

Selected Publications / Seznam Publikací

1. Sigwalt, D.; Šekutor, M.; Cao, L., Zavalij, P. Y.; **Hostaš, J.**; Ajani, H.; Hobza, P.; Mlinarić-Majerski, K.; Glaser, R.; Isaacs, L.: Unraveling the Structure-Affinity Relationship Between Cucurbit[*n*]urils (*n* = 7, 8) and Cationic Diamondoids, *J. Am. Chem. Soc.* **2017**, DOI: 10.1021/jacs.7b00056.
2. **Hostaš, J.**; Sigwalt, D.; Šekutor, M.; Ajani, H.; Dubecký, M.; Řezáč, J.; Zavalij, P. Y.; Cao, L.; Wohlschlagel, C.; Mlinarić-Majerski, K.; Isaacs, L.; Glaser, R.; Hobza, P.: "A Nexus between Theory and Experiment: Non-empirical Quantum Mechanical Computational Methodology Applied to Cucurbit[*n*]uril•Guest Binding Interactions", *Chem. Eur. J.* **2016**, 22, 17226-17238, cover page.
3. Rai, S. K.; Singh, P.; Kumar, R.; Tewari, A. K.; **Hostaš, J.**; Gnanasekaran, R.; Hobza P.: "Experimental and Theoretical Study of the Assessment of the Conformational Stability of Polymethylene-Bridged Heteroaromatic Dimers: A case of Unprecedented Folding", *Cryst. Growth Des.* **2016**, 16, 1179-1180.
4. Cao, L.; Škalamera, D.; Zavalij, P. Y.; **Hostaš, J.**; Hobza, P.; Mlinarić-Majerski, K.; Glaser, R.; Isaacs L.: "Influence of hydrophobic residues on the binding of CB[7] toward diammonium ions of common ammonium-ammonium distance", *Org. Biomol. Chem.* **2015**, 13, 6249.
5. **Hostaš, J.**; Jakubec, D.; Laskowski, R. A.; Gnanasekaran, R.; Řezáč, J.; Vondrášek, J.; Hobza P.: "Representative Amino Acid Side-Chain Interactions in Protein-DNA Complexes: A Comparison of Highly Accurate Correlated Ab Initio Quantum Mechanical Calculations and Efficient Approaches for Applications to Large Systems", *J. Chem. Theory Comput.* **2015**, 11, 4086-4092.
6. Jakubec, D.; **Hostaš, J.**; Laskowski, R. A.; Hobza, P.; Vondrášek, J.: "Large-Scale Quantitative Assessment of Binding Preferences in Protein-Nucleic Acid Complexes", *J. Chem. Theory Comput.* **2015**, 11, 1939-1948.
7. Kolář, M.; **Hostaš, J.**; Hobza, P.: "The strength and directionality of a halogen bond are co-determined by the magnitude and size of the sigma-hole", *Phys. Chem. Chem. Phys.* **2014**, 16, 9987-9996.
8. **Hostaš, J.**; Řezáč, J.; Hobza, P.: "On the Performance of the Semiempirical Quantum Mechanical PM6 and PM7 Methods for Noncovalent Interactions", *Chem. Phys. Lett.* **2013**, 568, 161-166.

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

Charles University in Prague, Faculty of Science
Department of Physical and Macromolecular Chemistry

Doktorský studijní program: Modelování chemických
vlastností nano- a biostuktur

Ph.D. study program: Modeling of chemical properties of
nano- and bioststructures

Autoreferát disertační práce
Summary of the Ph.D. Thesis



Přesné Kvantově Mechanické Výpočty Nekovalentních
Interakcí: Racionalizace Rentgenových Krystalových
Geometrií Aparátrem Kvantové Chemie

Accurate Quantum Mechanical Calculations on Noncovalent
Interactions: Rationalization of X-ray Crystal
Geometries by Quantum Chemistry Tools

Mgr. Jirí Hostaš

Školitel: Prof. Ing. Pavel Hobza, DrSc., FRSC, Dr. h. c.
Konzultant: Doc. RNDr. Jan Řezáč, Ph.D.

Praha, 2017

Abstrakt

Spolehlivá a jednoduše aplikovatelná pravidla jsou potřebná v oblasti biochemie, supramolekulární chemie i materiálových vědách. Zároveň množství informací, které můžeme získat z rentgenových krystalových struktur o povaze rozpoznávacích procesů, je omezené. Lepší pochopení nekovalentních interakcí, které hrají nejdůležitější roli, je potřebné pro přezkoumání univerzálních pravidel, řídicích jakékoliv rozpoznávací procesy. V této práci je prezentován systematický vývoj a studium přesnosti výpočetních metod, doplněný aplikacemi na systémech bílkovina•DNA a hostitel•host. Ne-empirické kvantově mechanické nástroje (metody DFT-D, MP2.5, CCSD(T) *atd.*) byly využity v několika projektech. Našli a potvrdili jsme existenci unikátních nízko ležících interakčních energií, vzdálených od zbývajících distribucí v několika párech aminokyselina–báze, které otevírají cestu k univerzálním pravidlům řídicím selektivní navázání jakékoliv sekvence DNA. Dále byly v několika případech provedeny predikce a ověřeny změny Gibbsovy energie (ΔG) a jejich komponentů a nakonec byly pečlivě porovnány s experimenty. Stanovili jsme, že molekula cholinu je vázána o 2.8 kcal/mol silněji (vypočtením ΔG) než acetylcholin v samo- uspořádané tří helikální rigidní kleci, odpovídající Kch/Kach = 109, což je v poměrně dobrém souladu s experimentální hodnotou 20. Nakonec byla popsána výborná korelace mezi teoretickou a experimentální ΔG pro systémy hostitel•host s molekulou cucurbit[7]urilů (CB[n]). Byla provedena predikce, že vazba u CB[7]•Diam-4,9-di(NMe₂propanoNH₃) by se mohla stát novým světovým rekordem v nekovalentní vazbě. Výše zmíněný derivát diamantanu je nyní připravován experimentálně. Tyto výsledky jasně demonstrují pevnou pozici výpočetní chemie jako komplementárního zdroje informací pro experimenty.

Curriculum vitae

Education

- Ph.D. (group of prof. Pavel Hobza), Modeling Chemical Properties of Nano- and Biostructures, Faculty of Natural Sciences, Charles University in Prague, Czech Republic 2012–present
- Mgr. (M.Sc.), Modeling Chemical Properties of Nano- and Biostructures, Faculty of Natural Sciences, Charles University in Prague, Czech Republic 2010–2012
- Bc., Chemistry, Faculty of Chemical Engineering, Institute of Chemical Technology in Prague, Czech Republic 2007–2010

Additional Educational Activities

- Summer school: "Dynamical Processes in Atoms, Molecules and Solids", 4 days, Prague, Czech Republic, **2015**.
- Summer school: "European Summerschool in Quantum Chemistry 2015", 2 weeks, Sicily, Italy **2015**.
- Summer school: "Quantum Monte Carlo and the CASINO program IX", 1 week, Valico Sotto, Italy **2014**.
- Athens program, "On Quantum, Chaos and Daemons", 1 week, Ecole des Ponts, Paris, France **2010**.
- Course on Biomolecular Interactions by Computational Chemistry Tools, National Centre for Biomolecular Research, Brno, Czech Republic **2008**.

Poster Presentations

- "Accurate DFT-D Calculations in Small Basis Set" International CECAM-Workshop & Tutorial on Approximate Quantum Methods in the *ab initio* World, Beijing, China **2016**.
- "Non-empirical Quantum Mechanical Computational Methodology Applied to Cucurbit[7]uril•Guest Binding Interactions", Theory and Applications of Computational Chemistry, Seattle, USA **2016**.
- "Binding Preference in Protein-Nucleic Acid Complexes: Representative Amino Acid Side-Chain Interactions", European Summerschool in Quantum Chemistry 2015", Sicily, Italy **2015**.
- "On the Performance of the Semiempirical Quantum Mechanical PM6 and PM7 Methods for Noncovalent Interactions", Very Accurate and Large Computations and Applications 2013, Fevik, Norway **2013**.
- "Calculation of highly accurate total energies of biomacromolecules from energies of smaller fragments", Conference on Computer Simulation and Theory of Macromolecules, Huenfeld, Germany **2009**.

6 Použitá literatura

- (1) Laskowski, R. A.; Thornton, J. M. Understanding the molecular machinery of genetics through 3D structures. *Nature Reviews Genetics* **2008**, *9*, 141-151.
- (2) Grimme, S. Accurate description of van der Waals complexes by density functional theory including empirical corrections. *J. Comput. Chem.* **2004**, *25*, 1463-1473.
- (3) Berka, K.; Laskowski, R.; Riley, K. E.; Hobza, P.; Vondrasek, J. Representative Amino Acid Side Chain Interactions in Proteins. A Comparison of Highly Accurate Correlated ab Initio Quantum Chemical and Empirical Potential Procedures. *J Chem Theory Comput* **2009**, *5*, 982-992.
- (4) Laguna, J.; Mukhopadhyay, P.; Chakrabarti, S.; Isaacs, L. The cucurbit n uril family. *Angew. Chem.-Int. Edit.* **2005**, *44*, 4844-4870.
- (5) Rezac, J.; Hobza, P. Benchmark Calculations of Interaction Energies in Noncovalent Complexes and Their Applications. *Chemical Reviews* **2016**, *116*, 5038-5071.
- (6) Rezac, J. Cuby: An Integrative Framework for Computational Chemistry. *J. Comput. Chem.* **2016**, *37*, 1230-1237.
- (7) Sedlak, R.; Janowski, T.; Pitonak, M.; Rezac, J.; Pulay, P.; Hobza, P. Accuracy of Quantum Chemical Methods for Large Noncovalent Complexes. *J Chem Theory Comput* **2013**, *9*, 3364-3374.
- (8) Halkier, A.; Helgaker, T.; Jorgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. Basis-set convergence in correlated calculations on Ne, N-2, and H2O. *Chem Phys Lett* **1998**, *286*, 243-252.
- (9) Singh, J.; Thornton, J. M. SIRIUS - An Automated-Method for The Analysis of the Preferred Packing Arrangements Between Protein Groups. *Journal of Molecular Biology* **1990**, *211*, 595-615.

Abstract

There is a need for reliable rules of thumb for various applications in the area of biochemistry, supramolecular chemistry and material sciences. Simultaneously, the amount of information, which we can gather from X-ray crystal geometries about the nature of recognition processes, is limited. Deeper insight into the noncovalent interactions playing the most important role is needed in order to revise these universal rules governing any recognition process. In this thesis, systematic development and study of the accuracy of the computational chemistry methods followed by their applications in protein•DNA and host•guest systems, are presented. The non-empirical quantum mechanical tools (DFT-D, MP2.5, CCSD(T) etc. methods) were utilized in several projects. We found and confirmed unique low lying interaction energies distinct from the rest of the distributions in several amino acid–base pairs opening a way toward universal rules governing the selective binding of any DNA sequence. Further, the predictions and examination of changes of Gibbs energies (ΔG) and its subcomponents have been made in several cases and carefully compared with experiments. We determined that the choline guest is bound 2.8 kcal/mol stronger (calculated ΔG) than acetylcholine to self-assembled triple helicate rigid cage, corresponding a $K_{ch}/K_{ach} = 109$ that is in fairly good correlation with the experimental value of 20. Finally, excellent correlation between theoretical and experimental ΔG has been reported ($\rho^2 = 0.84$) for cucurbit[n]uril (CB[n]) host•guest systems. Here, prediction has been made that binding in CB[7]•Diam-4,9-di(NMe₂propanoNH₃) complex could become next world record in the world of noncovalent interactions. This diamantane derivative is now being synthesized. Clearly, these findings demonstrate that the computational chemistry has a solid position as the complementary source of information to the data obtained from the experiments.

1 Introduction

The understanding of binding selectivity principles in both, natural and artificial complexes, would constitute great leap forward in many applications in biochemistry, supramolecular chemistry, material sciences or molecular biology. Recently, increasing amount of X-ray crystallographic data opened a space for many studies in this direction.¹

In the presented thesis, the main focus is being put on the description and modelling of building blocks of biopolymers (proteins and DNA) and of various biomimetic complexes in its natural environments as well as isolated. While the experiments successfully assess the strength of the binding in many instances, they mostly provide only limited information about the nature of the noncovalent interactions involved in the binding processes. The computational results can complement, provide insight or even predict the information received from the experiment. In order to achieve these goals we utilize non-empirical quantum mechanical methods such as CCSD(T), MP2, DFT, EDA *etc.*

In the first part of the thesis, we discuss the performance of semiempirical quantum mechanical (SQM) PM6 and PM7 methods and we switch to density functional theory (DFT) approach. However, in order to apply the DFT methodology for treatment of the large systems, the smaller basis set and proper description of London dispersion are mandatory. The most recent empirical dispersion correction, Grimme's D3 version, is limited only to few TZ size and larger basis sets.² Therefore, we focus on basis sets of smaller size. We develop D3 correction parameters for various basis set and functional combinations. Here, large number of benchmark data sets of noncovalent interactions is utilized.

Next, we verify the bioinformatic findings gathered from available crystal geometries of protein•DNA systems using DFT-D3 method. We compare the interaction energies in distributions of amino acid residues around each of the DNA bases for all 20×4 amino acid–DNA base pair combinations. Afterwards, we analyze these distributions with clustering procedure in a similar way as reported by Berka *et al.* for amino acid sidechain interactions in proteins.³ The resulting cluster representatives are utilized as model systems for benchmarking of various *ab initio* and empirical methods against calculated CCSD(T)/CBS reference.

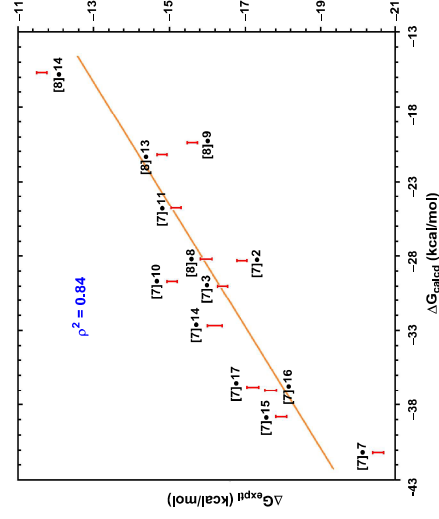
Finally, two projects concerning host•guest systems follow. Firstly, we study selective artificial receptor assembled from three helicates (from family of bis-bisurea ligands) functionalized into an "aromatic cage" that

Popsali jsme úroveň popisu různých SQM metod používaných pro odhady ΔG u systémů bílkovina•lék. Ukázali jsme, že když chceme dosáhnout přesnosti 1 kcal/mol a vyšší, např. velký rozsahem exotických nekovalentních interakcí, musí být pro to vybrány méně přibližné metody. Zde jsme využili velké množství post-HF metod a také metodu DFT rozšířenou o empirickou korekci na disperzní interakce. Obě tyto skupiny metod jsme použili ve zbylých prezentovaných projektech.

Následně jsme kvantitativně stanovili vazebné preference komplexů bílkovina•DNA. Bylo prokázáno, že geometrie aminokyselina–báze, schopné vzájemného rozpoznávání, korespondují unikátním energetickým minimům s interakčními energiemi vzdálenými od zbylé distribuce.

Na posledních dvou projektech jsme demonstrovali, že nástroje výpočetní chemie mohou pomoci k pochopení a interpretování rentgenových krystalografických dat. Reprodukovali jsme preferenci vazebnou selektivitu molekul cholinu nad acetylcholinem pro samo-uspořádání aromatické klece složené ze tří helikálních molekul a rozložili jsme vypočítané hodnoty ΔG do různých fyzikálně smysluplných členů. V podobném duchu jsme analyzovali 13 CB[n] hostitel•host systémů. Vypočítali jsme výrazný nárůst flexibility molekuly CB[8], což může indikovat, že může snadněji pojmout větší molekuly s většími požadavky na prostor. Toto zůstává stále experimentálně nepotvrzeno, avšak naše nedávné pilotní studie několika komplexů CB[8] s isodiamantany tomu nasvědčují.

- ❖ Byla provedena predikce nového světového rekordu v nekovalentní vazbě. V současnosti je připravován tento nový host, Diam-4,9-di(NiMe₂propanoNH₃), ve skupině Pavla Majera.



Obrázek 5. Korelace mezi ΔG_{calcd} a ΔG_{exptl} ($\Delta G_{exptl} = 0.247\Delta G_{calcd} - 9.179$, $\rho^2 = 0.84$, $n = 13$).

V této práci byl prezentován výzkum zaměřený na rozpoznávací procesy řízené nekovalentními interakcemi. U případů, kdy je vysoce pravděpodobné, že vazba je primárně entalpiicky řízená, kvantově mechanický popis přináší mnoho výhod do popisu vícečásticových efektů, přenosu elektronů a protonů, vzniku a rozpadu kovalentních vazeb *atd.*

Hlavní část snahy byla vynaložena na získání detailního pochopení rentgenových krystalových systémů hostitel•host, popisu přesnosti metod a vývoje empirické disperze pro metodu DFT. Studium vazebných preferencí mezi DNA a bílkovinami musí být považováno pouze jako počáteční studie na položení základu pro navazující detailnější práce na větších molekulárních fragmentech zvláště v rozpouštědlech. To stejné platí pro případ samospořádaných umělých klecí vázající molekuly cholimu a acetylcholinu.

5 Závěr

resembles ChoX protein. This artificial construct achieved experimental binding affinity of choline (Ch+) as high as 20 times of that of acetylcholine (ACh+). We try to determine the leading interaction motifs governing interplay between larger organic molecules. The second project is the application of quantum mechanical tools in the chemistry of cucurbit[*n*]uril (CB[*n*]).⁴ We study, describe and enhance the contemporary host•guest binding in present systems.

2 Aims of the Study

The aim of the thesis is to complement and predict experimental data of host•guest systems and search for universal rules governing DNA sequence recognition processes. In order to perform these tasks it proved to be mandatory to start with development and assessment of the accuracy of contemporary computational methods available for noncovalent interactions. Balanced treatment of different types of interactions such as hydrogen bond, halogen bond, X- π interactions, charge-transfer, dispersion *etc.* plays a key role.

3 Material and Methods

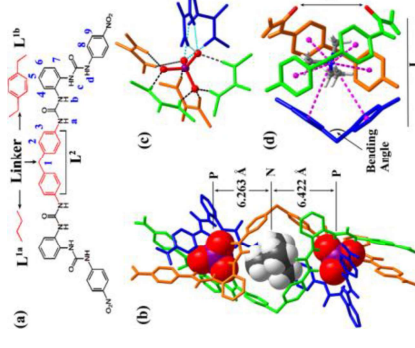
At first, we studied several groups of methods ranging from the force fields to SQM, DFT and *ab initio* methods. Various data sets of geometries and interaction energies of dimers played essential role in method accuracy assessment. The most prominent examples are: *S66(x8)*, *Wat5*, *PCONF*, *Goddard water*, *L7*, *AA-sidechain*, *Ihsg etc.* data sets.^{5,6} For most of these, the well-constructed benchmark CCSD(T)/CBS energies were published. The first exception is the *L7* set, for which Sedláč *et al.* published QCISD(T)/CBS energies calculated in a similar way to CCSD(T)/CBS estimates with the inclusion of the Δ QCISD(T) correction term instead of the ACCSD(T) one. For more details, go to the original paper.⁷ Second exception represents the *Wat5* data set. We calculated CCSD(T)/CBS estimates on the previously published structures with aug-cc-pVDZ basis set for the correction term and aug-cc-pVTZ and aug-cc-pVQZ basis sets for $\Delta E_{HF/CBS}^{MP2corr/CBS}$ and $\Delta E_{HF/CBS}^{HF/CBS}$ extrapolations according to Helgaker.⁸ Benchmark interaction energies were corrected for basis set superposition error (BSSE) using the counterpoise scheme of Boys and Bernardi.

Out of the 50,205 nucleotide–amino acid pairs gathered from the Protein•DNA interaction atlas we generated 20 distributions of amino acid residues around each of the DNA bases with a procedure similar to the SIRIUS set of scripts introduced by Singh and Thornton.⁹ Clustering technique resulted in 272 clusters and the same number of cluster representatives. The hydrogen atoms were added and their position optimized at the B3LYP-D3/def2-TZVPP level. The reference CCSD(T)/CBS interaction energies were calculated using composition scheme using basis sets described above.

In the studies of host-guest systems, we used density functional theory, more specifically BLYP-D3/def2-TZVPP/COSMO//BLYP-D/def2-SVP/COSMO setup. We tested the ability of COSMO solvation model to predict the solvation energies in host-guest complexes by comparison with other solvation models such as SMD and COSMO^{PM6}. Next, the accuracy of the BLYP-D3/def2-TZVPP method for interaction energies has been assessed against the more accurate methods such as CCSD(T)/CBS, MP2.5/CBS and FN-DMC for both small model systems and one whole CB[7]•Diamantane complex using Kernel Energy Method (KEM). The interaction energy was decomposed into physically meaningful terms using Energy Decomposition Analysis (EDA) method.

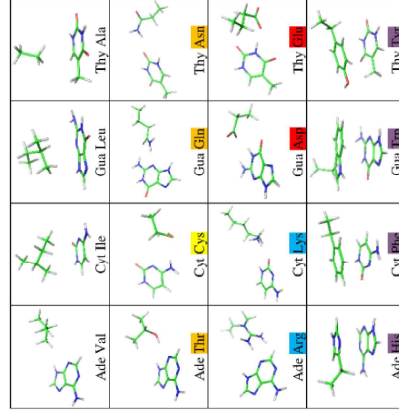
4 Results and Discussion

- ❖ None of the tested SQM methods showed a clear dominance and can thus be unambiguously recommended for the use in the field of noncovalent interactions. PM6 with D3H4(X) and DH2(X) corrections was slightly superior to the other setups (see Figure 1) while PM7 results for dispersion bound complexes and water clusters are a limiting factor in the possible applications of PM7.
- We passed to DFT approach and examined over 20 minimal and DZ size basis sets combined with several density functionals. We chose those with the smallest BSSE and for the most promising combinations we parameterised the Grimme's D3 correction on the S66x8 data set.
- The BLYP and B3LYP functionals yielded in average more accurate results than PBE, TPSS and PBE0 while the BLYP has lesser computational demands when compared to hybrid B3LYP functional.



Obrázek 4. (a) Struktury helikálních 4,4'-methylenebis(phenyl) molekul (b) Krystalová struktura (TMA)₅[(TMA)•(PO₄)₂(L₂)₃] komplexu; (c) Vodíkové vazby vytvořené mezi PO₄³⁻ aniontem and šesti močovinovými jednotkami L₂ molekul; (d) Aromatická klec držící TMA+ díky kationt-π interakci.

- ❖ Studovali jsme jedenáct krystalových struktur se systematicky rozdílnými molekulami hostů uvnitř (CB[7 or 8]) hostitelů. Struktury těchto biomimetických komplexů ukázaly, že žádné vysoko-energetické explicitní molekuly vody nejsou přítomny uvnitř kavit hostitelů a nejsou proto potřebné pro modelování těchto vazebných interakcí.
- ❖ Porovnáním pružnosti molekul CB[n] bylo nalezeno ostré zvýšení, při přechodu od CB[7] k CB[8] hostiteli. Solvatační vlastnosti byly dobře reprodukovány explicitním (WaterMap) a implicitními (COSMO and SMD) modely solvatace.
- ❖ BLYP-D3/def2-TZVPP interakční energie pro jeden z komplexů CB[7] s diamantem byla v blízkém souladu s hodnotou MP2.5/CBS (vypočtená s použitím KEM). Byla dosažena dobrá korelace mezi teoretickou (ΔG_{atad}) a experimentální (ΔG_{expt}) změnou Gibbsovy volné energie. (Obrázek 5, $\rho^2 = 0.84$)



Obázek 3. Geometrie bázi DNA a Ca reprezentací aminokyselin.

- Přepočítáním všech kontaktů spojených s jedním klastrem páru adenin–glutamin pomocí metody DFT-D/B3LYP-D3/def2-TZVPP, jsme ověřili naše pozorování ohledně distribuce párových interakčních energií párů aminokyselina–báze. Navzdory malé systematické odchylce, směrem k zápornějším hodnotám, příslušné DFT profily interakčních energií reprodukovaly empirické výsledky.
- Náš výpočetní přístup stanovil, že Ch⁺ je vázán 2.8 kcal/mol silněji (ΔG) než ACh⁺, což odpovídá $K(\text{Ch}^+)/K(\text{ACh}^+) = 109$. Toto je v dobré korelaci s dosaženou experimentální hodnotou rovnou 20.
- Ze srovnání vypočítaných entalpických a entropických příspěvků ke změně Gibbsovy energie můžeme vyvodit, že překvapivě právě relativní zvýšení entropického příspěvku od Ch⁺ k ACh⁺ způsobuje diskriminaci Ch⁺ nad ACh⁺. Avšak je třeba si uvědomit, že tyto malé rozdíly v ΔG jsou v rámci přesnosti metody silového pole.

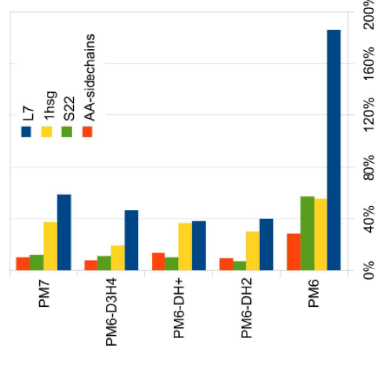


Figure 1. The RMSE between the respective SQM and CCSD(T)/CBS methods plotted as the percentage of the average stabilization energy.

- Large diversity of DFT results originated from the different diffuse character of basis sets and magnitude of BSSE. This was most profound in the case of fairly diffuse DZVP-DFT basis set having average RMSE 0.186 kcal/mol delivering superior results of almost TZ size basis set quality when newly parametrised D3 empirical dispersion is utilized. Substantial improvement at negligible additional cost can be achieved for X40 data set when larger basis set (def2-TZVP) is used only for heavier halogen atoms (Cl, Br, I).

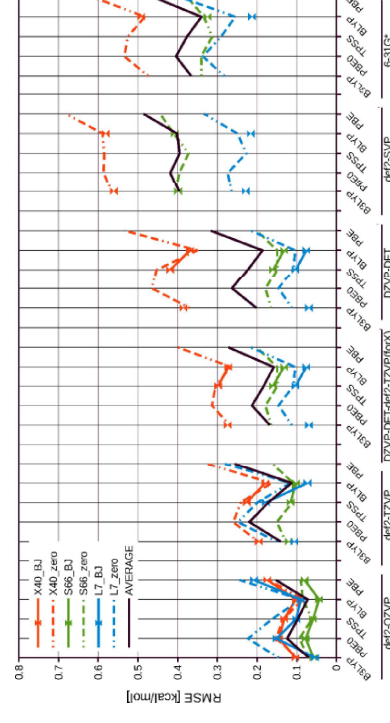


Figure 2. DFT-D3 results for various basis set and functional combinations

- BLYP and B3LYP functionals performed well for both geometry optimizations and energy calculations and are suitable for applications in large systems. It is important to note that even def2-SVP or 6-31G* basis sets combined with any tested functional clearly surpassed SQM methods tested in previous study.
- On representative geometries of Protein•DNA interaction we found that MP2.5 method achieved small RMSE of 0.11 kcal/mol (relative error of 2%) when compared to CCSD(T) method. This technique shows great promise for the future larger scale applications. DFT functional B3LYP systematically overestimated the strength of the binding by up to 0.31 kcal/mol in positively charged systems. The RMSE value of 0.25 kcal/mol is slightly increased to 0.35 kcal/mol, when the other tested functional BLYP was used instead. PM6-D3H4(X) method and Amber99SB-ILDN force field were in reasonable agreement with benchmark data.
- Next, a brief analysis of the interaction energy profiles of clusters associated with a certain amino acid–base pair followed. Following general observations could be drawn in several energy profiles (I.) the clusters accommodate most of the contacts within the interaction energy range (II.) the cluster consists of contacts with the highest stabilization energy found in the energy profile of the associated distribution (III) the low energetic contacts do not correspond to any cluster (IV.) the peaks of the interaction energy distributions corresponds well with the position of the cluster representatives.

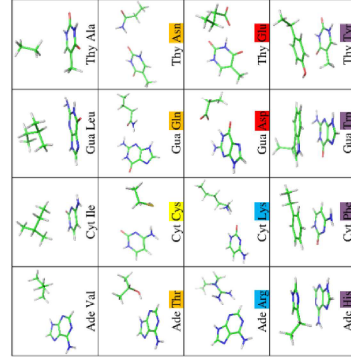
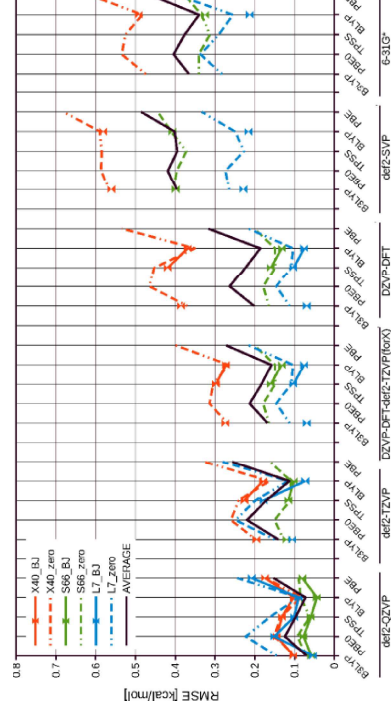
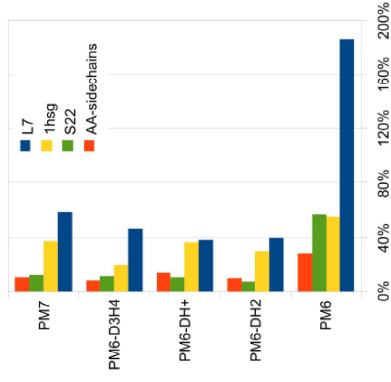


Figure 3: Geometries of DNA bases and C α representations of amino acid.



Obrázek 2. Výsledky DFT-D3 pro různé kombinace bází a funkcionálů.

- Na klastr reprezentantech geometrií interakcí bílkovina•DNA jsme určili, že zde metoda MP2.5 dosáhla malé chyby 0.11 kcal/mol (relativní chyba 2%) pro interakční energie, když byla srovnána s metodou CCSD(T). Tato technika vykazuje velký potenciál pro budoucí aplikace ve větších systémech. DFT funkcionál B3LYP systematicky nadhodnotil sílu vazeb až o 0.31 kcal/mol u kladně nabitých systémů. Hodnota RMSE 0.25 kcal/mol byla mírně zvýšena na 0.35 kcal/mol, když byly použity další testované funkcionály B3LYP a BLYP. Metoda PM6-D3H4(X) a metoda silového pole Amber99SB-ILDN byly v přijatelném souhlasu s referenčními hodnotami.
- Dále následovala stručná analýza interakčních profilů určitých párů aminokyselina–báze. Byla vyvozena následující obecná pozorování v několika profilech energií: (I.) klastry obsahující větší kontaktů v mezích daných interakčních energií (II.) klastry se skládají z kontaktů s nejvyššími stabilizačními energiemi nalezenými v energetických profilech příslušných distribucí (III.) nízké energetické kontakty nekorrespondují s žádným klastrem (IV.) píky distribucí interakčních energií dobře korespondují s pozicí reprezentantů klastrů.



Obrázek 1. Střední kvadratická odchylka mezi příslušnými SQM a CCSD(T)/CBS metodami zobrazená jako procenta průměrné stabilizační energie.

- Funkcionály BLYP and B3LYP dosáhly v průměru přesnějších výsledků než PBE, TPSS a PBE0, zatímco BLYP má menší výpočetní nároky, když je porovnán s hybridním funkcionálem B3LYP.
- Velká rozmanitost výsledků DFT vychází z různého difúzního charakteru bázi a velikosti BSSE. Toto bylo nejvíce patrné v případě poměrně difúzní DZVP-DFT báze mající průměrnou střední kvadratickou odchylkou (RMSE) 0.186 kcal/mol, přinášející skvělé výsledky téměř na úrovni bázi velikosti TZ, když byla zároveň použita nově parametrizovaná D3 empirická disperze. Výrazné zlepšení s minimálními náklady bylo dosaženo pro databázi X40, když větší báze (def2-TZVP) byla použita pouze pro těžší halogenové atomy (Cl, Br, I).
- Funkcionály BLYP and B3LYP se osvědčily pro optimalizace struktur i výpočty energií a jsou proto vhodné pro použití na větších systémech. Je důležité podotknout, že dokonce i báze def2-SVP or 6-31G* zkombinované s jakýmkoliv testovaným funkcionálem jasně překonaly výsledky SQM metod, které byly testovány v předchozím projektu.

- We verified our observations concerning the amino acid–base pair interaction energy distributions by recalculating all cluster-associated contacts for adenine–glutamine contacts using the DFT-D/B3LYP-D3/def2-TZVPP method. Despite the small shift toward more negative values, the respective interaction energy DFT profiles reproduced the empirical results very well.
- Our computational approach determined that the Ch⁺ is bound 2.8 kcal/mol stronger (ΔG) than ACh⁺, corresponding a $K(\text{Ch}^+)/K(\text{ACh}^+) = 109$. This is in nice correlation with the experimental value of 20.
- When comparing the calculated enthalpic and entropic contributions to the change of the Gibbs energy, we can conclude that surprisingly relative increase of the entropic contribution from Ch⁺ to ACh⁺ is causing the discrimination of Ch⁺ over ACh⁺ guest molecule. However, it is clear that these rather small differences in ΔG are within the accuracy of the force field.
- ❖ We studied eleven X-ray structures of systematically varied geometry guest molecules within a (CB17 or 8I) hosts. Structures of these biomimetic complexes showed that no high-energy explicit water molecules present inside of the host cavities are required for modeling their binding interactions.

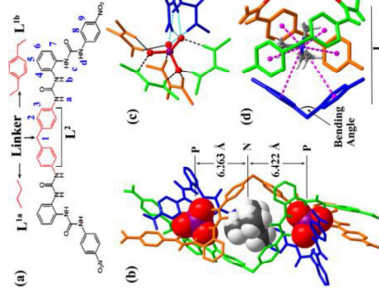


Figure 4. (a) Structures of helical 4,4'-methylenebis(phenyl) linkers (b) Crystal structure of $(\text{TMA})_5[(\text{TMA})_2(\text{PO}_4)_2(\text{L}_2)_3]$ complex; (c) Hydrogen bonds formed between a PO_4^{3-} ion and six urea units of L_2 molecules; (d) The aromatic cage trapping a TMA^+ through cation- π interactions.

- ❖ Sharp increase in deformability of CB[n] molecules has been found when passing from CB[7] to CB[8] hosts. Solvation properties have been well reproduced by explicit (WaterMap) and implicit (COSMO and SMD) solvent models.
- ❖ BLYP-D3/def2-TZVPP interaction energy for one of the CB[7] complex with diamantane has been found in close agreement with MP2.5/CBS value (utilizing KEM). Nice correlation between theoretical (ΔG_{calcd}) and experimental (ΔG_{expt}) changes of Gibbs free energies has been achieved. (Figure 5, $\rho^2 = 0.84$)
- ❖ Prediction has been made of the next world record noncovalent binder. Currently, the synthesis work on this new guest, Diam-4,9-di(NMe₂propanoNH₃), is in progress in group of Pavel Majer.

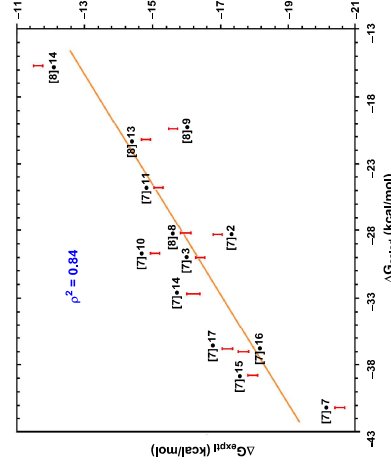


Figure 5. Correlation between ΔG_{calcd} and ΔG_{expt} ($\Delta G_{expt} = 0.247\Delta G_{calcd} - 9.179$, $\rho^2 = 0.84$, $n = 13$).

5 Conclusions

In this thesis, research concerning recognition processes governed by noncovalent interactions has been presented. In those cases when binding is

Využitím procedury podobné souboru programů SIRIUS popsané autory Singh and Thornton, jsme z celkového množství 50,205 párů nukleotid-aminokyselina, získané z Atlasu Bílkovina•DNA interakcí vygenerovali 20 distribuční aminokyselinyových reziduí okolo každé báze DNA.⁹ Klastrovací technika vyústila v 272 klastrů a stejné množství klastrů reprezentantů. Byly doplněny vodíkové atomy a optimalizovány jejich pozice na úrovni B3LYP-D3/def2-TZVPP. Referenční hodnoty CCSD(T) interakčních energií byly sestaveny pomocí členů a bází popsaných výše.

Ve studiích systémů hostitel•host jsme využili teorie funkcionálu hustoty, specifitěji BLYP-D3/def2-TZVPP/COSMO//BLYP-D/def2-SVP/COSMO. Testovali jsme schopnost solvatačního modelu COSMO^{DFT} predikovat solvatační energie v systémech hostitel•host srovnáním s dalšími modely jako je SMD a COSMO^{PM6}. Dále byla ověřena přesnost metody BLYP-D3/def2-TZVPP proti metodám CCSD(T)/CBS, MP2.5/CBS a FN-DMC, pro hodnoty interakčních energií malých modelových systémů i celého komplexu CB[7] s diamantánem, použitím tzv. metody energie kernelů (KEM). Interakční energie byly rozděleny do fyzikálně smysluplných členů použitím metody dekompozice analýzy energie (EDA).

4 Výsledky a diskuse

❖ Žádná z testovaných SQM metod neukázala jasnou dominanci nad ostatními, a proto nemůžeme žádnou jednoznačně doporučit pro použití na poli nekovalentních interakcí. PM6 s korekcí D3H4(X) a DH2(X) mimě vyčnívala nad ostatními postupy (viz. Obrázek 1), zatímco výsledky metody PM7 pro disperzně vázané komplexy a klastry vody jsou limitujícím faktorem pro možné aplikace metody PM7.

• Proto jsme přešli k přístupu DFT a ohodnotili přes 20 minimálních a DZ bází zkombinovaných s několika funkcionály hustoty. Vybrali jsme báze s nejmenší BSSE a pro ty kombinace s funkcionály, které se ukázaly jako nejslibnější, jsme parametrizovali Grimmeho D3 korekci na databázi S66x8.

složené ze tří helikálních molekul (z rodiny bis-bisurea ligandů) funkcionalizovaných do „aromatické klece“, která svým tvarem připomíná vazebné místo bílkoviny ChoX. Tato umělá konstrukce dosáhla experimentální vazebné afinity cholinu (Ch⁺) až 20ceti násobně vyšší, než je ta acetylcholinu (ACh⁺). Pokusili jsme se určit hlavní interakční motiv řídící tuto souhru velkých organických molekul. Druhý projekt je aplikaci kvantové mechanických nástrojů na chemii cucurbit[*n*]urilu (CB[*n*]). Studovali jsme popis a vylepšení současných vazebných motivů v těchto systémech hostitel•host.

2 Cíle Práce

Cílem této disertační práce je doplnit a předpovědět experimentální data systémů hostitel•host a hledat univerzální pravidla řídící procesy rozpoznávání sekvencí DNA. Abychom provedli tyto úkoly, ukázalo se být potřebné vyvinout a ověřit přesnost současných výpočetních technik dostupných pro nekovalentní interakce. Vyzvážený popis různých typů interakcí jako je vodíková vazba, halogenová vazba, X- π interakce, u systémů s přenosem náboje, disperze *atd.* hraje klíčovou roli.

3 Materiál a Metodika

Nejdříve jsme studovali několik skupin metod sahajících od metod silového pole k SQM, DFT až k metodám *ab initio*. Různé databáze geometrií a interakčních energií dimerů hrály důležitou roli u hodnocení přesnosti těchto metod. Nejdůležitější úlohu hrály databáze: *S66(a8)*, *Wat5*, *PCONF*, *Goddard water*, *L7*, *AA-sidechain*, *Ihsg* a další.^{5,6} Pro většinu z nich byly publikovány spolehlivé referenční hodnoty CCSD(T)/CBS energií. První výjímku představuje databáze *L7*, pro kterou Sedláč *et al.* publikoval energie vypočítané metodou QCISD(T)/CBS podobným způsobem, jako je tomu u odhadů CCSD(T)/CBS energií, kde místo ACCSD(T) korekčního členu zahrnuli Δ QCISD(T) korekční člen. Více informací je popsáno v původní publikaci.⁷ Druhou výjímku tvoří databáze *Wat5*. Vypočítali jsme odhady CCSD(T)/CBS energií na již publikovaných geometriích pomocí aug-cc-pVDZ báze použité pro korekční člen a aug-cc-pVTZ and aug-cc-pVQZ báze pro $\Delta E_{\text{MP2corr/CBS}}$ and $\Delta E_{\text{HF/CBS}}$ využitím extrapolace podle Helgakera.⁸ Referenční interakční energie byly zbaaveny superpoziční chyby (BSSE) využitím techniky Boyse and Bernardiho.

highly probably primarily enthalpy-driven, the QM description brings many advantages in description of many body effects, electron and proton transfers, formation and dissociation of a covalent bond *etc.*

Major part of the effort was made to obtain a detail understanding of the X-ray structures of host•guest systems, method performance and development of empirical dispersion for DFT methodology. While the studies of binding preferences between DNA and proteins must be considered as the initial investigation to set the stage for more elaborate analysis of larger molecular fragments especially in solvents. The same is true in case of self-assembled artificial cage binding choline and acetylcholine guest molecules.

We have described the performance of various SQM method (used nowadays for ΔG predictions in protein•ligand systems). We showed that, when one wants to achieve accuracy of 1 kcal/mol and higher across the wide range of exotic noncovalent interactions, the less approximate methods need to be chosen. Here, we utilized DFT methodology augmented with empirical correction for dispersion interaction and range of post-HF methods. Both of these classes of methods were used in the rest of the presented projects.

Next, we quantitatively assessed the binding preferences in protein•DNA complexes. It has been proven that amino acid–base geometries capable of one-to-one amino acid–base recognition correspond to unique energy minima with interaction energies distinct from the rest of the distribution.

Finally, in two last projects we have demonstrated that the computational chemistry tools can help us to understand and interpret the X-ray crystallographic data. We reproduced the preferential binding selectivity of choline guest over acetylcholine towards self assembled triple helicate aromatic cage and decomposed the calculated ΔG into various physically meaningful terms. In similar way we analyzed 13 CB[*n*] host•guest systems. We have predicted substantial increase of flexibility of CB[8] that could indicate that it can more readily encapsulate larger guests with higher spatial demands. It remains still to be proven experimentally however our recent pilot study of several CB[8] •isodiamantane complexes suggests this notion.

- (1) Laskowski, R. A.; Thornton, J. M. Understanding the molecular machinery of genetics through 3D structures. *Nature Reviews Genetics* **2008**, *9*, 141-151.
- (2) Grimme, S. Accurate description of van der Waals complexes by density functional theory including empirical corrections. *J. Comput. Chem.* **2004**, *25*, 1463-1473.
- (3) Berka, K.; Laskowski, R.; Riley, K. E.; Hobza, P.; Vondrasek, J. Representative Amino Acid Side Chain Interactions in Proteins. A Comparison of Highly Accurate Correlated *ab Initio* Quantum Chemical and Empirical Potential Procedures. *J. Chem. Theory Comput* **2009**, *5*, 982-992.
- (4) Lagona, J.; Mukhopadhyay, P.; Chakrabarti, S.; Isaacs, L. The cucurbit n uril family. *Angew. Chem.-Int. Edit.* **2005**, *44*, 4844-4870.
- (5) Rezac, J.; Hobza, P. Benchmark Calculations of Interaction Energies in Noncovalent Complexes and Their Applications. *Chemical Reviews* **2016**, *116*, 5038-5071.
- (6) Rezac, J. Cuby: An Integrative Framework for Computational Chemistry. *J. Comput. Chem.* **2016**, *37*, 1230-1237.
- (7) Sedlak, R.; Janowski, T.; Pitonak, M.; Rezac, J.; Pulay, P.; Hobza, P. Accuracy of Quantum Chemical Methods for Large Noncovalent Complexes. *J. Chem. Theory Comput* **2013**, *9*, 3364-3374.
- (8) Halkier, A.; Helgaker, T.; Jorgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. Basis-set convergence in correlated calculations on Ne, N-2, and H2O. *Chem Phys Lett* **1998**, *286*, 243-252.
- (9) Singh, J.; Thornton, J. M. SIRIUS - An Automated-Method for The Analysis of the Preferred Packing Arrangements Between Protein Groups. *Journal of Molecular Biology* **1990**, *211*, 595-615.

6 References

Pochopení principů, kterými se řídí selektivní navázání přírodou i člověkem vytvořených komplexů, by představovalo v aplikacích v oblasti biochemie, supermolekulární chemie, materiálových vědách nebo molekulární biologie obrovský skok kupředu. Zvětšující se množství rentgenových krystalografických struktur nedávno otevřelo prostor pro mnoho studií v tomto směru.¹

Ve zde prezentované disertaci je hlavním předmětem zájmu popis a modelování stavebních bloků biopolymerů (bílkovin a DNA) a různých biomimetických komplexů v reálném i izolovaném prostředí. Zatímco experimenty velmi často úspěšně stanovují sílu vazby, informace o povaze nekovalentních interakcí, hrající roli ve vazebných procesech, jsou velmi omezené. Výsledky výpočtů mohou doplnit, poskytnout náhled nebo dokonce predikovat informace získané z experimentu. Abychom dosáhli těchto cílů, použili jsme ne-empirické kvantově mechanické metody jako jsou CCSD(T), MP2, DFT, EDA *atd.*

V první části této práce jsme diskutovali výkonost semiempirických kvantově mechanických (SQM) metod PM6 a PM7 a přeusnuli jsme se k aplikaci teorii funkcionálu hustoty (DFT). Abychom použili DFT metodologii pro popis chování velkých systémů, ukazuje se jako důležité kombinovat malé báze a řádný popis Londonovy disperze. Nejnovější empirická korekce na disperzi, Grimmeho D3, je omezená pouze pro velké TZ a QZ báze.² Proto jsme se zaměřili na báze menší velikosti. Vyvinuli jsme D3 korekční parametry pro různé kombinaceází a funkcionálů. Zde bylo využito velké množství referenčních dat zaměřených na nekvalentní interakce.

Dále jsme za použití DFT-D3 metody ověřili bioinformatické poznatky získané z dostupných krystalových geometrií systémů bílkovina•DNA. Srovnali jsme distribuce interakčních energií reziduí aminokyselin okolo každé DNA báze pro všech 20×4 párů aminokyselina–DNA báze. Následně jsme tyto distribuce analyzovali a našli klustry podobných orientací, obdobným způsobem jako bylo popsáno v práci vypracované Berkou *et al.* pro interakce postranních řetězců u bílkovin.³ Výsledné reprezentanty klustrů byly použity jako modelové systémy pro ověření přesnosti různých *ab initio* a empirických metod, porovnané s námi vypočtenými referenčními hodnotami CCSD(T)/CBS.

Nakonec jsou popsány dva projekty zabývající se systémy hostitel•host. Nejdříve jsme studovali selektivní umělé připravené receptory