

**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: Makromolekulární chemie
Ph.D. study program: Macromolecular chemistry

Autoreferát disertační práce
Summary of the Ph.D. Thesis



Interakce proteinů s inhibitory: kvantověchemická studie
Interaction of proteins with inhibitors: quantum chemical study

Mgr. Petr Dobeš

Školitel:

prof. Ing. Pavel Hobza, DrSc.

Praha 2011

Úvod

Předložená disertační práce se zabývá teoretickým studiem interakcí mezi proteinkinázami a jejich inhibitory. Studované proteinkinázy, cyklin-dependentní kináza 2 (CDK2) a kináza CK2 (angl. *casein kinase 2*) hrají důležitou roli v regulaci buněčných procesů u eukaryotických organismů. Jejich nesprávná funkce v buňce může u člověka vést k závažným onemocněním. Tento proces je možné zvrátit vyřazením aberantních proteinkináz pomocí specifických nízkomolekulárních inhibitorů. Inhibitory proteinkináz se typicky vážou do aktivního místa enzymu pomocí nekovalentních interakcí. Teoretický popis těchto interakcí pomocí kvantověchemických a molekulárně mechanických metod může pomoci při pochopení biofyzikálních principů řídicích vazbu. Ty pak mohou být následně využity při racionálním návrhu účinnějších či specifitějších inhibitorů.

Nekovalentní interakce hrají klíčovou roli v biodisciplinách a jsou zodpovědné za strukturu a tedy i vlastnosti jak biomakromolekul (proteiny, nukleové kyseliny) tak i jejich komplexů s ligandy. Disperzní energie tvoří atraktivní část van der Waalsovy energie a jako taková hraje důležitou úlohu ve stabilizaci biomakromolekul. Komplexy s vodíkovou vazbou patří k nejsilněji vázaným nekovalentním komplexům. Mezi méně známé typy nekovalentních interakcí patří halogenová vazba (angl. *halogen bonding*), která se tvoří mezi těžším halogenovým atomem X (Cl, Br, I) s vlastností Lewisovy kyseliny, který je kovalentně vázaný na uhlík, a Lewisovou bází Y jako je např. atom kyslíku.

Cílem předložené disertace je zkoumat pomocí výpočetních metod nekovalentní interakce mezi proteinkinázami a jejich inhibitory, jež by mohly mít potenciální uplatnění při léčení závažných onemocnění. Práce je rozdělena na tři části. První se zabývá detailním popisem interakce nízkomolekulárního inhibitoru roskovitinu s CDK2. Ve druhé části byla soustředěna pozornost na teoretický popis interakce CDK2 s 15 strukturně odlišnými inhibitory za použití semiempirické kvantově-mechanické skórovací funkce (angl. *scoring function*). Konečně ve třetí části byl studován vztah mezi teoretickými a experimentálními hodnotami vazebných volných energií pro komplexy kukuřičné a lidské CK2 s halogenovanými inhibitory, které jsou s enzymem schopny tvořit halogenovou vazbu.

Metody

Molekulová mechanika popisuje molekuly na základě zákonů klasické fyziky. Tento popis je vyjádřen pomocí tzv. silového pole, což je množina parametrů a matematických funkcí, které popisují potenciální energii systému částic jako jsou atomy a molekuly v závislosti na jejich souřadnicích. Jedním z nejpoužívanějších silových polí je AMBER a to bylo také použito v předložené disertační práci.

Pomocí kvantověchemických výpočtů je v principu možné přesně popsat a předpovídat fyzikálně-chemické vlastnosti molekul i molekulových klastrů. Základní metodou kvantové chemie je metoda Hartreeho a Focka (HF). Metoda HF má svá omezení plynoucí ze zanedbání elektronové korelace a relativistických jevů. Korelované metody (nebo též post-SCF metody) jsou metody, které zlepšují popis vlnové funkce tím, že uvažují působení elektronové korelace. Mezi nejčastěji používanou korelovanou metodu patří poruchová metoda druhého řádu podle Møllera a Plesseta (MP2), která poskytuje přiměřené zahrnutí elektronové korelace. Metoda spřažených klastrů CCSD(T) poskytuje velmi přesné hodnoty energie a vlastností v širokém spektru molekul a molekulových klastrů.

Alternativním přístupem k řešení víceelektronové vlnové funkce je metoda funkcionálu elektronové hustoty DFT (angl. *density functional theory*), která podobně jako metoda HF slouží k výpočtu víceelektronové vlnové funkce v základním stavu. Fyzikální a chemické vlastnosti molekul jsou počítány pomocí funkcionálu elektronové hustoty. Výhodou metod DFT je jejich výpočetní nenáročnost ve srovnání s post-SCF metodami. Ještě jednodušší je metoda SCC-DFTB (angl. *self-consistent charge density-functional tight-binding*), což je aproximativní metoda DFT založená na principu těsné vazby. Po zahrnutí disperze empirickým členem vzniká metoda SCC-DFTB-D, která je schopna popsat i disperzní interakce. Dalšího zjednodušení a urychlení výpočtů je dosaženo použitím semiempirických kvantověchemických metod (SQM – angl. *semiempirical quantum mechanics*), které jsou založeny na podobném postupu výpočtu jako metoda HF. Výpočet je však zjednodušen za použití různých aproximací.

Jednou z nejvhodnějších SQM metod pro použití na biomolekulové systémy je metoda PM6-DH2 poskytující přesné výsledky pro různé typy nekovalentních interakcí jako jsou vodíkové vazby a disperzní interakce. V případě halogenové vazby tato metoda přeceňuje velikost interakce mezi halogenem a elektronegativním atomem. Po zahrnutí korekce vzniká metoda PM6-DH2X, která je schopna popsat halogenovou vazbu srovnatelně s náročnými metodami kvantové chemie.

Odhad vazebné volné energie nazývaný skóre slouží k srovnání vypočtených hodnot síly vazby v komplexech proteinu s inhibítorem s experimentem. Klasické skórovací funkce jsou založené na použití empirického potenciálu. V předložené práci byla použita semiempirická QM metoda PM6-DH2X. Skóre bylo počítáno jako součet interakční entalpie ve vodě, entropického členu, deformační energie inhibítora a korekce solvatační energie inhibítora.

Výsledky

V první publikaci byla studována interakce mezi CDK2 a inhibítorem roskovitinem. Pro výpočet interakční energie komplexu CDK2-roskovitin bylo nejprve nutné zmenšit popisovaný systém tak, aby jej bylo možné popsat kvantověchemickými metodami. Jako referenční metoda byla použita metoda MP2/aug-cc-pVDZ, pro kterou bylo již dříve ukázáno, že dává podobné výsledky jako velmi přesná metoda CCSD(T)/CBS. Příslušné interakční energie byly srovnány s energiemi určenými jinými výpočetními metodami - DFT (B3LYP/6-31G**), SCC-DBTB a SCC-DFTB-D (D – značí empirickou disperzi) a molekulová mechanika AMBER (ff99). Zjistili jsme, že pro vazbu roskovitinu s CDK2 je dominantní vliv disperzních sil. Korektní popis tudíž vyžaduje použít kvantověchemické metody zahrnující disperzní interakci, buď explicitně jako u metody MP2 nebo když je zahrnuta pomocí empirické korekce jako u metody SCC-DFTB-D. Dále je zřejmé, že metoda silového pole AMBER, která zahrnuje disperzní energii v Leonard-Jonesově členu, je též vhodná pro popis této interakce.

V druhé publikaci byla studována korelace vazebných volných energií získaných z teorie a experimentu na strukturně různorodé skupině 15 inhibitorů CDK2, které se kompetitivně vážou do aktivního místa CDK2. Příslušná skórovací funkce je založena na metodě PM6-DH2, a tak se zásadně liší od skórovacích funkcí založených na empirických potenciálech. Nejprve byla uvažována korelace experimentálně změřených inhibičních konstant s vypočítanou interakční entalpií ve vodním prostředí (ΔH_w) určené pomocí metody PM6-DH2

s implicitním modelem vody COSMO. Struktura komplexu inhibitoru s proteinem byla optimalizována stejnou metodou. Výsledkem byla silná korelace ($R^2=0,87$). Přidáním korekce na deformaci a desolvataci inhibitoru se korelace mírně zhoršila ($R^2=0,77$) a konečně započtením entropického členu došlo k dalšímu zhoršení korelace ($R^2=0,52$), zřejmě z důvodu nedostatků v popisu entropie. První uvedená korelace byla silnější než všechny dříve publikované korelace pro inhibitory CDK2 založené na empirických a semiempirických skórovacích funkcích.

Jak je uvedeno ve třetí publikaci, skórovací funkce založená na metodě PM6-D2X byla aplikována na dvě skupiny halogenovaných inhibitorů s CK2. První skupina byla vytvořena podle krystalových struktur kukuřičné CK2 a druhá pomocí namodelovaných struktur inhibitorů s lidskou CK2. Korelace mezi vypočítanou interakční entalpií ve vodě a experimentální inhibiční konstantou byla silná u skupiny vytvořené z krystalových struktur ($R^2=0,86$). Pro druhou skupinu struktur byla nalezena korelace výrazně slabší ($R^2=0,52$). Když byly k interakční entalpii přičteny korekce na desolvataci a deformační energie inhibitoru došlo v prvním případě k mírnému zhoršení ($R^2=0,81$), pro druhou skupinu došlo ale ke zlepšení ($R^2=0,71$). Přidání entropického členu založeného na empirickém potenciálu AMBER (ff99) vedlo k výraznému zhoršení korelace jak pro první skupinu ($R^2=0,24$) tak i pro druhou skupinu ($R^2=0,19$) a to z principiálního důvodu neschopnosti silového pole popsat halogenovou vazbu.

Závěry

Stabilizační energie komplexu CDK2 s roskovitinem je převážně tvořena disperzní energií. Metoda DFT, která nezahrnuje disperzní energii, je tak pro popis uvedeného komplexu zcela nevhodná. V případě zahrnutí disperze pomocí empirického členu dojde ke správnému popisu této interakce a metodu SCC-DFTB-D lze naopak doporučit pro uvedený komplex. Dominantní část celkové interakce je tvořena omezeným počtem aminokyselin, které přispívají k vazbě roskovitinu s CDK2 a jejich mutace může mít zásadní vliv na stabilizaci komplexu. Dále bylo ukázáno na schopnost silového pole popsat interakci roskovitinu s CDK2.

Při návrhu léčiv se k popisu interakce inhibitoru s proteinem používají skórovací funkce založené na principu silového pole. Metody silového pole ale nejsou schopny popsat kvantové jevy jako je polarizace nebo přenos náboje. Na skupině strukturně rozdílných CDK2 inhibitorů byla ukázána vhodnost použití skórovací funkce založené na semiempirické kvantověchemické metodě PM6-DH2. Nejsilnější korelace s inhibiční konstantou bylo dosaženo s interakční entalpií ($R^2=0,87$) spočítané pomocí metody PM6-DH2. Po zahrnutí všech členů skórovací funkce (včetně entropie počítané kvůli náročnosti silovým polem) se korelace zhoršila ($R^2=0,52$), zřejmě kvůli nedostatkům při výpočtu entropie v silovém poli AMBER. Díky velmi dobré korelaci s interakční entalpií se jeví nadějně, že výpočty pomocí skórovací funkce založené na metodě SQM mohou být v budoucnu vzhledem k rychlému rozvoji v oblasti informačních technologií použitelné při návrhu léčiv.

V poslední části práce byla studována možnost použití skórovací funkce založené na metodě PM6-DH2X pro halogenované inhibitory schopné tvořit halogenové vazby s proteinkinázou CK2. Struktury optimalizované pomocí PM6-D2X velmi dobře souhlasily s krystalovými strukturami, zatímco struktury optimalizované silovým polem AMBER nebyly schopny popsat halogenovou vazbu. Pro halogenované inhibitory CK2, založené jak na krystalových tak i namodelovaných strukturách, byla nalezena silná korelace mezi inhibiční konstantou a

interakční entalpií počítanou ve vodě. Po zahrnutí entropických členů došlo ke zhoršení korelace, což je způsobeno neschopností empirického potenciálu popsat halogenovou vazbu. Správný popis halogenovaných sloučenin v komplexech s halogenovou vazbou je jednou z podmínek k možnosti nalezení efektivnějších inhibitorů.

Publikace autora zahrnuté v této disertační práci

- [1] [Dobeš P, Otyepka M, Strnad M, Hobza P](#).: Interaction energies for the purine inhibitor roscovitine with cyclin-dependent kinase 2: correlated ab initio quantum-chemical, DFT and empirical calculations. *Chem. Eur. J.* 2006;12:4297-304.
- [2] [Dobeš P, Fanfrlík J, Řezáč J, Otyepka M, Hobza P](#).: Transferable scoring function based on semiempirical quantum mechanical PM6-DH2 method: CDK2 with 15 structurally diverse inhibitors. *J. Comput. Aided Mol. Des.* 2011;25:223-35.
- [3] [Dobeš P, Řezáč J, Fanfrlík J, Otyepka M, Hobza P](#).: Semiempirical QM Method PM6-DH2X Well Describes Geometry and Energetic of CK2-Inhibitor Complexes Involving Halogen Bonds while Empirical Potential Fails. *J. Phys. Chem. B*, under revision

Introduction

This dissertation focuses on theoretical studies of the interaction between protein kinases and their inhibitors. Studied protein kinases, cyclin-dependent kinase 2 (CDK2) and CK2 kinase (casein kinase 2) play an important role in regulating cellular processes in eukaryotic organisms. Their abnormal function in human cells can lead to serious diseases. This process can be stopped by blocking the aberrant protein kinases using specific low molecular weight inhibitors. Inhibitors of protein kinases typically bind to the active site of the enzyme by noncovalent interactions. Theoretical description of these interactions using quantum-chemical and molecular mechanical methods can help in understanding the biophysical principles governing the binding. These, in turn, can be subsequently used for a rational drug design of more effective and more specific inhibitors.

Noncovalent interactions play a key role in biodisciplines and they are responsible for the structure and hence properties of both biomacromolecules (proteins, nucleic acids) and their complexes with ligands. Dispersion energy is an attractive part of van der Waals energy, thus it plays an important role in the stabilization of biomacromolecules. Hydrogen bonding belongs to the strongest noncovalent interactions. Among the lesser-known type of noncovalent interaction is halogen bonding, which is formed between the heavier halogen atom X (Cl, Br, I) with Lewis acid properties, which is covalently bound to carbon and Lewis bases such as oxygen atom.

The aim of this thesis is to explore computational methods for noncovalent interactions between protein kinase and its inhibitors, which could have a potential application in treatment of serious diseases. The work is divided into three parts. The first chapter provides a detailed description of interaction for low molecular weight inhibitor roscovitine with CDK2. In the second part the focus lies on the theoretical description of the interaction of CDK2 with 15 structurally different inhibitors by using quantum mechanical semiempirical scoring function. Finally, in the third part the relationship has been studied between theoretical and experimental values of binding free energies for complexes of maize and human CK2 with halogenated inhibitors, which are capable of forming a halogen bonding with the enzyme

Methods

Molecular mechanics describes the molecule by using the laws of classical physics. This description is expressed by the so-called force field, which is a set of parameters and mathematical functions that describe the potential energy of particles such as atoms and molecules depending on their coordinates. One of the most widely used force fields is AMBER, which was used in this dissertation.

With help of quantum chemical calculations it is in principle possible to accurately predict the physico-chemical properties of molecules and molecular clusters. Basic method of quantum chemistry is Hartree-Fock method (HF). However HF method has limitations arising from the neglect of electron correlation. Correlated methods (also called post-SCF methods) are methods that improve the description of the wave function by considering the effect of electron correlation. The most commonly used one is correlated Møller-Plesset method with second-order perturbation (MP2). It provides adequate inclusion of electron correlation. One

of the best description of electron correlation, coupled cluster method CCSD(T) provides a highly accurate energy values in a wide range of molecules and molecular clusters.

An alternative approach to solving electronic wave function is density functional theory (DFT) method. Physical and chemical properties of molecules are calculated using electron density functional. The advantage of the DFT methods is their relative computational efficiency in comparison with the post-SCF methods. Even simpler, the SCC-DFTB (self-consistent charge density-functional tight-binding) method is an approximate DFT method based on the principle of tight-binding. After the inclusion of an empirical dispersion term, the SCC-DFTB-D method is able to describe the dispersion interaction. Other possible solution is to use semiempirical quantum mechanics methods (SQM), which are based on a similar calculation procedure as a HF method. The calculation is simplified by using different approximations.

One of the most suitable SQM methods for use in biomolecular systems is the PM6-DH2 method providing accurate results for different types of noncovalent interactions such as hydrogen bonds and dispersion interactions. In the case of the halogen bonding, this method overestimates the size and the interaction between the halogen atom and electronegative atom. The corrected method (PM6-DH2X) is able to describe the halogen bonding in comparison with computational expensive methods of quantum chemistry.

An estimate of the binding free energy is called the score and is calculated using a scoring function. It can be used for comparison of the experiment with the calculated scores for protein ligand complexes. Conventional scoring functions are based on the use of empirical potentials. In the present work we used the SQM method PM6-DH2X. The score was calculated as the sum of the interaction enthalpy in water, entropic term, deformation energy of an inhibitor and correction for solvation energy of inhibitor.

Results

In the first publication, the interaction between CDK2 and inhibitor roscovitine was studied. For the calculation of the interaction energy of the CDK2-roscovitine complex, firstly it was essential to reduce the described system so that it could be described by quantum chemical methods. As a reference, MP2/aug-cc-pVDZ method was used because it had previously been shown to give similar results as the very accurate CCSD(T)/CBS method. The reference interaction energies were compared with other methods - DFT (B3LYP/6-31G**), SCC-DFTB, SCC-DFTB-D and molecular mechanics AMBER (ff99). The dominant contribution comes from the dispersion forces. Thus it requires to use quantum chemical methods involving the dispersion interaction, either explicitly as in the MP2 method, or included in an empirical correction such as SCC-DFTB-D method. It is clear that the AMBER force field method, which includes the dispersion energy in the Leonard-Jones term, is also appropriate for the description of this interaction.

In the second publication, correlation of binding free energies obtained from experiment and theory was studied in the structurally diverse group of 15 competitive CDK2 inhibitors. The used scoring function is based on the PM6-DH2 method, thus it differs fundamentally from scoring functions based on empirical potentials. Firstly, we considered correlations of

experimental measurements of inhibition constants with the calculated interaction enthalpies in aqueous environment (ΔH_w) determined by the PM6-DH2 method with an implicit water model COSMO. The structure of the protein-inhibitor complex was optimized using the same method. There was a strong correlation ($R^2=0.87$). Secondly, adding a correction for deformation energy of inhibitor and correction for desolvation of inhibitor made the correlation slightly worse ($R^2=0.77$). Finally, addition of the entropic term to all the previous ones resulted in further deterioration in the correlation ($R^2=0.52$), probably due to the inaccuracies of the description of the entropy term. However, in the first case, the correlation was stronger than all the previously published correlations for CDK2 with inhibitors based mostly on empirical scoring functions.

As described in the third publication, scoring function based on the PM6-D2X method was used for two groups of halogenated inhibitors of CK2. The first group entailed the crystal structures of maize CK2 with inhibitors and the second group was formed from modeled structures of the human CK2 and inhibitors. Correlation between calculated interaction enthalpies in water and the inhibition constant was strong in the group created from the crystal structures ($R^2=0.86$). For the second group a much weaker correlation was found ($R^2=0.52$). When the interaction enthalpy term was augmented by the correction for desolvation of inhibitor and deformation energy of inhibitor, firstly there was a slight worsening in the correlation ($R^2=0.81$) but for the second group there was an improvement ($R^2=0.71$). Finally, the addition of the entropic term based on the empirical potential of AMBER (ff99) led to a significant deterioration in correlation as compared with the first group ($R^2=0.24$) and the second group ($R^2=0.19$). This is due to the inability of the force field to describe halogen bonding.

Conclusions

The stabilization energy of the complex of CDK2 with roscovitine is predominantly formed by the dispersion energy. DFT methods, which do not describe the dispersion energy was thus completely inappropriate for the treatment of such a system. When an empirical term is included to correct for the description of dispersion, such methods, as e.g. the SCC-DFTB-D, can be recommended for computation of this or similar complexes. The dominant part of the overall interaction is due to a limited number of amino acids, that contribute to binding of roscovitine with CDK2. Their mutations can thus be crucial to destabilize the complex. It was also shown that the AMBER force field was able to describe the interaction of roscovitine with CDK2.

The drug design community often employs scoring functions based on the principle of force field to describe the interaction between proteins and their inhibitors. However, the force fields methods are not able to describe quantum phenomena such as polarization or charge transfer. In contrast, the designed scoring functions based on the semiempirical method PM6-DH2 method was shown to give satisfactory results for the group of 15 structurally diverse inhibitors of CDK2. The strongest correlation with experimental inhibition constants was obtained using the interaction enthalpy ($R^2=0.87$) calculated by the PM6-DH2 method. After the inclusion of all the terms of the scoring functions (including the entropy term), the correlation deteriorated ($R^2=0.52$) probably due to the inaccuracies in the calculation of the entropy term in AMBER force field. Due to very good correlation with the interaction

enthalpy, it appears that the calculations using the scoring functions based on SQM method will be a promising approach for future drug design, enabled by the rapid development of information technology.

In last part of study, the scoring function based on the PM6-DH2X method was used for halogenated inhibitors that are able to form halogen bonding with CK2 protein kinase. Structures optimized using PM6-D2X agreed very well with the crystal structures, while the structures optimized by the AMBER force field were dissimilar, due to the inability to describe the halogen bonding. For halogenated CK2 inhibitors, based on the crystal and modeled structures, a strong correlation was found between the inhibitory constant and the interaction enthalpy calculated in water. After the inclusion of the entropic term, the correlation deteriorated, due to the inability of the empirical potential to describe the halogen bonding. The correct description of halogenated compounds in complexes with halogen bonding is one of the conditions for the possibility of finding more effective inhibitors.

Publications of the author included in this PhD thesis

- [1] [Dobeš P](#), [Otyepka M](#), [Strnad M](#), [Hobza P](#).: Interaction energies for the purine inhibitor roscovitine with cyclin-dependent kinase 2: correlated ab initio quantum-chemical, DFT and empirical calculations. *Chem. Eur. J.* 2006;12:4297-304.

- [2] [Dobeš P](#), [Fanfrlík J](#), [Řezáč J](#), [Otyepka M](#), [Hobza P](#).: Transferable scoring function based on semiempirical quantum mechanical PM6-DH2 method: CDK2 with 15 structurally diverse inhibitors. *J. Comput. Aided Mol. Des.* 2011;25:223-35.

- [3] [Dobeš P](#), [Řezáč J](#), [Fanfrlík J](#), [Otyepka M](#), [Hobza P](#).: Semiempirical QM Method PM6-DH2X Well Describes Geometry and Energetic of CK2-Inhibitor Complexes Involving Halogen Bonds while Empirical Potential Fails. *J. Phys. Chem. B*, under revision