhibitors frequently feature halogen substituents and consequently the sigma-hole bonding plays a role. Theoretical description of all these terms is tedious and force-field methods, frequently used in the realm of protein - ligand interactions, fail. The only alternative is thus represented by the quantum mechanical (QM) methods providing reliable descriptions of all energy terms mentioned. We standardly employ the SQM PM6-D3H4X method⁷⁰ that is designed to treat accurately van der Waals interactions, hydrogen bonding, and halogen bonding.^{76,256} This method can routinely be used for systems with up to 10,000 atoms. In contrast to fully empirical approaches like force fields, SQM methods are applicable and comparable throughout the chemical space.

The solvation/desolvation term represents the second most important term. While the gas-phase interaction energy is always attractive the solvation/desolvation term is systematically repulsive. As mentioned above the ligand is mostly highly polar or even charged and, consequently, it is strongly hydrated. Before entering the active side of protein it must be dehydrated and the respective (free) energy is very large, in absolute value even comparable to the gas-phase interaction energy. Calculation of ligand solvation free energy is difficult and is connected with much larger uncertainties than that of gas phase interaction energy. The SQM methods used for evaluation of interaction energy provide a reasonably small error below few kcal/mol while the continuum solvent based methods used mostly for evaluation of solvation energy provide much larger errors what is especially true for the charged systems. Nevertheless the solvation/desolvation term which is positive, is very important and cannot be neglected. It is important to mention that gas-phase interaction energies and solvation/desolvation free energies do not correlate simply because they are due to different physicochemical properties.

This first two terms are in absolute value comparable while other terms are smaller in magnitude. The accurate Score, where all terms are considered is computationally demanding. It can be used for accurate estimation of binding free energies for preselected (smaller) set of proteins and ligands where the structures are obtained from full gradient optimization. The most demanding part of the scoring is the SQM optimization of protein-ligand complexes. It must be done before scoring because (ΔE_{int}) is calculated on a structure optimized in water environment. The SQM optimization of protein-ligand complex last for several days and is the main limitation to use SQM based scoring in high throughput screening.

Here, we have simplified our SQM-based scoring function to make it usable in virtual screening on the basis of our previous experience. We defined the SQM/COSMO filter energy considering only first two terms of Eq.3.2 and compared its performance with several known scoring functions. Our novel scheme consists of a single-point rescoring of hydrogen-relaxed structures with no additional optimization of the systems. This approximate level can be successfully used for the fast ranking according to a physically meaningful estimate of binding free energies even for large sets of ligands and proteins, for example poses from a docking study. To this aim, we generated a large amount of sensible and non-redundant alternative ligand poses binding to four distinct proteins and checked how well the different scoring approaches were able to differentiate between the alternative and native states.

PDB	Resolution	Protein	Ligand	Features
1E66	2.10 Å	AChE	Huprine X	Two binding pockets, halogenated ligand
2IKJ	1.55 Å	AR	IDD393	Cofactor, halogenated ligand
3B92	2.00 Å	TACE	440	Metallo-protein, Zn ²⁺ cation coordinated by S ⁻ , three water molecules in binding site
1NH0	1.03 Å	HIV PR	KI2	Large, flexible and charged ligand, structural water molecule in binding site

Table 3.2: Four protein-ligand complexes with their characteristics.

Four unrelated difficult-to-handle protein ligand complexes that represent rather classical drug targets and that were resolved as high-resolution X-ray structures have been chosen for this study (Table 3.2), *i.e.* acetylcholine esterase (AChE, PDB: 1E66)²⁵⁷, TNF- α

converting enzyme (TACE, PDB: 3B92)²⁵⁸, aldose reductase (AR, PDB: 2IKJ)²⁵⁹ and HIV-1 protease (HIV PR; PDB: 1NH0)²⁰⁹.

For binding pose generation we used four different docking programs with overall 7 different scoring functions, *i.e.* empirical GlideScore XP (GlideXP), PLANTS PLP (PLP), AutoDock Vina (Vina), Chemscore (CS), Goldscore (GS) and ChemPLP and knowledge-based Astex Statistical Potential (ASP).²⁶⁰⁻²⁶⁵ Docking runs were started with 10 randomized ligand conformations and the original conformation extracted from the X-ray structure. For each docking run, up to 100 receptor-ligand poses were generated by each of the 7 docking setups. This hypothetical maximal number of 7,700 decoys per receptor-ligand pair was reduced by clustering with a cut-off of 0.5 Å in order to avoid redundant conformations. This yielded up to 2,865 ligand poses in total. The docking results are shown in Figure 1 in Appendix H.

All the poses were re-scored by nine scoring functions, 7 above mentioned functions plus two physics-based scoring functions: SQM/COSMO filter and AMBER/GB scoring function combining the ff03 and GAFF force fields with GB implicit solvent.^{71,204,266} For the latter two, hydrogens and close contacts were relaxed by the AMBER/GB method following our standard procedure of the structure preparation of protein-ligand complexes, where partial charges of ligands were derived from RESP fitting of the electrostatic potential calculated at the AM1-BCC level.^{267,268} To speed up the calculations, we defined a sphere of 8 to 12 Å (roughly 2,000 atoms) around aligned ligand poses as a region representing the binding site. This region was treated by SQM PM6D3H4X/COSMO method⁷⁰ and was the same for all the poses. These truncated systems (SQM/COSMO filter) were compared with full-sized systems (full SQM/COSMO) and shown that they behaved nearly identically. All scores coming from 9 scoring functions were normalized, using first and third quartiles. Normalized scores of each pose were thus comparable and their relative energy values were plotted against the RMSD of the crystal structure. The resulting clouds of points (shown in the Supplementary material of Appendix H) were further simplified to a single graph by showing only the lower boundary of all energies with respect to RMSD from the X-ray structure (Figure 3.21)

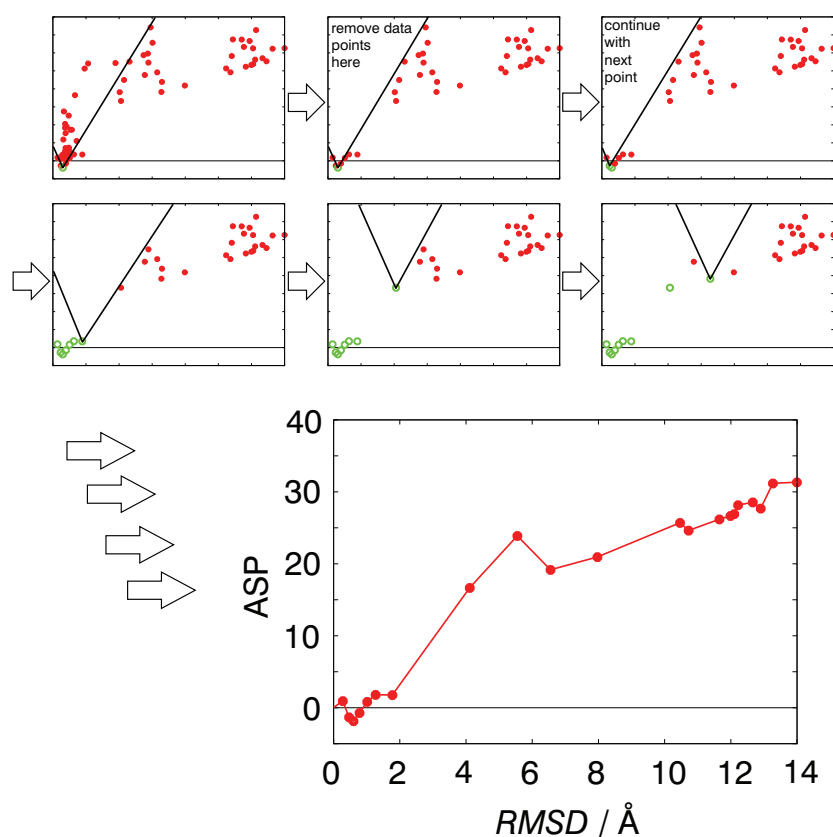


Fig. 3.21: Scheme of the algorithm to create the lower boundary from the whole data set. An iterative process reduces the large amount of data points to the most important points.

Based on the hypothesis of the native state being the lowest point in the accurate energy landscape of a given system, moving away from this state (internal motion, rotation, or translation of the ligand) should result in an increase of the free energy or the score. An ideal behavior of the graphs in Figure 3.22 would then be characterized by an increase of relative score with increasing RMSD to the x-ray structure, although local minima are possible. Small deviations from that behavior (lower energies for really small RMSD values) should be acceptable and might be explained by uncertainty of the crystal structure. The results revealed that the SQM/COSMO filter behaviour met this condition at superior level unlike the others scoring functions.

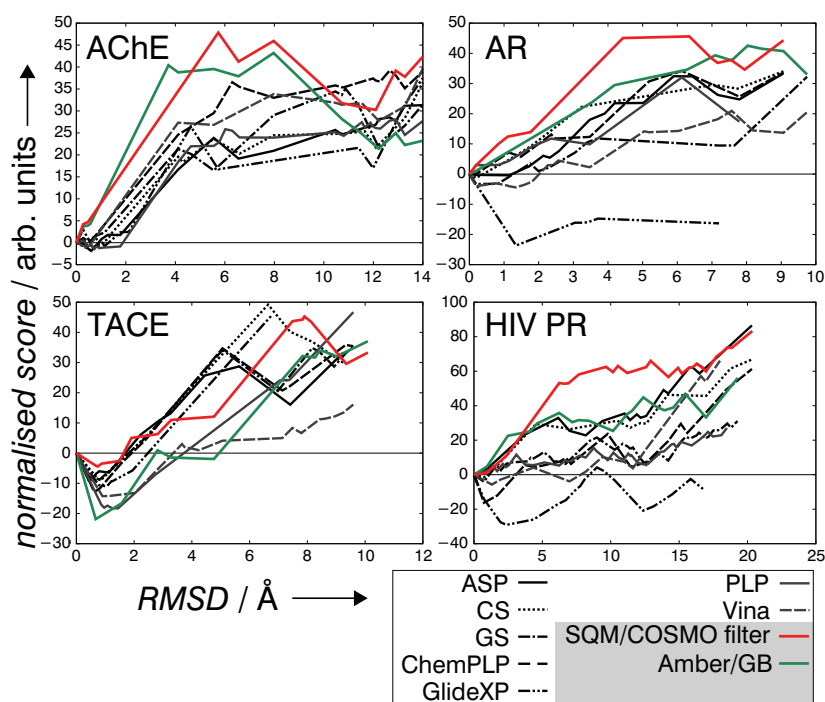


Fig. 3.22: The plots of normalized scores against RMSD values for all four protein-ligand systems.

In order to quantify the performance of studied scoring functions, we counted the false-positive solutions (Table 3.3). False positives are poses that deviated from the crystal position, however they were ranked lower in energy. The lowest number of false positives was found for SQM filter with 39 false positives in total, even zero for three protein-ligand complexes. It was followed by slightly worse performance of Gold CS with 54 false positives and ASP (59 false positives). Surprisingly high number of false positives was found also for AMBER/GB scoring function, behaving satisfyingly well for three systems, but yielded 171 false positives for TACE. All scoring functions performed satisfyingly well for AChE. In the case of AR and HIV PR, GlideScore XP generated the biggest number of false positive solutions, even shape-wise the energy landscape seemed ill-defined. The hardest case was the TACE metalloprotein. Here, all the scoring functions produced false-positive solutions but to a different extent. The SQM/COSMO filter performed best, followed by CS. The presence of the metal cation and the associated charge transfer effects between ligand and cation have shown the strength of an electronic-structure theory description of protein-ligand binding.

	Scoring function									
	SQM/ COSMO O	AMBER /GB	Glide				AutoDock			Gold
			XP	PLP	Vina	ASP	CS	GS	Chem PLP	
AChE	0	0	4	12	0	2	3	0	0	
AR	0	1	67	0	10	1	0	1	0	
TACE	39	171	181	294	63	56	49	78	111	
HIV PR	0	0	98	0	7	0	2	1	8	
Total	39	172	350	306	80	59	54	80	119	

Table 3.3: The numbers of false-positive solutions revealed from re-scoring of four protein-ligand systems

In order to quantify the behavior of individual scoring functions in more detail, we introduced the second criterion, i.e. RMSD^{max} referring about the maximum value of RMSD from the crystal structure revealed within all poses ranked in a defined interval of the normalized scores (see Table 3.4). First, we compared RMSD^{max} in the interval of normalized score up to 5. The SQM/COSMO filter behaved the best in this interval, showing RMSD^{max} of 0.88 Å on average. CS followed with the RMSD^{max} of 1.28 Å on average. ASP and AMBER/GB satisfied this condition of an averaged RMSD^{max} up to 2Å. AMBER/GB, however, failed in case of TACE with the RMSD^{max} of 4.76 Å.

	Scoring function								
	SQM/ COSMO	AMBER /GB	Glide		PLANTS		AutoDock		Gold
			XP	PLP	Vina	ASP	CS	GS	
Maximal RMSD within a window of 5 of the normalized Score									
AChE	0.47	0.57	2.13	0.78	0.78	1.78	1.43	1.14	0.78
AR	0.19	0.19	7.54	1.14	3.54	2.32	1.15	2.21	1.49
TACE	1.91	4.76	3.02	2.91	7.13	2.01	1.54	2.44	2.40
HIV PR	0.94	0.94	17.26	12.60	11.62	1.00	1.01	12.60	11.62
Average	0.88	1.62	7.49	4.61	5.77	1.78	1.28	4.60	4.55

Table 3.4: The maximum RMSD [Å] within all the poses in the defined range of the relative normalised score

When the interval of normalized score was increased up to 10, the lowest value RMSD^{max} about 1.32 Å was obtained again by SQM/COSMO filter, followed by AMBER/GB and CS (1.73 and 1.84 Å, respectively). Also ASP satisfied the condition of RMSD^{max} up to 2 Å. The other scoring functions were considerably worse (over 4 Å for both intervals). Behaviour of the scoring functions in bigger intervals is shown in Table S4 of Supplementary material of Appendix H.

To conclude, we have introduced very effective SQM-based tool for reliable ranking of binding poses from docking results. We have shown that the SQM/COSMO filter is able to recognize the correct binding pose (RMSD^{max} below 2 Å) and moreover to go beyond this limit and evaluate even small changes in the geometry of the ligand binding. We have successfully shown its superior performance among 8 widely used scoring functions on 4 unrelated protein-ligand systems. In contrast to standard scoring functions, no fitting against data sets has been involved in the SQM/COSMO filter. Furthermore, it offers generality and comparability across the chemical space and no system-specific parameterizations have to be performed. We have thus pushed the limits of the accuracy of scoring functions to estimate the energetics of protein-ligand complexes. Moreover the time requirements allow for calculations of thousands of docking poses. Therefore, we propose the SQM/COSMO filter as a tool for accurate medium-throughput refinement in later stages of virtual screening or as a reference method to judge the performance of other scoring functions.

Chapter 4

Summary and final remarks

The thesis aims to show the ability of QM-based approaches to contribute into the field of computer-aided drug design. The core of the thesis consists of 8 original papers that have been divided into three topics. The accompanying text introduces the reader into main aspects of drug design with a special focus on the methodology being used by the scientific community. The methods used in our studies are properly described in Chapter 2.

During my work I have focused on several different topics whose outcomes can be utilized in the rational drug design or even in the high throughput virtual screening of compound libraries. The first part of the thesis examines the origin of the σ -hole bonding using high level QM methods, such as CCSD(T)/CBS calculations, DFT-D interaction energy estimates and DFT-SAPT decompositions. The second part is devoted to three different protein-ligand systems, all representing pharmaceutically interesting targets. The last part focuses on the development and application of the effective SQM-based tool for a reliable ranking of docking poses in the process of virtual screening. We tested this SQM/COSMO filter on 4 different protein-ligand systems together with 8 standardly used scoring functions.

The first part of Chapter 3 has introduced a specific non-covalent interaction, so called σ -hole bonding. To shed light on the nature of bonding, we applied accurate quantum mechanical methods on the halogen-, chalcogen- and pnictogen bonded structures, *i.e.* on extended datasets of halogen bonded complexes, inorganic crystal structures of thaboranes stabilized by chalcogen bonds and a complete dataset of heteroboranes interacting with their organic partners by all three types of σ -hole bonding. We have shown that the only way how to elucidate the complete picture of σ -hole bonding is to relate the properties of monomers, *i.e.* σ -holes, with the properties of complexes. The importance of high level quantum mechanical methods was highlighted by the fact that strength of σ -hole bonding in isolated complexes is proportional to the magnitude of the σ -hole on

the atom may not be so straightforward. The DFT-SAPT decomposition of stabilization energies has revealed the concert action of polarization and dispersion energy to the stabilization of halogen bonding. We can thus conclude that the positive σ -hole and the negative electron donor interact by the electrostatic energy, which is responsible not only to the stability but also for the high directionality of the bond while the dispersion energy is responsible for its high stability. The contact atom pair (the halogen and electron donor) contributes by as much as 40% of the total dispersion energy and so plays a dominant role in bonding. The recent IUPAC definition of the halogen bond states that ‘the forces involved in the formation of the halogen bond are primarily electrostatic, but polarization, charge-transfer and dispersion contributions all play an important role.’ We have shown that such definition may not describe the unique phenomenon of the halogen bonding sufficiently enough.

In the study of the inorganic crystal of thiaborane we have shown the existence of five highly positive σ -holes on the positively charged pentacoordinated sulphur atom and consequently the ability of this structure to form B-S... π chalcogen bonds. These σ -hole bonds are considerably stronger than these in their organic counterparts. In order to gain a deeper insight into the nature of these σ -hole interactions, we have applied a detailed QM study to the majority of experimentally known *closo*-heteroboranes, where chalcogens and pnicogens are incorporated in the borane cage, together with *exo*-substituted halogens. As opposed to the classical electronegativity concept, we have shown that all these heteroatoms are centers of positive charges and so form very strong σ -hole bonds. DFT-SAPT decompositions of their total stabilization energies have revealed that chalcogen and pnicogen bonds come from dominating dispersion and electrostatic energy, followed by induction showing not negligible role of the charge transfer. We have also shown and quantified several ways of the modulation of σ -hole bonding which can be utilized in its applying in crystal engineering and drug design. The shown ability of heteroboranes to form all types of σ -hole bonds can be utilized in the design of heteroborane-based protein ligands, such as enzyme inhibitors or receptor antagonists/agonists.

Drug design efforts benefit greatly from knowledge of 3D structures of protein-ligand complexes. X-ray crystallography offers unprecedented insight into inhibitor binding modes and thus contributes considerably to the drug development. In the second part of Chapter 3 we have shown how structural information coupled with QM-based calculations can be effectively used for detail studies of three protein-ligand systems. All studied

proteins, *i.e.* HIV-1 protease, secreted aspartic protease and carbonic anhydrase, represent potential targets in drug design. We used a quantum mechanics/molecular mechanics (QM/MM) methodology to quantitatively describe the protein-ligand binding, to unveil features of the structure that are not accessible to the crystallographic experiment and to explain fundamental differences in the binding modes of inhibitors.

We have shown benefits of the QM-based approach in protein-ligand complexes involving proton-transfer phenomena, metal ions and unusual compounds such as boranes. We have used a this methodology to quantitatively describe the ligand binding and to explain fundamental differences in the binding modes such as in the case of *closo*- and *nido*-cages of carborane-based inhibitors of CAII. We have introduced the virtual glycine scan procedure that dissects the energy contributions to the total interaction energy of the sidechains of all important aminoacids in the particular active site. We have successfully determined the most probable protonation states of HIV-1 protease and secreted aspartic protease. We have identified the most stable isomers (conformers/tautomers) and rotamers of the studied ligands, *i.e.* the nonpeptidic inhibitor darunavir and phenylnorstatine-based peptidomimetic inhibitor KI2 of HIV-1 protease, pepstatin A in secreted aspartic protease and *closo*- and *nido*- carborane-based inhibitors of carbonic anhydrase II. These findings are very important for building the reliable computational model of the studied systems for further affinity estimates. All provided results are useful for an understanding and selectivity of the ligand binding to the particular protein target as well as for a further rational design of more potent/selective inhibitors.

The identification of productive binding poses between protein and ligand and the prediction of affinities by *in silico* experiments are key to the success in the field of drug design. Empirical scoring functions, the current standard in the field, have been failing due to their predictive power being limited by their parameterization or training. Moreover, quantum effects which undoubtedly play a key role in these processes are completely neglected. On contrary, advanced methods, based on the first principles of quantum mechanics, have up to recently been hampered by their substantial computational cost due to which various approximations had to be adopted. We have developed a physics-based filter composed of the semiempirical quantum mechanical description of protein-ligand interaction and solvation which shows a superior performance as compared to 8 standardly used methods, *i.e.* statistics-, knowledge- and force-field-based scoring functions. The last part of Chapter 3 has introduced the SQM/COSMO filter featuring two

dominant terms to describe protein-ligand interaction, namely the ΔE_{int} term at the PM6-D3H4X level for gas-phase non-covalent interactions and the $\Delta\Delta G_{solv}$ term at the COSMO level for implicit solvation. We have applied the SQM/COSMO filter on four unrelated protein-ligand systems. We have demonstrated its ability to recognize the correct binding pose and moreover to go beyond this limit and evaluate even small changes in the geometry of the ligand binding. Because of its advantages, *i.e.* generality, comparability across the chemical space, no need of any system-specific parameters, the SQM/COSMO filter has just pushed the limits of the accuracy of scoring functions to estimate the energetics of protein-ligand complexes. Together with its time requirements allowing calculations of thousands of docking poses, we propose the SQM/COSMO filter as a tool for accurate medium-throughput refinement in later stages of virtual screening or as a reference method to judge the performance of other scoring functions.

Bibliography

- (1) Adams, C. P.; Brantner, V. V. *Health Econ* **2010**, *19*, 130.
- (2) Kapetanovic, I. M. *Chem-Biol Interact* **2008**, *171*, 165.
- (3) Newman, D. J.; Cragg, G. M. *J Nat Prod* **2007**, *70*, 461.
- (4) Dror, O.; Shulman-Peleg, A.; Nussinov, R.; Wolfson, H. J. *Curr Med Chem* **2004**, *11*, 71.
- (5) Klebe, G. *J Mol Med-Jmm* **2000**, *78*, 269.
- (6) Jeremy M. Berg, J. L. T., Lubert Stryer *Biochemistry*; 5th edition; W. H. Freeman: New York, 2002.
- (7) Koshland, D. E. *P Natl Acad Sci USA* **1958**, *44*, 98.
- (8) Parsegian, V. A. *Van der Waals Forces: A Handbook for Biologists, Chemists, Engineers, and Physicists*; Cambridge University Press, 2005.
- (9) Stone, A. J. *The Theory of Intermolecular Forces*; Clarendon Press, 1997.
- (10) Hobza, P.; Müller-Dethlefs, K. *Non-covalent Interactions: Theory and Experiment*; Royal Society of Chemistry, 2010.
- (11) Cheng, Y.; Prusoff, W. H. *Biochem Pharmacol* **1973**, *22*, 3099.
- (12) Leavitt, S.; Freire, E. *Curr Opin Struc Biol* **2001**, *11*, 560.
- (13) Danielson, U. H. *Future Med Chem* **2009**, *1*, 1399.
- (14) Gohlke, H.; Klebe, G. *Curr Opin Struc Biol* **2001**, *11*, 231.
- (15) Karplus, M. *Annu Rev Bioph Biom* **2006**, *35*, 1.
- (16) Aqvist, J.; Medina, C.; Samuelsson, J. E. *Protein Eng* **1994**, *7*, 385.
- (17) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S. H.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. *Accounts Chem Res* **2000**, *33*, 889.
- (18) Shirts, M. R.; Pitner, J. W.; Swope, W. C.; Pande, V. S. *J Chem Phys* **2003**, *119*, 5740.
- (19) Genheden, S.; Ryde, U. *J Comput Chem* **2010**, *31*, 837.
- (20) Gohlke, H.; Case, D. A. *J Comput Chem* **2004**, *25*, 238.
- (21) Kamerlin, S. C. L.; Vicatos, S.; Dryga, A.; Warshel, A. *Annu Rev Phys Chem* **2011**, *62*, 41.
- (22) Laio, A.; Parrinello, M. *P Natl Acad Sci USA* **2002**, *99*, 12562.
- (23) Hamelberg, D.; Mongan, J.; McCammon, J. A. *J Chem Phys* **2004**, *120*, 11919.

- (24) Raha, K.; Peters, M. B.; Wang, B.; Yu, N.; WollaCott, A. M.; Westerhoff, L. M.; Merz, K. M. *Drug Discov Today* **2007**, *12*, 725.
- (25) Zhou, T.; Huang, D. Z.; Caflisch, A. *Curr Top Med Chem* **2010**, *10*, 33.
- (26) Lodola, A.; De Vivo, M. *Adv Protein Chem Str* **2012**, *87*, 337.
- (27) Mucs, D.; Bryce, R. A. *Expert Opin Drug Dis* **2013**, *8*, 263.
- (28) Merz, K. M. *Accounts Chem Res* **2014**, *47*, 2804.
- (29) Ponder, J. W.; Wu, C. J.; Ren, P. Y.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. *J Phys Chem B* **2010**, *114*, 2549.
- (30) Piquemal, J. P.; Chevreau, H.; Gresh, N. *J Chem Theory Comput* **2007**, *3*, 824.
- (31) Donchev, A. G.; Galkin, N. G.; Tarasov, V. I. *J Chem Phys* **2007**, *126*.
- (32) Popelier, P. L. A.; Smith, P. J. *Eur J Med Chem* **2006**, *41*, 862.
- (33) Chalk, A. J.; Beck, B.; Clark, T. *J Chem Inf Comp Sci* **2001**, *41*, 457.
- (34) Gogonea, V.; Suarez, D.; van der Vaart, A.; Merz, K. W. *Curr Opin Struc Biol* **2001**, *11*, 217.
- (35) Roy, R. K.; Krishnamurti, S.; Geerlings, P.; Pal, S. *J Phys Chem A* **1998**, *102*, 3746.
- (36) Torrent-Sucarrat, M.; De Proft, F.; Ayers, P. W.; Geerlings, P. *Phys Chem Chem Phys* **2010**, *12*, 1072.
- (37) Ryde, U.; Nilsson, K. *J Am Chem Soc* **2003**, *125*, 14232.
- (38) Nilsson, K.; Hersleth, H. P.; Rod, T. H.; Andersson, K. K.; Ryde, U. *Biophys J* **2004**, *87*, 3437.
- (39) Rajamani, R.; Reynolds, C. H. *J Med Chem* **2004**, *47*, 5159.
- (40) Piana, S.; Bucher, D.; Carloni, P.; Rothlisberger, U. *J Physic Chem B* **2004**, *108*, 11139.
- (41) Yu, N.; Hayik, S. A.; Wang, B.; Liao, N.; Reynolds, C. H.; Merz, K. M. *J Chem Theory Comput* **2006**, *2*, 1057.
- (42) He, X.; Wang, B.; Merz, K. M. *J Phys Chem B* **2009**, *113*, 10380.
- (43) Seebeck, B.; Reulecke, I.; Kamper, A.; Rarey, M. *Proteins* **2008**, *71*, 1237.
- (44) Kirton, S. B.; Murray, C. W.; Verdonk, M. L.; Taylor, R. D. *Proteins* **2005**, *58*, 836.
- (45) Wang, J. C.; Lin, J. H.; Chen, C. M.; Perryman, A. L.; Olson, A. J. *J Chem Inf Model* **2011**, *51*, 2528.
- (46) Fischer, B.; Fukuzawa, K.; Wenzel, W. *Proteins* **2008**, *70*, 1264.

- (47) Cho, A. E.; Guallar, V.; Berne, B. J.; Friesner, R. *J Comput Chem* **2005**, *26*, 915.
- (48) Chung, J. Y.; Hah, J. M.; Cho, A. E. *J Chem Inf Model* **2009**, *49*, 2382.
- (49) Illingworth, C. J. R.; Morris, G. M.; Parkes, K. E. B.; Snell, C. R.; Reynolds, C. A. *J Phys Chem A* **2008**, *112*, 12157.
- (50) Raha, K.; Merz, K. M. *J Med Chem* **2005**, *48*, 4558.
- (51) Fong, P.; McNamara, J. P.; Hillier, I. H.; Bryce, R. A. *J Chem Inf Model* **2009**, *49*, 913.
- (52) Chaskar, P.; Zoete, V.; Rohrig, U. F. *J Chem Inf Model* **2014**, *54*, 3137.
- (53) Senn, H. M.; Thiel, W. *Angew Chem Int Edit* **2009**, *48*, 1198.
- (54) Wichapong, K.; Rohe, A.; Platzer, C.; Slynko, I.; Erdmann, F.; Schmidt, M.; Sippl, W. *J Chem Inf Model* **2014**, *54*, 881.
- (55) Burger, S. K.; Thompson, D. C.; Ayers, P. W. *J Chem Inf Model* **2011**, *51*, 93.
- (56) Soderhjelm, P.; Aquilante, F.; Ryde, U. *J Phys Chem B* **2009**, *113*, 11085.
- (57) Hu, L. H.; Eliasson, J.; Heimdal, J.; Ryde, U. *J Phys Chem A* **2009**, *113*, 11793.
- (58) Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. *Chem Rev* **2012**, *112*, 632.
- (59) Antony, J.; Grimme, S.; Liakos, D. G.; Neese, F. *J Phys Chem A* **2011**, *115*, 11210.
- (60) Dixon, S. L.; Merz, K. M. *J Chem Phys* **1996**, *104*, 6643.
- (61) Hayik, S. A.; Dunbrack, R.; Merz, K. M. *J Chem Theory Comput* **2010**, *6*, 3079.
- (62) Soderhjelm, P.; Kongsted, J.; Ryde, U. *J Chem Theory Comput* **2010**, *6*, 1726.
- (63) Hennemann, M.; Clark, T. *J Mol Model* **2014**, *21*, 144.
- (64) Raha, K.; Merz, K. M. *J Am Chem Soc* **2004**, *126*, 1020.
- (65) Muddana, H. S.; Gilson, M. K. *J Chem Theory Comput* **2012**, *8*, 2023.
- (66) Mikulskis, P.; Genheden, S.; Wichmann, K.; Ryde, U. *J Comput Chem* **2012**, *33*, 1179.
- (67) Stewart, J. J. P. *J Mol Model* **2007**, *13*, 1173.
- (68) Rezac, J.; Fanfrlik, J.; Salahub, D.; Hobza, P. *J Chem Theory Comput* **2009**, *5*, 1749.
- (69) Rezac, J.; Hobza, P. *Chem Phys Lett* **2011**, *506*, 286.
- (70) Rezac, J.; Hobza, P. *J Chem Theory Comput* **2012**, *8*, 141.
- (71) Tsui, V.; Case, D. A. *Biopolymers* **2001**, *56*, 275.
- (72) Klamt, A.; Schuurmann, G. *J Chem Soc-Perkin Transactions 2* **1993**, 799.
- (73) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J Phys Chem B* **2009**, *113*, 6378.

- (74) Kolar, M.; Fanfrlik, J.; Lepsik, M.; Forti, F.; Luque, F. J.; Hobza, P. *J Phys Chem B* **2013**, *117*, 5950.
- (75) Fanfrlik, J.; Bronowska, A. K.; Rezac, J.; Prenosil, O.; Konvalinka, J.; Hobza, P. *J Phys Chem B* **2010**, *114*, 12666.
- (76) Lepsik, M.; Rezac, J.; Kolar, M.; Pecina, A.; Hobza, P.; Fanfrlik, J. *Chempluschem* **2013**, *78*, 921.
- (77) Dobes, P.; Rezac, J.; Fanfrlik, J.; Otyepka, M.; Hobza, P. *J Phys Chem B* **2011**, *115*, 8581.
- (78) Dobes, P.; Fanfrlik, J.; Rezac, J.; Otyepka, M.; Hobza, P. *J Comput Aid Mol Des* **2011**, *25*, 223.
- (79) Fanfrlik, J.; Kolar, M.; Kamlar, M.; Hurny, D.; Ruiz, F. X.; Cousido-Siah, A.; Mitschler, A.; Rezac, J.; Munusamy, E.; Lepsik, M.; Matejicek, P.; Vesely, J.; Podjarny, A.; Hobza, P. *Acs Chem Biol* **2013**, *8*, 2484.
- (80) Fanfrlik, J.; Ruiz, F. X.; Kadlcikova, A.; Rezac, J.; Cousido-Siah, A.; Mitschler, A.; Haldar, S.; Lepsik, M.; Kolar, M. H.; Majer, P.; Podjarny, A. D.; Hobza, P. *Acs Chem Biol* **2015**, *10*, 1637.
- (81) Fanfrlik, J.; Brahmshatriya, P. S.; Rezac, J.; Jilkova, A.; Horn, M.; Mares, M.; Hobza, P.; Lepsik, M. *J Phys Chem B* **2013**, *117*, 14973.
- (82) Boys, S. F.; Bernardi, F. *Mol Phys* **1970**, *19*, 553.
- (83) Mayer, I. *Int J Quantum Chem* **1983**, *23*, 341.
- (84) Halkier, A.; Helgaker, T.; Jorgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. *Chem Phys Lett* **1998**, *286*, 243.
- (85) Halkier, A.; Helgaker, T.; Jorgensen, P.; Klopper, W.; Olsen, J. *Chem Phys Lett* **1999**, *302*, 437.
- (86) Truhlar, D. G. *Chem Phys Lett* **1998**, *294*, 45.
- (87) Shin, I.; Park, M.; Min, S. K.; Lee, E. C.; Suh, S. B.; Kim, K. S. *J Chem Phys* **2006**, *125*.
- (88) Lee, E. C.; Kim, D.; Jurecka, P.; Tarakeshwar, P.; Hobza, P.; Kim, K. S. *J Phys Chem A* **2007**, *111*, 3446.
- (89) Zhao, Y.; Truhlar, D. G. *J Chem Theory Comput* **2007**, *3*, 289.
- (90) Goerigk, L.; Grimme, S. *J Chem Theory Comput* **2010**, *6*, 107.
- (91) Rezac, J.; Jurecka, P.; Riley, K. E.; Cerny, J.; Valdes, H.; Pluhackova, K.; Berka, K.; Rezac, T.; Pitonak, M.; Vondrasek, J.; Hobza, P. *Collect Czech Chem C* **2008**, *73*, 1261.

- (92) Jurecka, P.; Spomer, J.; Cerny, J.; Hobza, P. *Phys Chem Chem Phys* **2006**, *8*, 1985.
- (93) Rezac, J.; Riley, K. E.; Hobza, P. *J Chem Theory Comput* **2011**, *7*, 3466.
- (94) Rezac, J.; Riley, K. E.; Hobza, P. *J Chem Theory Comput* **2011**, *7*, 2427.
- (95) Sedlak, R.; Janowski, T.; Pitonak, M.; Rezac, J.; Pulay, P.; Hobza, P. *J Chem Theory Comput* **2013**, *9*, 3364.
- (96) Cizek, J. *J Chem Phys* **1966**, *45*, 4256.
- (97) Paldus, J.; Shavitt, I.; Cizek, J. *Phys Rev A* **1972**, *5*, 50.
- (98) Bartlett, R. J. *Annual Review of Physical Chemistry* **1981**, *32*, 359.
- (99) Jurecka, P.; Hobza, P. *Chem Phys Lett* **2002**, *365*, 89.
- (100) Spomer, J.; Jurecka, P.; Hobza, P. *J Am Chem Soc* **2004**, *126*, 10142.
- (101) Pitonak, M.; Riley, K. E.; Neogrady, P.; Hobza, P. *Chem Phys Chem* **2008**, *9*, 1636.
- (102) Pitonak, M.; Janowski, T.; Neogrady, P.; Pulay, P.; Hobza, P. *J Chem Theory Comput* **2009**, *5*, 1761.
- (103) Janowski, T.; Ford, A. R.; Pulay, P. *Mol Phys* **2010**, *108*, 249.
- (104) Pitonak, M.; Neogrady, P.; Hobza, P. *Phys Chem Chem Phys* **2010**, *12*, 1369.
- (105) Kohn, W.; Sham, L. J. *Phys Rev* **1965**, *140*, 1133.
- (106) Grimme, S.; Antony, J.; Schwabe, T.; Muck-Lichtenfeld, C. *Org Biomol Chem* **2007**, *5*, 741.
- (107) Sato, T.; Nakai, H. *J Chem Phys* **2009**, *131*, 224104.
- (108) Grafenstein, J.; Cremer, D. *J Chem Phys* **2009**, *130*, 124105.
- (109) Grimme, S. *J Comput Chem* **2004**, *25*, 1463.
- (110) Grimme, S. *J Comput Chem* **2006**, *27*, 1787.
- (111) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J Chem Phys* **2010**, *132*, 154104.
- (112) Chai, J. D.; Head-Gordon, M. *Phys Chem Chem Phys* **2008**, *10*, 6615.
- (113) Jurecka, P.; Cerny, J.; Hobza, P.; Salahub, D. R. *J Comput Chem* **2007**, *28*, 555.
- (114) Goerigk, L.; Kruse, H.; Grimme, S. *Chem Phys Chem* **2011**, *12*, 3421.
- (115) McNamara, J. P.; Hillier, I. H. *Phys Chem Chem Phys* **2007**, *9*, 2362.
- (116) Tuttle, T.; Thiel, W. *Phys Chem Chem Phys* **2008**, *10*, 2159.
- (117) Foster, M. E.; Sohlberg, K. *J Chem Theory Comput* **2010**, *6*, 2153.
- (118) Bernal-Uruchurtu, M. I.; Ruiz-Lopez, M. F. *Chem Phys Lett* **2000**, *330*, 118.
- (119) Repasky, M. P.; Chandrasekhar, J.; Jorgensen, W. L. *J Comput Chem* **2002**, *23*, 1601.

- (120) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys Rev B* **1998**, *58*, 7260.
- (121) Korth, M.; Pitonak, M.; Rezac, J.; Hobza, P. *J Chem Theory Comput* **2010**, *6*, 344.
- (122) Korth, M. *J Chem Theory Comput* **2010**, *6*, 3808.
- (123) Hostas, J.; Rezac, J.; Hobza, P. *Chem Phys Lett* **2013**, *568*, 161.
- (124) Stewart, J. J. P. *J Mol Model* **2009**, *15*, 765.
- (125) Schrödinger, E. *Annalen der Physik* **1926**, *385*, 437.
- (126) London, F. *Transactions of the Faraday Society* **1937**, *33*, 8b.
- (127) Hirschfelder, J.O.; Byron Bird C. F. C., R.; *Molecular Theory of Gases and Liquids*; John Wiley & Sons: New York, 1954.
- (128) Ahlrichs, R. *Theor Chim Acta* **1976**, *41*, 7.
- (129) Morgan, J. D.; Simon, B. *Int J Quantum Chem* **1980**, *17*, 1143.
- (130) Murrell, J. N.; Shaw, G. *J Chem Phys* **1967**, *46*, 1768.
- (131) Jeziorski, B.; Szalewicz, K.; Chalasinski, G. *Int J Quantum Chem* **1978**, *14*, 271.
- (132) Hirschfeld, J.; Silbey, R. *J Chem Phys* **1966**, *45*, 2188.
- (133) Jansen, G.; Hesselmann, A. *J Phys Chem A* **2001**, *105*, 11156.
- (134) Williams, H. L.; Chabalowski, C. F. *J Phys Chem A* **2001**, *105*, 646.
- (135) Hesselmann, A.; Jansen, G. *Chem Phys Lett* **2002**, *362*, 319.
- (136) Hesselmann, A.; Jansen, G. *Chem Phys Lett* **2002**, *357*, 464.
- (137) Hesselmann, A.; Jansen, G. *Chem Phys Lett* **2003**, *367*, 778.
- (138) Hesselmann, A.; Jansen, G.; Schutz, M. *J Chem Phys* **2005**, *122*, 14103.
- (139) Hesselmann, A.; Jansen, G.; Schutz, M. *J Am Chem Soc* **2006**, *128*, 11730.
- (140) Podeszwa, R.; Bukowski, R.; Szalewicz, K. *J Phys Chem A* **2006**, *110*, 10345.
- (141) Gruning, M.; Gritsenko, O. V.; van Gisbergen, S. J. A.; Baerends, E. J. *J Chem Phys* **2001**, *114*, 652.
- (142) Hesselmann, A. *J Phys Chem A* **2011**, *115*, 11321.
- (143) Warshel, A.; Levitt, M. *J Mol Biol* **1976**, *103*, 227.
- (144) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J Phys Chem* **1996**, *100*, 19357.
- (145) Cramer, C. J.; Truhlar, D. G. *Science* **1992**, *256*, 213.
- (146) Hassel, O.; Hvoslef, J. *Acta Chem Scand* **1954**, *8*, 873.
- (147) Bent, H. A. *Chem Rev* **1968**, *68*, 587.
- (148) Clark, T.; Hennemann, M.; Murray, J. S.; Politzer, P. *J Mol Model* **2007**, *13*, 291.

- (149) Politzer, P.; Lane, P.; Concha, M. C.; Ma, Y. G.; Murray, J. S. *J Mol Model* **2007**, *13*, 305.
- (150) Lommerse, J. P. M.; Stone, A. J.; Taylor, R.; Allen, F. H. *J Am Chem Soc* **1996**, *118*, 3108.
- (151) Kolar, M. H.; Carloni, P.; Hobza, P. *Phys Chem Chem Phys* **2014**, *16*, 19111.
- (152) Bader, R. F. W.; Carroll, M. T.; Cheeseman, J. R.; Chang, C. *J Am Chem Soc* **1987**, *109*, 7968.
- (153) Kolar, M.; Hostas, J.; Hobza, P. *Phys Chem Chem Phys* **2014**, *16*, 9987.
- (154) Desiraju, G. R.; Ho, P. S.; Kloo, L.; Legon, A. C.; Marquardt, R.; Metrangolo, P.; Politzer, P.; Resnati, G.; Rissanen, K. *Pure Appl Chem* **2013**, *85*, 1711.
- (155) Riley, K. E.; Murray, J. S.; Fanfrlik, J.; Rezac, J.; Sola, R. J.; Concha, M. C.; Ramos, F. M.; Politzer, P. *J Mol Model* **2011**, *17*, 3309.
- (156) Rezac, J.; Riley, K. E.; Hobza, P. *J Chem Theory Comput* **2012**, *8*, 4285.
- (157) Kozuch, S.; Martin, J. M. L. *J Chem Theory Comput* **2013**, *9*, 1918.
- (158) Deepa, P.; Sedlak, R.; Hobza, P. *Phys Chem Chem Phys* **2014**, *16*, 6679.
- (159) Riley, K. E.; Hobza, P. *Phys Chem Chem Phys* **2013**, *15*, 17742.
- (160) Munusamy, E.; Sedlak, R.; Hobza, P. *Chem Phys Chem* **2011**, *12*, 3253.
- (161) Riley, K. E.; Murray, J. S.; Fanfrlik, J.; Rezac, J.; Sola, R. J.; Concha, M. C.; Ramos, F. M.; Politzer, P. *J Mol Model* **2013**, *19*, 4651.
- (162) Sedlak, R.; Deepa, P.; Hobza, P. *J Phys Chem A* **2014**, *118*, 3846.
- (163) Lu, Y. X.; Zou, J. W.; Fan, J. C.; Zhao, W. N.; Jiang, Y. J.; Yui, Q. S. *J Comput Chem* **2009**, *30*, 725.
- (164) Wang, W. Z. *J Phys Chem A* **2011**, *115*, 9294.
- (165) Chudzinski, M. G.; Taylor, M. S. *J Org Chem* **2012**, *77*, 3483.
- (166) Trnka, J.; Sedlak, R.; Kolar, M.; Hobza, P. *J Phys Chem A* **2013**, *117*, 4331.
- (167) Mukherjee, A.; Desiraju, G. R. *Cryst Growth Des* **2011**, *11*, 3735.
- (168) Cincic, D.; Friscic, T.; Jones, W. *J Am Chem Soc* **2008**, *130*, 7524.
- (169) Cauliez, P.; Polo, V.; Roisnel, T.; Llusar, R.; Fourmigue, M. *Crysteng comm* **2010**, *12*, 558.
- (170) Cincic, D.; Friscic, T.; Jones, W. *Crysteng comm* **2011**, *13*, 3224.
- (171) Walsh, R. B.; Padgett, C. W.; Metrangolo, P.; Resnati, G.; Hanks, T. W.; Pennington, W. T. *Cryst Growth Des* **2001**, *1*, 165.
- (172) Politzer, P.; Murray, J. S.; Concha, M. C. *J Mol Model* **2008**, *14*, 659.
- (173) Sanz, P.; Yanez, M.; Mo, O. *J Phys Chem A* **2002**, *106*, 4661.

- (174) Werz, D. B.; Gleiter, R.; Rominger, F. *J Am Chem Soc* **2002**, *124*, 10638.
- (175) Iwaoka, M.; Takemoto, S.; Tomoda, S. *J Am Chem Soc* **2002**, *124*, 10613.
- (176) Wang, W. Z.; Ji, B. M.; Zhang, Y. *J Phys Chem A* **2009**, *113*, 8132.
- (177) Machacek, J.; Plesek, J.; Holub, J.; Hnyk, D.; Vsetecka, V.; Cisarova, I.; Kaupp, M.; Stibr, B. *Dalton Transactions* **2006**, 1024.
- (178) Fanfrlik, J.; Prada, A.; Padelkova, Z.; Pecina, A.; Machacek, J.; Lepsik, M.; Holub, J.; Ruzicka, A.; Hnyk, D.; Hobza, P. *Angew Chem Int Edit* **2014**, *53*, 10139.
- (179) Hobza, P.; Sponer, J. *Chem Rev* **1999**, *99*, 3247.
- (180) Metrangolo, P.; Resnati, G. *Chem-Eur J* **2001**, *7*, 2511.
- (181) Nguyen, H. L.; Horton, P. N.; Hursthouse, M. B.; Legon, A. C.; Bruce, D. W. *J Am Chem Soc* **2004**, *126*, 16.
- (182) Fedorov, O.; Huber, K.; Eisenreich, A.; Filippakopoulos, P.; King, O.; Bullock, A. N.; Szklarczyk, D.; Jensen, L. J.; Fabbro, D.; Trappe, J.; Rauch, U.; Bracher, F.; Knapp, S. *Chem Biol* **2011**, *18*, 67.
- (183) Hardegger, L. A.; Kuhn, B.; Spinnler, B.; Anselm, L.; Ecabert, R.; Stihle, M.; Gsell, B.; Thoma, R.; Diez, J.; Benz, J.; Plancher, J. M.; Hartmann, G.; Isshiki, Y.; Morikami, K.; Shimma, N.; Haap, W.; Banner, D. W.; Diederich, F. *Chem Med Chem* **2011**, *6*, 2048.
- (184) Wilcken, R.; Zimmermann, M. O.; Lange, A.; Joerger, A. C.; Boeckler, F. M. *J Med Chem* **2013**, *56*, 1363.
- (185) Wilcken, R.; Liu, X. R.; Zimmermann, M. O.; Rutherford, T. J.; Fersht, A. R.; Joerger, A. C.; Boeckler, F. M. *J Am Chem Soc* **2012**, *134*, 6810.
- (186) Lu, Y. X.; Liu, Y. T.; Xu, Z. J.; Li, H. Y.; Liu, H. L.; Zhu, W. L. *Expert Opin Drug Dis* **2012**, *7*, 375.
- (187) Bender, K.; Hennig, I.; Schweitzer, D.; Dietz, K.; Endres, H.; Keller, H. J. *Mol Cryst Liq Cryst* **1984**, *108*, 359.
- (188) Nagao, Y.; Hirata, T.; Goto, S.; Sano, S.; Kakehi, A.; Iizuka, K.; Shiro, M. *J Am Chem Soc* **1998**, *120*, 3104.
- (189) Taylor, J. C.; Markham, G. D. *J Biol Chem* **1999**, *274*, 32909.
- (190) Iwaoka, M.; Takemoto, S.; Okada, M.; Tomoda, S. *Chem Lett* **2001**, 132.
- (191) Zahn, S.; Frank, R.; Hey-Hawkins, E.; Kirchner, B. *Chem-Eur J* **2011**, *17*, 6034.
- (192) Bauza, A.; Mooibroek, T. J.; Frontera, A. *Angew Chem Int Edit* **2013**, *52*, 12317.

- (193) Zhu Yinghuai, J. A. M., and Narayan S. Hosmane *Recent Developments in the Boron Neutron Capture Therapy (BNCT) Driven by Nanotechnology*; CRC Press: Boca Raton: FL, USA, 2011.
- (194) Base, T.; Bastl, Z.; Plzak, Z.; Grygar, T.; Plesek, J.; Carr, M. J.; Malina, V.; Subrt, J.; Bohacek, J.; Vecernikova, E.; Kriz, O. *Langmuir* **2005**, *21*, 7776.
- (195) Cigler, P.; Kozisek, M.; Rezacova, P.; Brynda, J.; Otwinowski, Z.; Pokorna, J.; Plesek, J.; Gruner, B.; Doleckova-Maresova, L.; Masa, M.; Sedlacek, J.; Bodem, J.; Krausslich, H. G.; Kral, V.; Konvalinka, J. *P Natl Acad Sci USA* **2005**, *102*, 15394.
- (196) Brynda J., M. P., Šícha V., Fábry M., Poncová K., Bakardiev M., Grüner B., Cígler P., Řezáčová P. *Angew Chem Int Ed* **2013**, *52*, 13760.
- (197) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J Mol Biol* **1977**, *112*, 535.
- (198) Scapin, G. *Curr Pharm Design* **2006**, *12*, 2087.
- (199) Feyereisen, M.; Fitzgerald, G.; Komornicki, A. *Chem Phys Lett* **1993**, *208*, 359.
- (200) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G. M.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J. M.; Kollman, P. *J Comput Chem* **2003**, *24*, 1999.
- (201) Wang, J. M.; Cieplak, P.; Kollman, P. A. *J Comput Chem* **2000**, *21*, 1049.
- (202) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J Phys Chem* **1993**, *97*, 10269.
- (203) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J Comput Chem* **2004**, *25*, 1605.
- (204) D.A. Case, T. A. D., T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo,; M. Crowley, R. C. W., W. Zhang, K.M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I.; Kolossváry, K. F. W., F. Paesani, J. Vanicek, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke,; L. Yang, C. T., J. Mongan, V. Hornak, G. Cui, D.H. Mathews, M.G. Seetin, C. Sagui, V. Babin,; Kollman, a. P. A. University of California, San Francisco, 2008.
- (205) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. *J Chem Phys* **1984**, *81*, 3684.
- (206) Northrop, D. B. *Accounts Chem Res* **2001**, *34*, 790.
- (207) Lepsik, M.; Kriz, Z.; Havlas, Z. *Proteins* **2004**, *57*, 279.

- (208) Velazquez-Campoy, A.; Luque, I.; Todd, M. J.; Milutinovich, M.; Kiso, Y.; Freire, E. *Protein Sci* **2000**, *9*, 1801.
- (209) Brynda, J.; Rezacova, P.; Fabry, M.; Horejsi, M.; Stouracova, R.; Sedlacek, J.; Soucek, M.; Hradilek, M.; Lepsik, M.; Konvalinka, J. *J Med Chem* **2004**, *47*, 2030.
- (210) Kozisek, M.; Lepsik, M.; Saskova, K. G.; Brynda, J.; Konvalinka, J.; Rezacova, P. *Febs J* **2014**, *281*, 1834.
- (211) Wolfram Koch, M. C. H. *A Chemist's Guide to Density Functional Theory*; 2nd Edition; Wiley-VCH: Weinheim, 2000.
- (212) Piana, S.; Sebastiani, D.; Carloni, P.; Parrinello, M. *J Am Chem Soc* **2001**, *123*, 8730.
- (213) Tie, Y. F.; Boross, P. I.; Wang, Y. F.; Gaddis, L.; Hussain, A. K.; Leshchenko, S.; Ghoshl, A. K.; Louis, J. M.; Harrison, R. W.; Weber, I. T. *J Mol Biol* **2004**, *338*, 341.
- (214) Leibovitz, E.; Livshiz-Riven, I.; Borer, A.; Taraboulos-Klein, T.; Zamir, O.; Shany, E.; Melamed, R.; Rimon, O. F.; Bradenstein, R.; Chodick, G.; Golan, A. *Scand J Infect Dis* **2013**, *45*, 842.
- (215) Prysycz, L. P.; Nemeth, T.; Gacser, A.; Gabaldon, T. *Genome Biol Evol* **2013**, *5*, 2382.
- (216) Dostal, J.; Brynda, J.; Hruskova-Heidingsfeldova, O.; Pacht, P.; Pichova, I.; Rezacova, P. *J Enzym Inhib Med Ch* **2012**, *27*, 160.
- (217) Dostal, J.; Brynda, J.; Hruskova-Heidingsfeldova, O.; Sieglöva, I.; Pichova, I.; Rezacova, P. *J Struct Biol* **2009**, *167*, 145.
- (218) Dostal, J.; Pecina, A.; Hruskova-Heidingsfeldova, O.; Mareckova, L.; Pichova, I.; Rezacova, P.; Lepsik, M.; Brynda, J. *Acta Crystallogr D* **2015**, *71*, 2494.
- (219) Pecina, A.; Lepsik, M.; Rezac, J.; Brynda, J.; Mader, P.; Rezacova, P.; Hobza, P.; Fanfrlik, J. *J Phys Chem B* **2013**, *117*, 16096.
- (220) Massova, I.; Kollman, P. A. *J Am Chem Soc* **1999**, *121*, 8133.
- (221) Krishnamurthy, V. M.; Kaufman, G. K.; Urbach, A. R.; Gitlin, I.; Gudiksen, K. L.; Weibel, D. B.; Whitesides, G. M. *Chem Rev* **2008**, *108*, 946.
- (222) McDonald, P. C.; Winum, J. Y.; Supuran, C. T.; Dedhar, S. *Oncotarget* **2012**, *3*, 84.
- (223) Lock, F. E.; McDonald, P. C.; Lou, Y.; Serrano, I.; Chafe, S. C.; Ostlund, C.; Aparicio, S.; Winum, J. Y.; Supuran, C. T.; Dedhar, S. *Oncogene* **2013**, *32*, 5210.

- (224) Tafreshi, N. K.; Lloyd, M. C.; Bui, M. M.; Gillies, R. J.; Morse, D. L. *Subcell Biochem* **2014**, *75*, 221.
- (225) Supuran, C. T. *Nature Reviews Drug Discovery* **2008**, *7*, 168.
- (226) Monti, S. M.; Supuran, C. T.; De Simone, G. *Expert Opin Ther Pat* **2013**, *23*, 737.
- (227) Gitto, R.; Damiano, F. M.; Mader, P.; De Luca, L.; Ferro, S.; Supuran, C. T.; Vullo, D.; Brynda, J.; Rezacova, P.; Chimirri, A. *J Med Chem* **2012**, *55*, 3891.
- (228) Lipscomb, W. N. *Boron Hydrides*; W. A. Benjamin Inc.: New York, USA, 1963.
- (229) Grimes, R. N. *Carboranes, 2nd Edition*; Academic Press: London, 2011.
- (230) Valliant, J. F.; Guenther, K. J.; King, A. S.; Morel, P.; Schaffer, P.; Sogbein, O. O.; Stephenson, K. A. *Coord Chem Rev* **2002**, *232*, 173.
- (231) Lesnikowski, Z. J. *Collect Czech Chem C* **2007**, *72*, 1646.
- (232) Sivaev, I. B.; Bregadze, V. V. *Eur J Inorg Chem* **2009**, 1433.
- (233) Issa, F.; Kassiou, M.; Rendina, L. M. *Chem Rev* **2011**, *111*, 5701.
- (234) Scholz, M.; Hey-Hawkins, E. *Chem Rev* **2011**, *111*, 7035.
- (235) Endo, Y.; Iijima, T.; Yamakoshi, Y.; Fukasawa, H.; Miyaura, C.; Inada, M.; Kubo, A.; Itai, A. *Chem Biol* **2001**, *8*, 341.
- (236) Reynolds, R. C.; Campbell, S. R.; Fairchild, R. G.; Kisliuk, R. L.; Micca, P. L.; Queener, S. F.; Riordan, J. M.; Sedwick, W. D.; Waud, W. R.; Leung, A. K. W.; Dixon, R. W.; Suling, W. J.; Borhani, D. W. *J Med Chem* **2007**, *50*, 3283.
- (237) Rezacova, P.; Pokorna, J.; Brynda, J.; Kozisek, M.; Cigler, P.; Lepsik, M.; Fanfrik, J.; Rezac, J.; Saskova, K. G.; Sieglöva, I.; Plesek, J.; Sicha, V.; Gruner, B.; Oberwinkler, H.; Sedlacek, J.; Krausslich, H. G.; Hobza, P.; Kral, V.; Konvalinka, J. *J Med Chem* **2009**, *52*, 7132.
- (238) Fujii, S.; Goto, T.; Ohta, K.; Hashimoto, Y.; Suzuki, T.; Ohta, S.; Endo, Y. *J Med Chem* **2005**, *48*, 4654.
- (239) Fujii, S.; Masuno, H.; Taoda, Y.; Kano, A.; Wongmayura, A.; Nakabayashi, M.; Ito, N.; Shimizu, M.; Kawachi, E.; Hirano, T.; Endo, Y.; Tanatani, A.; Kagechika, H. *J Am Chem Soc* **2011**, *133*, 20933.
- (240) Julius, R. L.; Farha, O. K.; Chiang, J.; Perry, L. J.; Hawthorne, M. F. *Proc Natl Acad Sci U S A* **2007**, *104*, 4808.
- (241) Crabtree, R. H.; Siegbahn, P. E. M.; Eisenstein, O.; Rheingold, A. L. *Accounts Chem Res* **1996**, *29*, 348.
- (242) Custelcean, R.; Jackson, J. E. *Chem Rev* **2001**, *101*, 1963.

- (243) Fanfrlik, J.; Lepsik, M.; Horinek, D.; Havlas, Z.; Hobza, P. *Chem Phys Chem* **2006**, *7*, 1100.
- (244) Fanfrlik, J.; Hnyk, D.; Lepsik, M.; Hobza, P. *Phys Chem Chem Phys* **2007**, *9*, 2085.
- (245) Fanfrlik, J.; Brynda, J.; Rezac, J.; Hobza, P.; Lepsik, M. *J Phys Chem B* **2008**, *112*, 15094.
- (246) Matejicek, P.; Zednik, J.; Uselova, K.; Plestil, J.; Fanfrlik, J.; Nykanen, A.; Ruokolainen, J.; Hobza, P.; Prochazka, K. *Macromolecules* **2009**, *42*, 4829.
- (247) Sedlak, R.; Fanfrlik, J.; Hnyk, D.; Hobza, P.; Lepsik, M. *J Phys Chem A* **2010**, *114*, 11304.
- (248) Gutten, O.; Besseova, I.; Rulisek, L. *J Phys Chem A* **2011**, *115*, 11394.
- (249) Mader, P.; Pecina, A.; Cigler, P.; Lepsik, M.; Sicha, V.; Hobza, P.; Gruener, B.; Fanfrlik, J.; Brynda, J.; Rezacova, P. *Biomed Res Int* **2014**.
- (250) Alterio, V.; Hilvo, M.; Di Fiore, A.; Supuran, C. T.; Pan, P. W.; Parkkila, S.; Scaloni, A.; Pastorek, J.; Pastorekova, S.; Pedone, C.; Scozzafava, A.; Monti, S. M.; De Simone, G. *P Natl Acad Sci USA* **2009**, *106*, 16233.
- (251) Warren, G. L.; Andrews, C. W.; Capelli, A. M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. *J Med Chem* **2006**, *49*, 5912.
- (252) Leach, A. R.; Shoichet, B. K.; Peishoff, C. E. *J Med Chem* **2006**, *49*, 5851.
- (253) Gohlke, H.; Klebe, G. *Angew Chem Int Edit* **2002**, *41*, 2645.
- (254) Xu, M.; Lill, M. A. *Drug discovery today. Technologies* **2013**, *10*, e411.
- (255) Antony, J.; Grimme, S. *J Comput Chem* **2012**, *33*, 1730.
- (256) Riley, K. E.; Pitonak, M.; Jurecka, P.; Hobza, P. *Chem Rev* **2010**, *110*, 5023.
- (257) Dvir, H.; Wong, D. M.; Harel, M.; Barril, X.; Orozco, M.; Luque, F. J.; Munoz-Torrero, D.; Camps, P.; Rosenberry, T. L.; Silman, I.; Sussman, J. L. *Biochemistry* **2002**, *41*, 2970.
- (258) Bandarage, U. K.; Wang, T.; Come, J. H.; Perola, E.; Wei, Y.; Rao, B. G. *Bioorg Med Chem Lett* **2008**, *18*, 44.
- (259) Steuber, H.; Heine, A.; Klebe, G. *J Mol Biol* **2007**, *368*, 618.
- (260) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. *J Med Chem* **2006**, *49*, 6177.
- (261) Korb, O.; Stutzle, T.; Exner, T. E. *J Chem Inf Model* **2009**, *49*, 84.

- (262) Trott, O.; Olson, A. J. *J Comput Chem* **2010**, *31*, 455.
- (263) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. *J Comput Aid Mol Des* **1997**, *11*, 425.
- (264) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. *J Mol Biol* **1997**, *267*, 727.
- (265) Mooij, W. T. M.; Verdonk, M. L. *Proteins* **2005**, *61*, 272.
- (266) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J Comput Chem* **2004**, *25*, 1157.
- (267) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. *J Comput Chem* **2000**, *21*, 132.
- (268) Jakalian, A.; Jack, D. B.; Bayly, C. I. *J Comput Chem* **2002**, *23*, 1623.

List of Publications

Included in the Thesis:

1. Kolar, M. H.; Deepa, P.; Ajani, H.; Pecina, A.; Hobza, P.: Characteristics of a sigma-Hole and the Nature of a Halogen Bond; *Topics in Current Chemistry* **2015**, 359, 1.
2. Fanfrlik, J.*; Prada, A.*; Padelkova, Z. Pecina, A.; Machacek, J.; Lepsik, M.; Holub, J.; Ruzicka, A.; Hnyk, D.; Hobza, P.: The Dominant Role of Chalcogen Bonding in the Crystal Packing of 2D/3D Aromatics; *Angewandte Chemie International Edition* **2015**, 53, 10139.
* These authors contributed equally
3. Pecina, A.; Lepsik, M.; Hnyk, D.; Hobza, P.; Fanfrlik, J.: Chalcogen and Pnicogen Bonds in Complexes of Neutral Icosahedral and Bicapped Square-Antiprismatic Heteroboranes; *Journal of Physical Chemistry A* **2015**, 119, 1388.
4. Pecina, A.*; Prenosil, O.*; Fanfrlik, J.; Rezac, J.; Granatier, J.; Hobza, P.; Lepsik, M.: On the Reliability of the Corrected Semiempirical Quantum Chemical Method (PM6-DH2) for Assigning the Protonation States in Hiv-1 Protease/Inhibitor Complexes; *Collection of Czechoslovak Chemical Communications* **2011**, 76, 457.
* These authors contributed equally
5. Dostal, J.; Pecina, A.; Hruskova-Heidingsfeldova, O.; Mareckova, L.; Pichova, I.; Rezacova, P.; Lepsik, M.; Brynda, J.: Atomic resolution crystal structure of Sapp2p, a secreted aspartic protease from *Candida parapsilosis*; *Acta Crystallographica Section D-Biological Crystallography* **2015**, 71, 2494.
6. Pecina, A.; Lepsik, M.; Rezac, J.; Brynda, J.; Mader, P.; Rezacova, P.; Hobza, P.; Fanfrlik, J.: QM/MM Calculations Reveal the Different Nature of the Interaction of Two Carborane-Based Sulfamide Inhibitors of Human Carbonic Anhydrase II; *Journal of Physical Chemistry B* **2013**, 117, 16096.

7. Mader, P.*; Pecina, A.*; Cigler, P.; Lepsik, M.; Sicha, V.; Hobza, P.; Gruener, B.; Fanfrlik, J.; Brynda, J.; Rezacova, P.: Carborane-Based Carbonic Anhydrase Inhibitors: Insight into CAII/CAIX Specificity from a High-Resolution Crystal Structure, Modeling, and Quantum Chemical Calculations; *Biomed Research International* **2014**, 389869.

* These authors contributed equally

8. Pecina, A.*; Meier, R.*; Fanfrlik, J.; Lepsik, M.; Rezac, J.; Hobza, P.; Baldauf, C.: The SQM/COSMO Filter: Reliable Native Pose Identification Based on the Quantum-Mechanical Description of Protein–Ligand Interactions and Implicit COSMO Solvation; *Chemical Communications* **2016**, 52, 0000.

* These authors contributed equally

Not Included in the Thesis:

1. Lepsik, M.; Rezac, J.; Kolar, M. H.; Pecina, A.; Hobza, P.; Fanfrlik, J.: The Semiempirical Quantum Mechanical Scoring Function for In Silico Drug Design; *ChemPlusChem* **2013**, 78, 921.
2. Sedlak, R.; Fanfrlik, J.; Pecina, A.; Hnyk, D.; Hobza, P.; Lepsik, M.: Noncovalent Interactions of Heteroboranes; *Challenges and Advances in Computational Chemistry and Physics* **2015**, 20, 219.
3. Fanfrlik, J.; Sedlak, R.; Pecina, A.; Rulisek, L.; Dostal, L.; Moncol, J.; Ruzicka, A.; Hobza, P.: The non-planarity of the benzene molecule in the X-ray structure of the chelated bismuth(III) heteroboroxine complex is not supported by quantum mechanical calculations; *Dalton Transaction* **2016**, 45, 462.

Declaration of co-authorship

Prohlášení spoluautorů upřesňující podíl Mgr. Adama Peciny na publikacích přiložených k disertaci:

1. Kolar, M. H.; Deepa, P.; Ajani, H.; Pecina, A.; Hobza, P.; *Topics in Current Chemistry* **2015**, 359, 1.
2. Fanfrlik, J.; Prada, A.; Padelkova, Z. Pecina, A.; Machacek, J.; Lepsik, M.; Holub, J.; Ruzicka, A.; Hnyk, D.; Hobza, P.; *Angewandte Chemie International Edition* **2015**, 53, 10139.
3. Pecina, A.; Lepsik, M.; Hnyk, D.; Hobza, P.; Fanfrlik, J.; *Journal of Physical Chemistry A* **2015**, 119, 1388.
4. Pecina, A.; Prenosil, O.; Fanfrlik, J.; Rezac, J.; Granatier, J.; Hobza, P.; Lepsik, M.; *Collection of Czechoslovak Chemical Communications* **2011**, 76, 457.
5. Dostal, J.; Pecina, A.; Hruskova-Heidingsfeldova, O.; Mareckova, L.; Pichova, I.; Rezacova, P.; Lepsik, M.; Brynda, J.; *Acta Crystallographica Section D-Biological Crystallography* **2015**, 71, 2494.
6. Pecina, A.; Lepsik, M.; Rezac, J.; Brynda, J.; Mader, P.; Rezacova, P.; Hobza, P.; Fanfrlik, J.; *Journal of Physical Chemistry B* **2013**, 117, 16096.
7. Mader, P.; Pecina, A.; Cigler, P.; Lepsik, M.; Sicha, V.; Hobza, P.; Gruener, B.; Fanfrlik, J.; Brynda, J.; Rezacova, P.; *Biomed Research International* **2014**, 389869.
8. Pecina, A.; Meier, R.; Fanfrlik, J.; Lepsik, M.; Rezac, J.; Hobza, P.; Baldauf, C.; *Chemical Communications* **2016**, 52, 0000.

M.Sc. Adam Pecina carried out the work in publications 3, 4, 6, 8 with a high degree of independence (80%) and he contributed substantially (35-60%) also in other four publications.

Mgr. Adam Pecina je prvním (nebo sdíleným prvním) autorem pěti publikací (3, 4, 6, 7, 8) přiložených k disertaci, což jednoznačně vymezuje jeho podíl. V těchto případech je tento podíl dominantní a to ve všech fázích přípravy publikace, od zadání tématu až k jejímu sepsání. Podíl Adama Peciny je značný i ve třech dalších publikacích.

Prague, January 2016



RNDr. Martin Lepšík, Ph.D.



prof. Ing. Pavel Hobza, Dr.Sc., dr. h. c., FRSC