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 literaly 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 treatement of intermolecular interactions. It consists of eight original publications devided 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 pnicogen 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 spjaté 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 spjaté s přenosy protonu, s ionty kovů a s neobvyklými molekulami jakými jsou borany. Strukturní 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 čtyrech 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

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

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, transmemrane 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 1894[6] 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 fundametal for the existance of liquid states of matter and molecular clusters but also essential for nothing small than life itself.[9] The most important representants 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.

$$\Delta G_b = RT \ln K_i \qquad (Eq.\ 1.1)$$

,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 equation[11] (Eq. 1.2), whereas in the case of tight-binding inhibitors, the enzyme concentration [E] must be considered (Eq. 1.3)

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

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

, 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).

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

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 Lewitt and Arieh Warshel. Aside from routinely used molecular dynamic simulations (MD) and other extensions[15], 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 method[16] to widely used MM-GB/PBSA methods[17] 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 affinities[19] 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 timescales[21] Several approaches, for example metadynamics[22] or accelerated MD[23], 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 treatement 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 literature[24-28], so below I only give a brief overview.

There are various efforts toward the improvement of biomolecular force fields, such as AMOEBA[29] or QMPFF polarizable forcefields[30], 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 determination[37-42], 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 impoven 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 models[46-48] or include polarization effects[49] 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 schemes[56-58], DFT-D3 on truncated protein-ligand complexes[59] 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 binding[50] 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 space[67] 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 GB[71]) and QM-based (COSMO[72] or SMD[73]), have been carefully compared.[74] These methods are therefore used within the SQM-based scoring function[75], 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 ($\Delta G'^{w}_{conf}$) and the change of entropy upon ligand binding ($-T\Delta S$).

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

Its generality has been demonstrated in various non-covalent protein-ligand complexes[77-80] 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 essencial 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.

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

.

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

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

,where ΔE_{int}(A…B) is the interaction energy for the complex, E(A…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 disadvatage 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 Bernardi[82], 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).

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

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. Incompletness 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's[86] schemes in Eq. 2.4 and Eq. 2.5, respectively:

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

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

,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 aproximate 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 GMTKN30[90] 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):

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

,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 site 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 DFT[105] 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 desribe 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 interamolecular distances.

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

,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,…) 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 approach[111] 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-Gordon[112] 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 coeficients 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 s 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-D[115], OMxD[116], AM1-FS1[117] or PM3-PIF[118] and PM3-PDDG[119] however non 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 round 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 abbility 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 correction[113] the resulting PM6-DH method[68] 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 disadvatage 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

term[111], 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. Yelding 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 overcame 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 MOZYME[124], 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 corvengence 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 meaningfull 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 procedure[137], 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}, $E^2_{Ex\text{-}Ind}$ and $E^2_{Ex\text{-}Disp}$) and δHF term represents higher than second-order terms covered by the Hartree-Fock approach.

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

DFT-SAPT decomposition of the ineraction 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^{1}_{Ex}) 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 $E^{2}_{Ex\text{-}Ind}$ and $E^{2}_{Ex\text{-}Disp}$ 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:

$$E^S = E^I_{QM} + E^O_{MM} + E^{I-O}_{coupling} \qquad (Eq.2.9)$$

,where E^I_{QM} is the energy of the inner part of the system (e.g. the active site of the protein) calculated by quantum mechanics, E^O_{MM} is the energy of the rest of the system (the outer part) solved by molecular mechanics and the $E^{I-O}_{coupling}$ 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 mutualy 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 substractive 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:

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

,where E^{Model}_{QM} is the energy of the inner part of the system containing link atoms (model system) estimated at QM level, E^{Model}_{MM} is the energy of the same model system calculated by MM and E^{Real}_{MM} 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 substacted 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 inavitable 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 bz 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]

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

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 arised.

The *Generalized Born* (GB) implicit solvent model[71] 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 overlaping 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 pnicogen 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 pnicogen 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 transfer[147] and the concept of sigma-hole bonding[148,149] had began to be used later on with the knowledge of their electrostatic origin.[150]

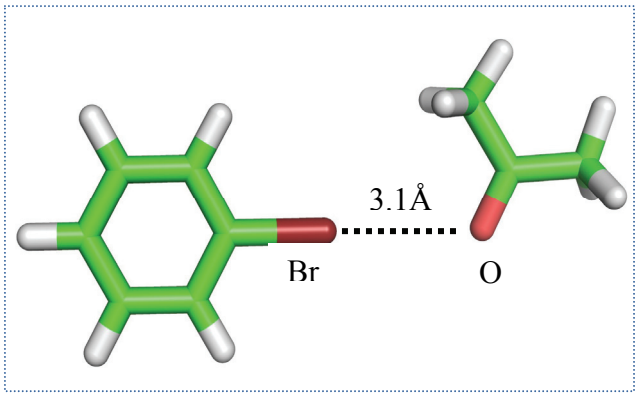


Fig. 3.1: Halogen bond of bromobenzene…acetone complex

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/bohr^3 electron density[152] 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*···*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*···*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*···*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…NH_3[157] 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.

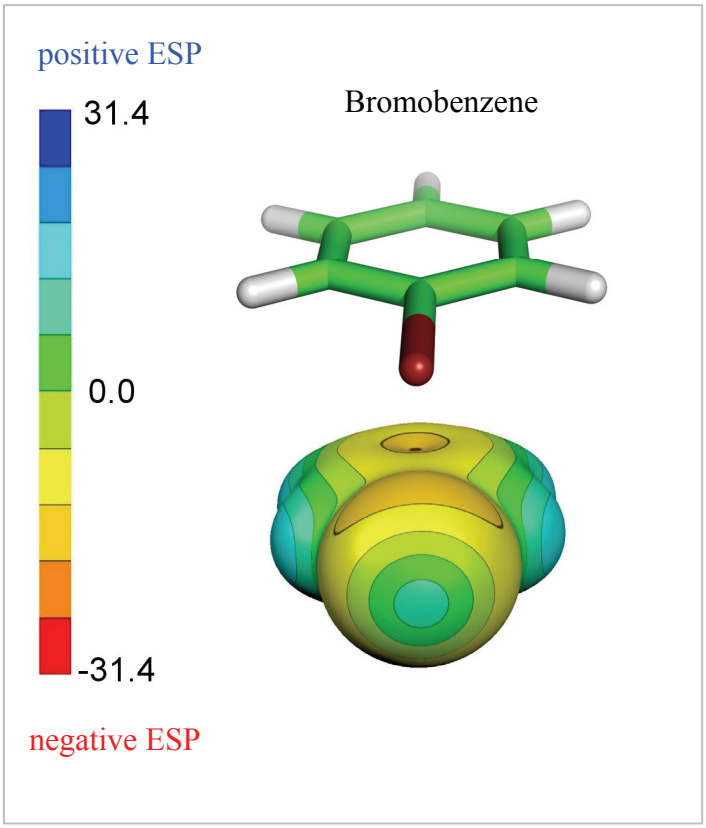


Fig. 3.2: The typical σ-hole, the area of positive ESP created on the top of bromine atom in bromobenzene molecule.

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 CH_3Cl, CH_3Br, CH_3I, CF_3Cl, CF_3Br and CF_3I with OCH_2, complexes of chlorobenzene, bromobenzene and iodobenzene with OC_3H_6, NC_3H_9 and $SHCH_3$, complexes of H_3CBr, H_3CI, F_3CBr and F_3CI with benzene)[156], 46 complexes from the XB51 dataset (HCN, NH_3 and HCP in complexes with ICF_3, BrF, ClF, $INC_4H_2O_2$, $BrO_2C_4H_2N$, BrC_6H_5, IC_6H_5 and Br_2, FI and H_3CI in complexes with FC_2H, FCH_3, NCH, NH_3, OCH_2, OPH_3, PCH, NC_5H_5 and H_3CI...LiH)[157], 11 complexes from the papers previously published by Hobza *et.al* (benzene with F_2, Cl_2, and Br_2, I_2...I_2, Br_2...Br_2, Cl_2...Cl_2, F_2...F_2, Br_2 with trimethylbenzene and hexamethylbenzene and twice CH_2BrOH...CH_2BrOH complex with Br-O halogen bonding and with Br-Br dihalogen bond)[160-162] and 13 complexes (C_2H_3Cl, C_2HCl, C_2H_3Br, C_2HBr, C_2H_3I and C_2HI in complexes with H_2CO, H_2O and NH_3) 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 (CH_3CN and CO in complex with BrF, ClF, BrF_3 and ClF_3)[164], 15 complexes (ICN, IBr, ICl and I_2 in complexes with NC_5H_5, complexes of I_2, C_4F_9I and C_6F_5I with OSC_2H_6, NC_6H_{15} and NC_7H_{13} and complexes of I_2 and C_4F_9I with OPC_3H_9) from Ref. [165] and finally 17 structures of organic crystals (I_2...1,4-diazabicyclo[2.2.2]octane, I_2...1,3-dithiole-2-thione-4-carboxyclic acid, C_6Cl_6...C_6Cl_6, C_6Br_6...C_6Br_6, $C_4N_3H_4Br$...$C_7F_4O_2HBr$, $C_6F_4I_2$...$I_2F_4C_6$, $C_7F_4O_2HBr$...$NBrC_4N_3H_2$, three different orientations of 1,2-TFIB...TMO, 2-mercapto-1-methylimidazole...1,2-TFIB, 4,4'-bipyridine ...1,2-TFIB, $(3,4,5\text{-trichlorophenol})_2$ 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 exchangecorrelation 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 dataset[156] 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 (F_2, Cl_2, ClF, ClF_3, F_3CCl, C_2H_3Cl, C_2HCl, chlorobenzene, C_6Cl_6, $C_6H_2OHCl_3$, Br_2, BrF, BrF_3, H_3CBr, F_3CBr, C_2H_3Br, C_2HBr, bromobenzene, C_6Br_6, $BrC_4H_2NO_2$, CH_2BrOH, $C_7F_4O_2HBr$, I_2, IF, ICl, IBr, ICN, H_3CI, F_3CI, C_2H_3I, C_2HI, iodobenzene, C_6F_5I, C_4F_9I, $INC_4H_2O_2$, $HO_2C_7F_4I$, TFIB) possess a positive σ-hole (with the exception of slightly negative σ-hole of H_3CCl. 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.

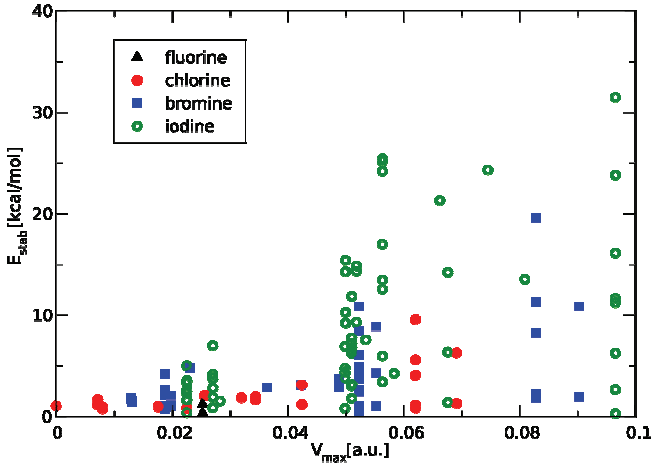


Fig. 3.3: The dependence of the stabilization energy E_{stab} on the magnitude of the σ-hole V_{max}

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.

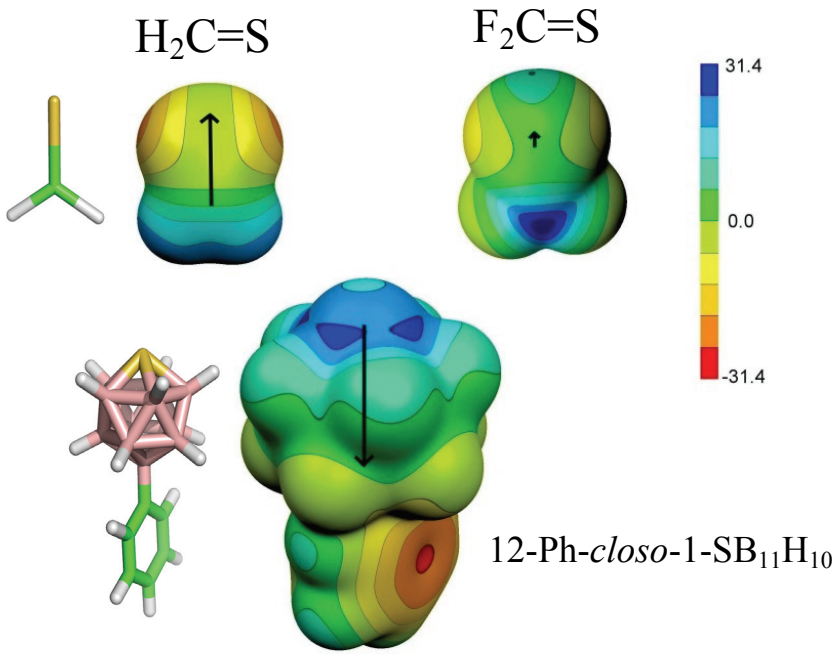


Fig. 3.4: Comparison of calculated ESPs and dipole moments (arrows) for a sulphur-bound 2D organic structures and 3D inorganic boron hydrid.

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-$SB_{11}H_{10}$), iodine atom (12-I-*closo*-1-$SB_{11}H_{10}$) and phenyl ring (12-Ph-*closo*-1-$SB_{11}H_{10}$).

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 dispersion[142] 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-$SB_{11}H_{10}$ is comparable to multisubstituted bromobenzene).[155] When going to 12-Cl-*closo*-1-$SB_{11}H_{10}$, 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.

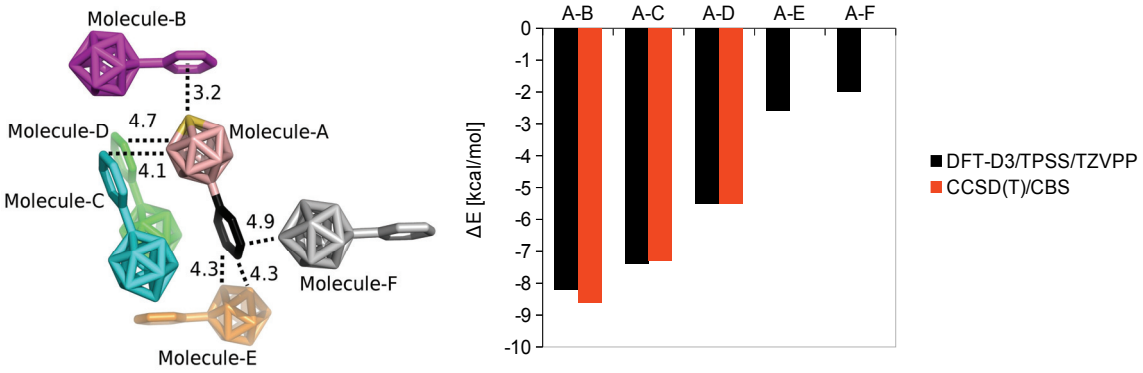


Fig. 3.5: All binding motifs of 12-Ph-*closo*-1-$SB_{11}H_{10}$ crystal model, where A…B motif is characterized as the B-S…π chalcogen bond (on the left) with the estimated interaction energies of all these motifs at DFT-D3 and CCSD(T)/CBS level (on the right).

The interaction energies of A-C and A-D stacking motifs of 12-Ph-*closo*-1-$SB_{11}H_{10}$ 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]

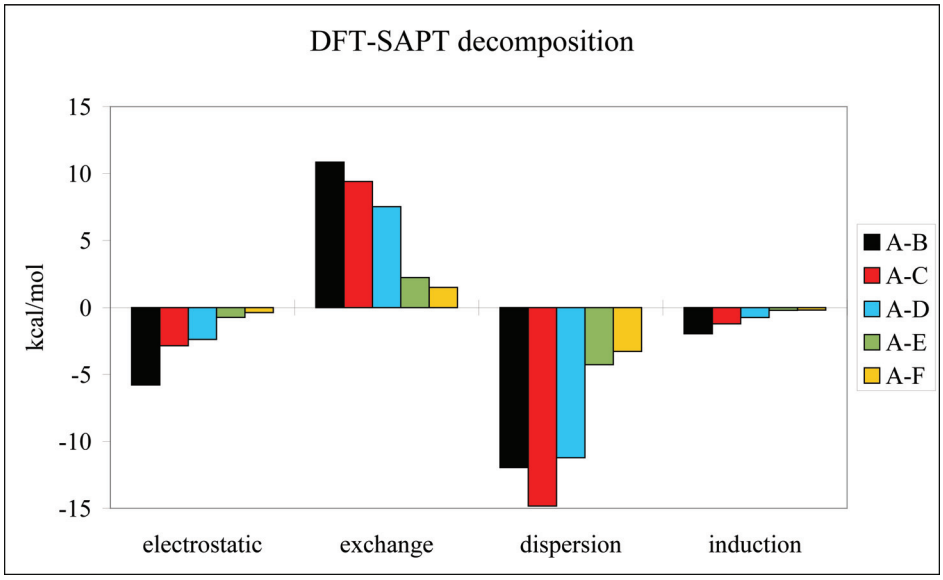


Fig. 3.6: The decomposition of the stabilization energy of all binding motifs of 12-Ph-*closo*-1-$SB_{11}H_{10}$ estimated by the DFT-SAPT method.

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 pnicogen atoms are incorporated in the 3D cages, whereas halogens are considered as *exo*-substituents of dicarbaboranes.

Our dataset included 12 heteroborane molecules (*closo*-1-$SB_{11}H_{11}$, 12-F-*closo*-1-$SB_{11}H_{10}$, 12-Cl-*closo*-1-$SB_{11}H_{10}$, 12-Br-*closo*-1-$SB_{11}H_{10}$, *closo*-1-$SeB_{11}H_{11}$, *closo*-1-SB_9H_9, *closo*-1,2-$P_2B_{10}H_{10}$, *closo*-1,2-$As_2B_{10}H_{10}$, *closo*-2,1-PCB_8H_9, *closo*-6,1-PCB_8H_9, 12-Br-*closo*-1,2-$C_2B_{10}H_{11}$ and 1-Br-*closo*-1,2-$C_2B_{10}H_{11}$). 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 level[138] we used the scaling factor estimated on a model system (*closo*-1-SB_9H_9…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.

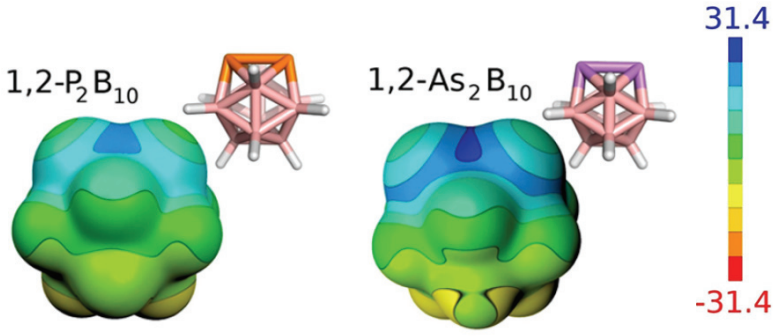


Fig. 3.7: ESP calculated on *closo*-1,2-$P_2B_{10}H_{10}$ and *closo*-1,2-$As_2B_{10}H_{10}$ shows the σ-holes position in the valley between pnicogen atoms.

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-PCB_8H_9 is the magnitude of the σ-hole higher about 5 kcal/mol than in *closo*-6,1-PCB_8H_9 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-$SB_{11}H_{11}$ cage interacting with benzene, trimethylamine, dimethyl ether, acetone and formamide partners. The complex 1-$SB_{11}H_{11}$…benzene represented the simplified model system of previously studied chalcogen bond in 12-Ph-*closo*-1-$SB_{11}H_{10}$ 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-$SB_{11}H_{11}$ ranges between 130° and 165° which agree with the positions of σ-holes. In all 1-$SB_{11}H_{11}$ 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-$SeB_{11}H_{11}$, 12-Cl-*closo*-1-$SB_{11}H_{10}$ and *closo*-1-SB_9H_9, and analized 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 ($SB_{11}H_{11}$ cage complexes lower down the ΔE by about 0.2 kcal/mol in contrast to SB_9H_9 cage) nor the *exo*-substitution of chlorine atom (12-Cl-*closo*-1-$SB_{11}H_{10}$ complexes have ΔE comparable to 1-$SB_{11}H_{11}$).

Pnicogen bonding was represented by 1,2-$P_2B_{10}H_{10}$ and *closo*-1,2-$As_2B_{10}H_{10}$ complexes with bond acceptors. The interaction energies of the 1,2-$P_2B_{10}H_{10}$ are of the similar strenght as in the 1-$SB_{11}H_{11}$ (pnicogen bonds were about 0.5 kcal/mol less stable). The difference can be seen for the interactions with benzene, where the $SB_{11}H_{11}$ 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-$C_2B_{10}H_{11}$ 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 anabled mainly by dispersion energy. In the case of 1-Br-*closo*-1,2-$C_2B_{10}H_{11}$ 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 pnicogen and chalcogen atoms incorporated in heteroborane cages are positively charged entities carring 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, pnicogen 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 definition[154] of halogen bonding has thus arisen.

Halogen, chalcogen and pnicogen 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 innorganic 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 impovement 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, aromacity, 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 prerequesite of the structure-based drug design. Protein structure is the 3D arrangments 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)-NH_2 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 treatement 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-like[144] substractive QM/MM approach with link atoms and mechanical embedding combined with GB[71] or COSMO[72] implicit solvent models. The QM part comprising up to 500 atoms is usually solved by corrected DFT methods[111,113] accelerated by density fitting[199] and the corrected semiempirical PM6 method[67] with linear scaling MOZYME algorithm[124] is used for even bigger QM parts covering thousands of atoms standardly. The MM part is usually treated using AMBER ff03 forcefield[200] for the protein and GAFF parameters[201] and RESP charges at HF/6-31G* level[202] 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 program[203] 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 thermostat[205] 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 finding 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 C_2-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 monoprotonated[207] 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-barier hydrogen bond.

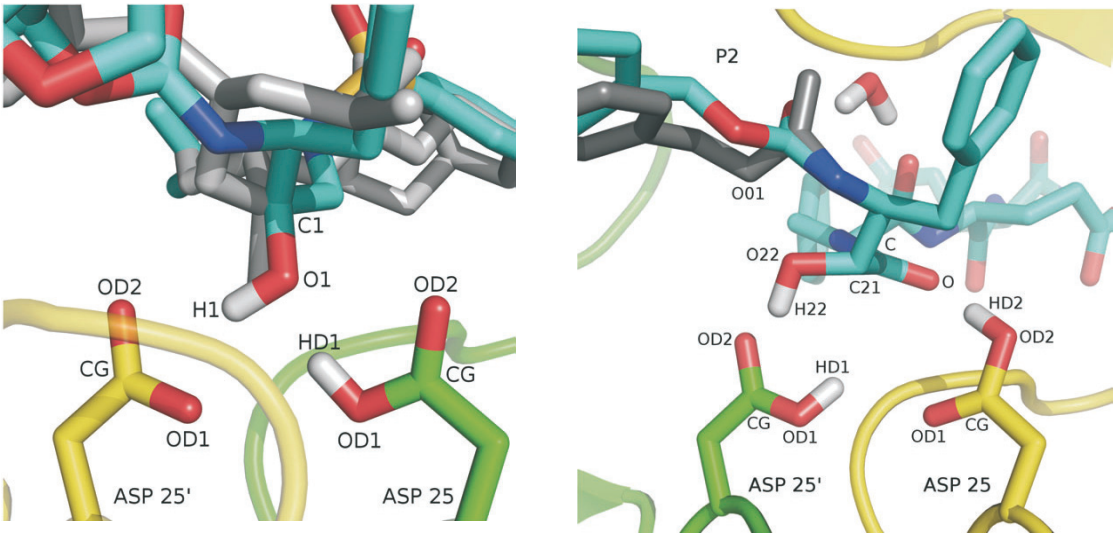


Fig. 3.8: On the left: the active site of HIV-1PR/DRV complex with monoprotonated aspartic dyad, where the first of four possible protonation variants is depicted (H1 of the ligand in the orientation A interacts with OD2 of Asp25′ whereas OD1 of Asp25 is protonated). On the right: KI2 inhibitor bound 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 method[121] 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-D[113] 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).

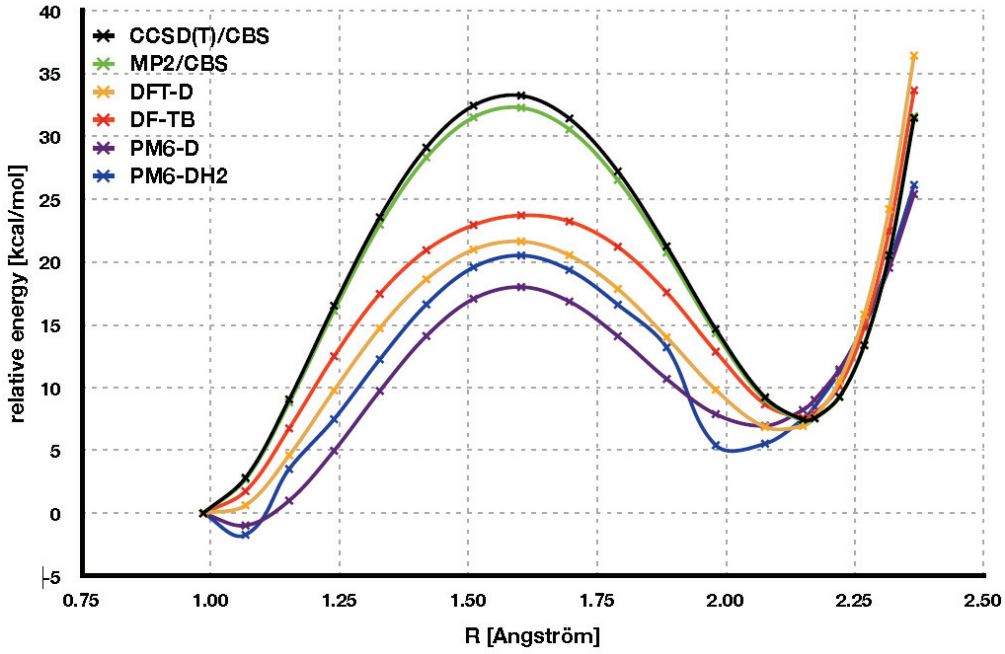


Fig. 3.9: The graph of proton-transfer barriers calculated in several QM-based methods.

Proton transfer has occured 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 barriers[211] 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, namelly 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 addopting 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)

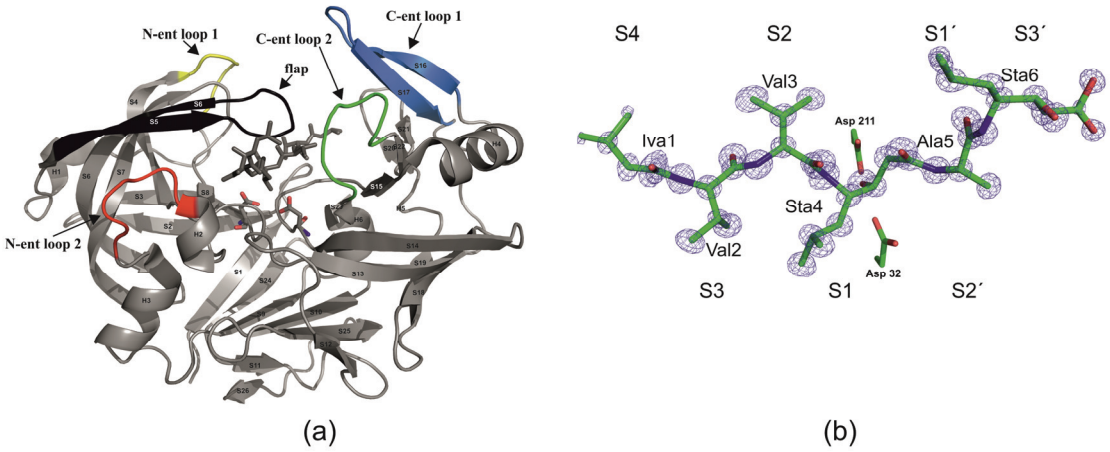


Fig. 3.10: (a) Sapp2p/pepstatin A complex, where protein is shown as ribbon and pepstatin A and catalytic aspartates as sticks. **(b)** Pepstatin A binding pose in Sapp2p with depicted electron density map in contours. Further names of peptidomimetic pepstatin A residues (Iva1-Val2-Val3-Sta4-Ala5-Sta6), 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 Cα 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=0.83Å) 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 Oδ2 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 (Oδ1) is protonated. The occupancy for these two states was estimated to be 60% and 40%, respectively, based on the C-Oδ1 distance. The suggested protonation of Asp32 Oδ1 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 Oδ1. The possible protonation variants of catalytic aspartates are depicted in Figure 3.11.

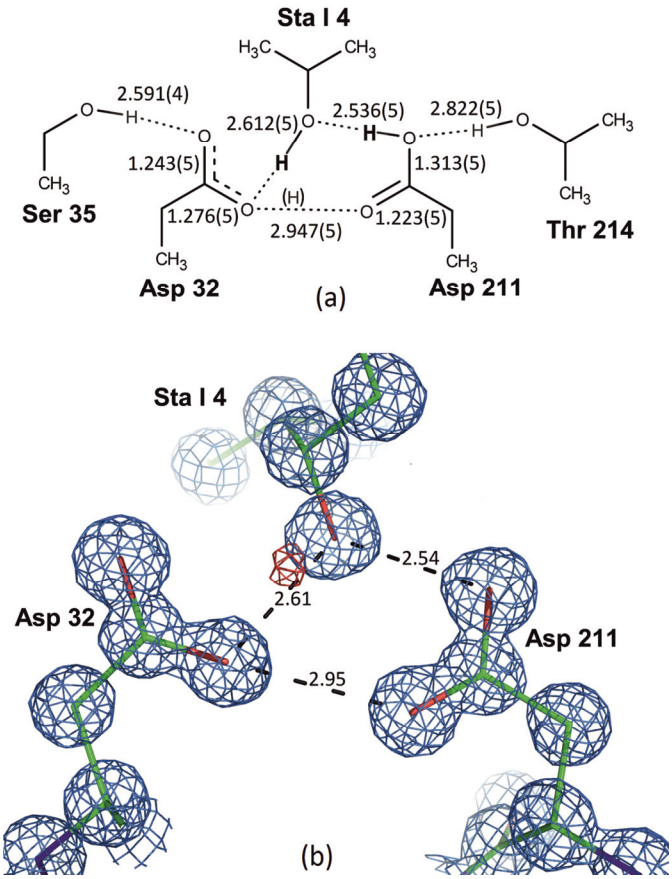


Fig. 3.11: (a) Schematic diagram of polar interactions in the active site of Sapp2p/pepstatin A complex. Clearly assigned hydrogens are in bold, hypothetical hydrogen atom is in parentheses. Distances are in Å with standard deviations in parentheses. **(b)** The active site in detail in stick

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=0.83Å)[218] and Sapp1p/pepstatin A (PDB code 3FV3, R=1.85Å)[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 Oδ2 oxygen and Asp32 either deprotonated or monoprotonated on the Oδ1 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 method[70] coupled with the COSMO implicit solvent model[72] using the linear scaling algorithm MOZYME[124]. 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 Oδ1 of Asp 32.

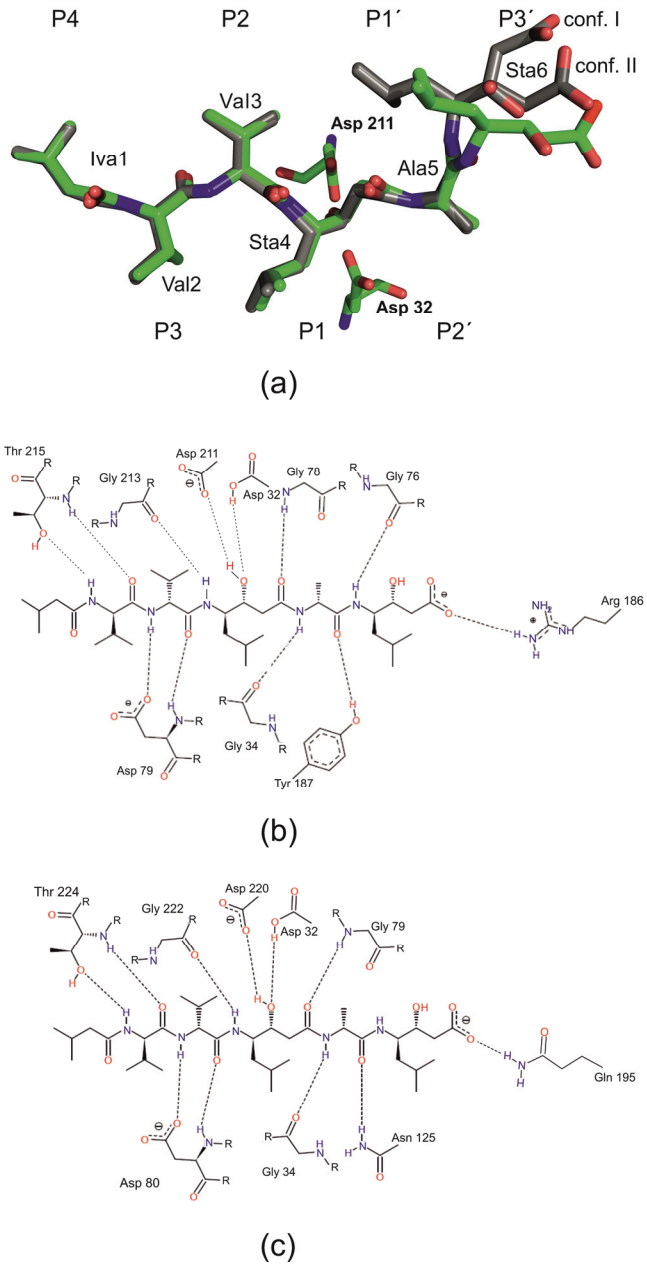


Fig. 3.12: (a) Aligned structures of pepstatin A bound to both isoenzymes, *i.e.* Sapp2p (carbon atoms in green color) and Sapp1p in I a II conformations (carbon atoms in grey color). **(b)** and **(c)** Schematic representation of hydrogen bonding of pepstatin A with aminoacid residues of Sapp2p and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

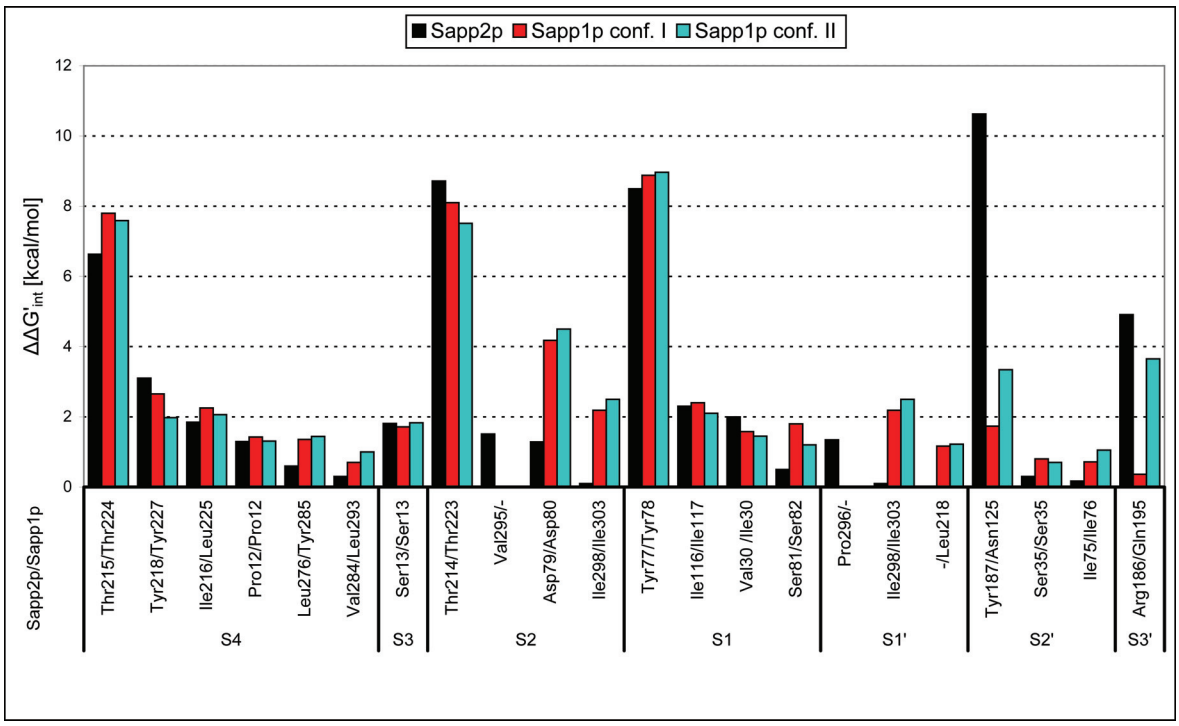


Fig. 3.13: The changes in free energy of interaction ($\Delta\Delta G'_{int}$) in kcal/mol upon mutation of a given amino acid residue to glycine to study the roles of individual amino acid side chains of Sapp2p and Sapp1p active sites in binding with pepstatin A.

As seen in Figure 3.13, individual substrate binding subsites of Sapp2p and Sapp1p (S4-S3´) possess aminoacid residues that can be devided 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 glycin 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 relativelly 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 isoenyzmes with the strognest 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 data[221], 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 SO_2-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/antagonists[235], boron-containing antifolates[236], 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 bonds[241-244] via B-H...Na^+ bridges[245,246] to $B_2H...\pi$ 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 inhibitors[236], 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).

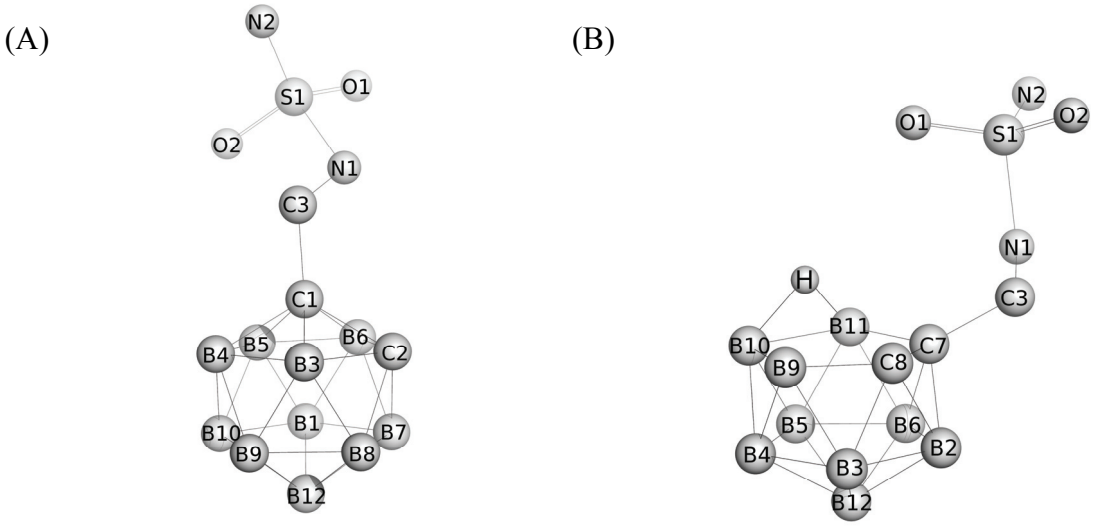


Fig. 3.14: (A) 1-methylenesulfamide-1,2-dicarba-*closo*-dodecaborane (**1a**) structure **(B)** 7-methylene sulfamide-(7,8-*nido*-dicarbaundecaborate) (**7a**) structure. Hydrogens are omitted for clarity with the exception of a B-H-B bridge in **7a**.

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).

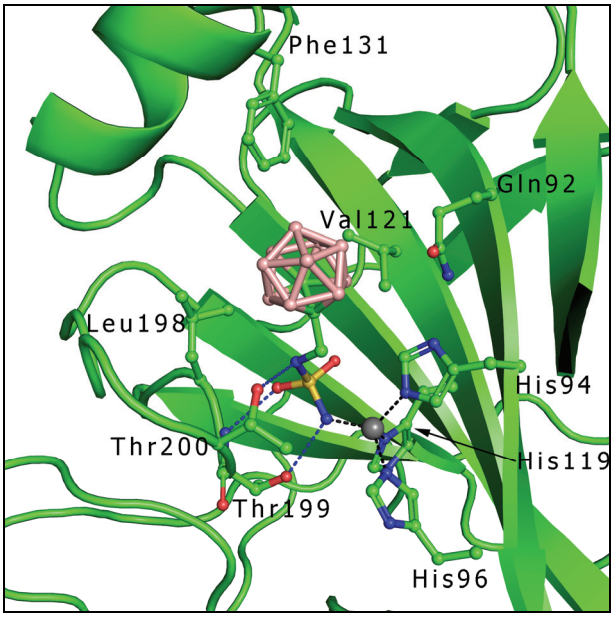


Fig. 3.15: The binding motif of the compound **1a** in CAII. Protein is shown in cartoon representation; residues involved in interactions with the Zn^{2+} ion (gray sphere) and **1a** are shown in stick representation. Polar interactions are represented by blue dashed lines; Zn^{2+} ion coordination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

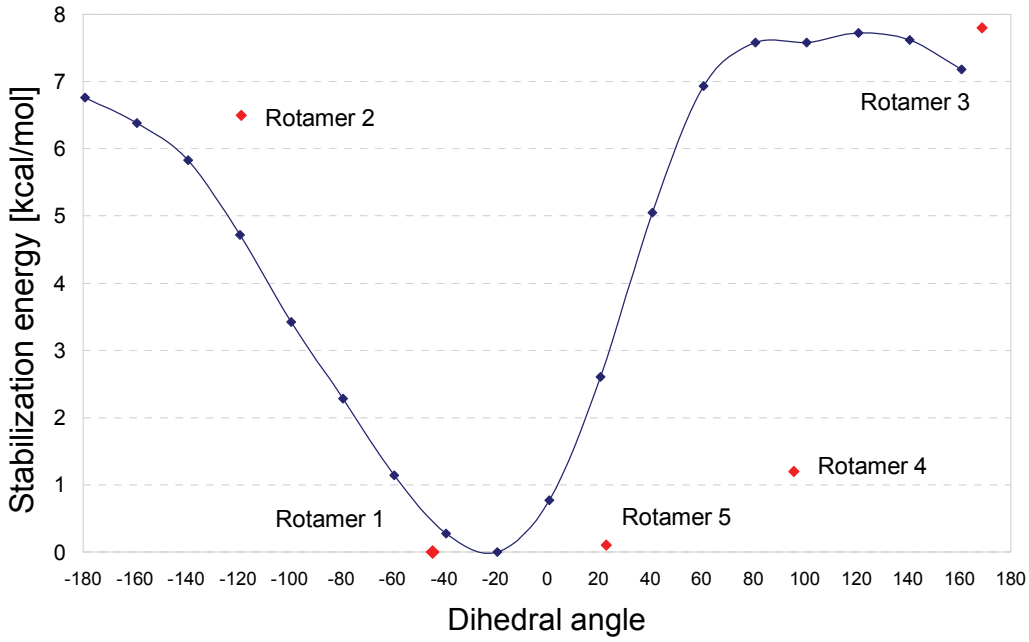


Fig. 3.16: The rotamer preferences of **1a** in the complex with CAII. The rotational profile of isolated inhibitor is shown by blue line, whereas relative stabilization energies of five rotamers in the complex with CAII are shown by red points.

The rigid scan showed that the carborane moiety of isolated **1a** preferred dihedral angle around -20° within the possible range of 80° to +21° 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 kca/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 find in the complex, however they are less populated.

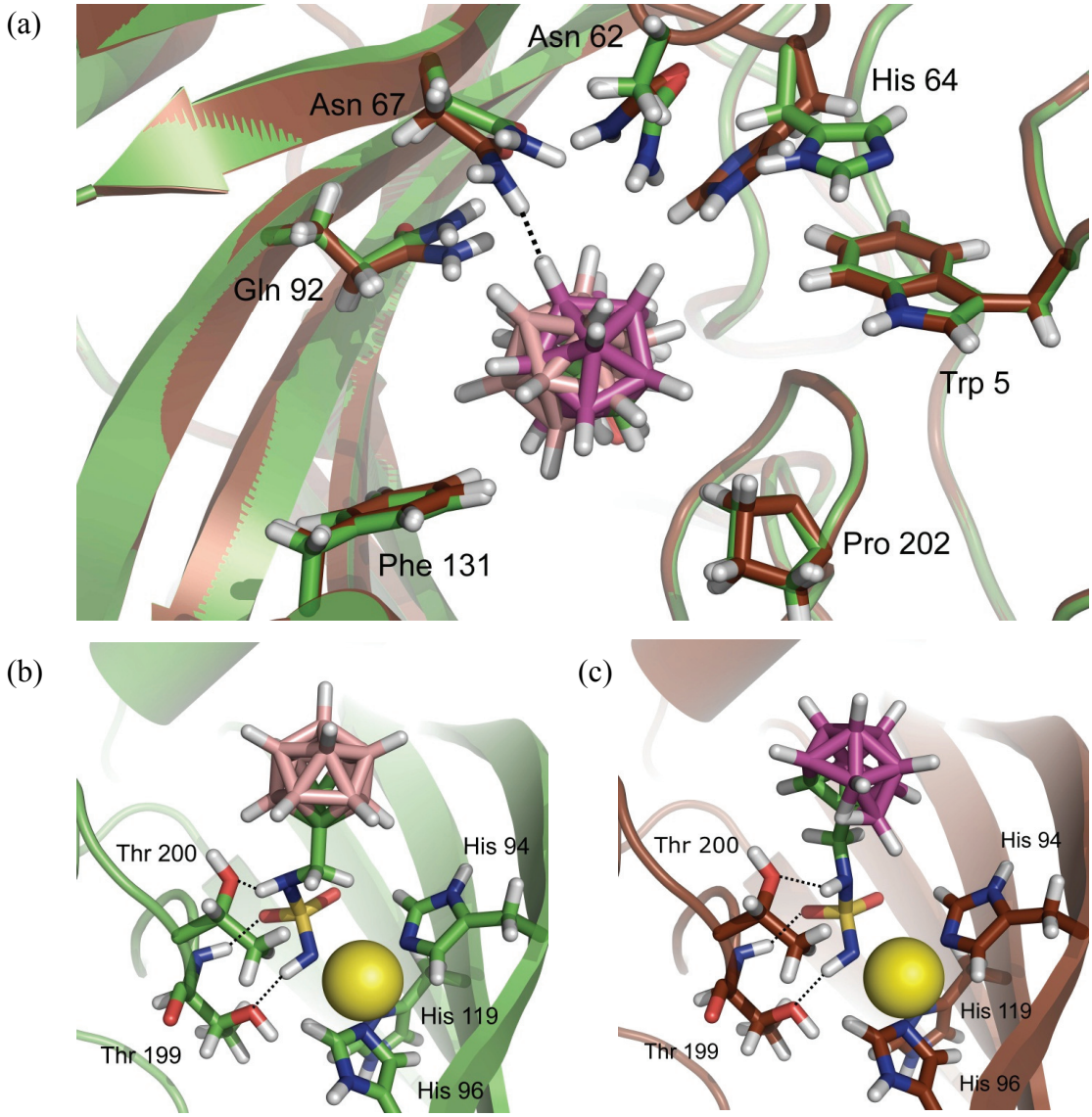


Fig. 3.17: Superposition of the QM/MM optimized complexes of CAII with **1a** and **7a** inhibitors, stressing in **(a)** the interactions of carborane cages (pink for **1a** and magenta for **7a**) with CAII aminoacid sidechains (CAII/**1a** complex in green and CAII/**7a** complex in brown) and the different interactions of sulfamide moiety of **1a (b)** and **7a (c)** where Zn^{2+} 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]

$$RSO_2NH^- + CAII{-}Zn^{2+}{-}OH_2^+ \rightarrow RSO_2NH^- Zn^{2+}{-}CAII + H_2O \qquad \text{(Eq. 3.1)}$$

, 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}	*Disp*	ΔG^o_b
1a *closo*-carborane based linhibitor	-37.7	-184.8	-42.0	-8.4±0.1
7a *nido*-carborane based linhibitor	-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 *closo* cage	-11.2	-22.4	-24.9	
7a *nido* 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 stabe **7a** isomer is shown in Figure 3.18.

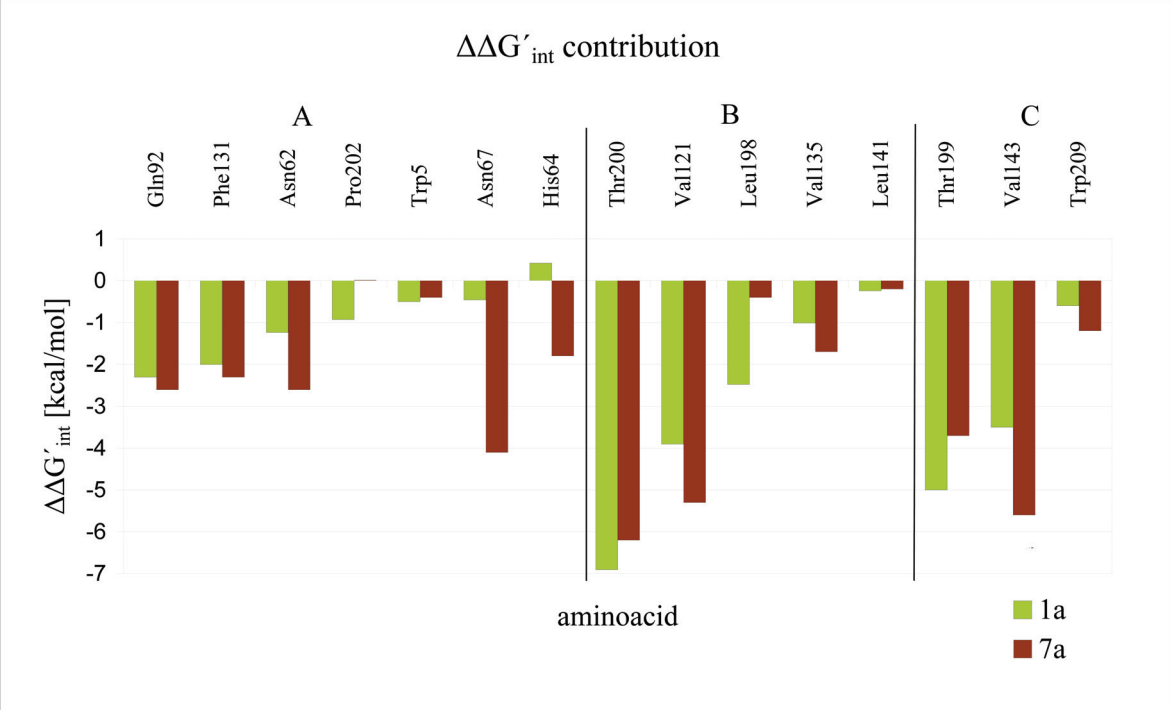


Fig. 3.18: The contribution of single amino acids to the interaction "free" energy $\Delta\Delta G'_{int}$ as obtained from a "virtual glycine scan". **A)** The first 7 amino acids (from Gln92 to His64) interact only with the carborane cage, **B)** the next 5 (from Thr200 to Leu141) have interactions 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 ($B\text{-}H^{\delta-}...O^{\delta-}$) was replaced by a weak hydrogen bond ($C\text{-}H^{\delta+}...O^{\delta-}$) 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 NH_2 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 model[71] using the ff03 AMBER forcefield followed by 10 ps annealing from 150 to 0K using Berendsen thermostat[205] 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.

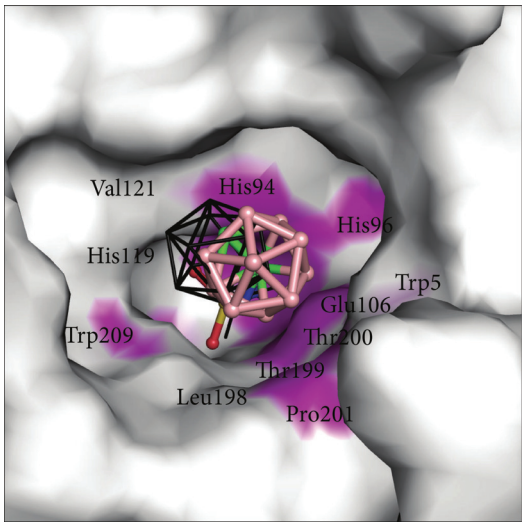


Fig. 3.19: Different binding poses of **1a** inhibitor in CAII (black lines) and CAIX (in pink). Atoms making contacts with **1a** are highlighted in magenta.

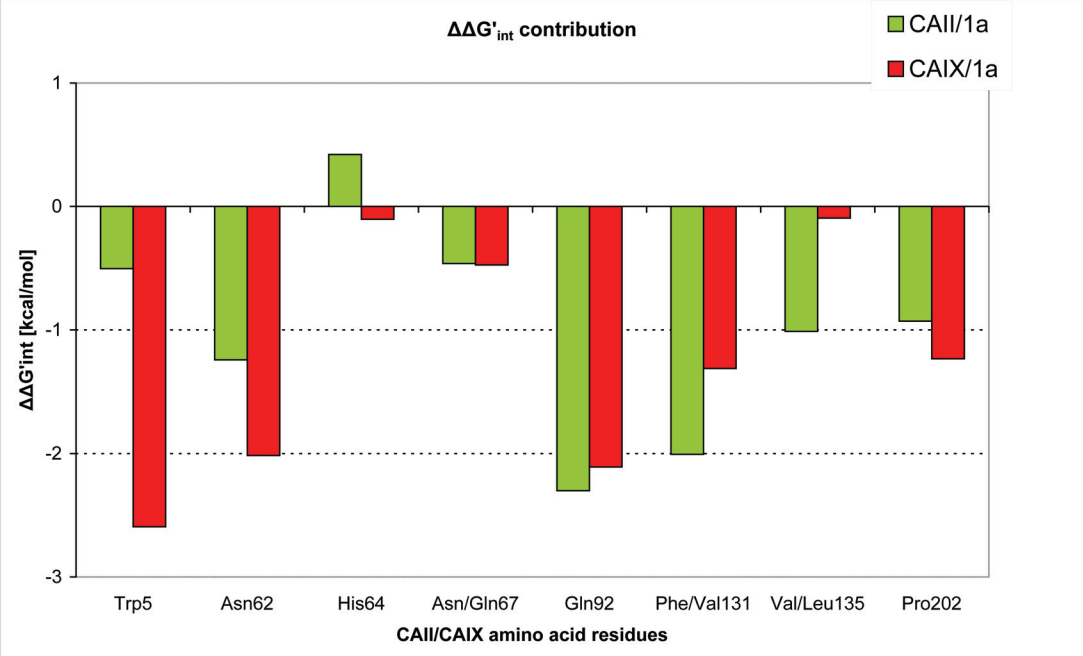


Fig. 3.20: Results of virtual glycine scan showing contributions of individual amino acid residues to the energy of binding of **1a** to CAII and CAIX, respectively.

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 conlcude, 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 strucure-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 poses[253] 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 ligands[219] 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):

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

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 ($\Delta G'^{w}_{conf}$) 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 method[70] 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 uncertainities than that of gas phase interaction energy. The SQM methods used for evaluation of interaction energy provide a reasonably small error bellow 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 computationaly 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 method[70] 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)

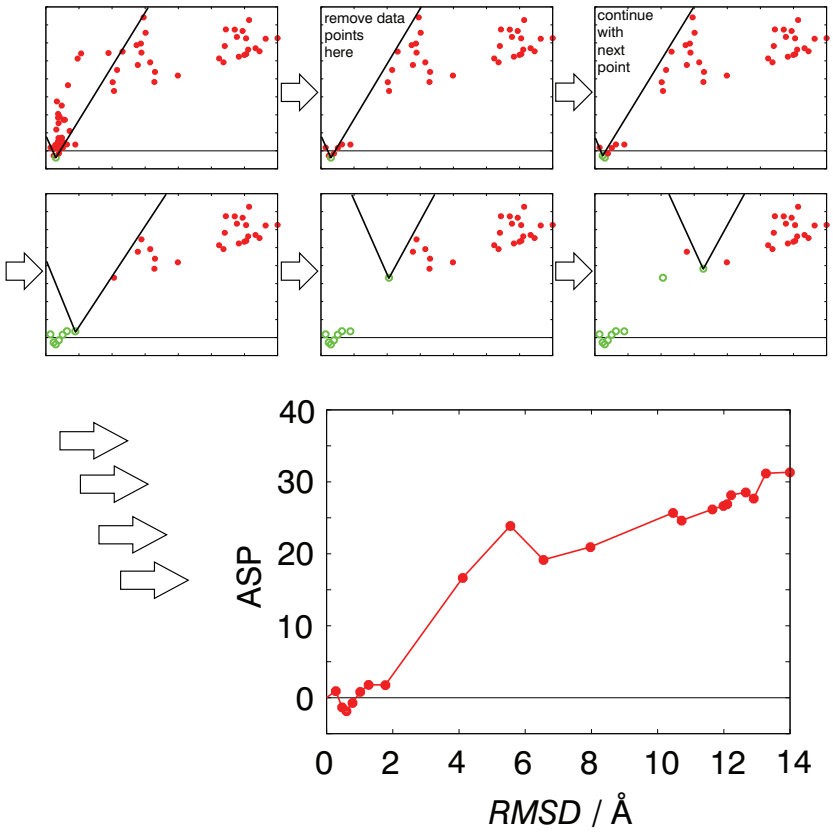


Fig. 3.21: Scheme of the algorithm to create the lower boundary from the whole data set. An iterative process reduces the large amount of data points to the most important points.

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 condicion at superior level unlike the others scoring functions.

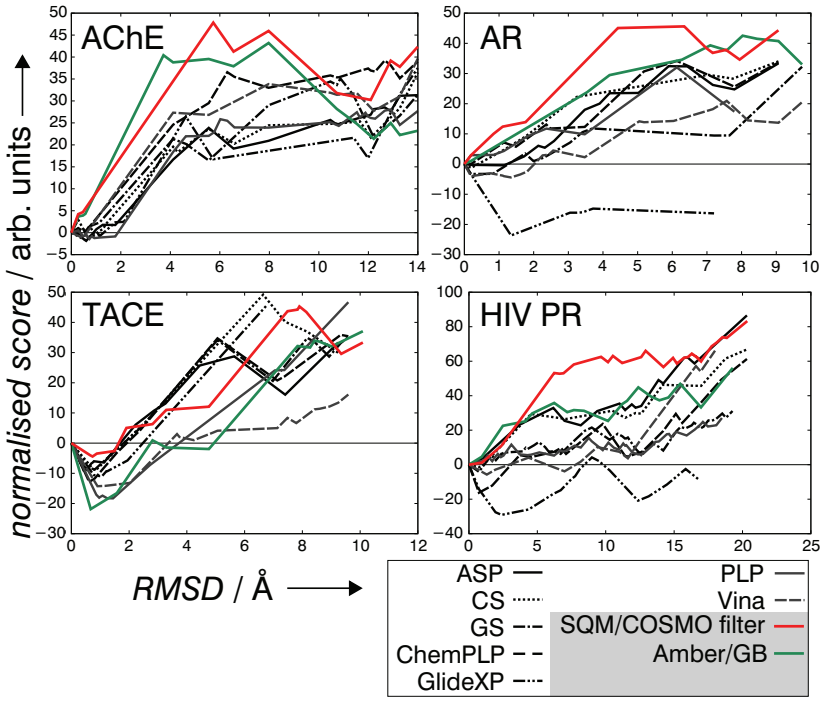


Fig. 3.22: The plots of normalized scores against RMSD values for all four protein-ligand systems.

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								
			Glide	PLANTS	AutoDock	Gold			
	SQM/ COSMO	AMBER /GB	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 behavied 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								
			Glide	PLANTS	AutoDock	Gold			
	SQM/ COSMO	AMBER /GB	XP	PLP	Vina	ASP	CS	GS	Chem PLP
	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 devided 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 hight 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 pnicogen bonded structures, *i.e.* on extended datasets of halogen bonded complexes, innorganic crystal structures of thiaboranes 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 innorganic 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 carborabe-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 of 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.; Pitera, 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.; Brahmkshatriya, 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.; Sponer, 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) Sponer, 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) Hirschfe.Jo; 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) Pryszcz, L. P.; Nemeth, T.; Gacser, A.; Gabaldon, T. *Genome Biol Evol* **2013**, *5*, 2382.

(216) Dostal, J.; Brynda, J.; Hruskova-Heidingsfeldova, O.; Pachl, P.; Pichova, I.; Rezacova, P. *J Enzym Inhib Med Ch* **2012**, *27*, 160.

(217) Dostal, J.; Brynda, J.; Hruskova-Heidingsfeldova, O.; Sieglova, 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.; Fanfrlik, J.; Rezac, J.; Saskova, K. G.; Sieglova, 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

Appendix A

Top Curr Chem (2015) 359: 1–26
DOI: 10.1007/128_2014_606

Published online: 20 March 2015

Characteristics of a σ-Hole and the Nature of a Halogen Bond

Michal H. Kolář, Palanisamy Deepa, Haresh Ajani, Adam Pecina, and Pavel Hobza

Abstract The nature of halogen bonding in 128 complexes was investigated using advanced quantum mechanical calculations. First, isolated halogen donors were studied and their σ-holes were described in terms of size and magnitude. Later, both partners in the complex were considered and their interaction was described in terms of DFT-SAPT decomposition. The whole set of complexes under study was split into two categories on the basis of their stabilisation energy. The first subset with 38 complexes possesses stabilisation energies in the range 7–32 kcal/mol, while the second subset with 90 complexes has stabilisation energies smaller than 7 kcal/mol. The first subset is characterised by small intermolecular distances (less than 2.5 Å) and a significant contraction of van der Waals (vdW) distance (sum of vdW radii). Here the polarisation/electrostatic energy is dominant, mostly followed

The online version of this chapter (doi:10.1007/128_2014_606) contains supplementary material, which is available to authorized users.

M.H. Kolář
Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo nám. 2, 166 10 Prague 6, Czech Republic

Institute for Advanced Simulations (IAS-5), Forschungszentrum Jülich GmbH, 52428 Jülich, Germany

Computational Biophysics, German Research School for Simulation Sciences GmbH, 52428 Jülich, Germany

P. Deepa, H. Ajani, and A. Pecina
Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo nám. 2, 166 10 Prague 6, Czech Republic

P. Hobza (✉)
Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo nám. 2, 166 10 Prague 6, Czech Republic

Department of Physical Chemistry, Regional Centre of Advanced Technologies and Materials, Palacky University, 771 46 Olomouc, Czech Republic
e-mail: pavel.hobza@uochb.cas.cz

by induction and dispersion energies. The importance of induction energy reflects the charge-transfer character of the respective halogen bonds. Intermolecular distances in the second subset are large and the respective contraction of vdW distance upon the formation of a halogen bond is much smaller. Here the dispersion energy is mostly dominant, followed by polarisation and induction energies. Considering the whole set of complexes, we conclude that the characteristic features of their halogen bonds arise from the concerted action of polarisation and dispersion energies and neither of these energies can be considered as dominant. Finally, the magnitude of the σ-hole and DFT-SAPT stabilisation energy correlates only weakly within the whole set of complexes.

Keywords CCSD(T) • DFT-SAPT • Dispersion energy • Electrostatic potential • Halogen bond • Noncovalent interactions • σ-Hole • σ-Hole magnitude • σ-Hole size

Contents

1 Introduction

The family of noncovalent interactions [1] has recently been augmented by a new type of bonding between a Lewis acid and a Lewis base where the Lewis base is an electron donor (O, N, S, P, . . .) and the Lewis acid is an atom which simultaneously contains an area of positive and negative electrostatic potential (ESP). The area of positive ESP, called the σ-hole [2–4], originates in an unequal occupation of valence orbitals. It was originally found on halogens but later also recognised on atoms of groups IV, V and VI. The bonds are referred to as halogen bonds, chalcogen and pnicogen bonds or, in general, σ-hole bonds [5].

For the σ-hole bonds, the σ-hole of the electron acceptor (i.e. Lewis acid) seems to be a key concept, although it concerns only one of the two interacting partners. To elucidate the complete picture of σ-hole bonding, it is inevitable to analyse the contributions to the total stabilisation energy of the entire complex as well. The text below focuses on both the analysis of electron acceptors in halogen bonds and, in more detail, on electron acceptor–donor pairs.

The stabilisation of an X–Y···D halogen bond, where Y is Cl, Br or I, X is an electronegative atom (mostly another halogen) or carbon and D is an electron donor (O, N, S, ...), is explained elegantly by the existence of a positive σ-hole. Energetically, the X–Y···D halogen bond is similar to the X–H···D hydrogen bond (H-bond). Following the reliable CCSD(T)/CBS calculations of the stabilisation energy, the most stable halogen-bonded complex (iodobenzene···trimethylamine) from the X40 dataset [6] (complexes containing halogens) amounts to 5.8 kcal/mol, and this stabilisation energy is comparable to the stabilisation of strong H-bonds. Much larger stabilisation energies were, however, calculated (at the same theoretical level) for complexes of small halogen donors, e.g. 17.1 and 15.3 kcal/mol for $FI\cdots NH_3$ and $FBr\cdots NH_3$, respectively [7]. Similarly, large stabilisation energies (8.0 and 15.0 kcal/mol) were also calculated (again at the same theoretical level) for the crystals of the complexes of the large organic molecules 1,3-dithiole-2-thione-4-carboxylic acid (DTCA) and 1,4-diazabicyclo[2.2.2]octane (DABCO) with diiodine I_2 [8]. Where do these large stabilisation energies come from? Is the nature of stabilisation in these complexes the same as in the previously mentioned ones?

The attraction in halogen-bonded complexes was originally assigned to electrostatic attraction between the positive σ-hole and a lone pair of the electron donor, which is reflected in a recently published IUPAC definition of halogen bond [9]. In our recent paper [10], however, we have pointed out the important role of dispersion interaction, which is easily explained by the fact that, in any halogen bond, two atoms with high polarisability (the halogen and electron donors) are located close to each other (closer than the sum of van der Waals (vdW) radii). In ten different halogen-bonded complexes investigated [10] by the symmetry-adapted perturbation theory (SAPT), the dispersion energy was dominant in eight cases while only in two cases was the electrostatic term slightly larger than the dispersion one. This is in contradiction to the previously mentioned definition of the halogen bond [9], which states that 'the forces involved in the formation of the halogen bond are primarily electrostatic'. Is it because of the fact that the complexes investigated in [10] were not typical halogen-bonded ones? Nevertheless, the list of the complexes studied (benzene···X_2, X=F, Cl, Br; formaldehyde···X, X=chloroform, halothane, enfurane, isofurane; bromomethanol dimer) justifies our choice.

This chapter is organised as follows. First, attention is paid to the characterisation of isolated halogen donors. Their σ-holes are described in terms of size and magnitude. These properties have recently been introduced to characterise such a rather complicated three-dimension object as the σ-hole [11]. In the second part of the chapter, both interacting partners are studied in terms of the SAPT decomposition of their total stabilisation energy. An extended set of SAPT decompositions calculated consistently at the same theoretical level is provided for different types of halogen-bonded complexes: ranging from weak/moderate complexes formed by standard electron donors (e.g. water, ammonia, formaldehyde, dimethyl ether or trimethylammonia) and standard halogen donors (e.g. halobenzenes or substituted halobenzenes) to strong halogen-bonded complexes with a significant charge transfer. An attempt is made to combine approaches to monomers and complexes to provide novel insight into halogen bonding.

2 Methods

2.1 Isolated Subsystems

The *magnitude* of the σ-hole was defined by Kolář et al. [11] as the value of the most positive (or the least negative) electrostatic potential (ESP) localised at the halogen boundary. The most positive ESP had been used previously to characterise σ-holes but the nomenclature was rather confusing. Further, the *size* of the σ-hole was defined as the spatial extent of the region of positive ESP on the halogen boundary. Such spatial characteristics were shown to be indications for attractive interaction: in [11], we concluded that the channel of attraction of the halogen bond, understood as an angular range with a positive total stabilisation energy with either hydrogen fluoride or argon atoms, is well reflected in the size of the σ-hole.

When limited to aromatic molecules with the C_{2v} symmetry point group, the size was initially defined in terms of the angular properties of the ESP profile (see [11] for details). The extension for non-symmetric cases has recently been provided [12]. The size was generalised as an area of positive ESP lying on the boundary of the halogen atom, defined arbitrarily as an isosurface of 0.001 e/bohr [3] electron density [13]. The area has to be refined to have an approximately rounded boundary, since the shape of positive ESP may be quite complicated for non-symmetric molecules [12].

The magnitude and size were calculated for all of the halogenated subsystems. Prior to the ESP calculations, all of the molecules were energy minimised. Both the minimisation and the ESP calculations were done at the PBE0/aug-cc-pVDZ level with the pseudopotentials on bromine and iodine atoms [14–17]. The calculations were performed in the Gaussian09 program package [18].

2.2 SAPT Decomposition

The SAPT method [19] provides an exact decomposition of the total interaction energies into various components of the first and second perturbation order. The DFT version of the SAPT (DFT-SAPT) [20–28] allows for the treatment of extended complexes (up to about 40 atoms) and the total interaction energy is decomposed into polarisation/electrostatic (E^{POL}), induction (E^{I}), dispersion (E^{D}) and exchange-repulsion (E^{ER}) terms. Here, the E^{I} and E^{D} terms include their exchange parts and induction energy further includes the δHF term, which accounts for higher than second-order terms covered by the Hartree–Fock approach. It should be mentioned that SAPT decomposition does not include the charge transfer energy, which is the energy stabilising complexes between electron donor (small ionisation potential) and electron acceptor (small electron affinity). This energy is covered in the induction energy and thus it contains not only the classical induction

energy term (permanent multipole/induced multipole) but also charge transfer (electron donor/electron acceptor) energy.

The greatest improvement of the DFT-SAPT method over the original SAPT is the acceleration of the calculations by one order of magnitude. The intramolecular treatment is conducted using the DFT and therefore suffers from inaccurate energies of the virtual orbitals. This drawback is corrected for in advance of the actual SAPT treatment by a gradient-controlled shift procedure, which uses the difference between the exact vertical ionisation potential (IP) and the energy of the (HOMO) [24]. In this work, PBE0/aug-cc-pVTZ and PBE0/aug-cc-pVDZ calculations were carried out to obtain the IP respective HOMO values and intermolecular terms were described by aug-cc-pVDZ and aug-cc-pVTZ basis sets. Bromine and iodine atoms were treated by pseudopotentials to describe relativistic effects of inner-core electrons correctly.

All the post Hartree–Fock calculations (including DFT-SAPT) were carried out using the Molpro 2010 package [29]. The DFT calculations were done utilising the Turbomole 6.3 package [30].

2.3 Complexes

Our goal was to collect a large set of halogen-bonded complexes of different size and origin. The common feature of all these complexes is the presence of halogen or dihalogen bonds [31, 32]. While in the halogen bond the halogen (Cl, Br or I) covalently bound to an electronegative atom or carbon is in contact with an electron donor (O, N, S, ...); in the case of the dihalogen bond one halogen atom is in contact with another halogen.

First, the complexes where the benchmark CCSD(T) stabilisation energies are known were utilised; in all these studies the complex geometry was determined at a lower theoretical level, mostly at DFT with an empirical dispersion correction [33] (DFT-D). We studied 18 complexes from our X40 dataset [6] (Table 2), 46 complexes from the XB51 dataset [7] (Table 3), 11 complexes from our previous papers [34–36] (Table 4) and 13 complexes from [37] (Table 5). Second, in the following halogen-bonded complexes, the stabilisation energy as well as the complex geometry were calculated at MP2 or DFT-D levels. Table 6 summarises eight complexes [38] of crystal motifs which were taken from the Cambridge Structure Database. Table 7 contains 15 complexes from [39] for which the binding free energy in nonpolar solvent was measured. Finally, Table 8 contains 17 structures of organic crystals, taken from [40–45]. Altogether, 128 halogen-bonded complexes were investigated. Structures of all investigated complexes are collected in the Electronic Supplementary Material Figs. S1, S2, S3, S4, S5, S6 and S7.

The structure of each of the halogen-bonded complexes was taken from the original references without any additional optimisation. For most of the complexes, the DFT-D (B97-D3/def2-QZVP) calculations [46] were also performed. All

interaction energies were corrected for the basis set superposition error (BSSE) utilising counterpoise correction [47].

As mentioned above, DFT-SAPT calculations were performed using the aug-cc-pVDZ and aug-cc-pVTZ basis sets. When passing to the larger basis set, all the SAPT energy terms remain practically unchanged with the exception of dispersion energy, which is underestimated with the smaller basis. This ratio was evaluated for 18 complexes from the X40 dataset and was used for scaling the aug-cc-pVDZ dispersion energy of the remaining complexes for which the DFT-SAPT/aug-cc-pVTZ calculation would be prohibitively expensive. For even larger crystal structures, for which the SAPT/aug-cc-pVDZ calculations of dispersion energy would be impractical, the dispersion and exchange-dispersion terms were approximated by an empirical atom-atom dumped dispersion term [48]. The ratio of the empirical dispersion energy and aug-cc-pVTZ perturbation dispersion energy evaluated again for 18 complexes from the X40 dataset was used for scaling the empirical dispersion energy for extended halogen-bonded complexes.

3 Results and Discussion

3.1 Halogenated Molecules

The properties of the subsystems, the magnitude and size of the σ-hole and the energy of the lowest unoccupied molecular orbital (LUMO) are shown in Table 1. The magnitude and size correlate well, with the correlation coefficient R being 0.86. This agrees with the previously presented dependence. Furthermore, the magnitude and size both increase with the atomic number of the halogen atom, which is also a well-known trend. All the molecules possess a positive σ-hole with the exception of H_3CCl, which has a slightly negative V_{max} of −0.0001 a.u. The most positive σ-hole can be found in FI, where two effects are combined, both increasing the magnitude of the σ-hole (activating the halogen for the halogen bond). These effects are the presence of a heavy halogen atom along with a strong electron withdrawing chemical group in its vicinity. Indeed, a comparison of, e.g. H_3CBr with F_3CBr or BzI with C_6F_5I reveals that the presence of fluorine atoms increases both the magnitude and the size of the σ-hole on iodine [2, 49]. In dihalogen molecules, the activation of the halogen participating in a halogen bond increases with the decreasing atomic number of the second halogen (iodine σ-hole magnitude IBr<ICl<IF). Hence, the fluorine has a positive σ-hole with a size of about 6 Å^2 when bound to another fluorine. The magnitude and size of its σ-hole are comparable with, e.g. H_3CI.

The magnitude and LUMO energy anticorrelate with $R = -0.76$. This means that strong electron acceptors (i.e. molecules with the most negative LUMO energy) have more positive σ-holes.

Table 1 The magnitude (in a.u.) and size (in Å^2) of the σ-holes of halogenated monomers and the energies of the lowest unoccupied molecular orbital (LUMO) (in a.u.)

Molecule	Magnitude	Size	LUMO
F_2	0.025	6.0	−0.134
Cl_2	0.042	10.7	−0.128
ClF	0.062	13.7	−0.127
ClF_3	0.069	13.6	−0.162
H_3CCl	0.000	0.0	−0.006
F_3CCl	0.032	12.3	−0.009
C_2H_3Cl	0.008	2.5	−0.016
C_2HCl	0.034	11.4	−0.001
C_6H_5Cl	0.007	2.1	−0.025
C_6Cl_6	0.026	8.6	−0.062
$C_6H_2OHCl_3$	0.018	5.8	−0.045
Br_2	0.052	12.5	−0.140
BrF	0.083	14.4	−0.136
BrF_3	0.090	15.6	−0.163
H_3CBr	0.013	3.3	−0.017
F_3CBr	0.042	14.2	−0.042
C_2H_3Br	0.020	5.6	−0.018
C_2HBr	0.049	13.7	−0.020
C_6H_5Br	0.019	5.2	−0.025
C_6Br_6	0.036	10.5	−0.085
$BrC_4H_2NO_2$	0.055	11.7	−0.114
CH_2BrOH	0.013	3.4	−0.019
$C_7F_4O_2HBr$	0.023	8.0	−0.064
I_2	0.056	9.7	−0.144
IF	0.097	17.3	−0.140
ICl	0.074	17.1	−0.140
IBr	0.066	16.5	−0.142
ICN	0.081	16.7	−0.077
H_3CI	0.022	6.5	−0.038
F_3CI	0.050	17.6	−0.073
C_2H_3I	0.028	8.6	−0.031
C_2HI	0.058	16.5	−0.043
C_6H_5I	0.027	8.1	−0.034
C_6F_5I	0.052	17.0	−0.071
C_4F_9I	0.050	19.2	−0.081
$INC_4H_2O_2$	0.068	15.0	−0.111
$HO_2C_7F_4I$	0.053	19.2	−0.082
TFIB	0.051	16.2	−0.074

3.2 Complexes

Tables 2, 3, 4, 5, 6, 7 and 8 summarise the energy characteristics of all complexes investigated and also show the Y···D and Δr distances, i.e. the distance between

Table 2 DFT-SAPT interaction energies (in kcal/mol) and Y···D/Δ*r* distances (Å) for halogen-bonded complexes from the X40 dataset [6]

No.	Basis	Complex	CCSDT/CBS	DFT-SAPT						Y···D/Δ*r*[a]	P:I:D[b]	Q[c]	Q[d]
			Δ*E*	E_{tot}	E_1^{Pol}	E_1^{Ex}	E_2^{Ind}	E_2^{Disp}	$E_2^{Disp.\ empirical}$				
1	aVDZ	$H_3CCl\cdots OCH_2$	−1.17	−0.79	−1.13	2.34	−0.36	−1.64	−1.38	3.30/0.03	0.7:0.2:1	1.18	1.40
	aVTZ			−1.07	−1.11	2.36	−0.38	−1.94			0.6:0.2:1		
2	aVDZ	$H_3CBr\cdots OCH_2$	−1.72	−1.44	−2.26	3.39	−0.48	−2.09	−1.94	3.17/0.2	1.1:0.2:1	1.19	1.28
	aVTZ			−1.79	−2.19	3.40	−0.50	−2.49			0.9:0.2:1		
3	aVDZ	$H_3CI\cdots OCH_2$	−2.38	−2.18	−3.66	4.78	−0.88	−2.41	−2.48	3.21/0.29	1.5:0.4:1	1.23	1.19
	aVTZ			−2.63	−3.53	4.79	−0.94	−2.96			1.2:0.3:1		
4	aVDZ	$F_3CCl\cdots OCH_2$	−2.25	−1.60	−2.92	3.99	−0.71	−1.97	−1.61	2.98/0.29	1.5:0.4:1	1.17	1.44
	aVTZ			−1.87	−2.83	3.99	−0.72	−2.31			1.2:0.3:1		
5	aVDZ	$F_3CBr\cdots OCH_2$	−3.1	−2.75	−4.66	5.29	−0.98	−2.41	−2.21	2.95/0.42	1.9:0.4:1	1.20	1.31
	aVTZ			−3.09	−4.50	5.27	−0.97	−2.88			1.6:0.3:1		
6	aVDZ	$F_3CI\cdots OCH_2$	−4.08	−3.86	−6.65	7.39	−1.78	−2.82	−2.89	3.01/0.49	2.4:0.6:1	1.23	1.20
	aVTZ			−4.35	−6.43	7.38	−1.83	−3.47			1.9:0.5:1		
7	aVDZ	$C_6H_5Cl\cdots OC_3H_6$	−1.49	−0.93	−1.55	3.64	−0.60	−2.42	−1.93	3.12/0.15	0.6:0.2:1	1.12	1.41
	aVTZ			−1.18	−1.49	3.62	−0.60	−2.71			0.5:0.2:1		
8	aVDZ	$C_6H_5Br\cdots OC_3H_6$	−2.43	−2.32	−3.39	5.03	−0.88	−3.08	−2.71	3.07/0.3	1.1:0.3:1	1.14	1.13
	aVTZ			−2.65	−3.29	5.02	−0.89	−3.50			0.9:0.3:1		
9	aVDZ	$C_6H_5I\cdots OC_3H_6$	−3.46	−3.71	−5.44	7.07	−1.80	−3.54	−3.44	3.09/0.41	1.5:0.5:1	1.18	1.03
	aVTZ			−4.18	−5.27	7.04	−1.82	−4.14			1.3:0.4:1		
10	aVDZ	$C_6H_5Cl\cdots NC_3H_9$	−2.11	−1.37	−3.37	6.27	−1.03	−3.24	−2.72	3.06/0.24	1.0:0.3:1	1.11	1.33
	aVTZ			−1.68	−3.26	6.21	−1.03	−3.61			0.9:0.3:1		
11	aVDZ	$C_6H_5Br\cdots NC_3H_9$	−3.78	−3.78	−7.47	9.69	−1.33	−4.67	−4.45	2.97/0.43	1.6:0.3:1	1.13	1.19
	aVTZ			−4.26	−7.27	9.58	−1.30	−5.28			1.4:1:1		
12	aVDZ	$C_6H_5I\cdots NC_3H_9$	−5.81	−6.74	−12.91	15.72	−3.39	−6.16	−6.61	2.97/0.56	2.0:0.6:1	1.06	0.98
	aVTZ			−7.02	−12.76	15.65	−3.42	−6.50			2.0:0.5:1		

13	aVDZ	$C_6H_5Br\cdots SHCH_3$	−2.32	2.00	−3.20	4.63	−0.34	−3.09	−2.62	3.54/0.11	1.0:0.1:1	1.15	1.36
	aVTZ			−2.41	−3.12	4.60	−0.33	−3.56			0.9:0.1:1		
14	aVDZ	$C_6H_5I\cdots SHCH_3$	−3.08	−2.70	−5.06	6.55	−0.57	−3.63	−3.37	3.56/0.22	1.4:0.2:1	1.05	1.13
	aVTZ			−2.83	−4.99	6.54	−0.57	−3.81			1.3:0.1:1		
15	aVDZ	$H_3CBr\cdots Bz$	−1.81	−1.47	−1.21	3.14	−0.13	−3.28	−3.41	3.55	0.4:0.04:1	1.13	1.09
	aVTZ			−1.88	−1.14	3.09	−0.12	−3.7			0.0:0.03:1		
16	aVDZ	$H_3CI\cdots Bz$	−2.48	−2.04	−1.99	3.89	−0.15	−3.79	−4.45	3.66	0.5:0.04:1	1.18	1.00
	aVTZ			−2.31	−2.21	4.55	−0.19	−4.46			0.5:0.04:1		
17	aVDZ	$F_3CBr\cdots Bz$	−3.11	−2.74	−2.60	4.09	−0.46	−3.78	−4.14	3.46	0.7:0.1:1	1.13	1.03
	aVTZ			−3.12	−2.60	4.16	−0.44	−4.27			0.6:0.1:1		
18	aVDZ	$F_3CI\cdots Bz$	−3.91	−3.29	−3.68	5.45	−0.63	−4.43	−5.38	3.53	0.8:0.1:1	1.16	0.95
	aVTZ			−4.78	−3.61	5.53	−1.59	−5.12			0.7:0.3:1		

[a]Δr is the difference of the distance between halogen and electron donor (Y···D) and the sum of the respective vdW radii
[b]Ratio of polarisation, induction and dispersion energies
[c]Ratio of dispersion energies determined with aug-cc-pVTZ and aug-cc-pVDZ basis sets
[d]Ratio of dispersion energies determined with SAPT/aug-cc-pVTZ and empirical dispersion energy

Table 3 DFT-SAPT/aVTZ-PP interaction energies (in kcal/mol) and Y···D (Å) for the XB51 dataset [7]

No.	Complex	CCSDT/CBS	DFT-SAPT						
		ΔE	E_{tot}	E_1^{Pol}	E_1^{Ex}	E_2^{Ind}	E_2^{Disp}	Y···D/Δr^a	P:I:D[b]
1	HCN···ICF_3	3.61	−3.78	−5.49	5.84	−1.38	−2.76	3.1/0.43	2.0:0.5:1
2	HCN···BrF	7.53	−8.24	−14.76	17.22	−5.41	−5.30	2.52/0.88	2.8:1.0:1
3	HCN···ClF	4.81	−4.08	−8.30	11.54	−3.62	−3.70	2.61/0.69	2.2:1.0:1
4	HCN···$BrO_2C_4H_2N$	4.32	−4.28	−6.76	7.60	−1.76	−3.35	2.81/0.59	2.0:0.5:1
5	HCN···I $NC_4H_2O_2$	5.91	−6.39	−9.78	10.82	−3.21	−4.22	2.86/0.67	2.3:0.8:1
6	HCN···BrC_6H_5	1.15	−1.10	−1.54	2.76	−0.47	−1.85	3.18/0.22	0.8:0.3:1
7	HCN···IC_6H_5	1.87	−1.93	−2.73	3.88	−0.81	−2.28	3.25/0.28	1.2:0.4:1
8	H_3N···ICF_3	5.88	−6.94	−12.03	12.06	−2.90	−4.07	2.99/0.54	3.0:0.7:1
9	H_3N···BrF	15.30	−19.60	−39.14	43.48	−15.20	−8.74	2.35/1.05	4.5:1.7:1
10	H_3N···ClF	10.54	−9.59	−28.31	42.61	−16.49	−7.40	2.34/0.96	3.8:2.2:1
11	H_3N···$BrC_4H_2NO_2$	8.02	−8.91	−17.52	18.54	−4.56	−5.36	2.66/0.74	3.3:0.9:1
12	H_3N···$IC_4H_2NO_2$	10.99	−14.25	−25.37	26.16	−8.30	−6.74	2.69/0.84	3.8:1.2:1
13	H_3N···BrC_6H_5	2.02	−2.13	−3.92	5.02	−0.74	−2.46	3.13/0.27	1.6:0.3:1
14	H_3N···IC_6H_5	3.33	−3.79	−6.53	7.34	−1.42	−3.18	3.17/0.36	2.1:0.4:1
15	HCP···ICF_3	0.89	−0.83	−0.85	2.45	−0.34	−2.09	3.72/0.06	0.4:1.7:1
16	HCP···BrF	2.07	−2.24	−3.25	7.62	−2.50	−4.11	3.08/0.57	0.8:0.6:1
17	HCP···ClF	1.16	−0.86	−1.71	5.57	−1.93	−2.79	3.16/0.39	0.6:0.7:1
18	HCP···$BrC_4H_2O_2N$	1.19	−1.04	−1.20	3.21	−0.53	−2.52	3.41/0.24	0.5:0.2:1
19	HCP···$IC_4H_2O_2N$	1.53	−1.43	−1.58	4.12	−0.94	−3.03	3.49/0.29	0.5:0.3:1
20	HCP···BrC_6H_5	0.85	−0.78	−0.82	2.02	−0.08	−1.90	3.63/0.02	0.4:0.04:1
21	HCP···IC_6H_5	0.92	−0.89	−0.94	2.36	−0.16	−2.15	3.76/0.02	0.4:0.07:1
22	Br_2···FC_2H	0.74	−0.51	−0.28	0.88	−0.12	−0.99	3.19/0.01	0.3:0.1:1
23	Br_2···FCH_3	3.61	−2.62	−4.28	5.58	−1.07	−2.85	2.81/0.39	1.5:0.4:1
24	Br_2···NCH	2.87	−3.68	−5.79	6.80	−1.65	−3.04	2.86/0.54	1.9:0.5:1

25	$Br_2\cdots NH_3$	5.95	−8.45	−16.83	18.97	−5.25	−5.34	2.66/0.74	3.1:1:1
26	$Br_2\cdots OCH_2$	4.41	−4.22	−7.48	9.40	−2.27	−3.87	2.75/0.62	1.9:0.6:1
27	$Br_2\cdots OPH_3$	7.29	−6.09	−10.76	13.04	−3.44	−4.93	2.68/0.69	2.2:0.7:1
28	$Br_2\cdots PCH$	1.18	−1.06	−1.19	3.07	−0.59	−2.35	3.44/0.21	0.5:0.3:1
29	$Br_2\cdots NC_5H_5$	9.00	−10.89	−19.97	24.15	−7.60	−7.47	2.57/0.83	2.7:1.0:1
30	$FI\cdots FC_2H$	0.29	−0.29	−0.02	0.28	−0.03	−0.52	3.02/0.31	0.04:0.06:1
31	$FI\cdots FCH_3$	9.33	−6.29	−9.95	11.60	−3.72	−4.22	2.66/0.67	2.4:0.9:1
32	$FI\cdots NCH$	5.97	−11.24	−18.37	20.97	−7.72	−6.13	2.61/0.92	3.0:1.3:1
33	$FI\cdots NH_3$	13.36	−23.82	−40.62	41.56	−16.07	−8.68	2.5/1.03	4.7:1.9:1
34	$FI\cdots OCH_2$	9.94	−11.66	−18.81	21.74	−8.43	−6.15	2.57/0.93	3.1:1.4:1
35	$FI\cdots OPH_3$	17.11	−16.13	−27.19	30.47	−11.39	−8.02	2.5/1.0	3.4:1.4:1
36	$FI\cdots PCH$	2.74	−2.68	−4.13	9.50	−3.39	−4.66	3.17/0.61	0.9:0.7:1
37	$FI\cdots NC_6H_5$	17.66	−31.50	−47.19	51.80	−24.25	−11.85	2.42/1.11	3.9:2.0:1
38	$H_3CI\cdots FC_2H$	0.50	−0.38	−0.30	1.05	−0.09	−1.04	3.39/0.06	0.3:0.09:1
39	$H_3CI\cdots FCH_3$	1.70	−1.69	−2.50	3.86	−0.59	−2.46	3.18/0.15	1.0:0.2:1
40	$H_3CI\cdots LiH$	3.62	−3.50	−8.13	9.06	−0.54	−3.89	2.74/1.06	2.1:0.1:1
41	$H_3CI\cdots NCH$	1.42	−1.60	−2.28	3.44	−0.68	−2.08	3.28/0.25	1.1:0.3:1
42	$H_2CI\cdots NH_3$	2.73	−3.34	−5.72	6.48	−1.19	−2.91	3.21/0.32	1.9:0.4:1
43	$H_3CI\cdots OCH_2$	2.39	−2.48	−3.83	5.35	−0.98	−3.03	3.17/0.33	1.3:0.3:1
44	$H_3CI\cdots OPH_3$	3.34	−3.57	−5.63	7.56	−1.50	−4.00	3.14/0.36	1.4:0.4:1
45	$H_3CI\cdots PCH$	0.85	−0.84	−0.91	2.23	−0.15	−2.01	3.77/0.01	0.5:0.08:1
46	$H_3CI\cdots NC_5H_5$	3.61	−5.01	−6.83	8.32	−2.23	−4.28	3.11/0.42	1.6:0.5:1

[a]Δr is the difference of the distance between halogen and electron donor (Y···D) and the sum of the respective vdW radii
[b]The ratio of polarisation, induction and dispersion energies

Table 4 DFT-SAPT/aug-cc-pVTZ interaction energies (in kcal/mol) and Y···D (Å) for halogen-bonded complexes

No.	Complex	CCSDT/CBS	DFT-SAPT					Y···D/Δr[a]	P:I:D[b]	Ref.[c]
		ΔE	E_{tot}	E_1^{Pol}	E_1^{Ex}	E_2^{Ind}	E_2^{Disp}			
1	F_2···Bz	−1.19	−1.22	−0.96	2.08	−0.49	−1.77	–	0.5:0.3:1	[34]
2	Cl_2···Bz	−2.86	−3.14	−3.17	6.52	−1.73	−4.53	–	0.7:0.4:1	[34]
3	Br_2···Bz	−3.66	−4.23	−5.01	9.83	−2.87	−5.85	–	0.9:0.5:1	[34]
4	CH_2BrOH···CH_2BrOH(Br–O)	−1.48	−1.56	−2.48	4.30	−0.50	−2.88	3.1/0.27	0.9:0.2:1	[35]
5	CH_2BrOH···CH_2BrOH(Br–Br)	−1.22	−1.44	−0.98	1.94	−0.31	−2.09	3.8/0.1	0.5:0.1:1	[35]
6	Trimethylbenzene···Br_2	−4.23	−4.89	−4.04	7.31	−1.50	−6.34	–	0.6:0.2:1	[34]
7	Hexamethylbenzene···Br_2	−5.66	−6.06	−5.43	9.60	−2.07	−7.79	–	0.8:0.3:1	[34]
8	I_2···I_2	−2.95	−3.44	−3.61	5.58	−1.09	−4.32	3.7/0.26	0.8:0.3:1	[36]
9	Br_2···Br_2	−2.28	−2.47	−2.53	4.27	−0.76	−3.45	3.4/0.3	0.7:0.2:1	[36]
10	Cl_2···Cl_2	−1.33	−1.20	−1.27	3.02	−0.63	−2.32	3.3/0.2	0.5:0.3:1	[36]
11	F_2···F_2	−0.39	−0.33	−0.17	0.46	−0.04	−0.58	2.9/0.2	0.3:0.1:1	[36]

[a]Δr is the difference of the distance between halogen and electron donor (Y···D) and the sum of the respective vdW radii
[b]Ratio of polarisation, induction and dispersion energies
[c]Structures from [34–36]

Table 5 DFT-SAPT/aVTZ-PP interaction energies (in kcal/mol) and Y···D (Å) for halogen-bonded complexes [37]

No.	Complex	CCSD(T)/CBS	DFT-SAPT						
		ΔE	E_{tot}	E_1^{Pol}	E_1^{Ex}	E_2^{Ind}	E_2^{Disp}	Y···D/Δr[a]	P:I:D[b]
1	$C_2H_3Cl\cdots OCH_2$	−1.49	−0.98	−0.93	1.96	−0.30	−1.70	3.53/0.26	0.5:0.1:1
2	$C_2HCl\cdots OH_2$	−2.08	−1.97	−2.25	2.61	−0.49	−1.84	3.07/0.2	1.1:0.3:1
3	$C_2HCl\cdots OCH_2$	−2.34	−1.69	−2.07	2.18	−0.42	−1.39	3.11/0.16	1.5:0.3:1
4	$C_2H_3Cl\cdots NH_3$	−0.89	−0.78	−1.04	2.10	−0.40	−1.44	3.33/0.03	0.7:0.3:1
5	$C_2HCl\cdots NH_3$	−2.65	−1.80	−2.68	3.16	−0.58	−1.71	3.20/0.1	1.6:0.3:1
6	$C_2H_3Br\cdots OH_2$	−1.36	−1.05	−0.85	1.19	−0.22	−1.17	3.36/0.01	0.7:0.2:1
7	$C_2H_3Br\cdots OCH_2$	−2.16	−1.04	−0.66	1.08	−0.25	−1.21	3.36/0.01	0.5:0.2:1
8	$C_2HBr\cdots OH_2$	−3.0	−3.68	−4.60	5.0	−1.18	−2.90	2.96/0.41	1.6:0.4:1
9	$C_2HBr\cdots OCH_2$	−3.37	−2.93	−4.17	4.11	−0.67	−2.2	2.99/0.38	1.9:0.3:1
10	$C_2H_3Br\cdots NH_3$	−1.87	−2.09	−2.84	3.34	−0.52	−2.13	3.24/0.16	1.3:0.2:1
11	$C_2HBr\cdots NH_3$	−4.12	−3.74	−5.14	4.71	−0.76	−2.55	3.10/0.3	2.0:0.3:1
12	$C_2H_3I\cdots OH_2$	−2.51	−1.56	−1.79	2.29	−0.41	−1.64	3.15/0.35	1.1:0.3:1
13	$C_2HI\cdots OH_2$	−4.38	−4.27	−6.15	5.80	−1.27	−2.66	3.06/0.44	2.3:0.5:1

[a]Δr is the difference of the distance between halogen and electron donor (Y···D) and the sum of the respective vdW radii
[b]Ratio of polarisation, induction and dispersion energies

Table 6 MP2/aug-cc-pVDZ, DFT-SAPT/aVTZ-PP interaction energies (in kcal/mol) and Y···D (Å) for crystal motifs [38]

No.	Complex	MP2/aug-cc-pVDZ	DFT-SAPT					Y···D/Δr[a]	P:I:D[b]
		ΔE	E_{tot}	E_1^{Pol}	E_1^{Ex}	E_2^{Ind}	E_2^{Disp}		
1	$CH_3CN\cdots BrF$	−8.85	−11.33	−19.05	22.22	−7.94	−6.56	2.46/0.94	2.9:1.2:1
2	$CH_3CN\cdots ClF$	−6.13	−5.62	−11.05	15.38	−5.44	−4.5	2.54/0.76	2.5:1.1:1
3	$CH_3CN\cdots BrF_3$	−9.30	−10.89	−16.45	17.10	−6.07	−5.46	2.61/0.79	3.0:1.1:1
4	$CH_3CN\cdots ClF_3$	−7.52	−6.31	−12.45	16.43	−5.47	−4.82	2.58/0.72	2.6:1.1:1
5	$CO\cdots BrF_3$	−1.25	−1.92	−2.09	2.91	−0.85	−1.88	2.97/0.4	1.1:0.5:1
6	$CO\cdots ClF_3$	−1.19	−1.30	−1.73	2.91	−0.69	−1.78	2.91/0.36	1.0:0.4:1
7	$CO\cdots BrF$	−1.27	−1.85	−1.97	2.81	−0.78	−1.91	2.90/0.47	1.0:0.4:1
8	$CO\cdots ClF$	−1.02	−1.14	−1.42	2.39	−0.57	−1.54	2.88/0.39	0.9:0.4:1

[a]Δr is the difference of the distance between halogen and electron donor (Y···D) and the sum of the respective vdW radii
[b]Ratio of polarisation, induction and dispersion energies

Table 7 DFT-SAPT/aVDZ-PP interaction energies and Y···D (Å) for halogen-bonded complexes [39]

No.	Complex	DFT-D[a]	DFT-SAPT							
		ΔE	E_{tot}	E_1^{Pol}	E_1^{Ex}	E_2^{Ind}	E_2^{Disp}	Y···D/Δr[b]	P:I:D[c]	E_{tot}[d]
1	ICN···NC_5H_5		−13.57	−20.73	20.73	−7.73	−6.72	2.77/0.76	3.1:1.2:1	−14.45
2	IBr···NC_5H_5		−21.32	−37.45	43.80	−18.68	−10.34	2.53/1.0	3.6:1.8:1	−22.67
3	ICl···NC_5H_5		−24.33	−41.08	46.84	−20.78	−10.70	2.49/1.04	3.8:1.9:1	−25.72
4	I_2···NC_5H_5		−16.99	−31.76	38.29	−15.56	−9.61	2.59/0.94	3.3:1.6:1	−18.64
5	I_2···NC_7H_{13}	−18.74	−25.44	−50.02	60.53	−23.30	−14.54	2.48/1.05	3.4:1.6:1	−27.33
6	I_2···OSC_2H_6	−9.39	−12.57	−22.44	27.22	−10.22	−8.20	2.62/0.88	2.7:1.2:1	−13.64
7	I_2···OPC_3H_9	−9.34	−13.49	−23.18	26.89	−10.24	−8.00	2.62/0.88	2.9:1.3:1	−14.53
8	I_2···NC_6H_{15}	−18.75	−25.20	−46.63	57.85	−22.21	−16.34	2.55/0.95	2.9:1.1:1	−27.33
9	C_4F_9I···OSC_2H_6	−5.01	−9.24	−14.44	15.97	−5.27	−6.38	3.75/0.25	2.3:0.8:1	−10.12
10	C_4F_9I···NC_6H_{15}	−11.14	−14.31	−24.34	29.43	−9.11	−11.83	2.81/0.72	2.4:0.9:1	−15.85
11	C_4F_9I···NC_7H_{13}	−10.05	−15.43	−29.12	33.80	−10.73	−10.79	2.69/0.84	2.7:1.0:1	−16.84
12	C_4F_9I···OPC_3H_9	−5.67	−10.30	−15.46	16.08	−5.53	−6.21	3.56/0.06	2.5:0.9:1	−11.12
13	C_6F_5I···OSC_2H_6	−4.93	−9.31	−14.14	15.44	−5.03	−6.42	2.80/0.7	2.2:0.8:1	−10.15
14	C_6F_5I···NC_6H_{15}	−10.91	−14.35	−23.75	28.44	−8.64	−11.96	2.81/0.72	2.3:0.8:1	−15.91
15	C_6F_5I···NC_7H_{13}	−10.79	−14.85	−27.82	31.94	−9.73	−10.63	2.70/0.83	2.6:0.9:1	−16.24

[a]DFT-D calculations at B97-D3/def2-QZVP level

[b]Δr is the difference of the distance between halogen and electron donor (Y···D) and the sum of the respective vdW radii

[c]Ratio of polarisation, induction and dispersion energies

[d]aug-cc-pVDZ: dispersion energy is multiplied by a factor of 1.15. This was determined as the mean value of the ratio of dispersion energies calculated at aug-cc-pVTZ to aug-cc-pVDZ levels (cf. Table 2)

Table 8 DFT-SAPT/aVDZ-PP interaction energies (kcal/mol) and Y···D (Å) for crystal structures

No.	Complex	DFT-D[a]	DFT-SAPT					X···Y/Δr[b]	P:I:D[c]	E^{a}_{tot}	Ref.
		ΔE	E_{tot}	E_1^{Pol}	E_1^{Ex}	E_2^{Ind}	E_2^{Disp}				
1	I_2···DABCO		−24.19	−65.36	83.02	−26.94	−17.15[d]	2.36/1.17	3.8:1.6:1	−26.43	[6]
2	I_2···DTCA		−5.98	−42.06	59.00	−11.82	−12.78[d]	2.73/1.05	3.3:0.9:1	−7.66	[6]
3	C_6Cl_6···C_6Cl_6		−2.1	−1.2	–	−0.1	−4.37[d]	–	0.3:0.02:1	−2.67	[25]
4	C_6Br_6···C_6Br_6		−2.9	−2.3	–	−0.3	−6.10[d]	–	0.4:0.05:1	−3.70	[25]
5	C_4N_3 H_4Br···$C_7F_4O_2HBr$	−4.44	−4.83	−7.58	9.34	−1.77	−5.59[d]	2.89/0.50	1.4:0.3:1	−5.60	[27]
6	$C_6F_4I_2$···$I_2F_4C_6$	−7.03	−7.59	−10.11	11.59	−3.31	−6.85[e]	3.01/0.32 4.06/0.73	1.5:0.5:1	−8.68	[28]
7	$C_7F_4O_2HBr$···$NBrC_4N_3H_2$	−4.44	−4.56	−7.58	9.38	−1.78	−5.45[e]	2.90/0.5 4.03/0.33	1.4:0.3:1	−5.43	
8	1,2-TFIB···TMO	−8.06	−6.47	−16.24	20.35	−3.05	−8.97[e]	3.170.61	1.8:0.4:1	−7.91	[27, 28]
9	1,2-TFIB···TMO	−10.07	−11.87	−19.86	23.34	−6.55	−10.46[e]	3.97/0.44	1.9:0.6:1	−13.53	[27]
10	2-Mercapto-1-methylimidazole···1,2-TFIB	−8.97	−6.78	−12.09	12.58	−2.54	−5.63[e]	3.31/0.47	2.1:0.5:1	−7.68	[30]
11	4,4′-Bipyridine···1,2-TFIB	−7.08	−7.77	−12.70	14.89	−4.35	−6.68[e]	2.91/0.62	1.9:0.7:1	−8.84	[28]
12	1,2-TFIB···TMO	−7.42	−6.28	−10.05	12.20	−2.21	−7.40[e]	3.39/0.39	1.4:0.3:1	−7.46	[27, 28]
13	1,2-TFIB···1,2-TFIB	−2.50	−3.04	−2.22	3.47	−0.43	−4.61[e]	4.01/0.05 3.55/0.22	0.5:0.09:1	−3.79	[28]

14	(3,4,5-Trichlorophenol)$_2$	−1.78	−1.03	−2.00	4.49	−0.41	−3.70[e]	3.11/0.16	0.5:0.1:1	−1.62	[31]
15	1,2-TFIB···1,2-TFIB-(A)	−2.38	−3.17	−2.33	3.87	−0.39	−5.14[e]	3.26/0.7	0.5:0.08:1	−3.99	[27, 28]
16	1,2-TFIB···1,2-TFIB-(B)	−6.21	−7.25	−5.25	9.19	−0.79	−12.34[e]	3.74/0.22	0.4:0.06:1	−9.23	[27, 28]
17	1,2-TFIB···1,2-TFIB-(C)	−1.23	−1.81	−1.39	2.88	−0.30	−3.56[e]	3.43/0.1 3.50/0.17	0.4:0.08:1	−2.37	[27, 28]

[a]DFT-D calculations at B97-D3/def2-QZVP level
[b]Δr is the difference between the X···Y distance and the van der Waals distance
[c]Ratio of polarisation, induction and dispersion energies
[d]aug-cc-pVDZ: dispersion energy is multiplied by a factor of 1.15. This was determined as the mean value of the ratio of dispersion energies calculated at aug-cc-pVTZ to aug-cc-pVDZ levels (cf. Table 2)
[e]Empirical dispersion energy is multiplied by a factor of 1.19. This was determined as the mean value of the ratio of dispersion energies calculated with aug-cc-pVTZ to E^{D}_{empir}

the halogen and the electron donor and the difference between this distance and the sum of the respective vdW radii.

Table 2 collects 18 complexes from the X40 dataset, for which the benchmark CCSD(T)/CBS energies were determined. For all these complexes, DFT-SAPT calculations were performed with both smaller (aug-cc-pVDZ) and larger (aug-cc-pVTZ) basis sets. The aug-cc-pVTZ DFT-SAPT total energies agree better with the CCSD(T)/CBS benchmark energies than the aug-cc-pVDZ ones. The average relative differences amount to 16% and 11%, respectively. The aug-cc-pVTZ DFT-SAPT total energies vary between −1.07 and −7.02 kcal/mol, but the stabilisation energies for most (12) complexes lie in a narrower interval, between 2 and 5 kcal/mol. Following expectations, the largest stabilisation energy was found for complexes containing heavy halogens and trimethylammonium as an electron donor. A comparison of the single energy terms showed that the first-order polarisation and exchange-repulsion, and the second-order induction energies determined with both basis sets are very similar and deviate by less than a few per cent. Dispersion energy is different, and here the aug-cc-pVTZ values are systematically larger than those calculated with the aug-cc-pVDZ basis set, on average by 15% (the largest difference, 23%, was found for the $F_3CI\cdots OCH_2$ complex and the smallest, 5%, for the $BzI\cdots SHCH_3$ complex). This value was used for scaling the dispersion energy calculated with a smaller aug-cc-pVDZ basis set. Only the aug-cc-pVTZ values are utilised in the subsequent discussion.

Investigating the aug-cc-pVTZ single energies, we found that in most (10) cases the dispersion energy is the largest (the most negative), followed by polarisation and induction energies. Only in eight complexes is the polarisation energy larger than the dispersion energy, but the difference is not large (on relative average by 30%). Induction energy is, in all 18 complexes, systematically the smallest, which indicates that with these complexes the charge transfer does not play an important role. All of the Y···D distances are shorter than the sum of the respective vdW radii (vdW distance), which amounts to 3.27, 3.37, 3.50, 3.30, 3.40, 3.53, 3.55, 3.65, 3.78, 2.70, 3.50, 3.70 and 3.96 Å for Cl···O, Br···O, I···O, Cl···N, Br···N, I···N, Cl···S, Br···S, I···S, F···F, Cl···Cl, Br···Br and I···I, respectively. The shortest distances (2.95 and 2.97 Å) were found for the complexes of trifluorobromomethane with formaldehyde and bromo- and iodobenzene with trimethylammonium and the longest distance (3.66 Å) for iodomethane···benzene. The largest contractions of the vdW distance (0.56, 0.49 and 0.43 Å, respectively) were detected for $BzI\cdots NC_3H_9$, $F_3CI\cdots OCH_2$ and $BzBr\cdots NC_3H_9$ complexes. Following expectations, the stabilisation energies of these complexes are among the largest.

Table 3 collects energies for 46 complexes of the XB51 dataset. As in the previous case, the DFT-SAPT/aug-cc-pVTZ total energies agree well with the CCSD(T)/CBS stabilisation energies. The average relative difference (18%) is larger than given previously but still reasonable. In the present case, the DFT-SAPT total energies lie in the broader interval, between −0.51 and −31.5 kcal/mol. The largest DFT-SAPT total energy in the Table 2 amounted to −7.02 kcal/mol. We (arbitrarily) consider this value to be the border between weak and medium, and

strong halogen-bonded complexes. A total of 34 complexes in this table have their stabilisation energies in the range 0.51–7.0 kcal/mol while 12 complexes are characterised by even more favourable stabilisation energy (in the range 7.0–31.5 kcal/mol). Only 14 of the 34 weaker complexes have the dispersion energy larger than the polarisation energy. In the remaining 20 cases, the polarisation is dominant. The dominant stabilisation in the 12 strongest complexes originates in polarisation energy, which is, in all cases, followed by induction energy. Dispersion energy is systematically the smallest one here. Such a combination of these three stabilisation energies (polarisation > induction > dispersion) is unique and was not detected in either the 18 complexes collected in the Table 2 or in the 34 weaker complexes in Table 3. Investigating these complexes, we immediately realise that the large induction energy cannot originate in classical permanent dipole-induced dipole induction energy but rather in charge-transfer energy. This is confirmed by the negative values of the LUMO of these electron acceptors (cf. Table 1), which indicates that they are exceptionally good electron acceptors. All of these strong complexes possess short Y···D distances, even below 2.5 Å. As in the previous case, the largest contractions of the vdW distances (1.11, 1.05 and 1.03 Å, respectively) were found for the strongest complexes, $FI\cdots NC_6H_5$, $H_3N\cdots BrF$ and $FI\cdots NH_3$. It should be noted that the contractions of the vdW distances are in the present case about twice as large as those in the Table 2. Similarly, the stabilisation energies are also much larger in the present complexes. It is apparent that the contractions of the vdW distances of more than 1 Å are connected with large stabilisation energies of more than 20 kcal/mol.

Table 4 collects 11 halogen-bonded complexes which differ from those previously investigated. Four of them are dihalogen dimers possessing a dihalogen bond, another five are complexes of benzene (or methylated benzene) with dihalogen (all having a X···π halogen bond) and, finally, the last two are halogen-bonded and dihalogen-bonded structures of bromomethanol dimer. In this case, the agreement between CCSD(T)/CBS stabilisation energies and DFT-SAPT total energies is comparable with previous cases (the average relative error amounts to 11%). DFT-SAPT stabilisation energies are moderate and are similar to those in Table 2 and part of Table 3 and are in the range of 0.3–6.1 kcal/mol. In all 11 cases, the dispersion energy is dominant and the induction energy is systematically the smallest. Evidently, none of these complexes correspond to the charge-transfer type, and thus all Y···D distances are larger than 2.9 Å. In all of these complexes, the contraction of the vdW distance is only small (less than 0.3 Å).

The complexes shown in Table 5 represent typical model halogen-bonded complexes between standard electron donors (OH_2, NH_3, and OCH_2) and halogen donors (halogen alkenes and alkynes). These complexes are characterised by a modest stabilisation energy between 0.78 and 4.27 kcal/mol and by relatively large halogen-bond lengths (more than 2.96 Å). DFT-SAPT total energies agree moderately with the CCSD(T) benchmark data (the average relative error is larger than previously and amounts to 22%). For complexes 1, 4, 6 and 7 belonging to the weakest group, the dispersion energy is dominant. For the nine remaining complexes, the first order polarisation energy is the largest (the most negative) energy

term. Induction energy is systematically the smallest here. The largest contractions of the vdW distances (0.44 and 0.41 Å, respectively) were again found for the strongest complexes (iodo- and bromomethane with water), and these contractions and stabilisation energies basically agree with those from Table 2.

For the complexes from Table 6, the benchmark CCSD(T) calculations are not available and the DFT-SAPT/aug-cc-pVTZ values are clearly more reliable than the MP2 ones. BrF, ClF, BrF_3 and ClF_3 are the halogen donors, whereas the CH_3CN and CO molecules are used as electron donors. Evidently, the strongest complexes, with a stabilisation energy of more than 10 kcal/mol, are formed between the CH_3CN electron donor and the BrF and BrF_3 electron acceptors (halogen donors). Table 1 shows that BrF_3 is the best electron acceptor (with the lowest LUMO) and BrF is still a very good electron acceptor. The chloro- analogues of these two acceptors exhibit relatively low LUMO values and are thus good acceptors as well. The two strongest complexes with a stabilisation energy of more than 10.9 kcal/mol have dominant polarisation energy followed by induction and dispersion terms. As mentioned above, such a decomposition is characteristic for strong charge-transfer halogen-bonded complexes. The six remaining complexes, with stabilisation energy in the range of 1.1 and 6.3 kcal/mol, belong to weaker halogen-bonded complexes. Here, the polarisation energy is four times more dominant and the dispersion energy twice, and in two cases the induction energy is larger than the dispersion energy. The intermolecular distances are in agreement with the stabilisation energies: for the four most stable ones the distance is short (below 2.61 Å) while in all the remaining cases it is considerably longer. The contractions of vdW distances in these complexes are also among the largest (0.72–0.94 Å).

Several complexes in Table 7 are too large and the DFT-SAPT calculations with the aug-cc-pVTZ basis set would be computationally inaccessible for them. Thus for all of the complexes from Table 7 we have used the smaller, aug-cc-pVDZ basis set, and the resulting dispersion energy was scaled by a factor of 1.15, which had been determined as the average ratio between dispersion energies at aug-cc-pVTZ and aug-cc-pVDZ levels (cf. Table 2). The resulting DFT-SAPT stabilisation energies are very large (between 10.1 and 27.3 kcal/mol) and are much larger than the DFT-D ones. Evidently, the former energies are more reliable. The decomposition of the total DFT-SAPT energy is in line with these values and the polarisation energy is systematically dominant. In eight cases the polarisation energy is followed by the induction energy, which proves the importance of the role of the charge transfer, and these complexes are mostly more stable than the others. In these complexes, the contraction of the vdW distance is very large (about 1 Å or even more) and also here it is valid that a contraction of about 1 Å is connected with a large stabilisation energy of more than 20 kcal/mol. In the remaining seven, mostly weaker complexes, the role of the induction and dispersion energies is reversed, but the polarisation energy remains dominant. The intermolecular distances here are in accord with the total energies and are larger than in the previous case.

Several crystal structures from Table 8 are even larger than those in Table 7 with as many as 32 atoms. Hence, even DFT-SAPT/aug-cc-pVDZ calculations would be

prohibitively expensive. Since SAPT decomposition is necessary for the assignment of the role of electrostatic, induction and dispersion energies (and thus the nature of binding) in extended complexes as well, a hybrid DFT-SAPT method was used here. All the energy terms with the exception of the dispersion one were evaluated using the aug-cc-pVDZ basis set while the dispersion energy was determined empirically (see the original paper [47]). These empirical dispersion energies were scaled by 1.19, which is the average ratio between DFT-SAPT/aug-cc-pVTZ and empirical dispersion energy determined for 18 complexes from Table 2. The dispersion energy for the complexes from Table 8 was thus scaled by 1.15 (complexes 1–6) or by 1.19 (complexes 7–17). Five complexes from Table 8 belong to a group of strongly stable halogen-bonded complexes with a stabilisation energy larger than 7 kcal/mol and in four (out of five) cases the polarisation energy is dominant. The dispersion energy is dominant in only one complex. Among these five complexes, the induction energy is mostly the smallest one and only in one case is the induction term larger than the dispersion term. This concerns the most stable complex (complex 1), having diiodine as an electron acceptor and DABCO as an electron donor. Diiodine is a very good electron acceptor (see Table 1), which is manifested by a large charge-transfer energy and, consequently, induction energy. In this case, the induction term is considerably larger than the dispersion energy and, further, the dominant (polarisation) term is the largest among all 128 complexes investigated. In this group of complexes, the intermolecular distances are all about 3 Å with the exception of diiodine-containing complexes, where the distance is well below 2.8 Å, and the contraction of the vdW distance is the largest (more than 1 Å). The second group of twelve complexes possesses stabilisation energies in the range of 1.5–6.8 kcal/mol and thus belongs among the weak/moderate halogen-bonded complexes. In eight out of twelve cases, the dispersion energy is dominant and only in two cases does the polarisation represent the largest attractive term. The induction energy is systematically the smallest one. All intermolecular distances are rather large and the respective contractions of the vdW distance are small or moderate.

Summarising results from previous tables, we can state that all of the halogen-bonded complexes investigated can be split into two different classes. The 38 complexes in the first group are the strongest with total stabilisation energy larger than 7 kcal/mol. Relatively small intermolecular distances (even below 2.4 Å) and a significant contraction of the vdW distance (up to 1.2 Å) are connected with the important role of induction energy, which is here mostly (in 21 cases) larger than the dispersion energy. In these complexes, the polarisation (electrostatic) energy is almost systematically dominant and only in one complex is the dispersion energy the largest. The second group of 90 less stable halogen-bonded complexes have stabilisation energies between 0.3 and 7 kcal/mol. Their intermolecular distances are contracted much less upon the formation of halogen bonds (mostly less than 0.5 Å; only in diatomic halogen donors could the contraction be larger). In 48 complexes out of the second group (53%), the dispersion energy is dominant, followed by the polarisation and the induction energies. In the rest of the complexes (47%), the polarisation energy is dominant, followed by the dispersion and induction

terms, and only in two cases is the induction energy larger than the dispersion energy.

The electrostatic terms in halogen and hydrogen bonds should be more or less comparable. The contribution of the dispersion energy to the stability of the halogen bond is much larger than that of the hydrogen bond. This is clearly caused by the fact that in the halogen bond two heavy atoms (the halogen and electron donors) with high polarisability are in close contact, while in the case of the hydrogen bond it is only the light hydrogen and electron donor which are close together. To demonstrate the importance of this contact atom pair, we evaluated, besides the total (empirical) dispersion energy, the contribution to the dispersion energy coming from this atom pair. In the case of 14 complexes from Table 2 (complexes with benzene were omitted), the contact atom pair dispersion energy forms on average 39% of the total dispersion energy. This ratio even increased (40%) when eight extended complexes from Table 8 were considered.

In the previous studies, it was demonstrated that the strength of halogen bonding in isolated complexes is proportional to the maximum of the ESP on the halogen [50]. In biological systems, however, this relation may not be so straightforward, because other effects such as solvation/desolvation come into play and the maximum of ESP is related to enthalpy changes rather than to free energies [51, 52]. Here we attempt to relate the properties of monomers, i.e. σ-holes, with the properties of complexes.

Surprisingly enough, the magnitude of the σ-hole correlates weakly with the stabilisation energy, with the correlation coefficient R being 0.52. The dependence of the stabilisation energy on the magnitude of the σ-hole is shown in Fig. 1. Of course, both interacting partners affect the stabilisation of a complex. However, the

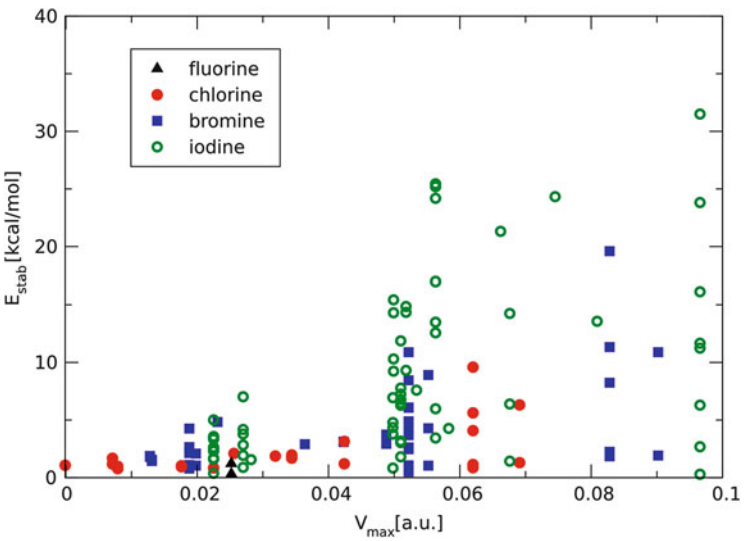


Fig. 1 The dependence of the stabilisation energy E_{stab} on the magnitude of the σ-hole V_{max}

points are spread in a triangular region with columns, distinguishing various halogenated monomers. When selecting the most stable complex of particular halogenated monomers, the correlation between the monomer's magnitude of σ-holes and the stabilisation energy increases to $R = 0.77$. Therefore, it seems that the magnitude of the σ-hole tells us something about the ability of a monomer to create a halogen bond but cannot provide the complete picture on halogen bonding. Indeed, when comparing complexes with the same electron donor, a strong correlation should be expected [50].

When the total stabilisation energy depends on both interacting molecules, the same should be true about the components of stabilisation energy. We did not observe any relation between the magnitudes of the σ-hole and the polarisation or induction terms of DFT-SAPT decomposition, most likely because of the large effect of the electron donor.

4 Conclusions

The analysis of electron acceptors (halogenated molecules) revealed a correlation between the extreme of ESP and the spatial extent of the positive region on top of the halogen boundary. The magnitude and size of a halogen σ-hole suggest a possible strength of the halogen bond in noncovalent complexes.

It was shown that all halogen-bonded complexes investigated could be split into two groups on the basis of their stabilisation energies. The complexes in the first group are stronger (their stabilisation energy is larger than 7 kcal/mol) and can be characterised as halogen-bonded complexes with a strong charge-transfer contribution. In practically all cases in this class, the polarisation (electrostatic) term is dominant and the induction term is mostly the second most important term, reflecting the important role of charge-transfer energy. The second class of halogen-bonded complexes is characterised by weaker stabilisation energies (below 7 kcal/mol) and represents rather standard halogen-bonded complexes. In this class of complexes, the dispersion energy is mostly dominant. In the whole set of 128 halogen-bonded complexes investigated, the polarisation (electrostatic) energy is dominant in 62% while in the remaining 38% it is the dispersion energy, which represents the dominant attractive term. We can thus state that the concerted action of polarisation and dispersion energies is responsible for the characteristic properties of halogen bonding. The electrostatic interaction between the positive σ-hole and the negative electron donor is responsible not only for the stability but also for the high directionality of the bond, while dispersion energy is responsible for its high stability. A dominant role is played by the contact atom pair (the halogen and electron donors), which contributes as much as 40% of the total dispersion energy. This significant contribution, which is characteristic for the halogen bond, is a consequence of two factors: first, the attractive electrostatic interaction between the halogen positive σ-hole and the negative electron donor and, second, the lower exchange-repulsion between the two subsystems, which is

also manifested as so-called polar flattening [53]. The recent IUPAC definition [9] of the halogen bond states that 'the forces involved in the formation of the halogen bond are primarily electrostatic, but polarisation, charge-transfer and dispersion contributions all play an important role'. A question thus arises as to whether the definition is sufficiently accurate and describes the unique phenomenon of the σ-hole of halogen bonding satisfactorily and fully.

Acknowledgements This work was part of the Research Project RVO: 61388963 of the Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic. It was also supported by the Czech Science Foundation [P208/12/G016] and the operational program Research and Development for Innovations of the European Social Fund (CZ 1.05/2.1.00/03/0058). MHK acknowledges the kind support provided by the Alexander von Humboldt Foundation.

References

1. Hobza P, Müller Dethlef K (2009) Non-covalent interactions. Royal Society of Chemistry, Cambridge
2. Clark T, Hennemann M, Murray JS, Politzer P (2007) J Mol Model 13:291–296
3. Politzer P, Murray JS, Clark T (2013) Phys Chem Chem Phys 15:11178–11189
4. Politzer P, Murray JS (2013) ChemPhysChem 14:278–294
5. Murray J, Lane P, Politzer P (2009) J Mol Model 15:723–729
6. Řezáč J, Riley KE, Hobza P (2012) J Chem Theory Comput 8:4285–4292
7. Kozuch S, Martin JML (2013) J Chem Theory Comput 9:1918–1931
8. Deepa P, Sedlak R, Hobza P (2014) Phys Chem Chem Phys 16:6679
9. Desiraju GR, Ho PS, Kloo L, Legon AC, Marguardt R, Metrangolo P, Politzer P, Resnati G, Rissanen K (2013) IUPAC definition of halogen bond. Pure Appl Chem 8:1711–1713
10. Riley KE, Hobza P (2013) Phys Chem Chem Phys 15:17742
11. Kolář M, Hostaš J, Hobza P (2014) Phys Chem Chem Phys 16:9987
12. Kolář M, Carloni P, Hobza P (2014) Phys Chem Chem Phys 16:19111
13. Bader RFW, Carroll MT, Cheeseman JR, Chang C (1987) J Am Chem Soc 109:7968–7979
14. Dunning TH Jr (1989) J Chem Phys 90:1007
15. Woon DE, Dunning TH Jr (1993) J Chem Phys 98:1358
16. Adamo C, Barone V (1999) J Chem Phys 110:6158–6169
17. Peterson KA, Figgen D, Goll E, Stoll H, Dolg M (2003) J Chem Phys 119:11113
18. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, Nakatsuji H, Caricato M, Li X, Hratchian HP, Izmaylov AF, Bloino J, Zheng G, Sonnenberg JL, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Montgomery JA Jr, Peralta JE, Ogliaro F, Bearpark M, Heyd JJ, Brothers E, Kudin KN, Staroverov VN, Kobayashi R, Normand J, Raghavachari K, Rendell A, Burant JC, Iyengar SS, Tomasi J, Cossi M, Rega N, Millam JM, Klene M, Knox JE, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Martin RL, Morokuma K, Zakrzewski VG, Voth GA, Salvador P, Dannenberg JJ, Dapprich S, Daniels AD, Farkas Ö, Foresman JB, Ortiz JV, Cioslowski J, Fox DJ (2009) Gaussian 09, revision D.01. Gaussian, Inc., Wallingford
19. Jeziorski B, Moszynski R, Szalewicz K (1994) Chem Rev 94:1887–1930
20. Hesselmann A, Jansen G (2002) Chem Phys Lett 357:464–470
21. Hesselmann A, Jansen G (2002) Chem Phys Lett 362:319–325

22. Misquitta AJ, Szalewicz K (2002) Chem Phys Lett 357:301–306
23. Jansen G, Hesselmann A (2001) J Phys Chem A 105:11156–11157
24. Hesselmann A, Jansen G (2003) Chem Phys Lett 367:778–784
25. Hesselmann A, Jansen G, Schütz M (2005) J Chem Phys 122:014103–014119
26. Hesselmann A, Jansen G, Schütz M (2006) J Am Chem Soc 128:11730–11731
27. Podeszwa R, Bukowski R, Szalewicz K (2006) J Phys Chem A 110:10345–10354
28. Williams HL, Chabalowski CF (2001) J Phys Chem A 105:646–659
29. Werner H-J, Knowles PJ, Manby FR, Schuetz M, Celani P, Knizia G, Korona T, Lindh R, Mitrushenkov A, Rauhut G, Adler TB, Amos RD, Bernhardsson A, Berning A, Cooper DLO, Deegan MJ, Dobbyn AJ, Eckert F, Goll F, Hampel C, Hesselmann A, Hetzer G, Hrenar T, Jansen G, Koeppl C, Liu Y, Lloyd AW, Mata RA, May AJ, McNicholas SJ, Meyer W, Mura ME, Nicklass A, Palmieri P, Pflueger K, Pitzer K, Reiher M, Shiozaki T, Stoll H, Stone AJ, Tarroni R, Thorsteinsson T, Wang M, Wolf A (2010) MOLPRO, version 2010.1, a package of ab initio programs. See http://www.molpro.net
30. TURBOMOLE V6.3 2011, a development of the University of Karlsruhe and the Forschungszentrum Karlsruhe GmbH, 1989–2007, TURBOMOLE GmbH, since 2007; http://www.turbomole.com
31. Price SL, Stone AJ, Lucas J, Rowland RS, Thornley AE (1994) J Am Chem Soc 116:4910–4918
32. Awwadi FF, Willett RD, Peterson KA, Twamley B (2006) Chem Eur J 12:8952–8960
33. Grimme S, Antony J, Ehrlich S, Krieg H (2010) J Chem Phys 132:154104
34. Munusamy E, Sedlak R, Hobza P (2011) ChemPhysChem 12:3253
35. Riley KE, Řezáč J, Hobza P (2013) J Mol Model 19:2879
36. Sedlak R, Deepa P, Hobza P (2014) J Phys Chem A 118:3846–3855
37. Lu YX, Fan J-C, Zaho W-N, Jiang Y-J, Yu Q-S (2009) J Comput Chem 30:725
38. Wang W (2011) J Phys Chem A 115:9294–9299
39. Chudzinski MG, Taylor MS (2012) J Org Chem 77:3483–3491
40. Trnka J, Sedlak R, Kolář M, Hobza P (2013) J Phys Chem A 117:4331–4337
41. Mukherjee A, Desiraju GR (2011) Cryst Growth Des 11:3735–3739
42. Cincic D, Friscic T, Jones W (2008) J Am Chem Soc 130:7524–7525
43. Cauliez P, Polo V, Roisnel T, Llusar R, Fourmigue M (2010) CrystEngComm 12:558–566
44. Cincic D, Friscic T, Jones W (2011) CrystEngComm 13:3224–3231
45. Jay JI, Padgett CW, Walsh RDB, Hanks TW, Pennington WT (2001) Cryst Growth Des 1:501–507
46. Zhao Y, Truhlar DG (2008) Theor Chem Acc 120:215–241
47. Boys SF, Bernardi F (2007) Mol Phys 19:553–566
48. Hesselmann A (2011) J Phys Chem A 115:11321–11330
49. Riley K, Murray J, Fanfrlík J, Řezáč J, Solá R, Concha M, Ramos F, Politzer P (2011) J Mol Model 17:3309–3318
50. Riley KE, Murray JS, Politzer P, Concha MC, Hobza P (2008) J Chem Theory Comput 5:155–163
51. Hardegger LA, Kuhn B, Spinnler B, Anselm L, Ecabert R, Stihle M, Gsell B, Thoma R, Diez J, Benz J, Plancher J-M, Hartmann G, Banner DW, Haap W, Diederich F (2011) Angew Chem Int Ed 50:314–318
52. Fanfrlík J, Kolář M, Kamlar M, Hurný D, Ruiz FX, Cousido-Siah A, Mitschler A, Řezáč J, Munusamy E, Lepšík M, Matějíček P, Veselý J, Podjarny A, Hobza P (2013) ACS Chem Biol 8:2484–2492
53. El Kerdawy A, Murray JS, Politzer P, Bleiziffer P, Heßelmann A, Görling A, Clark T (2013) J Chem Theory Comput 9:2264–2275

SUPPLEMENTARY MATERIAL FOR

The Characteristics of a σ-hole and the Nature of a Halogen Bond

Michal Kolář[1,2,3,+], Palanisamy Deepa[1,+], Haresh Ajani[1], Adam Pecina[1] and Pavel Hobza*[1,4]

[1]Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo nám. 2, 166 10 Prague 6, Czech Republic.

E-mail: pavel.hobza@uochb.cas.cz

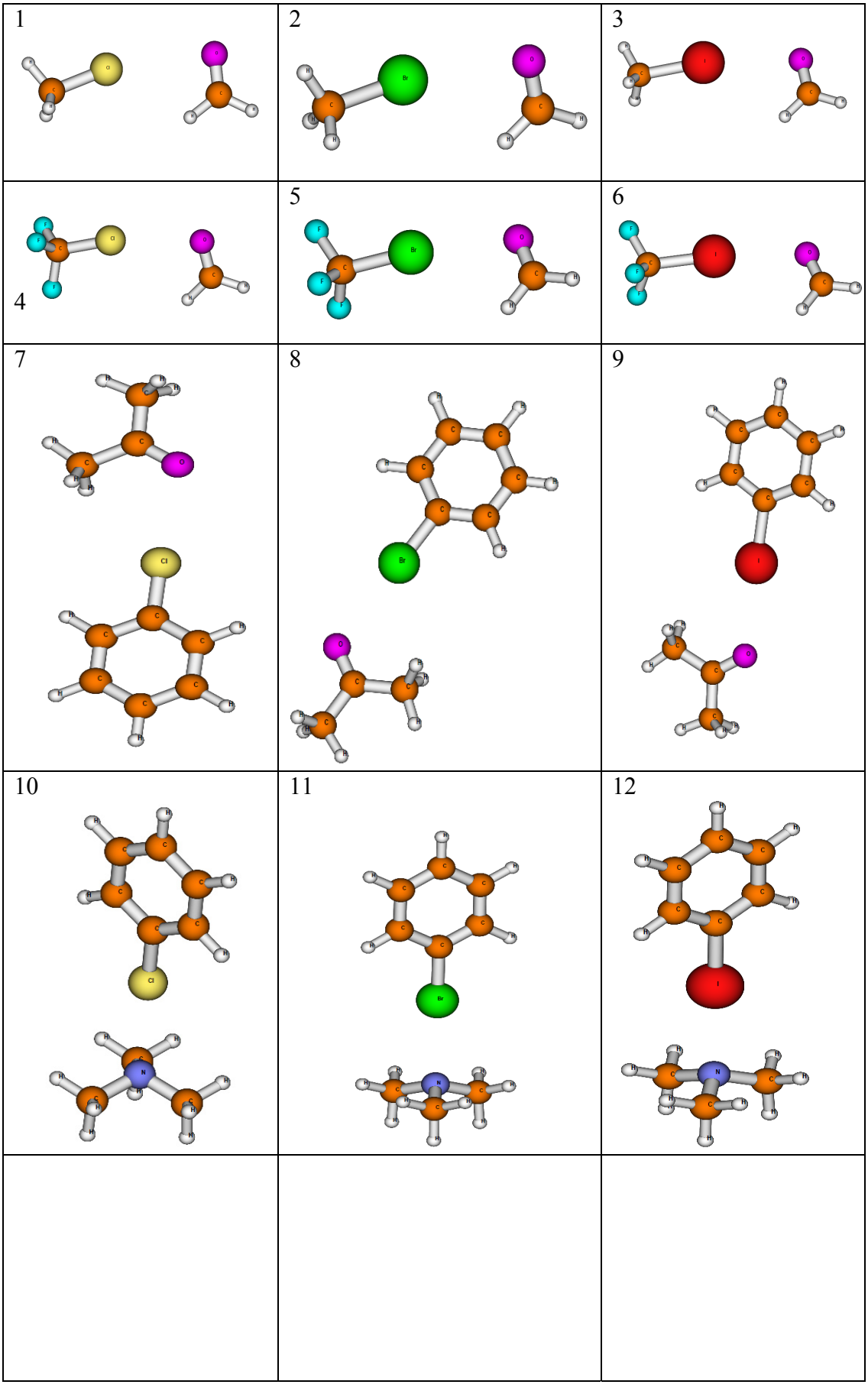
[2]Institute for Advanced Simulations (IAS-5), Forschungszentrum Jülich GmbH, 52428 Jülich, Germany.

[3]Computational Biophysics, German Research School for Simulation Sciences GmbH, 52428 Jülich, Germany.

[4]Regional Centre of Advanced Technologies and Materials, Department of Physical Chemistry, Palacky University, 771 46 Olomouc, Czech Republic.

[+]These authors have contributed equally to this work.

Fig. S1 Structures of 18 complexes from the X40 dataset[1]



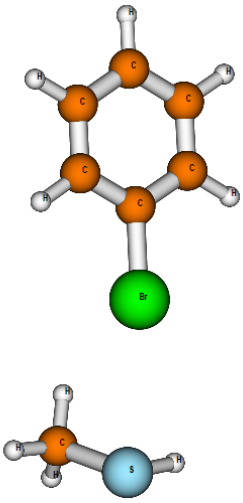
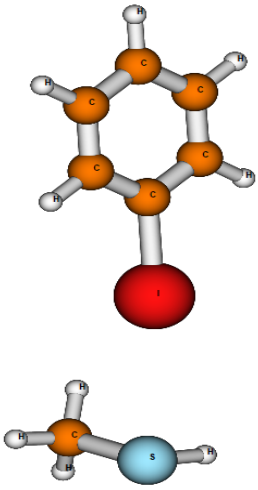
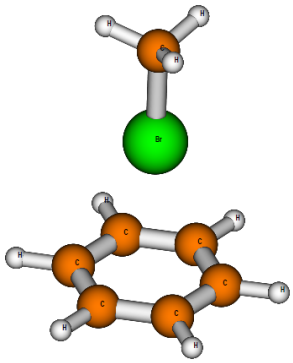
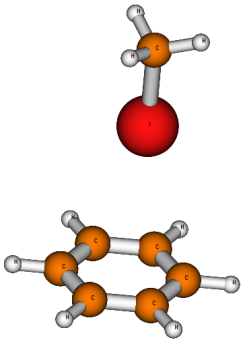
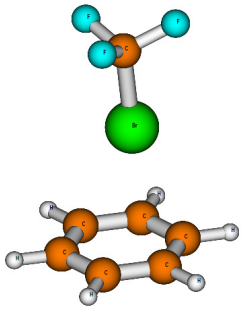
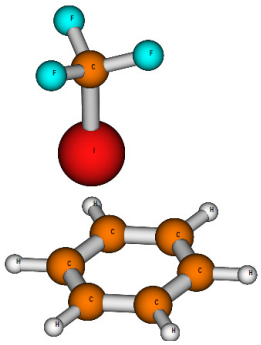
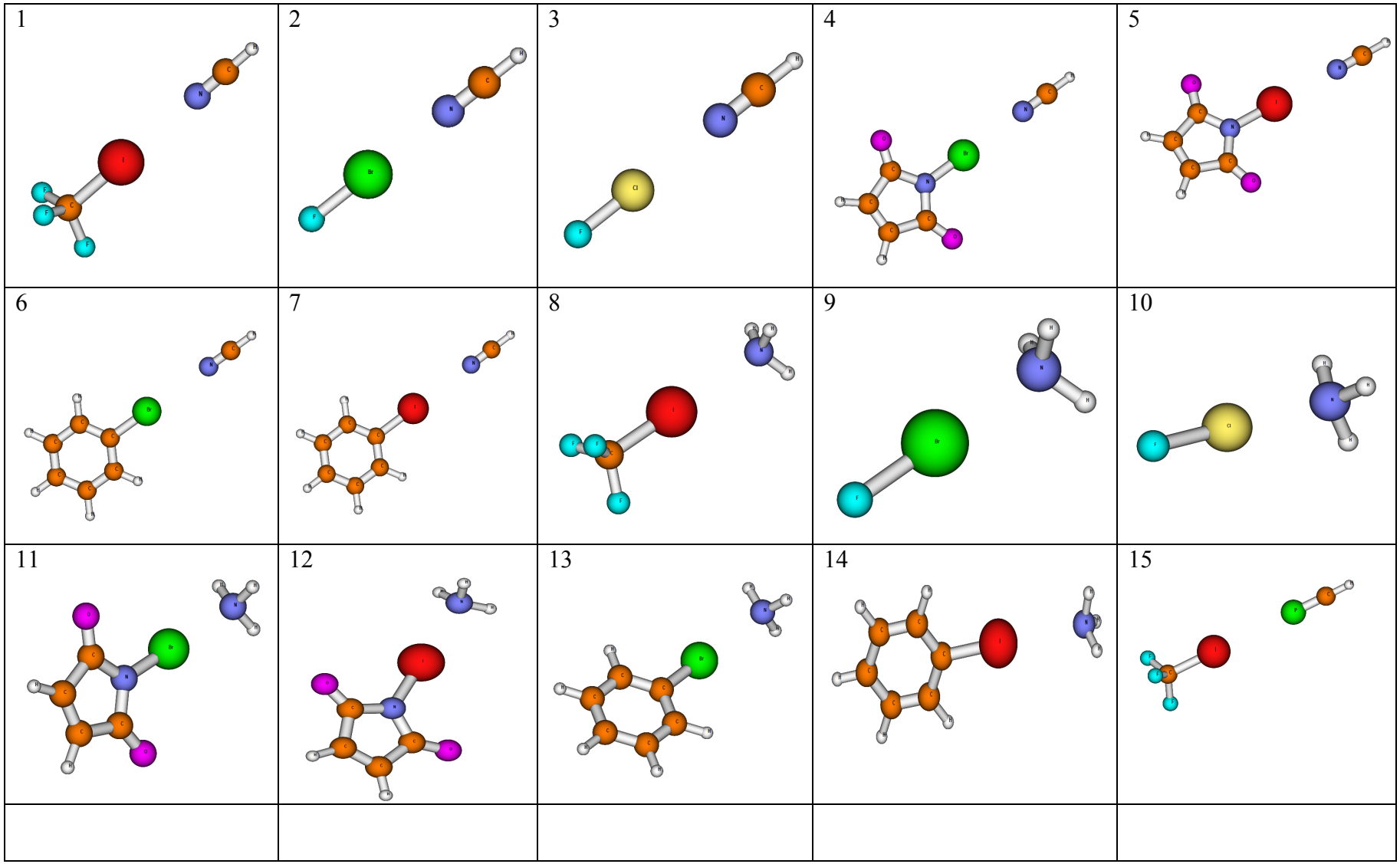
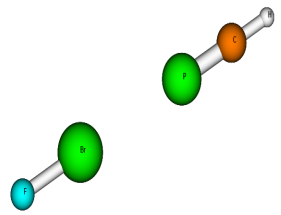
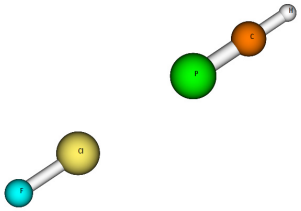
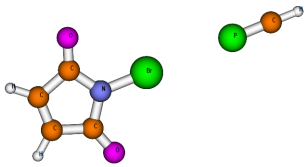
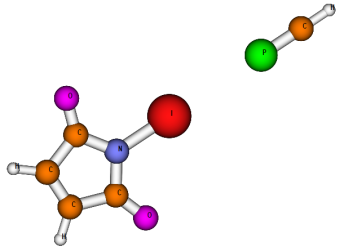
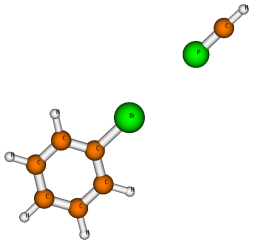
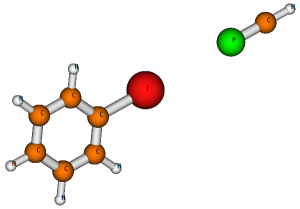
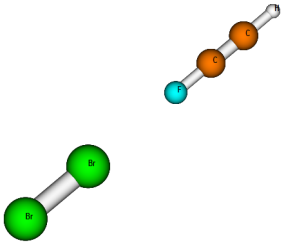
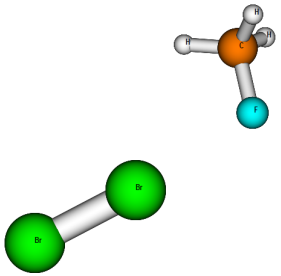
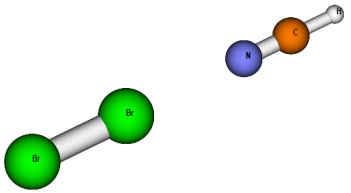
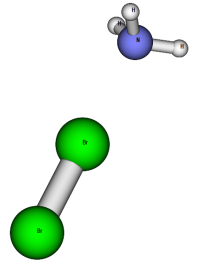
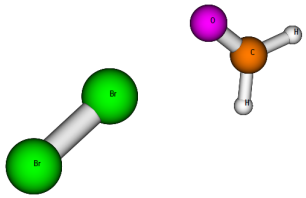
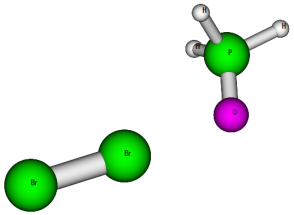
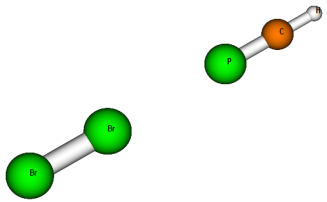
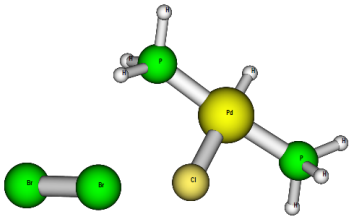
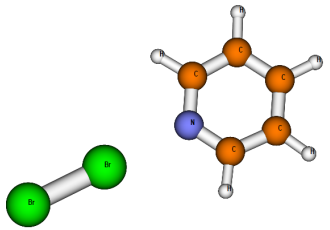
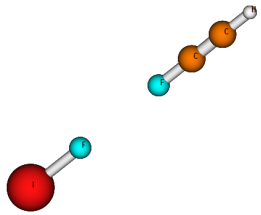
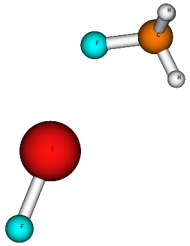
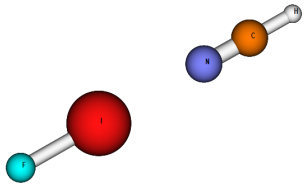
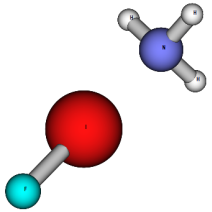
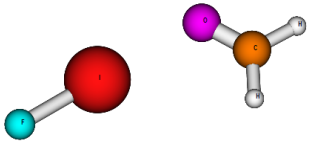
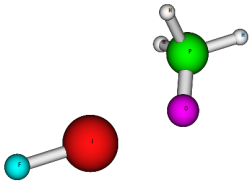
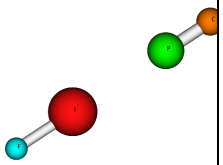
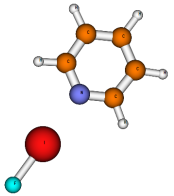
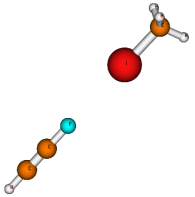
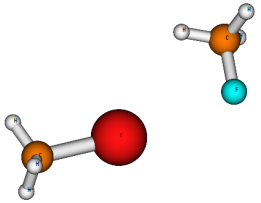
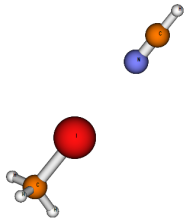
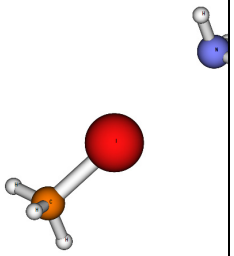
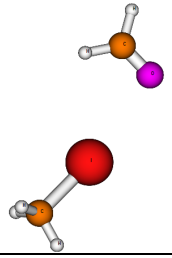
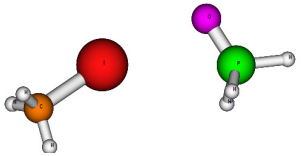
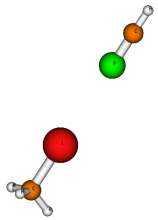
13 	14 	15 
16 	17 	18 

Fig. S2 Structures of 46 complexes from the XB51 dataset[2]



16 	17 	18 	19 	20 
21 	22 	23 	24 	25 
26 	27 	28 	29 	30 

31 	32 	33 	34 	35 
36 	37 	38 	39 	40 
41 	42 	43 	44 	45 

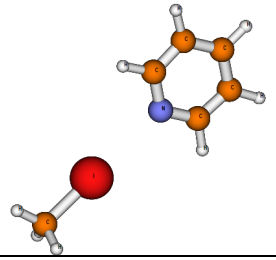
46 				

Fig. S3 Structure of 11 complexes from our previous papers[3-5]

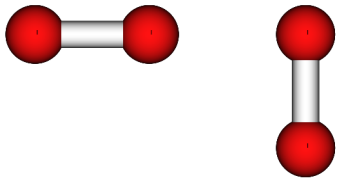
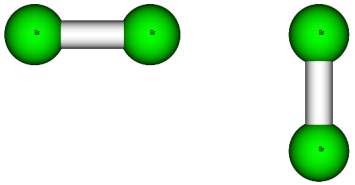
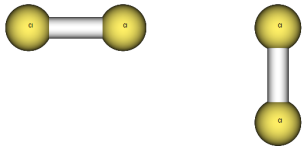
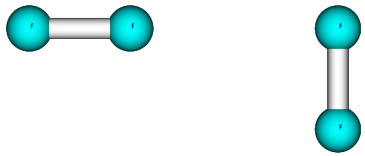
8 	9 	10 	11 

Fig. S4 Structure of 13 complexes from Ref.6

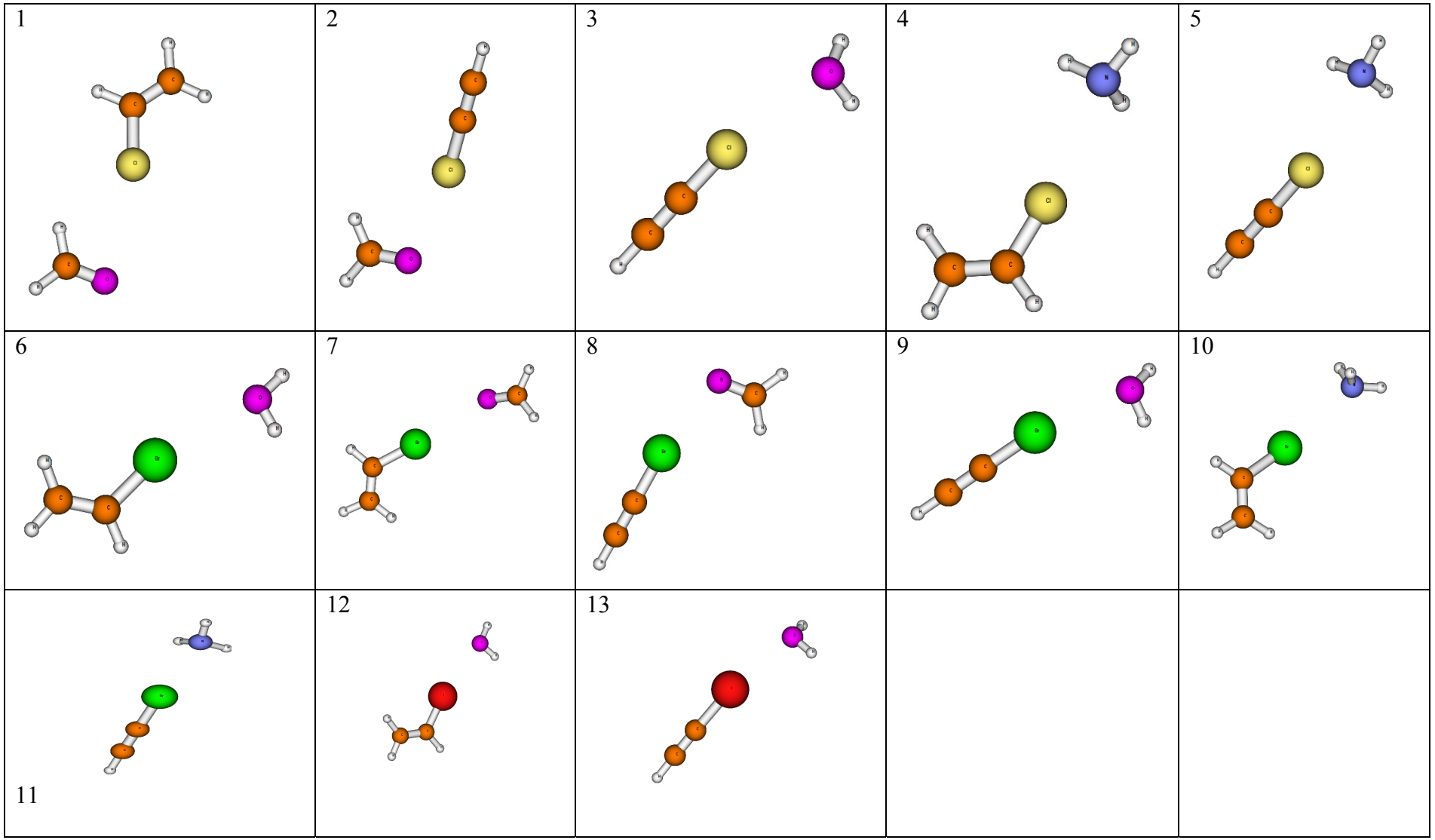


Fig. S5 Structures of 8 complexes of the crystal motifs[7] taken from the Cambridge Structure Database

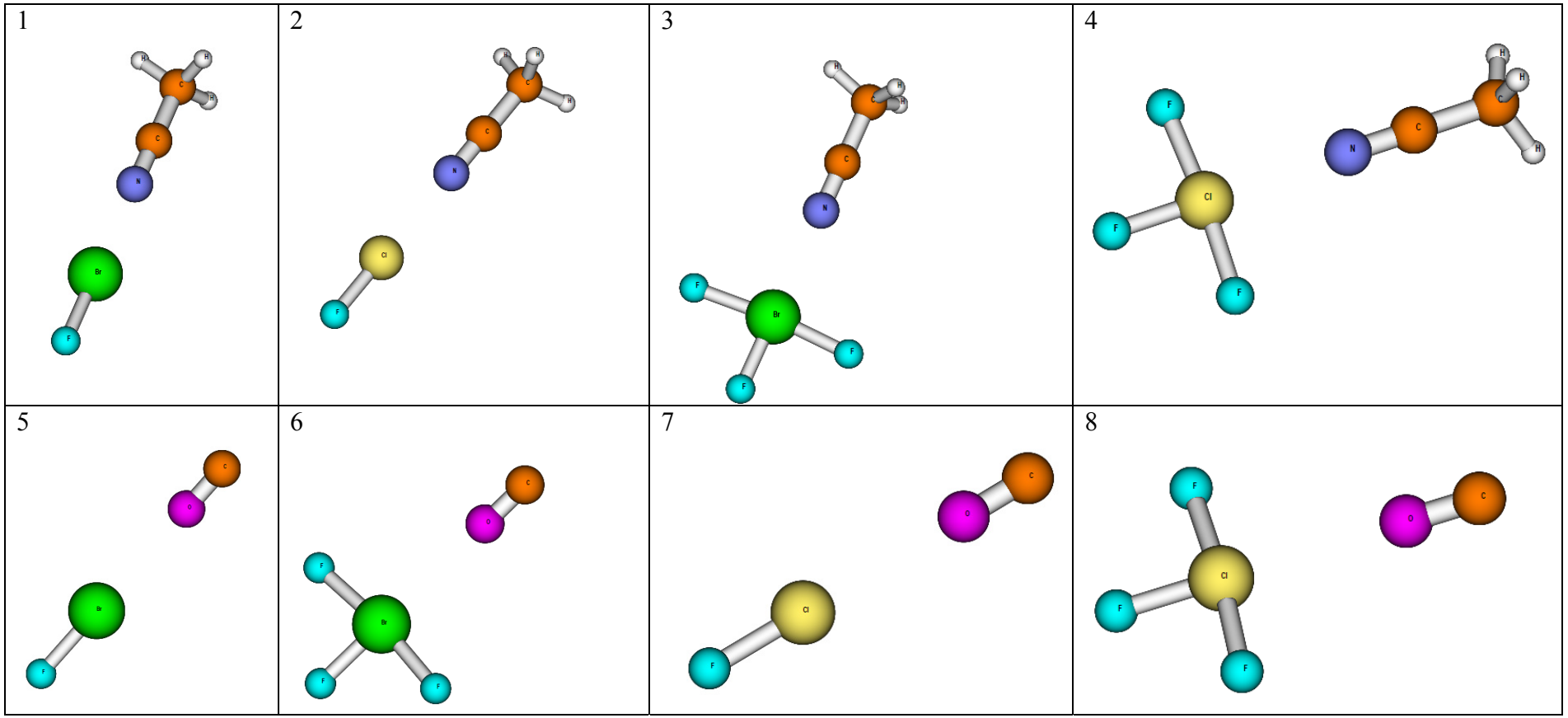
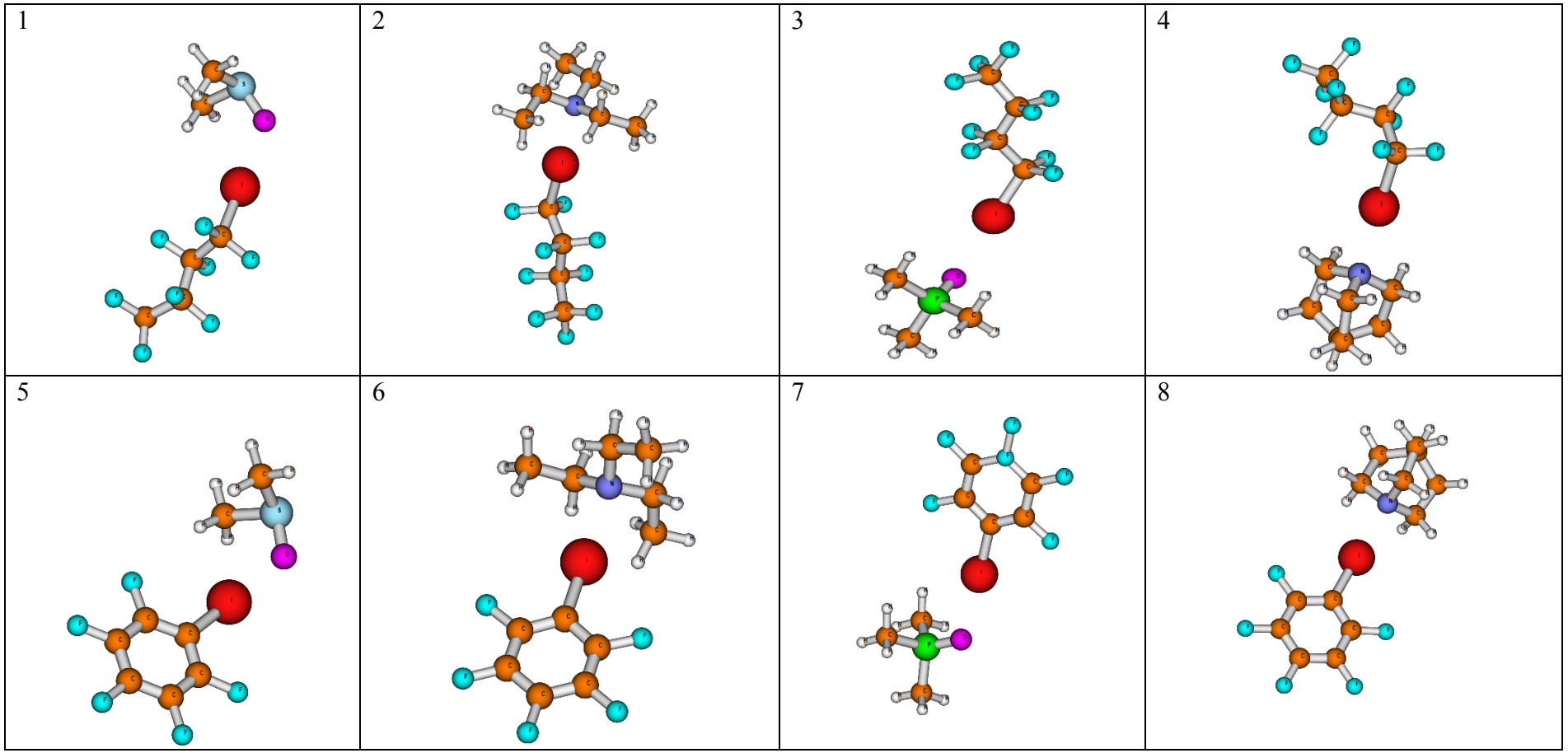


Fig. S6 Structures of 15 complexes from Ref. 8 for which the binding free energy (in the non-polar solvent) was measured.



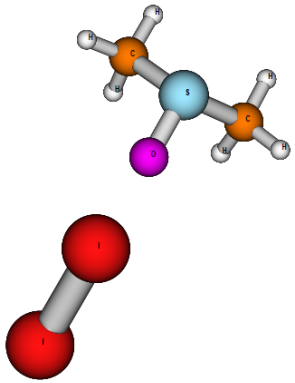
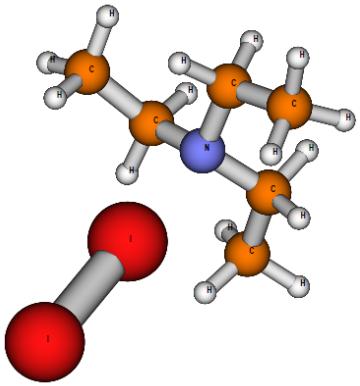
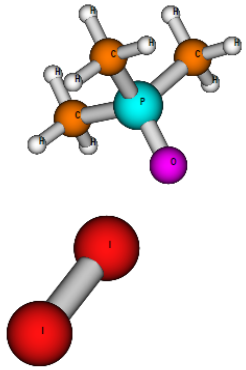
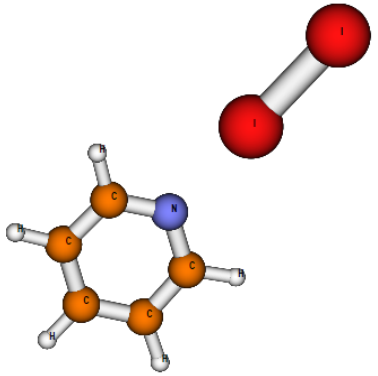
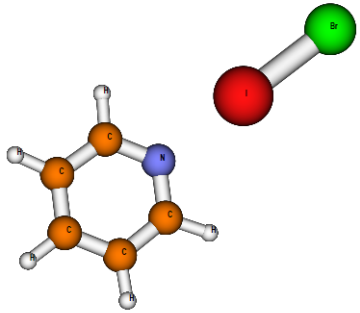
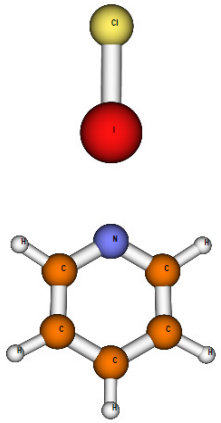
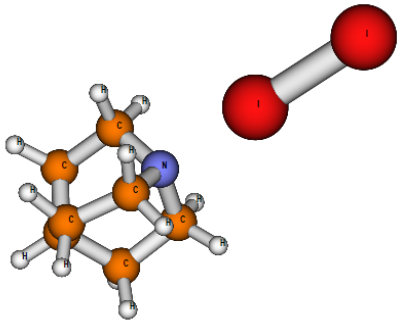
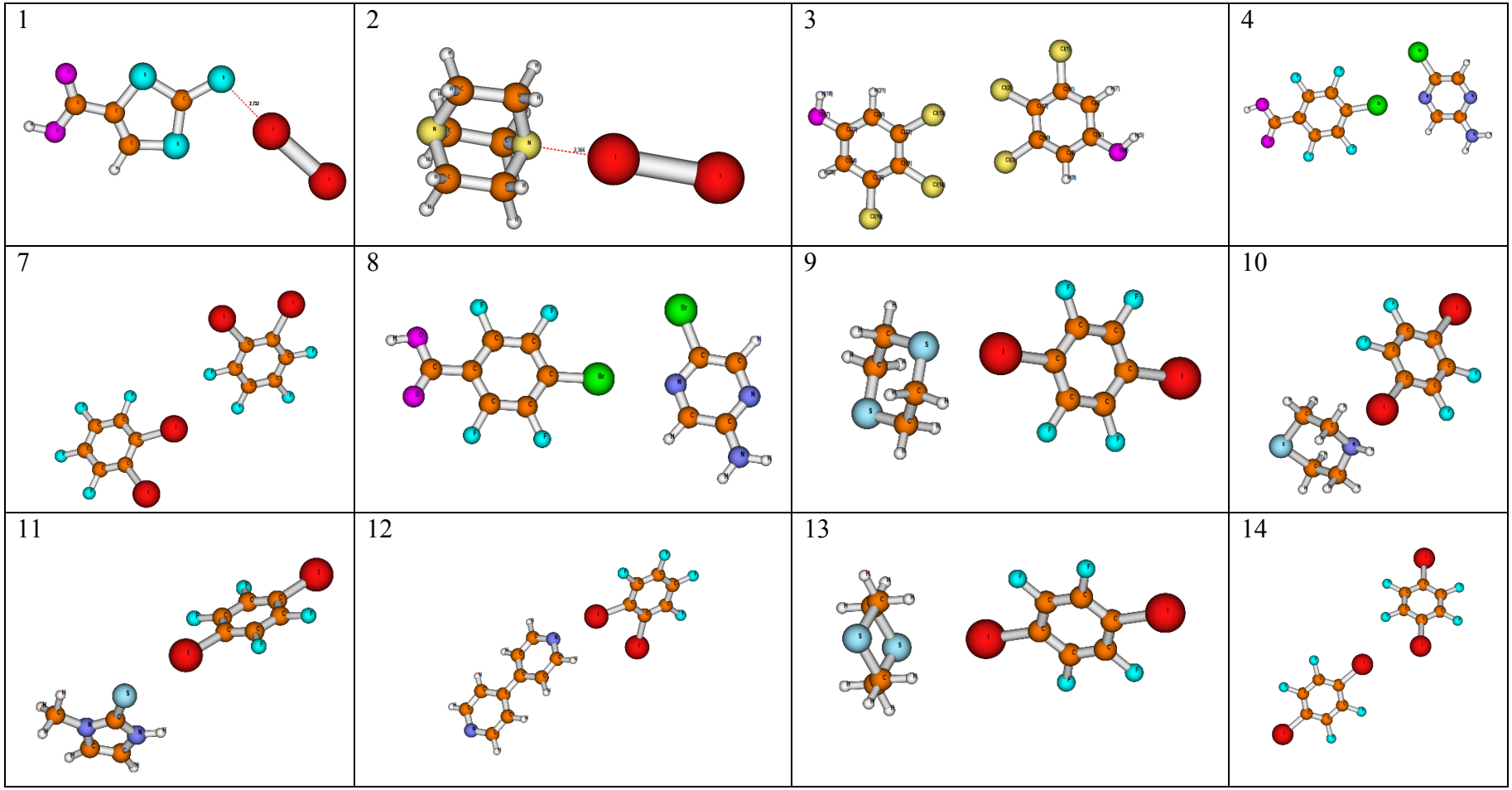
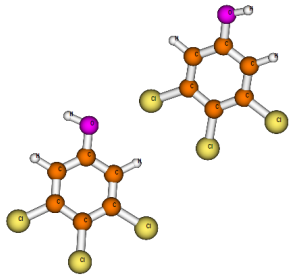
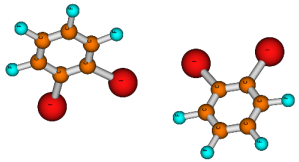
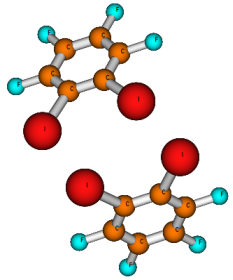
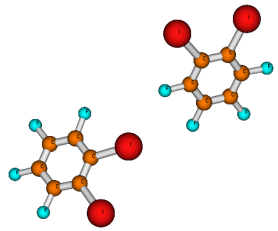
9 	10 	11 	12 
13 	14 	15 	

Fig. S7 Structures of 17 structures of organic crystals taken from Refs. 9-15.



15 	16 	17 	18 

References

1. J. Řezáč, K. E. Riley and P. Hobza, *J. Chem. Theory Comput.*, 2012, **8**, 4285-4292.
2. S. Kozuch and J. M. L. Martin, *J. Chem. Theory Comput.*, 2013, **9**, 1918-1931
3. E. Munusamy, R. Sedlak and P. Hobza, *ChemPhysChem* 2011, **12**, 3253.
4. K. E. Riley, J. Řezáč and P. Hobza, *J. Mol. Model.* 2013, **19**, 2879.
5. R. Sedlak, P. Deepa and P. Hobza, *J. Phys. Chem A* 2014 doi: 10.1021/jp502648e.
6. Y. X. Lu, J-C. Fan, W-N. Zaho, Y.-J. Jiang and Q.-S. Yu, *J. Compt. Chem.*, 2009, **30**, 725.
7. W. Wang, J. Phys. Chem. A 2011, **115**, 9294-9299.
8. M. G. Chudzinski and M. S. Taylor, *J. Org. Chem*, 2012, **77**, 3483-3491.
9. J. Trnka, R. Sedlak, M. Kolář and P. Hobza, *J. Phys. Chem*. A 2013, **117**, 4331-4337.
10. Mukherjee, G. R. Desiraju, *Crys. Growth Design* 2011, **11**, 3735-9.
11. D. Cincic, T. Friscic and W. Jones, *J. Am. Chem. Soc*. 2008, **130**, 7524–7525.
12. P. Cauliez, V. Polo, T. Roisnel, R. Llusar and M. Fourmigue. *Cryst. Eng. Comm*. 2010, **12**, 558-566.
13. D. Cincic, T. Friscic and W. Jones, *Cryst. Eng.Comm*. 2011, **13**, 3224-3231.
14. J. I. Jay, C. W. Padgett, R. D. B. Walsh, T. W. Hanks and W. T. Pennington. *Cryst. Growth Des*. 2001, **1**, 501-507.
15. Mukherjee and G. R. Desiraju *Cryst. Growth Des*. 2011, **11**, 3735-3739.

Appendix B

Chalcogen Bonding

DOI: 10.1002/anie.201405901

The Dominant Role of Chalcogen Bonding in the Crystal Packing of 2D/3D Aromatics**

Jindřich Fanfrlík, Adam Přáda, Zdeňka Padělková, Adam Pecina, Jan Macháček, Martin Lepšík, Josef Holub, Aleš Růžička, Drahomír Hnyk,* and Pavel Hobza**

Abstract: *The chalcogen bond is a nonclassical σ-hole-based noncovalent interaction with emerging applications in medicinal chemistry and material science. It is found in organic compounds, including 2D aromatics, but has so far never been observed in 3D aromatic inorganic boron hydrides. Thiaboranes, harboring a sulfur heteroatom in the icosahedral cage, are candidates for the formation of chalcogen bonds. The phenyl-substituted thiaborane, synthesized and crystalized in this study, forms sulfur···π type chalcogen bonds. Quantum chemical analysis revealed that these interactions are considerably stronger than both in their organic counterparts and in the known halogen bond. The reason is the existence of a highly positive σ-hole on the positively charged sulfur atom. This discovery expands the possibilities of applying substituted boron clusters in crystal engineering and drug design.*

The chalcogen bond is a novel type of noncovalent interaction between a chalcogen atom and an electron (e^-) donor.[1–4] It belongs to the family of σ-hole bonding, where the halogen bond (X-bond) is by far the most known. Since chalcogen atoms are electronegative elements, they are usually negatively charged in organic compounds. The chalcogen bond, that is, the bond between a (mostly) negatively charged chalcogen atom and a negatively charged e^- donor, is thus as counterintuitive as the X-bond. Because of the unequal occupation of the valence orbitals at the chalcogen (halogen), the electrostatic potential (ESP) around the chalcogen atom is strongly anisotropic. Therefore, besides the expected negative areas, there are also areas of positive ESP, called σ-holes.[5]

Although the chalcogen bond is not so well researched compared to the X-bond, it plays an important role in crystal engineering and in interactions of drugs or biological molecules.[2,3,6–10] It is important for the biological activity of several organic molecules.[8] An analysis of the Protein Data Bank suggests that the S···O interactions influence protein structures.[3,6]

A σ-hole is characterized by its magnitude ($V_{s,max}$) and size.[11] $V_{s,max}$ is defined as the value of the most positive ESP of an e^- density surface and the size as the spatial extent of the positive region. $V_{s,max}$ and the size increases on going from Cl to I or from S to Te. The chemical environment also plays a role. $V_{s,max}$ can be increased by e^--withdrawing groups in the vicinity of the X or chalcogen atom.[4,11,12] As also noticed, the higher the $V_{s,max}$ value is, the stronger is the X-bond. A modulation of the X-bond in protein–inhibitor complexes was used to reduce the IC_{50} values accordingly.[13,14]

Reference interaction energies (ΔE) for the X-bond are obtained using the highly accurate CCSD(T) calculations. Hartree–Fock (HF) and density funtional theory (DFT) usually give too low ΔE values.[12] Their use should thus be verified by the CCSD(T) calculations.

Inherently e^--deficient polyhedral boron clusters (boranes) exhibit an astonishing variety of stable structures. Numerous applications include radioactive waste extraction, nanotechnology and medicinal chemistry.[15–17] The properties which make boranes such suitable entities include their hydrophobicity, shape, 3D aromaticity, stability and ability to form dihydrogen bonds.[18,19] An important class of boranes comprises *closo*-$B_nH_n^{2-}$ (known for $n = 5$–12) dianions. The

[*] Dr. J. Fanfrlík,[+] A. Pecina, Dr. M. Lepšík, Prof. P. Hobza
Gilead Sciences Research Center and
Institute of Organic Chemistry and Biochemistry
Academy of Sciences of the Czech Republic, v.v.i.
Flemingovo nám. 2, 16610 Prague 6 (Czech Republic)
E-mail: hobza@uochb.cas.cz
Prof. P. Hobza
Regional Center of Advanced Technologies and Materials
Department of Physical Chemistry, Palacký University
77146 Olomouc (Czech Republic)
A. Přáda,[+] Dr. J. Macháček, Dr. J. Holub, Dr. D. Hnyk
Institute of Inorganic Chemistry
Academy of Sciences of the Czech Republic, v.v.i.
25068 Řež u Prahy (Czech Republic)
E-mail: hnyk@iic.cas.cz
Dr. Z. Padělková,[+] Prof. A. Růžička
University of Pardubice
Studentská 573, 53210 Pardubice (Czech Republic)
E-mail: ales.ruzicka@upce.cz

[+] These authors contributed equally to this work.

[**] This work was supported by research projects RVO 61388963 awarded by the Academy of Sciences of the Czech Republic. We acknowledge the financial support of the Czech Science Foundation (J.F., A.P., M.L., P.H.: grant number P208/12/G016; D.H.: grant number P208/10/2269). We also thank the Gilead Sciences and IOCB Research Centre for financial support. This work was also supported by the Operational Program Research and Development for Innovations—European Science Fund (grant number CZ.1.05/2.1.00/03.0058). This work was supported by the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070), funded by the European Regional Development Fund and the national budget of the Czech Republic via the Research and Development for Innovations Operational Programme, as well as Czech Ministry of Education, Youth and Sports via the project Large Research, Development and Innovations Infrastructures (LM2011033).

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201405901.

BH^{2-} vertices can be formally replaced by isoelectronic CH^- or S units and can thus form carboranes or neutral thiaboranes, respectively. Apart from the parent thiaborane, *closo*-1-$SB_{11}H_{11}$, its 12-Cl- and 12-I-variants (**Cl-SB$_{11}$** and **I-SB$_{11}$**, respectively) were previously prepared in our laboratories.[20] This prompted us to prepare the thiaborane with an antipodal phenyl *exo*-substitution, 12-Ph-*closo*-1-$SB_{11}H_{10}$ (**Ph-SB$_{11}$**), in which 2D and 3D aromatics are connected (see the Supporting Information). These substituted compounds enabled us to analyze the ability of thiaboranes to form chalcogen bonds.

Conceivably, the charge distribution of thiaboranes differs significantly from that of sulfur-containing organic compounds. The comparison of calculated ESPs and dipole moments is shown in Figure 1.

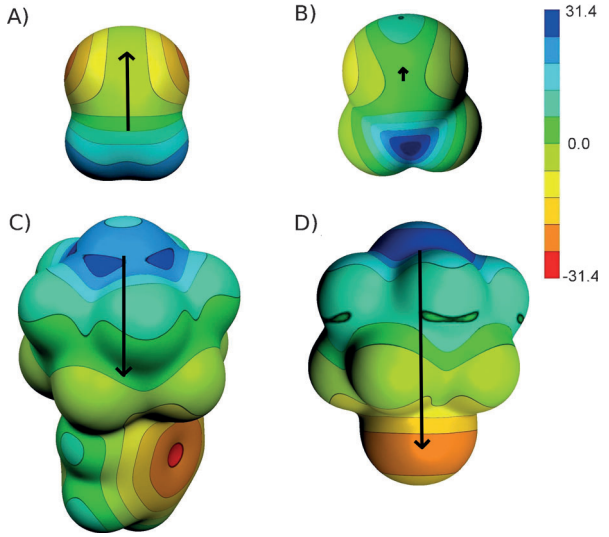


Figure 1. Electrostatic potentials (ESPs) on 0.001 a.u. and dipole moments (arrows) computed at the HF/cc-pVDZ level. The molecular surfaces of A) $H_2C{=}S$, B) $F_2C{=}S$, C) **Ph-SB$_{11}$** and D) **Cl-SB$_{11}$**. ESP in kcal mol^{-1}.

The S atom in $H_2C{=}S$ and $F_2C{=}S$ is divalent and negatively charged (Table 1). The evidence of the σ-hole is clearly visible in Figure 1 and Table 1, which show less negative and positive σ-holes in $H_2C{=}S$ and $F_2C{=}S$, respectively. The key role in this respect is played by two e^--withdrawing F atoms here. The most positive regions are localized at the top of the S atom. In contrast, the S atom in thiaboranes is bound to five B atoms and is thus positively charged (Table 1). Nevertheless, the ESP of **Ph-SB$_{11}$** clearly shows the existence of a less positive top of the S atom and of five highly positive σ-holes on its sides (Figure 1 C). The $V_{s,max}$ is even higher here than in the majority of halogenated compounds. The $V_{s,max}$ of **Ph-SB$_{11}$** is comparable to the $V_{s,max}$ of Br-benzene with several e^--withdrawing groups.[11,12] In **Cl-SB$_{11}$**, the σ-holes are even bigger and joined into a more positive belt, which ranges from 120° to 150° from the B12-S axis. It shows that the properties of σ-holes can be tuned by introducing e^--withdrawing groups on the 3D aromatic cage. Similar tuning of σ-hole properties on the 2D aromatic are well known for X-bonds. These findings have important consequences—the resulting chalcogen bond should be strong and bent, unlike the linear X-bonds. A detailed computational study on σ-hole bonding of heteroboranes is under preparation.

Table 1: The magnitude of σ-holes ($V_{s,max}$), Mulliken partial atomic charges on the S atom (q_s), and dipole moment (μ) in kcal mol^{-1}, e^-, and *D*, respectively.

Compound	$V_{\sigma,max}$	q_s	μ
$H_2C{=}S$	−3.1	−0.12	2.12
$F_2C{=}S$	12.6	−0.11	0.31
Ph-SB$_{11}$	26.7	0.13	3.29
Cl-SB$_{11}$	30.7	0.13	5.31[a]

[a] The experimental μ of **Cl-SB$_{11}$** is 5.49 D.[20]

The **Ph-SB$_{11}$** compound was synthesized and crystallized (see the Supporting Information). The crystal structure (Figure 2 A) showed several interaction motifs (Figure 2 D) which were investigated using advanced methods of quantum chemistry (QM) and compared with crystal structures of **Cl-SB$_{11}$** and **I-SB$_{11}$** (Figure 2).[20]

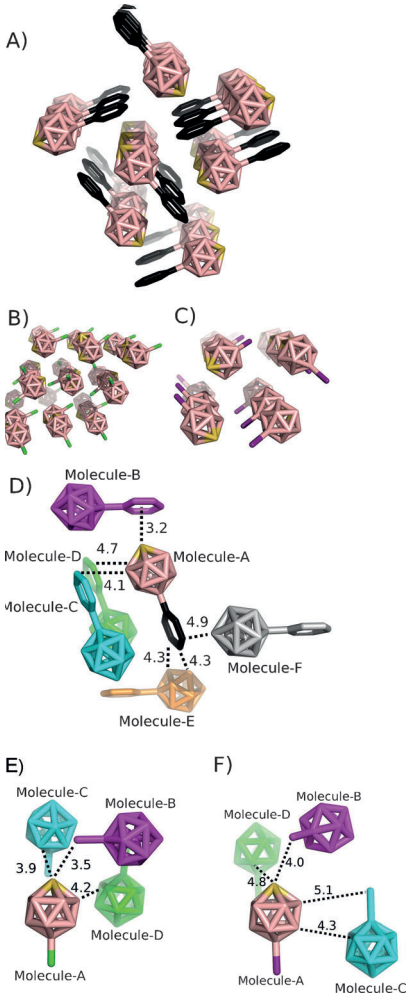


Figure 2. The crystal structures (A, B, C) and the most stable interactions (D, E, F) in the crystals of **Ph-SB$_{11}$** (A, D), **Cl-SB$_{11}$** (B, E), and **I-SB$_{11}$** (C, F). A, B, C and the central molecules in D, E, F colored by element (pink: B, yellow: S, black: C, green: Cl, magenta: I). H atoms omitted for clarity.

Table 2: The interaction energies (ΔE) and their decomposition into the electrostatic (E_1^{pol}), exchange-repulsion (E_1^{exch}), dispersion (E^{disp}), and induction (E_1^{ind}) terms; all in kcal mol^{-1}.

Motif	DFT-D3/CCSD(T)	DFT-SAPT				
	ΔE	E_1^{pol}	E_1^{exch}	E^{disp}	$E^{ind[a]}$	ΔE
		Ph-SB$_{11}$				
A-B	−8.2/−8.6	−5.8	10.8	−12.0	−2.0	−8.8
A-C	−7.4/−7.3	−2.8	9.4	−14.8	−0.8	−9.1
A-D	−5.5/−5.4	−2.4	7.5	−11.2	−0.5	−6.6
A-E	−2.6	−0.7	2.2	−4.3	−0.2	−3.0
A-F	−2.0	−0.4	1.5	−3.3	−0.2	−2.4
		Cl-SB$_{11}$				
A-B	−4.9	−3.5	5.6	−6.7	−0.9	−5.5
A-C	−3.9	−1.5	2.9	−6.1	−0.3	−5.0
A-D	−3.8	−1.5	3.9	−6.5	−0.5	−4.6
		I-SB$_{11}$				
A-B	−5.4	−3.5	6.6	−8.4	−1.1	−6.4
A-C	−4.3	−1.7	3.2	−6.7	−0.4	−5.6
A-D	−3.4	−0.8	3.4	−6.5	−0.5	−4.3

[a] $E^{ind} = E_2^{ind} + E_2^{exch\text{-}ind} + \delta HF$.

The interaction energies of the binding motifs found in the crystals are shown in Table 2. The strongest ΔE (−8.2 kcal mol^{-1}) was found for the A-B motif of **Ph-SB$_{11}$**, characterized by the B–S···π chalcogen bond. The DFT-D3 results were verified by the benchmark CCSD(T)/complete basis set (CBS) calculations, with a fair agreement. Notice that the B12-S axes of two **Ph-SB$_{11}$** molecules are not perpendicular (B12-S-Ph angle is 155°), in agreement with the prediction of nonlinearity of the chalcogen bond of the thiaboranes. The ΔE in the A-C and A-D stacking motifs of **Ph-SB$_{11}$**, which have no chalcogen bond, are weaker (−7.4 and −5.5 kcal mol^{-1}, respectively). The A-E and A-F motifs are considerably less stable because of the longer distances between Ph and the thiaborane cage.

Passing Cl or I, the chalcogen bonds are disfavored by about 3 kcal mol^{-1} compared to Ph. The B–S···π chalcogen bond is thus considerably more stable than the B–S···X one. The most stable motif features the chalcogen bond in all the crystals studied. The other motifs have head-to-tail and stacking interactions, and their ΔE are only slightly weaker.

The total ΔE is decomposed here using the DFT-SAPT technique in order to determine the nature of the respective binding (Table 2). The total ΔE values at the DFT-SAPT level are all slightly more negative than the DFT-D3 ones. In the case of **Ph-SB$_{11}$**, the A-C motif became more stable than the chalcogen-bonded A-B motif but based on the comparison with the CCSD(T) values this is probably an artifact of the method. The chalcogen bond in the motifs is dominated by the dispersion energy. The electrostatic stabilization is also important. Further, the induction energy is systematically larger for structures with a chalcogen bond, because of charge transfer in this motif.

To summarize, the most stable binding motif in the crystal of **Ph-SB$_{11}$** corresponds to a very strong B–S···π chalcogen bond exceeding −8 kcal mol^{-1}. It is considerably stronger than known C–X···π X-bonds. For comparison, the ΔE of the trifluoroiodomethane..benzene complex, possessing the X-bond, is −3.9 kcal mol^{-1}.[21] Dominant stabilization of the chalcogen bond investigated comes from dispersion and electrostatic energies. The phenyl group occurs frequently in proteins (in phenylalanine). The chalcogen bond can thus be used for designing protein–ligand interactions as well as for crystal engineering.

Experimental Section

Syntheses, NMR spectroscopies, X-ray diffraction analysis, and quantum mechanical calculations: The synthetic procedure of **Ph-SB$_{11}$** is based on the iodination of *closo*-1-SB$_{11}$H$_{11}$ followed by Negeshi coupling with Br-benzene. The shielding tensor of **Ph-SB$_{11}$** was computed and ^{1}H along with ^{11}B NMR spectra were recorded. The X-ray diffraction structure was established. QM calculations were performed using this structure. All the details are given in the Supporting Information.

Received: June 4, 2014
Published online: July 25, 2014

Keywords: boranes · chalcogen bonds · crystal structures · sulfur · X-ray diffraction

[1] P. Sanz, M. Yanez, O. Mo, *J. Phys. Chem. A* **2002**, *106*, 4661 – 4668.
[2] D. B. Werz, R. Gleiter, F. Rominger, *J. Am. Chem. Soc.* **2002**, *124*, 10638 – 10639.
[3] M. Iwaoka, S. Takemoto, S. Tomoda, *J. Am. Chem. Soc.* **2002**, *124*, 10613 – 10620.
[4] W. Wang, B. Ji, Y. Zhang, *J. Phys. Chem. A* **2009**, *113*, 8132 – 8135.
[5] T. Clark, M. Hennemann, J. S. Murray, P. Politzer, *J. Mol. Model.* **2007**, *13*, 291 – 296.
[6] M. Iwaoka, S. Takemoto, M. Okada, S. Tomoda, *Chem. Lett.* **2001**, 132 – 133.
[7] F. T. Burling, B. M. Goldstein, *J. Am. Chem. Soc.* **1992**, *114*, 2313 – 2320.
[8] Y. Nagao, T. Hirata, S. Goto, S. Sano, A. Kakehi, K. Iizuka, M. Shiro, *J. Am. Chem. Soc.* **1998**, *120*, 3104 – 3110.
[9] J. C. Taylor, G. D. Markham, *J. Biol. Chem.* **1999**, *274*, 32909 – 32914.
[10] K. Bender, I. Hennig, D. Schweitzer, *Mol. Cryst. Liq. Cryst.* **1984**, *108*, 359 – 371.
[11] a) M. Kolář, J. Hostaš, P. Hobza, *Phys. Chem. Chem. Phys.* **2014**, *16*, 9987 – 9996; b) R. Wilcken, M. O. Zimmermann, A. Lange, A. C. Joerger, F. M. Boeckler, *J. Med. Chem.* **2013**, *56*, 1363 – 1388.
[12] K. E. Riley, J. S. Murray, J. Fanfrlík, J. řezáč, R. J. Sola, M. C. Concha, F. M. Ramos, P. Politzer, *J. Mol. Model.* **2011**, *17*, 3309 – 3318.
[13] L. A. Hardegger, B. Kuhn, B. Spinnler, L. Anselm, R. Ecabert, M. Stihle, B. Gsell, R. Thoma, J. Diez, J. Benz, et al., *Angew. Chem.* **2011**, *123*, 329 – 334; *Angew. Chem. Int. Ed.* **2011**, *50*, 314 – 319.
[14] J. Fanfrlík, M. Kolář, M. Kamlar, D. Hurny, F. X. Ruiz, A. Cousido-Siah, A. Mitschler, J. Řezáč, E. Munusamy, M. Lepšík, et al., *ACS Chem. Biol.* **2013**, *8*, 2484 – 2492.
[15] a) N. S. Hosmane, *Boron Science*, CRC, Boca Raton, FL, **2011**; b) T. Baše, Z. Bastl, Z. Plzák, T. Grygar, J. Plešek, V. Malina, J. Šubrt, J. Boháček, E. Večerníková, O. Kříž, *Langmuir* **2005**, *21*, 7776 – 7785.
[16] P. Cígler, M. Kožíšek, P. Řezáčová, J. Brynda, Z. Otwinowski, J. Pokorná, J. Plešek, B. Grüner, L. Dolečková-Marešová, M. Máša, et al., *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15394 – 15399.

[17] J. Brynda, P. Mader, V. Šícha, M. Fábry, K. Poncová, M. Bakardiev, B. Grüner, P. Cígler, P. Řezáčová, *Angew. Chem.* **2013**, *125*, 14005 – 14008; *Angew. Chem. Int. Ed.* **2013**, *52*, 13760 – 13763.
[18] J. Fanfrlík, M. Lepšík, D. Horinek, Z. Havlas, P. Hobza, *ChemPhysChem* **2006**, *7*, 1100 – 1105.
[19] Z. F. Chen, R. B. King, *Chem. Rev.* **2005**, *105*, 3613 – 3642.
[20] J. Macháček, J. Plešek, J. Holub, D. Hnyk, V. Všetečka, I. Císařová, M. Kaupp, B. Štíbr, *Dalton Trans.* **2006**, *8*, 1024 – 1029.
[21] J. Řezáč, K. E. Riley, P. Hobza, *J. Chem. Theory Comput.* **2012**, *8*, 4285 – 4292.

Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

69451 Weinheim, Germany

The Dominant Role of Chalcogen Bonding in the Crystal Packing of 2D/3D Aromatics**

Jindřich Fanfrlík, Adam Přáda, Zdeňka Padělková, Adam Pecina, Jan Macháček, Martin Lepšík, Josef Holub, Aleš Růžička, Drahomír Hnyk,* and Pavel Hobza**

anie_201405901_sm_miscellaneous_information.pdf

Syntheses

0.5 g (3 mmol) of *closo*-1-$SB_{11}H_{11}$ obtained from Katchem was transformed to **I-SB_{11}** according to Reference 1 and the iodine atom was substituted with phenyl by Negishi coupling.[2] The coupling was conducted in dry THF. 1.2 ml of bromobenzene (4 times excess) was refluxed in 20 ml of dry THF with 0.3 g of Mg filings to produce PhMgBr. 1.7 g of $ZnCl_2$ was dissolved in 10 ml of dry THF added slowly to the PhMgBr solution and then refluxed. The solution of phenylzinc reagent was separated from the white precipitate of $MgCl_2$. The **I-SB_{11}** was dissolved in 10 ml of dry THF, and a catalytic amount of [$Pd(PPh_3)_2Cl_2$] was added, resulting in lightly orange solution. Then the PhZnBr solution was poured into the reaction mixture. Under reflux the mixture turned dark brown and a lighter precipitate was formed. After about 1 hour of refluxing,the reaction mixture was poured slowly into a stirred mixture of 40 ml of distilled water and 20 ml of concentrated HCl, covered by 20 ml of hexane. Then the top clear layer of hexane was separated from the water layer and a thin dark organic layer. Hexane was partially evaporated, until the precipitation of a solid residue started, than it was boiled under refluxed and left crystalise. The crystals of **Ph-SB_{11}** were subsequently formed.

X-Ray crystallography of Ph-SB_{11}

Methods. Data for colourless crystal were collected at 150(1)K on a Nonius KappaCCD diffractometer using MoK_α radiation (λ= 0.71073 Å), and graphite monochromator. The structures were solved by direct methods (SIR92).[3] All reflections were used in the structure refinement based on F^2 by full-matrix least-squares technique (SHELXL97).[4] Hydrogen atoms were mostly localized on a difference Fourier map, however to ensure uniformity of treatment of crystal, all hydrogens were recalculated into idealized positions (riding model) and assigned temperature factors of $1.5U_{eq}$ (pivot atom). Absorption correction was carried on using Gaussian integration from crystal shape (Coppens).[5]

Crystallographic data for $C_6H_{15}B_{11}S$, M = 238.15, monoclinic, $P2_1/c$, a = 6.6800(3), b = 14.0860(9), c = 15.2001(9), β = 110.178(5), Z = 4, V = 1342.47(14) $Å^3$, D_c = 1.178 $g.cm^{-3}$, μ = 0.204 mm^{-1}, T_{min} = 0.960, T_{max} = 0.973; 13574 reflections measured (Θ_{max} = 27.5°), 3061 independent (R_{int} = 0.028), 2605 with $I > 2\sigma(I)$, 163 parameters, S = 1.114, R_1(obs. data) = 0.0395, wR_2(all data) = 0.0935; max., min. residual electron density = 0.239, -0.278 $eÅ^{-3}$. CCDC Deposition number: 1000796.

A full list of crystallographic data and parameters including fractional coordinates is deposited with the Cambridge Crystallographic Data Center, 12 Union Road, Cambridge, CB2 1EZ, UK [Fax: int. code +44(1223)336-033; e-mail: deposit@ccdc.cam.ac.uk].

Results. Structure of **Ph-SB_{11}** is shown Figure S1. Note that C-C(B12)-C *ipso* angle in phenyl ring is computed to be 117.8 degrees at MP2 level. An angle which is typical of silicon substitution in benzene derivatives.[6] Almost the same value of this angle was computed in the hypothetical C_6H_5-BH_2.[7] Hammett constants for silicon substituents suggest them to act as slight electron donors. This geometric criterion offers the same behavior for **Ph-SB_{11}**.

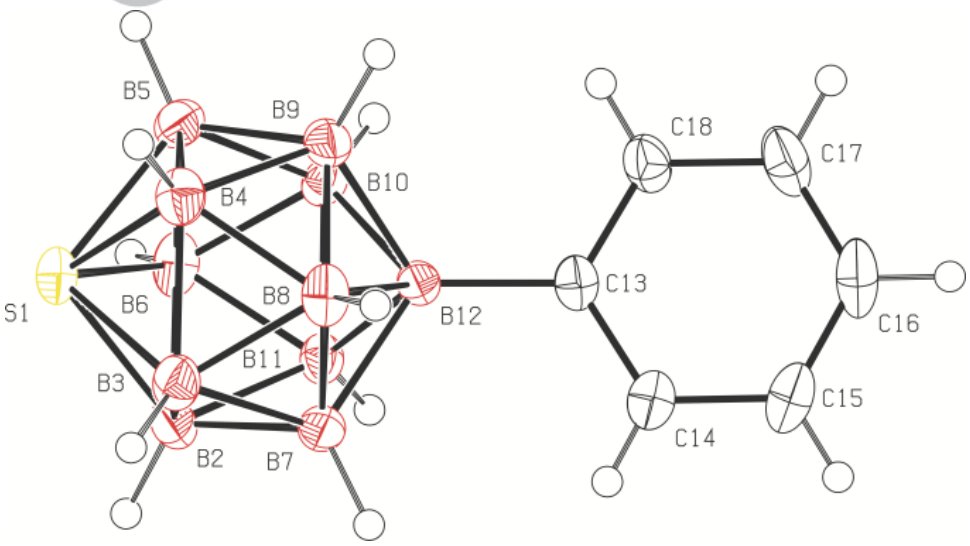


Figure S1: Molecular structure (ORTEP 50% probability level) of P**h-SB$_{11}$**. Selected interatomic distances [Å] and angles [°]: S1 B3 1.9932(19), S1 B5 1.9974(18), S1 B4 1.9976(19), S1 B2 2.0023(19), S1 B6 2.0124(18), B12 C13 1.580(2), B3 S1 B5 98.86(8), B3 S1 B4 56.22(8), B5 S1 B4 55.98(8), B3 S1 B2 55.96(8), B5 S1 B2 98.43(7), B4 S1 B2 98.84(8), B3 S1 B6 98.39(8), B5 S1 B6 55.90(8), B4 S1 B6 98.64(7), B2 S1 B6 55.42(8).

^{11}B and ^{1}H NMR Chemical Shifts

Measurements. Shielding tensor was calculated and ^{11}B and ^{1}H NMR spectra were measured on a Varian Mercury Plus 400 NMR spectrometer under standard conditions on freshly prepared samples, using deuterated chloroform as solvent. ^{11}B chemical shifts are given relative to $BF_3 \cdot OEt_2$, ^{1}H chemical shifts relative to TMS. The ^{11}B signals were assigned to the individual positions in the molecule on the basis of ^{11}B –^{11}B correlated spectrum.

δ (^{11}B) exp./calc.: 26.3/28.9 ppm (B12), -4.6/-4.7 ppm (B7–B9), -7.2/-6.8 ppm(B2–B6). Note that NMR is a "slow" technique and detects only three signals as if the molecule was C_{5v}-symmetrical. In accordance with this observation, calculated values are C_{5v}-symmetrized.

Calculations. The so-called antipodal atom B(12) in the parent *closo*-1-$SB_{11}H_{11}$ resonates at 18.6 ppm in its ^{11}B NMR spectrum. The same atom in **Ph-SB_{11}** resonates at 26.3 ppm in its NMR pattern. Such a shift to higher frequency (ca. 8 ppm) is caused by deshielding of this atom in the latter with respect to the former. Electron transfer from the 3D aromatic towards the 2D aromatic benzene ring can account for this experimental observation, well reproduced by the calculations. Note that the so-called nucleus-independent chemical shifts, acting as one of the aromaticity criterion, are computed to be -33.1 ppm and -15.0 ppm using GIAO-MP2/II level of theory for the parent 3D and 2D aromatics, respectively. When these two aromatics are connected to form **Ph-SB_{11}**, the corresponding NICS values are computed to be -29.4 ppm and -11.6 ppm, respectively. All four values strongly support aromatic behavior of the separate and connected thiaborane and benzene ring and also indicate electron transfer between these two moieties.

QM calculations

Using the crystal structures of **Ph-SB$_{11}$** (see above), **Cl-SB$_{11}$** and **I-SB$_{11}$**[1] all the molecules within 5 Å of the central molecules were considered. In the obtained crystal model (about 300 atoms), the H atoms were optimized using DFT-D/BLYP/SVP method.[8] Subsequently, interaction energy of the central molecule with each surrounding molecule was evaluated at DFT-D3/TPSS/TZVPP[9] (with the ecp-28-mdf-TZVPP pseudopotential for iodine) using the Turbomole 6.3 program.[10]

Interaction energy was decomposed by the density functional theory based symmetry adapted perturbation theory (DFT-SAPT)[11] in a hybrid scheme, where dispersion is modeled empirically,[12] DFT part of SAPT is treated using the localized and asymptotically corrected LPBE0AC exchange-correlation functional with density fitting and the aug-cc-pVDZ basis set. A gradient-controlled shift is obtained by PBE1PBE/aug-cc-pVDZ and PBE1PBE/TZVP calculations. The frozen core approximation was used for all calculations and pseudopotential for iodine was used to cover relativistic effects.

DFT-D3 interaction energies were compared to benchmark CCSD(T)/CBS [13] for the most stable motifs of **Ph-SB$_{11}$**. CCSD(T)/CBS calculated as the sum of HF/CBS energy and MP2/CBS correlation energy, both extrapolated from aug-cc-pVDZ and aug-cc-pVTZ basis sets. The CCSD(T) correction term is calculated with modified 6-31G* basis set for which the exponents of polarization functions are changed from their original values 0.8 to 0.25, 0.6 to 0.19 and 0.65 to 0.20 for carbon, boron and sulfur, respectively. For more details see the original paper.[14] Counterpoise corrections for basis set superposition error (BSSE) are used for all calculations and density fitting is used for acceleration of MP2 and HF calculations.

Magnetic shielding was calculated using the GIAO-MP2 method incorporated into Gaussian09 utilizing the IGLO-II basis with the MP2 optimized structures and frozen core electrons.

Electrostatic potentials were computed at HF/cc-pVDZ level using Gaussian09[15] and Molekel4.3[16] programs.

References

[1] J. Machacek, J. Plesek, J. Holub, D. Hnyk, V. Vsetecka, I. Cisarova, M. Kaupp, B. Stibr, *Dalton Trans.* **2006**, *8*, 1024–1029.

[2] A. O. King, N. Okukado, E. Negishi, *J. Chem. Soc. Chem. Comm.* **1977**, 683–684.

[3] A. Altomare, G. Cascarone, C. Giacovazzo, A. Guagliardi, M.C. Burla, G. Polidori,

M.Camalli, *J. Appl. Crystallogr.* **1994**, *27*, 1045.

[4] G. M. Sheldrick, SHELXL-97, A Program for Crystal Structure Refinement. University of Göttingen, Germany (**1997**).

[5] P. Coppens in: F. R. Ahmed, S. R. Hall, C. P. Huber (Eds.), Crystallographic Computing, Copenhagen, Munksgaard, **1970**, 255–270.

[6] A. Domenicano, A. Vaciago, C. A. Coulson *Acta Crystallogr.* **1975**, *B31*, 1630–1641.

[7] A. R. Campanelli, A. Domenicano, F. Ramondo, *J. Phys. Chem. A.* **2003**, *107*, 6429–6440.

[8] P. Jurecka, J. Cerny, D. R. Salahub, P. Hobza, *J. Comput. Chem.* **2007**, *28*, 555-569.

[9] S. Ehrlich, J. Moellmann, W. Reckien, T. Bredow, S. Grimme, *ChemPhysChem,* **2011**, *12*, 3414-3420.

[10] R. Ahlrichs, M. Bar, M. Haser, H. Horn, C. Kolmel, *Chem. Phys. Lett.* **1989**, *162*, 165-169.

[11] a) B. Jeziorski, R. Moszynski, K. Szalewicz, *Chem. Rev.* **1994**, *94*, 1887–1930. b) G. Jansen, A. Hesselmann, *J. Phys. Chem. A* **2001**, *105*, 11156–11157. c) A. Hesselmann, G. Jansen, M. Schutz, *J. Chem. Phys.* **2005**, *122*. d) A. Hesselmann, G. Jansen, *Chem. Phys. Lett.* **2002**, *357*, 464–470. e) A. Hesselmann, G. Jansen, M. Schutz, *J. Am. Chem. Soc.* **2006**, *128*, 11730–11731. f) A. Hesselmann, G. Jansen, *Chem. Phys. Lett.* **2003**, *367*, 778–784. g) A. Hesselmann, G. Jansen, *Chem. Phys. Lett.* **2002**, *362*, 319–325. h) H. L. Williams, C. F. Chabalowski, *J. Phys. Chem. A* **2001**, *105*, 646–659. i) A. J. Misquitta, K. Szalewicz, *Chem. Phys. Lett.* **2002**, *357*, 301–306. j) R. Podeszwa, R. Bukowski, K. Szalewicz, *J. Phys. Chem. A* ***2006***, *110*, 10345–10354. k) S. Rybak, K. Szalewicz, B. Jeziorski, G. Corongiu, Chem. Phys. Lett. **1992**, *199*, 567–573.

[12] A . Heselmann, *J. Phys. Chem. A* **2011**, *115*, 11321–11330.

[13] P. Jurecka, P. Hobza, *Chem. Phys. Lett.* **2002**, *365*, 89-94.

[14] P. Hobza, J. Sponer, *Chem. Rev.* **1999**, *99*, 3247 – 3276.

[15] Gaussian 09, Revision D.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, et al. Gaussian, Inc., Wallingford CT, **2009**.

[16] a) MOLEKEL 4.3, P. Flükiger, H.P. Lüthi, S. Portmann, J. Weber, Swiss Center for Scientific Computing, Manno (Switzerland), 2000-2002. b) S. Portmann, H. P. Lüthi. MOLEKEL: *CHIMIA*, **2007**, *28*, 555-569.

Appendix C

THE JOURNAL OF PHYSICAL CHEMISTRY A

Article

pubs.acs.org/JPCA

Chalcogen and Pnicogen Bonds in Complexes of Neutral Icosahedral and Bicapped Square-Antiprismatic Heteroboranes

Adam Pecina,[†] Martin Lepšík,[†] Drahomír Hnyk,*[,‡] Pavel Hobza,[†,§] and Jindřich Fanfrlík*[,†]

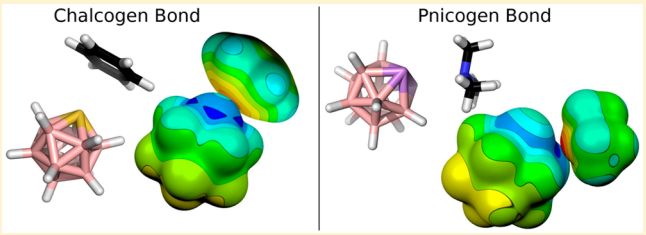
[†]Gilead Sciences and IOCB Research Center and Institute of Organic Chemistry and Biochemistry (IOCB), Academy of Sciences of the Czech Republic, v.v.i., Flemingovo nám. 2, 16610 Prague 6, Czech Republic

[‡]Institute of Inorganic Chemistry, Academy of Sciences of the Czech Republic, v.v.i., 250 68 Řež near Prague, Czech Republic

[§]Regional Center of Advanced Technologies and Materials, Department of Physical Chemistry, Palacký University, 77146 Olomouc, Czech Republic

Supporting Information

ABSTRACT: A systematic quantum mechanical study of σ-hole (chalcogen, pnicogen, and halogen) bonding in neutral experimentally known *closo*-heteroboranes is performed. Chalcogens and pnicogens are incorporated in the borane cage, whereas halogens are considered as *exo*-substituents of dicarbaboranes. The chalcogen and pnicogen atoms in the heteroborane cages have partial positive charge and thus more positive σ-holes. Consequently, these heteroboranes form very strong chalcogen and pnicogen bonds. Halogen atoms in dicarbaboranes also have a highly positive σ-hole, but only in the case of C-bonded halogen atoms. In such cases, the halogen bond of heteroboranes is also strong and comparable to halogen bonds in organic compounds with several electron-withdrawing groups being close to the halogen atom involved in the halogen bond.



1. INTRODUCTION

The noncovalent interactions of halogens, chalcogens, and pnicogens (abbreviated as X, E, Pn, respectively) are known to be directional. This is caused by the anisotropic distribution of electrons in the *p*-orbitals when these elements are covalently bound. This gives rise to a region of positive electrostatic potential (ESP) called σ-hole,[1] located on the axis of the covalent bond and distal from the bonded atom (Figure 1). For

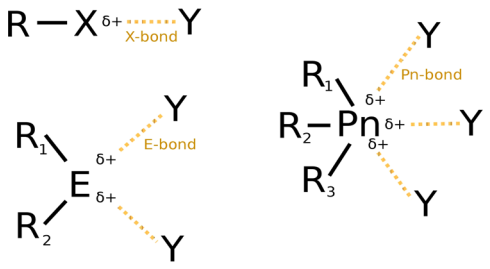


Figure 1. Schematic figure of halogen (X), chalcogen (E), and pnicogen (Pn) bonds. Y stands for an electron donor.

the most extensively studied halogen bond (X-bond), the σ-hole[2] acts as a X-bond donor and enables the X-bond formation. X-bonds play an important role in molecular recognition, crystal engineering[3,4] and interactions of drugs or biological molecules.[5−9] Since the E and Pn atoms are multivalent, the respective σ-holes are localized on all the axes of the covalent bonds, distal to the bonded atom, contrary to the monovalent X atoms, which have the σ-hole localized at the top of the atom. Although the interactions of these elements have not yet attracted as much attention as X-bonding, their importance is expected to be growing.[2] Chalcogen bonds (E-bonds) play a role in crystal engineering and in drug design.[10−16] An analysis of the Protein Data Bank suggests that the E-bonds also influence protein structures.[11,12] Recently, pnicogen bonds (Pn-bond) have been recognized as new supramolecular linkers.[17,18]

A σ-hole is characterized by its magnitude ($V_{S,max}$) and size.[19] $V_{S,max}$ is defined as the value of the most positive ESP of an electron (e^-) density surface and the size as the spatial extent of the positive region. $V_{S,max}$ and the size increase with the atomic number. $V_{S,max}$ can also be increased by electron-withdrawing groups in the vicinity of the X, E, or Pn atom. As also noticed, the higher the $V_{S,max}$ value, the stronger the σ-bond.[20−22] Apart from small model complexes, this has already been demonstrated in more complex biomolecular systems. A modulation of the X-bond in protein-inhibitor complexes has been used to reduce the IC_{50} values accordingly. Specifically, Cl-to-Br and Cl-to-I substitutions have enhanced the X-bond in the cathepsin-inhibitor complex and reduced IC_{50} from 30 nM to 6.5 and 4.3 nM.[23] In addition, we have recently modulated the IC_{50} of aldose reductase inhibitors from 1900 nM to 190 nM by fluorination close to the X atom that is involved in a X-bond.[24]

Received: November 6, 2014
Revised: January 30, 2015
Published: January 30, 2015

DOI: 10.1021/jp511101n
J. Phys. Chem. A 2015, 119, 1388−1395

Detailed understanding of sigma-hole bonding has been gained using quantum chemical (QM) computations. Hartree–Fock (HF) and density functional theory (DFT), which properly describe the σ-hole, usually give too low stabilization energies for σ-hole bonded complexes.[21] The DFT-D3/TPSS/TZVPP method has already been shown to reliably describe B–S$\cdots\pi$ type E-bonds.[25] However, reference interaction energy (ΔE) is only obtained using the highly accurate but also time-consuming CCSD(T) calculations.

It is known that binary boron hydrides (boranes) exhibit an astonishing variety of stable three-dimensional structures. Their building blocks are triangles of boron atoms which are held together by delocalized electron-deficient three-center two-electron (3c2e) bonding.[26,27] An important class of boranes comprises [*closo*-B_nH_n]$^{2-}$ (known for n = 5–12) dianions. The systematic replacement of BH vertices (formally neutral) leads to a variety of *closo* heteroboranes with different total charges. For example, formal incorporation of a single S^{2+}, Se^{2+}, or CH^+ vertex gives rise to the neutral *closo*-1-SB_nH_n, *closo*-1-SeB_nH_n, or anionic [*closo*-1-CB_nH_{n+1}]$^-$ heteroboranes, respectively. Since incorporation a single V group element into icosahedral cage is difficult to achieve experimentally, we opted for diphospha- and diarsaboranes. Conceivably, P–P and As–As linkages make the overall charge neutral as well as two CH^+ groups. A combination of CH^+ and P^+ makes the overall charge neutral, too, as exemplified for n = 10. As opposed to the classical electronegativity concept, the heteroatoms incorporated in the borane cages are centers of positive charge(s), as also proved experimentally.[28,29] Terminal hydrogens in all these heteroboranes, which due to the electropositivity of boron bear a partial negative charge (hydridic character), can be replaced by substituents, such as X atoms or aryls. It should be mentioned that heteroboranes have already been used in radioactive waste extraction, nanotechnology, and medicinal chemistry.[30−33] The properties which make boranes such suitable entities include their hydrophobicity, shape, 3D aromaticity, stability, and ability to form dihydrogen bonds.[34−36]

On that basis, we have recently tackled an inherently electron-deficient heteroborane in order to examine the E-bond effects first. Namely, we synthesized and crystallized the phenyl-substituted thiaborane (12-Ph-*closo*-1-$SB_{11}H_{10}$ abbreviated as **Ph-SB_{11}**) and observed the formation of S$\cdots\pi$ type E-bonds.[25] Quantum chemical analysis revealed that $V_{S,max}$ is more positive for these inorganic S-containing molecules than for organic ones. The reason is the existence of five highly positive σ-holes on the already positively charged pentacoordinated S atom. Consequently, these interactions are considerably stronger than those in their organic counterparts and in the known X-bond. In order to gain a deeper insight into the nature of these noncovalent interactions, we report here a systematic quantum mechanical study of σ-hole bonding in neutral experimentally known icosahedral and square-antiprismatic *closo*-heteroboranes, in which carbon, E and Pn atoms are incorporated in the borane cage. In the case of dicarbaboranes, X *exo*-substituents are also considered.

The ability to form very strong σ-hole bonds might significantly increase the range of applications of heteroboranes.

2. METHODS

2.1. The Systems Studied. In this paper, we chose these heteroborane molecules: *closo*-1-$SB_{11}H_{11}$, 12-F-*closo*-1-$SB_{11}H_{10}$, 12-Cl-*closo*-1-$SB_{11}H_{10}$, 12-Br-*closo*-1-$SB_{11}H_{10}$, *closo*-1-$SeB_{11}H_{11}$, *closo-1-SB_9H_9*, *closo*-1,2-$P_2B_{10}H_{10}$, *closo*-1,2-$As_2B_{10}H_{10}$, *closo*-2,1-PCB_8H_9, *closo*-6,1-PCB_8H_9, 12-Br-closo-1,2-$C_2B_{10}H_{11}$, and 1-Br-*closo*-1,2-$C_2B_{10}H_{11}$ (abbreviated as **SB_{11}**, **12-F-SB_{11}**, **12-Cl-SB_{11}**, **12-Br-SB_{11}**, **SeB_{11}**, **SB_9**, **1,2-P_2B_{10}**, **1,2-As_2B_{10}**, **2,1-PCB_8**, **6,1-PCB_8**, **12-Br-C_2B_{10}**, and **1-Br-C_2B_{10}**, respectively; numbering of 10 and 12 vertex cages shown in Figure 2A).

For selected boron clusters (**SB_{11}**, **12-Cl-SB_{11}**, **SeB_{11}**, **SB_9**, **1,2-P_2B_{10}**, **1,2-As_2B_{10}**, **12-Br-C_2B_{10}** and **1Br-C_2B_{10}**), we studied noncovalent complexes with five σ-hole acceptors—specifically with benzene (**BEN**), trimethylamine (**TMA**), dimethyl ether (**DME**), acetone (dimethyl ketone **DMK**), and formamide (**FA**). In order to find the minimum of the

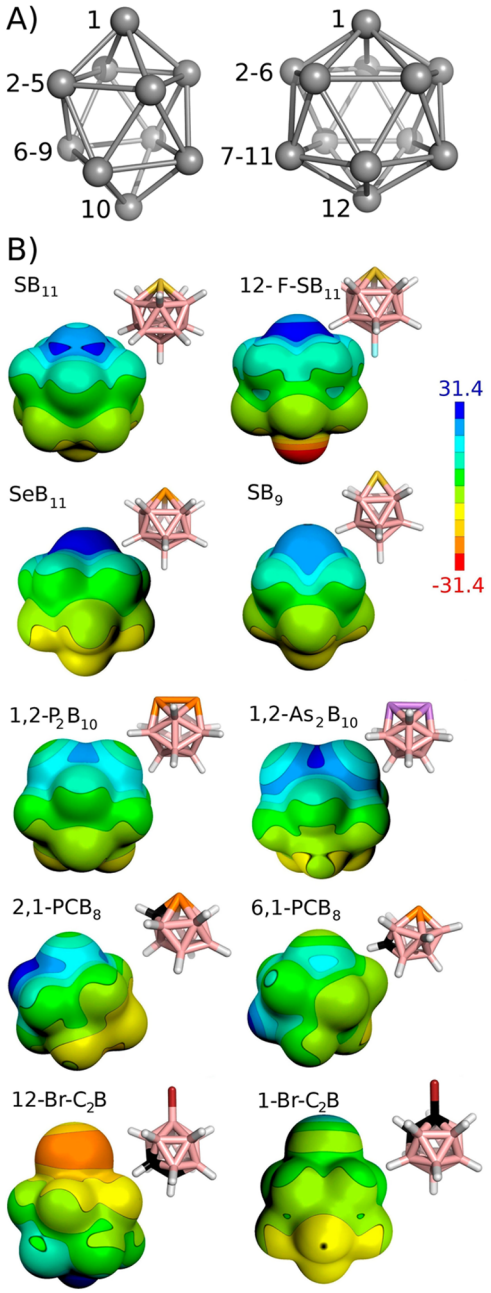


Figure 2. (A) Structure and numbering scheme of 10 and 12 vertex cages; (B) Schematic figures and computed electrostatic potentials on 0.001 au molecular surfaces of the studied molecules. Color of ESP ranges in kcal/mol and colors of atoms are as follows: black, C; light-pink, B; yellow, S; yellow-orange, Se; cyan, F; white, H; orange, P; purple-blue, As; dark-red, Br.

DOI: 10.1021/jp511101n
J. Phys. Chem. A 2015, 119, 1388−1395

complexes studied, we optimized them with various fixed angles of σ-hole bonds ranging from 90° to 180° with a 5° step.

2.2. The Setup of the Calculations. The structures and geometries of all the studied systems were obtained by the density functional theory augmented with empirically parametrized dispersion (DFT-D3) with the default zero-damping function,[37] the TPSS functional[38] and TZVPP[39] basis set. For gradient optimizations, we used the LBFGS algorithm with the strict optimization criteria (energy change <0.0006 kcal mol^{-1}, the largest gradient component <0.12 kcal mol^{-1} Å^{-1} and the root-mean-square gradient <0.06 kcal mol^{-1} Å^{-1}). Vibrational frequencies were calculated numerically at the above-mentioned level to confirm that the complexes represented the true minima. Reference interaction energy (ΔE) was obtained using the highly accurate and time-consuming CCSD(T) calculations.

The CCSD(T)/complete basis set (CBS) limit interaction energy (ΔE) was determined using a previously described extrapolation scheme (eq 1).[40]

$$\Delta E_{CBS}^{CCSD(T)} = \Delta E_{CBS}^{MP2} + (\Delta E^{CCSD(T)} - \Delta E^{MP2})_{aug-cc-pVDZ} \quad (1)$$

Here the MP2/CBS interaction energy was calculated as the sum of HF/CBS ΔE and MP2/CBS correlated ΔE, extrapolated to the CBS using two-point extrapolation methods[41] with the aug-cc-pVDZ and aug-cc-pVTZ basis sets of Dunning. The CCSD(T) correction term, determined as a difference between CCSD(T) and MP2 ΔE, was evaluated with the aug-cc-pVDZ basis set.[40]

The interaction energy was decomposed by using the density functional theory-based symmetry-adapted perturbation theory (DFT-SAPT).[42−45] The inaccurate energies of the virtual orbitals obtained when using the DFT method were corrected by a gradient-controlled shift procedure. PBE1PBE/aug-cc-pVDZ and PBE1PBE/TZVP calculations were carried out to obtain the desired shift value. The DFT part was treated using the localized and asymptotically corrected LPBE0AC exchange-correlation functional with density fitting and the aug-cc-pVDZ basis set. This combination of the functional and the basis set has been shown to provide a reasonably good description of electrostatic and induction energies, but the dispersion term is underestimated by approximately 10−20%.[46] For the present complexes, the use of a larger basis set (e.g., aug-cc-pVTZ) for all the studied complexes is too demanding. We have thus calculated DFT-SAPT/CBS (using two-point extrapolation methods[41] with the aug-cc-pVDZ and aug-cc-pVTZ basis sets) only for the **SB$_9$···FA** complex. The obtained scaling factor, specific for heteroborane cages, of 1.148 was used to scale the dispersion energy in order to obtain results comparable with CBS data.

The total ΔE 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. The first two terms represent polarization (electrostatic) (E_1^{Pol}), induction (E_2^{Ind}), and dispersion (E_2^{D}) together with exchange-repulsion terms (E_1^{Ex}, $E_2^{Ex\text{-}Ind}$ and $E_2^{Ex\text{-}D}$), whereas the δHF term represents higher than second-order terms covered by the Hartree−Fock approach.

Core electrons were kept frozen for intermolecular correlation contributions and ΔE was corrected for the basis set superposition error (BSSE).[47]

ESP was calculated on isolated molecules at the HF/cc-pVDZ level of theory.

All the calculations were carried out using the Turbomole 6.3 program suite,[48] Gaussian09,[49] and Molpro 2010[50] quantum chemistry programs.

3. RESULTS AND DISCUSSION

3.1. ESP of Isolated Molecules. The calculated ESPs of isolated heteroboranes are shown in Figure 2B. $V_{S,max}$ and dipole moments are summarized in Table 1. The results

Table 1. Magnitude of σ-Holes ($V_{S,max}$) and the Dipole Moments (μ) in kcal mol^{-1} and *D*, respectively

compound	$V_{S,max}$	μ
	Chalcogens	
SB$_{11}$	28.2	3.4
SeB$_{11}$	29.5	4.1
12-Br-SB$_{11}$	30.4	5.5
12-Cl-SB$_{11}$	30.7	5.3
12-F-SB$_{11}$	29.2	4.8
SB$_9$	22.4	3.5
	Pnicogens	
1,2-P$_2$B$_{10}$	22.6	2.7
1,2-As$_2$B$_{10}$	26.9	3.6
2,1-PCB$_8$	20.7	3.6
6,1-PCB$_8$	15.7	1.9
	Halogens	
1-Br-C$_2$B$_{10}$	27.0	3.3
12-Br-C$_2$B$_{10}$	−8.8	6.4

indicate that the $V_{S,max}$ of the σ-holes of chalcogen (E) and pnicogen (Pn) atoms are more positive in boranes than in organic molecules.[25] The borane clusters are electron deficient and heteroatom vertices are positively charged. The σ-holes on E and Pn atoms in boranes are thus areas with highly positive ESP on already positively charged E and Pn atoms. In the cases of **1,2-P$_2$B$_{10}$** and **1,2-As$_2$B$_{10}$**, the most positive ESP is not located at the Pn atoms themselves, but in the valley between them. This might play an important role in the structure of Pn-bonded complexes. The $V_{S,max}$ of σ-holes can be further modulated in several ways: by changing the atomic number of the E and Pn atoms, by changing the skeleton of the borane cage and by changing the chemical environment (introducing *exo*-substituents). S-to-Se and P-to-As substitutions increased the value of $V_{S,max}$ by about 1.3 and 4.3 kcal mol^{-1}, respectively. The $V_{S,max}$ on the S atom was also higher by 5.9 kcal mol^{-1} in the 12-vertex cage than in the ten-vertex cage. In **PCB$_8$** compounds, $V_{S,max}$ depends significantly on the position of the CH vertex. If the P vertex is not next to the CH vertex, $V_{S,max}$ strongly decreases. Finally, H-to-X substitutions in the *para* position to the heteroatom increase the value of $V_{S,max}$ by about 1.8 kcal mol^{-1} on average. Interestingly, fluorination has a smaller effect than chlorination and bromination. The changes in $V_{S,max}$ are supported by the calculated dipole moments. Compounds with the most positive $V_{S,max}$ (30.7 and 30.4 kcal mol^{-1} for **12-Cl-SB$_{11}$** and **12-Br-SB$_{11}$**, respectively) have the largest dipole moment (5.3 and 5.5 D, respectively), whereas the compound with the least positive $V_{S,max}$ (15.7 kcal mol^{-1} for **6,1-PCB$_8$**) has the smallest dipole moment (1.9 D) .

In contrast to E and Pn atoms incorporated in heteroborane cages, X atoms appear as *exo*-substituents in the case of dicarbaboranes. Like in PCB$_8$ molecules, the values of $V_{S,max}$ depend significantly on the position of the C vertex in the carborane cage (i.e., to which vertex the X atom is bound). The

DOI: 10.1021/jp511101n
J. Phys. Chem. A 2015, 119, 1388−1395

B-bound X atom is negatively charged and the σ-hole is an area of a less negative ESP. On the other hand, the C-bound X atom is positively charged and, consequently, the σ-hole is highly positive.

3. 2. σ-Hole Bonded Complexes. *Chalcogen Bond (E-Bond).* The SB_{11}···benzene (**BEN**) complex was used as a simple model of 12-Ph-1-$SB_{11}H_{10}$ (**12-Ph-SB_{11}**), which has only recently been shown to form very strong B−S···π type E-bonds.[25] To study this complex, we optimized the **SB_{11}···BEN** complex with various fixed values of the B12−S−Phe angle (measured to the centroid of the **BEN** ring, from 90° to 180°). The results (Figure S1) show that the minimum of the B12−S-**BEN** angle is about 165° (Figure 3A). The minimum is broad,

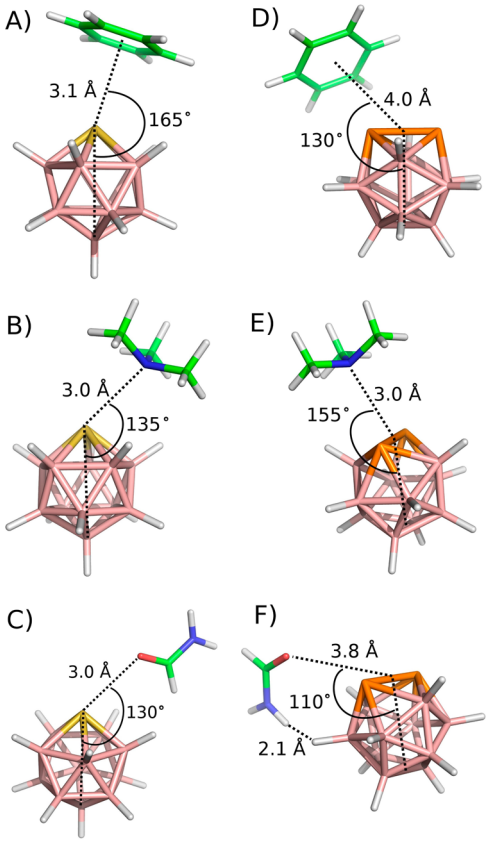


Figure 3. Structure of the (A) **SB_{11}···BEN**, (B) **SB_{11}···TMA**, (C) **SB_{11}···FA**, (D) **1,2-P_2B_{10}···BEN**, (E) **1,2-P_2B_{10}···TMA**, and (F) **1,2-P_2B_{10}···FA** complexes. Color of atoms as follows: green, C; light-pink, B; yellow, S; white, H; orange, P.

and energy increases only when the B12−S···**BEN** angle drops below 150°. This finding is in agreement with the experimental value of 155° in X-ray structure of **12-Ph-SB_{11}**. The ΔE of −6.2 kcal mol^{-1} is close to the value reported for **12-Ph-SB_{11}** (−8.2 kcal mol^{-1}),[25] which confirms the dominant role of E-bonding in **12-Ph-SB_{11}** crystal packing.

Besides the E-bond to **BEN**, we have also considered other E-bond acceptors−trimethylamine (**TMA**), dimethyl ether (**DME**), acetone (dimethyl ketone, **DMK**), and formamide (**FA**). The ΔE of the studied complexes of **SB_{11}** is summarized in Table 2, and structures of selected complexes are shown in Figure 3, parts B and C. The data show that the optimum angle of the E-bonds of **SB_{11}** ranges between 130° and 165°, which corresponds to the positions of the σ-holes. The strength of the E-bond varies according to the chalcogen-bond acceptor. The best E-bond acceptors are **BEN** and **TMA** while the weakest is **DME** (ΔE of −6.2, −6.6, and −4.2 kcal mol^{-1}, respectively). The frequency calculations showed that the obtained structures are true minima. All ΔE are due to the σ-hole interactions (no dihydrogen bond shorter than 2.5 Å). The DFT-D3 results were verified by the benchmark CCSD(T)/CBS calculations, with a fair agreement. The ΔE was decomposed using the DFT-SAPT method to determine the nature of the binding of **SB_{11}** to the E-bond acceptors. Like in the case of the B−S··· π type E-bond reported recently,[25] dispersion also plays a dominant role in the other **SB_{11}** complexes studied. Polarization (electrostatic) energy appears to be the second most important contribution to the overall ΔE. In the case of the **SB_{11}···FA** complex, polarization (electrostatic) energy even becomes the dominant term. Induction energy is systematically the least important term, which indicates that charge transfer does not play any important role here.

As demonstrated above, the $V_{S,max}$ of σ-holes can be modulated by changing the atomic number of the E atom, by changing the skeleton of the borane cage and by changing the chemical environment (introducing *exo*-substitutions). Having obtained the knowledge of the energetic balance of the **SB_{11}**··· benzene (**BEN**) complex, we selected other experimentally known[28,51,52] icosahedral and also bicapped square-antiprismatic heteroboranes for forming such complexes in which the cluster moieties are represented by **SeB_{11}**, **12-Cl-SB_{11}**, and **SB_9**. The ΔE of these complexes is summarized in Table 3 and the Supporting Information. Even though **12-Cl-SB_{11}** has a higher value of $V_{S,max}$ than **SeB_{11}**, the **Cl-SB_{11}** complexes have ΔE comparable with the parental **SB_{11}** (ΔE by about 0.2 kcal mol^{-1} more negative on average). The S-to-Se substitution has a large impact on ΔE. Upon the S-to-Se substitution, the ΔE becomes more negative by 1.3 kcal mol^{-1} on average. SAPT analyses have shown that all the terms (polarization, dispersion and induction) become more negative. Finally, the change of the skeleton has only a small effect on the strength of the respective E-bond. The ΔE of **SB_9** complexes was by about 0.2 kcal mol^{-1} less negative than the ΔE of the **SB_{11}** complexes (see Table S1).

The results of this study show that the E-bonds are significantly stronger in heteroboranes than in the related organic complexes known in the literature. For example, the ΔE of the **BEN**···Se=CF_2, **BEN**···Se−$(CF_3)_2$, **TMA**···Se=CF_2 and **TMA**···Se−$(CF_3)_2$ complexes possessing the chalcogen bond amounts to −3.5, −4.5, −4.6, and −3.7 kcal mol^{-1}, respectively,[53] whereas the ΔE of the above-mentioned **SeB_{11}···BEN** and **SeB_{11}···TMA** complexes is −7.0 and −8.1 kcal mol^{-1}, respectively.

Pnicogen Bond (Pn-Bond). For the study of Pn-bonds, we selected the experimentally known complexes of **1,2-P_2B_{10}** and **1,2-As_2P_{10}**.[54] Their geometrical and energy data are summarized in Table 4 and Figure 3D−F. The DFT-D3 results of complexes with Pn-bond were also verified by the DFT-D frequency calculations and benchmark CCSD(T)/CBS calculations. In contrast to E-bonds, we have found two minima in the scan. It is most significant for the complex with **BEN** (see Figure S2). The Pn-bond is more bent in the most stable structure of the **1,2-P_2B_{10}···BEN** complex (an optimal angle about 130 deg). The other complexes have an optimal angle similar to the E-bonded complexes. The ΔE of the **1,2-P_2B_{10}** complexes is comparable to the ΔE of the **SB_{11}** complexes. The difference in ΔE was 0.6 kcal mol^{-1} on average (with E-bonds

DOI: 10.1021/jp511101n
J. Phys. Chem. A 2015, 119, 1388−1395

Table 2. Structural and Energetic Characteristics of SB_{11} Complexes[a]

			DFD-D3/CCSD(T)	DFT-SAPT				
complex	α	d	ΔE	E_1^{Pol}	E_1^{Ex}	E_2^{D}	E^{Ind}[b]	ΔE
SB_{11}···BEN	180	3.10	−6.0/−6.1	−5.5	9.6	−8.5	−2.0	−6.4
	165	3.12	−6.2/−6.3	−5.5	9.8	−8.8	−2.0	−6.5
SB_{11}···TMA	180	3.22	−4.0/−3.8	−4.0	6.1	−5.0	−1.1	−4.0
	135	2.96	−6.6/−6.5	−8.9	14.7	−9.4	−2.6	−6.3
SB_{11}···DME	180	3.22	−3.1/−3.3	−2.5	3.4	−3.2	−0.6	−3.0
	140	3.10	−4.2/−4.3	−3.5	4.4	−4.2	−0.9	−4.2
SB_{11}···DMK	180	3.21	−3.5/−3.4	−3.3	4.1	−3.5	−0.9	−3.6
	135	3.03	−5.1/−5.4	−5.4	6.7	−5.3	−1.5	−5.5
SB_{11}···FA	180	3.18	−3.4/−2.5	−3.4	3.5	−2.6	−0.9	−3.4
	130	3.02	−5.6/−5.9	−6.3	7.7	−5.4	−1.8	−5.8

[a]The B12−S−chalcogen bond acceptor angle (α) in degrees. The S···chalcogen bond acceptor distance (d) in Å. The interaction energy (ΔE) and its decomposition into electrostatic (E_1^{Pol}), exchange-repulsion (E_1^{Exc}), dispersion (E_2^{D}) and induction (E^{ind}) terms; energy in kcal mol^{-1}. [b]$E^{Ind} = E_2^{Ind} + E_2^{Ex\text{-}Ind} + \delta HF$

Table 3. Structural and Energetic Characteristics of SeB_{11} Complexes[a]

			DFD-D3/CCSD(T)	DFT-SAPT				
complex	α	d	ΔE	E_1^{Pol}	E_1^{Ex}	E_2^{D}	E^{Ind}[b]	ΔE
			SeB_{11}					
SeB_{11}···BEN	180	3.16	−6.8/−6.5	−6.5	10.5	−8.4	−2.7	−7.0
	160	3.17	−7.0	−6.6	10.9	−9.0	−2.6	−7.4
SeB_{11}···TMA	180	3.23	−4.9/−4.4	−5.6	7.4	−5.1	−1.6	−4.8
	130	2.94	−8.1	−12.2	18.8	−10.3	−4.0	−7.7
SeB_{11}···DME	180	3.22	−3.7/−3.4	−3.5	4.4	−3.5	−0.9	−3.6
	130	3.03	−5.9	−6.4	8.6	−6.2	−1.7	−5.8
SeB_{11}···DMK	180	3.18	−4.3/−4.0	−4.8	5.4	−3.7	−1.4	−4.5
	125	3.06	−6.2	−7.3	8.9	−6.4	−2.0	−6.8
SeB_{11}···FA	180	3.16	−4.2/−4.1	−4.8	4.9	−3.0	−1.3	−4.3
	130	2.92	−7.1	−9.6	11.7	−6.4	−3.0	−7.3

[a]The B12−Se-chalcogen-bond acceptor angle (α) in degrees. The Se···chalcogen-bond acceptor distance (d) in Å. The interaction energy (ΔE) and its decomposition into electrostatic (E_1^{Pol}), exchange−repulsion (E_1^{Ex}), dispersion (E_2^{D}), and induction (E^{Ind}) terms; energy in kcal mol^{-1}. [b]$E^{Ind} = E_2^{Ind} + E_2^{Ex\text{-}Ind} + \delta HF$

being more stable) and exceeded 1 kcal mol^{-1} only in the complex with **BEN**, where the ΔE of the SB_{11} complex was more negative by 1.4 kcal mol^{-1}. The geometrical and energetic changes in the complexes with **BEN** might be caused by the position of the σ-hole, which is located in the valley between two Pn atoms and hence less accessible for more bulky Pn-bond acceptors. The P-to-As substitution made the ΔE more negative by 0.7 kcal mol^{-1} on average. The P-to-As substitution thus had a smaller effect than the S-to-Se substitution. The SAPT analysis showed similarities between the studied Pn-bonds and E-bonds. The contribution from dispersion energy plays a dominant role here as well. Polarization (electrostatic) and induction energies are not negligible. Is some complexes, polarization (electrostatic) energy was even comparable to dispersion.

It should be mentioned that interactions in two Pn-bonded complexes were not exclusively σ-hole bonded, but contained also dihydrogen bonds (See Figure 3). Specifically, **P_2B_{10}···FA** and **As_2B_{10}···DME**, contains dihydrogen bonds with H···H distance of 2.1 and 2.4 Å, respectively. Comparison with the starting geometries however shows that ΔE are mainly due to the σ-hole interactions even here. ΔE of the **P_2B_{10}···FA** and **As_2B_{10}···DME** complexes in optimal geometries are −5.3 and −5.1 kcal mol^{-1}, respectively. The linear structures of the same complexes do not have dihydrogen bonds and have comparable ΔE values (−4.3 and −4.9 kcal mol^{-1},respectively; see Table 4).

Like E-bonds, the Pn-bonds with geometrically suitable Pn-bond acceptors are stronger in the heteroboranes than in other neutral Pn-bonded complexes known in the literature. For comparison, the ΔE of the **TMA**···$AsCl_3$ complex possessing a Pn-bond was reported to be −4.9[53] kcal mol^{-1} while that of the **1,2-As_2B_{10}···TMA** complex was −6.9 kcal mol^{-1}. In the case of complexes with **BEN**, compounds with a geometrically more accessible σ-hole form stronger Pn-bonds. For comparison, the ΔE of the **1,2-As_2B_{10}···BEN** complex is −5.6 kcal mol^{-1} while that of the $AsCl_3$···**BEN** complex has been reported to be −6.5 kcal mol^{-1}.[53]

Halogen Bond (X-bond). It was demonstrated above that the σ-hole on the Br atom in a carborane molecule is positive only if the Br is bound to the C atom. In the case of the Br bound to the B atom, the $V_{S,max}$ has a negative value, which is however less negative than that at the belt. This means that the σ-hole is still there but it does not have a positive value. Consequently, the **12-Br-C_2B_{10}** does not form a X-bond with most of the studied complexes. Only in the case of the **12-Br-C_2B_{10}···TMA** complex, a X-bond shorter than the sum of van der Waals radii is formed. The SAPT decomposition of ΔE shows that this rather weak but interesting interaction is enabled mainly by dispersion (see Table 5).

DOI: 10.1021/jp511101n
J. Phys. Chem. A 2015, 119, 1388−1395

Table 4. Structural and Energetic Characteristics of P_2B_{10} and As_2B_{10} Complexes[a]

complex	α	d	DFD-D3/CCSD(T) ΔE	DFT-SAPT E_1^{Pol}	E_1^{Ex}	E_2^{D}	E^{Ind}[b]	ΔE
				P_2B_{10}				
1,2-P_2B_{10}···BEN	180	3.63	−4.0/−4.3	−3.1	6.1	−6.9	−0.9	−4.8
	130	3.95	−4.8	−4.5	9.2	−8.7	−1.6	−5.6
1,2-P_2B_{10}···TMA	180	2.99	−5.8/−6.0	−8.3	14.0	−9.2	−2.5	−6.1
	155	2.99	−6.3	−8.8	14.7	−10.1	−2.5	−6.7
1,2-P_2B_{10}···DME	180	3.00	−4.1/−4.3	−4.7	8.1	−6.3	−1.4	−4.3
	145	3.04	−4.4	−4.8	8.2	−6.7	−1.3	−4.6
1,2-P_2B_{10}···DMK	180	3.00	−4.2/−4.3	−5.0	8.1	−5.7	−1.8	−4.5
	145	3.08	−4.3	−4.8	7.7	−6.1	−1.5	−4.7
1,2-P_2B_{10}···FA	180	2.98	−4.3/−4.5	−5.3	7.2	−4.7	−1.9	−4.6
	110	3.75	−5.3	−5.9	9.4	−6.4	−2.4	−5.3
				As_2B_{10}				
1,2-As_2B_{10}···BEN	180	3.61	−4.9	−4.1	7.1	−7.4	−1.2	−5.6
	160	3.64	−5.6/−5.4	−5.0	9.1	−8.6	−1.7	−6.2
1,2-As_2B_{10}···TMA	180	3.06	−6.5	−8.2	12.5	−8.7	−2.4	−6.8
	145	3.07	−6.9/−7.0	−7.9	12.4	−9.6	−2.2	−7.3
1,2-As_2B_{10}···DME	180	2.99	−4.9	−5.5	8.2	−6.1	−1.6	−5.0
	145	3.03	−5.1/−5.1	−5.3	7.7	−6.1	−1.5	−5.2
1,2-As_2B_{10}···DMK	180	3.00	−5.2	−6.2	8.6	−5.9	−2.0	−5.5
	140	3.08	−5.4/−5.3	−6.6	9.4	−6.6	−2.1	−5.8
1,2-As_2B_{10}···FA	180	2.98	−5.2	−6.3	7.4	−4.7	−2.1	−5.6
	150	3.00	−5.8/−5.8	−6.7	8.1	−5.5	−2.1	−6.1

[a]The center of B9 and B12-center of the bond between two pnicogen−pnicogen bond acceptor angle (α) in deg. The center of two pnicogen··· pnicogen bond acceptor distance (d) in Å. The interaction energy (ΔE) and its decomposition into electrostatic (E_1^{Pol}), exchange−repulsion (E_1^{Ex}), dispersion (E_2^{D}) and induction (E^{ind}) terms; energy in kcal mol^{-1}. [b]$E^{ind} = E_2^{Ind} + E_2^{Ex\text{-}Ind} + \delta HF$

Table 5. Structural and Energetic Characteristics of Br-C_2B_{10} Complexes[a]

complex	α	d	DFD-D3/CCSD(T) ΔE	DFT-SAPT E_1^{Pol}	E_1^{Ex}	E_2^{D}	E^{Ind}[b]	ΔE
				1-Br-C_2B_{10}				
1-Br-C_2B_{10}..BEN	180	3.40	−3.3	−2.6	4.7	−4.9	−0.8	−3.5
1-Br-C_2B_{10}···TMA	180	2.63	−9.6/−8.1	−20.2	30.9	−9.2	−8.0	−6.5
1-Br-C_2B_{10}···DME	180	2.79	−4.6/−4.6	−7.7	11.5	−5.2	−2.4	−3.8
1-Br-C_2B_{10}···DMK	180	2.80	−4.6/−4.5	−8.1	12.0	−5.1	−2.8	−4.0
1-Br-C_2B_{10}···FA	180	2.76	−5.0/−5.0	−8.5	11.7	−4.5	−3.0	−4.3
				12-Br-C_2B_{10}				
12-Br-C_2B_{10}···TMA	180	3.29	−2.1/−1.8	−2.2	4.7	−3.6	−0.8	−1.8

[a]The B12−Br halogen-bond acceptor angle (α) in deg. Br···halogen bond acceptor distance (d) in Å. The interaction energy (ΔE) and its decomposition into electrostatic (E_1^{pol}), exchange-repulsion (E_1^{exch}), dispersion (E_2^{disp}), and induction (E^{ind}) terms; energy in kcal mol^{-1}. [b]$E^{Ind} = E_2^{Ind} + E_2^{Exch\text{-}Ind} + \delta HF$

The **1-Br-C_2B_{10}···TMA** complex has a very strong X-bond. It is due to the very large polarization (electrostatic) term and moderately large induction term (see Table 5). The very strong ΔE of this complex is also confirmed by the CCSD(T) calculation. **1-Br- C_2B_{10}** had the least negative ΔE with **BEN**. **TMA** has been known to be a considerably better electron donor than **BEN** for the X-bond for organic compounds as well. For example, the ΔE of the **BEN**···Br-CF_3 and **TMA**···Br-CF_3 X-bonded complexes has been calculated to be −3.3 and −7.6 kcal mol^{-1}, respectively.[53]

A comparison of the strength of the X-bond of heteroboranes and that of already known organic compounds reveals that they are similar provided that the organic compounds have several electron-withdrawing groups adjacent to the X atom involved in the X-bond. For example, the ΔE of the **BEN**···$BrCH_3$, **BEN**···$BrCF_3$, and **BEN**···**1-Br-C_2B_{10}** complexes possessing a Br···π type X-bond is −1.8, −3.1,[55] and −3.3 kcal mol^{-1}, respectively.

4. CONCLUSIONS

The chalcogens (E) and pnicogens (Pn) incorporated in heteroborane cages carry a partial positive charge and have highly positive σ-holes. Consequently, these heteroboranes form very strong E- and Pn-bonds. Halogen (X) atoms in *exo*-substituted dicarbaboranes also have a highly positive σ-hole, but only in the case of a C-bound X atom. In such cases, the X-bond of heteroboranes is strong and comparable to X-bonds in organic compounds with several electron-withdrawing groups being close to the X atom involved in the X-bond. The fact that heteroboranes can form strong σ-hole bonds to various electron donors can be utilized in the design of heteroborane-based protein ligands, such as enzyme inhibitors or receptor agonists/antagonists, and in crystal engineering.

DOI: 10.1021/jp511101n
J. Phys. Chem. A 2015, 119, 1388−1395

■ ASSOCIATED CONTENT

Supporting Information

The structural and energetic characteristics of **Cl-SB$_{11}$** and **SB$_9$** complexes and plots of interaction energies. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Authors

*(J.F.) Telephone: (+420) 220 410 318. Fax: (+420) 220 410 320. E-mail: fanfrlik@uochb.cas.cz.

*(D.H.) Telephone: (+420) 266 173 218. E-mail: hnyk@iic.cas.cz.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by Research Project RVO 61388963, awarded by the Academy of Sciences of the Czech Republic. We acknowledge the financial support of the Czech Science Foundation (J.F., A.P., M.L., P.H., Grant No. P208/12/G016; D.H., Grant No. P208/10/2269). We also thank the Gilead Sciences and IOCB Research Centre for financial support. This work was also supported by the Operational Program Research and Development for Innovations—European Science Fund (Grant No. CZ.1.05/2.1.00/03.0058). This work was supported by the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070), funded by the European Regional Development Fund, and the national budget of the Czech Republic via the Research and Development for Innovations Operational Programme, as well as Czech Ministry of Education, Youth and Sports via the project Large Research, Development and Innovations Infrastructures (LM2011033).

■ REFERENCES

(1) Clark, T.; Hennemann, M.; Murray, J. S.; Politzer, P. Halogen Bonding: The σ-hole. *J. Mol. Model.* **2007**, *13*, 291−296.

(2) Murray, J. S.; Politzer, P. Halogen Bonding: An Interim Discussion. *ChemPhysChem* **2013**, *14*, 278−294.

(3) Metrangolo, P.; Resnati, G. Halogen Bonding: A Paradigm in Supramolecular Chemistry. *Chem.—Eur. J.* **2001**, *7*, 2511−2519.

(4) Nguyen, L. H.; Horton, P. N.; Hursthouse, M. B.; Legon, A. C.; Bruce, D. W. Halogen Bonding: A New Interaction for Liquid Crystal Formation. *J. Am. Chem. Soc.* **2004**, *126*, 16−17.

(5) Lu, Y.; Liu, Y.; Xu, Z.; Li, H.; Liu, H.; Zhu, W. Halogen Bonding for Rational Drug Design and new Drug Discovery. *Expert Opin. Drug Discovery* **2012**, *7*, 375−383.

(6) Wilcken, R.; Zimmermann, M. O.; Lange, A.; Joerger, A. C.; Boeckler, F. M. Principles and Applications of Halogen Bonding in Medicinal Chemistry and Chemical Biology. *J. Med. Chem.* **2013**, *56*, 1363−1388.

(7) Wilcken, R.; Liu, X.; Zimmermann, M. O.; Rutherford, T. J.; Fersht, A. R.; Joerger, A. C.; Boeckler, F. M. Halogen-Enriched Fragment Libraries as Leads for Drug Rescue of Mutant p53. *J. Am. Chem. Soc.* **2012**, *134*, 6810−6818.

(8) Fedorov, O.; Huber, K.; Eisenreich, A.; Filippakopoulos, P.; King, O.; Bullock, A. N.; Szklarczyk, D.; Jensen, L. J.; Fabbro, D.; Trappe, J.; et al. Specific CLK Inhibitors from a Novel Chemotype for Regulation of Alternative Splicing. *Chem. Biol.* **2011**, *18*, 67−76.

(9) Hardegger, L. A.; Kuhn, B.; Spinnler, B.; Anselm, L.; Ecabert, R.; Stihle, M.; Gsell, B.; Thoma, R.; Diez, J.; Benz, J.; et al. Halogen Bonding at the Active Sites of Human Cathepsin L and MEK1 Kinase: Efficient Interactions in Different Environments. *Chem. Med. Chem.* **2011**, *54*, 2048−2054.

(10) Werz, D. B.; Gleiter, R.; Rominger, F. Nanotube Formation Favored by Chalcogen-Chalcogen Interactions. *J. Am. Chem. Soc.* **2002**, *124*, 10638−10639.

(11) Iwaoka, M.; Takemoto, S.; Tomoda, S. Statistical and Theoretical Investigations on the Directionality of Nonbonded S···O Interactions. Implications for Molecular Design and Protein Engineering. *J. Am. Chem. Soc.* **2002**, *124*, 10613−10620.

(12) Iwaoka, M.; Takemoto, S.; Okada, M.; Tomoda, S. Statistical Characterization of Nonbonded S···O Interactions in Proteins. *Chem. Lett.* **2001**, 132−133.

(13) Burling, F. T.; Goldstein, B. M. Computational Studies of Nonbonded Sulfur-Oxygen and Selenium-Oxygen Interactions in the Thiazole and Selenazole Nucleosides. *J. Am. Chem. Soc.* **1992**, *114*, 2313−2320.

(14) Nagao, Y.; Hirata, T.; Goto, S.; Sano, S.; Kakehi, A.; Iizuka, K.; Shiro, M. Intramolecular Nonbonded S···O Interaction Recognized in (Acylimino)thiadiazoline Derivatives as Angiotensin II Receptor Antagonists and Related Compounds. *J. Am. Chem. Soc.* **1998**, *120*, 3104−3110.

(15) Taylor, J. C.; Markham, G. D. The Bifunctional Active Site of S-Adenosylmethionine Synthetase. Roles of the Active Site Aspartates. *J. Biol. Chem.* **1999**, *274*, 32909−32914.

(16) Bender, K.; Hennig, I.; Schweitzer, D.; Dietz, K.; Endres, H.; Keller, H. J. Synthesis, Structure and Physical Properties of a Two-Dimensional Organic Metal, Di[bis(ethylenedithiolo)-tetrathiofulvalene] triiodide, (BEDT-TTF)$^{+}_{2}$I$^{-}_{3}$. *Mol. Cryst. Liq. Cryst.* **1984**, *108*, 359−371.

(17) Zahn, S.; Frank, R.; Hey-Hawkins, E.; Kirchner, B. Pnicogen Bonds: A New Molecular Linker? *Chem.—Eur. J.* **2011**, *17*, 6034−6038.

(18) Bauzá, A.; Mooibroek, T. J.; Frontera, A. Tetrel-Bonding Interaction: Rediscovered Supramolecular Force? *Angew. Chem., Int. Ed.* **2013**, *52*, 12317−12321.

(19) Kolář, M.; Carloni, P. Statistical Analysis of σ-Holes: A Novel Complementary View on Halogen Bonding. *Phys. Chem. Chem. Phys.* **2014**, *16*, 19111−19114.

(20) Wang, W.; Ji, B.; Zhang, Y. Chalcogen Bond: A Sister Noncovalent Bond to Halogen Bond. *J. Phys. Chem. A* **2009**, *113*, 8132−8135.

(21) Riley, K. E.; Murray, J. S.; Fanfrlik, J.; Rezac, J.; Sola, R. J.; Concha, M. C.; Ramos, F. M.; Politzer, P. Halogen Bond Tunability I: The Effects of Aromatic Fluorine Substitution on the Strengths of Halogen-Bonding Interactions Involving Chlorine, Bromine, and Iodine. *J. Mol. Model.* **2011**, *17*, 3309−3318.

(22) Kolář, M.; Hostaš, J.; Hobza, P. The Strength and Directionality of a Halogen Bond are Co-determined by the Magnitude and Size of the σ-hole. *Phys. Chem. Chem. Phys.* **2014**, *16*, 9987−9996.

(23) Herdegger, L. A.; Kuhn, B.; Spinnler, B.; Anselm, L.; Ecabert, R.; Stihle, M.; Gsell, B.; Thoma, R.; Diez, J.; Benz, J.; et al. Systematic Investigation of Halogen Bonding in Protein-Ligand Interactions. *Angew. Chem., Int. Ed.* **2011**, *50*, 314−319.

(24) Fanfrlík, J.; Kolář, M.; Kamlar, M.; Hurny, D.; Ruiz, F. X.; Cousido-Siah, A.; Mitschler, A.; Řezáč, J.; Munusamy, E.; Lepšík, M.; et al. The Modulation of Aldose Reductase Inhibition by Halogen Bond Tuning. *ACS Chem. Biol.* **2013**, *8*, 2484−2492.

(25) Fanfrlík, J.; Přáda, A.; Padělková, Z.; Pecina, A.; Macháček, J.; Lepšík, M.; Holub, J.; Růžička, A.; Hnyk, D.; Hobza, P. The Dominant Role of Chalcogen Bonding in the Crystal Packing of 2D/3D Aromatics. *Angew. Chem., Int. Ed.* **2014**, *53*, 10139−10142.

(26) Lipscomb, W. N. *Boron Hydrides*; W. A. Benjamin: New York, 1963.

(27) King, R. B. Three-Dimensional Aromaticity in Polyhedral Boranes and Related Molecules. *Chem. Rev.* **2001**, *101*, 1119−1152.

(28) Macháček, J.; Plešek, J.; Holub, J.; Hnyk, D.; Všetečka, V.; Císařová, I.; Kaupp, M.; Štíbr, B. New Route to 1-thia-closo-dodecaborane(11), *closo*-1-SB$_{11}$H$_{11}$, and its Halogenation Reactions. The Effect of the Halogen on the Dipole Moments and the NMR Spectra and the Importance of Spin−Orbit Coupling for the ^{11}B Chemical Shifts. *Dalton Trans.* **2006**, *8*, 1024−1029.

DOI: 10.1021/jp511101n
J. Phys. Chem. A 2015, 119, 1388−1395

(29) Hnyk, D.; Všetečka, V.; Drož, L.; Exner, O. Charge Distribution within 1,2-dicarba-closo-dodecaborane: Dipole Moments of its Phenyl Derivatives. *Collect. Czech. Chem. Commun.* **2001**, *66*, 1375−1379.
(30) Hosmane, N. S. *Boron Science*; CRC Press: Boca Raton FL, 2011.
(31) Baše, T.; Bastl, Z.; Plzák, Z.; Grygar, T.; Plešek, J.; Malina, V.; Šubrt, J.; Boháček, J.; Večerníková, E. Kříž, O. Carboranethiol-Modified Gold Surfaces. A Study and Comparison of Modified Cluster and Flat Surfaces. *Langmuir* **2005**, *21*, 7776−7785.
(32) Cígler, P.; Kožíšek, M.; Řezáčová, P.; Brynda, J.; Otwinowski, Z.; Pokorná, J.; Plešek, J.; Grüner, B.; Doleckova-Marešová, L.; Masa, M.; et al. From Nonpeptide Toward Noncarbon Protease Inhibitors: Metallacarboranes as Specific and Potent Inhibitors of HIV Protease. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 15394−15399.
(33) Brynda, J.; Mader, P.; Šícha, V.; Fábry, M.; Poncová, K.; Bakardiev, M.; Grüner, B.; Cígler, P.; Řezáčová, P. Carborane-Based Carbonic Anhydrase Inhibitors. *Angew. Chem., Int. Ed.* **2013**, *52*, 13760−13763.
(34) Fanfrlík, J.; Lepšík, M.; Horinek, D.; Havlas, Z.; Hobza, P. Interaction of Carboranes with Biomolecules: Formation of Dihydrogen Bonds. *ChemPhysChem* **2006**, *7*, 1100−1105.
(35) Chen, Z. F.; King, R. B. Spherical Aromaticity: Recent Work on Fullerenes, Polyhedral Boranes, and Related Structures. *Chem. Rev.* **2005**, *105*, 3613−3642.
(36) Pecina, A.; Lepšík, M.; Řezáč, J.; Brynda, J.; Mader, P.; Řezáčová, P.; Hobza, P.; Fanfrlík, J. QM/MM Calculations Reveal the Different Nature of the Interaction of two Carborane-Based Sulfamide Inhibitors of Human Carbonic Anhydrase II. *J. Phys. Chem. B* **2013**, *117*, 16096−16104.
(37) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate *ab initio* Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104−154119.
(38) Tao, J. M.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. Climbing the Density Functional ladder: Nonempirical meta-generalized gradient approximation designed for molecules and solids. *Phys. Rev. Lett.* **2003**, *91*, 146401−146404.
(39) Schafer, A.; Huber, C.; Ahlrichs, R. Fully Optimized Contracted Gaussian-Basis Sets of Triple Zeta Valence Quality for Atoms Li to Kr. *J. Chem. Phys.* **1994**, *100*, 5829−5835.
(40) Jurečka, P.; Hobza, P. On the convergence of the (ΔECCSD-(T)−ΔEMP2) term for complexes with multiple H-bonds. *Chem. Phys. Lett.* **2002**, *365*, 89−94.
(41) Lee, E. C.; Kim, D.; Jurečka, P.; Tarakeshwar, P.; Hobza, P.; Kim, K. S. Understanding of Assembly Phenomena by Aromatic-Aromatic Interactions: Benzene Dimer and the Substituted Systems. *J. Phys. Chem. A* **2007**, *111*, 3446−3457.
(42) Jeziorski, B.; Moszynski, R.; Szalewicz, K. Perturbation Theory Approach to Intermolecular Potential Energy Surfaces of van der Waals Complexes. *Chem. Rev.* **1994**, *94*, 1887−1930.
(43) Williams, H. L.; Chabalowski, C. F. Using Kohn−Sham Orbitals in Symmetry-Adapted Perturbation Theory to Investigate Intermolecular Interactions. *J. Phys. Chem. A* **2001**, *105*, 646−659.
(44) Jansen, G.; Hesselmann, A. First-Order Intermolecular Interaction Energies from Kohn−Sham Orbitals. *Chem. Phys. Lett.* **2002**, *357*, 464−470.
(45) Misquitta, J.; Szalewicz, K. Intermolecular Forces from Asymptotically Corrected Density Functional Description of Monomers. *Chem. Phys. Lett.* **2002**, *357*, 301−306.
(46) Hesselman, A.; Jansen, G.; Schütz, M. Density-Functional Theory-Symmetry-Adapted Intermolecular Perturbation Theory with Density Fitting: A New Efficient Method to Study Intermolecular Interaction Energies. *J. Chem. Phys.* **2005**, *122*, 14103−14119.
(47) Boys, S. F.; Bernardi, F. The Calculation of Small Molecular Interactions by the Differences of Separate Total Energies. Some Procedures with Reduces Errors. *Mol. Phys.* **1970**, *19*, 553−566.
(48) Ahlrichs, R.; Bar, M.; Haser, M.; Horn, H.; Kolmel, C. Electronic Structure Calculations on Workstation Computers: The Program System Turbomole. *Chem. Phys. Lett.* **1989**, *162*, 165−169.
(49) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A. et al. *Gaussian 09, Revision, D.01*; Gaussian, Inc.: Wallingford CT, 2009.
(50) Werner, H.-J.; Knowles, P. J.; Manby, F. R.; Schutz, M.; Celani, P.; Knizia, G.; Korona, T.; Lindh, R.; Mitrushenkov, A.; Rauhut, G. et al.. *MOLPRO, version 2010, a package of ab initio programs*; University College Cardi Consultants Limited; Wales, U.K., 2010.
(51) Hnyk, D.; Wann, D. A.; Holub, J.; Bühl, M.; Robertson, H. E.; Rankin, D. W. H. The Gas-Phase Structure of 1-Selena-*closo*-Dodecaborane(11), 1-$SeB_{11}H_{11}$, Determined by the Concerted Use of Electron Diffraction and Computational Methods. *Dalton Trans.* **2008**, *7*, 96−100.
(52) Holub, J.; Kennedy, J. D.; Štíbr, B. Nine-Vertex Polyhedral Monothiaborane Chemistry. The First *nido* Thianonaborane: Isolation and Characterization by NMR of the [*nido*-9-SB_8H_9]- Anion, and its Conversion to Compounds of the 6-L-*arachno*-4-SB_8H_{10} Series (where L = SMe_2 and Pyridine). *Collect. Czech. Chem. Commun.* **1994**, *59*, 367−373.
(53) Bauzá, A.; Quiñonero, D.; Deyà, P. M.; Frontera, A. Halogen Bonding versus Chalcogen and Pnicogen Bonding: A Combined Cambridge Structural Database and Theoretical Study. *CrystEngComm* **2013**, *15*, 3137−3144.
(54) McLellan, R.; Boag, N. M.; Dodds, K.; Ellis, D.; Macgregor, S. A.; McKay, D.; Masters, S. L.; Noble-Eddy, R.; Platt, N. P.; Rankin, D. W. H.; et al. New Chemistry of 1,2-*closo*-$P_2B_{10}H_{10}$ and 1,2-*closo*-$As_2B_{10}H_{10}$; In Silico and Gas Electron Diffraction Structures, and New Metalladiphospha- and Metalladiarsaboranes. *Dalton Trans.* **2011**, *40*, 7181−7192.
(55) Řezáč, J.; Riley, K. E.; Hobza, P. Benchmark Calculations of Noncovalent Interactions of Halogenated Molecules. *J. Chem. Theory Comput.* **2012**, *8*, 4285−4292.

DOI: 10.1021/jp511101n
J. Phys. Chem. A 2015, 119, 1388−1395

Supporting Information

Chalcogen and Pnicogen Bonds in Complexes of Neutral Icosahedral and Bicapped Square-Antiprismatic Heteroboranes

Adam Pecina,[1] Martin Lepšík,[1] Drahomír Hnyk,[2] Pavel Hobza,[1,3] and Jindřich Fanfrlík*[1]*

[1]Gilead Sciences and IOCB Research Center and Institute of Organic Chemistry and Biochemistry (IOCB), Academy of Sciences of the Czech Republic, v.v.i.; Flemingovo nám. 2, 16610 Prague 6 (Czech Republic)

[2]Institute of Inorganic Chemistry, Academy of Sciences of the Czech Republic, v.v.i.; 250 68 Řež (Czech Republic)

[3] Regional Center of Advanced Technologies and Materials Department of Physical Chemistry, Palacký University, 77146 Olomouc (Czech Republic)

Table S1. The structural and energetic characteristics of **Cl-SB$_{11}$** and **SB$_9$** complexes. The B12-chalcogen-chalcogen-bond acceptor angle (α) in degrees. The chalcogen...chalcogen-bond acceptor distance (d) in Å. The interaction energy (ΔE) and its decomposition into electrostatic (E_1^{Pol}), exchange-repulsion (E_1^{Ex}), dispersion (E_2^{D}) and induction (E^{Ind}) terms; energy in kcal mol^{-1}.

Complex	α	d	DFD-D3	DFT-SAPT				
			ΔE	E_1^{Pol}	E_1^{Ex}	E_2^{D}	E^{Ind}[a]	ΔE
Cl-SB$_{11}$								
Cl-SB$_{11}$...BEN	180	3.11	–6.2	–5.4	9.2	–8.4	–2.1	–6.7
	165	3.13	–6.3	–5.4	9.3	–8.6	–2.0	–6.7
Cl-SB$_{11}$...TMA	180	3.24	–4.2	–4.0	5.6	–4.8	–1.1	–4.3
	140	3.03	–6.3	–7.5	11.5	–7.8	–2.2	–6.1
Cl-SB$_{11}$...DME	180	3.18	–3.2	–2.8	3.7	–3.4	–0.7	–3.2
	135	3.07	–4.9	–4.5	5.8	–5.2	–1.1	–5.0
Cl-SB$_{11}$...DMK	180	3.21	–3.7	–3.4	3.7	–3.2	–0.9	–3.9
	135	3.04	–5.4	–5.8	7.1	–5.5	–1.6	–5.9
Cl-SB$_{11}$...FA	180	3.15	–3.7	–3.8	3.8	–2.8	–1.0	–3.8
	130	3.01	–5.9	–6.6	7.6	–5.3	–1.8	–6.1
SB$_9$								
SB$_9$...BEN	180	3.13	–5.0	–4.8	8.4	–7.2	–1.8	–5.5
	150	3.19	–5.3	–4.6	8.2	–7.8	–1.6	–5.8
SB$_9$...TMA	180	3.19	–3.6	–4.2	6.0	–4.4	–1.1	–3.7
	125	2.98	–6.1	–8.3	13.1	–8.4	–2.4	–6.0
SB$_9$...DME	180	3.23	–2.8	–2.2	3.3	–3.4	–0.6	–2.8
	125	3.08	–4.5	–4.3	6.0	–5.3	–1.1	–4.6
SB$_9$...DMK	180	3.20	–3.1	–3.2	4.2	–3.3	–0.9	–3.3
	120	3.11	–5.7	–5.2	6.6	–5.6	–1.4	–5.6
SB$_9$...FA	180	3.17	–3.2	–3.3	3.4	–2.4	–1.0	–3.4
	115	3.14	–5.3	–3.3	3.4	–2.4	–1.1	–3.6

[a] $E^{Ind} = E_2^{Ind} + E_2^{Ex\text{-}Ind} + \delta HF$

Figure S1. Interaction energy (ΔE) plotted against B12-S-chalcogen bond acceptor angle. Structures of complexes at an optimal angle (right) and at 180 degrees (left) are shown. Energy in kcal mol^{-1}, angle in degrees and Color coding as follows: black, C; light-pink, B; yellow, S; white, H.

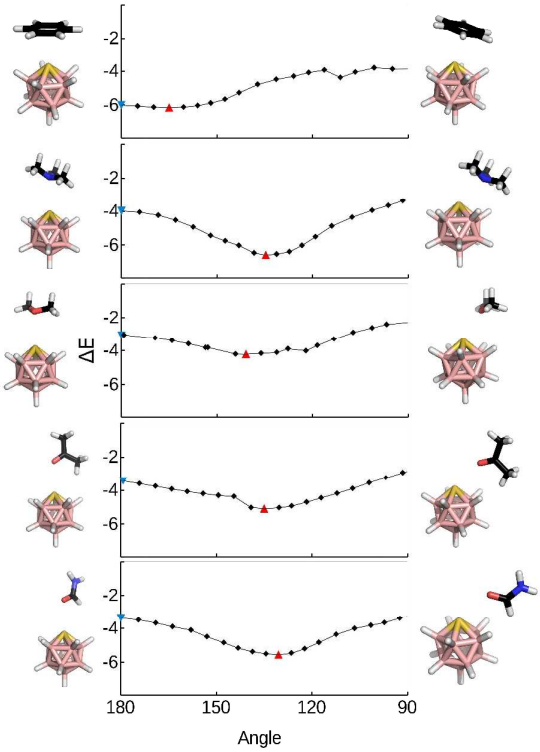
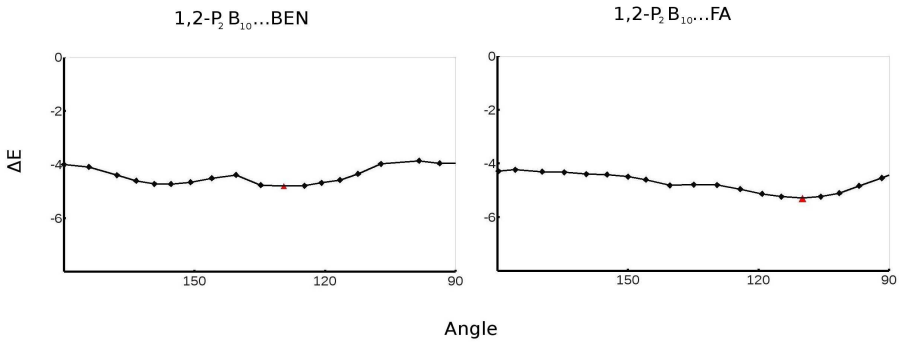


Figure S2. Interaction energy (ΔE) plotted against the centre of B9 and B12 -the centre of P1 and P2 -pnictogen-bond acceptor angle. Energy in kcal mol^{-1}, angle in degrees



Appendix D

ON THE RELIABILITY OF THE CORRECTED SEMIEMPIRICAL QUANTUM CHEMICAL METHOD (PM6-DH2) FOR ASSIGNING THE PROTONATION STATES IN HIV-1 PROTEASE/INHIBITOR COMPLEXES

Adam PECINA[1,#], Ondřej PŘENOSIL[2,#], Jindřich FANFRLÍK[3], Jan ŘEZÁČ[4], Jaroslav GRANATIER[5], Pavel HOBZA[6,*] and Martin LEPŠÍK[7,*]

Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, v.v.i., and Center for Biomolecules and Complex Molecular Systems, 166 10 Prague 6, Czech Republic; e-mail: [1] pecina@uochb.cas.cz, [2] ondrej.prenosil@uochb.cas.cz, [3] jindrich.fanfrlik@uochb.cas.cz, [4] rezac@uochb.cas.cz, [5] j.granatier@gmail.com, [6] hobza@uochb.cas.cz, [7] lepsik@uochb.cas.cz

Received February 3, 2011
Accepted March 11, 2011
Published online April 20, 2011

Dedicated to Dr. Zdeněk Havlas on the occasion of his 60th birthday.

A novel computational protocol for determining the most probable protonation states in protein/ligand complexes is presented. The method consists in treating large parts of the enzyme using the corrected semiempirical quantum chemical (QM) method – PM6-D2 for optimization and PM6-DH2 for single-point energies – while the rest is calculated using molecular mechanics (MM) within a hybrid QM/MM fashion. The surrounding solvent is approximated by an implicit model. This approach is applied to two model systems, two different carboxylate pairs in one general and one unique HIV-1 protease/inhibitor complex. The effect of the size of the movable QM part is investigated in a series of several sizes, 3-, 6-, 8- and 10-Å regions surrounding the inhibitor. For the smallest region (< 450 atoms) the computationally more costly DFT QM/MM optimizations are performed as a check of the correctness. Proton transfer (PT) phenomena occur at both the PM6-D2 and DFT levels, which underlines the requirement for a QM approach. The barriers of PT are checked in model carboxylic acid pairs using the highly accurate MP2 and CCSD(T) values. An important result of this study is the fine-tuning of the protocol which can be used in further applications; its limitations are also shown, pointing to future developments. The calculations reveal which protonation variants of the active site are the most stable. In conclusion, the presented protocol can also be utilized for defining probable isomers in biomolecular systems. It can also serve as a preparatory step for further interaction-energy and binding-score calculations.

Keywords: HIV-1 protease inhibition; Protonation; QM/MM calculations; Semiempirical quantum chemical method; Proton transfer; Drug design; Inhibitors; X-ray crystallography.

[#] Both authors contributed equally to the work.

Collect. Czech. Chem. Commun. 2011, Vol. 76, No. 5, pp. 457–479

doi:10.1135/cccc2011035

Protonation and thus also the charge of amino-acid residues in proteins are defined by their pK_a values and pH [1]. The determination of the pK_a values of titratable residues in enzyme active sites is thus a key prerequisite for a molecular understanding of the reaction mechanisms and inhibition. Indeed, it has been shown that the calculated protein/ligand interaction energy is sensitive to the protonation state of the active site[2]. However, owing to the interactions between protonation sites, the protein titration curves may deviate from the standard Henderson–Haselbalch curves[3,4]. A number of experimental and computational approaches have been devised to determine the pK_a values of amino-acid side chains in proteins. A few of them related to this work are mentioned here. A classical experimental approach is based on measuring the pH dependence of a reaction[5]. Atomic resolution (R < 1.1 Å) X-ray crystal structures have also been used to infer the protonation states of titratable residues in enzyme active sites[6,7]. From the theoretical side, electrostatic pK_a calculations[4,8], quantum-mechanics (QM)-based calculations[9–11] or force-field molecular dynamics simulations[12,13] have been utilized to determine the protonation states.

HIV protease (PR) is one of the most intensively studied pharmaceutical targets. Its C_2-symmetrical dimeric structure features two catalytic aspartates (Asp25/Asp25′) in its active site (Fig. 1). These two carboxyl moieties are coplanar and so close to each other that one Asp has a shifted pK_a of ~6 while the other stays at ~3.5 [14]. As a result of this, one proton connects this Asp dyad of an unliganded PR via a double-well low-barrier hydrogen bond[15]. Bidirectional proton hopping between the two aspartates in this system has been simulated in an ab initio molecular dynamics study using a six-residue fragment of the active site[16]. In complexes with inhibitors, the catalytic Asp dyad of PR is monoprotonated in most cases (inhibitors featuring hydroxyl isostere; ref.[17] and references therein) and less frequently diprotonated (statine-based inhibitors)[6,18].

Proton transfer (PT) is one of the most important quantum effects, which are, by definition, not covered by empirical force fields. Quantum mechanical calculations, in contrast, inherently describe not only PT but also other quantum effects like charge redistribution, electron transfer or halogen bonding. The QM methods are thus the proper tool to use in order to determine the protonation states in the active sites of proteins where these phenomena might be important. However, because of the high computational costs, usually only a few residues in the active site could be treated[11,19]. Recent progress in the development of linear-scaling semiempirical quantum chemical (SQM) methods has offered the possibility to treat the whole biomolecular system containing several thousand atoms. However, the ac-

curacy of such methods was quite low by QM standards. Therefore, corrected versions have recently been introduced.

In our laboratory, we have chosen the novel SQM method PM6 with a parametrization for 70 elements[20], which is very well suited for the modeling of protein/ligand complexes, thanks to among others its linear-scaling algorithm MOZYME[21]. To increase further its accuracy for noncovalent binding, we have corrected this method with dispersion and hydrogen-bonding corrections (PM6-DH2) to reproduce closely benchmark CCSD(T) data[22,23]. As the solvent effects influence the biomolecular structures and energies, we use an implicit solvent of the COSMO[24] or generalized Born type[25] around the proteins while some important explicit water molecules from the crystal structure may be added. Hybrid QM/MM calculations using PM6 in a large QM part are also possible.

For this pilot study, we have chosen two model systems, one general and one unique HIV-1 PR/inhibitor complex. The inhibitors are: (i) the clinically successful nonpeptidic inhibitor darunavir (DRV, TMC-114, UIC-94017; K_i value of 5.3 pM)[26] and (ii) a phenylnorstatine-based peptidomimetic in-

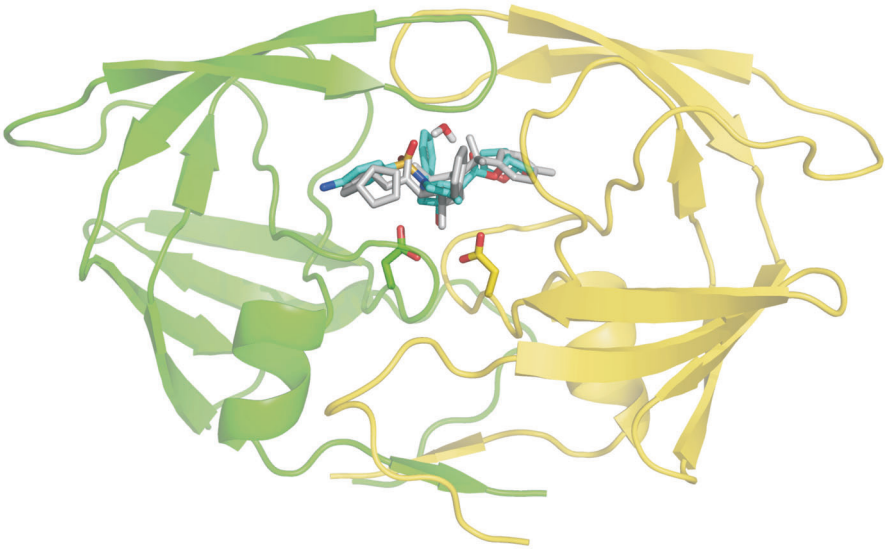
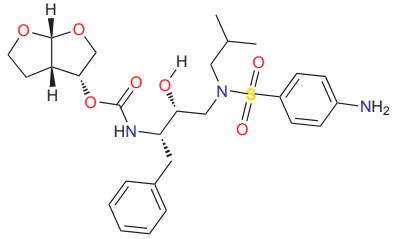


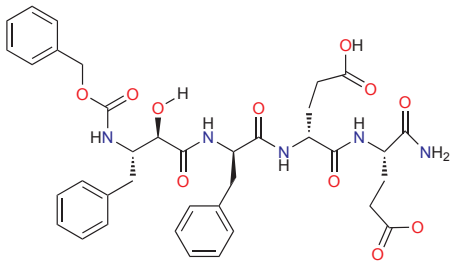
FIG. 1
The ribbon structure of the dimeric wild-type HIV-1 protease (PR) in complex with the darunavir (DRV) inhibitor (PDB code 3QOZ). One PR monomer is shown in green, the other in yellow. The two catalytic aspartates are depicted as sticks. Both of the crystallographic orientations of DRV are shown in cyan and grey. The structural flap water is shown with the hydrogens added. The oxygen atoms are in red, nitrogen atoms in blue, hydrogens in white

hibitor KI2 (K_i value of 180 pM)[27] (Fig. 2). The complex of wild-type (wt) PR with DRV crystallized in the hexagonal $P6_1$ space group yielded two orientations of the inhibitor[28] in the pseudo-C_2-symmetrical enzyme (the conformations of the A and B chain PR residues differ slightly owing to the binding of the asymmetrical inhibitor; Figs 1, 3A). Moreover, there are four possible variants of the monoprotonated catalytic Asp dyad for both inhibitor orientations (see the Methods).

The KI2 inhibitor formed a unique complex with the PR in which two molecules of KI2 bound to the enzyme; one was localized in the active site[6] and the other at the outer part of the PR, which allowed an atomic resolution of the crystal structure (1.03 Å)[29]. With such a high quality of the X-ray structure, the protonation state of the active site could be inferred from measuring the highly accurate CG-OD1/OD2 distances (Fig. 3B)[6]. Furthermore, the electron-density maps allowed a resolution of the P2 benzyloxycarbonyl group of KI2 to conformations (depicted in cyan and grey in Fig. 3B) with an alternative possibility of hydrogen bonding (H22 bonding inter-molecularly to the OD2 of the Asp25′ or intra-molecularly to the O01 of the KI2; Fig. 3B). In another part of the active site, another carboxyl–



DRV



KI2

FIG. 2
The structures of two potent protease inhibitors studied in this work

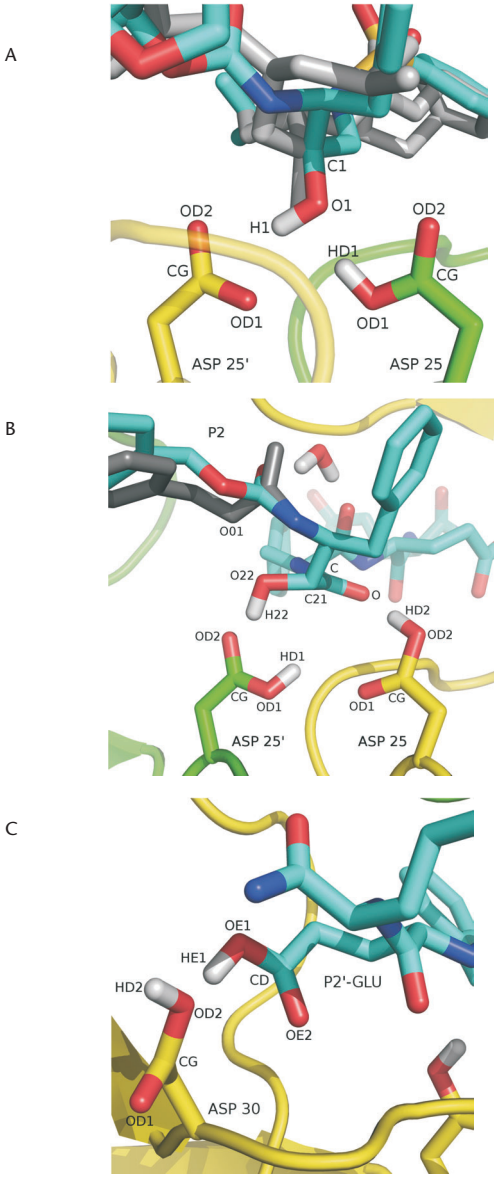


FIG. 3
Details of the active sites of protease/inhibitor complexes. The color coding is the same as in Fig. 1. The hydroxyl (O1–H1) of darunavir binding to a monoprotonated Asp25/25′ dyad of protease (A). The norstatine hydroxyl (O22–H22) and carbonyl (C–O) of KI2 binding to a diprotonated Asp25/25′ dyad of protease (B). The protonated Glu-P2′ of KI2 binding to Asp30 of protease (C)

carboxyl interaction was observed, namely between the side chain of the Asp30 of the PR and the P2′ glutamate moiety of the KI2 (Fig. 3C). The highly accurate CG-OD1/OD2 and CD-OE1/OE2 distances only reveal a preference for a protonation at the OD2 of the Asp30 and the OE1 of the Glu-P2′. This could be explained by either the fact that both of these oxygen atoms are protonated or the fact that a single proton connects them via

TABLE I
The protonation, orientation and conformation variants of the calculated structures

Variant	Inhibitor orientation	Active-site proton location
A. The PR/DRV complex		
D1	A	Asp25:OD1
D2	A	Asp25′:OD1
D3	A	Asp25:OD2
D4	A	Asp25′:OD2
D5	B	Asp25:OD1
D6	B	Asp25′:OD1
D7	B	Asp25:OD2
D8	B	Asp25′:OD2

Variant	Inhibitor chain/conformation	H22 hydrogen bond acceptor[a]	HE1/HD2 location in Asp30/Glu-P2′ pair[b]
B. The PR/KI2 complex			
K1	I/A	OD2	OD2
K2	I/A	OD2	OE1
K3	I/A	OD2	OD2, OE1
K4	I/B	OD2	OD2
K5	I/B	O01	OD2
K6	I/B	OD2	OE1
K7	I/B	O01	OE1
K8	I/B	OD2	OD2, OE1

[a] cf. Fig. 3B; [b] cf. Fig. 3C.

a hydrogen bond which can be either localized or mobile (low-barrier hydrogen bond). In order to shed light on the probability of these variants, we have constructed the respective molecular models and explored them computationally.

In this paper, several variants of the proton locations on the carboxylic moieties in question in both the wtPR/DRV and KI2 complexes have been investigated. Using QM/MM optimizations, these structures have been sorted by the relative energies of their QM parts; the lowest energy variants correspond to the most stable ones. The corrected PM6 method (PM6-D2 for optimization and PM6-DH2 for single-point energies) is used for the QM regions extending up to 10 Å from the inhibitors, while the results are checked using the DFT QM/MM optimizations on the 3-Å surroundings. The observed proton-transfer phenomena are further checked on small monoprotonated carboxylate pair models using high-level MP2 and CCSD(T) methods. In summary, we present a novel computational protocol which not only can be used for determining the protonation in the active site of the HIV PR but which also represents a general computational procedure enabling an objective decision on which potential isomers in a biomolecule/ligand complex will be populated.

METHODS

Systems Studied

The complexes of wild-type HIV-1 protease with two potent inhibitors, darunavir (DRV, TMC-114, UIC-94017)[26] and KI2 [27], were studied (Figs 1, 2). The current crystallographic structure of the wtPR/DRV complex (PDB code 3QOZ) represents a general case of the most common hydroxyl isostere inhibitors binding to the PR. The complex crystallized in the hexagonal $P6_1$ space group, which resulted in a superposition of two orientations of the inhibitor in the pseudo-C_2-symmetrical PR. It should be noted that there exist two other wtPR/DRV structures (PDB codes 2IEN [30] and 1T3R [31]) that crystallized in a less common orthorhombic $P2_12_12$ space group with a single inhibitor orientation. These structures superpose with our structure with a root-mean-square deviation (RMSD) of Cα atoms of the protein of 0.3 and 0.4 Å, respectively, showing their high similarity. The one proton present in the catalytic aspartate dyad can be placed on either the OD1 or OD2 atoms of either the Asp25 or Asp25′ (Fig. 3A). These four variants were studied for both of the inhibitor orientations (Table IA). The flap water was included in the model; the others were discarded.

The wtPR/KI2 crystal structure was determined at an atomic resolution of 1.03 Å (PDB code 1NH0), which allowed the deduction of the protonation state of the catalytic aspartates (Fig. 3B)[6]. Further in the binding cavity, another carboxyl–carboxyl interaction was observed, namely between the Asp30 of the PR and the P2′ glutamate moiety of the KI2 (Fig. 3C). The highly accurate CG-OD1/OD2 and CD-OE1/OE2 distances only revealed a preference for a protonation at the OD2 of the Asp30 or the OE1 of the Glu-P2′ or both (cf. the K1, K2 and K3 variants in Table IB). Furthermore, two conformations, A and B, of the inhibitor benzyloxycarbonyl group with relative occupancies of 54 and 46%, respectively, were fitted to the electron density maps (EDM). Several PR residues were also refined to alternate conformations – the active-site examples are the Asp 30′ (54:47), Val 32′ (55:45), Val 82/Val 82′ (65:35) or Ile 84/Ile 84′ (61:39). The pairs of the major (K1-K3 variants, Table IB) and minor (K4-K8) conformations of all of the PR residues with alternate conformations of the KI2 inhibitor were constructed. In the B conformation of the inhibitor, the acceptor of the H22 atom of the KI2 could be either the OD2 atom of the Asp25′ (K4, K6 and K8 variants) or the O01 atom of the KI2 (K5, K7). Taken together, a set of eight variants was prepared for the PR/KI2 complex (Table IB). The flap water was included in the model; the others were discarded.

System Setup

The structures with the protonation variants shown in Table I were prepared using a special protocol developed with the aim of enabling a comparison of the stabilities of differentially protonated structures. Thus the steps of hydrogen-atom addition and relaxation (see below) were performed only for the D1, K1 and K4 variants. Only then were the protonation states exchanged to include all of the variants and the position of the added proton was optimized.

Hydrogen atoms were added to D1, K1 and K4 variant structures using the UCSF Chimera program[32] for the ligand and the Reduce[33] and LEaP programs available in the AMBER 10 simulation package[34] for the protein. To mimic the pH of 5.0 and 5.6 used for the crystallization experiments of the wtPR/DRV [28] and wtPR/KI2 [6] complexes, respectively, the arginine, lysine and histidine residues as well as the N-termini were modeled as positively charged, whereas the aspartic and glutamic acid side-chains (with the exceptions of the Asp25/25′, Asp30 and the P2′-Glu of the KI2) as well as the C-termini were in their anionic forms. The DRV inhibitor was neutral and the KI2 was in a mono- or dianionic state depending on the protonation

variant. The positions of the added hydrogens were relaxed using the SANDER module of AMBER by up to 10,000 steps of the steepest descent, and conjugate gradient optimizations until the root-mean-square gradient fell below 0.001 kcal/mol/Å. The position of the flap water was then optimized using the same criteria as above, because in the case of the DRV, the crystallographic position of the flap water was an average of two slightly differing positions for each inhibitor orientation.

The parameters for these AMBER calculations were as follows: the ff03 force field[35] was used for the protein and the General AMBER force field (GAFF)[36] for the ligands. The charges for the ligand were obtained using a restrained fit to the electrostatic potential (RESP) calculated at the HF/6-31G* level[37].

QM/MM Setup

The hybrid QM/MM calculations were set up as follows: for large QM regions (6-, 8- and 10-Å surroundings of the inhibitor), the QM part was treated using the semiempirical quantum chemical PM6 method[20] corrected for dispersion (PM6-D2; optimization) and hydrogen bonding (PM6-DH2; single-point energies)[22,23]. The hydrogen-bond corrections could not be used for optimizations because of their inability to tackle PT. All of the PM6 calculations were performed using the MOZYME linear-scaling algorithm available in the MOPAC code[21]. The benchmark QM/MM calculations were performed on the smallest region of the 3.0- (DRV) or 2.5-Å (KI2) surroundings of the inhibitor using the calculations on the DFT and compared with the PM6 level. An RI-DFT-D approach (the accelerated resolution-of-the-identity variant[38] enhanced with empirical dispersion[39] was used with the TPSS/TZVP//B-LYP/SVP functional/basis set combination for single-point and optimization, respectively) using the Turbomole program, version 6.2 [40]. To accelerate the SCF convergence, a levelshift of 0.25 a.u. was applied. The QM part was calculated in vacuum. The MM part was treated using AMBER and the parameters listed above.

The coupling between the QM and MM parts was done with an in-house program using a subtractive scheme of an ONIOM-type[41]. The protein/inhibitor complex was surrounded by the generalized Born (GB)[25] implicit solvent model. To speed up the convergence of optimizations, the outer part of the protein was kept frozen.

RESULTS AND DISCUSSION

DFT and PM6 QM/MM Optimizations

The RI-DFT-D QM/MM calculations (the B-LYP/SVP optimizations followed by the TPSS/TZVP single-point energies) were used as a reference for both the PR/DRV and PR/KI2 complexes (eight variants for each, see Table I). Thus, if the structures and energies of the DFT and PM6 QM/MM calculations differed qualitatively, conclusions were drawn from the former ones. For the feasibility of the DFT calculations, the size of the QM parts was selected not to exceed 450 atoms. Thus, the 3-Å surroundings of the Asp25/Ap25′ pair for the PR/DRV case and the 2.5-Å surroundings of the Asp30/Glu-P2′ pair including the Asp25/Asp25′ dyad for the PR/KI2 case were chosen. The single-point energies of the QM part in vacuum were sorted with respect to the most stable one (Table II).

The RI-DFT-D calculations on the PR/DRV system have revealed that either orientation of the DRV in the complex (A and B) yields two stable structures (D1, D2, and D5, D6 variants; see Table IIA) within 3 kcal/mol (for a discussion of the energy cutoff, see ref.[42]). The other variants are 4.3–13.0 kcal/mol less stable. It is important to bear in mind that because of the C_2-pseudosymmetrical structure of the PR/DRV complex, the D1–D6, D2–D5, D3–D8 and D4–D7 pairs are symmetry equivalents as regards the protonation state relative to the inhibitor orientation.

In contrast to the DFT, in the PM6-D2 QM/MM optimizations four of the eight structures resulted in a proton transfer in the active site (which transformed D1 and D3 to D2 and D6 and D8 to D5; denoted with an asterisk, Table IIA). To check the validity of such an observation, we investigated the heights of the PT barriers given by these two methods on the model systems derived from the PR/DRV and PR/KI2 complexes and compared their values to the benchmark values at the RI-MP2 and CCSD(T) levels. The preliminary data confirmed the well-known tendency of the DFT generalized gradient approximation (GGA) functionals to underestimate the reaction barrier heights[43] but also showed an even greater underestimation using the PM6-D2 [44]. Taking the PT into account, two energetically best structures at the PM6-DH2//PM6-D2 level corresponded to the D5 and D2 variants, the symmetrically equivalent pair. These two structures were the most stable ones at the DFT level as well. The minor consequences of the inability of the PM6-DH2//PM6-D2 approach to localize also the D1–D6 pair as equally stable are discussed below. It is of interest that even after the PT occurred, the D1, D3, D5 and D6 variants remained by 5.4–7.1 kcal/mol less

stable than the D8* structure. A visual inspection of the optimized structures showed that these variants became trapped in the local minima, differing in the geometry of the active site from the D2 and D8* structures. In the next section, we have investigated whether allowing more relaxation in the more distant surroundings of the active site could help bring these structures to the global minimum.

In the second system, the PR/KI2 complex, we started by studying the mono- and diprotonated variants of the Asp30/Glu-P2′ pair. For the former case (all of the variants except for the K3 and K8), the DFT QM/MM optimizations resulted in a PT from the Glu-P2′ of the KI2 to the Asp30 PR residue (Table IIB). Nevertheless, the PM6-D2 QM/MM optimizations re-

TABLE II
The relative energies (kcal/mol) of the QM parts (3-Å surroundings for DRV, 2.5 Å for KI2) of the protonation variants on the QM/MM optimized geometries

Variant	DFT	PM6
A. The wtPR/DRV complex		
D1	2.7	6.7[a]
D2	2.0	2.6
D3	6.0	7.1[a]
D4	13.0	20.2
D5	0.0	5.4
D6	1.1	5.5[a]
D7	6.7	16.7
D8	4.3	0.0[a]
B. The wtPR/KI2		
K1	0.0	0.3
K2	0.1[b]	0.0
K4	10.5	8.5
K5	20.2	9.8
K6	10.1[b]	9.0
K7	19.3[b]	7.9

[a] Denotes a proton transfer which transforms the D1 and D3 structures to D2 and the D6 and D8 structures to D5. [b] Denotes a proton transfer which transforms the K2 structure to K1, K6 to K4 and K7 to K5

sulted in an intermediate structure in which the HE1 proton is localized between the two oxygen atoms of the Glu-P2′ and Asp30 with typical O–H distances of 1.2–1.3 Å. This again can be explained by the shape of the PT curve for the PM6-D2 method, which for this model system has an energy minimum at the intermediate positions, unlike the DFT. However, the preliminary data show that even the DFT curve differs from the MP2 and CCSD(T) ones in the details of the shape and energetics[44]. In the diprotonated case (K3 and K8 variants), a PT occurred in neither the DFT nor PM6-D2 QM/MM optimizations. A comparison of the mono- and diprotonated variants to the crystal structure revealed large deviations of the Asp30/Glu-P2′ pair for the diprotonated variants as opposed to the energetically most stable monoprotonated variants (RMSD of 0.89 vs 0.14 and 0.94 vs 0.19 Å, respectively). We thus conclude that the monoprotonated variant is going to be more probable than the diprotonated one.

Comparing the relative stabilities of the A and B conformations of the KI2 and the PR residues in the PR/KI2 complex, the PM6-D2 and DFT QM/MM optimizations consistently show that the former is more stable by roughly 10 or 10–20 kcal/mol, respectively (Table IIB). Although the energy difference is too high (several possible reasons are discussed below) to allow the population of the B conformations, this result qualitatively agrees with the higher occupation of 54–65% for the A conformations observed in the crystal structure[6].

Another structural detail investigated in the PR/KI2 complex was the hydrogen bond formed by the hydroxyl of the KI2 (O22–H22 atoms). In the intermolecular case, i.e. binding to the OD2 atom of the Asp25′ (K4, K6 and K8 variants), the atoms involved in this hydrogen bond did not undergo any sizeable movements. However, in the intra-molecular case, i.e. binding to the O01 oxygen of the KI2 (K5, K7 variants), the O01 oxygen moved away (in the direction of its position in the A conformation) to increase its distance to the O22 atom from 2.28 Å in the B-conformation of the X-ray structure to 3.01 Å in the DFT QM/MM optimized structure. This shift of the O01 oxygen suggests that this intramolecular hydrogen bond would not be stable.

Size of the QM Moving Region in the QM/MM Calculations

To investigate the effect of the size of the moving QM part on the determination of the most stable protonation states, a region comprising the 6-Å surroundings of the inhibitor in the PR/DRV complex was set up and extended to include the 8- and 10-Å surroundings. The respective sizes of

these QM regions were 967, 1325 and 1696 atoms for the PR/DRV system, and 1050, 1374 and 1733 atoms for the PR/KI2. It should be noted that such an extension in a molecular system is not a smooth one as charged groups can be included in the QM part upon its extension and affect substantially its electrostatics.

In the PR/DRV case, the results were qualitatively similar to those found on the small QM part of 3 Å (cf. Table IIA); the same four variants underwent PT (D1 and D3 to D2 and D6 and D8 to D5; denoted with an asterisk, Table IIIA). Again, considering the PT, the D2–D5 equivalent pair had the lowest energy, whereas the other variants were less stable by 15.9–39.0 kcal/mol (Table IIIA) as compared to 16.7–20.2 kcal/mol and 1.1–13.0 kcal/mol for the smallest 3-Å region in the PM6-DH2//PM6-D2 and DFT QM/MM calculations, respectively (cf. Table IIA). This comparison shows that even allowing large parts of the protein move does not alleviate the problem of trapping the unstable D4 and D7 variants in the local minima of higher energy.

A similar set of calculations has been conducted for eight variants (K1–K8) of the PR/KI2 complex. For the monoprotonated Glu-P2′/Asp30 pair, the relative energies of the QM part are shown in Table IIIB. The sizes of the 6- and 8-Å surroundings of the inhibitor are energetically consistent with the DFT and corrected PM6 QM/MM optimizations on a small 2.5-Å region in that the QM parts of the A conformations (K1, K2 variants) are by 13.9–20.8 kcal/mol more stable than the B conformations (K4–K7 variants). In the larger region of 10 Å, however, another variant (K5 variant) approached the stability of the most stable variant, K1. Due to the large size of the QM region (1733 atoms), we wanted to verify whether the energy differences stemmed from the differences in the active site because of the different protonation variants (that is the goal of our investigation) or whether some structurally unrelated changes occurred at more distant parts of the PR (these would be unwanted effects that we would wish to avoid). We therefore reoptimized the 10-Å region optimized geometries using a smaller 8-Å region and compared the relative single-point energies using the 8-Å region. The last column of Table IIIB shows that the K5 variant again became less stable, which suggests that unrelated structural changes in the farther (in this case 8–10 Å) region can significantly influence the energetics of the QM part. A visual inspection of the optimized geometries revealed, like in the PM6-DH2//PM6-D2 QM/MM calculations in the smallest 2.5-Å region, that in all the monoprotonated structures the HE1 proton ended in an intermediate position between the OE1 of the Glu-P2′ and the OD2 of the Asp30. This corresponds to the shift of the position of the mini-

mum on the PM6-D2 hypersurface (HE1–OE1 and HE1–OD2 distances of 1.2 and 1.3 Å, respectively) relative to DFT (1.6 and 1.1 Å)[44].

The diprotonated variant of the A and B models of the PR/KI2 complex (K3 and K8 variants, respectively) showed a large structural deviation from the crystal structure, during which the Asp30 side chain turned outwards

TABLE III
The relative energies (kcal/mol) of the QM parts for the protonation variants on the PM6-D2 QM/MM optimized geometries with a varying size of the moving QM part

Variant	6 Å	8 Å	10 Å
A. The wtPR/DRV complex			
D1	17.6[a]	3.4[a]	20.5[a]
D2	7.5	7.5	1.3
D3	0.0[a]	0.0[a]	21.8[a]
D4	28.0	23.0	16.4
D5	18.3	3.8	26.8
D6	19.8[a]	4.1[a]	0.0[a]
D7	35.3	33.4	39.0
D8	25.3[a]	2.9[a]	8.2[a]

Variant	6 Å	8 Å	10 Å	8 Å//10 Å[b]
B. The PR/KI2 complex				
K1	0.0	0.0	0.0	0.0
K2	3.7	2.9	5.7	6.7
K4	15.9	17.3	8.8	11.3
K5	13.9	18.9	0.9	10.5
K6	17.5	18.0	19.8	21.0
K7	20.8	20.7	14.8	20.3

[a] Denotes a proton transfer which transforms the D1 and D3 structures to D2 and the D6 and D8 structures to D5. [b] The 8 Å//10 Å column means an 8-Å region reoptimization and single-point energies on the 10-Å region optimized geometries.

from its original position. Table IV shows the RMSDs with respect to the X-ray structure of the non-hydrogen atoms of the mono- and diprotonated Asp30/Glu-P2′ dyad obtained using corrected PM6 QM/MM optimizations using different sizes of the QM part. The small values for the monoprotonated variants found consistently for both the A and B conformations (cf. Table IVA and IVB) suggest that either the models of the diprotonated variants are more sensitive to the used approximations such as the lack of explicit water molecules or that the Asp30/Glu-P2′ dyad is only singly protonated.

Methodological Issues

In this study, we present hybrid QM/MM calculations on a biomolecular system in which protons play a pivotal role. For two molecular complexes, a QM region of approximately 400 atoms has been chosen, including not only the protonated carboxylate pairs but also their close (~3-Å) surroundings. The QM part was treated with the DFT method, which has been used

TABLE IV
The root-mean-square deviations (Å) with respect to the crystal structure of the non-hydrogen atoms of the mono- and diprotonated Asp30/Glu-P2′ dyad obtained using QM/MM optimizations with different sizes of the QM parts

QM region, Å	K1:OD2 or OE1, monoprotonated	K3:OD2, OE1, diprotonated
A. The K1 and K3 variants		
2.5	0.14	1.45
6	0.24	0.89
8	0.21	1.05
10	0.24	0.93
QM region, Å	K4–K7:OD2 or OE1, monoprotonated	K8:OD2, OE1, diprotonated
B. The most stable of K4-K7 and K8 variants		
2.5	0.16	1.10
6	0.22	0.52
8	0.21	0.52
10	0.24	0.63

frequently in biomolecular QM/MM calculations[45–47]. Yet, even for the accelerated RI variant[38] of the DFT, this size currently represents the upper limit. However, it is important to test the effect of increasing the QM part further.

To be able to treat the biomolecular systems of several thousand atoms, we have turned to semiempirical methods and because of its superior performance we have chosen the PM6 method[20] with a linear-scaling algorithm[21]. However, its description of the noncovalent interactions had to be enhanced by introducing empirical corrections for dispersion and hydrogen-bonding[22,23]. The newly developed method has been successfully applied for two biomolecular systems[48,49].

The present study is the first one to study the protonation phenomena in biomolecules using the corrected PM6-DH2 method. It should be stressed that a QM approach is the only one to be used (in contrast to the MM methods) to describe correctly a molecular system in which PT phenomena can occur[11]. For the two molecular complexes studied here, a PT occurred in one of them (PR/KI2 case) on the DFT level, while in the other system (PR/DRV; D1–D8 variants) no PT did take place. The results from the corrected PM6-DH2 method differed qualitatively; in one system (PR/KI2), the proton ended between the two oxygen atoms, whereas in the second system (PR/DRV) a PT was observed. We have therefore conducted a preliminary study of the PT barrier heights on a model of monoprotonated carboxylate pairs using high-level computational chemistry methods. The results have not only confirmed a well-known tendency of DFT GGA functionals to underestimate the reaction barriers[43] but also showed an even greater underestimation on the PM6-D2 level and shifting of the minimum toward the intermediate positions of the proton between the two oxygens[44]. This finding thus points to the need for better corrections or even new reparametrizations of the PM6 method which would also describe PT. Moreover, our results show that the frequently used DFT calculations must be taken with caution and preferably checked against higher-level QM calculations.

Owing to recent developments in linear-scaling semiempirical quantum chemical methods, we have been able to increase the size of the QM part stepwise up to approximately 1733 atoms (10-Å region surrounding the ligand) and optimize it at the PM6-D2 level. Although there were quantitative differences, the most stable variants (D2, D5 and K1, taking PT into account; cf. Table III) were found consistently in the 2.5–3-, 6-, 8- and 10-Å regions. However, in the 10-Å region (and significantly more in 12-Å regions, not shown) optimizations, unrelated structural changes occurred far

from the active site that affected the relative stabilities. Poised between the Scylla of allowing sufficient relaxation and the Charybdis of avoiding structural changes far from the active site, we recommend an optimal size of the QM region for HIV protease studies of ~8-Å surrounding the ligand. We add that this size may differ for other protein/ligand systems, depending on the flexibility of the complex, the hydration of the active site, etc.

A small note regarding the preparation of the structures should be made here. We have endeavored to develop a protocol which would be useful not only for a comparison of various protonation variants in the HIV protease but more generally of possible constitutional isomers (tautomers, conformers) in complex biomolecular systems. We therefore urge that the variant structures be prepared carefully and consistently (cf. the hydrogen-atom addition and relaxation performed solely for representative structures, only then setting up the protonation variants) to eliminate unwanted geometrical and energy differences.

As mentioned, the current protocol can be utilized in a QM-based scoring of the HIV protease/ligand complexes to select the most probable protonation variants for further scoring calculations. Including the PT, the PM6-DH2//PM6-D2 method has correctly found the D2–D5 pair as the most stable (Table IIA). However, the D1–D6 pair, which was the second most stable on the DFT level, transformed due to PT into the former one. This was caused by the underestimation of the PT barriers on the PM6-D2 level, which is even greater than that of the DFT level as shown by comparison with high-level MP2 and CCSD(T) calculations. The less stable D3–D8 pair again transformed in the PM6-D2 QM/MM optimizations to the most stable D2–D5 pair owing to PT, indicating its low stability. Finally, the least stable D4–D7 pair on the DFT level was also the least stable on the PM6-DH2//PM6-D2 level. Taken together, a semi-quantitative agreement of the corrected PM6 energies with the DFT ones can be obtained in cases where an incorrect PT does not occur. This points to a need of further adjustment of the corrected PM6 protocol for biomolecular systems by either restraining the O–H bonds in question or introducing another reparametrization.

Biomolecular Findings

In order to draw conclusions for the two HIV protease/inhibitor complexes studied in this work, we must be aware of the strengths and weaknesses of the crystallographic structures and computational methods/protocols employed. As already mentioned, the wtPR/DRV complex crystallized in the

common hexagonal $P6_1$ space group, in which the electron-density maps (EDM) for the two orientations of the inhibitor in the pseudo-C_2-symmetrical enzyme overlapped. Some inaccuracies of the starting structure may hence stem from fitting the two inhibitor orientations into these EDMs. In contrast, the highly accurate X-ray structure of the PR/KI2 complex allowed an inference of the proton locations of the catalytic Asp25/25′ while providing hints for the Asp30/Glu-P2′ pair. The P2 benzyloxycarbonyl of the KI2 inhibitor, on the other hand, had a poor omit EDM, which was explained by its higher mobility and an alternative conformation[6].

As regards the accuracy of the PM6-DH2 method, it has been established on several datasets of noncovalently interacting model complexes that it performs equally well as the DFT-D within a chemical accuracy[22,23]. However, we should bear in mind that these values hold for equilibrium geometries obtained on accurate MP2 and CCSD(T) geometries[50]. For less accurate geometries (as for example the PM6-D2 level used here), the error will increase.

For the PR/DRV complex, which represents a general case of the HIV PR/inhibitor complexes in which two orientations of the inhibitor were refined, we have found that the symmetry-related pairs of the protonation variants are also energy-related. The structural similarities of the two inhibitor orientations in PR were acknowledged in analyses of PR/DRV X-ray structures[30]. However, in the calculations of HIV protease/inhibitor complexes, it has been a common practice to use only the first orientation of the inhibitor (see e.g. ref.[17] and the references therein or ref.[51]). To the best of our knowledge, we present in this paper the very first study to confirm that, if using a QM-relaxed region, only one inhibitor orientation is sufficient to correctly describe the energetics in the active site of HIV PR/inhibitor complexes. In the DFT QM/MM calculations, the D1–D6 and D2–D5 symmetry-related pairs proved to be the most stable, separated from the less stable pairs by 4.3–13.0 kcal/mol. In an atomic-resolution (1.1 Å) crystal structure of DRV in complex with the PR Val82Ala mutant, a streak of positive electron density in the omit map appeared, suggesting the location of a proton[30]. This finding presents an experimental verification of our approach, as this corresponds to our stable D2 variant. In a molecular mechanics-based study of the DRV and a related inhibitor amprenavir (APV) binding to the PR, several protonation variants were tried[51]; although an equivalent of the D2 variant (beware of the fact that the PR chain notation is reversed with respect to our study) had the most favorable interaction energy with the APV (see the Supporting Information to ref.[51]), another variant (an equivalent of D3) was chosen because of its structural

similarity to the crystal structure after a molecular dynamics run. However, force-field-based methods may be unreliable for a structural description of the active site of enzymes, in which quantum effects such as PT or charge redistribution can occur.

The X-ray structure of the PR/KI2 complex is unique in that the atomic resolution of 1.03 Å enabled the deduction of the protonation states of the catalytic aspartates[6]. However, despite the high quality of the crystal structure, three molecular features remained questionable: (i) the protonation state of the Asp30/Glu-P2′ carboxylate pair, (ii) the relative stabilities of the A and B conformations of the P2 group of KI2 and several PR residues, and (iii) the acceptor of the hydrogen bond from the KI2 hydroxyl. The DFT QM/MM calculations revealed that of the three possible variants of the Asp30/Glu-P2′ protonation, the diprotonated variants (K3 and K8) could be excluded based on geometrical criteria, whereas the inhibitor OE1 oxygen-protonated variants (K2, K6 and K7) transformed during optimizations into the respective Asp30:OD2 protonated variants of K1, K4 and K5. The higher stability of the K1, K4 and K5 variants was also corroborated by the preliminary high-level QM calculations on a model system derived from the Asp30/Glu-P2′ pair of this crystal structure. We have thus determined using our computations that the OD2 atom of Asp30, and not the OE1 of Glu-P2′of KI2, will be protonated in the PR/KI2 complex. This is an interesting and farther reaching conclusion since the Glu residue is present as the P2′ moiety not only in several inhibitors[27,52,53] but also in the substrate derived from the CA-p2 cleavage site[54,55].

The second molecular feature of the PR/KI2 complex which deserved attention was the stability of the major (A) and minor (B) conformations of the P2 moiety of the KI2 and surrounding PR residues. The DFT QM/MM calculations have identified the A conformation as more stable than B, which is in qualitative agreement with the higher occupancy of the former over the latter observed in the crystal structure. However, the energy difference of 10.1–20.2 kcal/mol (Table IIB) is too high to interpret the crystallographic occupancy ratio of 54:46. Several limitations of the presented computational approach as well as crystallographic issues can be responsible. We could envisage that allowing structural relaxation of the active site surroundings would bring the two alternative conformations closer in energy, but this possibility was disproved by the corrected PM6 QM/MM calculations in the larger regions (Table IIIB). The lack of the explicit description of the vibrational energy and the dynamics may be another reason for the high energy difference. Regarding the X-ray structure, the experimental electron density maps (EDM) reveal that the P2 moiety is quite

flexible (among others reflected by the presence of two alternative conformations) as compared to the rest of the inhibitor. Upon a closer inspection of the EDM of the PR/KI2 complex, we observed that whereas the A conformation of the P2 was fitted into a well-defined EDM, the alternative B conformation could be fitted in several ways. It may be that some of these possible alternative B conformations would be lower in energy than the one present in the crystal structure.

The third molecular feature was the identity of the acceptor oxygen for the hydrogen bond of the KI2 hydroxyl in the B conformation. It was consistently found, for both the inter- and intra-molecular variants, that the O01 oxygen deviated from its crystallographic position toward a position found in the A conformation. This can either be a proof that (i) the hydrogen-bond acceptor of the KI2 hydroxyl would rather be OD2 of Asp25′ than O01 of the KI2 or (ii) the B conformation of the P2 moiety present in the crystal may be less stable than a potential other alternative conformation which could be fitted into EDM (see above). In summary, even very high quality crystal structures, such as that of the PR/KI2 complex, pose some unknowns for computational chemists. However, these may be elucidated by means of calculations.

CONCLUSIONS

Based on the results of this pilot computational study, several methodological and biomolecular conclusions have been drawn.

1) We have presented a novel computational protocol for determining the probable protonation states based on the quantum mechanical energy. This approach is general and can be utilized for assessing the stabilities of various conformers/tautomers in biomolecular systems.

2) A comparison with the benchmark MP2 and CCSD(T) data on protonated carboxylate pair model systems revealed that the DFT using a GGA functional and even more the PM6-D2 underestimate the PT barriers.

3) The corrected PM6 QM/MM calculations using a QM region extending up to 3 Å from the inhibitor found the same stable protonation states in the two HIV protease complexes as DFT. The extension of the QM region from 3 to 8 Å gave the same qualitative picture on the corrected PM6 level.

4) Allowing relaxation of overly large regions in the QM part (>10 Å) increases the risk of distant unrelated structural changes occurring, which can affect the energetics of the active site.

5) The symmetry-related pairs of the HIV PR/inhibitor complexes with two orientations of the inhibitors are also energy-related. It has been shown here on the PR/DRV complex.

6) The Asp30/Glu-P2′ carboxylate pair is monoprotonated on the Asp30 as shown for the PR/KI2 complex by (i) the geometrical instability of the diprotonated variants and (ii) the PT in the QM/MM calculations, corroborated by PT transfer barriers in model systems obtained with high-level QM calculations. This finding has consequences for other HIV PR inhibitors and substrates containing a Glu moiety at P2′.

7) The acceptor of the hydrogen bond from the hydroxyl group of the KI2 is most probably the OD2 oxygen of the Asp25′. There is, however, a possibility that an intramolecular hydrogen bond could form transiently with a structure from a dynamic equilibrium of alternative P2 conformations.

8) The major A conformation of the KI2 and the surrounding PR residues is more stable than the B conformation. This agrees with the experimental crystallographic finding of its higher occupation.

We thank Dr. J. Brynda for the electron-density maps of wtPR/KI2, Dr. P. Řezáčová and Dr. K. Grantz-Šašková for releasing the unpublished wtPR/DRV structure prior to the publication. Additionally, we are indebted to the Reviewer 1 for the valuable comments and suggestions. This work was a part of Research Project No. Z40550506 of the Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic and was supported by Grants No. LC512 and No. 1M0508 from the Ministry of Education, Youth and Sports of the Czech Republic and the Grant No. P208/11/0295 from the Grant Agendy of the Czech Republic. The support of the Praemium Academiae, Academy of Sciences of the Czech Republic, awarded to P.H. in 2007, is also acknowledged.

REFERENCES

1. Bombarda E., Ullmann G. M.: *J. Phys. Chem. B* **2010**, *114*, 1994.
2. Tawa G. J., Topol I. A., Burt S. K., Erickson J. W.: *J. Am. Chem. Soc.* **1998**, *120*, 8856.
3. Bashford D., Karplus M.: *J. Phys. Chem.* **1991**, *95*, 9556.
4. Ullmann G. M.: *J. Phys. Chem. B* **2003**, *107*, 1263.
5. Berg J. M., Tymoczko J. L., Stryer L.: *Biochemistry*, 5th ed. W. H. Freeman, New York 2002.
6. Brynda J., Řezáčová P., Fábry M., Hořejší M., Stouračová R., Sedláček J., Souček M., Hradilek M., Lepšík M., Konvalinka J.: *J. Med. Chem.* **2004**, *47*, 2030.
7. Podjarny A., Cachau R. E., Schneider T., Van Zandt M., Joachimiak A.: *Cell. Mol. Life Sci.* **2004**, *61*, 763.
8. Trylska J., Antosiewicz J., Geller M., Hodge C. N., Klabe R. M., Head M. S., Gilson M. K.: *Protein Sci.* **1999**, *8*, 180.
9. Rajamani R., Reynolds C. H.: *J. Med. Chem.* **2004**, *47*, 5159.

10. Yu N., Hayik S. A., Wang B., Liao N., Reynolds C. H., Merz K. M.: *J. Chem. Theory Comput.* **2006**, *2*, 1057.
11. Piana S., Sebastiani D., Carloni P., Parrinello M.: *J. Am. Chem. Soc.* **2001**, *123*, 8730.
12. Harte W. E., Beveridge D. L.: *J. Am. Chem. Soc.* **1993**, *115*, 3883.
13. Park H., Lee S.: *J. Am. Chem. Soc.* **2003**, *125*, 16416.
14. Hyland L. J., Tomaszek T. A. J., Meek T. D.: *Biochemistry* **1991**, *30*, 8454.
15. Northrop D. B.: *Acc. Chem. Res.* **2001**, *34*, 790.
16. Piana S., Carloni P.: *Proteins Struct. Funct. Genet.* **2000**, *39*, 26.
17. Lepšík M., Kříž Z., Havlas Z.: *Proteins Struct. Funct. Bioinf.* **2004**, *57*, 279.
18. Velazquez-Campoy A., Luque I., Todd M. J., Milutinovich M., Kiso Y., Freire E.: *Protein Sci.* **2000**, *9*, 1801.
19. Sirois S., Proynov E. I., Truchon J. F., Tsoukas C. M., Salahub D. R.: *J. Comput. Chem.* **2003**, *24*, 1110.
20. Stewart J. J. P.: *J. Mol. Mod.* **2007**, *13*, 1173.
21. Stewart J. J. P.: *J. Mol. Mod.* **2009**, *15*, 765.
22. Řezáč J., Fanfrlík J., Salahub D., Hobza P.: *J. Chem. Theory Comput.* **2009**, *5*, 1749.
23. Korth M., Pitoňák M., Řezáč J., Hobza P.: *J. Chem. Theory Comput.* **2010**, *6*, 344.
24. Klamt A., Schuurmann G.: *J. Chem. Soc., Perkin Trans. 2* **1993**, 799.
25. Tsui V., Case D. A.: *Biopolymers* **2001**, *56*, 275.
26. Koh Y., Nakata H., Maeda K., Ogata H., Bilcer G., Devasamudram T., Kincaid J. F., Boross P., Wang Y. F., Ties Y. F., Volarath P., Gaddis L., Harrison R. W., Weber I. T., Ghosh A. K., Mitsuya H.: *Antimicrob. Agents Chemother.* **2003**, *47*, 3123.
27. Rinnová M., Hradilek M., Bařinka C., Weber J., Souček M., Vondrášek J., Klimkait T., Konvalinka J.: *Arch. Biochem. Biophys.* **2000**, *382*, 22.
28. Kožíšek M., Šašková K. G., Lepšík M., Brynda J., Konvalinka J., Řezáčová P.: Manuscript in preparation.
29. Brynda J., Řezáčová P., Fábry M., Hořejší M., Stouračová R., Souček M., Hradilek M., Konvalinka J., Sedláček J.: *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 1943.
30. Tie Y. F., Boross P. I., Wang Y. F., Gaddis L., Hussain A. K., Leshchenko S., Ghosh A. K., Louis J. M., Harrison R. W., Weber I. T.: *J. Mol. Biol.* **2004**, *338*, 341.
31. Surleraux D. L. N. G., Tahri A., Verschueren W. G., Pille G. M. E., de Kock H. A., Jonckers T. H. M, Peeters A., De Meyer S., Azijn H., Pauwels R., de Bethune M. P., King N. M., Prabu-Jeyabalan M., Schiffer C. A., Wigerinck P. B. T. P.: *J. Med. Chem.* **2005**, *48*, 1813.
32. 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.
33. Word J. M., Lovell S. C., Richardson J. S., Richardson D. C.: *J. Mol. Biol.* **1999**, *285*, 1735.
34. Case D. A., Darden T. A., Cheatham III T. E., Simmerling C. L., Wang J., Duke R. E., Luo R., Crowley M., Walker R. C., Zhang W., Merz K. M., Wang B., Hayik S., Roitberg A., Seabra G., Kolossváry I., Wong K. F., Paesani F., Vanicek J., Wu X., Brozell S. R., Steinbrecher T., Gohlke H., Yang L., Tan C., Mongan J., Hornak V., Cui G., Mathews D. H., Seetin M. G., Sagui C., Babin V., Kollman P. A.: *AMBER 10*. University of California, San Francisco 2008.
35. 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.
36. Wang J. M., Cieplak P., Kollman P. A.: *J. Comput. Chem.* **2000**, *21*, 1049.
37. Bayly C. I., Cieplak P., Cornell W. D., Kollman P. A.: *J. Phys. Chem.* **1993**, *97*, 10269.

38. Feyereisen M., Fitzgerald G., Komornicki A.: *Chem. Phys. Lett.* **1993**, *208*, 359.
39. Jurečka P., Černý J., Hobza P., Salahub D. R.: *J. Comput. Chem.* **2007**, *28*, 555.
40. Ahlrichs R., Bär M., Häser M., Horn H., Kölmel C.: *Chem. Phys. Lett.* **1989**, *162*, 165.
41. a) Svensson M., Humbel S., Froese R. D. J., Matsubara T., Sieber S., Morokuma K.: *J. Phys. Chem.* **1996**, *100*, 19357; b) Dapprich S., Komaromi I., Byun K. S., Morokuma K., Frisch M. J.: *J. Mol. Struct. – THEOCHEM* **1999**, *461–462*, 1.
42. Fanfrlík J., Brynda J., Řezáč J., Hobza P., Lepšík M.: *J. Phys. Chem. B* **2008**, *112*, 15094.
43. Koch W., Holthausen M. C.: *A Chemist's Guide to Density Functional Theory*, 2nd ed. Wiley-VCH, Weinheim 2000.
44. Granatier J., Fanfrlík J., Hobza P., Lepšík M.: Unpublished results.
45. Banáš P., Jurečka P, Walter N. G., Šponer J., Otyepka M: *Methods* **2009**, *49*, 202.
46. Cui Q., Karplus M.: *J. Phys. Chem. B* **2002**, *106*, 1768.
47. Senn H. M., Thiel W.: *Angew. Chem. Int. Ed.* **2009**, *48*, 1198.
48. Fanfrlík J., Bronowska A. K., Řezáč J., Přenosil O., Konvalinka J., Hobza P.: *J. Phys. Chem. B* **2010**, *114*, 12666.
49. Dobeš P., Fanfrlík J., Řezáč J., Otyepka M., Hobza P.: *J. Comput.-Aided Mol. Des.* **2011**, *25*, 223.
50. Jurečka P., Šponer J., Černý J., Hobza P.: *Phys. Chem. Chem. Phys.* **2006,** *8*, 1985.
51. Hou T. J., Yu R.: *J. Med. Chem.* **2007**, *50*, 1177.
52. Urban J., Konvalinka J., Stehlíková J., Gregorova E., Majer P., Souček M., Andreansky M., Fábry M., Štrop P.: *FEBS Lett.* **1992**, *298*, 9.
53. Weber I. T., Wu J., Adomat J., Harrison R. W., Kimmel A. R., Wondrak E. M., Louis J. M.: *Eur. J. Biochem.* **1997**, *249*, 523.
54. Richards A. D., Phylip L. H., Farmerie W. G., Scarborough P. E., Alvarez A., Dunn B. M., Hirel P. H., Konvalinka J., Štrop P., Pavlíčková L., Kostka V., Kay J.: *J. Biol. Chem.* **1990**, *265*, 7733.
55. Prabu-Jeyabalan M., Nalivaika E., Schiffer C. A.: *J. Mol. Biol.* **2000**, *301*, 1207.

Appendix E

BIOLOGICAL CRYSTALLOGRAPHY
ISSN 1399-0047

Atomic resolution crystal structure of Sapp2p, a secreted aspartic protease from *Candida parapsilosis*

Jiří Dostál,[a,b]* Adam Pecina,[a,b] Olga Hrušková-Heidingsfeldová,[a,b] Lucie Marečková,[a] Iva Pichová,[a,b] Pavlina Řezáčová,[a,c] Martin Lepšík[a,b] and Jiří Brynda[a,c]*

[a]Institute of Organic Chemistry and Biochemistry (IOCB), Academy of Sciences of the Czech Republic, Flemingovo náměstí 2, 166 10 Prague 6, Czech Republic, [b]Gilead Sciences and IOCB Research Centre, Flemingovo náměstí 2, 166 10 Prague 6, Czech Republic, and [c]Institute of Molecular Genetics, Academy of Sciences of the Czech Republic v.v.i., Vídeňská 1083, 142 20 Prague 6, Czech Republic. *Correspondence e-mail: dostalj@uochb.cas.cz, brynda@uochb.cas.cz

Received 2 September 2015
Accepted 13 October 2015

Edited by Z. Dauter, Argonne National Laboratory, USA

Keywords: aspartic protease; *Candida parapsilosis*; Sapp2p; crystal structure; ultrahigh resolution; interaction energy; quantum mechanics.

PDB reference: Sapp2p, 4y9w

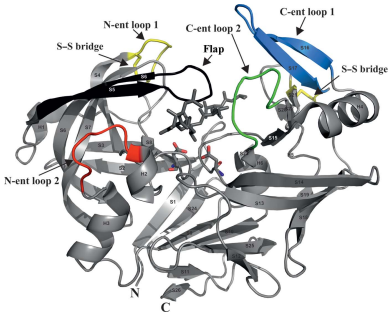
Supporting information: this article has supporting information at journals.iucr.org/d

The virulence of the *Candida* pathogens is enhanced by the production of secreted aspartic proteases, which therefore represent possible targets for drug design. Here, the crystal structure of the secreted aspartic protease Sapp2p from *Candida parapsilosis* was determined. Sapp2p was isolated from its natural source and crystallized in complex with pepstatin A, a classical aspartic protease inhibitor. The atomic resolution of 0.83 Å allowed the protonation states of the active-site residues to be inferred. A detailed comparison of the structure of Sapp2p with the structure of Sapp1p, the most abundant *C. parapsilosis* secreted aspartic protease, was performed. The analysis, which included advanced quantum-chemical interaction-energy calculations, uncovered molecular details that allowed the experimentally observed equipotent inhibition of both isoenzymes by pepstatin A to be rationalized.

1. Introduction

Candida parapsilosis is an opportunistic fungal pathogen. Although it is less common and less virulent than *C. albicans*, it causes a wide variety of hospital-acquired infections and presents a serious problem, particularly in neonatal intensive-care units (Leibovitz *et al.*, 2013; Pammi *et al.*, 2013; Trofa *et al.*, 2008). *C. parapsilosis* is an exogenous pathogen that often forms biofilms on catheters and other inserted devices, and it has been isolated from the hands of healthcare workers more frequently than other yeast species (Pammi *et al.*, 2013; Pfaller *et al.*, 2010). *C. parapsilosis* has also been isolated from a variety of natural sources, including soil, insects and domestic animals (Pryszcz *et al.*, 2013; Trofa *et al.*, 2008).

The success of pathogenic *Candida* species in colonizing and infecting various host niches relies on several specific features, such as efficient adherence to host surfaces, morphological diversity, biofilm formation, adaptability of metabolism and secretion of hydrolytic enzymes. The extracellular hydrolases, namely aspartic proteases, lipases and phospholipases, facilitate the penetration of the pathogens through host tissues. Secreted aspartic proteases (Saps) of pathogenic *Candida* spp. have broad substrate specificities and degrade a wide variety of host protein substrates ranging from structural proteins to immunoglobulins (Hrušková-Heidingsfeldová, 2008; Hube & Naglik, 2001). *SAP* genes usually occur



in gene families, which enables differential regulation according to ambient conditions and the stage of infection (Naglik *et al.*, 2004). The properties, regulation and evolution of the *SAP* gene family have been extensively studied, particularly in *C. albicans*, which possesses ten Sap isoenzymes. Four of the *C. albicans* Saps have been crystallized and structurally characterized (Abad-Zapatero *et al.*, 1996; Borelli *et al.*, 2007, 2008; Cutfield *et al.*, 1995). *C. parapsilosis* has long been considered to have only three genes encoding Saps. However, sequencing of the full *C. parapsilosis* genome enabled an *in silico* analysis that revealed up to 14 potential Sap-encoding sequences (Parra-Ortega *et al.*, 2009). This raised questions about the regulation of the individual *SAPP* genes and the properties of the putative protease isoenzymes.

Characterization of the first *C. parapsilosis* isoenzyme, Sapp1p, was facilitated by the fact that its expression can be induced by the presence of an exogenous protein as a sole source of nitrogen. Sufficient amounts of Sapp1p for crystallization can thus be easily obtained directly from *C. parapsilosis* culture supernatant. In our previous studies, we determined the crystal structures of Sapp1p in complex with pepstatin A, a classical aspartic protease inhibitor, and with ritonavir, a clinically used HIV protease inhibitor (Dostál *et al.*, 2009, 2012).

Expression of the second *C. parapsilosis* isoenzyme, Sapp2p, cannot be induced by a particular nitrogen source and its abundance is much lower than that of Sapp1p. When a protein is used as a source of nitrogen, Sapp2p usually constitutes less than 10% of the proteases recovered from the medium (Fusek *et al.*, 1993; Hrušková-Heidingsfeldová, 2008). In addition, two homologues of Sapp2p sharing 91.5% identity occur in the *C. parapsilosis* genome (Dostál *et al.*, 2015). Nevertheless, we succeeded in purifying and crystallizing one of the Sapp2p homologues, namely Sapp2p/CPAR2_102580 (entry CPAR2_102580 in the *Candida* Genome Database is identical to entry A47701 in the NCBI). Here, we report its structure in complex with pepstatin A determined at an atomic resolution of 0.825 Å. To understand the differences in pepstatin A binding to Sapp1p and Sapp2p on an accurate quantitative basis, we employed quantum-mechanical (QM) calculations to evaluate the interactions of the active-site residues with the inhibitor.

2. Materials and methods

2.1. Protein preparation

Sapp2p was purified from its natural source. *C. parapsilosis* strain P-69 was obtained from the mycological collection of the Faculty of Medicine, Palacky University, Olomouc, Czech Republic. The yeast was cultivated in YCB–BSA medium [1.2%(*w*/*v*) yeast carbon base, 0.2%(*w*/*v*) BSA, 15 m*M* sodium citrate pH 4.0] for 72 h at 303 K in a rotation shaker. The cells were harvested by centrifugation (5000*g* for 15 min). Isolation and purification of the mixture of Sapp2p and Sapp1p isoenzymes was performed as described in Hrušková-Heidingsfeldová *et al.* (2009) and Dostál *et al.* (2009). The efficiency of the purification steps was analyzed using SDS–PAGE, Western blotting and activity assays. Protein analyses and proteolytic activity assays were carried out as described previously (Pichová *et al.*, 2001; Dostál *et al.*, 2003; Merkerová *et al.*, 2006).

2.2. Mass-spectrometric analysis

Prior to identification by mass spectrometry (MS), proteins were separated using SDS–PAGE, stained with Coomassie Brilliant Blue R-250 (Thermo Scientific), excised from the gel and digested in-gel with either Trypsin Gold (Promega) or Endoproteinase Asp-N (Roche Applied Science). The resulting peptides were solubilized in 30 µl 0.1%(*w*/*v*) formic acid and injected into an Ultimate 3000 RSCL Nano LC (Thermo Scientific). The peptides were trapped on an Acclaim PepMap100 C18 trap column (3 µm particles, 100 Å, 75 µm × 2 cm; Thermo Scientific) and separated using an Acclaim PepMap RSCL C18 column (2 µm particles, 100 Å, 75 µm × 15 cm; Thermo Scientific). The mobile phase consisted of 0.1% formic acid in water (*A*) and 0.1% formic acid in acetonitrile (*B*) and the sample-loading solution consisted of 2% acetonitrile and 0.1% formic acid in water. All chemicals were Optima LC/MS grade (Thermo Scientific). The nano LC was coupled online with a TripleTOF 5600 system (AB Sciex). The MS scan was in the range *m*/*z* 350–1200 in high-resolution mode (>30 000) and the top 25 precursor ions were selected for subsequent MS/MS scans in high-sensitivity mode (>15 000). The data were processed using the *ProteinPilot* software 4.0 with the Paragon Algorithm 4.0.0.0 (AB Sciex). The software used only unique peptide sequences with greater than or equal to 95% confidence as evidence for protein identification. The data were searched against the UniProt database with the *BioWorks Browser* 3.3.1 SP1 and *SEQUEST* 2.0 software (Thermo Scientific). Only peptides identified with a confidence of ≥95% were taken into account.

2.3. Protein crystallization

The Sapp2p–pepstatin A complex was prepared by mixing the enzyme with a fivefold molar excess of pepstatin A (dissolved in dimethyl sulfoxide) and concentrated by ultrafiltration to 18 mg ml^{-1} using Amicon Ultra 0.5 ml 30K filters (Millipore). Screening for crystallization conditions was performed with the help of a Gryphon crystallization workstation (ArtRobbins) by the vapour-diffusion method in sitting-drop mode at 292 K in 96-well plates. The protein solution (0.2 µl) was mixed with 0.2 µl well solution and the mixture was equilibrated over 200 µl reservoir solution. The PEGs Suite (Qiagen) was used for the initial crystallization-condition screen. Initial microcrystals appeared in several days in various conditions containing 0.1 *M* MES pH 6.5 and 20–30% PEG 200–400 as precipitant.

Further optimization was performed manually and involved changing to the hanging-drop mode in 24-well crystallization plates (EasyXtal DG-Tool, Qiagen). Crystals were obtained by mixing 3 µl protein–pepstatin A complex solution with 1 µl reservoir solution composed of 0.1 *M* MES pH 6.5, 30% PEG

Table 1
Crystal data and diffraction data-collection and refinement statistics for the Sapp2p–pepstatin A complex.

For the data-collection statistics, the values in parentheses are for the highest resolution shell. For the refinement statistics, the values in parentheses are for reflections stronger then $4\sigma(F_o)$. The Friedel pairs were not merged in refinement.

Data-collection statistics	
Wavelength (Å)	0.91841
Space group	$P2_1$
Unit-cell parameters (Å, °)	a = 48.25, b = 57.58, c = 54.32, α = 90.0, β = 93.0, γ = 90.0
No. of molecules in asymmetric unit	1
Resolution range (Å)	12.820–0.825 (0.850–0.825)
No. of unique reflections	261700 (13705)
Multiplicity	8.1 (3.1)
Completeness (%)	92.3 (65.5)
R_{merge}†	0.087 (0.378)
Average $I/\sigma(I)$	13.2 (2.0)
Wilson B (Å^2)	6.2
Refinement statistics	
Resolution range (Å)	11.368–0.825
No. of reflections in working set	515542 (404510)
No. of reflections in test set	5199 (4050)
R_{work}‡ (%)	10.69 (9.32)
R_{free}§ (%)	13.15 (11.14)
R_{all} (%)	10.71 (9.33)
R.m.s.d., bond lengths (Å)	0.0217
R.m.s.d., bond angles¶ (Å)	0.0523
No. of non-H atoms in asymmetric unit	3044
No. of water molecules in asymmetric unit	393
Mean ADP (Å^2)	
Main chain	8.5
Side chain and water	14.0
Residues in alternative conformations	41
Ramachandran plot statistics	
Residues in favoured regions (%)	96.8
Residues in allowed regions (%)	2.5

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where the average intensity $\langle I(hkl)\rangle$ is taken over all symmetry-equivalent measurements and $I_i(hkl)$ is the intensity of the ith measurement of reflection hkl. ‡ $R = \sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively. § R_{free} is equivalent to R but is calculated for 5% of the reflections that were chosen at random and omitted from the refinement process. ¶ Accurate bond angle is defined as the optimal distance between two atoms that are both bonded to the same atom.

400 and equilibrating the drop over 0.5 ml reservoir at 292 K. Crystals appeared after 1 d in the form of crystal clusters and reached their full size of 300 × 250 × 150 µm within 1 d. For data collection, the crystal was divided into three parts using a scalpel, and these were individually cryocooled in liquid nitrogen without additional cryoprotection.

2.4. Data collection and structure determination

The diffraction data set for the Sapp2p–pepstatin A complex was collected at 100 K on the MX14.2 beamline at BESSY, Berlin, Germany (Mueller *et al.*, 2012). Data were integrated, reduced and scaled with *XDS* (Kabsch, 2010) using the *xdsgui* interface (Diederichs, 2010). To collect as complete high-resolution data as possible, we merged and scaled data sets from two parts of the original crystal using *XSCALE* (Kabsch, 2010). The second part of the crystal was mounted in a different orientation to the first, and the merged data reached 92.3% completeness (65.5% for the highest shell); the mosaicity estimated by the program was 0.3° for both parts of the crystal. The low completeness at high resolution was caused by the physical limitations of the beamline at the closest distance of the detector. Moreover, for resolutions of up to 0.825 Å it was necessary to use diffraction spots in the corners of the MAR Mosaic 225 detector, and part of the detector area was shadowed by the cryodevice and beamstop support. Crystal parameters and data-collection statistics are given in Table 1.

The phase problem was solved by molecular replacement using *MOLREP* (Vagin & Teplyakov, 2010). The search model was derived from the structure of an Sapp1p–pepstatin A complex (PDB entry 3fv3; Dostál *et al.*, 2009). *MOLREP* found one molecule in the asymmetric unit using diffraction data in the resolution range 25.46–3.8 Å. The resulting R factor was 46.1%.

The model was rebuilt in *Coot* (Emsley & Cowtan, 2004) into the map calculated using phases from the molecular-replacement solution. This initial model of the Sapp2p–pepstatin A complex was submitted to *REFMAC* for isotropic refinement. Further model refinement was carried out in *SHELXL*2013 (Gruene *et al.*, 2014) using isotropic and anisotropic refinement protocols. The default *SHELXL* restraints ISOR, SIMU and DELU were applied to the anisotropic atomic displacement parameters (ADPs). The H-atom positions were recalculated at every refinement cycle in idealized positions, and their isotropic ADPs were fixed at values 20% higher than the ADPs of their parent atoms. The occupancies of side chains adopting alternative conformations were refined with their sums constrained to unity. Finally, the occupancies of O atoms of the solvent water were also refined. If the occupancy parameter was refined to a value exceeding 0.95, it was fixed at an occupancy value of 1. Cycles of refinement were interspersed with visual inspection sessions using *Coot*, and if necessary the model was corrected manually, for example, by introducing alternative conformations of several side chains. After applying the conjugate-gradient least-squares (CGLS) minimization method, the last round of refinement was performed using the full-matrix least-squares option, with the parameter shifts damped to zero, to obtain reliable estimations of all refined and derived parameters of the model. The Friedel pairs were not merged for the *SHELXL* refinement; the command MERG 2 was used. The final model and the corresponding structure factors have been deposited in the PDB with identification code 4y9w.

2.5. Molecular modelling

The crystal structure of pepstatin A in complex with Sapp2p (resolution of 0.83 Å, PDB entry 4y9w; this work) was compared with that of pepstatin A in complex with Sapp1p (resolution of 1.85 Å, PDB entry 3fv3; Dostál *et al.*, 2009). For the latter complex, two conformations of pepstatin A were observed in the eight molecules present in the asymmetric unit (conformation I was found in chains *A*, *B*, *C*, *D* and *F*, and conformation II was found in chains *E*, *G* and *H*). Both conformations were considered in this study. For further computations, all water molecules and ions were omitted. Protonation of histidines was assigned based on visual

inspection of their surroundings (all His resides were monoprotonated on N^{ε}). The protein N-terminus and the side chains of lysines and arginines were positively charged, while the C-terminus and the side chains of glutamates and aspartates (with the exception of the catalytic dyad) were negatively charged to reflect the predominant state at the experimental pH of 6.5–7.0. The active site was treated according to the crystallographic findings from the Sapp2p–pepstatin A complex, *i.e.* Asp211 was monoprotonated on the $O^{\delta 2}$ atom and Asp32 was either deprotonated or monoprotonated on the $O^{\delta 1}$ atom. The inhibitors were protonated using *UCSF Chimera* (Pettersen *et al.*, 2004). The positions of the added H atoms were relaxed *in vacuo* using the *FIRE* algorithm (Bitzek *et al.*, 2006) followed by molecular dynamics-based simulated annealing (3 ps from 1700 to 0 K) using the Berendsen thermostat (Berendsen *et al.*, 1984) in the *SANDER* module of the *AMBER* 10 package (Case *et al.*, 2008). Similarly, amino-acid residues that were not well defined in the electron-density maps (Ala208, Asn252, Pro253, Thr279 and Asn281) were relaxed by annealing (3 ps from 300 to 0 K) using the Berendsen thermostat. Atomic charges for the inhibitors were obtained by the *RESP* procedure (Bayly *et al.*, 1993) at the HF/6-31G* level. The protein parameters were obtained from the ff03 force field (Duan *et al.*, 2003), while GAFF parameters were used for the ligands (Wang *et al.*, 2004).

2.6. Quantum-mechanical calculations

2.6.1. Setup. All four model complexes, Sapp2p–pepstatin A with Asp32 deprotonated or monoprotonated (see above) and Sapp1p–pepstatin A in conformations I and II, were optimized using the following setup. Only surroundings of the ligands within 8 Å (approximately 1560 atoms in total) were taken into account. Of these atoms, only atoms belonging to amino acids within 6 Å of the ligand (approximately 1200 atoms) were allowed to move. The energies and gradients were obtained by the semi-empirical quantum-mechanical (SQM) method PM6-D3H4 coupled with the COSMO implicit solvent model using the linear scaling method *MOZYME* in *MOPAC* (Lepšík *et al.*, 2013). The SQM optimizations were performed in several rounds until the energy and gradient convergence criteria (ΔE = 0.005 kcal mol^{-1}, maximum gradient of 1 kcal mol^{-1} Å^{-1}, root mean square of the gradient of 0.5 kcal mol^{-1} Å^{-1}) were met. Interaction 'free' energies ($\Delta G'_{\rm int}$) of all of the studied systems were determined on the whole optimized structures using the PM6-D3H4 method and the COSMO solvent model.

2.6.2. Interaction energies. The differential contribution of the amino acids in the active sites of Sapp1p and Sapp2p to pepstatin A binding was examined by 'virtual glycine scanning' (Pecina *et al.*, 2013), *i.e.* the interacting amino acids in the active site were substituted by glycine. The energy contributions of their side chains ($\Delta\Delta G'_{\rm int}$) were calculated as the difference between the original $\Delta G'_{\rm int}$ with the wild-type amino acid and the new $\Delta G'_{\rm int}$ with the mutated glycine residue. The $\Delta G'_{\rm int}$ values were obtained on the whole optimized structures as single-point energies at the PM6-D3H4/COSMO level (Lepšík *et al.*, 2013).

3. Results and discussion

3.1. Purification of Sapp2p

Sapp2p was purified from *C. parapsilosis* cultivation medium, where it was present along with large amounts of Sapp1p, as described previously (Fusek *et al.*, 1993; Merkerová *et al.*, 2006; Hrušková-Heidingsfeldová *et al.* (2009). Two *SAPP2* homologues occur in the *C. parapsilosis* genome. They share 91.5% identity and differ mainly in their C-terminal sequence (Dostál *et al.*, 2015). We purified Sapp2p/CPAR2_102580 (the standard and systematic name according to http://www.candidagenome.org), which is the shorter of the two Sapp2p molecular species, consisting of 395 amino acids.

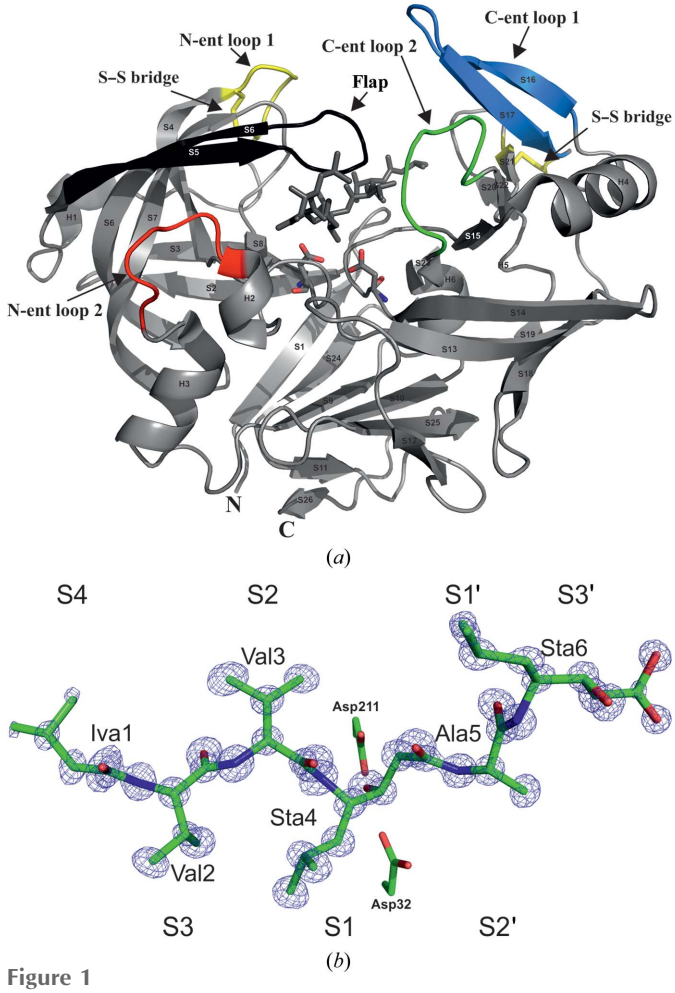


Figure 1
Overall structure of Sapp2p in complex with pepstatin A. (*a*) Overall three-dimensional structure and secondary-structural elements of Sapp2p in complex with pepstatin A. The protein is shown in ribbon representation; pepstatin A and the catalytic aspartates are shown in stick representation. The flap and entrance loops covering the active site are coloured and labelled. (*b*) Structure of pepstatin A bound to Sapp2p. The $2F_{\rm o} - F_{\rm c}$ electron-density map is contoured at 1.5σ. Residue names (Iva1-Val2-Val3-Sta4-Ala5-Sta6) and corresponding substrate-binding subsites (S4–S3′) are indicated. Catalytic aspartates are also shown as sticks.

Protein identity was confirmed by mass spectrometry. We did not detect any Sapp2p/P32950, and thus the question of the natural occurrence and role of this variant remains open.

3.2. Overall structure description and quality

The crystal structure of Sapp2p (Fig. 1) was solved by molecular replacement using the structure of Sapp1p (PDB entry 3fv3; Dostál *et al.*, 2009) as a search model and was refined to a resolution of 0.825 Å (Table 1). The crystal structure of Sapp2p in complex with pepstatin A belonged to space group $P2_1$, with a solvent content of 44%. The asymmetric unit contained one molecule of Sapp2p. All 395 residues could be modelled into the electron-density map, with the exception of the side chains of several surface-exposed residues (Ser33, Asn252, Pro253, Thr279, Ala208 and Asn281).

The Sapp2p structure comprises two topologically similar N- and C-terminal domains with a large substrate-binding cleft located between them (Fig. 1). The structure is stabilized by two disulfide bridges (Cys47–Cys52 and Cys249–Cys283). The conserved sequence DT(S)G, which is present as one copy in each domain and contains the catalytic aspartate residues (Asp32 and Asp211), is the signature motif of aspartic proteases (Rao *et al.*, 1991). Similar to other aspartic proteases, the Sapp2p active site is covered by an antiparallel β-sheet (residues 71–89), commonly known as the active-site flap, which plays an important role in substrate binding. Because the substrate-binding site in our structure is occupied by the substrate-mimicking inhibitor pepstatin A, the flap adopts a closed conformation. The substrate-binding site is lined by four entrance loops. Two N-terminal entrance loops, N-ent loop 1 (Cys47–Cys52) and N-ent loop 2 (Glu124–Asp132), flank the flap. Two C-terminal entrance loops, C-ent loop 1 (Ser289–Pro297) and C-ent loop 2 (Ala233–Ile247), are located across the binding cleft, facing the flap and the N-ent loop, respectively.

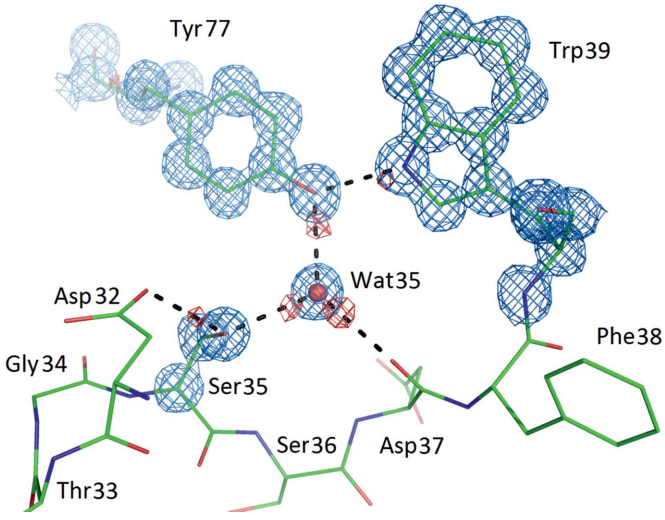


Figure 2
The network of hydrogen bonds stabilizing the closed flap. Detail of the hydrogen-bonding network between the flap and the catalytic site. The Tyr77 side chain and the stretch of amino-acid residues 32–39 are shown in stick representation; hydrogen bonds are shown as black dashed lines. The $2F_o - F_c$ electron-density map contoured at the 1.5σ level is shown in light blue and the $F_o - F_c$ difference electron-density map contoured at the 1.8σ level is shown in red. Both maps are calculated from refinement cycles prior to adding H atoms to the model.

The electron-density map for the active-site-bound ligand was of excellent quality, allowing us to model pepstatin A with full occupancy (Fig. 1*b*).

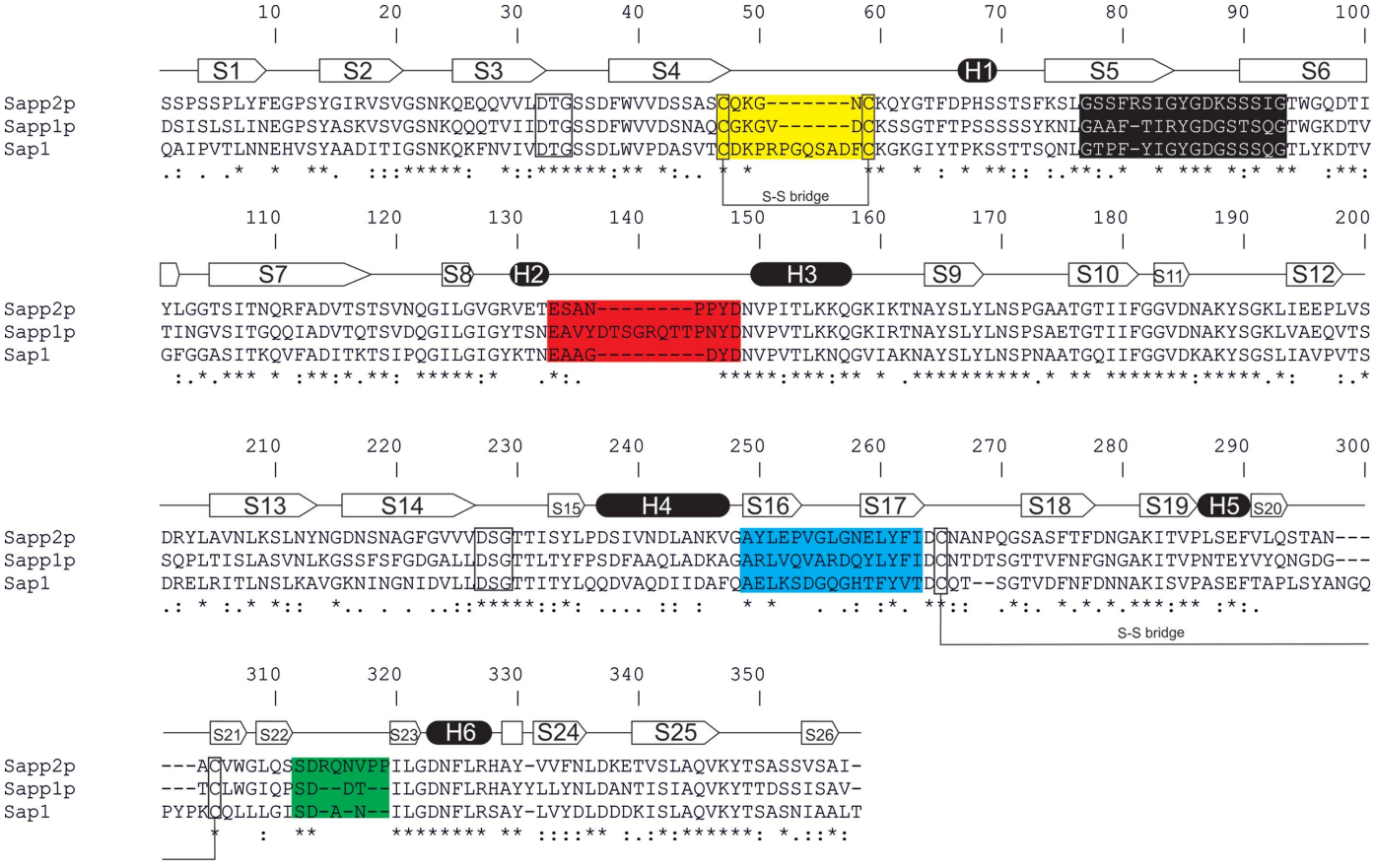
The atomic resolution achieved for the Sapp2 structure allowed the localization of numerous H atoms in the difference density maps and thus uncovered hydrogen-bonding networks. An example depicted in Fig. 2 shows part of the stabilizing hydrogen-bonding network of the flap closing over the active site and the hydrogen-bonding network leading to the catalytic residue. The side chain of Tyr77, a residue located next to the tip of the flap (Asp79), interacts with Trp39 through its O^{η} atom *via* a hydrogen bond to $H^{\varepsilon 1}$ of Trp39. The H^{η} atom of Tyr77 interacts with the O atom of water molecule 35 (Wat35). Wat35 donates one of its H atoms to a hydrogen bond to the main-chain carbonyl O atom of Asp37 and the second hydrogen to a hydrogen bond to the side-chain hydroxyl O atom (O^{γ}) of Ser35. H^{γ} of Ser35 donates a hydrogen bond to the $O^{\delta 2}$ atom of the catalytic residue Asp32. This hydrogen bond is critical for positioning the carboxyl group of Asp32 in a plane with the carboxyl group of the second catalytic residue Asp211 (see Fig. 6*a*). The positions of these H atoms are indicated by the presence of positive electron density (Fig. 2). Moreover, these densities are perfectly located between the hydrogen-bond acceptor and donor atoms.

3.3. Comparison of Sapp2p with related structures

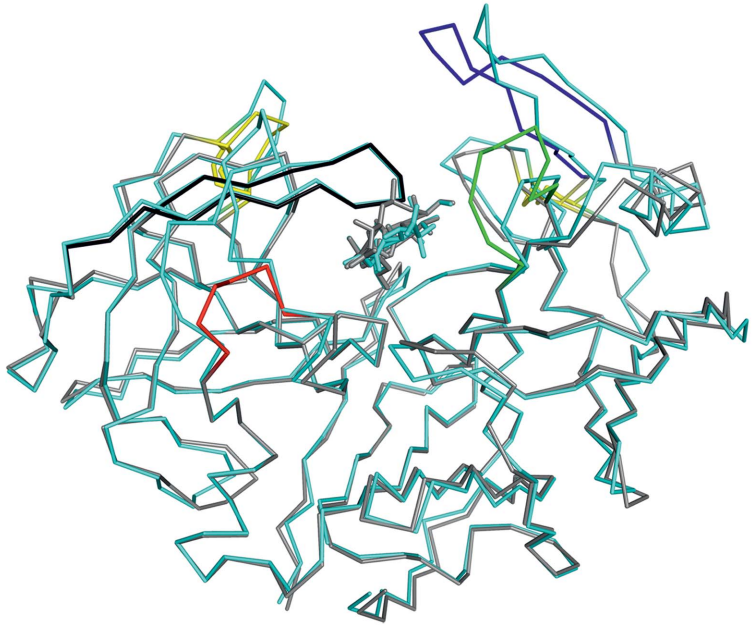
The overall fold and topology of Sapp2p is similar to those of the other Sap family enzymes. The closest sequence and structural homologue of Sapp2p is Sapp1p. The sequence homology of these two isoenzymes is over 80% and their structures superpose with a root-mean-square deviation (r.m.s.d.) of 1.25 Å for 330 aligned C^{α} atoms (Fig. 3). Among the Saps from *C. albicans*, the most similar to Sapp2p in sequence and structure is Sap1 (sequence homology of 48%, r.m.s.d. of 2.64 Å for superposition of 331 aligned C^{α} atoms).

Both Sapp1p and Sapp2p contain two pairs of cysteine residues, and the S–S bridge topology is similar in both proteins (Cys47–Cys53 and Cys258–Cys292 in Sapp1p; Cys47–Cys52 and Cys249–Cys283 in Sapp2p). In addition, both the Sapp1p and Sapp2p isoenzymes contain one serine residue (Ser193 in Sapp1p and Ser184 in Sapp2p) encoded by the ambiguous CUG codon. These serines occur within loops (Leu182–Leu188 in Sapp2p; Val191–Thr198 in Sapp1p) that are topologically similar in both isoenzymes, although their sequence homology is quite low.

Despite the high structural similarity, there are noticeable differences between Sapp2p and Sapp1p. The major differences in backbone superposition between Sapp1p and Sapp2p are located in the loops that line the entrance to the substrate-binding cleft. There are two major differences in the entrance



(*a*)



(*b*)

Figure 3
Comparison of Sapp2p with the homologous proteins Sapp1p (*C. parapsilosis*) and Sap1 (*C. albicans*). (*a*) Sequence alignment of Sapp2p, Sapp1p and Sap1. Secondary-structure elements (H, helices; S, strands) found in Sapp2p are shown. Identity is indicated by an asterisk (*), strong similarity is indicated by a colon (:) and weak similarity is indicated by a point (.). The residues of the active site are framed and the S–S bridges are indicated. The entrance-loop sequences are coloured as follows: N-ent loop 1, yellow; N-ent loop 2, red; C-ent loop 1, blue; C-ent loop 2, green. The flap region (Gly78–Gly104) is coloured black. (*b*) Superposition of the crystal structures of Sapp1p and Sapp2p in complex with pepstatin A. Sapp1p in complex with pepstatin A (PDB entry 3fv3; Dostál *et al.*, 2009) is coloured cyan. Sapp2p in complex with pepstatin A (PDB entry 4y9w; this work) is coloured grey with highlighted entrance loops and flap; the colour coding corresponds to that in (*a*). Pepstatin A is shown in stick representation and the S–S bridges are indicated as yellow sticks.

loops: a deletion in the Sapp2p sequence in N-ent loop 2 and an insertion in C-ent loop 2 of Sapp2p (Figs. 1 and 3). These loops are in direct contact with the C-terminal residue of pepstatin A, and their conformations significantly affect the character, shape and size of the substrate-binding cleft (Fig. 4).

Compared with Sapp1p, Sapp2p contains a deletion of eight amino acids in N-ent loop 2, resulting in the substrate-binding cleft being more open than that of Sapp1p. On the other hand, an amino-acid insertion in C-ent loop 1 of Sapp2p results in closure of the binding cavity, with a tighter embrace of the central part of the substrate/inhibitor (Fig. 4). Tight closing of the substrate-binding cleft of Sapp2p is mediated by interaction between the C-ent loops (residues Gln293, Val295 and Asn242) and the flap (residues Gly78 and Lys80).

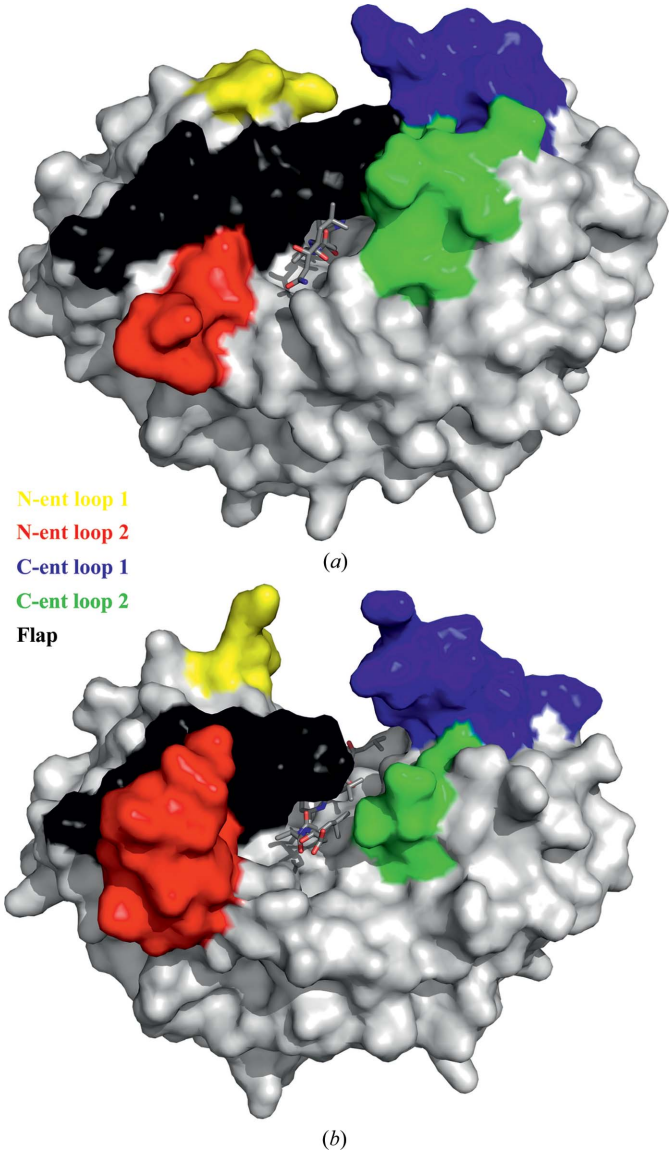


Figure 4
Overall structures of Sapp2p (*a*) and Sapp1p (*b*) represented by solvent-accessible surfaces. N-ent loop 1 and N-ent loop 2 are coloured yellow and red, respectively. C-ent loop 1 and C-ent loop 2 are coloured blue and green, respectively. The flap is coloured black and pepstatin A is shown in stick representation.

3.4. Pepstatin A binding to Sapp2p substrate-binding pockets

The active site, which is located between the two domains of the molecule at the bottom of a large cleft, is one of the most highly conserved regions in the Sap family. Pepstatin A, a peptide-like inhibitor containing six amino-acid residues in positions P4–P3′ (Iva1-Val2-Val3-Sta4-Ala5-StaOH6), bound to Sapp2p in an extended conformation, occupying the S4–S3 substrate-binding pockets of the active site of the enzyme. The pepstatin A conformation in Sapp2p is very similar to that observed in the previously reported structure of Sapp1p (Fig. 5*a*), with the exception of the Sta6 residue in position P3′. In the Sapp1p crystal structure, two alternative conformations of the Sta6 residue were observed (denoted I and II).

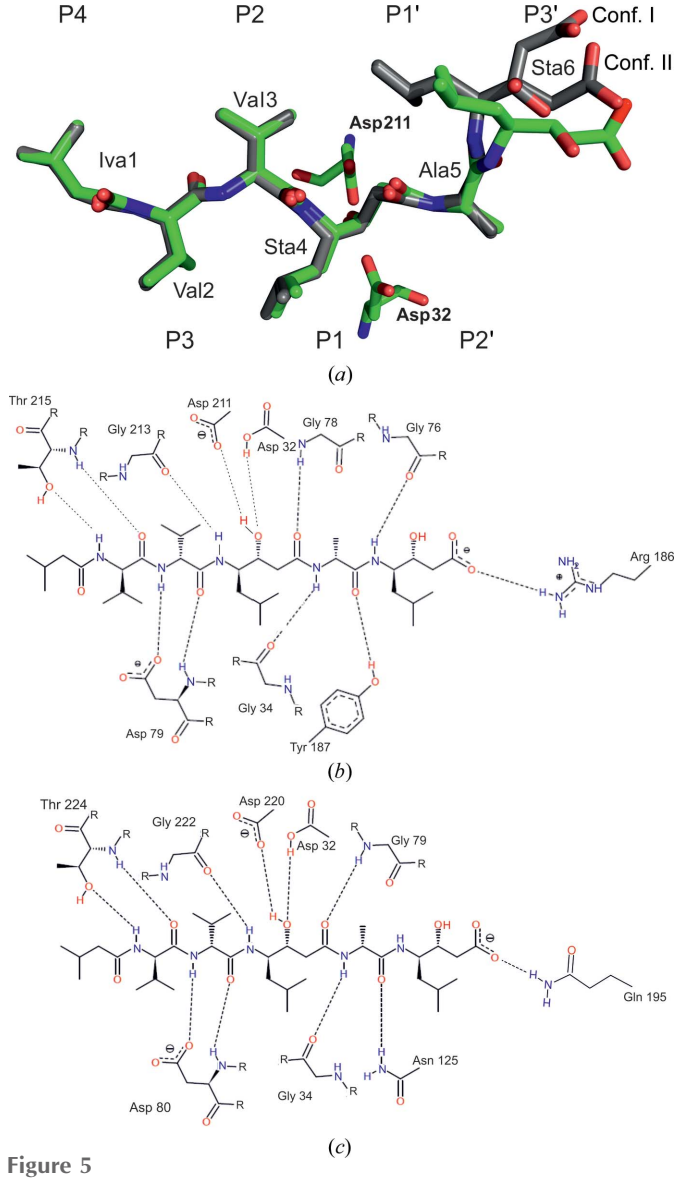


Figure 5
Pepstatin A binding to Sapp2p and Sapp1p. (*a*) Superposition of pepstatin A bound to Sapp2p (green C atoms) and Sapp1p in conformations I and II (grey C atoms). Two catalytic aspartates of Sapp2p are also depicted. O and N atoms are coloured red and blue, respectively. Schematic representations of hydrogen-bonding interactions of pepstatin A with the protein atoms of Sapp2p (*b*) and Sapp1p (*c*) are shown.

Table 2
Residues in contact (<4.2 Å) with pepstatin A in the Sapp1p and Sapp2p structures.

Residues that form hydrogen bonds to pepstatin A are shown in bold.

Subsite	Sapp2p	Sapp1p
S4	**Thr215**	**Thr224**
	Ile216	Leu225
	Tyr218	Tyr227
	Leu276	Tyr285
	Val284	Leu293
S3	Pro12	Pro12
	Ser13	
	Asp79	**Asp80**
	Gly213	**Gly222**
	Thr215	Thr224
S2	Tyr77	Tyr78
	Gly78	**Gly79**
	Asp79	**Asp80**
	Thr214	Thr223
	Tyr218	Tyr227
	Val295	
	Ile298	Ile303
S1	Val30	Ile30
	Asp32	**Asp32**
	Gly34	Gly34
	Tyr77	Tyr78
	Gly78	Gly79
	Asp79	Asp80
	Ser81	Ser82
	Ile116	Ile117
	Asp211	
	Gly213	Gly222
	Thr214	Ile303
S1′	**Gly34**	**Gly34**
	Gly76	
	Tyr187	**Asn125**
		Leu218
		Asp220
	Ile298	Ile303
S2′	Ser35	Ser35
	Ile75	Ile76
	Gly76	Gly79†
	Gly78	
	Arg186	Leu218†
	Tyr187	**Asn125**
S3′	Arg186	
		Gly79‡
		Gln195‡
		Asp301‡

† Residues interacting with pepstatin A conformation I. ‡ Residues interacting with pepstatin A conformation II.

The r.m.s.d.s for superposition of pepstatin A atoms bound to Sapp2p and Sapp1p are 1.06 and 1.47 Å for conformations I and II, respectively. When the first five residues of pepstatin A are compared, the r.m.s.d.s are 0.46 and 0.49 Å for conformations I and II, respectively. The structurally different binding of pepstatin A in Sapp2p compared with Sapp1p (Fig. 5*a*) is the result of three changes in the hydrogen bonding of the P2′ and P3′ inhibitor moieties (Fig. 5).

The side chain of the P4 residue Iva1 points towards the opening of the active site. The S3 and S2 subsites are occupied by Val2 and Val3, respectively. The side chain of Sta4 in the P1 position is closely packed against the side chain of Val2 in the P3 position. The Sta4 hydroxyl group is engaged in hydrogen-bonding interactions with the catalytic aspartates Asp32 and Asp211. Interestingly, the S1′ subsite is not occupied by the residue immediately following the P1 Sta4 but by a backwards-turned Sta6 side chain owing to a shift of register caused by the longer backbone of the statin moiety. The P2′ subsite is occupied by Ala5, which follows the P1 Sta4 in the sequence (Fig. 5*a*). The inverse γ-turn involving both Sta4 and Sta6 changes the direction of the inhibitor chain, leading the carboxylate of Sta6 towards the protein surface and occupying the S3′ subsite. As a result, the backbones of the P2′ Ala5 and the P3′ Sta6 residues deviate from the regular extended conformation. The different conformations of the P3′ Sta6 when bound to Sapp2p and Sapp1p, respectively, are caused by differences in the structure of the entrance to the active site, namely N-ent loop 2 and C-ent loop 1 (Fig. 4).

3.5. Polar and van der Waals interactions of pepstatin A with Sapp2p and Sapp1p

Direct hydrogen bonds to Sap isoenzymes are only supplied by the pepstatin A backbone (Figs. 5*b* and 5*c*). There are 12/11 direct hydrogen bonds between pepstatin A and Sapp1p/Sapp1p, respectively. Furthermore, the polar atoms of the P4 and P3′ residues are involved in water networks that help to hold the inhibitor in the enzyme cavity. Pepstatin A forms analogous hydrogen bonds to the Sapp2p and Sapp1p isoenzymes, with the following exceptions: (i) the carbonyl O atom of the P2′ Ala accepts a hydrogen from the phenolic hydroxyl of Tyr187 of Sapp2p but from the amide NH2 group of Asn125 in Sapp1p, (ii) the hydrogen donated by the NH group of Sta6 to Gly76 in Sapp2p is lost in Sapp1p owing to the presence of the bulky Arg77 side chain and (iii) the terminal carboxylate of Sta6 accepts a hydrogen bond from Arg186 of Sapp2p but from Gln195 in conformation II of Sapp1p (no hydrogen bond is formed in conformation I) (Figs. 5*b* and 5*c*).

The first difference (Tyr187/Asn125 in the S2′ pocket) creates an energy difference of up to 9 kcal mol^{-1} (see Fig. 7) that will partially be offset by the second difference (the presence and absence of Gly76). The third difference (Arg186/Gln195 in the S3′ pocket) favours Sapp2p binding by approximately 1 kcal mol^{-1}. An additional hydrogen bond (pepstatin A to Asp79 in Sapp2p or Asp80 in Sapp1p) appears to be structurally conserved but presents an energy difference of 3–4 kcal mol^{-1} in favour of Sapp1p. Additional calculations showed that this can be ascribed to a markedly more positive charge in the surroundings (Lys49 and Lys80) in Sapp2p. Furthermore, differences in the stabilization of the surrounding water cluster may come into play. Although the interactions of pepstatin A with Sapp2p and Sapp1p are mostly similar, we noted several mutually compensating differences, in line with the similar values of the measured inhibition constants of 0.4 and 0.3 n*M*, respectively, for the binding of pepstatin A to Sapp2p and Sapp1p.

In addition to polar interactions, pepstatin A makes numerous van der Waals interactions with the residues listed in Table 2. We carried out computational analysis of pepstatin A binding to the Sapp isoenzymes to assess the energy contributions of individual residue side chains (see §3.7).

3.6. Catalytic site protonation

The high resolution of our Sapp2p–pepstatin A structure and the low e.s.d. of the bond lengths (on the subpicometre scale) allowed us to suggest the protonation states of the catalytic aspartates (Fig. 6*a*). By comparing the distances between C^{γ} and $O^{\delta1/2}$ of catalytic aspartates with the optimal distances for a C—O single bond (1.3 Å) and a C═O double bond (1.2 Å), we identified which O^{δ} atoms are protonated (Wlodawer *et al.*, 2001). Moreover, the position of the Sta4 hydroxyl H atom, which mimics the transition state, is clearly visible in the difference electron-density map contoured at the 2σ level (Fig. 6*b*). The C—O interatomic distances in Asp211 indicate protonation of the $O^{\delta2}$ atom. The situation is slightly more complicated for Asp32 because the C—O distances in the carboxylic moiety are very similar (1.24 and 1.27 Å). The crystal structure is likely to reflect a superposition of two states: one in which both O^{δ} atoms are deprotonated (the optimal distance for a delocalized C—O partial double bond is 1.249 Å) and a minor form in which $O^{\delta1}$ is protonated. The occupancies of these two states were estimated to be 60 and 40%, respectively, based on the C—$O^{\delta1}$ distance.

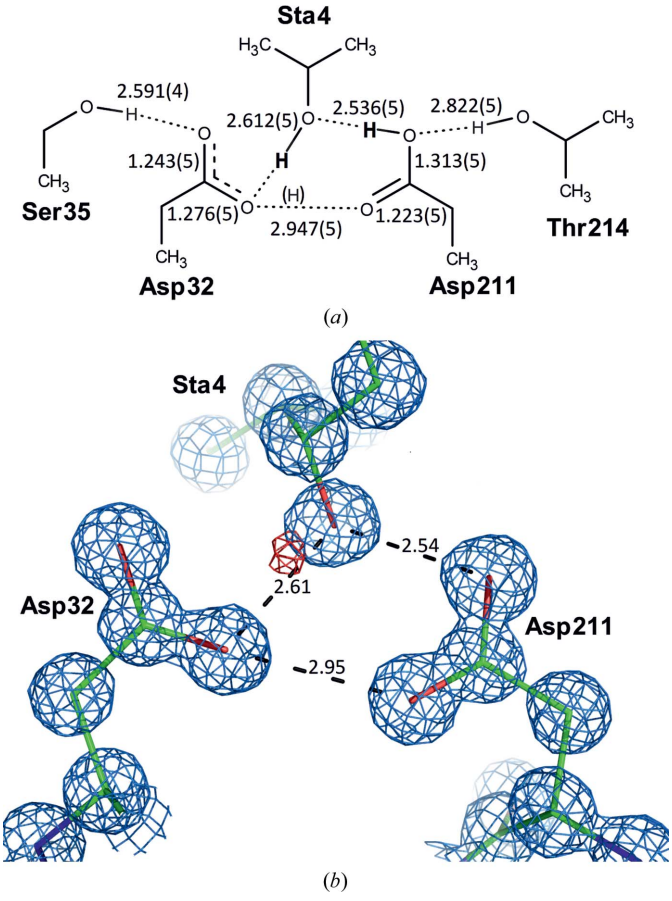


Figure 6
Polar interactions in the active site of the atomic resolution Sapp2p structure. (*a*) Schematic diagram with bonding distances (values are in Å; estimated standard deviations of distances are in parentheses). H atoms in bold are clearly assigned (the hydroxyl H atom of pepstatin was visible in difference electron density; the H atom on the protonated carboxyl of Asp211 was deciphered from C—O bonding distances). The H atom in parentheses is hypothetical. (*b*) Detailed structure of the active site in stick representation; hydrogen bonds are shown as dotted lines (numbers represent distances in Å). The $2F_o - F_c$ electron-density map contoured at the 1.5σ level is shown in light blue; the $F_o - F_c$ difference electron-density map contoured at the 2σ level is shown in red.

Protonation of Asp32 $O^{\delta1}$ can be achieved *via* the presence of the proton shared between the $O^{\delta1}$ atoms of the two catalytic aspartates or by a transient shift of the Sta4 hydroxyl H atom towards Asp32 $O^{\delta1}$. Proton sharing between the $O^{\delta1}$ atoms of two catalytic aspartates has been observed for a related aspartic protease, HIV-1 protease, in complex with a norstatine-based inhibitor (Brynda *et al.*, 2004). However, we found that a shared proton cannot be accommodated in the Sapp2p active site for steric reasons (the position of this hypothetical proton is shown in parentheses in Fig. 6*a*). Moreover, the shared-proton arrangement has far less favourable interaction energy with the statin inhibitor than the model with only two protons in the active site. We therefore suggest that the protonated state of Asp32 may indicate a transient shift of the statin hydroxyl proton towards $O^{\delta1}$ of Asp32.

The arrangement of the Sapp2p active site in our structure is very similar to the structures of other aspartic proteases in complex with an inhibitor with a hydroxyl group located between the active-site aspartates, such as the structures of endothiapepsin from the fungus *Endothia parasitica* (Coates *et al.*, 2002) and HIV-1 protease (Adachi *et al.*, 2009). In all of these atomic resolution X-ray structures, a proton is shared between the hydroxyl O atom of the inhibitor and the $O^{\delta1}$ atom of the catalytic aspartate.

3.7. Comparison of pepstatin A interactions with Sapp2p and Sapp1p

The residues involved in pepstatin A binding in Sapp2p and Sapp1p and the hydrogen-bonding pattern for the central binding pocket (S3–S1′) are mostly conserved (Figs. 5*b* and 5*c*). Most of the amino acids (75%) that form the Sapp1p and Sapp2p substrate-binding sites are conserved. Only residues involved in pepstatin A binding in the S2′ and S3′ subsites are significantly different between Sapp1p and Sapp2p. We used a virtual glycine scan (Pecina *et al.*, 2013) to study the roles of individual amino-acid side chains in the Sapp2p and Sapp1p active sites in the binding of pepstatin A. We used a fast and reliable semi-empirical quantum-mechanical (QM) method, PM6-D3H4X (Lepšík *et al.*, 2013). We needed a QM approach to quantitatively describe the strength of noncovalent interactions (Riley *et al.*, 2010), including quantum effects such as proton transfer. At the same time, using the semi-empirical approximation, we were able to include over 1000 atoms in the QM part and thus capture the long-range effects, such as electrostatic interactions.

The changes in the free energy of interaction ($\Delta\Delta G'_{\rm int}$) upon the mutation of a given amino-acid residue to glycine are shown in Fig. 7.

The amino-acid residues in the individual substrate-binding subsites of Sapp2p and Sapp1p fell into one of four categories: (i) identical residues, (ii) similar residues, (iii) different

residues and (iv) residues that do not form corresponding pairs owing to different tracing of the protein backbone. A fourth category comprises residues that do not form corresponding pairs owing to different tracing of the protein backbone. It is worth mentioning that the energy contributions inherently contain the effect of hydrogen bonding mediated by the residue side chain. The contributions of glycine residues (Gly213/Gly222 in S1, Gly78/Gly79 in S1/S1′ and Gly34/Gly34 in S1′) cannot be evaluated by the virtual glycine-scanning method. However, the contributions of these residues are likely to be very similar in Sapp2p and Sapp1p because they have similar conformations in both isoforms. The only exception in this category is the Gly76···Sta6 hydrogen bond, which is present in Sapp2p and absent in Sapp1p.

In the S4 subsite, we identified the following Sapp2p/Sapp1p residue pairs in the three categories defined above: (i) Pro12/Pro12 (also contributing to S3), Thr215/Thr224 (also contributing to S3) and Tyr218/Tyr227 (also contributing to S2); (ii) Ile216/Leu225 and Val284/Leu293; and (iii) Leu276/Tyr285. All of these pairs feature similar energy contributions within a difference of approximately 1 kcal mol^{-1}, even in category (iii) (Fig. 7). The strongest contribution to binding, of around 7 kcal mol^{-1}, is mediated by Thr215/Thr224. This is achieved by a combination of aliphatic···aliphatic interactions in the S4 subsite and hydrogen bonding in the S3 subsite.

The S3 subsite is relatively solvent-exposed. The only interacting residue in Sapp2p and Sapp1p is Ser13, which makes an almost identical contribution to the interaction energy in both enzymes (of close to 2 kcal mol^{-1}).

The S2 subsite is formed by several Sapp2p/Sapp1p conserved residues in category (i), Thr214/Thr223, Tyr218/Tyr227, Ile298/Ile303 (S2/S1′) and Asp79/Asp80 (S2/S1), and by Val296 of Sapp2p, which has no counterpart in Sapp1p. The conserved residues have similar contributions in Sapp2p and Sapp1p. Thr214/Thr223 has very strong interactions (around 7 kcal mol^{-1}) owing to a combination of aliphatic···aliphatic dispersion interactions in the S2 pocket and hydrogen bonding in the S1 pocket. Ile298/Ile303 has very weak (0.5 kcal mol^{-1}) methyl···methyl dispersion interactions (Jurecka *et al.*, 2006) in S2 that are identical for both enzyme isoforms. The contribution of this residue is, however, different in S1′. Interestingly, the conserved Asp79/Asp80 residue with an identical interaction pattern (two hydrogen bonds and a van der Waals interaction) has a difference of 3–4 kcal mol^{-1} in the energy of binding of pepstatin A to Sapp2p and Sapp1p, respectively. This difference can be ascribed to the long-range electrostatic influence of Lys49 and Lys80, which are present only in Sapp2p and are located approximately 7 Å from the charged Asp79 side chain. Additional interaction in the Sapp2p S2 subsite is mediated by Val296, which is located in the C-ent loop insertion and has no counterpart in Sapp1p.

The S1 subsite features the following Sapp2p/Sapp1p residues: (i) Ile116/Ile117, Tyr77/Tyr78 and Ser81/Ser82, and (ii) Val30/Ile30. All of these interacting residues provide similar contributions in Sapp2p and Sapp1p. Tyr77/Tyr78 provides a very large contribution of almost 9 kcal mol^{-1} owing to main-chain/main-chain hydrogen bonding combined with CH···π interactions.

The P1′ side chain of the Sta6 residue of pepstatin A features one nonpolar interaction of aliphatic···aliphatic type with the Pro296 side chain of Sapp2p. As this residue is located in the C-ent loop insert, Sapp1p does not have a counterpart. However, owing to the different pose of the P1′ moiety in Sapp1p there is a favourable interaction with Leu218. The Ile298/Ile303 pair only has interactions in Sapp1p.

In the S2′ subsite, there are the following Sapp2p/Sapp1p interacting residues: (i) Ile75/Ile76 and Ser35/Ser35 and (iii) Tyr187/Asn125. While the former two interact with pepstatin A only weakly, Tyr187 makes the strongest contribution to the energy of binding among all of the calculated interactions. The reason 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 Cζ···Oη bond length of 1.337 (9) Å suggests that the proton is shared between the two O atoms. In Sapp1p, however, interaction with the carbonyl of the Ala in P2′ is mediated by a medium-strong hydrogen bond to Asn125.

The P3′ terminal carboxylate of pepstatin A Sta6 is exposed to the solvent. In addition to hydrogen bonds to water molecules, it forms a salt bridge with Arg186 in Sapp2p, which is functionally

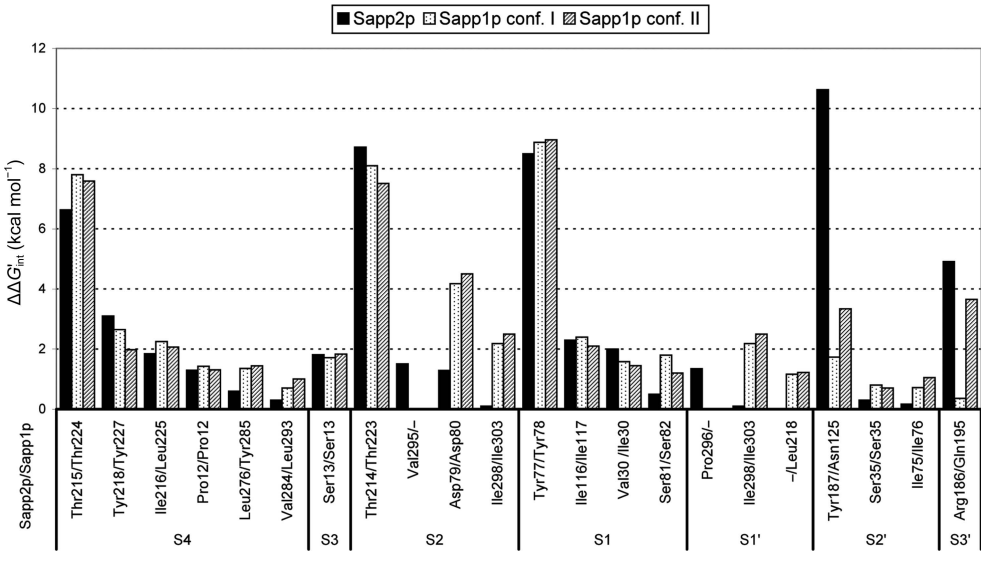


Figure 7
Energy contributions (kcal mol^{-1}) of the amino acids in Sapp2p and Sapp1p (two conformations of pepstatin A, I and II, were evaluated; *cf.* Fig. 5*a*) from a virtual glycine scan.

replaced in Sapp1p in conformation II by a charge-assisted hydrogen bond to Gln195.

Overall, our quantification of pepstatin A interactions with the Sapp2p and Sapp1p isoenzymes yielded similar results ($\sim$60 kcal mol^{-1} for Sapp2p and $\sim$55 kcal mol^{-1} for Sapp1p). This is the result either of similar interaction strengths or some weaker and other compensating stronger interactions. Our finding is in line with the similar values of the measured inhibition constants of 0.4 and 0.3 n*M*, respectively, for the binding of pepstatin A to Sapp2p and Sapp1p.

In summary, we have analyzed the binding of pepstatin A to the Sapp2p and Sapp1p isoenzymes by a combination of high-resolution crystallography and advanced quantum-chemical interaction energy calculations.

Acknowledgements

This work was supported by grant GA14-23022S from the Czech Science Foundation and by projects RVO 61388963 and RVO 68378050 awarded by the Academy of Sciences of the Czech Republic. The research leading to these results has received funding from the European Community's Seventh Framework Programme under BioStruct-X (grant agreement No. 283570). The authors would like to thank the IOCB MS facility for performing the peptide mass-fingerprinting analysis and Elena Dolejší for technical assistance.

References

Abad-Zapatero, C., Goldman, R., Muchmore, S. W., Hutchins, C., Stewart, K., Navaza, J., Payne, C. D. & Ray, T. L. (1996). *Protein Sci.* **5**, 640–652.
Adachi, M. *et al.* (2009). *Proc. Natl Acad. Sci. USA*, **106**, 4641–4646.
Bayly, C. I., Cieplak, P., Cornell, W. D. & Kollman, P. A. (1993). *J. Phys. Chem.* **97**, 10269–10280.
Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. (1984). *J. Chem. Phys.* **81**, 3684–3690.
Bitzek, E., Koskinen, P., Gähler, F., Moseler, M. & Gumbsch, P. (2006). *Phys. Rev. Lett.* **97**, 170201.
Borelli, C., Ruge, E., Lee, J. H., Schaller, M., Vogelsang, A., Monod, M., Korting, H. C., Huber, R. & Maskos, K. (2008). *Proteins*, **72**, 1308–1319.
Borelli, C., Ruge, E., Schaller, M., Monod, M., Korting, H. C., Huber, R. & Maskos, K. (2007). *Proteins*, **68**, 738–748.
Brynda, J., Rezacova, P., Fabry, M., Horejsi, M., Stouracova, R., Sedlacek, J., Soucek, M., Hradilek, M., Lepsik, M. & Konvalinka, J. (2004). *J. Med. Chem.* **47**, 2030–2036.
Case, D. A. *et al.* (2008). *AMBER* 10. University of California, San Francisco.
Coates, L., Erskine, P. T., Crump, M. P., Wood, S. P. & Cooper, J. B. (2002). *J. Mol. Biol.* **318**, 1405–1415.
Cutfield, S. M., Dodson, E. J., Anderson, B. F., Moody, P. C., Marshall, C. J., Sullivan, P. A. & Cutfield, J. F. (1995). *Structure*, **3**, 1261–1271.
Diederichs, K. (2010). *Acta Cryst.* D**66**, 733–740.
Dostál, J., Brynda, J., Hrušková-Heidingsfeldová, O., Pachl, P., Pichová, I. & Řezáčová, P. (2012). *J. Enzyme Inhib. Med. Chem.* **27**, 160–165.
Dostál, J., Brynda, J., Hrušková-Heidingsfeldová, O., Sieglová, I., Pichová, I. & Řezáčová, P. (2009). *J. Struct. Biol.* **167**, 145–152.
Dostál, J., Hamal, P., Pavlícková, L., Souček, M., Ruml, T., Pichová, I. & Hrusková-Heidingsfeldová, O. (2003). *J. Clin. Microbiol.* **41**, 712–716.
Dostál, J., Merkerová, M., Vinterová, Z., Pichová, I. & Hrušková-Heidingsfeldová, O. (2015). *Folia Microbiol. (Praha)*, **60**, 373–374.
Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T., Caldwell, J., Wang, J. & Kollman, P. (2003). *J. Comput. Chem.* **24**, 1999–2012.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Fusek, M., Smith, E. A., Monod, M. & Foundling, S. I. (1993). *FEBS Lett.* **327**, 108–112.
Gruene, T., Hahn, H. W., Luebben, A. V., Meilleur, F. & Sheldrick, G. M. (2014). *J. Appl. Cryst.* **47**, 462–466.
Hrušková-Heidingsfeldová, O. (2008). *Front. Biosci.* **13**, 7227–7242.
Hrušková-Heidingsfeldová, O., Dostál, J., Majer, F., Havlíková, J., Hradilek, M. & Pichová, I. (2009). *Biol. Chem.* **390**, 259–268.
Hube, B. & Naglik, J. (2001). *Microbiology*, **147**, 1997–2005.
Jurecka, P., Sponer, J., Cerný, J. & Hobza, P. (2006). *Phys. Chem. Chem. Phys.* **8**, 1985–1993.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
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. (2013). *Scand. J. Infect. Dis.* **45**, 842–848.
Lepšík, M., Řezáč, J., Kolář, M., Pecina, A., Hobza, P. & Fanfrlík, J. (2013). *Chempluschem*, **78**, 921–931.
Merkerová, M., Dostál, J., Hradilek, M., Pichová, I. & Hruskova-Heidingsfeldova, O. (2006). *FEMS Yeast Res.* **6**, 1018–1026.
Mueller, U., Darowski, N., Fuchs, M. R., Förster, R., Hellmig, M., Paithankar, K. S., Pühringer, S., Steffien, M., Zocher, G. & Weiss, M. S. (2012). *J. Synchrotron Rad.* **19**, 442–449.
Naglik, J., Albrecht, A., Bader, O. & Hube, B. (2004). *Cell. Microbiol.* **6**, 915–926.
Pammi, M., Holland, L., Butler, G., Gacser, A. & Bliss, J. M. (2013). *Pediatr. Infect. Dis. J.* **32**, e206–e216.
Parra-Ortega, B., Cruz-Torres, H., Villa-Tanaca, L. & Hernández-Rodríguez, C. (2009). *Mem. Inst. Oswaldo Cruz*, **104**, 505–512.
Pecina, A., Lepšík, M., Řezáč, J., Brynda, J., Mader, P., Řezáčová, P., Hobza, P. & Fanfrlík, J. (2013). *J. Phys. Chem. B*, **117**, 16096–16104.
Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). *J. Comput. Chem.* **25**, 1605–1612.
Pfaller, M. A. *et al.* (2010). *Diagn. Microbiol. Infect. Dis.* **67**, 162–171.
Pichová, I., Pavlíčková, L., Dostál, J., Dolejší, E., Hrušková-Heidingsfeldová, O., Weber, J., Ruml, T. & Souček, M. (2001). *Eur. J. Biochem.* **268**, 2669–2677.
Pryszcz, L. P., Németh, T., Gácser, A. & Gabaldón, T. (2013). *Genome Biol. Evol.* **5**, 2382–2392.
Rao, J. K., Erickson, J. W. & Wlodawer, A. (1991). *Biochemistry*, **30**, 4663–4671.
Riley, K. E., Pitoňák, M., Jurečka, P. & Hobza, P. (2010). *Chem. Rev.* **110**, 5023–5063.
Trofa, D., Gácser, A. & Nosanchuk, J. D. (2008). *Clin. Microbiol. Rev.* **21**, 606–625.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. (2004). *J. Comput. Chem.* **25**, 1157–1174.
Wlodawer, A., Li, M., Gustchina, A., Dauter, Z., Uchida, K., Oyama, H., Goldfarb, N. E., Dunn, B. M. & Oda, K. (2001). *Biochemistry*, **40**, 15602–15611.

Appendix F

THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

QM/MM Calculations Reveal the Different Nature of the Interaction of Two Carborane-Based Sulfamide Inhibitors of Human Carbonic Anhydrase II

Adam Pecina,[†] Martin Lepšík,[†] Jan Řezáč,[†] Jiří Brynda,[†,‡] Pavel Mader,[‡] Pavlína Řezáčová,[†,‡] Pavel Hobza,*[,†,§] and Jindřich Fanfrlík*[,†]

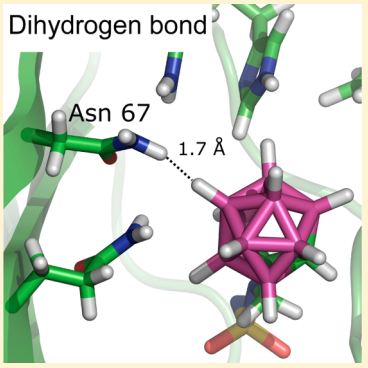
[†]Institute of Organic Chemistry and Biochemistry (IOCB), Academy of Sciences of the Czech Republic, v.v.i., Gilead Sciences and IOCB Research Center, Flemingovo nám. 2, 166 10, Prague 6, Czech Republic

[‡]Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Videnska 1083, Prague 4, Czech Republic

[§]Regional Center of Advanced Technologies and Materials, Department of Physical Chemistry, Palacký University, Olomouc 771 46, Olomouc, Czech Republic

Supporting Information

ABSTRACT: The crystal structures of two novel carborane-sulfamide inhibitors in the complex with human carbonic anhydrase II (hCAII) have been studied using QM/MM calculations. Even though both complexes possess the strongly interacting sulfamide···zinc ion motif, the calculations have revealed the different nature of binding of the carborane parts of the inhibitors. The neutral *closo*-carborane cage was bound to hCAII mainly via dispersion interactions and formed only very weak dihydrogen bonds. On the contrary, the monoanionic *nido* cage interacted with the protein mainly via electrostatic interactions. It formed short and strong dihydrogen bonds (stabilization of up to 4.2 kcal/mol; H···H distances of 1.7 Å) with the polar hydrogen of protein NH_2 groups. This type of binding is unique among all of the classical organic and inorganic inhibitors of hCAII. Virtual glycine scanning allowed us to identify the amino-acid side chains, which made important contributions to ligand-binding energies. In summary, using QM/MM calculations, we have provided a detailed understanding of the differences between the interactions of two carborane sulfamides, identified the amino acids of hCAII with which they interact, and thus paved the way for the computer-aided rational design of selective boron-cluster-containing hCAII inhibitors.



1. INTRODUCTION

Carboranes (or carbaboranes in formal nomenclature) are inorganic boron hydrides in which one or more BH^- units are replaced by an isoelectronic CH group. Boron clusters (including heteroatoms, such as carbon) form an astonishing variety of three-dimensional structures stabilized by 3-center 2-electron delocalized bonding. Two most extensively studied classes are represented by *closo* and *nido* carboranes, the former group having closed cages and the latter missing one vertex, optionally possessing a B−H−B hydrogen bridge.[1]

Carborane derivative chemistry has been rapidly expanding in the last decades,[2−4] and the compounds have found many novel applications in catalysis, nanomaterial science, or medicine as hydrophobic pharmacophores.[5−16] Several examples from the last category include boron neutron capture therapy (BNCT),[17−21] estrogen receptor agonist and antagonist, and the inhibition of HIV protease.[22−24] The properties that make carboranes such suitable entities for drug design are their hydrophobicity, geometry, stability, resistance to catabolism, and charge delocalization over their surface. The nature of the noncovalent binding of carboranes to (bio)macromolecules has been studied extensively in our previous works using mostly quantum mechanical (QM) calculations. Specific carborane−protein interactions range from C−H··H−B dihydrogen bonds[25,26] via B−H···Na^+ bridges[27,28] to $B_2H\cdots\pi$ and C−H···π hydrogen bonds[29] (reviewed in refs 22 and 23). However, it was not clear which of these interactions play roles in new cases of carborane−biomolecule complexes. For example, van der Waals and electrostatic interactions were found in the interaction of dihydrofolate reductase carborane inhibitors,[30] whereas for a carborane ligand of the vitamin D receptor, a "hydrophobic interaction" was postulated.[31]

Carbonic anhydrase (CA) belongs to a family of monomeric zinc metalloenzymes, which catalyze the reversible reaction of carbon dioxide hydration and bicarbonate dehydration. The control of this reaction is crucial in several very important physiological processes, such as maintaining the acid−base balance in blood and other tissues, and facilitating the transport of carbon dioxide and protons in the intracellular space, across biological membranes, and in the layers of the extracellular space.[32]

Received: October 15, 2013
Revised: November 27, 2013
Published: November 29, 2013

dx.doi.org/10.1021/jp410216m | J. Phys. Chem. B 2013, 117, 16096−16104

In the human organism, 12 catalytically active CA isoenzymes have been identified; they differ in their cellular localization and their expression in various tissues. In this study, we focus on hCAII, a ubiquitous enzyme essential for the maintenance of general acid−base balance. This enzyme is one of the most studied with a wealth of structural and biochemical data.[32]

The active site of hCAs is well conserved in sequence among various isoforms. It has a shape of a deep conical cleft and contains a Zn^{2+} ion with a bound hydroxyl group (Zn^{2+}−OH^-) coordinated by three histidine residues (His94, His96, His119), which are held in a distorted tetrahedral geometry. During the catalysis, the CO_2 substrate is weakly bound in a hydrophobic region located 3−4 Å from the zinc ion, and the enzyme works via a two-step mechanism: first, a nucleophilic attack of the zinc-bound hydroxide on the carbon dioxide forms a metal-bound bicarbonate followed by a displacement of the bicarbonate by a water molecule. Second, the regeneration of the active site occurs by the ionization of the zinc-bound water molecule and the removal of a proton from the active site. This happens via a series of proton-transfer steps, where several amino acids (Tyr7, Asn62, Asn67, Thr199, and Thr200) form a solvent network with His64, whose imidazole ring swaps between an inward and outward conformation and plays the role of a proton shuttle from the protein active site to the bulk solvent.[32]

Clinical regulation of the activity of human carbonic anhydrase (hCA) by small-molecule inhibitors proved to be a reliable therapeutic method for a number of human diseases and for several decades has been a major component of therapy for high blood pressure, glaucoma, hyperthyrosis, hypoglycemia, and recently also cancer.[33] The most important classes of hCA inhibitors are aryl-sulfonamides and inorganic anions (reviewed in ref 32). The sulfonamide/sulfamate SO_2−NH_2 headgroup is weakly acidic (pK_a is about 9−10); upon approaching the zinc ion, however, it leaves the proton to coordinate Zn^{2+} via electrostatic interactions. The tail of the inhibitor molecule can be substituted by specific functional groups to provide further interactions with the amino acids of hCA. The selectivity against the 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.[34]

Recently, novel carborane-based sulfamide inhibitors of hCAII and hCAIX have been designed, prepared, and shown to inhibit the enzymes in submicromolar range.[35] Two parent compounds possessing the *closo*- and *nido*-carborane cages, 1-methylenesulfamide-1,2-dicarba-*closo*-dodecaborane (**1a**) and 7-methylenesulfamide-(7,8-*nido*-dicarbaundecaborate) (**7a**) (Figure 1), have been crystallized in complex with hCAII, and crystal structures were determined using data at resolutions of 1.35 and 1.55 Å, respectively (PDB codes 4MDG and 4MDM).[35] These high-resolution structures revealed a binding mode in which the sulfamide moiety coordinated the zinc ion and the carborane cluster filled the conical enzyme's binding pocket. In contrast to the extensively studied ubiquitous interactions between Zn^{2+} and sulfamide moiety, a theoretical explanation of binding of various carborane cages to a protein is unique. Specifically, it was not clear which physical forces drive the binding of the carborane cages to hCAII. It could either be the hydrophobicity of the carborane cage,[36] dispersion interactions, an effect of the cage on the pK_a of the sulfamide moiety, or the formation of dihydrogen bonds.[25] To test these possibilities and gain deeper insight into the nature of the interactions, we performed DFT-D QM/MM calculations on the hCAII/**1a** and hCAII/**7a** complexes. By coupling the calculations with pK_a predictions using COSMO-RS, we were able to partition the binding energetics and elucidate the nature of the interaction of carborane-based sulfamide inhibitors with hCAII.

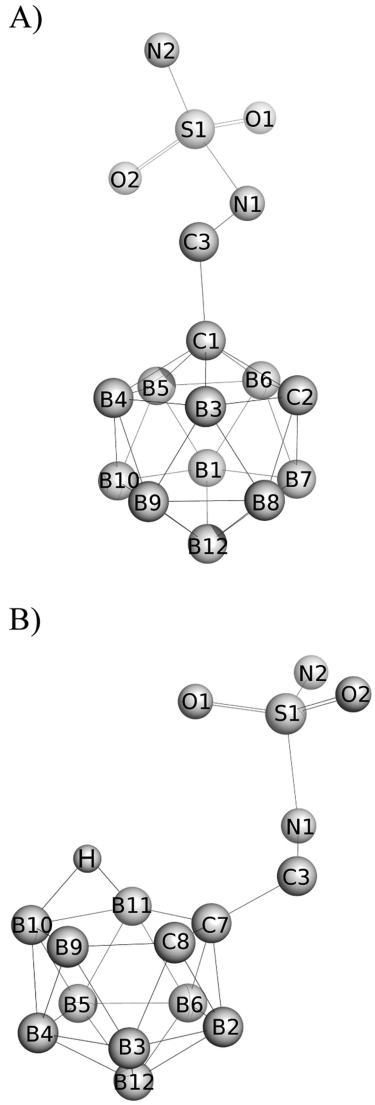


Figure 1. Structures of the (A) 1-methylenesulfamide-1,2-dicarba-*closo*-dodecaborane (**1a**) and (B) 7-methylenesulfamide-(7,8-*nido*-dicarbaundecaborate) (**7a**) compounds. Hydrogens are omitted for clarity with the exception of a hydrogen bridge in **7a**. This hydrogen bridge is located between either B10 and B11 atoms (shown) or B9 and B10 (not shown).

2. METHODS

2.1. Model Systems and Structure Preparation. Two crystal structures of hCAII in complex with **1a** and **7a** were determined at high resolutions of 1.35 and 1.55 Å, respectively (PDB codes 4MDG and 4MDM).[35] Hydrogens were added by the Reduce[37] and LEaP modules in the AMBER10 package[38] for the protein with an individual protonation of all of the histidines assigned on the basis of the visual inspection of their surroundings. The protein N-terminus and all of the lysines and arginines were positively charged, whereas the C-terminus and

dx.doi.org/10.1021/jp410216m | J. Phys. Chem. B 2013, 117, 16096−16104

all of the glutamates and aspartates were negatively charged to reflect the predominant state at pH 7. The inhibitors were protonated by the UCSF Chimera program for the sulfamide headgroup and manually for carborane cages.[39] The sulfamide moiety binds to the Zn^{2+} of hCA in a deprotonated NH^- form[32] and was thus modeled accordingly.

Compound **1a** has five possible rotational isomers (rotamers), differing in the positions of the carbon atom (C2) in the lower pentagon of the cage (five rotamers of the *closo* cage), while compound **7a** has two possible positions of the carbon (C8) atom (two enantiomers) combined with two positions of the B−H−B bridge (B9−H−B10 or B10−H−B11) (see Figure 1). The energies of the isolated isomers of compound **1a** and the enantiomers of compound **7a** were computed at the DFT method augmented with dispersion correction (D)[40] combined with the TPSS functional and the TZVP basis set. In addition, a rigid scan of the N1−C3−C1−C2 dihedral (Figure 1) at the same level was performed to shed light on the relation between the energy and the carbon position in the **1a** compound.

Complexes of all of these isomers in the complex with hCAII were prepared and fully optimized using the QM/MM procedure (see below). One crystal water molecule (Wat265) bridging the inhibitors and hCAII residues Thr198, Glu106, and Tyr7 was retained to maintain the integrity of the active site. Other waters were discarded. The atomic charges for the inhibitors were obtained by the RESP procedure[41] at the HF/6-31G* level. We have shown previously that RESP can be applied to carboranes to describe their interactions.[25] The protein parameters were obtained from the ff03 force field,[42] whereas for the ligands GAFF parameters were used.[43]

The positions of the added hydrogen atoms were relaxed in vacuo using the FIRE algorithm[44] followed by annealing (10 ps) from 600 to 0 K using the Berendsen thermostat[45] in the SANDER module of the AMBER 10 package.[38]

2.2. QM/MM Setup and Optimization. We used our in-house QM/MM program (CUBY3), which works as an interface between the Turbomole package[46] for QM calculations and the AMBER package[38] used for MM calculations. We applied an ONIOM-like subtractive scheme[47] with link atoms and mechanical embedding. The QM part was defined as the 4 Å surroundings of the ligands (480 atoms in total), which is around the current limit for DFT-D optimizations in a reasonable time. The MM part constituted the remainder of the protein, and the surrounding solvent was approximated by a generalized Born (GB) implicit model.[48]

All of the prepared complexes (see section 2.1) were optimized (specifically, the QM part was optimized while the MM part was frozen) using the DFT-D method.[40] We applied the resolution of the identity (RI) approximation[49] to the DFT method combined with the B-LYP functional and the SVP (3s2p1d/2s1p) basis set.[46] The QM/MM optimizations were performed in several rounds until the energy and gradient convergence criteria (ΔE = 0.005 kcal/mol, the maximum gradient of 1 kcal/mol/Å, the root-mean-square of the gradient of 0.5 kcal/mol/Å) were met. The interaction energies of all of the studied systems were determined on the optimized structures using the RI-DFT-D methodology at the TPSS/TZVP level.[46]

2.3. Strategy of the Calculations. The first step was to identify the most stable isomers of the **1a** and **7a** inhibitors in the complex with hCAII. This was done on the basis of the QM/MM energies of the optimized structures. The most stable complexes were selected and further used.

2.3.1. Carborane versus Sulfamide Interaction. To address the questions of the driving force of the binding of carborane-based inhibitors, we fragmented the ligands into two parts (the carborane cage and the sulfamide headgroup) and capped them by hydrogen atoms. Subsequently, their interaction "free" energies ($\Delta G'_{int}$) with the hCAII active site (QM part) were calculated using the RI-DFT-D TPSS/TZVP method. The DFT-D method was used with the more reliable[50] COSMO solvent model[51] instead of the GB model[48] utilized in the QM/MM optimizations.

The sulfamide headgroup interacted directly with the Zn^{2+} cation. An accurate calculation of the desolvation free energy of the bare cation is, however, a very difficult task, with the results depending on the method, atomic radii, etc.[52] To decrease the error of the calculated $\Delta G'_{int}$ between hCAII Zn^{2+} and the sulfamide headgroup, a single explicit water molecule (i.e., the first solvation shell) was considered to screen the Zn^{2+} charge following eq 1:[32]

$$RSO_2NH^- + hCAII{-}Zn^{2+}{-}OH_2^+ \rightarrow RSO_2NH^-Zn^{2+}{-}hCAII + H_2O \quad (1)$$

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

2.3.2. Interactions with the Active Site. The contribution of the amino acids in the active site to the binding was examined by a "virtual glycine scanning" approach, which was inspired by the "computational alanine scanning" procedure.[53] Single amino acids were substituted by glycine. The energy contributions ($\Delta\Delta G'_{int}$) were calculated as the difference between the original $\Delta G'_{int}$ at the QM/MM level with the wild-type amino acid and the new $\Delta G'_{int}$ with the mutated glycine residue.

2.3.3. Effect of Carborane Substituents on the pK_a of the Sulfamide Head Group. Further, we examined the role of the carborane moiety on the pK_a of the sulfamide, which in turn has profound effects on the inhibition characteristics (generally, the lower is the pK_a, the stronger is the binding).[32] Therefore, we calculated the pK_a of the sulfamide headgroup using the highly accurate COSMO-RS[54,55] at the BP86/TZVP level. We started with a series of substituted phenylsulfonamides with experimentally determined pK_a's (Supporting Information Table S2) and followed with all of the isomers of the sulfamide compounds **1a** and **7a**.

Moreover, the charge distribution within the inhibitors and their fragments (the headgroup and the cages) was calculated by RESP procedure[41] at the HF/6-31G* level to estimate a possible polarization between the two parts, which could influence the pK_a value.

3. RESULTS AND DISCUSSION

3.1. Comparing the X-ray and QM/MM-Optimized Structures. The studied hCAII/**1a** and hCAII/**7a** complexes obtained from X-ray crystallography (PDB codes: 4MDG and 4MDM)[35] were optimized using the QM/MM methodology. The differences between the two complexes were captured by the QM/MM optimizations. Importantly, the differences between the QM/MM and X-ray structures are significantly smaller than the differences between the X-ray structures (see Figure 2). The QM/MM optimized geometries are thus reliable and can be further used in this study.

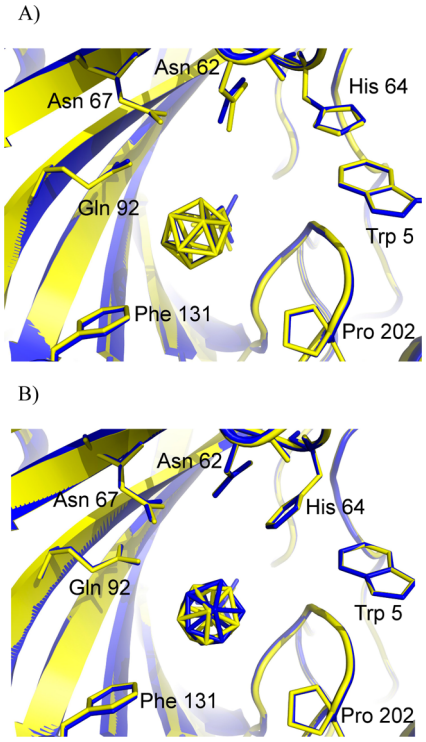


Figure 2. An overlay of the X-ray (in blue) and QM/MM optimized (in yellow) structures of the (A) hCAII/**1a** and (B) hCAII/**7a** complexes.

The X-ray structures of the hCAII/**1a** and hCAII/**7a** complexes slightly differ in the active site. The most significant difference was in the position of His64 and subsequently in the positions of the neighboring amino acids Asn62, Asn67, and Gln92 (Figure 3a). The *nido* carborane cage of compound **7a** interacted with the inward conformation of the imidazole ring of His64 via a strong dihydrogen bond (the H···H distance of 2.1 Å), whereas in the hCAII/**1a** complex, the side chain of His64 was shifted outward from the inhibitor **1a**.

Besides the changes in the orientation of the amino-acid side chains interacting with the carborane cages, compounds **1a** and **7a** also differed in the interactions of the C3−N1 linker with the enzyme's amino-acid residues. The hydrogen bond with Thr200 was shorter in hCAII/**1a** than in the hCAII/**7a** complex (Figure 3b,c). Moreover, a weak hydrogen bond (C−H···N) was also formed with the side chain of Leu198 in the hCAII/**1a** complex. In contrast, the C3−N1 part of the head of **7a** interacted with Leu198 via dispersion aliphatic−aliphatic interactions (see Supporting Information Figure S1).

3.2. Carborane Isomer Stabilities. Substituted carboranes may occur in several rotameric states. If unconstrained, these states can interchange, like in the case of the rotational freedom of cobalta bis(dicarbollide).[57] However, in the cases of inter-[27] or intra-[58] molecular interactions, a single rotational state may be preferred.[30] To obtain insight into the rotamer preferences of **1a** and how they are influenced by the protein surroundings, we studied the rotational profile of the isolated compound **1a** using QM calculations in vacuum as well as by the use of QM/MM in the GB solvent in the complex with hCAII. For **7a**, we performed similar calculations but focused on the two enantiomers of **7a** (racemic mixture) and the two positions of the hydrogen-bond bridge instead. Hydrogen atoms are not present in the X-ray structures at the given resolution and may be fluxional;[26] altogether, there were thus four isomers of **7a**.

*Rotational Profile of Isolated **1a**.* The rigid scan of the N1−C3−C1−C2 dihedral (Figure 1a) showed that the carborane moiety in the isolated **1a** preferred the N1−C3−C1−C2 dihedral angle of about −20°. At room temperature, the N1−C3−C1−C2 dihedral of **1a** could range from about −80° to +21° (with the energy difference being smaller than 2 kcal/mol) (Figure 4). The rotational barrier for a complete 360° rotation was about 8 kcal/mol high. The most stable rotamers were stabilized by a weak hydrogen bond between the C2−H vertex and the oxygen of the sulfamide headgroup (a distance of about 2.5 Å). The unfavorable energies of the less stable rotamers were caused by an electrostatic repulsion between the B−H group and the oxygen of the sulfamide head.

*Rotational Profile of **1a** in the Complex with hCAII.* The relative QM/MM energies of the rotamers of **1a** in the complex with hCAII are shown in Figure 4. Although the barrier for the 360° rotation did not change (about 8 kcal/mol high), the well around the minimum is broader in the complex than in isolation, thus allowing wider rotation at room temperature (from −60° to +100°). The computed minimum also agreed with the experimentally determined position of the carbon atoms of **1a** structural data (a dihedral angle of −44°).

*Isomer Stabilities of **7a** − Isolated and Complexed with hCAII.* The DFT-D (TPSS/TZVP) calculations on the isolated **7a** molecule showed that both positions of the B−H−B bridge should be considered for both enantiomers, because their relative energies differed only by up to 1.5 kcal/mol.

In the complex with hCAII, the energy differences between the isomers were only slightly larger (below 3 kcal/mol). We can thus assume that in addition to the most stable isomer, the other isomers might also be found in the complex with hCAII. However, they are supposed to be less populated. The most stable isomer was the ρ enantiomer with the hydrogen bridge between B10 and B11 (Figure 1b). The position of the C8 of this isomer is also in agreement with crystallographic observations.

3.3. Driving Force of the Carborane-Sulfamide Inhibitor Binding to hCAII. To address the question of the driving force of the binding of carborane-based inhibitors into hCAII, we fragmented the ligands into the carborane cage and the sulfamide head. We calculated the interaction "free" energies ($\Delta G'_{int}$), and the gas-phase interaction energies on the structures optimized in the water environment (ΔE_{int}) using the DFT-D method and COSMO implicit solvent model (for details, see section 2.3.1). The influence of amino acids beyond the QM region is comparable ($\Delta G'_{int}$ of −9.3 and −8.7 kcal/mol for **1a** and **7a**, respectively), which justifies their neglect when comparing binding of these compounds. The calculated values as well as the experimental binding affinities are summarized in Table 1. The calculated relative binding free energies of **1a** and **7a** to hCAII are in very good agreement with the experimental ones ($\Delta\Delta G'_{int}$ of −0.9 kcal/mol as compared to $\Delta\Delta G^{\circ}_{b}$ of −1.4 kcal/mol). However, it should be stressed here that several important terms of the binding affinity, such as the binding entropy and protein deformation, are not calculated here (the problems associated with the calculations of these terms are discussed in detail in refs 59 and 60). Thus, the interplay of these other terms is difficult to estimate. Definitely, however, they would disfavor binding and therefore bring the calculated absolute values closer to the experimental ones. Despite these limitations, the reproduced energy difference

dx.doi.org/10.1021/jp410216m | J. Phys. Chem. B 2013, 117, 16096−16104

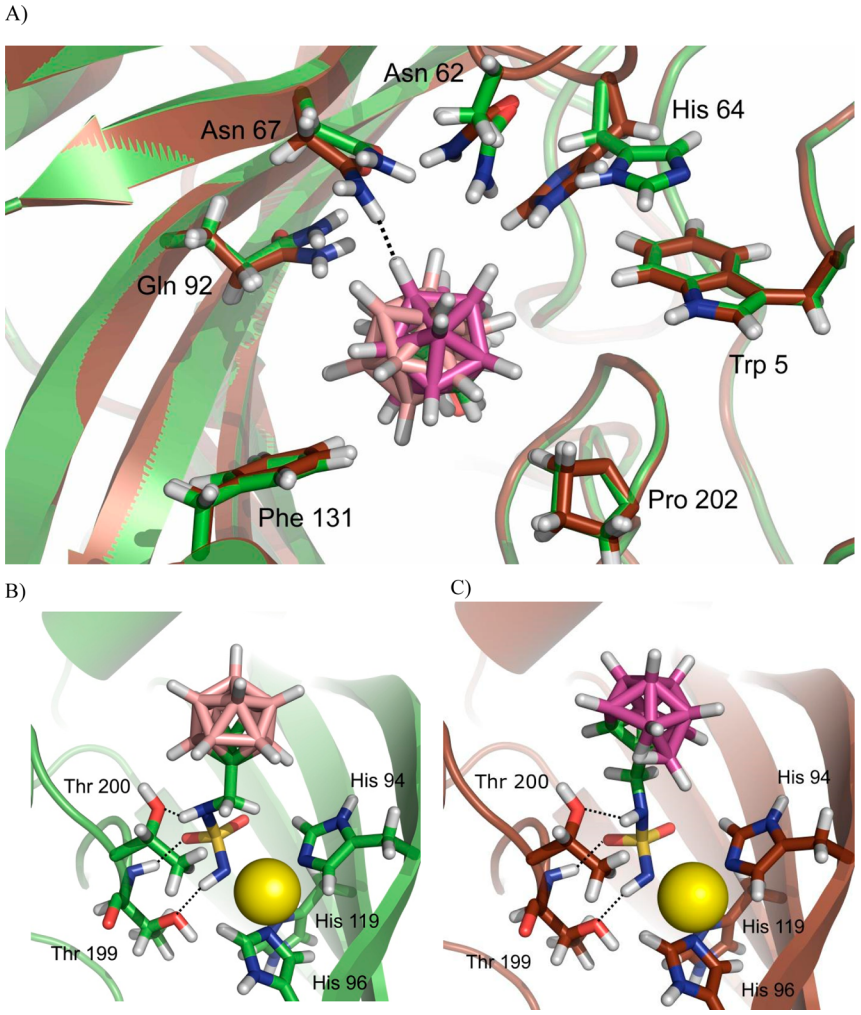


Figure 3. (A) An overlay of hCAII/**1a** and hCAII/**7a** QM/MM optimized structures. Compound **1a** is in pink, **7a** in magenta, hCAII of hCAII/**1a** in green, and hCAII of hCAII/**7a** in brown. The hydrogen-bond interactions of the sulfamide headgroup of **1a** (B) and **7a** (C) with the hCAII active site, where Zn^{2+} is visualized as a yellow sphere. Figure was prepared with PyMol, version 2006.[56]

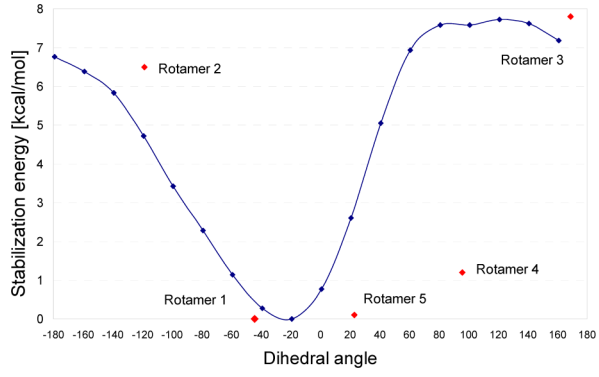


Figure 4. A dihedral rigid scan of the isolated compound **1a** calculated at the DFT-D:TPSS/TZVP level in vacuum (blue) and the graph of the dependence of relative stabilization energy on the dihedral angle of **1a** in the hCAII/**1a** complex (red).

Table 1. Decomposition of the DFT-D (TPSS/TZVP) Interaction Energy between the Ligand Fragments and hCAII[a]

	$\Delta G'_{int}$	ΔE_{int}	D	ΔG°_{b}
1a	−37.7	−184.8	−42.0	−8.4 ± 0.1
7a	−36.8	−246.6	−38.3	−7.0 ± 0.2
1a headgroup	−26.5	−162.4	−17.1	
7a headgroup	−26.1	−165.3	−16.8	
1a cage	−11.2	−22.4	−24.9	
7a cage	−10.7	−81.3	−21.5	

[a]The structures were cut out of the hCAII/**1a** and hCAII/**7a** complexes optimized by the QM/MM method. The interaction "free" energy ($\Delta G'_{int}$) calculated in the COSMO solvent model, the gas-phase interaction energy (ΔE_{int}), the dispersion energy (D) contribution to the interaction energy, and the experimental binding free energy ΔG°_{b} (calculated by $\Delta G^{\circ}_{b} = RT^{*}\ln K_i$, where dissociation constant (K_i) is taken from ref 35) are all in kcal/mol.

enables us to analyze the calculated binding affinities in more detail. The $\Delta G'_{int}$ of the carborane cages was significantly weaker than that of the sulfamide moiety for both **1a** and **7a** (−26.5 vs −11.2 and −26.1 vs −10.7 kcal/mol, respectively), which agrees with previous observations concerning the energetic importance of the sulfonamide (or isosteric) moiety for the binding to hCAII.[32]

More importantly, the difference between the ΔE_{int} of the *closo* and *nido* cages was large. The *closo*-carborane cage interacts 58.9 kcal/mol less strongly than the *nido* cage (Table 1). In the case of the *closo*-carborane cage, the dispersion

dx.doi.org/10.1021/jp410216m | J. Phys. Chem. B 2013, 117, 16096−16104

energy itself was larger than the total ΔE_{int} and thus 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. The driving force of the *nido* cage hence seemed to be of an electrostatic character. The desolvation penalty of the neutral *closo* cage was small; consequently, the resulting $\Delta G'_{int}$ values of the *closo* and *nido* cages were comparable.

3.4. Sulfamide pK_a in Aryl- and Carborane-Based hCA Inhibitors. The pK_a calculations of sulfamide inhibitors using the COSMO-RS method[54] were first performed on a small series of aryl-based compounds with experimentally measured pK_a values (Supporting Information Table S2). The computed values are down-shifted from the experimental ones by 0.6−1.3 units, but the relative trend is excellent with a correlation coefficient of R^2 = 0.88 (Figure 5).

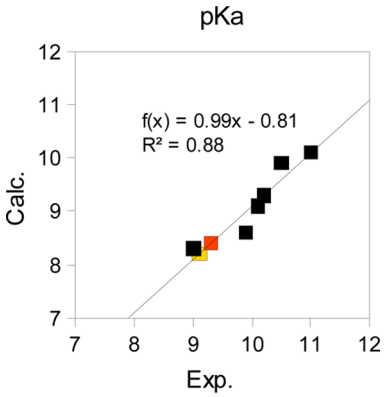


Figure 5. The calculated pK_a values plotted against the experimental values. The experimental values are taken from ref 32 and colored black (see also Supporting Information Table S2). For calculations, the COSMO-RS methodology[54] was used. The pK_a values of **1a** (in orange) and **7a** (in yellow) are estimated using the COSMO-RS results and the correlation equation.

Further, we estimated the pK_a of the sulfamide in the carborane-based inhibitors. The COSMO-RS results showed that the carborane cages of **1a** and **7a** had a very similar effect on the sulfonamide moiety. The COSMO-RS pK_a values of the **1a** and **7a** inhibitors were calculated to be about 8.4 and 8.2. Using the correlation equation (Figure 5), the pK_a values for the sulfamide moieties in **1a** and **7a** can be estimated to be about 9.3 and 9.1, respectively. When we compare **1a** and **7a** with the other compounds (Supporting Information Table S2 and Figure 5; i.e., carborane cages changed to phenyl), we conclude that the *closo*- and *nido*-carborane cages lower the pK_a of the sulfamide group. This may be connected with the electron deficiency of the boron clusters, thus contributing to the potency of carborane-based sulfamide inhibitors.

3.5. Contributions of hCA Active-Site Amino Acids to Binding. The binding role of the individual amino-acid side chains in the active site was studied by a "virtual glycine scanning" approach. The $\Delta\Delta G'_{int}$ upon single amino-acid mutation into glycine for the most stable **1a** rotamer (a dihedral angle of −44°) is shown in Figure 6. The *closo* cage of compound **1a**, whose binding was driven mainly by dispersion energy (see above), had the strongest dihydrogen bonding interaction with Gln92 (the H···H distance of 2.0 Å; $\Delta\Delta G'_{int}$ of −2.3 kcal/mol) and with Phe131 (the H···H distance of 2.2 Å; $\Delta\Delta G'_{int}$ of −2.0 kcal/mol). In general, the dihydrogen bonds of the *closo* cage were weak, that is, only with nonpolar C−H groups and rather long and at the margin of the range of H···H distances (with Phe131 2.2 Å; with Pro202 2.9 Å; with Asn62 3.0 Å). All other important amino acids such as Thr200, Val121, and Leu198 had, besides the dihydrogen interactions (2.5 Å; 2.5 and 2.7 Å; 2.3 and 2.6 Å, respectively), also a strong contribution to the binding via an interaction with the sulfamide headgroup.

To explain the differences in rotational profile of **1a** in isolation and in the complex with hCAII, a "virtual glycine scan" was also performed for rotamer 4 (a dihedral angle of +96°), which is considerably more stable in the complex than in isolation. Rotamer 4 had stronger (more negative) $\Delta\Delta G'_{int}$ with Thr200, Val121, and Gln92 (by 1.7, 0.9, and 0.5 kcal/mol, respectively; see Supporting Information Graph S3). In the case of Thr200, the increase in $\Delta\Delta G'_{int}$ was due to a weak hydrogen bond ($C-H^{\delta+}\cdots O^{\delta-}$), which replaced repulsion between the B−H vertex and the O atom of Thr200 ($B-H^{\delta-}\cdots O^{\delta-}$).

The *nido* carborane cage (compound **7a**) differs significantly from the *closo* carborane cage. Here, we found a short dihydrogen bond between a B5−H vertex and the polar NH_2 group of Asn67 with the H···H distance of 1.7 Å. This dihydrogen bond presented the largest contribution to the binding of the *nido* cage. $\Delta\Delta G'_{int}$ was calculated to be −4.1 kcal/mol (see Figure 6). It should be stressed here that Asn67 interacted with compound **7a** only via the single dihydrogen bond and had no other contacts. The calculated interaction can thus be directly assigned to the single dihydrogen bond and is in agreement with both the distance and the energetic ranges found for carborane−biomolecule interactions.[25] The neighboring Asn62 had also a more attractive interaction with the *nido* carborane cage than it had with the *closo* cage (−1.8 vs −0.6 kcal/mol, respectively). Another larger change concerned the flexible His64, which provided $\Delta\Delta G'_{int}$ with the *nido* cage of −2.0 kcal/mol (a single dihydrogen bond with the H···H distance of 2.1 Å). In the case of the *closo* carborane cage, His64 was far away from the inhibitor and did not have any interaction ($\Delta\Delta G'_{int}$ of 0.4 kcal/mol). The complexes of compounds **1a** and **7a** differed also in the position of the head (specifically the −C3−N1− part, see section 3.1). The hCAII/**1a** had a more favorable interaction with Leu198 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** (Supporting Information Figure S1).

3.6. Charge Analysis. Partial atomic charges were calculated using RESP (a correct description of dihydrogen bonding, see ref 25) and analyzed. The *nido*-carborane cage of **7a** was negatively charged, which translated into more negative charges on boron-bound hydrogens (of −0.15) and consequently also the formation of stronger dihydrogen bonds. The *closo*-carborane cage of **1a** was neutral, and boron-bound hydrogens only had a slightly negative charge (of −0.08). This explained why the closed cage of **1a** formed considerably weaker dihydrogen bonds. It seems that the total negative charge of the carborane cage is a prerequisite for the formation of strong and short dihydrogen bonds (cf., ref 25). This knowledge may help fine-tune the binding affinity of carborane-containing ligands in rational drug design.

4. CONCLUSIONS

We have conducted the very first QM/MM study of two novel carborane-sulfamide inhibitors of human carbonic anhydrase II (hCAII). The outcome is a detailed atomistic and energetic understanding of the nature of inhibitor binding. Although the

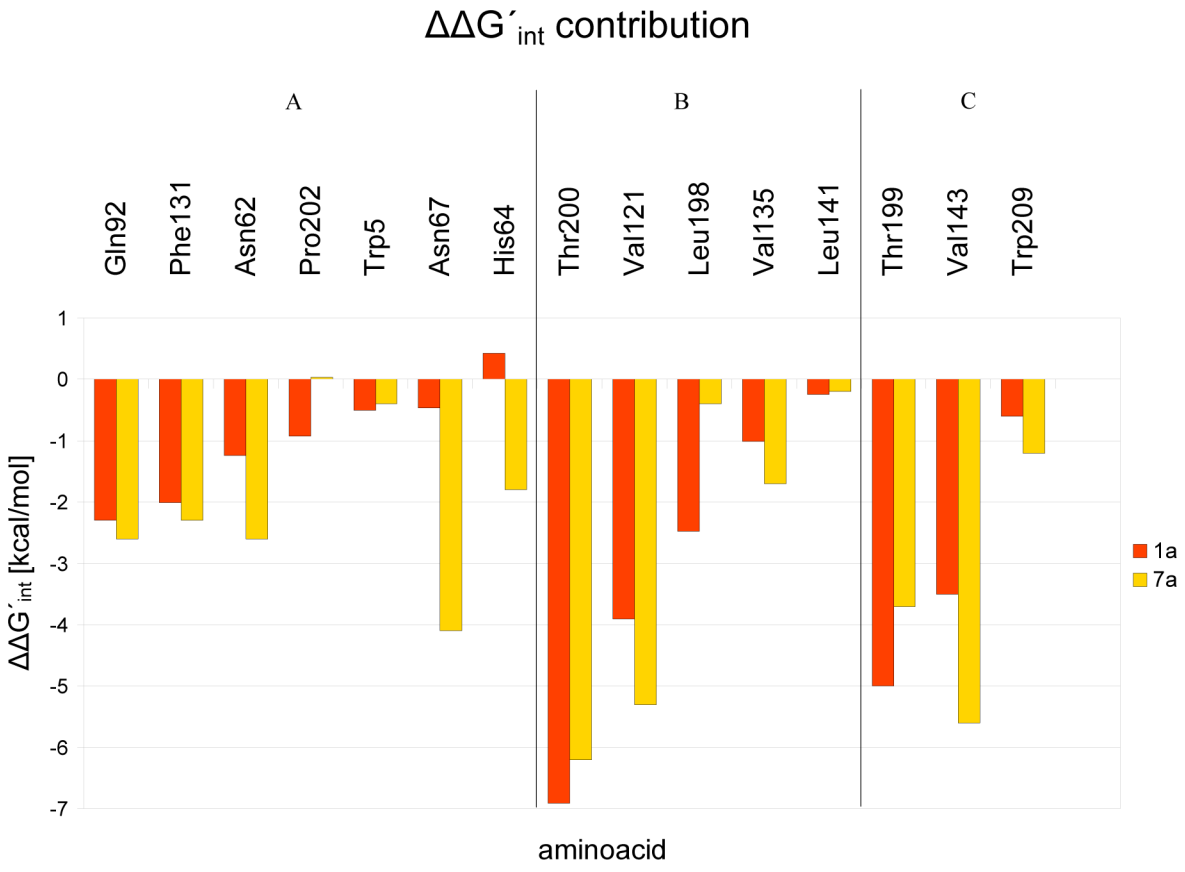


Figure 6. The contribution of single amino acids to the interaction "free" energy $\Delta\Delta G'_{int}$ as obtained from a "virtual glycine scan". (A) The first seven amino acids (from Gln92 to His64) interact only with the carborane cage, (B) the next five (from Thr200 to Leu141) have interactions with both the cage and the sulfamide head, and (C) the last three interact only with the sulfamide head (from Thr199 to Trp209).

studied inhibitors bind mainly via the sulfamide moiety to the zinc ion, the different nature of binding of the carborane part of the inhibitors was revealed. The neutral *closo*-carborane cage was bound mainly via dispersion interactions and formed only very weak dihydrogen bonds (the H···H distance greater than 2.2 Å; only with nonpolar C−H groups). In contrast, the negatively charged *nido* cage interacted with the protein mainly via electrostatic interactions. It formed short and strong dihydrogen bonds (with an energy up to −4.2 kcal/mol; the H···H distances as short as 1.7 Å) with the polar hydrogen of NH_2 groups. Both electron-deficient *closo*- and *nido*-carborane cages lowered the pK_a of the sulfamide anchor as compared to phenyl and thus also contributed to the binding affinity. A detailed understanding of the differences in the interactions of various carboranes is important for their future use in rational drug design.

■ ASSOCIATED CONTENT

Supporting Information

Figure of the interaction of the side chain of Leu198 with compounds **1a** and **7a**, table of pK_a values calculated using COSMO-RS, comparison with experimental data, and a graph of "virtual glycine scan" results. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: jindrich.fanfrlik@uochb.cas.cz.
*E-mail: pavel.hobza@uochb.cas.cz.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was part of the Research Projects RVO 61388963 and 68378050 awarded by the Academy of Sciences of the Czech Republic. We acknowledge the financial support from the Czech Science Foundation [P208/12/G016] and the operational program Research and Development for Innovations of the European Social Fund (CZ 1.05/2.1.00/03/0058).

■ REFERENCES

(1) Grimes, R. N. *Carboranes*, 2nd ed.; Academic Press: London, 2011.
(2) Korbe, S.; Schreiber, P. J.; Michl, J. Chemistry of the carba-closo-dodecaborate(−) anion, $CB_{11}H_{12}^-$. *Chem. Rev.* **2006**, *106*, 5208.
(3) Bregadze, V. I. Dicarba-closo-dodecaboranes $C_2B_{10}H_{12}$ and their derivatives. *Chem. Rev.* **1992**, *92*, 209.
(4) Štíbr, B. Carboranes other than $C_2B_{10}H_{12}$. *Chem. Rev.* **1992**, *92*, 225.
(5) Serrano-Andrés, L.; Oliva, J. M. Photochemical window mechanism for controlled atom release in carborane endohedral boxes: Theoretical evidence. *Chem. Phys. Lett.* **2006**, *432*, 235.
(6) Crespo, O.; Gimeno, M. C.; Laguna, A.; Ospino, I.; Aullon, G.; Oliva, J. M. Organometallic gold complexes of carborane. Theoretical comparative analysis of ortho, meta, and para derivatives and luminescence studies. *Dalton Trans.* **2009**, 3807.
(7) Oliva, J. M.; Serrano-Andrés, L.; Klein, D. J.; Schleyer, P. V.; Michl, J. Design of Carborane Molecular Architectures via Electronic Structure Computations. *Int. J. Photoenergy* **2009**.

(8) Grimes, R. N. Metallacarboranes in the new millennium. *Coord. Chem. Rev.* **2000**, *200*, 773.
(9) Teixidor, F.; Nuñez, R.; Flores, M. A.; Demonceau, A.; Viñas, C. Forced exo-nido rhoda and ruthenacarboranes as catalyst precursors: a review. *J. Organomet. Chem.* **2000**, *614*, 48.
(10) Bednarska, K.; Olejniczak, A. B.; Wojtczak, B. A.; Sulowska, Z.; Lesnikowski, Z. J. Adenosine and 2′-deoxyadenosine modified with boron cluster pharmacophores as new classes of human blood platelet function modulators. *ChemMedChem* **2010**, *5*, 749.
(11) Lesnikowski, Z. J. Boron clusters - A new entity for DNA-oligonucleotide modification. *Eur. J Org. Chem.* **2003**, 4489.
(12) Issa, F.; Kassiou, M.; Rendina, L. M. Boron in drug discovery: carboranes as unique pharmacophores in biologically active compounds. *Chem. Rev.* **2011**, *111*, 5701.
(13) Scholz, M.; Hey-Hawkins, E. Carbaboranes as pharmacophores: properties, synthesis, and application strategies. *Chem. Rev.* **2011**, *111*, 7035.
(14) Satapathy, R.; Dash, B. P.; Maguire, J. A.; Hosmane, N. S. New Developments in the Medicinal Chemistry of Carboranes. *Collect. Czech. Chem. Commun.* **2010**, *75*, 995.
(15) Lesnikowski, Z. J. Boron units as pharmacophores - New applications and opportunities of boron cluster chemistry. *Collect. Czech. Chem. Commun.* **2007**, *72*, 1646.
(16) Sivaev, I. B.; Bregadze, V. V. Polyhedral Boranes for Medical Applications: Current Status and Perspectives. *Eur. J. Inorg. Chem.* **2009**, 1433.
(17) Hawthorne, M. F.; Maderna, A. Applications of radiolabeled boron clusters to the diagnosis and treatment of cancer. *Chem. Rev.* **1999**, *99*, 3421.
(18) Bialek-Pietras, M.; Olejniczak, A. B.; Tachikawa, S.; Nakamura, H.; Lesnikowski, Z. J. Towards new boron carriers for boron neutron capture therapy: Metallacarboranes bearing cobalt, iron and chromium and their cholesterol conjugates. *Bioorg. Med. Chem.* **2013**, *21*, 1136.
(19) Schaffran, T.; Lissel, F.; Samatanga, B.; Karlsson, G.; Burghardt, A.; Edwards, K.; Winterhalter, M.; Peschka-Suss, R.; Schubert, R.; Gabel, D. Dodecaborate cluster lipids with variable headgroups for boron neutron capture therapy: Synthesis, physical-chemical properties and toxicity. *J. Organomet. Chem.* **2009**, *694*, 1708.
(20) Justus, E.; Awad, D.; Hohnholt, M.; Schaffran, T.; Edwards, K.; Karlsson, G.; Damian, L.; Gabel, D. Synthesis, liposomal preparation, and in vitro toxicity of two novel dodecaborate cluster lipids for boron neutron capture therapy. *Bioconjugate Chem.* **2007**, *18*, 1287.
(21) Nakamura, H.; Ueno, M.; Lee, J. D.; Ban, H. S.; Justus, E.; Fan, P.; Gabel, D. Synthesis of dodecaborate-conjugated cholesterols for efficient boron delivery in neutron capture therapy. *Tetrahedron Lett.* **2007**, *48*, 3151.
(22) Řezáčová, P.; Cígler, P.; Matìjíček, P.; Lepšík, M.; Pokorná, J.; Grunner, B.; Konvalinka, J. *Boron Science - New Technologies and Applications*; CRC Press: New York, 2011.
(23) Farras, P.; Juarez-Perez, E. J.; Lepšík, M.; Luque, R.; Nuñez, R.; Teixidor, F. Metallacarboranes and their interactions: theoretical insights and their applicability. *Chem. Soc. Rev.* **2012**, *41*, 3445.
(24) Cígler, P.; Kožíšek, M.; Řezáčová, P.; Brynda, J.; Otwinowski, Z.; Pokorná, J.; Plešek, J.; Gruner, B.; Doleckova-Marešova, L.; Máša, M.; et al. From nonpeptide toward noncarbon protease inhibitors: metallacarboranes as specific and potent inhibitors of HIV protease. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 15394.
(25) Fanfrlík, J.; Lepšík, M.; Horinek, D.; Havlas, Z.; Hobza, P. Interaction of carboranes with biomolecules: formation of dihydrogen bonds. *ChemPhysChem* **2006**, *7*, 1100.
(26) Fanfrlík, J.; Hnyk, D.; Lepšík, M.; Hobza, P. Interaction of heteroboranes with biomolecules. Part 2. The effect of various metal vertices and exo-substitutions. *Phys. Chem. Chem. Phys.* **2007**, *9*, 2085.
(27) Fanfrlík, J.; Brynda, J.; Řezáč, J.; Hobza, P.; Lepšík, M. Interpretation of Protein/Ligand Crystal Structure using QM/MM Calculations: Case of HIV-1 Protease/Metallacarborane Complex. *J. Phys. Chem. B* **2008**, *112*, 15094.
(28) Matějíček, P.; Zedník, J.; Uselova, K.; Pleštil, J.; Fanfrlík, J.; Nykanen, A.; Ruokolainen, J.; Hobza, P.; Procházka, K. Stimuli-Responsive Nanoparticles Based on Interaction of Metallacarborane with Poly(ethylene oxide). *Macromolecules* **2009**, *42*, 4829.
(29) Sedlák, R.; Fanfrlík, J.; Hnyk, D.; Hobza, P.; Lepšík, M. Interactions of Boranes and Carboranes with Aromatic Systems: CCSD(T) Complete Basis Set Calculations and DFT-SAPT Analysis of Energy Components. *J. Phys. Chem. A* **2010**, *114*, 11304.
(30) 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.; et al. Novel boron-containing, nonclassical antifolates: Synthesis and preliminary biological and structural evaluation. *J. Med. Chem.* **2007**, *50*, 3283.
(31) Fujii, S.; Masuno, H.; Taoda, Y.; Kano, A.; Wongmayura, A.; Nakabayashi, M.; Ito, N.; Shimizu, M.; Kawachi, E.; Hirano, T.; Endo, Y.; et al. Boron cluster-based development of potent nonsecosteroidal vitamin D receptor ligands: direct observation of hydrophobic interaction between protein surface and carborane. *J. Am. Chem. Soc.* **2011**, *133*, 20933.
(32) Krishnamurthy, V. M.; Kaufman, G. K.; Urbach, A. R.; Gitlin, I.; Gudiksen, K. L.; Weibel, D. B.; Whitesides, G. M. Carbonic anhydrase as a model for biophysical and physical-organic studies of proteins and protein-ligand binding. *Chem. Rev.* **2008**, *108*, 946.
(33) Supuran, C. T. Carbonic anhydrases: novel therapeutic applications for inhibitors and activators. *Nat. Rev. Drug Discovery* **2008**, *7*, 168.
(34) Gitto, R.; Damiano, F. M.; Máder, P.; De Luca, L.; Ferro, S.; Supuran, C. T.; Vullo, D.; Brynda, J.; Řezáčová, P.; Chimirri, A. Synthesis, Structure-Activity Relationship Studies, and X-ray Crystallographic Analysis of Arylsulfonamides as Potent Carbonic Anhydrase Inhibitors. *J. Med. Chem.* **2012**, *55*, 3891.
(35) Brynda, J.; Máder, P.; Šícha, V.; Fábry, M.; Poncová, K.; Bakardiev, M.; Grüner, B.; Cígler, P.; Řezáčová, P. Carborane-Based Carbonic Anhydrase Inhibitors. *Angew. Chem., Int. Ed.*, DOI: 10.1002/anie.201307583.
(36) Plešek, J. Potential Applications of the Boron Cluster Compounds. *Chem. Rev.* **1992**, *92*, 269.
(37) Word, J. M.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol.* **1999**, *285*, 1735.
(38) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; et al. *AMBER 10*; University of California: San Francisco, CA, 2008.
(39) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605.
(40) Jurečka, P.; Černý, J.; Hobza, P.; Salahub, D. Density functional theory augmented with an empirical dispersion term. Interaction energies and geometries of 80 noncovalent complexes compared with ab initio quantum mechanics calculations. *J. Comput. Chem.* **2007**, *28*, 555.
(41) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges - the RESP model. *J. Phys. Chem.* **1993**, *97*, 10269.
(42) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G. M.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; et al. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24*, 1999.
(43) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157.
(44) Bitzek, E.; Koskinen, P.; Gahler, F.; Moseler, M.; Gumbsch, P. Structural relaxation made simple. *Phys. Rev. Lett.* **2006**, *97*, 170201.
(45) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684.

(46) Ahlrichs, R.; Bar, M.; Haser, M.; Horn, H.; Kolmel, C. Electronic structure calculations on workstation computers − the program system Turbomole. *Chem. Phys. Lett.* **1989**, *162*, 165.
(47) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. A multilayered integrated MO+MM method for geometry optimizations and single point energy predictions. A test for Diels-Alder reactions and Pt(P(t-Bu)(3))(2)+H-2 oxidative addition. *J. Phys. Chem.* **1996**, *100*, 19357.
(48) Tsui, V.; Case, D. A. Theory and applications of the generalized Born solvation model in macromolecular Simulations. *Biopolymers* **2001**, *56*, 275.
(49) Feyereisen, M.; Fitzgerald, G.; Komornicki, A. Use of approximate integrals in ab initio theory. An application in MP2 energy calculations. *Chem. Phys. Lett.* **1993**, *208*, 359.
(50) Kolář, M.; Fanfrlík, J.; Lepšík, M.; Forti, F.; Luque, F. J.; Hobza, P. Assessing the Accuracy and Performance of Implicit Solvent Models for Drug Molecules: Conformational Ensemble Approaches. *J. Phys. Chem. B* **2013**, *117*, 5950.
(51) Klamt, A.; Schuurmann, G. COSMO - A new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799.
(52) Gutten, O.; Besseová, I.; Rulíšek, L. Interaction of metal ions with biomolecular lignds: How accurate are calculated free energies associated with metal ion complexation? *J. Phys. Chem. A* **2011**, *115*, 11394.
(53) Massova, I.; Kollman, P. A. Computational alanine scanning to probe protein - protein interactions: A novel approach to evaluate binding free energies. *J. Am. Chem. Soc.* **1999**, *121*, 8133.
(54) Klamt, A.; Ecklert, F.; Hornig, M. COSMO-RS: A novel view to physiological solvation and partition questions. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 355.
(55) Klamt, A. Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *J. Phys. Chem.* **1995**, *99*, 2224.
(56) DeLano, W. L. *The PyMOL Molecular Graphics System*; Scientific: Palo Alto, CA, 2002.
(57) Buhl, M.; Hnyk, D.; Machacek, J. Computational study of structures and prperties of metalloboranes: cobalt bis(dicarbollide). *Chem.-Eur. J.* **2005**, *11*, 4109.
(58) Juarez-Perez, E. J.; Viñas, C.; Teixidor, F.; Nuñez, R. First example of the formation of a Si−C bond from an intramolecular Si−H···H−C diyhydrogen interaction in a metallacarborane: A theoretical study. *J. Organomet. Chem.* **2009**, *694*, 1764.
(59) Lepšík, M.; Řezáč, J.; Kolář, M.; Pecina, A.; Hobza, P.; Fanfrlík, J. The semiempirical quantum mechanical scoring function for in silico drug design. *ChemPlusChem* **2013**, *78*, 921.
(60) Fanfrlík, J.; Brahmkshtriya, P. S.; Řezáč, J.; Jílková, A.; Horn, M.; Mareš, M.; Hobza, P.; Lepšík, M. Quantum mechanics-based scoring rationalizes the irreversible inactivation of parasitic Schistosoma mansoni cysteine peptidase by vinyl sulfone inhibitors. *J. Phys. Chem. B* **2013**, *117*, 14973.

SUPPLEMENTARY MATERIAL FOR

QM/MM Calculations Reveal the Different Nature of the Interaction of Two Carborane-Based Sulfamide Inhibitors of Human Carbonic Anhydrase II

Adam Pecina[1], Martin Lepšík[1], Jan Řezáč[1], Jiří Brynda[1,2], Pavel Mader[2], Pavlína Řezáčová[1,2], Pavel Hobza[1,3] and Jindřich Fanfrlík[1]**

[1] Institute of Organic Chemistry and Biochemistry (IOCB), Academy of Sciences of the Czech Republic, v.v.i., Gilead Sciences and IOCB Research Center, Flemingovo nám. 2, 166 10, Prague 6, Czech Republic

[2] Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Videnska 1083, Prague 4, Czech Republic

[3] Regional Center of Advanced Technologies and Materials, Department of Physical Chemistry, Palacký University, Olomouc, 771 46 Olomouc, Czech Republic

Figure S1: The interaction of the side chain of Leu198 with the **1a** compound (a) and the **7a** compound (b).

a b

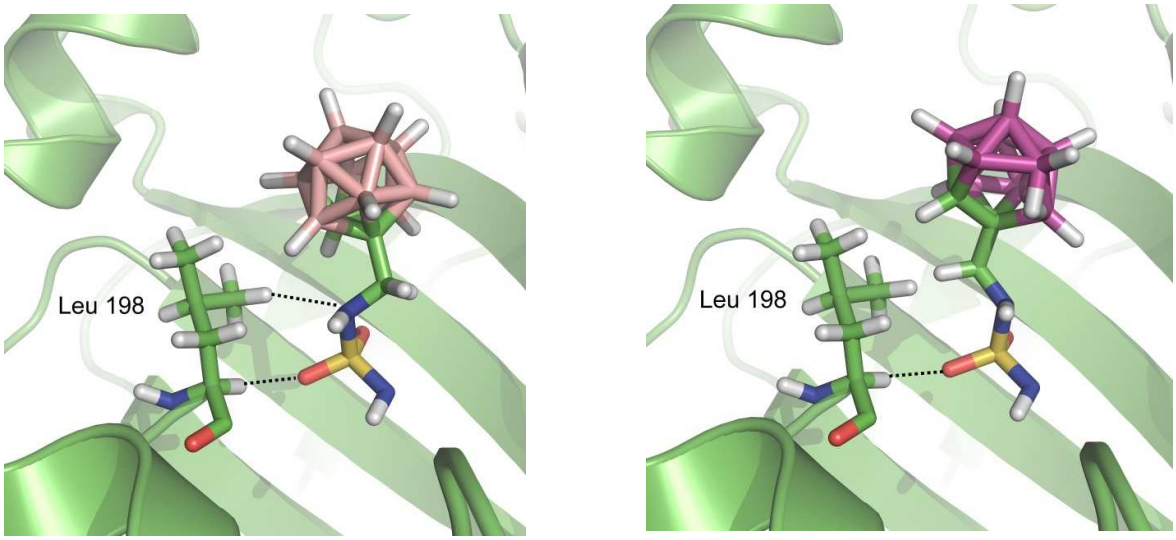
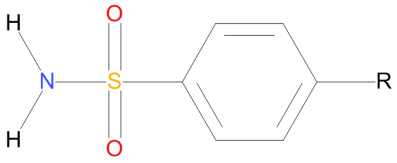
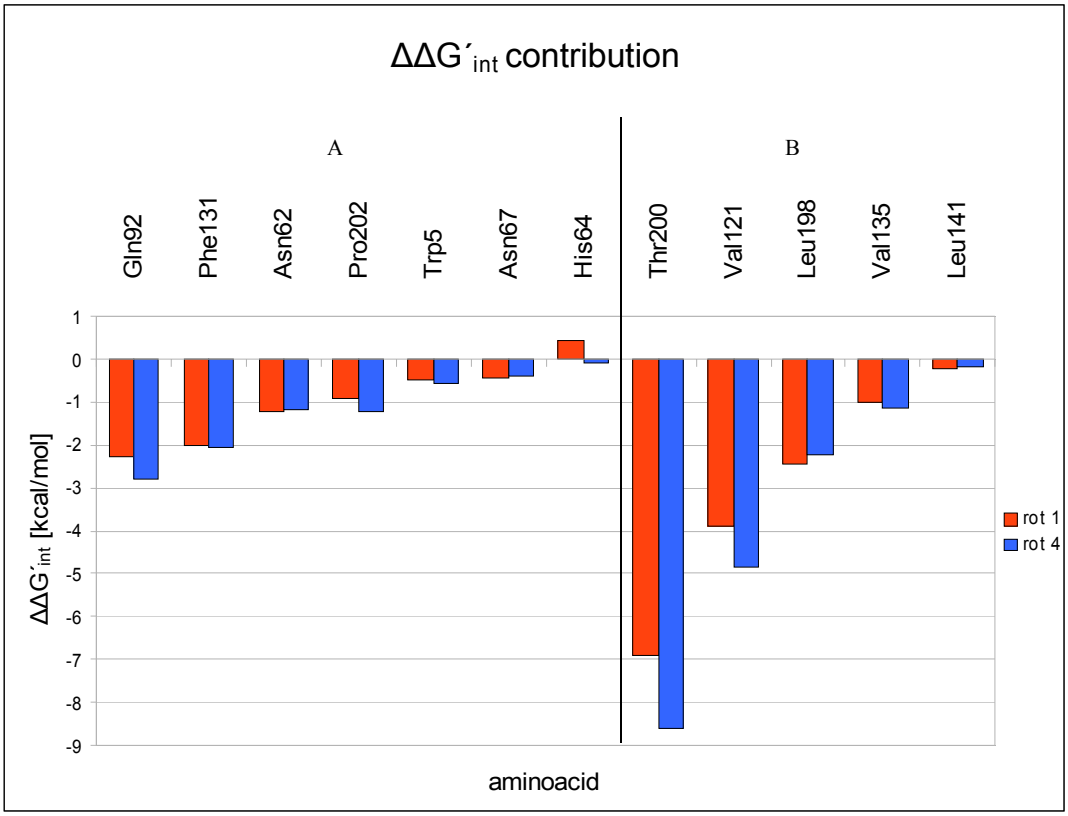


Table S2. The experimental and calculated pKa values. The experimental values taken from Ref.[1]. For calculations, the COSMO-RS methodology[2] was used.

	pKa	
Compound	Exp.	Calc.
R = CH_3	10.2	9.3
R = Cl	9.9	8.6
R = H	10.1	9.1
R = NH_2	10.5	9.9
R = $NHCH_3$	11.0	10.1
R = NO_2	9.0	8.3



Graph S3. 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 and B) the next 5 (from Thr200 to Leu141) have interactions with both, the cage and the sulfamide head, respectively.



References

(1) Krishnamurthy, V. M.; Kaufman, G. K.; Urbach, A. R.; Gitlin, I.; Gudiksen, K. L.; Weibel, D. B.; Whitesides, G. M. Carbonic anhydrase as a model for biophysical and physical-organic studies of proteins and protein-ligand binding. *Chem. Rev.* **2008**, *108*, 946.

(2) Klamt, A. Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena *J. Phys. Chem.* **1995**, *99*, 2224.

Appendix G

Hindawi Publishing Corporation
BioMed Research International
Volume 2014, Article ID 389869, 9 pages
http://dx.doi.org/10.1155/2014/389869

Research Article

Carborane-Based Carbonic Anhydrase Inhibitors: Insight into CAII/CAIX Specificity from a High-Resolution Crystal Structure, Modeling, and Quantum Chemical Calculations

Pavel Mader,[1,2] Adam Pecina,[3] Petr Cígler,[3] Martin Lepšík,[3] Václav Šícha,[4] Pavel Hobza,[3,5] Bohumír Grüner,[4] Jindřich Fanfrlík,[3] Jiří Brynda,[1,3] and Pavlína Řezáčová[1,3]

[1] *Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Vídeňská 1083, 140 00 Prague 4, Czech Republic*
[2] *Structural Genomics Consortium, University of Toronto, Toronto, ON, Canada M5G 1L7*
[3] *Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Gilead Sciences and IOCB Research Center, Flemingovo nam. 2, 166 10 Prague 6, Czech Republic*
[4] *Institute of Inorganic Chemistry, Academy of Sciences of the Czech Republic, v.v.i., Hlavní 1001, 250 68 Řež near Prague, Czech Republic*
[5] *Regional Center of Advanced Technologies and Materials, Department of Physical Chemistry, Palacký University, 77146 Olomouc, Czech Republic*

Correspondence should be addressed to Jindřich Fanfrlík; fanfrlik@uochb.cas.cz, Jiří Brynda; brynda@img.cas.cz, and Pavlína Řezáčová; rezacova@uochb.cas.cz

Received 24 April 2014; Accepted 8 June 2014; Published 18 September 2014

Academic Editor: Mariya Al-Rashida

Carborane-based compounds are promising lead structures for development of inhibitors of carbonic anhydrases (CAs). Here, we report structural and computational analysis applicable to structure-based design of carborane compounds with selectivity toward the cancer-specific CAIX isoenzyme. We determined the crystal structure of CAII in complex with 1-methylenesulfamide-1,2-dicarba-*closo*-dodecaborane at 1.0 Å resolution and used this structure to model the 1-methylenesulfamide-1,2-dicarba-*closo*-dodecaborane interactions with CAIX. A virtual glycine scan revealed the contributions of individual residues to the energy of binding of 1-methylenesulfamide-1,2-dicarba-*closo*-dodecaborane to CAII and CAIX, respectively.

1. Introduction

Carbonic anhydrases (CAs) play important roles in many physiological and pathophysiological processes. For example, extracellular CAs participate in tumor growth and progression [1]. CAIX, which is selectively expressed in a range of hypoxic tumors, is a validated diagnostic and therapeutic target (recently reviewed in [2–4]). There are 15 human CA isoenzymes, and due to the ubiquity of these enzymes in human tissues, selective inhibition is a very important aspect of drug design.

Three main classes of CA inhibitors have been described to date (reviewed in [5]): (i) metal ion binders (sulfonamides, sulfamides, sulfamates, dithiocarbamates, thiols, and hydroxamates); (ii) compounds that anchor the zinc-coordinated water molecule/hydroxide ion (phenols, carboxylates, polyamines, esters, and sulfocoumarins); and (iii) coumarins and related compounds that bind further away from the metal ion.

CA inhibitors from the first class (metal ion binders) contain specific functional groups that interact with the catalytic Zn^{2+} ion in the CA active site. These metal-binding functionalities are typically joined to a "ring" structure. This moiety is not necessarily aromatic; however, it is usually consisting of a 5- or 6-membered hydrocarbon ring or conjugated ring system containing nitrogen, oxygen, and/or sulfur. Numerous functional groups have been added to the ring structure scaffold to modify inhibitor properties such as specificity toward a particular CA isoenzyme, pK_a, or solubility (reviewed in [6]). Recently, we reported design of CA inhibitors containing

space-filling carborane clusters in place of the typical ring structure [7]. We showed that various carborane clusters act as CA inhibitors and that modifying these clusters with an appropriately attached sulfamide group and other substituents leads to compounds with selectivity toward the cancer-specific CAIX isoenzyme.

Boron is one of few chemical elements that can form binary hydrides composed of more than two atoms, which leads to formation of boron cluster compounds (boron hydrides or boranes). Their basic structural feature is formation of a polyhedron with triangular facets held together by 3-center 2-electron bonds with an extensive electron delocalization [8]. A typical structural archetype is represented by the divalent *closo*-$B_{12}H_{12}{}^{2-}$ anion, an extremely stable compound with a symmetrical 12-vertex icosahedron structure [9]. Replacement of one or more $\{BH^-\}$ in borane cage with $\{CH\}$ leads to series of carboranes and removal of BH vertices leads to various open-cage (*nido*-) species. Carboranes thus offer a large variety of structural archetypes that provide interesting counterparts to organic compounds [10].

Many features of icosahedral 12-vertex carboranes are useful in the design of biologically active compounds. Carboranes have high thermal and chemical stability; therefore, they generally do not undergo catabolism and are nontoxic to the host organism [11, 12]. The basic *closo*-$C_2B_{10}H_{12}$ carborane cluster is highly hydrophobic [13]; however, its controlled deboronation can generate water soluble 11-vertex *nido*-$C_2B_9H_{12}{}^-$. These anions represent important intermediates in the synthesis of a family of mainly anionic metal bis(dicarbollides) accessible via metal insertion. Incorporation of carborane cages into the structures of certain substances of medicinal interest can enhance hydrophobic interactions between the boron cluster-coupled pharmaceuticals and their protein targets, increase *in vivo* stability, and facilitate uptake through cellular membranes [14, 15]. The successful use of boron clusters as hydrophobic pharmacophores has recently been increasing [16, 17]. Examples of carborane pharmacophores include boron-containing antifolates [18], HIV protease inhibitors [19, 20], and estrogen receptor agonists and antagonists [21], among others [16, 22, 23].

Drug design efforts benefit greatly from knowledge of the 3D structures of protein-ligand complexes. X-ray crystallography has contributed considerably to the development of CA inhibitors; more than 500 structures of human CA isoenzymes (wild-type and mutant forms) in complex with various inhibitors have offered unprecedented insight into inhibitor binding modes (reviewed in [24]). Structural information coupled with experimental inhibition data can be used to validate various computational approaches to assess inhibitor binding strength. Once a particular theoretical approach reproduces the known data well, it can be used for prospective design. For studies involving metal ions and unusual compounds such as boranes, the use of quantum chemistry (QM) is warranted [25, 26]. Indeed, we recently used a quantum mechanics/molecular mechanics (QM/MM) methodology to quantitatively describe the binding of two carborane-based sulfamides to CAII [7] and to explain fundamental differences in the binding modes of *closo*- and *nido*-cages [27].

Here, we report the X-ray structure of CAII with bound 1-methylenesulfamide-1,2-dicarba-*closo*-dodecaborane (compound **1**, Figure 1(a)) determined at 1.0 Å resolution. This atomic-level resolution allowed us to assess in detail the positions of carbon and boron atoms in the carborane cage of **1**. Additionally, we modeled the complex of **1** with CAIX. We employed a virtual glycine scan to analyze the differences between the interactions of **1** with CAII and CAIX.

2. Materials and Methods

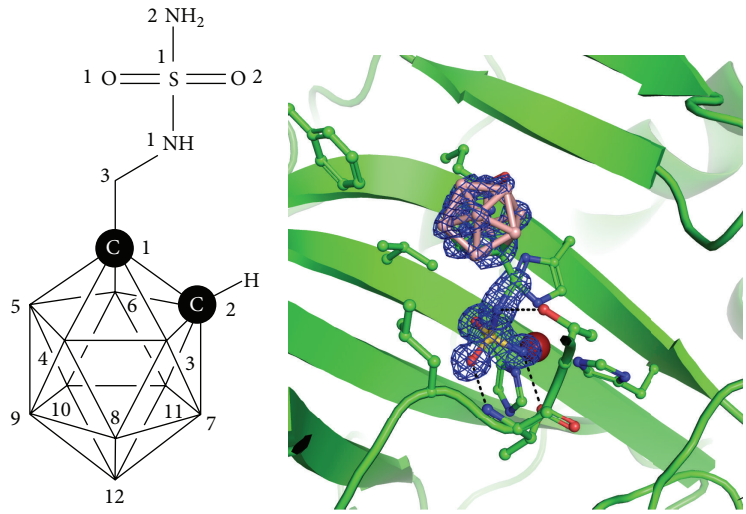
2.1. Protein Crystallization and Diffraction Data Collection. For crystallization of human CAII (Sigma, catalogue number C6165) in complex with 1-methylenesulfamide-1,2-dicarba-*closo*-dodecaborane (compound **1**), we adapted a previously described procedure [28]. CAII (at a concentration of 4 mg·mL^{-1}, dissolved in water) was incubated in aqueous solution containing a 2-fold molar excess of *p*-hydroxymercuribenzoate (Sigma, catalogue number 55540). The protein was concentrated to 10 mg·mL^{-1} and unbound *p*-hydroxymercuribenzoate was removed with Amicon Ultra-4 concentrators (Merck-Millipore MWCO 10 kDa).

The complex of CAII with **1** was prepared by adding a 1.1-fold molar excess of **1** (in DMSO) to the 10 mg·mL^{-1} solution of CAII in water without pH adjustment (the final DMSO concentration did not exceed 5% v/v).

The best diffracting crystals were obtained using the hanging-drop vapor diffusion method under the following conditions: 2 μL protein-inhibitor complex solution was mixed with 2 μL precipitant solution [2.5 M $(NH_4)_2SO_4$, 0.3 M NaCl, and 100 mM Tris-HCl, pH 8.2] and equilibrated over a reservoir containing 1 mL of precipitant solution at 18°C. Crystals with dimensions of 0.3 mm × 0.1 mm × 0.1 mm grew within 7 days.

For cryoprotection, the crystals were incubated in mother liquor supplemented with 25% glycerol for approximately 30 s, flash-frozen, and stored in liquid nitrogen. Diffraction data for the CAII complex were collected at 100 K at the X14.2 BESSY beamline in Berlin, Germany [29]. Data were collected in two passes: the high-resolution range (11.75–1.00 Å) and the low-resolution range (21.08–1.20 Å). The two datasets were integrated with iMOSFLM [30] and merged and scaled with SCALA [31]. Data collection and refinement statistics are summarized in Table 1.

2.2. Structure Determination, Refinement, and Analysis. Crystal structures were solved by difference Fourier method using the CAII structure (PDB code 3IGP [34]) as a starting model. The model was refined using REFMAC5 [35], part of the CCP4 program suite [36]. The model was initially refined with isotropic atomic displacement parameters (ADPs); hydrogen atoms in riding positions were added later. For the final rounds of refinement, we used a mixed isotropic-anisotropic model of ADPs: anisotropic ADPs were used for all atoms, and only atoms in alternative conformations were refined isotropically. Atomic coordinates for the structure of **1** were generated by quantum mechanics computation with DFT-D methodology [37] using the B-LYP functional and SVP basis set [38] in the Turbomole program [39].



(a) (b)

Figure 1: (a) Structural formula of **1** with atom numbers used in the crystal structure coordinate file. The vertices in carborane cluster represent BH groups. (b) Crystal structure of CAII in complex with **1**. The CAII active site is shown in cartoon representation; residues involved in interactions with the Zn^{2+} ion (purple sphere) and **1** are shown in stick representation with carbon atoms colored green. Boron atoms are colored pink, and other heteroatoms are colored according to standard color coding: oxygen, red; nitrogen, blue; sulfur, yellow. The 2*Fo-Fc* electron density map for **1** is contoured at 1 σ.

A geometric library for **1** was generated using the Libcheck program from the CCP4 suite. Coot [40] was used for rebuilding. The quality of the refined model was assessed using MolProbity [33]. The coordinates and structure factors were deposited in the PDB under accession code 4Q78. Final refinement statistics are summarized in Table 1. All structural figures were prepared using PyMOL 1.4.1 [41].

2.3. Model of CAIX-1 Complex. The complex of CAIX and **1** was modeled by aligning the existing crystal structures of the CAIX catalytic domain (PDB code 3IAI [42]) with the CAII-**1** complex (PDB code 4MDG [7]) using PyMOL version 1.2 [43]. Preparation of structure coordinate files for further calculations was performed as described before for CAII [27].

The complex was fully optimized using a QM/MM procedure. We used ONIOM-like subtractive scheme [44] with link atoms and mechanical embedding to be consistent with our previous studies [27, 45–48]. The QM part is described at the DFT-D TPSS/TZVP//BLYP/SVP level of theory [39] and comprises 218 atoms including the atoms present in **1** and 8 amino acids (Trp5, Asn62, His64, Gln67, Gln92, Val131, Leu135, and Pro202). The MM part constituted the remainder of the protein, and the surrounding solvent was approximated by a generalized Born (GB) implicit model. Detailed description of the procedure was published in [27]. One crystal water molecule (Wat272) bridging the inhibitors and CAII residues Thr199, Glu106, and Tyr7 was retained to maintain the integrity of the active site. Other water molecules present in the crystal structures were omitted.

The positions of the added hydrogen atoms, **1**, and 15 amino acids surrounding the ligand (Trp5, Asn62, Gly63, His64, Gln67, Leu91, Gln92, Leu123, Val131, Leu135, Leu141, Thr200, Pro201, Pro202, and Ala204) were relaxed in a GB implicit solvent model using the FIRE algorithm followed by 10 ps annealing from 100 K or 150 K to 0 K using the Berendsen thermostat [49] in the SANDER module of the AMBER 10 package [50].

2.4. Virtual Glycine Scan. The contribution of the active site amino acids to inhibitor binding was examined by virtual glycine scanning. Individual amino acids in contact with **1** in the CAIX-**1** model and CAII-**1** crystal structure were substituted with glycine. The energy contributions ($\Delta\Delta G'_{int}$) were calculated as the difference between the original $\Delta G'_{int}$ at the QM/MM level with the wild-type amino acid and the new $\Delta G'_{int}$ with the mutated glycine residue [27].

3. Results and Discussion

*3.1. Crystal Structure of CAII in Complex with **1** at Atomic Resolution.* The overall structure of CAII in complex with **1** was refined to 1.0 Å resolution. This high resolution allowed us to observe details that could not be fully resolved in the complex structure determined previously at lower resolution. Atomic resolution was achieved by derivatization of CAII using the 4-(hydroxymercury)benzoic acid (abbreviated MBO in the cif library of small molecules) method described by [28]. The mercury atom of MBO covalently binds to Sγ of Cys206. This modification allows formation of a hydrogen bond between the OZ1 oxygen of the MBO carboxyl group and the main-chain amino group of Tyr40 in the neighboring protein molecule, reinforcing the crystal lattice and increasing the diffraction quality of the crystal. In our structure, MBO is modeled in two alternative conformations with occupancies of 0.6 and 0.2.

When our atomic resolution structure is compared with the structure of the CAII-**1** complex determined at 1.35 Å resolution (PDB code 4MDG [7]), the RMSD value for

Table 1: Data collection and refinement statistics.

Data collection statistics	
Space group	$P2_1$
Cell parameters (Å; °)	42.20, 41.73, 72.16; 90.0, 104.4, 90.0
Wavelength (Å)	0.9184
Resolution (Å)	21.08–1.00 (1.05–1.00)
Number of unique reflections	108,781 (15,490)
Multiplicity	3.5 (2.5)
Completeness (%)	83.1 (81.4)
R_{merge} [a]	0.056 (0.375)
Average $I/\sigma(I)$	10.8 (2.3)
Wilson B (Å^2)	6.5
Refinement statistics	
Resolution range (Å)	69.90–1.00 (1.03–1.00)
No. of reflections in working set	97,856 (7,831)
No. of reflections in test set	5,426 (412)
R value (%) [b]	17.5 (24.4)
R_{free} value (%) [c]	20.0 (26.2)
RMSD bond length (Å)	0.011
RMSD angle (°)	1.53
Number of atoms in AU	2297
Number of protein atoms in AU	2081
Number of water molecules in AU	176
Mean ADP value protein/inhibitor (Å^2)	12.0/17.6
Ramachandran plot statistics [d]	
Residues in favored regions (%)	96.56
Residues in allowed regions (%)	3.44

The data in parentheses refer to the highest-resolution shell.
[a] $R_{merge} = \sum_{hkl} \sum_i I_i(hkl) - \langle I(hkl)\rangle \mid / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the individual intensity of the ith observation of reflection hkl and $\langle I(hkl)\rangle$ is the average intensity of reflection hkl with summation over all data.
[b] R value $= ||F_o| - |F_c||/|F_o|$, where F_o and F_c are the observed and calculated structure factors, respectively.
[c] R_{free} is equivalent to R value but is calculated for 5% of reflections chosen at random and omitted from the refinement process [32].
[d] as determined by Molprobity [33].

superposition of the Cα atoms of residues 4–261 is 0.142 Å, a value typical for superposition of identical structures [51]. The N-terminal residue His3 is traced differently in the two structures; double conformations of numerous side chains (e.g., Glu14, His64, and Gln74) are resolved in the atomic resolution structure. We found an additional difference in the loop formed by amino acid residues 124–139, with a maximum difference of 0.738 Å for the position of Gln136 Cα. Gln136 forms van der Waals contacts with the MBO covalently attached to Cys206. The positions of Phe131 and Val135, which form a hydrophobic rim at the active site, are also influenced by MBO binding. This results in a subtle positional shift of the inhibitor, with an RMSD of 0.145 Å for superposition of 12 atoms in the carborane cage of **1** bound to CAII and CAII derivatized by MBO. This value is below the value observed for superposition of identical structures [51].

Atomic-level resolution allowed us to resolve the carbon and boron atom positions in the symmetrical carborane cage of **1**. When analyzing the values of the electron density map at positions of atoms bonded to the C1 atom, we can assume that positions with higher density levels are more likely to be carbon than boron atoms. Similar analysis was done by others for boron-containing inhibitor of human dihydrofolate reductase [18]. The C2 atom of the carborane cage (Figure 1(a)) was modeled into the position with an electron density value of 1.16 e/Å^3, which was approximately 0.15 e/Å^3 higher than those for the B3, B4, B5, and B6 atoms. To exclude the possibility that higher density is caused by model bias, we altered the composition of the cage by replacing the C2 atom with a boron atom. Electron density values did not change significantly after several rounds of refinement cycles.

Thus, we can conclude that the most probable position of the second carbon atom in the carborane cage of **1** is the position assigned to the C2 atom in our crystal structure. This is in good agreement with the recently published QM/MM modeling study [27].

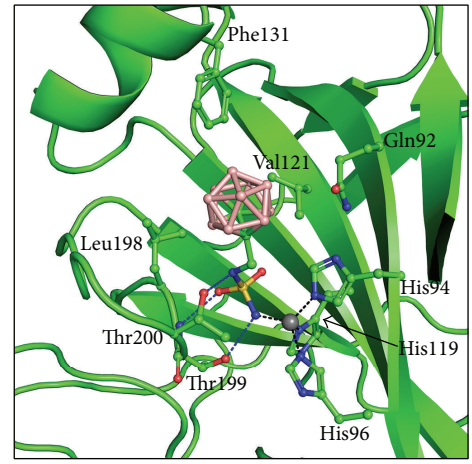
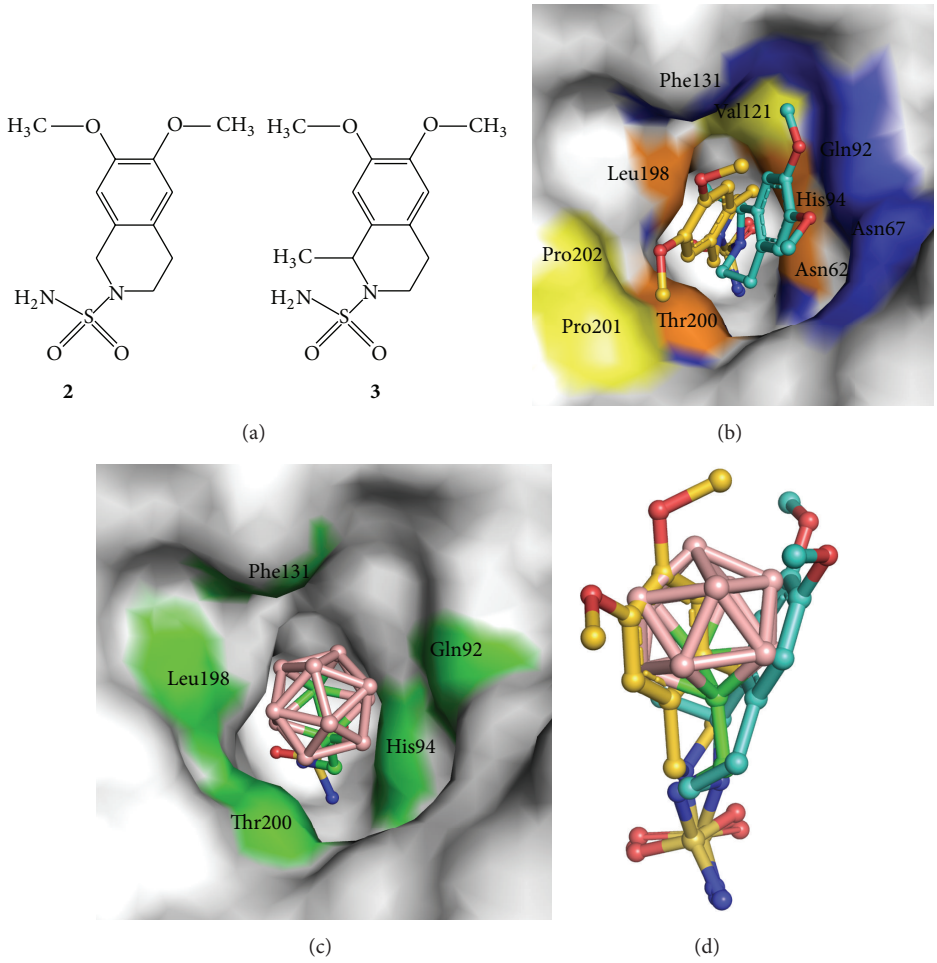


Figure 2: Interactions of **1** with CAII. The protein is shown in cartoon representation; residues involved in interactions with the Zn^{2+} ion (gray sphere) and **1** are shown in stick representation. Polar interactions are represented by blue dashed lines; Zn^{2+} ion coordination is shown as black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

3.2. Detailed Analysis of Inhibitor Interactions with CAII. The crystal structure of human CAII in complex with **1** determined at 1.0 Å resolution confirmed the key interactions that our group observed previously [7]. The compound fits very well into the CAII active site cavity and makes numerous polar and nonpolar interactions with the residues in the enzyme active site. The sulfamide moiety, which forms key polar interactions with the active site Zn^{2+} ion, also makes polar interactions with Thr199 typical of other sulfamide inhibitors of CAII (Figure 2). The linker NH group forms an additional polar interaction with Oγ of Thr200. The compound makes several van der Waals interactions with residues Gln92, His94, His96, His119, Val121, Phe131, Leu198, and Thr200 (Figure 2). All interactions between the inhibitor and protein are summarized in Table 2.



(a) (b) (c) (d)

FIGURE 3: (a) Structural formulas of **2** and **3**. (b) Interactions of **2** and **3** with the CAII active site. Compound **2** is represented with golden carbon atoms, while the carbon atoms of **3** are colored turquoise. Surface of residues making contacts with the isoquinoline moiety of **2** and **3** are highlighted in yellow and blue, respectively. Surface of residues colored orange make contacts with both compounds. Atoms involved in contacts with the sulfonamide groups are not highlighted. (c) Interactions of **1** with the CAII active site. Surface of residues making contacts with the carborane and linker moiety of **1** are highlighted in green. Atoms involved in contacts with the sulfonamide groups are not highlighted. (d) Superposition of binding poses of **1**, **2**, and **3** in the CAII active site. Superposition of the complex structures was based on the best fit for Cα atoms of CAII residues 6–261.

The idea of designing CA inhibitors containing a carborane cluster moiety originated from our previous structural studies of isoquinoline-containing sulfonamide inhibitors (Figure 3(a)). Structural analysis of CAII in complex with 6,7-dimethoxy-1,2,3,4-tetrahydroisoquinolin-2-ylsulfonamide (**2**, PDB code 3IGP, [34]) and 6,7-dimethoxy-1-methyl-1,2,3,4-tetrahydroisoquinolin-2-ylsulfonamide (**3**, PDB code 3PO6, [52]) revealed two distinct binding modes that engage two opposite sides of the enzyme active site cavity (Figure 3(b)). Following this analysis, we hypothesized that the binding space within the enzyme active site cavity could be effectively filled with a bulky hydrophobic molecule with a spherical structure. This led to design of **1** which exhibited inhibitory property to CAII and CAIX with Ki values in submicromolar range. Structural analysis of CAII-**1** indicates that our structure-based design was sound. We found that the carborane cluster interacts with both sides of the enzyme active site as predicted (Figure 3(c), Table 3) and that the position of **1** in the CAII active site superposes well with the two binding modes observed for 2 and 3 (Figure 3(d)).

3.3. Model of the CAIX-1 Complex. The CAII-**1** crystal structure was used to model binding of compound **1** into the CAIX active site using QM/MM methods (Figure 4).

The substrate binding sites of CAII and CAIX differ by only six amino acids: 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. These variations result in a differently shaped active site cavity, which accommodated **1** in a slightly different pose (Figure 4). While the position of the sulfamide anchor remained unchanged, the carborane cluster shifted by 2.1 Å (expressed as a difference in the position of B12) away from the central β-sheet. In CAIX-**1**, the carborane interacts more with the opposite site of the active site, specifically with amino acid residues His94, His96, Glu106, Leu198, Thr199, Thr200, and Pro201 (Figure 4, Table 3). All polar and van der

Table 2: List of contacts between CAII and **1**.

CAII			**1**	
Residue		Atom	Atom[a]	Distance [Å][b]
	Zn	**ZN**	**N2**	**1.87**[c]
	Zn	ZN	S	3.04
	Zn	ZN	O2	3.05
92	Gln	OE1	B6	3.47
92	Gln	OE1	B11	3.52
92	Gln	CD	B6	3.84
94	His	CE1	O2	2.97
94	His	NE2	N2	3.23
94	His	NE2	O2	3.31
94	His	CE1	C3	3.67
94	His	NE2	S	3.81
94	His	CE1	N2	3.82
94	His	CE1	S	3.84
94	His	NE2	C3	3.94
96	His	NE2	N2	3.14
96	His	CE1	N2	3.56
119	His	ND1	N2	3.39
119	His	ND1	O2	3.88
119	His	CE1	N2	3.96
121	Val	CG2	O2	3.82
131	Phe	CZ	B8	3.83
131	Phe	CZ	B7	3.97
198	Leu	CA	O1	3.09
198	Leu	C	O1	3.36
198	Leu	CB	O1	3.60
198	Leu	CD2	O1	3.63
198	Leu	CD1	B3	3.86
199	**Thr**	**N**	**O1**	**2.70**
199	**Thr**	**OG1**	**N2**	**2.74**
199	Thr	OG1	O1	3.58
199	Thr	OG1	S	3.78
199	Thr	N	S	3.83
199	Thr	CA	O1	3.83
199	Thr	CB	N2	3.98
200	**Thr**	**OG1**	**N1**	**3.02**
200	Thr	OG1	C3	3.14
200	Thr	OG1	B4	3.36
200	Thr	OG1	B3	3.56
200	Thr	OG1	C1	3.66

[a]Atom labels correspond to those shown in Figure 1(a).
[b]All contacts with a distance between ligand and protein (or Zn) atoms less than or equal to 4 Å are listed.
[c]Polar interactions are highlighted in bold.

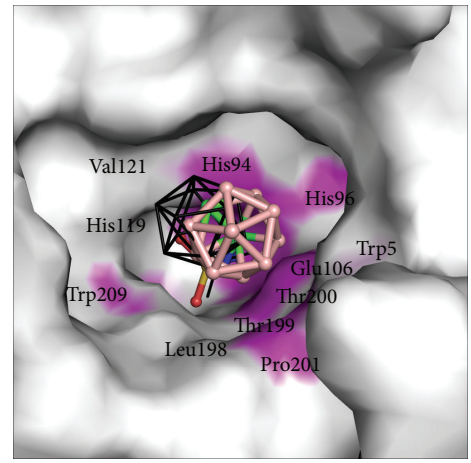


Figure 4: Interactions of **1** with the CAIX active site. Atoms making contacts with the carborane and linker moiety of **1** are highlighted in magenta. Atoms involved in contacts with the sulfonamide groups are not highlighted. Superposition of the binding pose of **1** in CAII is shown as black lines. Superposition is based on the best fit for Cα atoms of all residues of CAII onto CAIX.

Table 3: CAII or CAIX residues interacting with **1**, **2**, and **3**.

	CAII		CAIX
1[a]	**2**[b]	**3**[c]	**1**[d]
			Trp5
		Asn62	
		Asn67[e]	
Gln92	Gln92	Gln92	
His94	His94	His94	His94
His96	His96	His96	His96
			Glu106
His119	His119	His119	His119
Val121		Val121	Val121
Phe131	Phe131	Phe131	
	Val143	Val143	
Leu198	Leu198	Leu198	Leu198
Thr199	**Thr199**	**Thr199**	**Thr199**
Thr200	Thr200	Thr200	**Thr200**
	Pro201		Pro201
	Pro202		
	Trp209		Trp209

Interacting residues were identified from [a]crystal structure 4Q78 (this work); [b]crystal structure 3IGP [34]; [c]crystal structure 3PO6 [52]; [d]computational model (this work); [e]residues making polar interactions are highlighted in bold.

Waals interactions between CAIX and **1** are summarized in Table 4.

We used a virtual glycine scan to study the roles of individual amino acid side chains in the active sites of CAII and CAIX in binding of **1**. The changes in free energy of interaction ($\Delta\Delta G'_{int}$) upon mutation of a given amino acid residue to glycine are shown in Figure 5.

The largest energy change (2.6 kcal/mol) occurred for Trp5, which is positioned closer to **1** in CAIX-**1** than in CAII-**1**. The side chain of Trp5 forms several dihydrogen bonds with the carborane cage of **1**. The shortest one has a H⋯H distance of 2.3 Å. The other major contributor to strong

Table 4: Interactions between CAIX and **1**.

CAIX Residue		Atom	**1** Atom[a]	Distance [Å][b]
	Zn	**ZN**	**N2**	**2.1**[c]
	Zn	ZN	S	3.3
	Zn	ZN	O2	3.5
5	Trp	CZ2	B5	3.74
5	Trp	CZ2	B10	3.81
94	His	CE1	O2	3.15
94	His	CE1	C3	3.74
94	His	NE2	N2	3.36
94	His	NE2	S	3.88
94	His	NE2	O2	3.45
94	His	NE2	C3	3.76
96	His	CE1	N2	3.99
96	His	NE2	N2	3.49
106	Glu	OE2	N2	3.71
119	His	ND1	N2	3.37
119	His	CE1	N2	3.83
121	Val	CG2	O2	3.58
198	Leu	CA	O1	3.04
198	Leu	CB	O1	3.4
198	Leu	CD2	O1	3.43
198	Leu	C	O1	3.38
199	Thr	N	S	3.88
199	**Thr**	**N**	**O1**	**2.79**
199	Thr	CA	O1	3.96
199	Thr	CB	N2	3.85
199	**Thr**	**OG1**	**N2**	**2.63**
199	Thr	OG1	S	3.69
199	Thr	OG1	O1	3.65
200	Thr	OG1	C1	3.77
200	Thr	OG1	B5	3.56
200	**Thr**	**OG1**	**N1**	**3.13**
200	Thr	OG1	C3	3.31
200	Thr	OG1	B4	3.64
201	Pro	O	B4	3.6
201	Pro	O	B10	3.49
201	Pro	O	B8	3.96
209	Trp	CZ2	O1	3.74

[a]Atom labels correspond to those shown in Figure 1(a).
[b]All contacts with a distance less than or equal to 4 Å between ligand and protein (and Zn) atoms are listed.
[c]Polar interactions are highlighted in bold.

CAIX-**1** binding was Asn62; the energy of binding exceeded that in CAII-**1** by nearly 1 kcal/mol. These contributions were cancelled out by differences in binding energy contributions of amino acid residues 131 (Phe/Val) and 135 (Val/Leu), which were lower in CAIX by 0.7 and 0.9 kcal/mol, respectively. The energy changes of other residues were small.

When we compared binding of **1** to CAII and CAIX, we noted that the favorable energy changes in CAIX-**1** due to the binding of residues Trp5, Asn62, and His64 were slightly larger than the unfavorable changes in binding caused by the different amino acids at residues 131 and 135. This is in qualitative agreement with the experimental K_i values, which are 700 ± 141 nm for inhibition of CAII and 380 ± 111 nM for inhibition of CAIX [7].

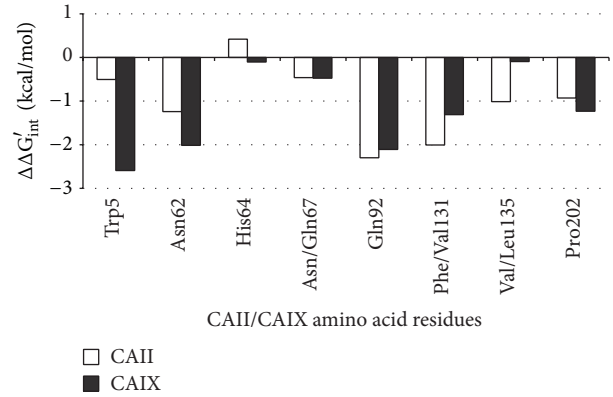


Figure 5: Results of virtual glycine scan showing contributions of individual residues to the energy of binding of **1** to CAII and CAIX, respectively.

4. Conclusions

We determined to atomic resolution the crystal structure of CAII in complex with 1-methylenesulfamide-1,2-dicarba-*closo*-dodecaborane (**1**), a parent compound of a recently reported series of CA inhibitors containing carborane cages [7]. Comparing this crystal structure with those of CAII complexes with conventional organic inhibitors showed that the three-dimensional cluster fills the enzyme active site cavity. Atomic-level resolution allowed us to distinguish the positions of carbon and boron atoms in the carborane cage. The crystal structure also served as a model for construction of the CAIX-**1** computational model. Virtual glycine scan enabled us to quantify the contributions of individual residues to the energy of binding of **1** to CAII and CAIX and uncover differences of the enzyme active site cavities. Structural and computational analysis will be used in future structure-based design of carborane compounds with selectivity toward the cancer-specific CAIX isoenzyme.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Authors' Contribution

Pavel Mader and Adam Pecina contributed equally to this work.

Acknowledgments

This work was supported by the Technology Agency of the Czech Republic (Project TE01020028) and in part by Research Projects RVO 68378050 and 61388963 awarded by

the Academy of Sciences of the Czech Republic. The authors acknowledge the financial support of the Czech Science Foundation (P208/12/G016) and the operational program Research and Development for Innovations of the European Social Fund (CZ 1.05/2.1.00/03/0058). They also thank Gilead Sciences and IOCB Research Centre for financial support. The use of MX14.2 operated by the Helmholtz-Zentrum Berlin at the BESSY II electron storage ring (Berlin-Adlershof, Germany) to collect diffraction data is acknowledged.

References

[1] P. Swietach, S. Patiar, C. T. Supuran, A. L. Harris, and R. D. Vaughan-Jones, "The role of carbonic anhydrase 9 in regulating extracellular and intracellular pH in three-dimensional tumor cell growths," *The Journal of Biological Chemistry*, vol. 284, no. 30, pp. 20299–20310, 2009.

[2] P. C. McDonald, J. Winum, C. T. Supuran, and S. Dedhar, "Recent developments in targeting carbonic anhydrase IX for cancer therapeutics," *Oncotarget*, vol. 3, no. 1, pp. 84–97, 2012.

[3] F. E. Lock, P. C. McDonald, Y. Lou et al., "Targeting carbonic anhydrase IX depletes breast cancer stem cells within the hypoxic niche," *Oncogene*, vol. 32, no. 44, pp. 5210–5219, 2013.

[4] N. K. Tafreshi, M. C. Lloyd, M. M. Bui, R. J. Gillies, and D. L. Morse, "Carbonic anhydrase IX as an imaging and therapeutic target for tumors and metastases," in *Carbonic Anhydrase: Mechanism, Regulation, Links to Disease, and Industrial Applications*, vol. 75 of *Subcellular Biochemistry*, pp. 221–254, 2014.

[5] S. M. Monti, C. T. Supuran, and G. de Simone, "Anticancer carbonic anhydrase inhibitors: a patent review (2008–2013)," *Expert Opinion on Therapeutic Patents*, vol. 23, no. 6, pp. 737–749, 2013.

[6] R. McKenna and C. T. Supuran, "Carbonic anhydrase inhibitors drug design," *Subcellular Biochemistry*, vol. 75, pp. 291–323, 2014.

[7] J. Brynda, P. Mader, V. Sicha et al., "Carborane-based carbonic anhydrase inhibitors," *Angewandte Chemie International Edition in English*, vol. 52, pp. 13760–13763, 2013.

[8] W. N. Lipscomb, *Boron Hydrides*, W.A. Benjamin, New York, NY, USA, 1963.

[9] P. R. Schleyer and K. Najafian, "Stability and three-dimensional aromaticity of closo-monocarbaborane anions, $Cb_{n-1}H_n^-$, and closo-dicarboranes, $C_2B_{n-2}H_n$," *Inorganic Chemistry*, vol. 37, pp. 3454–3470, 1998.

[10] R. E. Williams, in *The Borane , Carborane and Carbocation Continuum*, J. Casanova, Ed., pp. 3–57, John Wiley & Sons, New York, NY, USA, 1997.

[11] J. F. Valliant, K. J. Guenther, A. S. King et al., "The medicinal chemistry of carboranes," *Coordination Chemistry Reviews*, vol. 232, no. 1-2, pp. 173–230, 2002.

[12] Z. J. Lesnikowski, "Boron units as pharmacophores—new applications and opportunities of boron cluster chemistry," *Collection of Czechoslovak Chemical Communications*, vol. 72, no. 12, pp. 1646–1658, 2007.

[13] J. Plešek, "Potential applications of the boron cluster compounds," *Chemical Reviews*, vol. 92, no. 2, pp. 269–278, 1992.

[14] J. Seidler, S. L. McGovern, T. N. Doman, and B. K. Shoichet, "Identification and prediction of promiscuous aggregating inhibitors among known drugs," *Journal of Medicinal Chemistry*, vol. 46, no. 21, pp. 4477–4486, 2003.

[15] M. Sibrian-Vazquez, E. Hao, T. J. Jensen, and M. G. H. Vicente, "Enhanced cellular uptake with a cobaltacarborane-porphyrin-HIV-1 Tat 48-60 conjugate," *Bioconjugate Chemistry*, vol. 17, no. 4, pp. 928–934, 2006.

[16] M. Scholz, M. Steinhagen, J. T. Heiker, A. G. Beck-Sickinger, and E. Hey-Hawkins, "Asborin Inhibits Aldo/Keto Reductase 1A1," *ChemMedChem*, vol. 6, no. 1, pp. 89–93, 2011.

[17] F. Issa, M. Kassiou, and L. M. Rendina, "Boron in drug discovery: carboranes as unique pharmacophores in biologically active compounds," *Chemical Reviews*, vol. 111, no. 9, pp. 5701–5722, 2011.

[18] R. C. Reynolds, S. R. Campbell, R. G. Fairchild et al., "Novel boron-containing, nonclassical antifolates: synthesis and preliminary biological and structural evaluation," *Journal of Medicinal Chemistry*, vol. 50, no. 14, pp. 3283–3289, 2007.

[19] P. Cigler, M. Kozisek, P. Rezacova et al., "From nonpeptide toward noncarbon protease inhibitors: metallacarboranes as specific and potent inhibitors of HIV protease," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, pp. 15394–15399, 2005.

[20] P. Řezáčová, J. Pokorná, J. Brynda et al., "Design of HIV protease inhibitors based on inorganic polyhedral metallacarboranes," *Journal of Medicinal Chemistry*, vol. 52, no. 22, pp. 7132–7141, 2009.

[21] Y. Endo, T. Iijima, Y. Yamakoshi et al., "Potent estrogen agonists based on carborane as a hydrophobic skeletal structure: a new medicinal application of boron clusters," *Chemistry & Biology*, vol. 8, no. 4, pp. 341–355, 2001.

[22] R. L. Julius, O. K. Farha, J. Chiang, L. J. Perry, and M. F. Hawthorne, "Synthesis and evaluation of transthyretin amyloidosis inhibitors containing carborane pharmacophores," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 12, pp. 4808–4813, 2007.

[23] S. Fujii, H. Masuno, Y. Taoda et al., "Boron cluster-based development of potent nonsecosteroidal vitamin D receptor ligands: direct observation of hydrophobic interaction between protein surface and carborane," *Journal of the American Chemical Society*, vol. 133, no. 51, pp. 20933–20941, 2011.

[24] V. M. Krishnamurthy, G. K. Kaufman, A. R. Urbach et al., "Carbonic anhydrase as a model for biophysical and physical-organic studies of proteins and protein-ligand binding," *Chemical Reviews*, vol. 108, no. 3, pp. 946–1051, 2008.

[25] K. Raha, M. B. Peters, B. Wang et al., "The role of quantum mechanics in structure-based drug design," *Drug Discovery Today*, vol. 12, no. 17-18, pp. 725–731, 2007.

[26] M. Lepšík, J. Řezáč, M. Kolář, A. Pecina, P. Hobza, and J. Fanfrlík, "The semiempirical quantum mechanical scoring function for in silico drug design," *ChemPlusChem*, vol. 78, pp. 921–931, 2013.

[27] A. Pecina, M. Lepsik, J. Rezac et al., "QM/MM calculations reveal the different nature of the interaction of two carborane-based sulfamide inhibitors of human carbonic anhydrase II," *The Journal of Physical Chemistry B*, vol. 117, pp. 16096–16104, 2013.

[28] C. A. Behnke, I. Le Trong, J. W. Godden et al., "Atomic resolution studies of carbonic anhydrase II," *Acta Crystallographica D: Biological Crystallography*, vol. 66, no. 5, pp. 616–627, 2010.

[29] U. Mueller, N. Darowski, M. R. Fuchs et al., "Facilities for macromolecular crystallography at the Helmholtz-Zentrum Berlin," *Journal of Synchrotron Radiation*, vol. 19, no. 3, pp. 442–449, 2012.

[30] T. G. G. Battye, L. Kontogiannis, O. Johnson, H. R. Powell, and A. G. Leslie, "iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM," *Acta Crystallographica D*, vol. 67, no. 4, pp. 271–281, 2011.

[31] P. Evans, "Scaling and assessment of data quality," *Acta Crystallographica Section D: Biological Crystallography*, vol. 62, no. 1, pp. 72–82, 2006.

[32] A. T. Brunger, "Free *R* value: a novel statistical quantity for assessing the accuracy of crystal strucutres," *Nature*, vol. 355, no. 6359, pp. 472–475, 1992.

[33] S. C. Lovell, I. W. Davis, W. B. Arendall III et al., "Structure validation by Cα geometry: φ,ψ and Cβ deviation," *Proteins*, vol. 50, no. 3, pp. 437–450, 2003.

[34] R. Gitto, S. Agnello, S. Ferro et al., "Identification of 3,4-dihydroisoquinoline-2(1H)-sulfonamides as potent carbonic anhydrase inhibitors: synthesis, biological evaluation, and enzyme-ligand X-ray studies," *Journal of Medicinal Chemistry*, vol. 53, no. 6, pp. 2401–2408, 2010.

[35] G. N. Murshudov, A. A. Vagin, and E. J. Dodson, "Refinement of macromolecular structures by the maximum-likelihood method," *Acta Crystallographica D*, vol. 53, no. 3, pp. 240–255, 1997.

[36] "The CCP4 suite: programs for protein crystallography," *Acta Crystallographica D*, vol. 50, pp. 760–763, 1994.

[37] P. Jurečka, J. Černý, P. Hobza, and D. R. Salahub, "Density functional theory augmented with an empirical dispersion term. Interaction energies and geometries of 80 noncovalent complexes compared with ab initio quantum mechanics calculations," *Journal of Computational Chemistry*, vol. 28, no. 2, pp. 555–569, 2007.

[38] F. Weigend and R. Ahlrichs, "Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: design and assessment of accuracy," *Physical Chemistry Chemical Physics*, vol. 7, no. 18, pp. 3297–3305, 2005.

[39] R. Ahlrichs, M. Bär, M. Häser, H. Horn, and C. Kölmel, "Electronic structure calculations on workstation computers: the program system turbomole," *Chemical Physics Letters*, vol. 162, no. 3, pp. 165–169, 1989.

[40] P. Emsley and K. Cowtan, "Coot: model-building tools for molecular graphics," *Acta Crystallographica D: Biological Crystallography*, vol. 60, no. 12, pp. 2126–2132, 2004.

[41] W. L. DeLano, *The PyMOL Molecular Graphics System*, DeLano Scientific LLC, San Carlos, Calif, USA, 2002, http://www.pymol.org.

[42] V. Alterio, M. Hilvo, A. Di Fiore et al., "Crystal structure of the catalytic domain of the tumor-associated human carbonic anhydrase IX," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 38, pp. 16233–16238, 2009.

[43] W. L. DeLano, *The PymoL Molecular Graphics System*, DeLano Scientific, Palo Alto, Calif, USA, 2002.

[44] M. Svensson, S. Humbel, R. D. J. Froese, T. Matsubara, S. Sieber, and K. Morokuma, "ONIOM: a multilayered integrated MO + MM method for geometry optimizations and single point energy predictions. A test for Diels-Alder reactions and Pt(P(t-Bu)$_3$)$_2$ + H$_2$ oxidative addition," *The Journal of Physical Chemistry*, vol. 100, no. 50, pp. 19357–19363, 1996.

[45] A. Pecina, O. Přenosil, J. Fanfrlík et al., "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*, vol. 76, no. 5, pp. 457–479, 2011.

[46] P. S. Brahmkshatriya, P. Dobeš, J. Fanfrlík et al., "Quantum mechanical scoring: structural and energetic insights into cyclin-dependent kinase 2 inhibition by pyrazolo[1,5-a]pyrimidines," *Current Computer—Aided Drug Design*, vol. 9, no. 1, pp. 118–129, 2013.

[47] J. Fanfrlík, M. Kolář, M. Kamlar et al., "Modulation of aldose reductase inhibition by halogen bond tuning," *ACS Chemical Biology*, vol. 8, pp. 2484–2492, 2013.

[48] J. Fanfrlík, P. S. Brahmkshatriya, J. Řezáč et al., "Quantum mechanics-based scoring rationalizes the irreversible inactivation of parasitic *Schistosoma mansoni* cysteine peptidase by vinyl sulfone inhibitors," *Journal of Physical Chemistry B*, vol. 117, pp. 14973–14982, 2013.

[49] H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, A. Dinola, and J. R. Haak, "Molecular dynamics with coupling to an external bath," *The Journal of Chemical Physics*, vol. 81, no. 8, pp. 3684–3690, 1984.

[50] D. A. Case, T. E. Cheatham III, T. Darden et al., "The Amber biomolecular simulation programs," *Journal of Computational Chemistry*, vol. 26, no. 16, pp. 1668–1688, 2005.

[51] M. J. Betts and M. J. E. Sternberg, "An analysis of conformational changes on protein-protein association: implications for predictive docking," *Protein Engineering*, vol. 12, no. 4, pp. 271–283, 1999.

[52] P. Mader, J. Brynda, R. Gitto et al., "Structural Basis for the Interaction between Carbonic Anhydrase and 1,2,3,4-tetrahydroisoquinolin-2-ylsulfonamides," *Journal of Medicinal Chemistry*, vol. 54, no. 7, pp. 2522–2526, 2011.

Appendix H

Journal Name

COMMUNICATION

The SQM/COSMO Filter: Reliable Native Pose Identification Based on the Quantum-Mechanical Description of Protein–Ligand Interactions and Implicit COSMO Solvation

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Adam Pecina, †[a] René Meier, †[b] Jindřich Fanfrlík, [a] Martin Lepšík, [a] Jan Řezáč, [a] Pavel Hobza *[a,c] and Carsten Baldauf *[d]

Current virtual screening tools are fast, but reliable scoring is elusive. Here, we present the 'SQM/COSMO filter', a novel scoring function featuring quantitative semiempirical quantum mechanical (SQM) description of all types of noncovalent interactions coupled with implicit COSMO solvation. We show unequivocally that it outperforms eight widely used scoring functions. The accuracy and chemical generality of the SQM/COSMO filter make it a perfect tool for the late stages of virtual screening.

Despite the enormous advances in method development for structure-based *in silico* drug design, reliable predictions of the structures (docking) and affinities (scoring) of protein–ligand (P–L) complexes still remain an unsolved task.[1] A plethora of scoring functions (SFs) have been devised by utilising experimental data for regression analyses, by constructing knowledge-based potentials, or based on physical laws.[2-3] As none of the SFs is general enough to perform equally strongly for a diverse set of P–L complexes, utilising several SFs at once (consensus scoring) holds promise.[4] Regression-analysis and knowledge-based approaches to scoring are trained on a set of P–L complexes and rely on variable master equation terms. Their validity is limited to complexes similar to the training set. In principle, this problem has been overcome in physics-based methods. Because of computational cost, preference has been given to molecular mechanics (MM) methods, such as the combination of MM interaction energies with implicit solvation free energy terms (generalised Born, GB, or Poisson-Boltzmann, PB) to estimate affinities.[2] Additionally, the wide coverage of organic chemical space in the GAFF (general AMBER force field)[5] has made the parameterisation of ligands for MM straightforward. However, an explicit description of quantum mechanical (QM) effects in P–L interactions, such as charge transfer, polarisation, covalent-bond formation or σ-hole bonding, was missing. QM methods, which describe these effects qualitatively better than the energy functions used in MM-based SFs, were thus introduced into computational drug design.[6,7] Recent developments in QM methods and algorithms as well as the availability of a powerful computing infrastructure have paved the way to apply them for P–L complexes in numerous setups: linear scaling or efficient parallelisation of semi-empirical QM (SQM) methods,[7-10] QM/MM,[7,8,11,12] DFT-D3 on truncated P–L complexes[13] or various fragmentation methods.[11,14] Specifically, AM1, RM1, PM6 or DF-TB SQM methods have been used[7-9,12,15] as such or with empirical corrections for dispersion, hydrogen- and halogen-bonding[16] to describe the P-L noncovalent interactions. Merz et al. pioneered this area by introducing a QM-based SF (QMScore), a combination of the AM1 SQM method with an empirical dispersion (D) and the PB implicit solvent [Eq. 1].[17] The method was useful for describing metalloprotein–ligand binding, but further corrections were needed, especially for a quantitative treatment of dispersion and hydrogen bonding.[10]

$$Score = \Delta E_{\text{int}} + \Delta\Delta G_{solv} + \Delta G'^{w}_{conf} - T\Delta S \qquad (Eq.1)$$

Equation 1. A general physics-based SF. The terms are: 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 ($\Delta G'^{w}_{conf}$) and the change of entropy upon ligand binding ($-T\Delta S$).

Our approach is systematic. Using accurate calculations in small model systems as a benchmark, we developed corrections for SQM methods that provide reliable and accurate description of a wide range of noncovalent

[a] Institute of Organic Chemistry and Biochemistry (IOCB) and Gilead Sciences and IOCB Research Center, Flemingovo nám. 2, 16610 Prague 6 (Czech Republic)
[b] Institut für Biochemie, Fakultät für Biowissenschaften, Pharmazie und Psychologie, Universität Leipzig, Brüderstrasse 34, D-04109 Leipzig (Germany)
[c] Regional Centre of Advanced Technologies and Materials, Department of Physical Chemistry, Palacký University, 77146 Olomouc (Czech Republic)
[d] Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4-6, D-14195 Berlin (Germany)
† These authors have contributed equally to this work.
* Corresponding authors: hobza@uochb.cas.cz, baldauf@fhi-berlin.mpg.de
Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

interactions including dispersion, hydrogen and halogen bonding.[16] Coupled with the PM6 SQM method[18], the resulting PM6-D3H4X approach is applicable to wide chemical space and does not require any system-specific parameterisation. We use it here to calculate the ΔE_{int} term. Subsequently, we compared MM-based (PB or GB) and QM-based (COSMO[19] or SMD) implicit solvent models and found the latter group to be more accurate.[20] These are therefore used for the $\Delta\Delta G_{solv}$ term. These two dominant terms, ΔE_{int} and $\Delta\Delta G_{solv}$, are at the heart of our SQM-based SF.[15] We have demonstrated its generality in various noncovalent P–L complexes, such as aldose reductase or carbonic anhydrase and moreover extended it to treat covalent inhibitor binding (Refs. 15, 21, 22).

In this work, we adapt our SQM-based SF to make it usable in virtual screening on the basis of our previous experience. By taking the two dominant terms only, ΔE_{int} and $\Delta\Delta G_{solv}$, we define the 'SQM/COSMO filter' energy. Its performance is tested here against eight widely used SFs. GlideScore XP (GlideXP)[23], PLANTS PLP (PLP)[24], AutoDock Vina (Vina)[25], Chemscore (CS)[26], Goldscore (GS)[27] and ChemPLP[24] are empirical, regression-based functions which use different terms to describe vdW contacts, lipophilic surface coverage, hydrogen bonding, ligand strain, and desolvation. The Astex Statistical Potential (ASP)[28] is a knowledge-based potential. The classical physics-based AMBER/GB SF combines the ff03-GAFF MM force fields with GB implicit solvent.[5,29]

The goal is 'cognate docking'[30], i.e. the ability to identify sharply the known native X-ray P–L binding pose from a set of decoy structures generated by docking (Figure 1). To understand our results in detail, we have not opted for treating them in a statistical manner[31] as in the pose decoy test sets available.[32] Instead we cautiously selected four unrelated difficult-to-handle P–L systems, which comply with strict criteria for the selection of crystallographic structures for docking (details in SI).[33] These systems are: acetylcholine esterase (AChE, PDB: 1E66)[34], TNF-α converting enzyme (TACE, PDB: 3B92)[35], aldose reductase (AR, PDB: 2IKJ)[36] and HIV-1 protease (HIV PR; PDB: 1NH0)[37]. For the latter, the protonation of the active site is inferred from ultra-high resolution X-ray crystallography. Based on these P–L crystal structures, we have created a set of non-redundant poses (2,865 in total) by docking with four popular docking programs (Glide, PLANTS, AutoDock Vina and GOLD) coupled to seven widely used SFs[23-28] (Figure 1, Table S2).

All the poses were re-scored by all nine SFs. For the seven regression- and knowledge-based SFs, we used the recommended protocols. For the two physics-based SFs, only hydrogens and close contacts were relaxed by the AMBER/GB method. RMSD of the poses relative to the crystal were measured (details in S1.6). The scores were normalised and are shown relative to the score of the crystal pose.

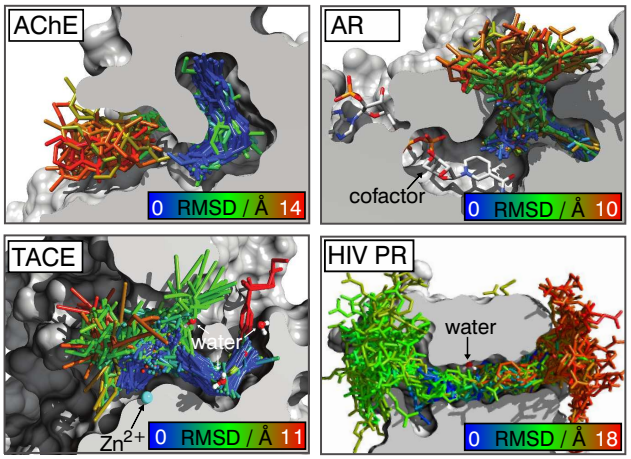


Figure 1. The ligand poses generated by the four docking programs. Ligand poses are color-coded by RMSD.

The identification of the X-ray pose as the minimum-free-energy structure is an unambiguous criterion for the performance of any SF. The ideal behaviour of such a score *vs.* the RMSD curve (Figure 2) is characterised by the positive values of energies for decoy poses. Small deviations (negative energies for very small RMSD values) are acceptable and might be explained by inaccuracies of the crystal structure. This condition is met by the SQM/COSMO filter, unlike the other SFs (Figure 2). The numbers of false-positive solutions as well as the maximum RMSD ($RMSD^{max}$) from the X-ray pose within a defined interval of the normalised score quantify the virtually ideal behaviour of the SQM/COSMO filter in comparison to the other SF.

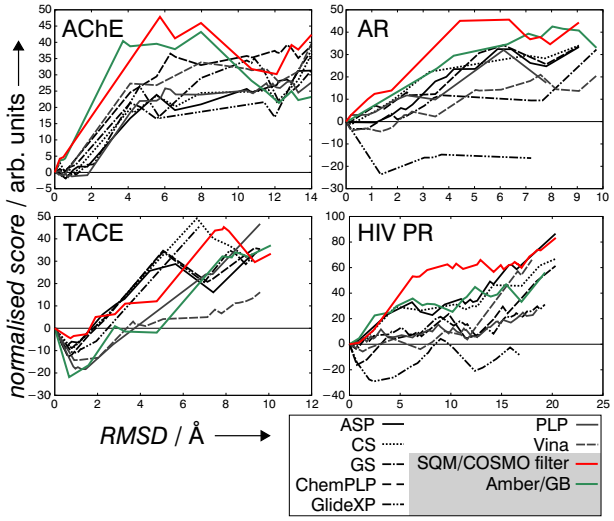


Figure 2. The plots of normalised scores against RMSD values for all four P–L systems.

Journal Name

COMMUNICATION

Table 1: The numbers of false-positive solutions, i.e. solutions that are scored better than the X-ray pose and have RMSD > 0.5 Å.

	Scoring function								
			Glide	PLANTS	AutoDock	Gold			
	SQM/COSMO	AMBER/GB	XP	PLP	Vina	ASP	CS	GS	ChemPLP
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 2: The maximum RMSD [Å] within all the poses in the defined range of the relative normalised score

	Scoring function								
			Glide	PLANTS	AutoDock	Gold			
	SQM/COSMO	AMBER/GB	XP	PLP	Vina	ASP	CS	GS	ChemPLP
	Maximal RMSD within a window of 5 of the normalised 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**

The number of false positives is lowest for the SQM/COSMO filter, even zero for three P–L systems (Table 1). CS and ASP perform slightly worse. AMBER/GB performs satisfyingly well for three systems but yields 171 false positives for TACE. For AChE, all the SFs perform satisfyingly well. For AR and HIV PR, GlideXP generates the highest number of false positive solutions and also shape-wise the free energy landscape looks ill-defined (Figure 2). In the case of AR, a plateau of negative relative scores is observed for GlideXP. The hardest case is the TACE metalloprotein. Here, all the SFs produce false-positive solutions but to a different extent. The SQM/COSMO filter performs best, followed by CS. This example in particular shows the strength of an electronic-structure theory description of P–L binding. The presence of the metal cation in this protein and the associated charge-transfer effects between the ligand and the cation are not adequately described by classical force-fields or statistical potentials, but they are well represented by the SQM/COSMO filter.

The second criterion, $RMSD^{max}$, is shown for the interval of the normalised relative scores below 5 (Table 2). The SQM/COSMO filter shows the lowest $RMSD^{max}$ of 0.88 Å on average. CS follows with 1.28 Å on average. ASP and AMBER/GB satisfy the condition of an averaged $RMSD^{max}$ up to 2 Å. AMBER/GB, however, fails in the difficult case of TACE with $RMSD^{max}$ of 4.76 Å. Analogous analyses at greater intervals have revealed a similar ordering of the SFs (Table S4).

The SQM/COSMO filter enables us not only to recognise the correct binding pose (RMSD below 2 Å) but also to go beyond this limit and evaluate even small changes in the geometry of the ligand binding.

The price for such a high accuracy is the increased computational time requirements. The SQM/COSMO filter is ca. 100-times slower than the statistics- and knowledge-based SFs and about 10-times slower than the classical physics-based AMBER/GB. However, compared to the standard SQM-based SF, it is ca. 100-times faster. The speed can be further enhanced by parallelisation.

To summarise, we have pushed the limits of the accuracy of SFs to judge the energetics of P–L noncovalent interactions. Based on our development and extensive experience with SQM-based scoring function[21], the SQM/COSMO filter has been introduced. It features two dominant terms to describe P–L interaction, namely the ΔE_{int} term at the PM6-D3H4X level for gas-phase noncovalent interactions and the $\Delta\Delta G_{solv}$ term at the COSMO level for implicit solvation. We showed previously that both these methods are very accurate at a reasonable speed.[16,20] The SQM/COSMO energy is calculated in four unrelated P–L complexes. The SQM/COSMO filter is compared to eight widely used SFs, which are statistics-, knowledge- or force-field-based. The SQM/COSMO scheme exhibits a superior performance as judged by two criteria, the number of false positives and $RMSD^{max}$. In contrast to standard SFs, no fitting against data sets has been involved. Furthermore, it offers generality and comparability across the chemical space and no system-specific parameterisations have to be performed. The time requirements allow for calculations of thousands of docking poses as we have demonstrated in this pilot study. 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. The proof of concept that reliable QM calculations can be now performed for tens of thousands of large biochemical entities opens way to progress in closely related disciplines such as materials design.

This work was supported by research projects RVO 61388963 awarded by the Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic. We acknowledge the financial support of the Czech Science Foundation (grant number P208/12/G016). The authors acknowledge the support by the project L01305 of the Ministry of Education, Youth and Sports of the Czech Republic. The computations were performed at the Center for Information Services and High Performance Computing (ZIH) at TU Dresden.

Keywords: semi-empirical quantum mechanics • scoring function • molecular docking • virtual screening • noncovalent interactions

References

1 G. L. Warren, C. W. Andrews, A. M. Capelli, B. Clarke, J. LaLonde, M. H. Lambert, M. Lindvall, N. Nevis, S. F. Semus, S. Senger et al., *J. Med. Chem.,* 2006, **49**, 5912; A. R. Leach, B. K. Shoichet, C. E. Peishoff, *J. Med. Chem.,* 2006, **49**, 5851.
2 H. Gohlke, G. Klebe, *Angew. Chem. Int. Ed.,* 2002, **41**, 2644.
3 R. Meier, M. Pippel, F. Brandt, W. Sippl, C. Baldauf, *J. Chem. Inf. Model.,* 2010, **50**, 879.
4 P. S. Charifson, J. J. Corkery, M. A. Murcko, W. P. Walters, J. *Med. Chem.,* 1999, **42**, 5100; R. Wang, S. J. Wang, *J. Chem. Inf. Comput. Sci.,* 2001, **41**, 1422.
5 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.,* 2004, **25**, 1157.
6 K. Raha, M. B. Peters, B. Wang, N. Yu, A. M. Wollacott, L. M. Westerhoff, K. M. Merz, *Drug Discov. Today,* 2007, **12**, 725; M. Xu, M. A. Lill*, Drug Discov. Today: Technologies,* 2013, **10**, 411.
7 D. Mucs, R. A. Bryce, *Exp. Opin. Drug Discov.,* 2013, **8**, 263.
8 S. A. Hayik, R. Dunbrack, K. M. Merz, *J. Chem. Theory Comput.,* 2010, **6**, 3079.
9 M. Hennemann, T. Clark, *J. Mol. Model*., 2014, **20**, 2331.
10 H. S. Muddana, M. K. Gilson*, J. Chem. Theory Comput.,* 2012, **8**, 2023; P. Mikulskis, S. Genheden, K. Wichmann, U. Ryde, *J. Comput. Chem.,* 2012, **33**, 1179.
11 P. Soderhjelm, J. Kongsted, U. Ryde, *J. Chem. Theory Comput.,* 2010, **6**, 1726.
12 K. Wichapong, A. Rohe, C. Platzer, I. Slynko, F. Erdmann, M. Schmidt, W. Sippl, *J. Chem. Inf. Model.,* 2014, **54**, 881; P. Chaskar, V. Zoete, U. F. Röhrig, J. Chem. Inf. Model., 2014, **54**, 3137; S. K. Burger, D. C. Thompson, P. W. Ayers*, J. Chem. Inf. Model.,* 2011, **51**, 93.
13 J. Antony, S. Grimme, D. G. Liakos, F. Neese, *J. Phys. Chem. A,* 2011, **115**, 11210.
14 M. S. Gordon, D. G. Fedorov, S. R. Pruitt, L. V. Slipchenko, *Chem. Rev.,* 2012, **112**, 632; J. Antony, S. Grimme, *J. Comput. Chem.,* 2012, **33**, 1730.
15 M. Lepšík, J. Řezáč, M. Kolář, A. Pecina, P. Hobza, J. Fanfrlík, *ChemPlusChem,* 2013, **78**, 921.
16 J. Řezáč, J. Fanfrlík, D. Salahub, P. Hobza, *J. Chem. Theory Comput.,* 2009, **5**, 1749; J. Řezáč, P. Hobza, *Chem. Phys. Lett.,* 2011, **506**, 286; J. Řezáč, P. Hobza, *J. Chem. Theory Comput.,* 2012, **8**, 141.
17 K. Raha, K. M. Merz*, J. Am. Chem. Soc.,* 2004, **126**, 1020; K. Raha, K. M. Merz, *J. Med. Chem.,* 2005, **48**, 4558.
18 J. J. P. Stewart, *J. Mol. Model,* 2007, **13**, 1173.
19 A. Klamt, G. Schüürmann*, J. Chem. Soc., Perkin Trans. 2,* 1993, 799.
20 M. Kolář, J. Fanfrlík, M. Lepšík, F. Forti, F. J. Luque, P. Hobza, *J. Phys. Chem. B,* 2013, **117**, 5950.
21 J. Fanfrlík, A. K. Bronowska, J. Řezáč, O. Přenosil, J. Konvalinka, P. Hobza, *J. Phys. Chem. B,* 2010, **114**, 12666; P. Dobeš, J. Řezáč, J. Fanfrlík, M. Otyepka, P. Hobza, *J. Phys. Chem. B,* 2011, **115**, 8581; A. Pecina, O. Přenosil, J. Fanfrlík, J. Řezáč, J. Granatier, P. Hobza, M. Lepšík, *Collect. Czech. Chem. Commun.,* 2011, **76**, 457; A. Pecina, M. Lepšík, J. Řezáč, J. Brynda, P. Mader, P. Řezáčová, P. Hobza, J. Fanfrlík, *J. Phys. Chem. B,* 2013, **117**, 16096; J. Fanfrlík, M. Kolář, M. Kamlar, D. Hurn, F. X. Ruiz, A. Cousido-Siah, A. Mitschler, J. Řezáč, E. Munusamy, E.; M. Lepšík, et al., *ACS Chem. Biol.,* 2013, **8**, 2484; J. Fanfrlík, F. X. Ruiz, A. Kadlčíková, J. Řezáč, A. Cousido-Siah, A. Mitschler, S. Haldar, M. Lepšík, M. H. Kolář, P. Majer, et al., *ACS Chem. Biol.,* 2015, **10**, 1637.
22 J. Fanfrlík, P. S. Brahmkshatriya, J. Řezáč, A. Jílková, M. Horn, M. Mareš, P. Hobza, M. Lepšík, *J. Phys. Chem. B,* 2013, **117**, 14973.
23 R. A. Friesner, R. B. Murphy, M. P. Repasky, L. L. Frye, J. R. Greenwood, T. A. Halgren, P. C. Sanschagrin, D. T. Mainz*, J. Med. Chem.,* 2006, **49**, 6177.
24 O. Korb, T. Stützle, T. E. Exner*, J. Chem. Inf. Model,* 2009, **49**, 84.
25 O. Trott, A. J. Olson, *J. Comput. Chem.*, 2010, **31**, 455.
26 M. D. Eldridge, C. W. Murray, T. R. Auton, G. V. Paolini, R. P. Mee, *J. Comput.-Aided Mol. Des.,* 1997, **11**, 425.
27 G. Jones, P. Willett, R. C. Glen, A. R. Leach, R. Taylor, *J. Mol. Biol.,* 1997, **267**, 727.
28 W. T. M. Mooij, M. L. Verdonk, *Proteins,* 2005, **61**, 272.
29 Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, R.; T. Lee*, J. Comput. Chem.,* 2003, **24**, 1999; V. Tsui, D. A. Case, *Biopolymers,* 2001, **56**, 275.
30 A. Nicholls, A. N. Jain*, J. Comput. Aided Mol. Des.,* 2008, **22**, 133.
31 B.Liu, S. Wang, X. Wang, *Sci. Rep*., 2015, **50**, 15479.
32 E. Perola, W. P. Walters, P. S. Charifson, *Proteins,* 2004, **56**, 235; J. W. M. Nissink, Ch. Murray, M. Hartshorn, M. L. Verdonk, J. C. Cole, R. Taylor, *Proteins,* 2002, **49**, 457; P. Ferrara, H. Gohlke, D. J. Price, G. Klebe, C. L. Brooks, *J. Med. Chem.,* 2004, **47**, 3032.
33 G. Klebe, *Drug Discov. Today,* 2006, **11**, 580.
34 H. Dvir, D. M. Wong, M. Harel, X. Barril, M. Orozco, F. J. Luque, D. Munoz-Torrero, P. Camps, T. L. Rosenberry, I. Silman et al., *Biochemistry,* 2002, **41**, 2970.
35 U. K. Bandarage, T. Wang, J. H. Come, E. Perola, Y. Wei, B. G. Rao, *Bioorg. Med. Chem. Lett.,* 2008, **18**, 44.
36 H. Steuber, A. Heine, G. Klebe, *J. Mol. Biol.,* 2007, **368**, 618.
37 J. Brynda, P. Řezáčová, M. Fábry, M. Hořejší, R. Štouračová, J. Sedláček, M. Souček, M. Hradílek, M. Lepšík, J. Konvalinka*, J. Med. Chem.,* 2004, **47**, 2030.

SUPPLEMENTARY MATERIAL

The SQM/COSMO Filter: Reliable Native Pose Identification Based on the Quantum-Mechanical Description of Protein–Ligand Interactions and Implicit COSMO Solvation

Adam Pecina[1+], René Meier[2+], Jindřich Fanfrlík[1], Martin Lepšík[1], Jan Řezáč[1], Pavel Hobza[1,3]* and Carsten Baldauf[4]*

[1] *Institute of Organic Chemistry and Biochemistry (IOCB) and Gilead Sciences and IOCB Research Center, Flemingovo nám. 2, 16610 Prague 6 (Czech Republic).*

[2] *Institut für Biochemie, Fakultät für Biowissenschaften, Pharmazie und Psychologie Universität Leipzig, Brüderstrasse 34, D-04109 Leipzig (Germany)*

[3] *Regional Centre of Advanced Technologies and Materials, Department of Physical Chemistry, Palacký University, 77146 Olomouc (Czech Republic)*

[4] *Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4-6, D-14195 Berlin (Germany)*

[+] These authors have contributed equally to this work

E-mail: hobza@uochb.cas.cz, baldauf@fhi-berlin.mpg.de

Abbreviations

AChE … Acetylcholine esterase
AR … Aldose reductase
ASP … Astex Statistical Potential
ChemPLP … GOLD ChemPLP score
COSMO ... Conductor-like Screening model
CS ... Chemscore
FIRE … Fast Inertial Relaxation Engine
GAFF … general AMBER force field
GB … generalised Born implicit solvent model
GlideXP ... GlideScore Extra Precision
GS … Goldscore
HIV PR ... HIV-1 protease
IQR … interquartile range
MAD … mean absolute deviation
MM … molecular mechanics
PB … Poisson-Boltzmann implicit solvent model
PLP ... PLANTS PLP score
P-L … protein–ligand
QM … quantum mechanical
Q1 and Q3 ... the first and the third quartile
RMSD … Root-mean-square deviation
$RMSD^{max}$ … maximal root-mean-square deviation
SD … Steepest descent
SF … scoring function
SMD ... Solvation Model based on Density
SQM… semiempirical quantum mechanical
TACE … TNF-α converting enzyme
TDOF... torsional degrees of freedom
Vina ... AutoDock Vina
vdW … van der Waals

1. Methods

1.1. Protein-ligand complexes

Four unrelated protein-ligand complexes that feature difficult-to-handle noncovalent interactions were chosen for this study. These were resolved by X-ray crystallography at reasonable resolution (Table S1) and the ligand electron density was well distinguishable. The ligands are shown in Figure S1.

Table S1. Protein-ligand complexes used in this study

PDB	Reference	Resolution	Protein	Ligand	Features
1E66	[1]	2.10 Å	**AChE**	Huprine X	Two binding pockets, halogenated ligand
2IKJ	[2]	1.55 Å	**AR**	IDD393	Cofactor, halogenated ligand
3B92	[3]	2.00 Å	**TACE**	440	Metallo-protein, Zn^{2+} cation coordinated by S^-, three water molecules in binding site
1NH0	[4]	1.03 Å	**HIV PR**	KI2	Large, flexible and charged ligand, structural water molecule in binding site

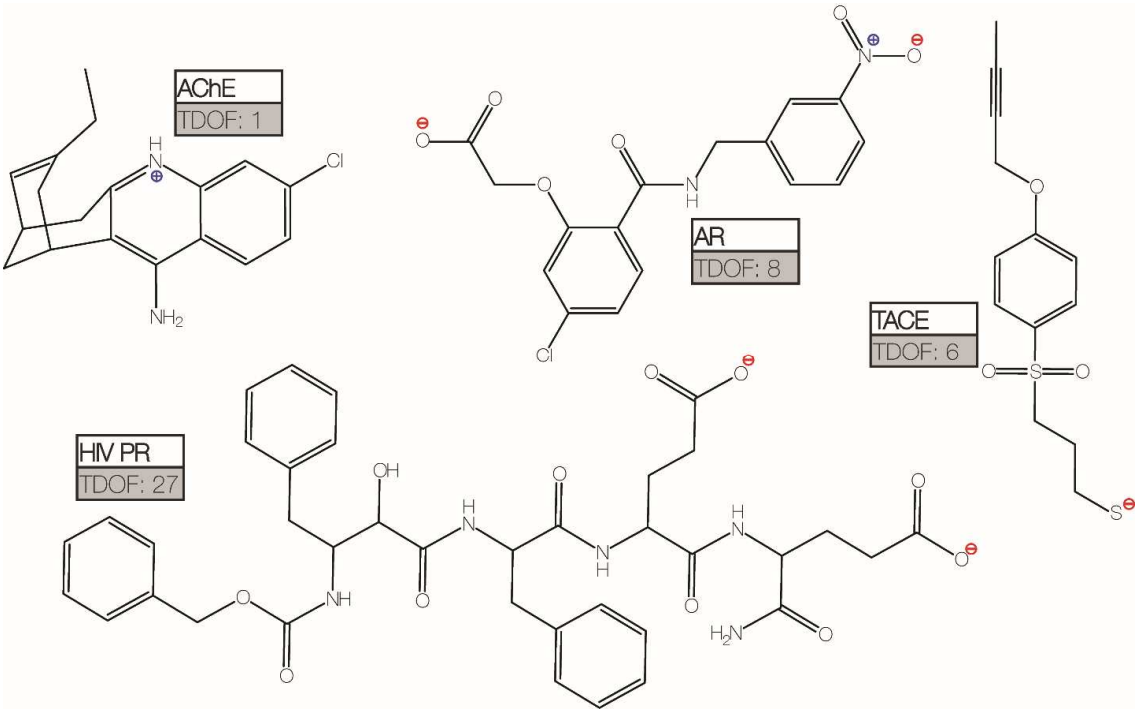


Figure S1. 2D structures of the studied ligands (labelled by their target protein, see table S1) with their charges and the numbers of torsional degrees of freedom (TDOF).

1.2. Generation of Protein-Ligand Poses via Docking

Four different docking programs with overall 7 different scoring functions (Table S2) were used to generate protein-ligand poses, the workflow is summarized in Figure S2. The individual docking runs were started from the structure of the ligand in the respective X-ray structure and in addition from up to 10 randomized ligand conformations. These starting conformations were created with the conformation search in MOE[5] with at least 2 Å RMSD between the conformations and an energy window of 7 kcal/mol using the Amber 10[6]+EHT force field.[7] For each docking run, up to 100 receptor-ligand poses were generated by each of the 7 docking setups. If the docking program supports removal of redundant results, this option was used. The hypothetical maximal number of 7,700 decoys per receptor-ligand pair was reduced by clustering with a cut-off of 0.5 Å for decoys up to 2 Å RMSD to the crystal structure and a cut-off of 2 Å for all other decoys in order to avoid redundant conformations. This yielded more than 2,800 ligand-receptor poses; exact numbers are given in Table S2.

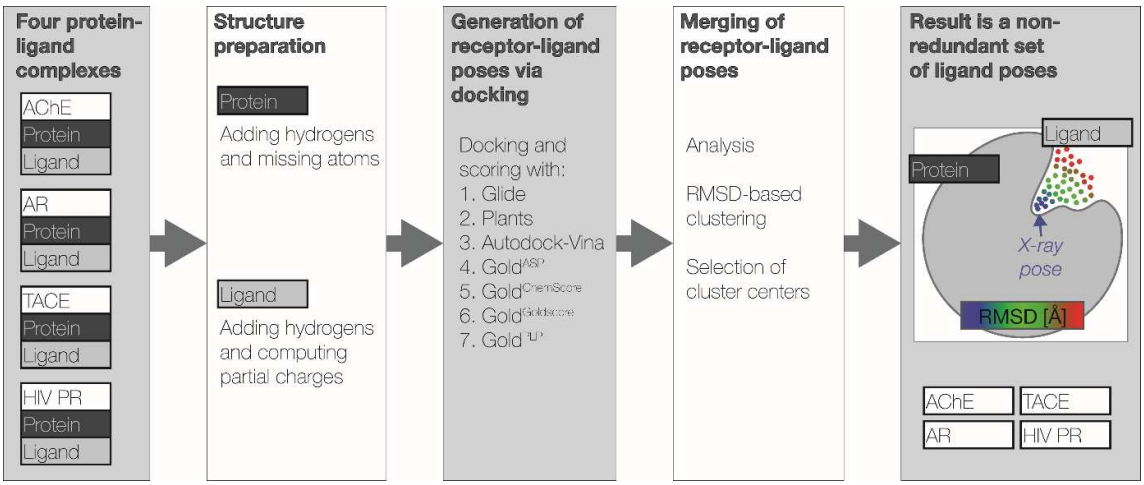


Figure S2. Schematic representation of the workflow that was used to generate sets of alternative and non-redundant binding poses of protein-ligand complexes.

Table S2. Docking protocols and numbers of generated decoy poses.

Setup	Software	Energy function	Number of generated poses			
			AChE	AR	TACE	HIV PR
1	Glide	GlideScore XP	4	19	27	38
2	PLANTS	PLANTS PLP	200	1,100	1,100	700
3	Autodock Vina	Vina	2	168	220	140
4	GOLD	ASP	200	1,100	1,100	700
5		Chemscore	200	1,100	1,100	700
6		Goldscore	200	1,100	1,100	700
7		ChemPLP	200	1,100	1,100	700
Poses after clustering		sum = 2,865	67	163	734	1901

1.3. Physics-based scoring

1.3.1. Structure preparation

Careful preparation of the protein-ligand structures was carried out as physics-based methods (AMBER/GB and SQM/COSMO) are sensitive to molecular details, e.g. protonation states and geometrical clashes generated by the docking procedures.

Ligands were prepared by adding hydrogen atoms with UCSF Chimera.[8] Force-field parameters for the ligands were taken from GAFF[9] and partial charges were derived from RESP fitting of the electrostatic potential (ESP) calculated at the AM1-BCC level.[10]

The protein structures were prepared using the Reduce[11] and LEaP programs[12] that are part of the AMBER 10 package[6]. The protonation states of histidine side chains were manually assigned based on the hydrogen-bonding patterns and pH of the crystallization conditions.

Acetylcholine esterase (AChE). For the 1E66 X-ray structure (Table S1), the carbohydrate modifications of the enzyme were not considered. Based on the experimental pH of crystallization of 5.6,[1] His471 is modeled as doubly protonated. The ligand Huprine X is protonated (charge +1, Figure S1) and forms a hydrogen bond with the backbone carbonyl of His440.

Aldose reductase (AR). The structure 2IKJ (Table S1) features the NADP cofactor (charge -3), singly protonated histidines, and a ligand with charge -1. The O1 of the inhibitor carbonyl group forms a hydrogen-bond with Nε1 of Trp111 and the O2 binds to the side-chain of His110 and Tyr48.[2] The nitrophenyl group of the inhibitor is placed in the specificity pocket of the enzyme where it forms an interaction to Leu300 NH via the nitro oxygen and a face-to-face oriented π…π stacking with the side-chain of Trp111.

TNF-α converting enzyme (TACE). It is a metallo-protease whose structure (PDB code 3B92, Table S1) features a Zn^{2+} cation that is coordinated with the inhibitor thiol moiety and the three histidine side-chains of the protein. The thiol group was modeled as thiolate (S^-) in analogy with deprotonated sulfonamide (SO_2NH^-) group that we studied earlier.[13] Three structural water molecules from the crystal structure

were considered throughout this study. Three water molecules (W524, W538, W676) were required to achieve sensible docking results. The first water molecule is bound by Ala439 and the sulfonyl group of the inhibitor, the second is bound by Glu398 and Val440 and the third is bound by Tyr436 and Ile438.

HIV-1 Protease (HIV PR). This homodimeric enzyme (Table S1) features a structural water molecule in the flap region of the active site that was considered in all the calculations. The Asp25/25' dyad is considered doubly protonated based on the crystallographic findings.[4] The Asp30 side chain is protonated on Oδ2 according to the QM calculations of protein-ligand stabilities and proton transfer barriers.[14]

1.3.2. Geometry Optimization

Hydrogen positions were subjected to steepest-descent optimization (SD) and simulated annealing with the SANDER module of the AMBER package.[6] In the *protein-ligand complexes*, the positions of the hydrogen atoms within 6 Å around the ligand position were optimized in three steps: (i) 50 optimization steps using SD, (ii) simulated annealing for this part of the protein/ligand complex, (iii) optimization of hydrogen positions with the FIRE algorithm. For poses with close contacts between ligand and protein below 1.5 Å, 50 SD optimization steps of the ligand embedded in the fixed protein were performed.

1.3.3. Scoring

In the Pavel Hobza's group, we have been developing an SQM scoring function[15] which correctly describes all types of noncovalent interactions, viz. dispersion, hydrogen and halogen-bonding. We have demonstrated its applicability for various protein-ligand systems, such as protein kinases, aldose reductase, HIV-1 protease and carbonic anhydrase.[13-16] As a special case, we have also extended it to treat covalent inhibitor binding.[17] Recently, there have been several attempts to make QM methods applicable in virtual screening, especially by their acceleration and simplification.[18]

1.3.4 SQM region

To make the calculations faster, we defined a sphere of 8 to 12 Å (roughly 2,000 atoms) around the aligned ligand poses as a region representing the binding site.

This region was treated by SQM 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 (see later, Figure S4).

1.4. Score Normalisation

The calculated scores are on different scales and thus are not straightforwardly comparable. In order to generate comparable numbers, they were converted to a normalised score. For each data set, i.e. all poses of a protein-ligand complex ranked by a scoring/energy function, the first quartile (Q_1) and the third quartile (Q_3) were calculated. The interquartile range (IQR) is defined as:

$$IQR = Q_3 - Q_1$$

All poses with energies greater than $Q_3 + 1.5\,IQR$ were considered as high energy outliers and were removed from the dataset. Finally, the relative energies of poses with respect to the energy of the X-ray pose were scaled with a factor F:

$$F = \frac{100}{(Q_3 + 1.5\,IQR) - (Q_1 - 1.5\,IQR)}$$

The resulting normalised scores are comparable between the different energy functions.

1.5. RMSD Measurements

For all the ligand poses generated, the distances in Cartesian space (root-mean square deviation, RMSD) from the X-ray structure were determined. The RMSD values were calculated without considering hydrogens. The algorithm takes full molecule symmetry into account, based on a graph depth-first-search[5] and atom priorities following the Cahn-Ingold-Prelog rules. All RMSD values were calculated without superposition so that the resulting values truly express a distance in the multi-dimensional energy landscape.

1.6. Normalised Scores vs. RMSD

The energy values of every pose were plotted *vs.* the RMSD value to the crystal structure. The cloud of points (see Appendix of the SI) was 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 S3). The graph was constructed by removing

all data points above a point if the connecting line would have a |slope| > 12 starting with the lowest energy data point. This was repeated on all points in the order of increasing energy until the whole data set was processed. The remaining points were connected with lines.

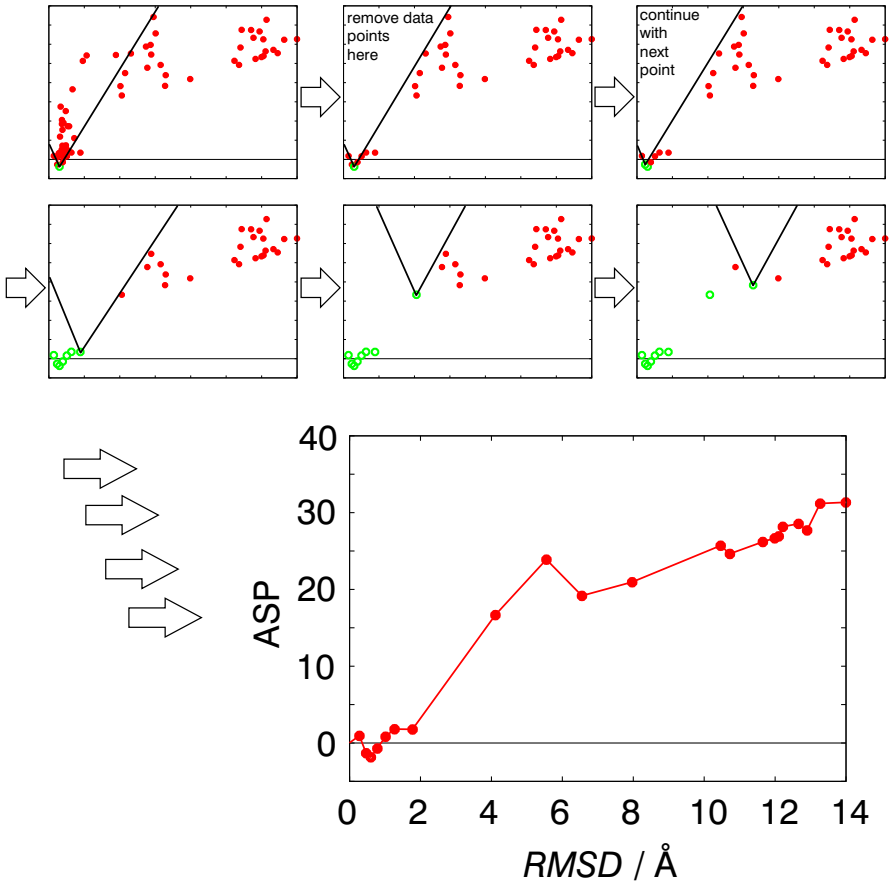


Figure S3. Scheme of the algorithm to create the lower boundary from the whole data set. An iterative process reduces the large amount of data points to the most important points for this study.

2. Results

2.1. Convergence of SQM region size

In all four systems we compared the influence of applying the truncation scheme to covering the full protein-ligand complex in a SQM calculation. Table S3 shows gives the mean absolute deviation (MAD). The MAD values of up to 4 kcal/mol are, however, not visible in the overall shape of the lower-bound representation of the binding energy landscape (see Figure 2). The results of SQM/COSMO filter and full SQM/COSMO are in good agreement (Figure S4). The use of SQM/COSMO filter can thus be recommended for use due its speed.

Table S3. Mean absolute deviations (MAD/kcal.mol^{-1}) between SQM and full-SQM energy approaches.

Protein	AChE	AR	TACE	HIV PR
Overall atoms	8,388	5,160	4,064	3,230
Atoms in SQM region	1,843	1,960	1,489	2,200
MAD / kcal/mol	2.8±1.6	2.5±0.8	0.5±0.9	3.9±0.8

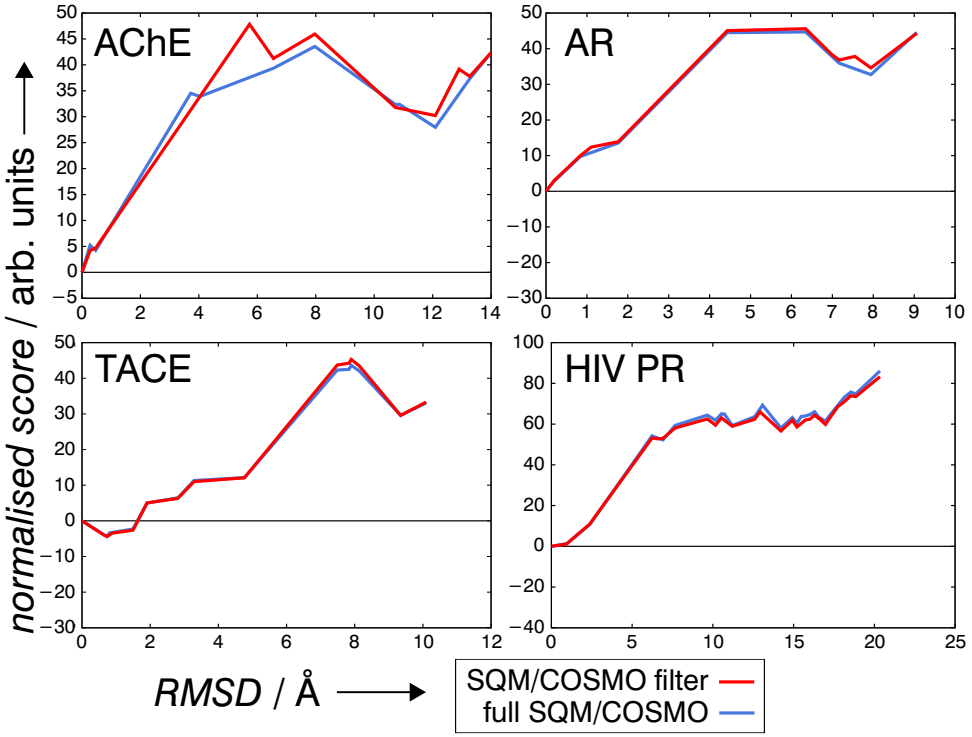


Figure S4. Comparison of the full-size SQM/COSMO vs. SQM/COSMO filter plots

2. 2. Quality Criterion – RMSD^{max}

Here, were present the results of of second criterion, RMSD^{max}, for larger score windows of 10 and 20 (Table S4). In the former, 3 scoring functions (SQM/COSMO, AMBER/GB and Gold CS) recognized the correct binding pose (RMSD < 2 Å). The SQM/COSMO showed the lowest RMSD^{max} (1.32 Å). In the score window of 20, no scoring function met the limit of 2 Å. However, AMBER/GB and SQM/COSMO were close (RMSD^{max} of 2.04 and 2.49 Å).

Table S4. Behaviour of the scoring function within normalised scores up to 10 and 20

	Scoring function								
			Glide	PLANTS	AutoDock			Gold	
	SQM/COSMO	AMBER/GB	XP	PLP	Vina	ASP	CS	GS	ChemPLP
	Maximal RMSD within a window of 10 of the normalised Score								
AchE	0.63	1.01	2.13	2.13	1.01	1.78	1.78	1.14	1.01
AR	0.84	0.19	7.54	3.47	3.54	2.59	1.77	7.66	1.81
TACE	2.81	4.76	3.13	2.91	8.06	2.86	2.63	2.44	2.73
HIV PR	1.01	0.94	17.74	13.13	11.62	1.00	1.08	14.20	12.64
Average	**1.32**	**1.62**	**7.64**	**5.41**	**6.06**	**2.06**	**1.81**	**6.36**	**4.55**
	Maximal RMSD within a window of 20 of the normalised Score								
AChE	1.06	1.14	11.99	4.11	19.85	7.97	6.58	5.55	1.43
AR	1.77	1.16	9.06	7.79	9.75	3.90	2.32	8.18	3.54
TACE	2.37	1.10	18.22	16.51	12.60	1.94	1.93	16.90	14.20
HIV PR	4.76	4.76	3.13	2.91	9.59	7.41	2.63	6.98	7.13
Average	**2.49**	**2.04**	**10.60**	**7.83**	**12.95**	**5.31**	**3.37**	**9.40**	**6.58**

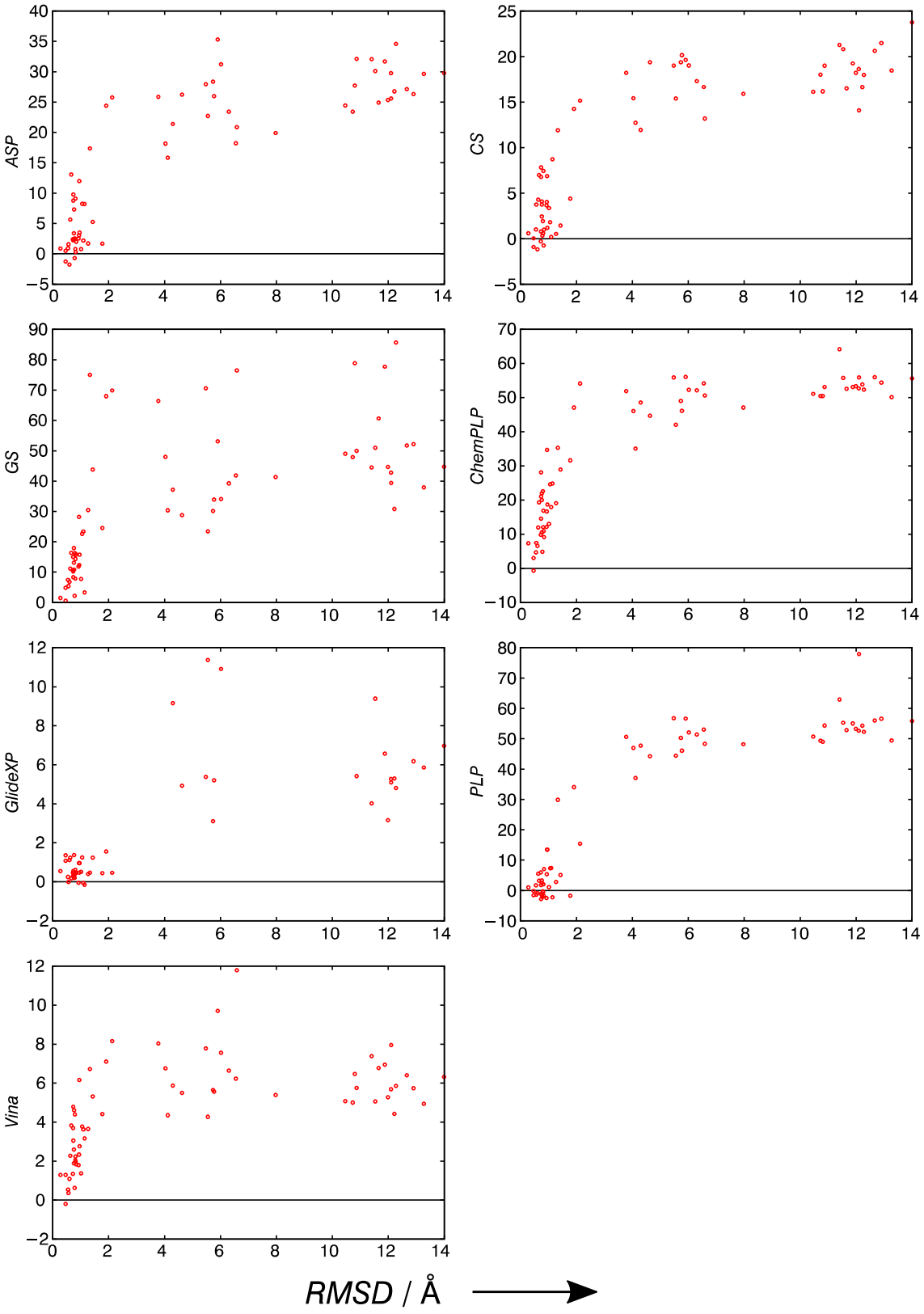
3. References

[1] H. Dvir, D. M. Wong, M. Harel, X. Barril, M. Orozco, F. J. Luque, D. Munoz-Torrero, P. Camps, T. L. Rosenberry, I. Silman et al., *Biochemistry* **2002**, *41*, 2970–2981.
[2] H. Steuber, A. Heine, G. Klebe, *J. Mol. Biol.* **2007**, *368*, 618–638.
[3] U. K. Bandarage, T. Wang, J. H. Come, E. Perola, Y. Wei, B. G. Rao, *Bioorg. Med. Chem. Lett.* **2008**, *18*, 44–48
[4] J. Brynda, P. Řezáčová, M. Fábry, M. Hořejší, R. Štouračová, J. Sedláček, M. Souček, M. Hradílek, M. Lepšík, J. Konvalinka, *J. Med. Chem.* **2004**, *47*, 2030–2036.
[5] Chemical Computing Group Inc., Molecular Operating Environment (MOE) 2013.08, 1010 Sherbooke St. West, Suite #910, Montreal, QC(Canada), **2015**.
[6] D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, M. Crowley, R.C. Walker, W. Zhang, K.M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D.H. Mathews, M.G. Seetin, C. Sagui, V. Babin, and P.A. Kollman, AMBER 10, University of California, San Francisco, **2008**.
[7] P. R. Gerber, K. Muller, *J. Comput.Aided Mol. Des.* **1995**, *9*, 251–268.
[8] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J. Comput. Chem.*, **2004**, *25*, 1605–1612.
[9] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.*, **2004**, *25*, 1157–1174.
[10] a) A. Jakalian, B. L. Bush, D. B. Jack, C. I. Bayly, *J. Comput. Chem.* **2000**, *21*, 132–146; b) A. Jakalian, D. B. Jack, C. I. Bayly, *J. Comput. Chem.* **2002**, *23*, 1623–1641.
[11] J. M. Word, S. C. Lovell, J. S. Richardson, D. C. Richardson, *J. Mol. Biol.* **1999**, *285*, 1735–1747.
[12] C. E. A. F. Schafmeister, W. S. Ross, V. Romanovski, LEAP, University of California, San Francisco, **1995**.
[13] A. Pecina, M. Lepšík, J. Řezáč, J. Brynda, P. Mader, P. Řezáčová, P. Hobza, J. Fanfrlík, *J. Phys. Chem. B* **2013**, *117*, 16096-16104
[14] A. Pecina, O. Přenosil, J. Fanfrlík, J. Řezáč, J. Granatier, P. Hobza, M. Lepšík, *Collect. Czech. Chem. Commun.* **2011**, *76*, 457-479.
[15] a) J. Fanfrlík, A. K. Bronowska, J. Řezáč, O. Přenosil, J. Konvalinka, P. Hobza, *J. Phys. Chem. B* **2010**, *114*, 12666-12678, b) M. Lepšík, J. Řezáč, M. Kolář, A. Pecina, P. Hobza, J. Fanfrlík, *ChemPlusChem* **2013**, *78*, 921-931.
[16] a) P. Dobeš, J. Řezáč, J. Fanfrlík, M. Otyepka, P. Hobza, *J. Phys. Chem. B* **2011**, *115*, 8581-8589; b) J. Fanfrlík, M. Kolář, M. Kamlar, D. Hurn, F. X. Ruiz, A. Cousido-Siah, A. Mitschler, J. Řezáč, E. Munusamy, E.; M. Lepšík, et al., *ACS Chem. Biol.* **2013**, *8*, 2484-2492; c) J. Fanfrlík, F. X. Ruiz, A. Kadlčíková, J. Řezáč, A. Cousido-Siah, A. Mitschler, S. Haldar, M. Lepšík, M. H. Kolář, P. Majer, et al., *ACS Chem. Biol.* **2015**, *10*, 1637-1642.
[17] J. Fanfrlík, P. S. Brahmkshatriya, J. Řezáč, A. Jílková, M. Horn, M. Mareš, P. Hobza, M. Lepšík, *J. Phys. Chem. B* **2013**, *117*, 14973-14982.
[18] a) K. Wichapong, A. Rohe, C. Platzer, I. Slynko, F. Erdmann, M. Schmidt, W. Sippl, *J. Chem. Inf. Model.* **2014**, *54*, 881–893; b) P. Chaskar, V. Zoete, U. F. Röhrig, *J. Chem. Inf. Model.* **2014**, *54*, 3137–3152; c) S. K. Burger, D. C. Thompson, P. W. Ayers, *J. Chem. Inf. Model* **2011**, *51*, 93–101.

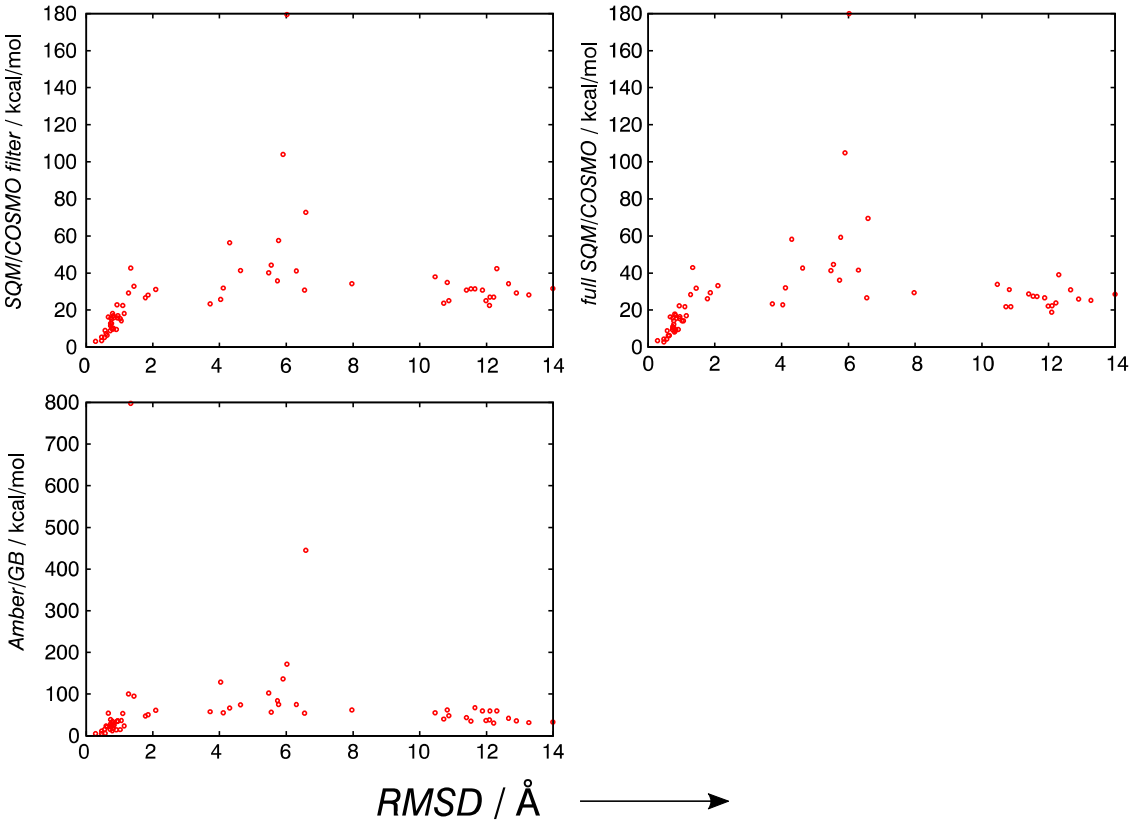
4. Appendix

Raw energy and score values plotted against RMSD values for the tested scoring functions for all poses of AChE, AR, TACE and HIV PR.

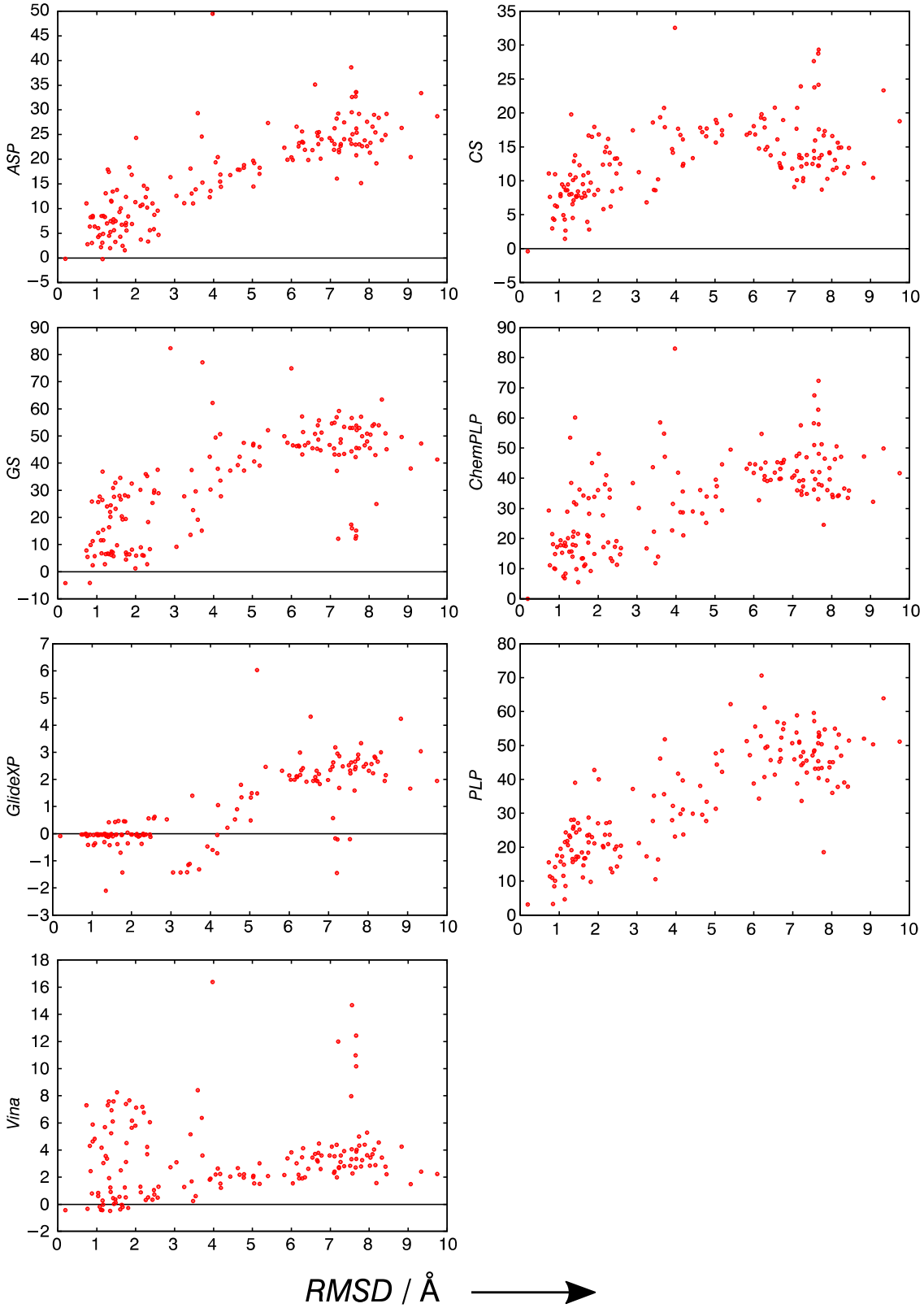
A1: Raw scores and energies for **AChE**.



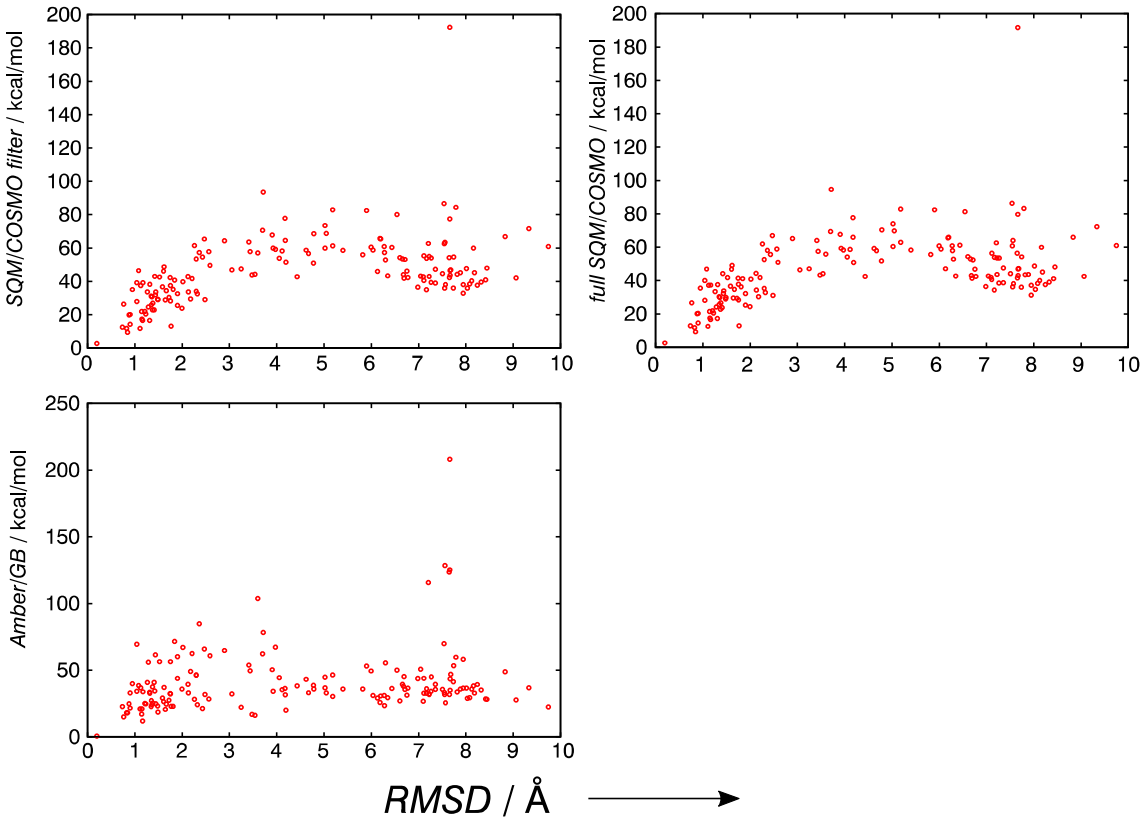
A1 continued: Raw scores and energies for **AChE**.



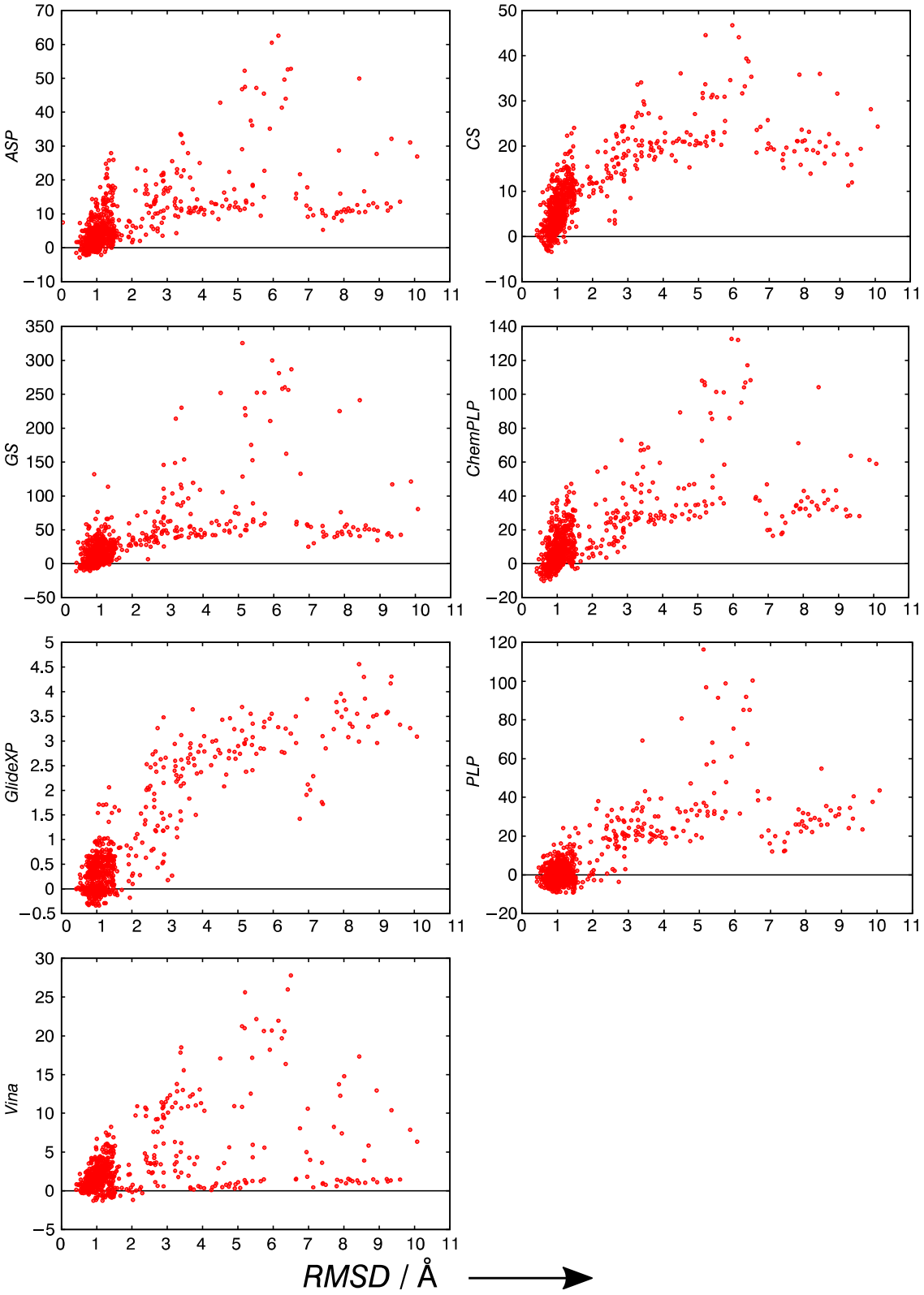
A2: Raw scores for **AR**.



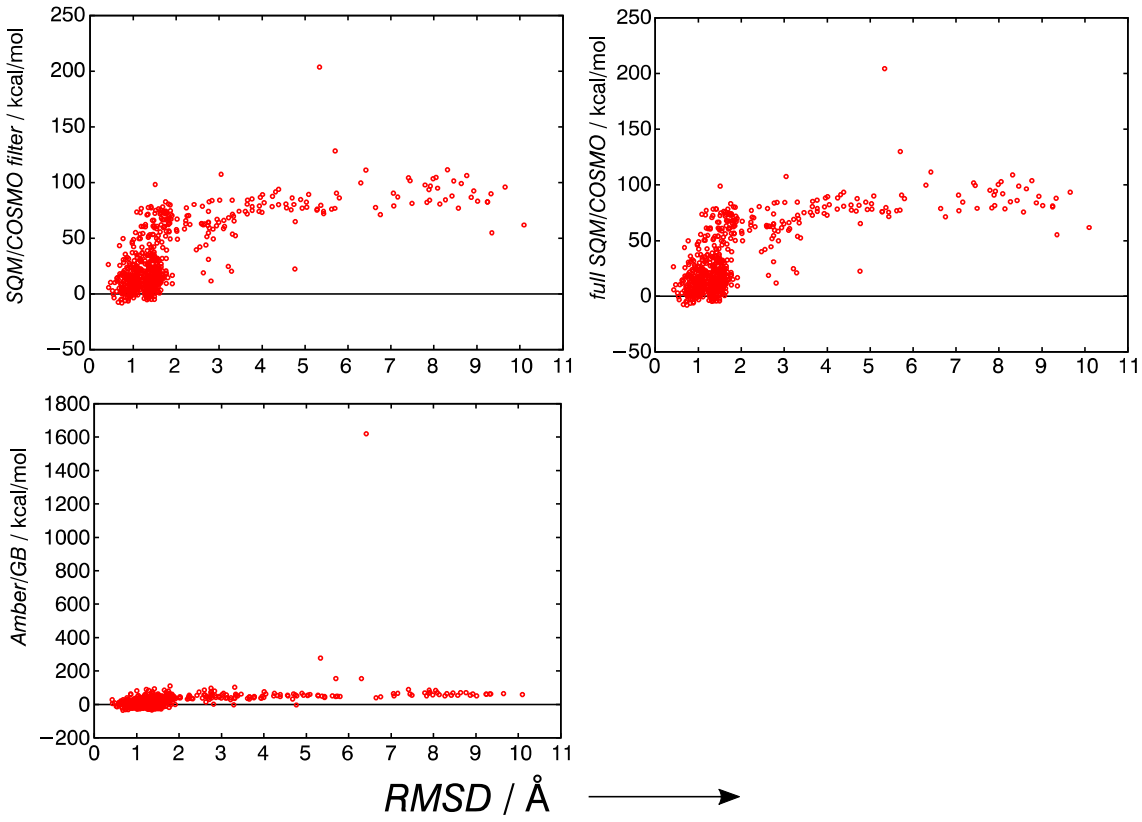
A2 continued: Raw scores and energies for **AR** continued.



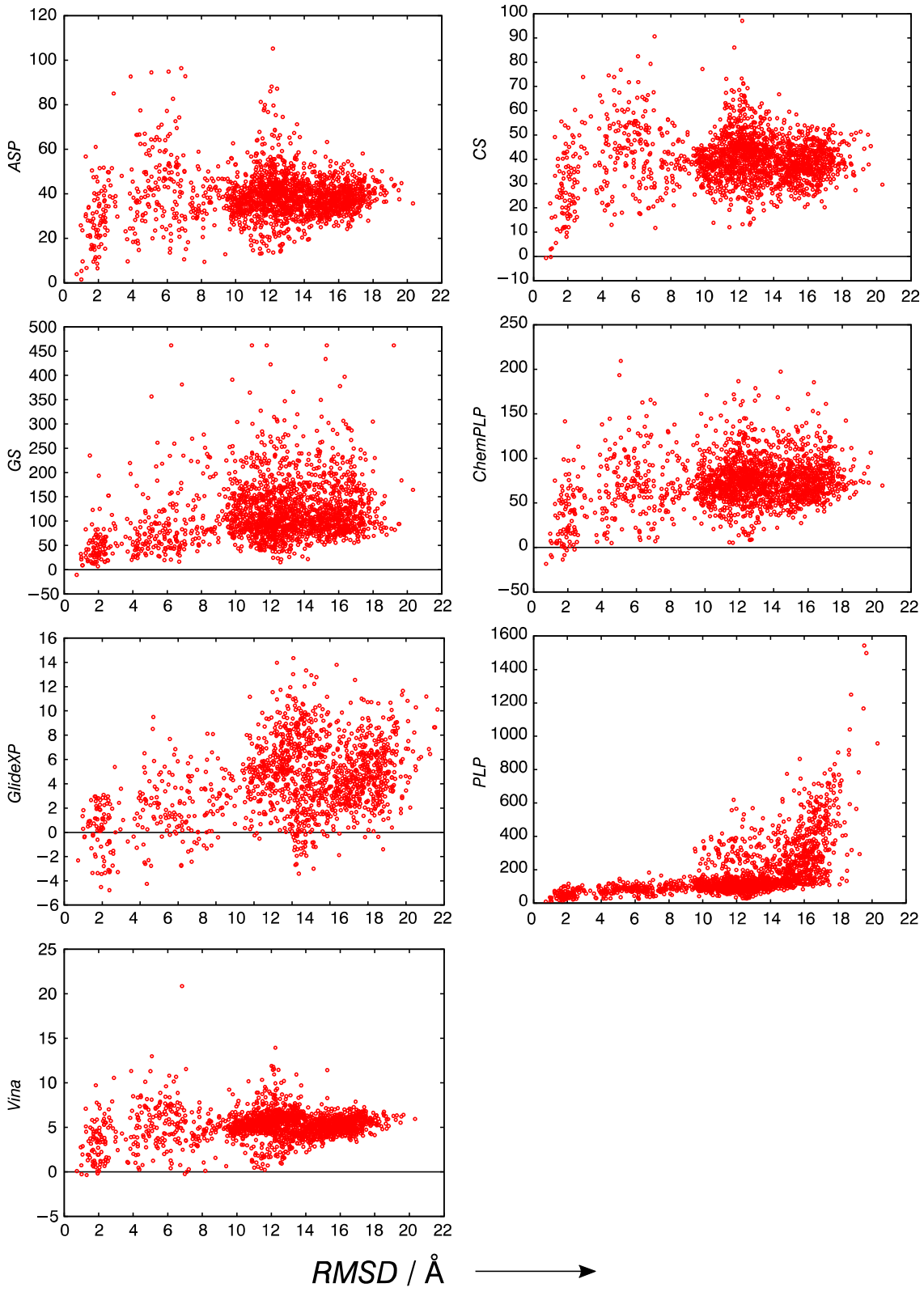
A3: Raw scores and energies for **TACE**.



A3 continued: Raw scores and energies for **TACE** continued.



A4: Raw scores for **HIV PR**.



A4 continued: Raw scores and energies for **HIV PR** continued.