

**Univerzita Karlova v Praze, Přírodovědecká fakulta
Katedra fyzikální a makromolekulární chemie**

**Charles University in Prague, Faculty of Science
Department of physical macromolecular chemistry**

Doktorský studijní program: Modelování chemických vlastností nano- a biostruktur
Ph.D. study program: Modelling of Chemical Properties of Nano- and Biostructures

Autoreferát disertační práce
Summary of the Ph.D. Thesis



**Kvantově chemické pojetí návrhu léčiv
Quantum Chemical Approach for *In Silico* Drug Design**

Adam Pecina

Školitel/Supervisor:

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

Praha, 2016

Abstrakt

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

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

Abstract

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

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

1. Úvod

Proces vývoje léčiva je nejen velmi nákladný, ale také s sebou nese nesmírně velké riziko selhání.¹ Těmto selháním lze předejít užitím různých výpočetních metod v jakékoli preklinické fázi vývoje, což je čím dál tím populárnější hlavně díky obrovskému pokroku ve výkonu softwaru i hardwaru. Na první pohled je tedy zřejmé, že počítačový návrh léčiv (CADD z angl. computer-aided drug design) může dosti zásadně ušetřit čas i náklady nutné na vývoj léčiva.²

Většina léčiv funguje tak, že kompetitivně inhibují jejich biologický cíl. Vůbec nejběžnějším příkladem je reverzibilně nekovalentně navázaný ligand v aktivním místě enzymu, zabraňující tak vazbě původního substrátu. Účinnost ligandu lze předpokládat úměrnou volné vazebné energii ligandu v proteinu. Tato volná vazebná energie (ΔG_b) souvisí s měřitelnou disociační konstantou (K_i) nebo s hodnotou koncentrace IC_{50} :

$$\Delta G_b = RT \ln K_i \quad (1)$$

Výpočetně lze volnou vazebnou energii stanovit mnoha způsoby, tj. od statistických metod po fyzikálně chemické přístupy. Pro správný popis všech příspěvků vazebné energie je nutné užití metod jež se umí s velkou přesností vypořádat s dostatečně velkou částí systému (tisíce atomů) a to v rozumném čase. Tyto metody musí spolehlivě popsat všechny typy nekovalentních interakcí, jakými jsou například vodíkové vazba, disperzní interakce či halogenová vazba.³ Výpočetní nástroje jsou důležité nejen pro samotnou identifikaci potencionálního léčiva (v rámci procesu tzv. virtual screeningu) či výběr modifikací jež zlepší účinnost a další vlastnosti těchto látek, ale také pro detailnější porozumění mechanismu jejich působení.

Narozdíl od velmi rozšířených molekulově mechanických (MM) metod (které jsou shrnuty v Ref. 4), jsou kvantově mechanické (QM) metody ze své podstaty schopny popsat nekovalentní interakce spolehlivě.⁵ QM metody jsou schopny popsat kvantové efekty jakými jsou například přenos protonu či náboje, polarizace nebo vazebné interakce pomocí sigma díry, zahrnují většinu organických a anorganických entit a nepotřebují žádné přídatné parametry.

2. Cíle práce

Tato práce si klade za cíl ukázat způsobilost některých kvantově mechanických přístupů na poli počítačového návrhu léčiv. Práce zdůrazňuje závěry osmi původních článků, které

jsou rozděleny do třech částí. První část se věnuje studiu síly a povahy všech typů vazebných interakcí pomocí sigma díry, tj. halogenové, chalkogenové a pniktogenové vazby, a to přesnými kvantově mechanickými metodami. V druhé části se aplikací hybridního QM/MM přístupu zabýváme komplexy proteinů s ligandy. Snažíme se zde hlavně odhalit strukturní vlastnosti systémů, jež jsou experimentálně nedosažitelné. Nakonec představujeme účinný nástroj založený na semiempirické kvantově mechanické (SQM) metodě, jenž může být použit pro virtuální prohledávání látek z databází.

3. Metody

Všechny zde použité metody jsou založeny na kvantové mechanice a jejich různá přesnost a náročnost pak předurčily jejich použití v jednotlivých studiích. Konkrétně byly použity následující metody: od velmi přesné metody CCSD(T)/CBS⁶ pro srovnávací výpočty modelových systémů, přes vyhovující DFT-SAPT⁷ metodu pro rozložení interakčních energií, až po rychlé DFT a PM6 metody rozšířené o empirickou korekci pro disperzní energii a korekce pro vodíkovou a halogenovou vazbu.⁸ Hybridní QM/MM přístup⁹ kombinovaný s implicitně aproximovaným vodním prostředím¹⁰ (s výjimkou explicitně zahrnutých strukturně důležitých molekul vody) nám dovolil zkoumat komplexy proteinů s ligandy sestávající z tisíců atomů.

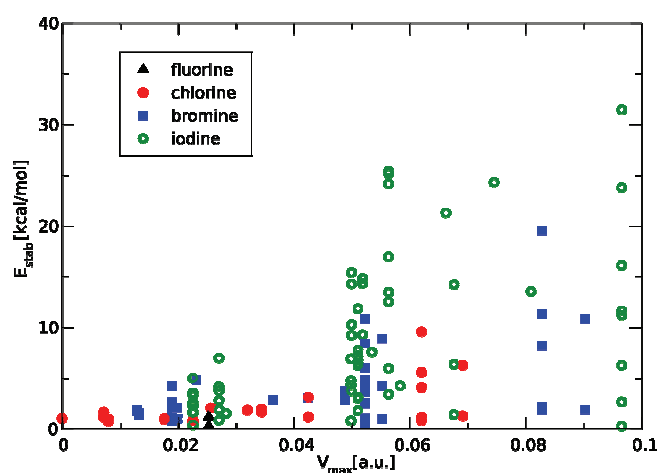
4. Výsledky a závěry

4.1 Původ interakcí pomocí σ -díry

V této první části výsledků jsme si kladli za cíl komplexně osvětlit podstatu halogenové vazby. Vytvořili jsme databázi 128 halogenově a dihalogenově vázaných komplexů různé velikosti a původu. Takto jsme do studie zahrnuli zástupce slabých a středně silně halogenově vázaných komplexů vytvořených ze standardních elektronových donorů (jakými jsou např. voda, amoniak, formaldehyd, dimethyl ether či trimethylamin), zástupce standardních halogenových donorů (např. halogenbenzeny nebo substituované halogenbenzeny) a silně halogenově vázané komplexy se značným přenosem náboje.

Sigma díry všech subsystémů nesoucí atom halogenu byly popsány z hlediska velikosti a síly.¹¹ Dále jsme spočítali celkové stabilizační energie všech komplexů na DFT-D3 a CCSD(T)/CBS úrovni a provedli rozklad interakčních energií metodou DFT-SAPT.

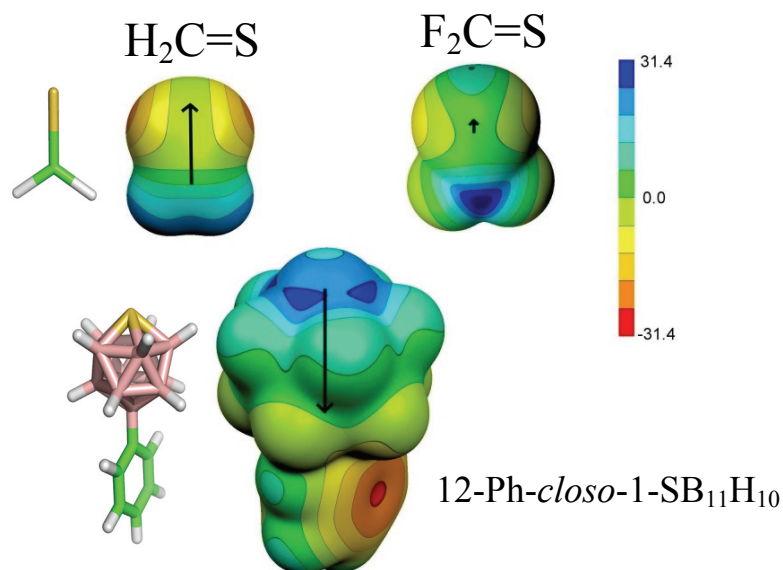
Ukázali jsme, že všechny studované subsystemy mají pozitivní sigma díru. Velikost a síla sigma díry dobře korelovaly ($R=0,86$) a byly zvýšeny se vzrůstajícím atomovým číslem halogenu a také s prezencí atomů fluorů které k sobě stahují elektrony. Síla sigma díry u dihalogenové vazby vzrůstala s klesajícím atomovým číslem druhého halogenu. Vlastnosti monomerů byly vztaženy k vlastnostem komplexů (Obrázek 1) a bylo ukázáno, že stabilizační energie a síla sigma díry korelují překvapivě slabě ($R=0,52$). Tato korelace nicméně vzrostla na 0,77 při výběru nejsilnějšího komplexu konkrétního monomeru. Síla sigma díry tedy informuje spíše o schopnosti daného monomeru tvořit halogenovou vazbu, než o síle halogenově vázaného komplexu.



Obr. 1: Závislost stabilizační energie E_{stab} na síle sigma díry V_{max} .

Z výsledků DFT-SAPTu je patrné, že dominantním příspěvkem k celkové stabilizační energii je u 62% komplexů elektrostatická energie, zatímco u zbývajících 38% případů je to energie disperzní. Lze tedy konstatovat, že oba příspěvky hrají stejně důležitou roli v charakterizaci halogenové vazby, přičemž elektrostatická energie odpovídá stabilizaci a směrovosti vazby a disperzní energie vysoké stabilizaci.

Dále jsme studovali vazebné interakce pomocí sigma díry u heteroboranů. Ukázali jsme, že zatímco sigma díra u CH_2S lokalizovaná na vrcholku dvojmocné síry je jen méně záporná než okolní elektrostatický potenciál (ESP) a již kladná u CF_2S , tak pentakoordinovaná síra v krystalu thiaboranů je kladně nabitá s méně pozitivní částí na vrcholku atomu a pěti vysoce kladně nabitými sigma dírami po stranách atomu (Obr. 2).

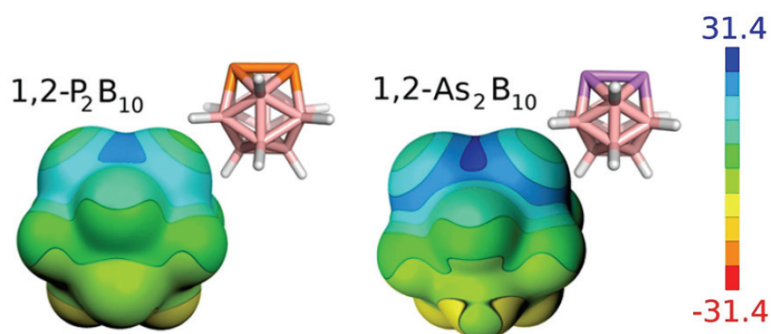


Obr. 2: Porovnání ESP a dipólových momentů (šípky) pro síru obsahující 2D organické sloučeniny a 3D anorganické heteroborany.

U těchto anorganických krystalů jsme kvantově mechanicky analyzovali všechny vazebné motivy thiaboranů s *exo*-substituovaným atomem chlóru, jódu a fenylovou skupinou. Zde jsme ukázali stěžejní roli nelineární chalcogenové B-S... π vazby. DFT-SAPT dekompozice celkové stabilizační energie pak odhalila dominantní úlohu disperzní energie následovanou elektrostatikou. Indukční člen pak napověděl o nezanedbatelné úloze přenosu náboje u těchto sloučenin.

K rozkrytí kompletního obrázku o interakcích heteroboranů pomocí sigma díry jsme provedli systematickou výpočetní studii většího setu 12 heteroboranových molekul obsahujících atom chalcogenu nebo pniktogenu v 3D *closo*-kleci a *exo*-substituovaného atomu halogenu u dikarbaboranů. Výsledky prokázaly, že sigma díry chalcogenů i pniktogenů u heteroboranů jsou vysoce kladně nabitě části na již kladně nabitých atomech a že jsou i kladněji nabitě než u organických molekul. V případě dvou pniktogenů je oblast s nejkladnějším ESP umístěna mezi oba atomy, jak je ukázáno na obrázku 4.

Bylo prokázáno, že síla sigma díry může být vylepšena: *i*) změnou atomového čísla, tj. substitucí Se místo S a As místo P, *ii*) změnou skeletonu klece – od desetivrcholové k dvanáctivrcholové kleci a *iii*) změnou chemického okolí, např. *exo*-substitucí halogenu v *para* pozici heteroatomu.



Obr. 4: Vypočtený ESP pro *closo*-1,2-P₂B₁₀H₁₀ a *closo*-1,2-As₂B₁₀H₁₀

Síla odpovídajících nekovalentních interakcí byla prozkoumána CCSD(T), DFT-D3 a DFT-SAPT výpočty komplexů vybraných heteroboranů s pěti akceptory sigma díry (benzen, trimethylamin, dimethyl ether, aceton a formamid). Bylo prokázáno, že všechny tyto molekuly tvoří vazby pomocí sigma díry jež jsou silnější než u sigma dírou vázaných organických molekul. U chalkogenové vazby jsme ukázali, že ji lze modulovat substitucí síry selenem, zatímco u pniktogenové vazby substituce fosforu arsenem nepřináší tížené vylepšení. Z výsledků DFT-SAPTu u obou typů vazeb je patrná dominantní role disperzní energie, následovaná elektrostatikou, bez příspěvku přenosu náboje.

4.2 Protein-ligandové interakce

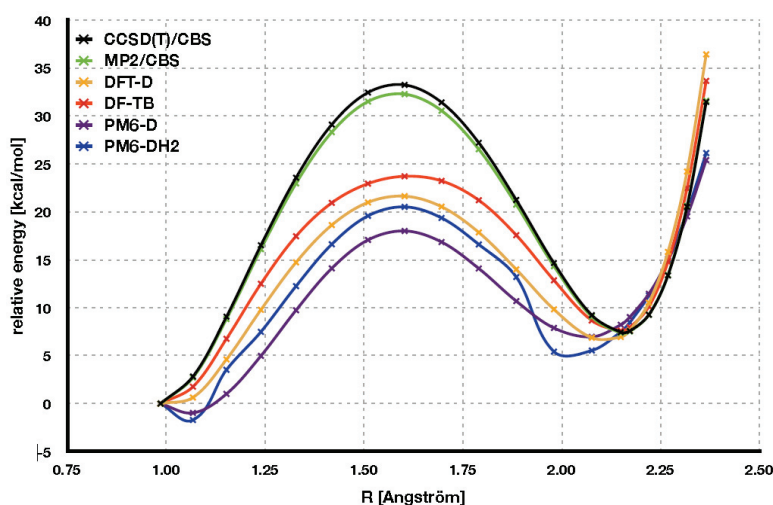
Tato druhá část výsledků se zaměřuje na detailní QM analýzu protein-ligandových interakcí tří medicínálně důležitých proteinů – HIV-1 proteasy, dvou isoenzymů sekretované aspartátové proteasy a karbanhydrasy II a IX. Zde jsme si hlavně kladli za cíl odhalit strukturní vlastnosti systémů, jež jsou experimentálně nedosažitelné: *i)* určení protonačních stavů a identifikace nejstabilnějších konformerů/tautomerů, *ii)* rozbor energetických příspěvků jednotlivých aminokyselin k celkové interakci ligandu v proteinu a *iii)* charakterizace vazby podobných či různých ligandů.

4.2.1 HIV-1 Proteasa

U dvou modelových systémů komplexu HIV-1 PR s ligandem – darunavir (DRV)¹² a peptidomimetický inhibitor KI2¹³, jsme aplikovali DFT-D a PM6-DH2 metody v hybridním schématu. Osvětlili jsme tak pravděpodobnost protonačních stavů těchto komplexů s ohledem na všechny orientace i konformace studovaných inhibitorů. Správnost použití kvantových metod byla zdůrazněna výskytem jevu přenosu vodíku v aktivním místě proteinu.

V případě dvou orientací ligandu v HIV-1 protease jsme demonstrovali symetrickou a energetickou spřízněnost párů protonačních variant. Výpočetně se nám podařilo identifikovat nejstabilnější protonační variantu katalytických aspartátů (Asp25/Asp25') u HIV-1 PR/DRV komplexu, což bylo následně potvrzeno experimentem. U komplexu HIV-1 PR s KI2 jsme odhalili monoprotonaci na Asp30 u karboxylového páru Asp30/Glu-P2' a také to, že akceptorem hydroxylové skupiny KI2 inhibitoru je OD2 kyslík z Asp25'.

Z metodologického hlediska z práce vyplynulo, že PM6 výpočty s QM částí rozšířenou na 8 Å (což je cca 1300 atomů) dávají kvalitativně stejné odpovědi jako DFT-D výpočty na menší části (QM obsahující cca 500 atomů). Ukázali jsme také, že při zvětšení QM části se riskuje, že strukturní změny ve velkých vzdálenostech od aktivního místa ovlivní celkovou energetiku. Dále jsme porovnávali několik QM metod pro popis přenosu vodíku u malého modelového systému karboxylového páru. Ukázali jsme zde, že DFT GGA funkcionály podceňují energetickou bariéru jen o něco méně než PM6-D a PM6-DH2 metody (Obr. 5) a tím jsme poukázali na potřebu zavedení nových korekcí PM6 metody.



Obr. 5: Graf energetických bariér proton transferu.

4.2.2 Sekretovaná aspartátová proteasa kvasinky *Candida Parapsilosis*

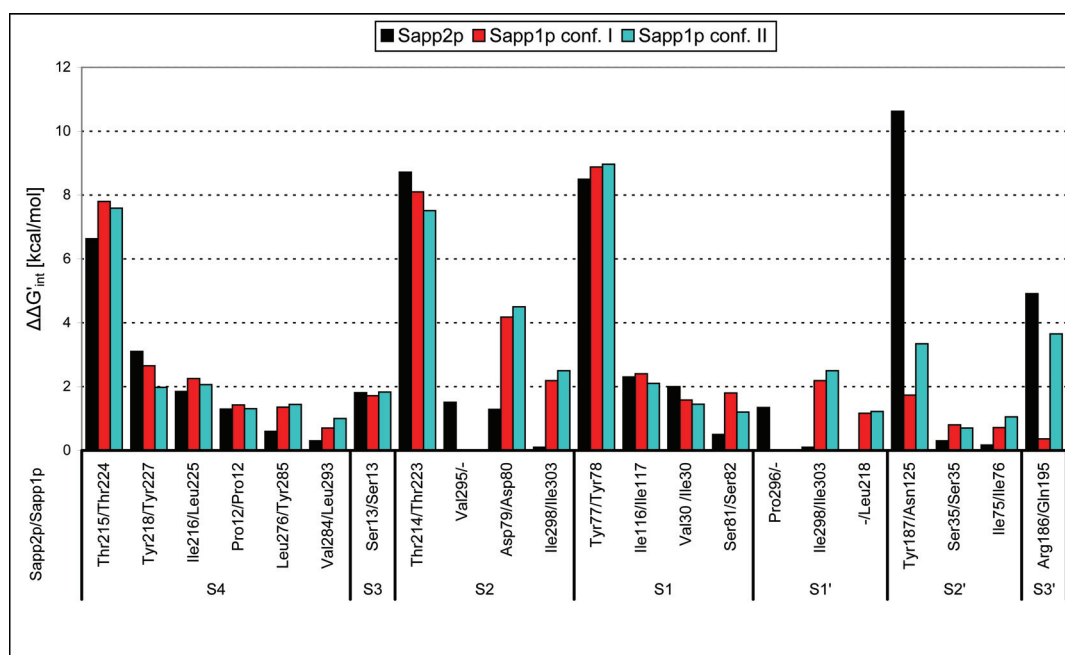
Zde jsme aplikovali již nově upravenou PM6 metodu v hybridním QM/MM schématu pro detailní porozumění rozdílu vazby pepstatinu A do dvou isoenzymů sekretované aspartátové proteasy – Sapp1p a Sapp2p.¹⁴

Konkrétně jsme usilovali o rozšíření závěrů z krystalografie o fakta, která nejsou dosažitelná ani z krystalů s velkým rozlišením: i) vyvrácení přítomnosti třetího protonu

v aktivním místě a *ii*) analýza energetických příspěvků všech důležitých aminokyselin v aktivním místě.

Zprvė jsme zde ukázali, že varianta sdíleného (třetího) protonu mezi katalytickými aspartáty (Asp32/Asp211) má interakční energii o 10 kcal/mol menší než varianta se dvěma protony v aktivním místě. To bylo ověřeno i optimalizací, kdy měl třetí proton problémy i ze stérických důvodů.

Pro stanovení energetických příspěvků všech aminokyselin v aktivním místě jsme užili tzv. virtuální glycinový sken (Obr. 6). Zkoumaná aminokyselina je zde substituována glycinem a energetický příspěvek jejího postranního řetězce ($\Delta\Delta G'_{int}$) je vypočten z rozdílu $\Delta G'_{int}$ s původní aminokyselinou a nové $\Delta G'_{int}$ s mutovaným glycinovým residuem. Interakční “volná” energie ($\Delta G'_{int}$) všech studovaných systémů byla stanovena na optimalizovaných QM částech jako single-point výpočet na PM6-D3H4/COSMO úrovni.



Obr. 6: Virtuální glycinový sken pro Sapp2p a Sapp1p aktivní místa s navázaným pepstatinem A.

Ačkoli byla síla interakce pepstatinu A s oběma isoenzymy téměř stejná, virtuální glycinový sken zachytil několik vzájemně se kompenzujících rozdílů vazby pepstatinu A v Sapp2p a Sapp1p.

4.2.3 Karboranové inhibitory karbanhydrasy (CA)

V této části jsme zkoumali vazbu dvou karboranových inhibitorů s *closo*- a *nido*-klecí do CAII isoenzymu.¹⁵ Zaměřili jsme se na teoretické vysvětlení: *i*) charakteru vazby těchto inhibitorů, *ii*) stability všech rotamerů a enantiomerů těchto látek a také na *iii*) analýzu energetických příspěvků jednotlivých aminokyselin v aktivním místě proteinu.

V první řadě jsme modelovali a výpočetně identifikovali nejstabilnější isomery z pěti možných rotačních isomerů *closo*-karboranového inhibitoru (**1a**) a ze čtyř možných isomerů *nido*-karboranového inhibitoru (**7a**), tj. dva enantiomery se dvěma možnými pozicemi vodíkového B-H-B můstku.

Vazba **1a** a **7a** ligandů se v CAII proteinu mírně liší a to v orientaci postranních řetězců aminokyselin interagujících jak s karboranovými klecemi tak také se sulfoamidovým spojovacím článkem. Abychom osvětlili tento rozdíl, fragmentovali jsme ligandy na dvě části a vypočítali jejich interakční “volnou” energii ($\Delta G'_{int}$) s aktivním místem CAII. Hodnoty vypočtené relativní vazebné volné energie byly ve shodě s experimentálně stanovenými hodnotami, přičemž sulfonamidová skupina interagovala významně silněji než karboranové klece, což je ve shodě s dřívějšími poznatky. Sulfonamidové části ligandů **1a** a **7a** přitom interagovaly se stejnou silou, ale rozdíl v interakční energiích (ΔE_{int}) *closo* a *nido* karboranové klece byl signifikantní. Ukázali jsme, že *closo* klec interagovala o 58,9 kcal/mol slaběji než *nido* klec, přičemž hlavní roli při vazbě hrála disperzní energie. Na druhou stranu, disperzní energie u *nido* klece tvořila jen 26,5% z celkové ΔE_{int} . Zásadním prvkem při interakci *nido* klece se tedy jeví být elektrostatický charakter vazby. Výsledná interakční “volná” energie obou klecí je však srovnatelná a to díky vysoké desolvatační penalizaci nabitě *nido* karboranové klece.

Role jednotlivých postranních řetězců všech aminokyselin aktivního místa byly kvantifikovány glycinovým skenem, který ukázal zásadní rozdíly ve vazbě obou ligandů. *Closo* karboranová klece interagovala pomocí disperzní energie a vytvářela jen slabé divodíkové vazby s nepolárními C-H skupinami. Naproti tomu, záporně nabitá *nido* klec interagovala s proteinem hlavně pomocí elektrostatiky a vytvářela velmi silné a krátké divodíkové vazby s polárními vodíky NH₂ skupin. Nejsilnějším příspěvkem k celkové vazebné energii ($\Delta\Delta G'_{int} = -4.1$ kcal/mol) byla přitom krátká divodíková vazba s Asn67 s H...H vzdáleností 1,7 Å.

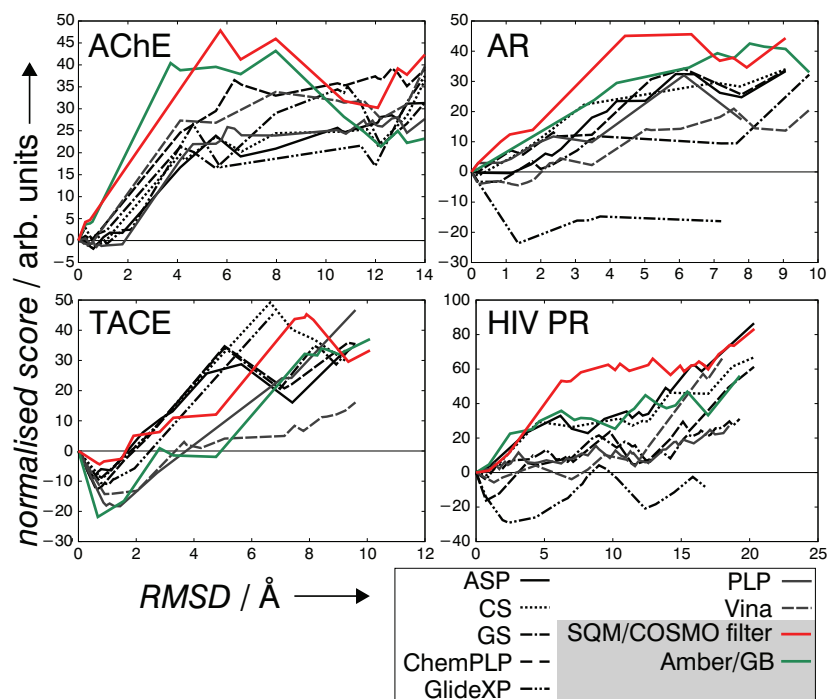
Ligand **1a** byl následně namodelován i ve vazbě s CAIX isoenzymem, jenž se liší od CAII ve tvaru dutiny aktivního místa způsobeném variací šesti aminokyselin. Přitom jsme ukázali odlišnou pozici *closo* karboranové klece v CAII a CAIX a pomocí glycinového skenu kvantifikovali jednotlivé vazebné příspěvky všech aminokyselin v této opačné části aktivního místa CAIX a porovnali s příspěvky aminokyselin CAII. Tím jsme představili strukturní a výpočetní analýzu aplikovatelnou ve strukturním návrhu (z angl. structure-based drug design) karboranových inhibitorů selektivních vůči CAIX isoenzymu, jenž je mimojiné spjatý s rakovinovým onemocněním.

4.3 SQM/COSMO filtr

V této práci jsme zjednodušením semiempirické skórovací funkce, tak aby byla použitelná ve virtuálním prohledávání látek (z angl. virtual screening), zavedli “SQM/COSMO filtr”. Ten se skládá ze dvou dominantních členů pro popis protein-ligandových nekovalentních interakcí, a to z interakční energie ve vakuu (ΔE_{int}) na PM6-D3H4X úrovni a ze změny solvatační energie při vazbě ligandu do proteinu ($\Delta\Delta G_{solv}$) na COSMO implicitní úrovni. Naše nové schéma zahrnuje pouze single-point výpočty struktur s relaxovanými pozicemi vodíků a to bez jakékoli další optimalizace. Chování SQM/COSMO filtru bylo testováno spolu s osmi známými skórovacími funkcemi (GlideScore XP, PLANTS PLP, AutoDock Vina, Gold ChemScore a GoldScore, ChemPLP, Astex Statistical Potential a AMBER/GB)¹⁶ na čtyřech nepříbuzných netriviálních protein-ligandových systémech – acetylcholin esterasa (AchE), TNF- α adaptovaný enzym (TACE), aldosová reduktasa (AR) a HIV-1 proteasea (HIV PR).^{17,13} Celkově bylo dokovacími algoritmy vygenerováno 2865 smysluplných rozlišných ligandových póz. Všechny tyto pózy byly přepočítány devíti skórovacími funkcemi a normalizované výsledné energie byly zaneseny do grafů v závislosti na RMSD krystalové pózy ligandu. Grafy byly následně zjednodušeny a přeloženy do jednoho výsledného grafu (Obr. 7). Pro kvantifikaci chování jednotlivých skórovacích funkcí jsme zavedli dvě kritéria – počet nesprávně vyhodnocených příznivých řešení (z angl. false-positive solutions) a $RMSD^{max}$ kritérium vztahující se k maximální hodnotě RMSD všech ligandových póz od krystalové struktury v předem definovaném rozpětí normalizovaného skóre.

Výsledky ukázaly jak chování SQM/COSMO filtru jenž splnil všechna kritéria předčí ostatní použité skórovací funkce. Za SQM/COSMO filtrem, majícím nejméně “false-positive” řešení, následuje Gold CS a ASP skórovací funkce.

Touto studií jsme tedy demonstrovali schopnost SQM/COSMO filtru rozpoznat správnou vazebnou pózu ligandu a také schopnost jít za tuto hranici a správně ohodnotit dokonce i velmi malé změny v geometrii navázaného ligandu.



Obr.7: Grafy závislosti normalizovaných hodnot skóre na RMSD hodnotách od krystalové pozice ligandů všech čtyř protein-ligandových systémů.

5. Závěr

Zaprvé jsme úspěšně aplikovali přesné kvantově mechanické metody na komplexy vázané sigma dírou, tj. na rozšířený dataset halogenově vázaných komplexů, na anorganické krystaly thiaboranů stabilizovaných chalkogenovou vazbou a na kompletní dataset heteroboranů interagujících se svými organickými partnery pomocí všech třech typů vazby pomocí sigma díry. Bylo ukázáno, že jediným způsobem jak správně rozkrýt kompletní obrázek o vazebné interakci sigma-díry je spřažení vlastností monomerů, tj. sigma děr, s vlastnostmi komplexů. Analýza energetických příspěvků k celkové stabilizační energii ukázala společnou roli polarizační a disperzní energie na stabilizaci halogenové vazby a nadhodila tak otázku adekvátnosti stávající IUPAC definice.¹⁸ Dále jsme se zabývali experimentálně známými heteroborany a studovali jejich dovednost tvořit chalkogenové a pniktogenové vazby. DFT-

SAPT rozložil jejich celkovou stabilizaci na jednotlivé příspěvky a ukázal dominující roli disperzní energie a elektrostatiky, následovanou indukční energií. Ukázali jsme a také kvantifikovali několik způsobů modulace těchto interakcí, což může být dále využito v krystalovém inženýrství nebo vývoji léčiv.

Zadruhé jsme ukázali jak lze ruku v ruce se strukturálními znalostmi efektivně využít teoretické výpočty pro detailní popis protein-ligandových systémů. Využitím QM/MM přístupu jsme kvantitativně popsali vazbu ligandu do proteinu, odkryli experimentálně nedostupné strukturální vlastnosti a vysvětlili fundamentální rozdíly vazby různých ligandů. Výhody kvantově mechanických metod byly zejména ukázány u protein-ligandových systémů s přenosem protonu, ionty kovu či s neobvyklými ligandy, tj. borany. Zavedli jsme virtuální glycinový sken, jenž dovoluje podrobnou analýzu energetických příspěvků postranních řetězců aminokyselin daného aktivního místa k celkové interakční energii proteinu s ligandem. Pomocí opravených semiempirických a DFT metod jsme také představili výpočetní protokol užitečný pro stanovení nejpravděpodobnějších protonačních stavů a identifikaci nejstabilnějších isomerů (konformerů/tautomerů) v biomolekulárních komplexech. Tato zjištění jsou velmi důležitá pro vytvoření správného výpočetního modelu studovaného systému, na němž pak závisí výsledky výpočtů vazebných afinit. Všechny prezentované výsledky jsou užitečné nejen pro pochopení interakcí a selektivity specifických ligandů ve vazbě do daného proteinového cíle, ale také pro další racionální návrh účinnějších/selektivnějších inhibitorů.

Následně je zaveden SQM/COSMO filtr – účinný nástroj pro spolehlivou identifikaci správných vazebných póz, založený na semiempirickém PM6D3H4X popisu protein-ligandových interakcí a na implicitní COSMO solvataci. SQM/COSMO filtr díky svým vlastnostem tj. obecnou použitelností, srovnatelností napříč chemickým spektrem a absencí potřeby systémově specifických parametrů, posunul hranice přesnosti skórovacích funkcí pro určování energie vazby ligandů v proteinech. Společně s časovou nenáročností dovolující výpočty tisíců dokovaných póz se námi předložený SQM/COSMO filtr může stát efektivním nástrojem pro zpřesňování v pozdějších fázích virtuálního prohledávání nebo nástrojem pro posuzování výkonu jiných skórovacích funkcí.

1. Introduction

Drug discovery (DD) process is not only very expensive but it also carries an extremely high risk of failure.¹ The use of various *in silico* techniques can help avert those failures in all pre-clinical phases of DD which 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 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 protein active site, thus preventing a native substrate from entering the site. The ligand potency is expected to be proportional to 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).

$$\Delta G_b = RT \ln K_i \quad (1)$$

The change of the binding free energy can be estimated computationally by using many tools, from statistics-based methods to physical chemistry-based approaches. In order to correctly describe all phenomena contributing to the binding energy, the methods must be able to deal with a sufficient large parts of the system (thousands of atoms) with high accuracy within a reasonable time. The methods must reliably describe all types of non-covalent interactions, *e.g.* hydrogen bonding, dispersion interaction or halogen bonding.³ Computational tools are not only important for the identification of drug candidates (“virtual screening”) but also for the selection of modifications that would improve the potency and other properties of the compounds and also for bringing a deeper insight into the mechanism of their action.

In contrast of widely used molecular mechanics (reviewed in Ref. 4), quantum mechanical (QM) methods are by the definition able to reliably describe non-covalent interactions.⁵ QM 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.

2. Aims of the Study

The thesis aims to show the ability of some QM-based approaches to contribute to the field of computer-aided drug design. It emphasizes outcomes of eight original papers that have been divided into three parts. The first part is devoted to high-level QM methods used to study accurately the strength and origin of the stabilization for σ -hole bonded model systems, going from halogen through chalcogen to pnictogen bonding. The second part explores protein-ligand binding by using a hybrid QM/MM approach. It aims mostly to unveil features of the structure which are not accessible to the crystallographic experiments. The last part introduces an effective semiempirical quantum mechanical (SQM) based tool applicable for virtual screening.

3. Methods

The thesis is based on QM methods with different accuracy/computational cost performance determining their use in presented studies. It comprises particularly from highly accurate CCSD(T)/CBS calculations⁶ used in model systems as benchmark data, through suitable DFT-SAPT decomposition⁷ of the interaction energy, to the fast DFT and semiempirical PM6 methods augmented with empirical dispersion correction and hydrogen- and halogen- bond corrections.⁸ All calculations of the protein-ligand complexes comprising thousands of atoms are speeded up by using a hybrid QM/MM approach⁹ with water environment approximated by a combination of implicit solvent model¹⁰ and explicitly treated important structure water molecules.

4. Results and discussion

4.1 Nature of σ -hole bonding

In the first part we aimed to shed light on the nature of halogen bonding comprehensively. We made up a significantly large dataset of 128 halogen- or dihalogen- bonded complexes of different size and origin. Thus we covered 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.

Firstly, we described σ -holes of all halogenated subsystems in terms of size and magnitude.¹¹ Secondly, we calculated total stabilization energies of all complexes at DFT-D3 and CCSD(T)/CBS level and performed the DFT-SAPT decomposition.

It was shown that all studied subsystems possess a positive σ -hole. 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. In the case of the dihalogen bonding, the magnitude increases with the decreasing atomic number of the second halogen. When relating the properties of monomers with properties of complexes (Figure 1), we showed that the stabilization energy surprisingly correlates with the magnitude of the σ -holes only weakly ($R=0.52$). However when we selected the most stable complex of particular halogenated monomers, the correlation coefficient increased to 0.77. The magnitude of σ -hole therefore informs about the ability of the monomer to create halogen bond rather than about the strength of the halogen-bonded complex.

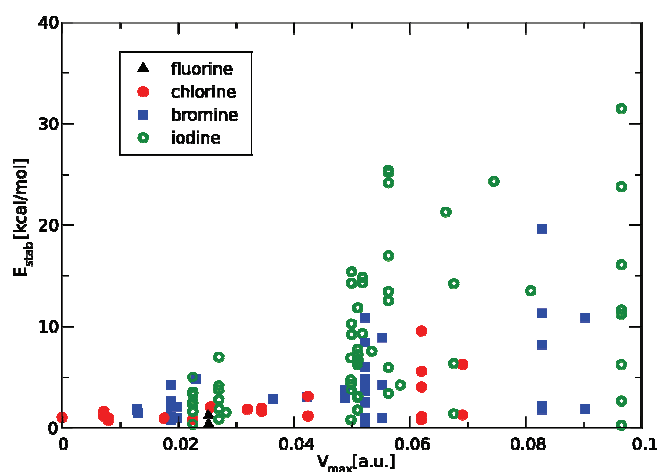


Fig. 1: The dependence of the stabilization energy E_{stab} on the magnitude of the σ -hole V_{max} .

The results of DFT-SAPT decomposition showed that the most dominant contribution to the total stabilization energy was in 62% of complexes the electrostatic energy, whereas in remaining 38% of cases the dominant term was 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.

In the second part we studied σ -hole bonding of heteroboranes. We showed that while the σ -hole localized at the top of divalent sulphur atom in CH_2S is just less negative than

the surrounding electrostatic potential (ESP) and already positive in the case of CF_2S , the pentacoordinated sulphur atom in crystallized thiaborane structure is positively charged with the less positive area at the top of the atom and five highly positive σ -holes on its sides (Fig. 2)

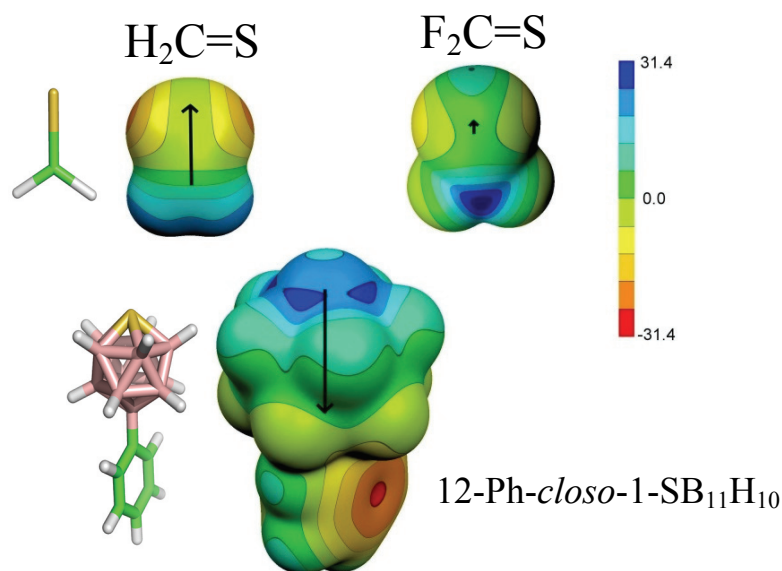


Fig. 2: Comparison of calculated ESPs and dipole moments (arrows) for a sulphur-bound 2D organic structures and 3D inorganic boron hydrid.

We performed quantum mechanical analysis of all binding motifs in inorganic crystals of thiaboranes with an *exo*-substituted chlorine atom, iodine atom and phenyl ring. We showed the pivotal role of non-linear chalcogen B-S... π bond. The DFT-SAPT decomposition of total stabilization energy revealed the dominant role of dispersion contribution, followed by electrostatic term and induction, showing not negligible charge transfer phenomenon.

To elucidate the complete picture of the σ -hole bonding of heteroboranes we performed a systematic computational study on bigger dataset of 12 *closo*-heteroborane molecules in which chalcogen and also pnicogen atoms were incorporated in the 3D cages, whereas halogens were considered as *exo*-substituents of dicarbaboranes. The results showed that the σ -holes of chalcogens and pnicogens are highly positive areas formed on already positively charged atoms and are both more positive in heteroboranes than in organic molecules. In the case of two pnicogens, the area of the most positive ESP is placed between these atoms as shown in Figure 4.

We showed that the magnitude of σ -holes can be tuned by *i*) changing the atomic number, *e.g.* going from S to Se and from P to As, *ii*) changing the skeleton of the borane cage, *i.e.*

going from 10-vertex to 12-vertex cage and *iii*) changing the chemical environment, *e.g.* by *exo*-substitutions of halogen atoms (instead of hydrogens) in *para* position to the heteroatom.

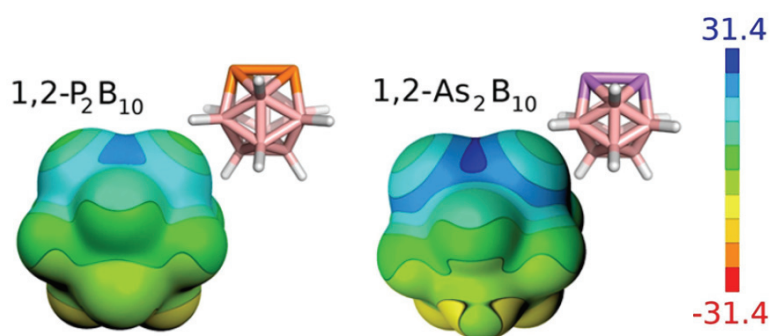


Fig. 4: ESP calculated on *closo*-1,2-P₂B₁₀H₁₀ and *closo*-1,2-As₂B₁₀H₁₀

The strength of non-covalent interactions of selected boron clusters with five σ -hole acceptors (benzene, trimethylamine, dimethyl ether, acetone and formamide) was elucidated by CCSD(T), DFT-D3 and DFT-SAPT calculations and it was shown that these molecules can form very strong σ -hole bonds that are stronger than in other neutral σ -hole bonded organic complexes. It was shown that the chalcogen bonding of heteroboranes can be modulated by S to Se substitution, whereas P-to-As substitution has smaller impact on modulation in pnictogen bonding. DFT-SAPT decomposition revealed the dominant role of dispersion energy, followed by electrostatic term in both cases. Small induction energy contributions showed that charge transfer does not contribute.

4.2 Protein-ligand binding

The second part of the thesis focused on a detailed quantum mechanical analysis of protein-ligand interactions of three medically important targets – HIV-1 protease, two isoenzymes of secreted aspartic protease and carbonic anhydrase II and IX. We aimed to unveil features of the structure which are not accessible to the crystallographic experiment: *i*) the determination of the protonation states and the identification of the most stable conformers/tautomers *ii*) the dissection of energy contributions of individual amino acids toward the total interaction and *iii*) the characterization of the nature of different binding of similar/different ligands.

4.2.1 HIV-1 Protease

In two model systems of HIV-1 PR/inhibitor complex, particularly with inhibitors darunavir (DRV)¹² and peptidomimetic KI2 inhibitor¹³, we applied DFT-D and PM6-DH2 methods in a hybrid scheme with MM to shed light on the probability of all protonation variants of complexes with regard to all orientations and conformations of studied ligands. The need of QM-based methods was underlined by the occurrence of proton transfer phenomena.

We showed 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 successfully identified the most stable protonation variant of catalytic aspartic dyad (Asp25/Asp25') of HIV-1 PR/DRV complex that was confirmed by the experiment. In the case of HIV-1 PR/KI2 complex we revealed that the Asp30/Glu-P2' carboxylate pair is monoprotinated 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'.

From the methodological point of view we concluded 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 the size of QM region (up to 500 atoms). We showed that bigger QM parts increase the risk of unrelated distant structural changes that can affect the energetics of the active site. Moreover we compared QM methods on the small model of carboxylate pair and showed that proton transfer barriers are underestimated by DFT using GGA functional and even more by the PM6-D and PM6-DH2 method (Fig. 5). We thus pointed to the need for better corrections for PM6 method.

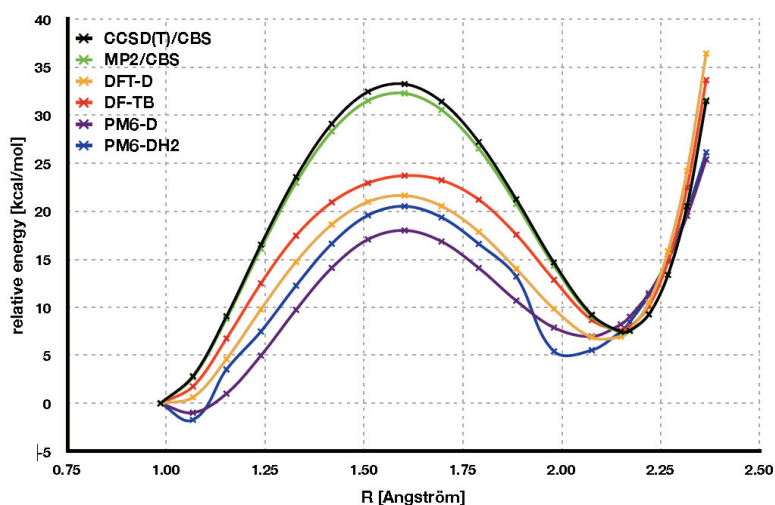


Fig. 5: The graph of proton-transfer barriers calculated in several QM-based methods.

4.2.2 Secreted Aspartic Protease of *Candida Parapsilosis*

We employed newly corrected PM6 method in a hybrid scheme to understand the difference of binding of the pepstatin A ligand to two *C. parapsilosis* isoenzymes, Sapp1p and Sapp2p.¹⁴ We specifically augmented the crystallographical findings by features not accessible even in high resolution: *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.

We calculated that the shared (third) proton between catalytic aspartate dyad (Asp32/Asp211) has far less favorable interaction energy (by about 10 kcal/mol) with the pepstatin inhibitor than the model with only two protons in the active site. It was also proved by steric reasons revealed from the optimization.

To dissect the energy contributions of the all important amino acids in the active site we applied a virtual glycine scanning (Fig. 6). The interacting amino acids in the active site were substituted with glycine and 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 studied systems were determined on optimized structures as single-point energies at the PM6-D3H4/COSMO level.

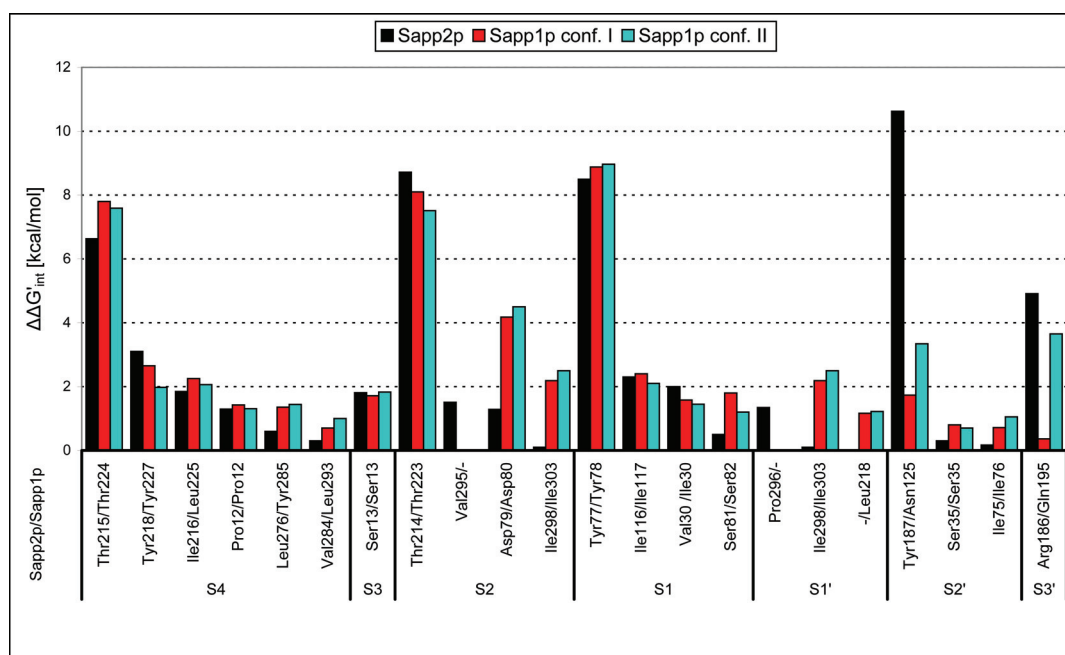


Fig. 6: The virtual glycine scan for of Sapp2p and Sapp1p active sites in binding with pepstatin A

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.

4.2.3 Carborane-based Inhibitors of Carbonic Anhydrases (CA)

We studied the nature of binding of two carborane-based sulfamide inhibitors bearing *closo*- and *nido*-carborane cages to CAII isoenzyme.¹⁵ We aimed to theoretically explain: *i*) the nature of binding of these carborane-based inhibitors, *ii*) the stability of all possible rotamers and enantiomers of studied inhibitors and *iii*) contributions of all important aminoacids in the active site of CAII.

First of all we modeled and computationally identified the most stable isomers from all five possible rotational isomers (rotamers) of *closo*-carborane inhibitor (**1a**) and four possible isomers of *nido*-based inhibitor (**7a**), *i.e.* two enantiomers combined with two positions of the B-H-B bridge.

The binding of **1a** and **7a** to the CAII protein slightly differed in the orientation of amino acid sidechains interacting with carborane cages and also in the interaction of sulfamide linker. We thus fragmented ligands into two parts and calculated their interaction ‘free’ energies ($\Delta G'_{int}$) with the CAII active site. The results revealed a good agreement of calculated and experimental relative binding free energies and the $\Delta G'_{int}$ of sulfonamide moiety was significantly stronger than of the carborane cages that is in agreement with previous findings. However 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. It was shown that 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.

We also quantified the roles of individual amino acid sidechains in the active site by virtual glycine scan procedure and showed the significant differences. *Closo*-carborane cage were bound mainly via dispersion interactions and formed only very weak dihydrogen bonds with non-polar C-H groups. Contrarily, negatively charged *nido*-carborane interacted with the protein mainly via electrostatic interactions and formed very strong and short dihydrogen

bonds with polar hydrogens of NH₂ groups. The biggest contribution to the total binding ($\Delta\Delta G'_{int} = -4.1$ kcal/mol) was mediated by a short dihydrogen bond with Asn67 with the H...H distance of 1.7 Å.

We also modeled the binding of compound **1a** into CAIX isoenzyme that differs from CAII in a shape of the active site cavity caused by variations of six amino acids. We showed the different position of the *closo*-cage in CAII and CAIX. The interacting contributions of all important amino acids in the opposite site of the CAIX active site were determined by virtual glycine scan and compared with CAII. We thus reported structural and computational analysis applicable to structure-based design of carborane compounds with selectivity towards the cancer-specific CAIX isoenzyme.

4.3 The SQM/COSMO filter

We simplified the SQM-based scoring function to make it usable in virtual screening and defined the SQM/COSMO filter energy. It considers 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. Our novel scheme consists of a single-point rescoring of hydrogen-relaxed structures with no additional optimization of the systems. We compared its performance with eight known scoring functions (GlideScore XP, PLANTS PLP, AutoDock Vina, Gold ChemScore and GoldScore, ChemPLP, Astex Statistical Potential and AMBER/GB)¹⁶ on four unrelated difficult-to-handle protein ligand systems – acetylcholine esterase (AChE), TNF- α converting enzyme (TACE), aldose reductase (AR) and HIV-1 protease (HIV PR).^{17,13} We generated 2,865 sensible and non-redundant alternative ligand poses in total by docking algorithms. All the poses were re-scored by nine scoring functions and resulting scores were normalized and plotted against the RMSD of the crystal structure. The resulting clouds of points were simplified to a single graph (see Fig. 7). In order to quantify the performance of studied scoring functions we introduced two criteria - the number of false-positive solutions and $RMSD^{max}$ referring about the maximum value of RMSD from the crystal structure revealed within all poses ranked in a defined interval of the normalized scores.

The results showed that the SQM/COSMO filter behaviour met all conditions at superior level unlike the others scoring functions with the lowest number of false positives, followed by slightly worse performance of Gold CS and ASP.

We thus 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.

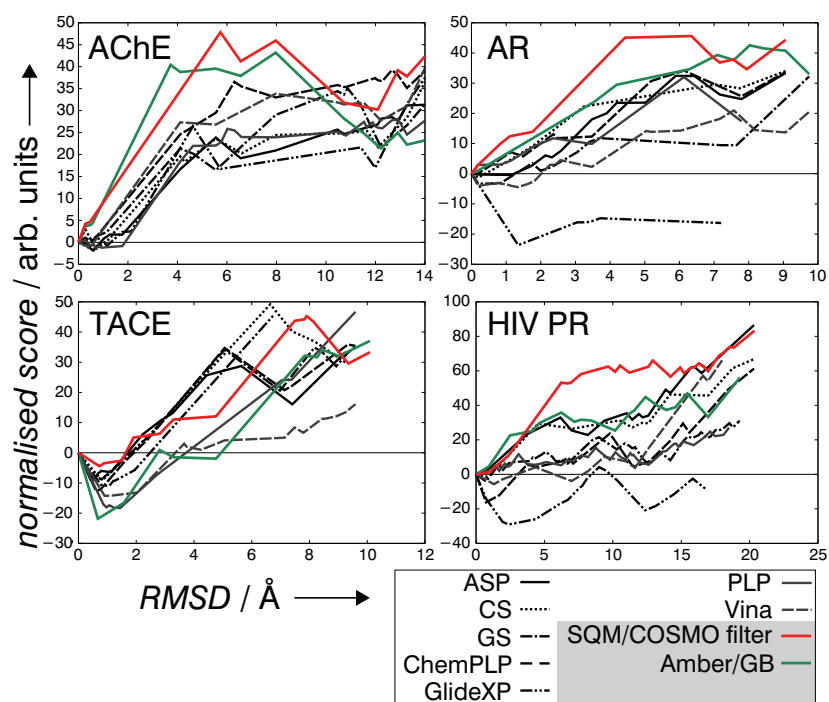


Fig.7: The plots of normalized scores against RMSD values for all four protein-ligand systems

5. Conclusions

We successfully applied accurate quantum mechanical methods on the σ -hole bonded complexes, *i.e.* on extended datasets of halogen bonded complexes, inorganic 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. It was 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 analysis of contributions to the total stabilization energies has revealed the concert action of polarization and dispersion energies to the stabilization of halogen bonding and thus has opened the question of adequacy of the recent IUPAC definition of halogen bonding.¹⁸ We also tackled experimentally-known heteroboranes in order to examine their ability to form chalcogen and pnicogen bonding. DFT-SAPT decomposed their total stabilization energies and have shown the dominating role of dispersion and electrostatic energy, followed by induction. 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.

We have shown how structural information coupled with theoretical calculations can be effectively used for detail studies of protein-ligand systems. We used 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 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 introduced the novel computational protocol using corrected SQM and DFT methods 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. 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 understanding and selectivity of the ligand binding to the particular protein target as well as for further rational design of more potent/selective inhibitors

We have introduced the SQM/COSMO filter – the effective tool for reliable native pose identification based on the semiempirical PM6D3H4X description of protein–ligand interactions and implicit COSMO solvation. 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.

References:

- (1) Adams, C. P.; Brantner, V. V. *Health Econ* **2010**, *19*, 130.
- (2) Kapetanovic, I. M. *Chem-Biol Interact* **2008**, *171*, 165.
- (3) Hobza, P.; Müller-Dethlefs, K. *Non-covalent Interactions: Theory and Experiment*; Royal Society of Chemistry, 2010.
- (4) Gohlke, H.; Klebe, G. *Curr Opin Struc Biol* **2001**, *11*, 231.
- (5) Lepsik, M.; Rezac, J.; Kolar, M. et.al. *Chempluschem* **2013**, *78*, 921
- (6) Lee, E. C.; Kim, D.; Jurecka, P. et.al. *J Phys Chem A* **2007**, *111*, 3446
- (7) Jansen, G.; Hesselmann, A. *J Phys Chem A* **2001**, *105*, 11156; Podeszwa, R.; Bukowski, R.; Szalewicz, K. *J Phys Chem A* **2006**, *110*, 10345; Hesselmann, A. *J Phys Chem A* **2011**, *115*, 11321.
- (8) Grimme, S.; Antony, J.; Ehrlich, S., et.al. *J Chem Phys* **2010**, *132*; Stewart, J. J. P. *J Mol Model* **2009**, *15*, 765; Jurecka, P.; Cerny, J.; Hobza, P., et.al. *J Comput Chem* **2007**, *28*, 555; Korth, M.; Pitonak, M.; Rezac, J. et.al. *J Chem Theory Comput* **2010**, *6*, 344; Rezac, J.; Hobza, P. *J Chem Theory Comput* **2012**, *8*, 141.
- (9) Svensson, M.; Humbel, S.; Froese, R. D. J., et.al. *J Phys Chem* **1996**, *100*, 19357.
- (10) Tsui, V.; Case, D. A. *Biopolymers* **2001**, *56*, 275; Klamt, A.; Schuurmann, G. *J Chem Soc-Perkin Transactions 2* **1993**, 799.
- (11) Kolar, M. H.; Carloni, P.; Hobza, P. *Phys Chem Chem Phys* **2014**, *16*, 19111.
- (12) Kozisek, M.; Lepsik, M.; Saskova, K. G. et.al. *Febs J* **2014**, *281*, 1834.
- (13) Brynda, J.; Rezacova, P.; Fabry, M. et.al. *J Med Chem* **2004**, *47*, 2030
- (14) Dostal, J.; Brynda, J.; Hruskova-Heidingsfeldova, O. et.al. *J Enzym Inhib Med Ch* **2012**, *27*, 160; Dostal, J.; Brynda, J.; Hruskova-Heidingsfeldova, O., et.al. *J Struct Biol* **2009**, *167*, 145.
- (15) Brynda J., M. P., Šícha V., Fábry M., et.al. *Angew Chem Int Ed* **2013**, *52*, 13760.
- (16) Friesner, R. A.; Murphy, R. B.; Repasky, M. P. et.al. *J Med Chem* **2006**, *49*, 6177; Korb, O.; Stutzle, T.; Exner, T. E. *J Chem Inf Model* **2009**, *49*, 84; Trott, O.; Olson, A. J. *J Comput Chem* **2010**, *31*, 455; Eldridge, M. D.; Murray, C. W.; Auton, T. R.; et.al. *J Comput Aid Mol Des* **1997**, *11*, 425; Jones, G.; Willett, P.; Glen, R. C.; et.al. *J Mol Biol* **1997**, *267*, 727; Mooij, W. T. M.; Verdonk, M. L. *Proteins* **2005**, *61*, 272.
- (17) Dvir, H.; Wong, D. M.; Harel, M.; Barril, X.; et.al. *Biochemistry* **2002**, *41*, 2970; Bandarage, U. K.; Wang, T.; Come, J. H. et.al. *Bioorg Med Chem Lett* **2008**, *18*, 44; Steuber, H.; Heine, A.; Klebe, G. *J Mol Biol* **2007**, *368*, 618.
- (18) Desiraju, G. R.; Ho, P. S.; Kloo, L.; et.al. *Pure Appl Chem* **2013**, *85*, 1711

Curriculum vitae
ADAM PECINA

Department of Computational Chemistry, Institute of Organic Chemistry & Biochemistry (IOCB) and Gilead Science & IOCB Research Center
Academy of Sciences of the Czech Republic, Flemingovo nám. 2, CZ-166 10 Prague 6, Czech Republic
Phone: (+420) 220 410 318, Adam.Pecina@uochb.cas.cz

Education

- 2013 Doctoral exam, Faculty of Science, Charles University in Prague
- since 2009 Ph.D. student in Modelling of Chemical Properties of Nano- and Biostructures,
Faculty of Science, Charles University in Prague
thesis: “Quantum chemical approach for in silico drug design“
- 2007-2009 MSc. in Biophysical Chemistry, Faculty of Science, Palacký University
Olomouc, thesis: “Quantum chemical investigation of thermodynamic
characteristics of formation of methylated DNA basis in vacuo“
- 2004-2007 BSc. in Applied Chemistry, Faculty of Science, Palacký University Olomouc,
thesis: “Chemiluminescent determination of antioxidant capacity of seminal
plasma“

Experience

- 2010 World Class University Project, Pohang University of Science and
Technology, Pohang, Republic of Korea, three-month stay
- since 2008 Research group of prof. Pavel Hobza, Institute of Organic Chemistry and
Biochemistry, Academy of Sciences of the Czech Republic, v.v.i.

Selected 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.