Univerzita Karlova v Praze
Přírodovědecká fakulta

Studijní program: Makromolekulární chemie
Studijní obor: Makromolekulární chemie

Mgr. Petr Dobeš

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

Typ závěrečné práce

Disertační

Školitel: prof. Ing. Pavel Hobza, DrSc.

Praha, 2011

Prohlášení:

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

V Praze, 15.04.2011

Podpis

Poděkování. Děkuji svému školiteli prof. Ing. Pavlu Hobzovi, DrSc. za zadání velice zajímavého tématu, skvělé vedení a pomoc při vypracování disertační práce. Rád bych taky poděkoval za skvělou spolupráci doc. RNDr. Michalu Otyepkovi, Ph.D., RNDr. Jindřichu Fanfrlíkovi, Ph.D. a RNDr. Janu Řezáčovi, Ph.D., se kterými jsem spolupracoval na publikovaných pracích a při odborných konzultacích. Ještě bych chtěl poděkovat RNDr. Martinu Lepšíkovi, Ph.D za cenné připomínky a inspirativní debaty. Dále bych rád poděkoval členům Centra komplexních molekulových systémů za vytvoření příjemného inspirativního pracovního prostředí.

Obsah

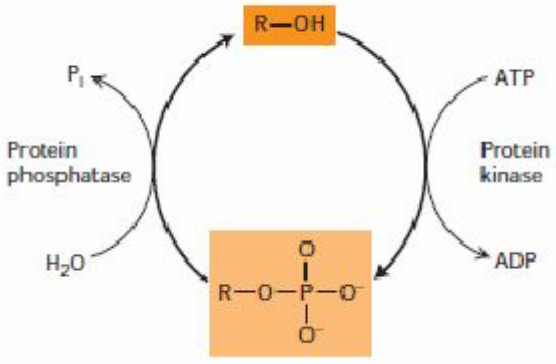
1 Úvod

1.1 Předmět disertační práce

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 řídících vazbu. Ty pak mohou být následně využity při racionálním návrhu účinnějších či specifičtějších inhibitorů.

1.2 Proteinkinázy

Proteinkinázy jsou enzymy, které přenášejí fosfátovou skupinu z molekuly bohaté na energii (např. adenosin-5'-trifosfát, ATP) na specifický proteinový substrát. Tento proces se nazývá fosforylace. Proteinkinázy, společně s enzymy specificky odstraňujícími fosfátové skupiny z proteinů, tzv. fosfatázami, jsou klíčové v mnohých regulačních biologických drahách, jako jsou regulace buněčného cyklu, diferenciace, mezimembránový přenos a sekrece buněčných proteinů (např. růstových hormonů)[1,2,3]. (Obr. 1)

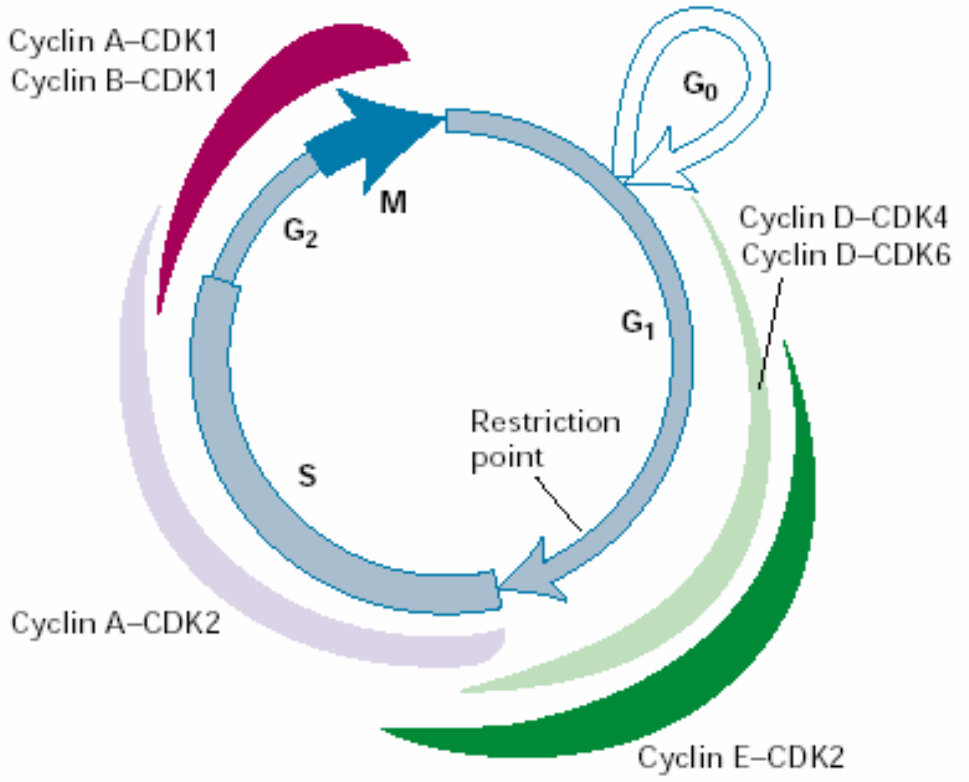


Obrázek 1 Schéma fosforylace proteinů kinázami a defosforylace proteinů fosfatázami, převzato z [4]

Proteinkináz je v lidské eukaryotické buňce přibližně 500 druhů a tvoří asi 2 % proteinů kódovaných v lidském genomu. Podle typu cílové aminokyseliny, která je v substrátu fosforylována rozeznáváme serin-threoninové kinázy, tyrosinové kinázy nebo smíšené kinázy se schopnosti fosforylovat všechny zmíněné aminokyseliny. Malou skupinu tvoří histidinkinázy.

Kvůli klíčové roli proteinkináz v regulaci buněčného cyklu jsou již desetiletí intenzivně studovány možnosti specificky blokovat jednu nebo více proteinkináz pomocí nízkomolekulárních inhibitorů. Takové látky by mohly nalézt uplatnění jako např. protirakovinné léky[5]. Byly navrženy tisíce látek, z nichž nejlepší dosahovaly inhibice proteinkináz v nanomolární koncentraci. Stovky látek byly zkoušeny v klinických testech a desítky byly schváleny pro používání v klinické praxi pro léčbu onkologických onemocnění. Patrně nejznámějším případem z klinické praxe je úspěšné používání inhibitoru Gleevec pro inhibici Abl tyrosinkinázy, která hraje důležitou roli u chronické myeloidní leukemie[6].

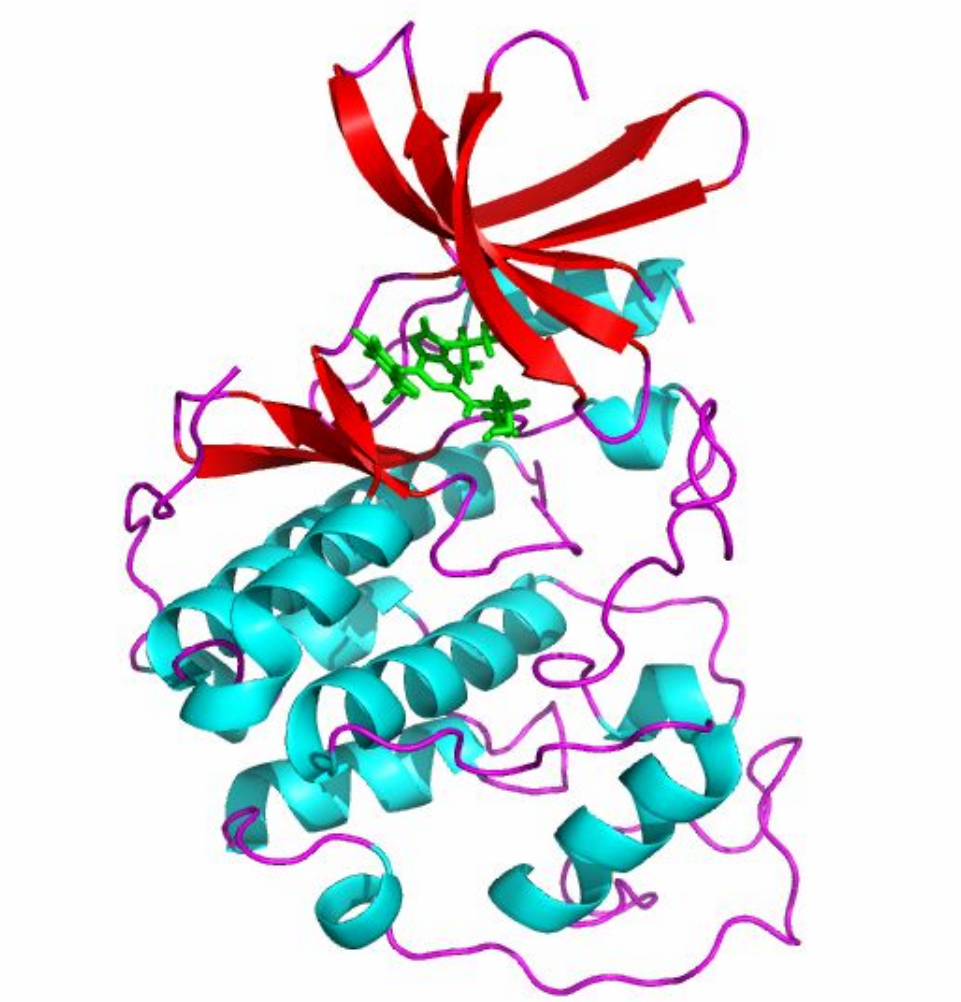
Mezi důležité proteinkinázy patří cyklin-dependentní kinázy (CDK) hrající zásadní roli při řízení buněčného cyklu u eukaryotických organismů.



Obrázek 2 Role CDK2 v buněčném cyklu působící během G1 a G1/S fáze (kontrolní bod – ang. restriction point) Převzato z [4]

Řízení aktivit CDK je zprostředkováno spojením s regulační podjednotkou – proteinem cyklinem nebo fosforylací jinými proteinkinázami, což vede k funkční změně v proteinu[2,7]. Rozeznáváme několik enzymů CDK lišících se v lokalizacích a rolích v regulaci buňky. Důležitou a nejlépe studovanou z pohledu biochemie a strukturní biologie je cyklin-depentní kináza 2 (CDK2)[8], která je jedním z nejdůležitějších regulátorů buněčného cyklu[9,10], působící během G1 a G1/S fáze (Obr. 2).

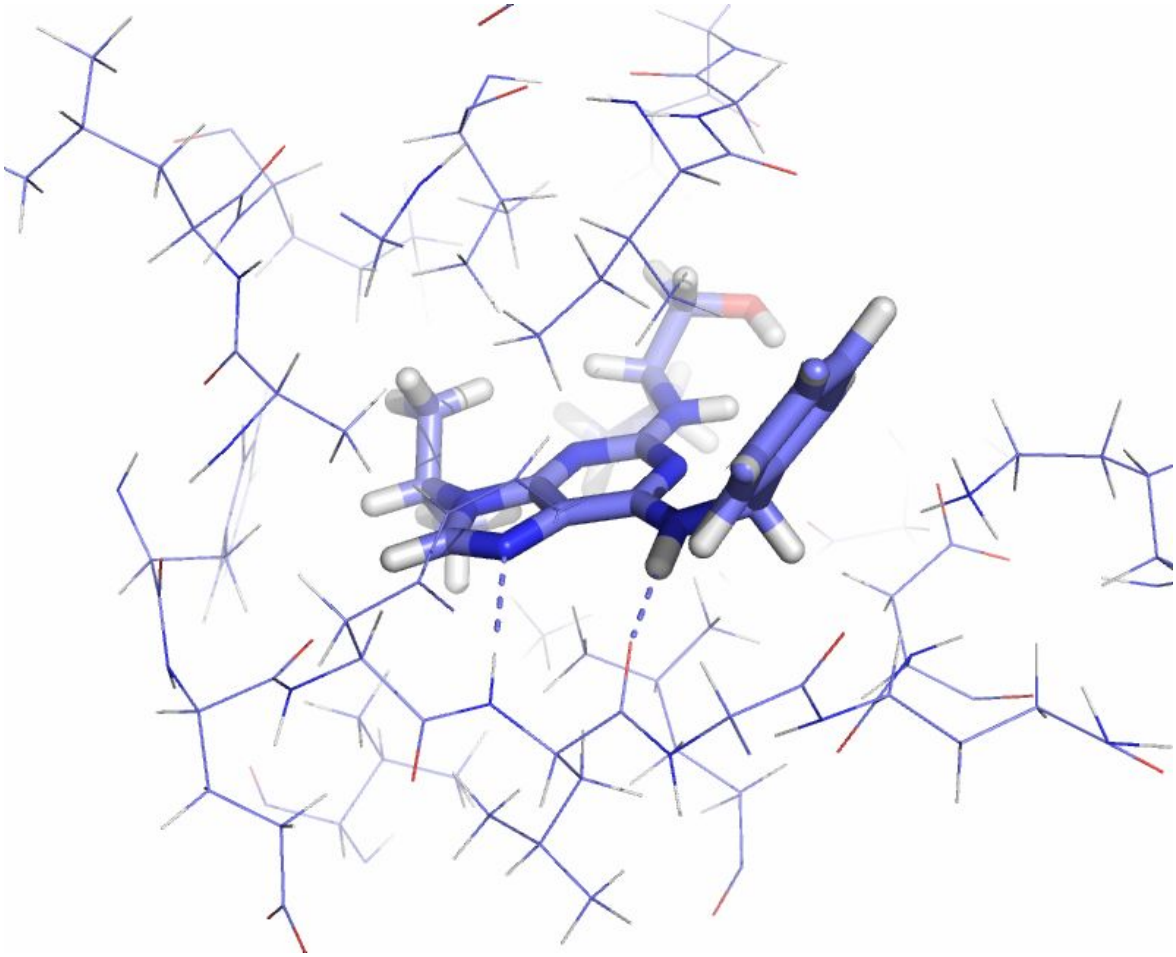
Narušení správné funkce cyklin-dependentních kináz může vyústit ve vývoj nádorů[11]. Deregulace konkrétního enzymu CDK2 je často spojena s mnoha závažnými onkologickými onemocněními[12]. Tato okolnost je důvodem dlouhodobé pozornosti věnované vývoji účinných inhibitorů CDK2[13,14,15,5]. Několik látek je v současné době v klinických zkouškách (např. flavopiridol)[16,17]. Vývoj léčiv, které působí jako inhibitory CDK2 s sebou nese řadu specifických problémů, např. schopnost nádorových buněk množit se, přestože jejich CDK2 je inhibována[18].



Obrázek 3 Struktura proteinkinázy CDK2 s inhibitorem roskovitinem (zeleně, tyčkový model) (PDB kód 2A4L [19]): inhibitor se váže do aktivního místa, které má tvar kavity mezi dvěma laloky, N-terminální lalok je v horní části obrázku a je tvořen převážně β-skládanými listy (červené pruhy). C-terminální lalok je v dolní části obrázku a je tvořen převážně α-šroubovicemi (světle modře). Úseky polypeptidového řetězce bez sekundární struktury (angl. *coil*) jsou znázorněny fialově.

Lidská CDK2 je tvořena 298 aminokyselinami a má supersekundární strukturu (angl. *fold*) charakteristickou pro proteinkinázy. Aktivní místo je mezi N-terminální a C-terminální doménou – lalokem (angl. *lobes*), které jsou spojeny peptidovým řetězcem bez sekundární struktury (angl. *coil*) (Obr. 3). Aktivní místo je umístěno v hluboké dutině, která je schopna vázat ATP a purinové analogy (inhibitory). Je tvořeno třemi oblastmi pojmenovanými po částech ATP, tj. purinové vazebné místo, místo pro fosfátovou skupinu a kapsu pro cukerný zbytek. Doposud byly publikovány desítky struktur komplexů CDK2 s inhibitory založenými na purinovém jádře i inhibitory apurinovými[20,21,22]. Po dlouhodobém studiu inhibice

CDK2 byly ve skupině prof. Strnada v Laboratoři růstových regulátorů v Olomouci objeveny nízkomolekulární purinové inhibitory CDK2 olomoucin[23] a roskovitin[24], které se vyznačují strukturní podobností s ATP. Z důvodu pochopení těchto vlastností do atomárních detailů, věnovali jsme se v první studii předkládané v této disertační práci komplexu CDK2-roskovitin.

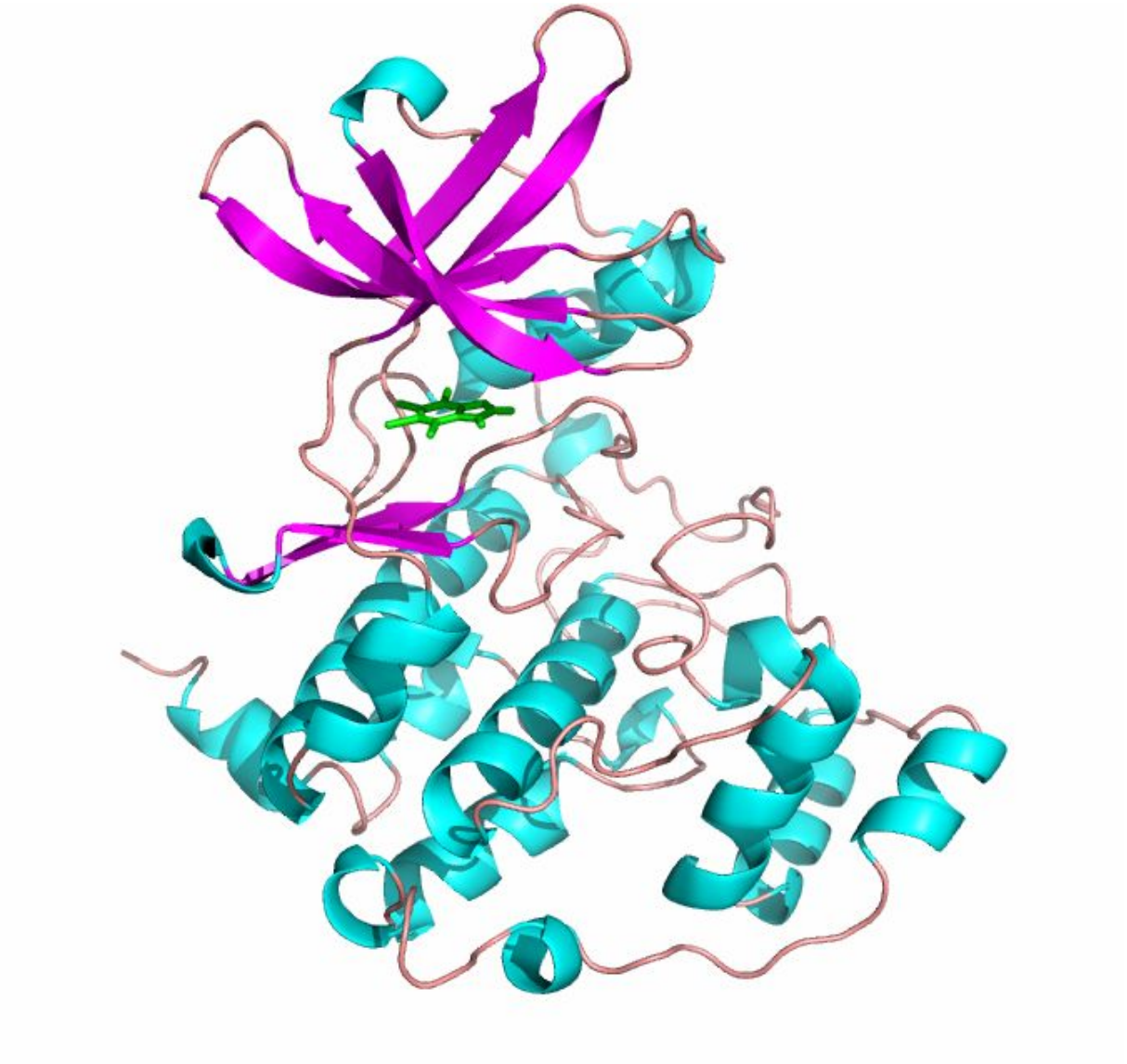


Obrázek 4 Struktura kavity proteinkinázy lidské CDK2 s inhibitorem roskovitinem (PDB kód 2A4L [19])

Vazebný mód CDK2 inhibitorů je charakterizován několika specifickými vodíkovými vazbami k peptidové páteři v aktivním kinázovém místě (Obr. 4). Vazebný mód roskovitinu je charakterizován dvěma vodíkovými vazbami mezi roskovitinem a peptidovou vazbou leucinu 83.

Další zkoumaná proteinkináza CK2 (angl. *casein kinase 2*) je pleotropní (působící na více substrátů) serin-threoninová kináza s více než stovkou cílových proteinů majících různé regulační buněčné funkce, např. buněčná exprese, syntéza a degradace proteinů, případně signalizace a potlačení apoptózy[25,26]. CK2 jsou více exprimovány v nádorových

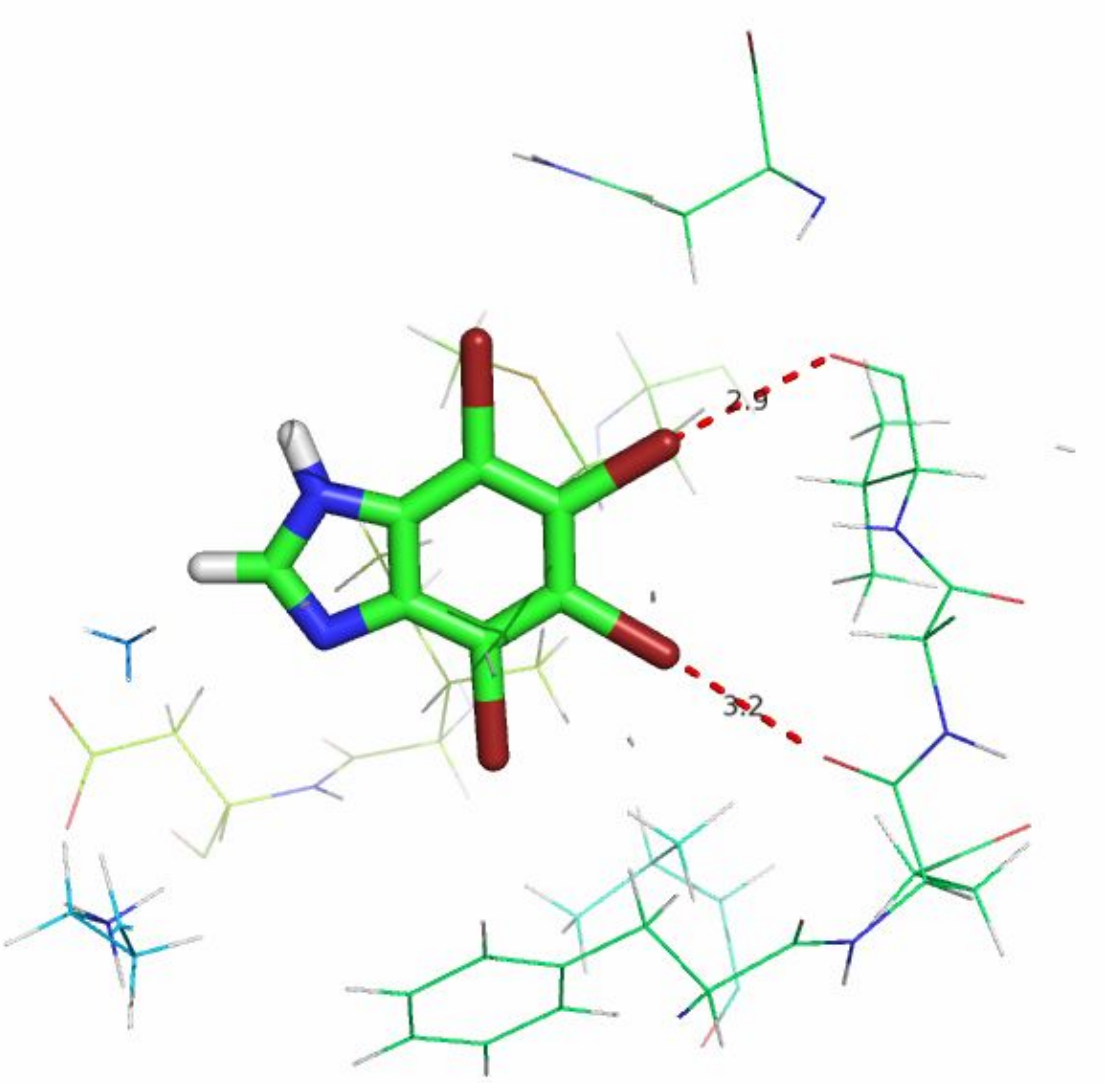
buňkách[27]. Snížení aktivity CK2 může být jedním z možných přístupů k léčení nebo k zpomalení růstu nádorů[28,29]. CK2 se skládá ze dvou katalytických (α) a dvou regulačních domén (β) (Obr. 5)[30]. Aktivní místo α-podjednotky CK2 leží mezi N-terminální a C-terminální doménou, které jsou spojeny peptidovým řetězcem bez sekundární struktury podobně jako u CDK2. Aktivní místo CK2 je umístěno v kavitě, která je schopna vázat ATP.



Obrázek 5 Struktura komplexu CK2 α-podjednotky s inhibitorem (zeleně, tyčkový model) (PDB kód: 2OXY): inhibitor se váže do hluboké kavity aktivního místa mezi N-terminální (nahoře) a C-terminálním (dole) lalokem. Barevné značení sekundárních struktur je stejné jako na Obr. 3.

Byly vyvinuty desítky CK2 inhibitorů, jejichž střední inhibiční koncentrace (IC50) jsou v řádu mikromolárním až nanomolárním [31,32]. Mezi nejčastěji používaný inhibitor CK2 patří tetrabromobenzotriazol (TBB), který má inhibiční konstantu (K_i) 0,4 μM. V současnosti je

popsána řada krystalových struktur tetrabromovaných a tretrajodovaných inhibitorů podobných TBB s kukuřičnou CK2 mající 77 % homologii (sekvenční identita) s lidskou α-podjednotkou CK2[33], pro kterou však nejsou krystalové struktury tohoto typu inhibitorů (Obr. 6). Tyto inhibitory se vážou do vazebného místa pro ATP katalytické α-podjednotky CK2. Inhibitory jsou vázané v malé hydrofobní kavitě[34], což vysvětluje jejich velkou selektivitu vůči dalším proteinkinázám. Vedle nepolárních interakcí významných pro zvýšení síly vazby této řady inhibitorů hrají důležitou roli také polární interakce. Ty jsou zodpovědné za orientaci inhibitoru v kavitě. Není to způsobeno klasickou vodíkovou vazbou, ale přítomností dvou halogenových vazeb typu kyslík - halogen[35] (Obr. 6).

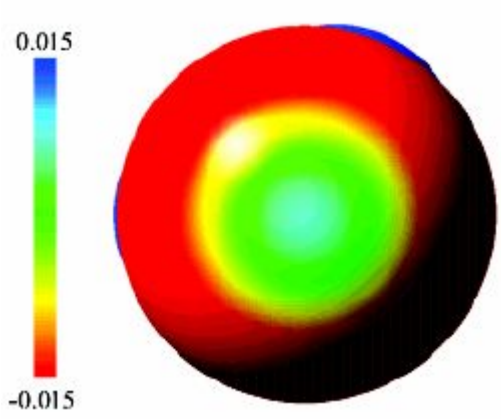


Obrázek 6 Kavita katalytického místa α-podjednoty kukuřičné proteinkinázy CK2 s inhibitorem (PDB kód – 2OXY) Dvě halogenové vazby mezi dvěmi atomy bromu inhibitoru a karbonylovými kyslíky peptidové páteře.

Po nahrazení bromu jódem došlo u těchto tetrahalogenovaných inhibitorů ke zvýšení inhibiční konstanty, až o jedno desetinné místo[32]. Uvedená řada inhibitorů tak představuje vhodný model pro studium halogenové vazby u komplexů proteinu s inhibitory.

1.3 Nekovalentní interakce

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. Celková interakční energie se skládá z atraktivní (elektrostatické, indukční a disperzní) a repulzní složky[36]. Elektrostatické síly působí mezi permanentními elektrickými multipóly, indukční síly působí mezi permanentním multipólem a indukovaným multipólem. Disperzní (Londonova) síla vzniká mezi časově proměnnými a indukovanými multipóly v molekulách a tak může existovat i mezi molekulami bez permanentního elektrického multipólu. Disperzní energie tvoří atraktivní část van der Waalsovy energie a jako taková hraje důležitou úlohu ve stabilizaci biomakromolekul[37]. Repulzní interakce se uplatní vždy, když vzdálenost mezi molekulami je dostatečně malá a překryv jejich elektronových hustot je nenulový. Komplexy s vodíkovou vazbou patří k nejsilněji vázaným nekovalentním komplexům. Vodíková vazba X-H…Y je charakterizována interakcí mezi donorem X-H obsahujícím vodík kovalentně vázaný na elektronegativní atom X (O, N, F...) a akceptorem Y, tj. místem s přebytkem elektronů, jako jsou volné elektronové páry atomů O, N, F, atp. nebo násobné vazby [38]. Mezi méně známé typy nekovalentních interakcí patří halogenová vazba C-X…Y (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 např. atom kyslíku.



Obrázek 7 Ukázka σ-díry pro molekulu CH_3Br pomocí vypočítaného molekulárního elektrostatického potenciálu. Převzato z [39]

Schopnost být Lewisovou kyselinou je u halogenových atomů způsobena přítomností tzv. σ-díry (angl. *σ-hole*), tj. pozitivně nabité oblasti na atomu halogenu v ose vazby C-X ve

vzdálenější části od atomu uhlíku. (Obr. 7). Halogenová vazba se tvoří v důsledku kvantových efektů a jako taková nemůže být popsána klasickými metodami molekulové mechaniky (tj. bez explicitního popsání distribuce elektronové hustoty na atomu). Kvantově chemické metody halogenovou vazbu popisují uspokojivě.

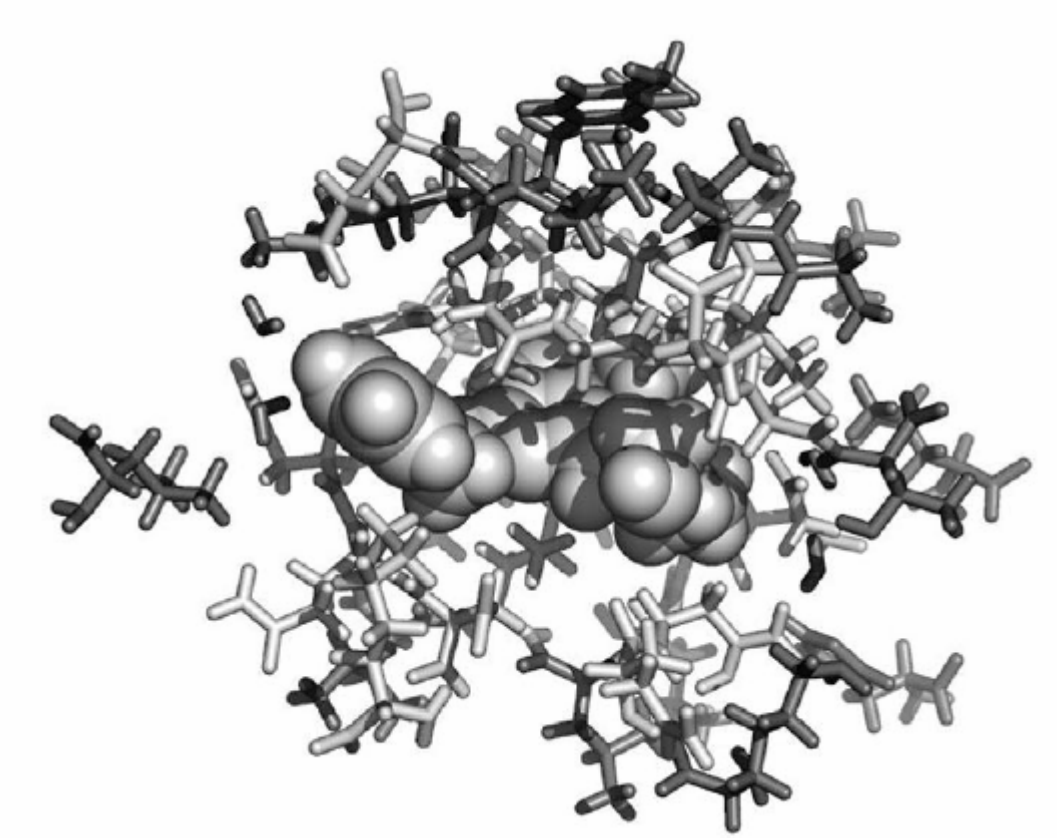
1.4 Cíle

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 energii pro komplexy kukuřičné a lidské CK2 s halogenovanými inhibitory, které jsou s enzymem schopny tvořit halogenovou vazbu.

2 Studované systémy

2.1 CDK2 a inhibitor roskovitin

V první části práce byl studován komplex CDK2 s inhibitorem roskovitinem vázaným v kinázové kavitě (Obr. 8). Studovaný systém o velikosti okolo pěti tisíc atomů je pro přesné metody kvantové chemie příliš velký. Bylo však zjištěno, že pro uspokojivé popsání interakcí v komplexech proteinu s inhibitorem je dostatečné uvažovat pouze ty aminokyseliny, které jsou vzdáleny od inhibitoru roskovitinu o méně než 5 Å[40].



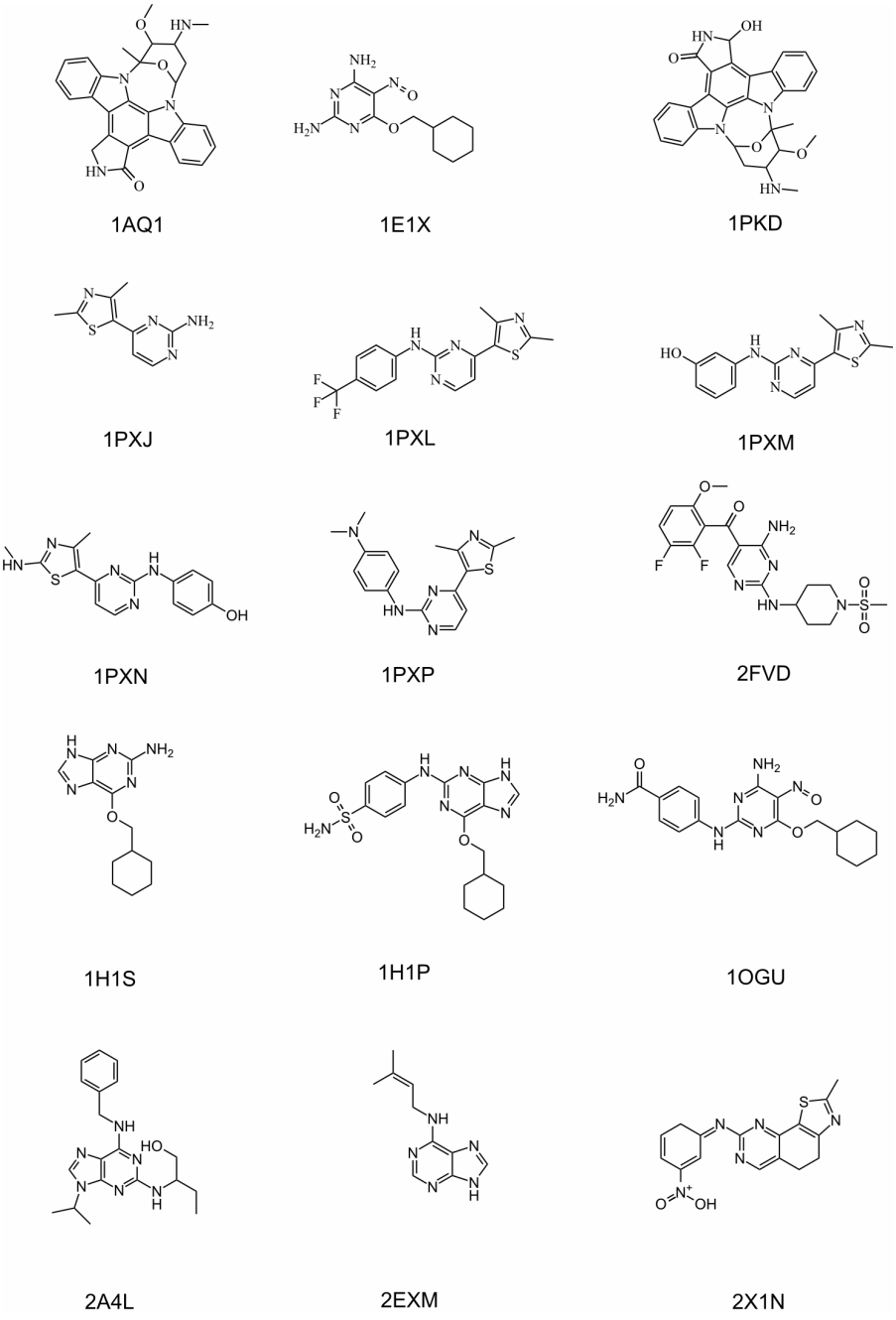
Obrázek 8 Aminokyseliny kavity CDK2 s navázaným inhibitorem roskovitinem

Vzhledem k tomu, že v blízkosti inhibitoru roskovitinu se nachází peptidová páteř spojující blízké aminokyseliny do příliš velkého systému, byla peptidová páteř přerušena mezi atomy uhlíku C_α a dusíku. Takto bylo vytvořeno 14 fragmentů s roskovitinem a s jednou nebo více aminokyselinami: kyselinou glutamovou 8 (E8); isoleucinem 10 (I10); valinem 18 (V18); alaninem 31 (A31); lysinem 33 a asparagovou kyselinou 145 (K33D145); valinem 64 (V64); fenylalaninem 80 (F80); kyselinou glutamovou 81, fenylalaninem 82 a leucinem 83 (E81F82L83); leucinem 83, histidinem 84 glutaminem 85 (L83H84Q85); glutaminem 85,

kyselinou asparagovo 86 and lysinem 86 (Q85D86K89); glutaminem 131 and asparginem 132 (Q131N132); leucinem 134 (L134); alaninem 144 (A144); glycinem 11, kyselinou glutamovou 12 a glycinem 13 (G11E12G13). Na těchto fragmentech byla počítaná interakční energie pomoci kvantověchemických metod. (Podrobnější popis přípravy studovaných struktur je uveden v metodách článku u přílohy 1).

2.2 CDK2 a inhibitory

Bylo uvažováno 15 struktur CDK2 se strukturně odlišnými inhibitory (Obr. 9), které byly vzaty z PDB databáze struktur (angl. *Protein Data Bank*). Studované systémy CDK2 měly dva konformační stavy, jeden v neaktivním stavu a druhý v aktivním stavu s fosforylovaným threoninem 160 a proteinem cyklinem, který byl pro vlastní výpočet odstraněn. Všechny uvažované inhibitory se vázaly do kinázové kavity CDK2 směrem k peptidové páteři proteinu prostřednictvím specifických motivů tvořených vodíkovými vazbami. (Podrobnější popis přípravy studovaných struktur je uveden v metodách článku u přílohy 2).

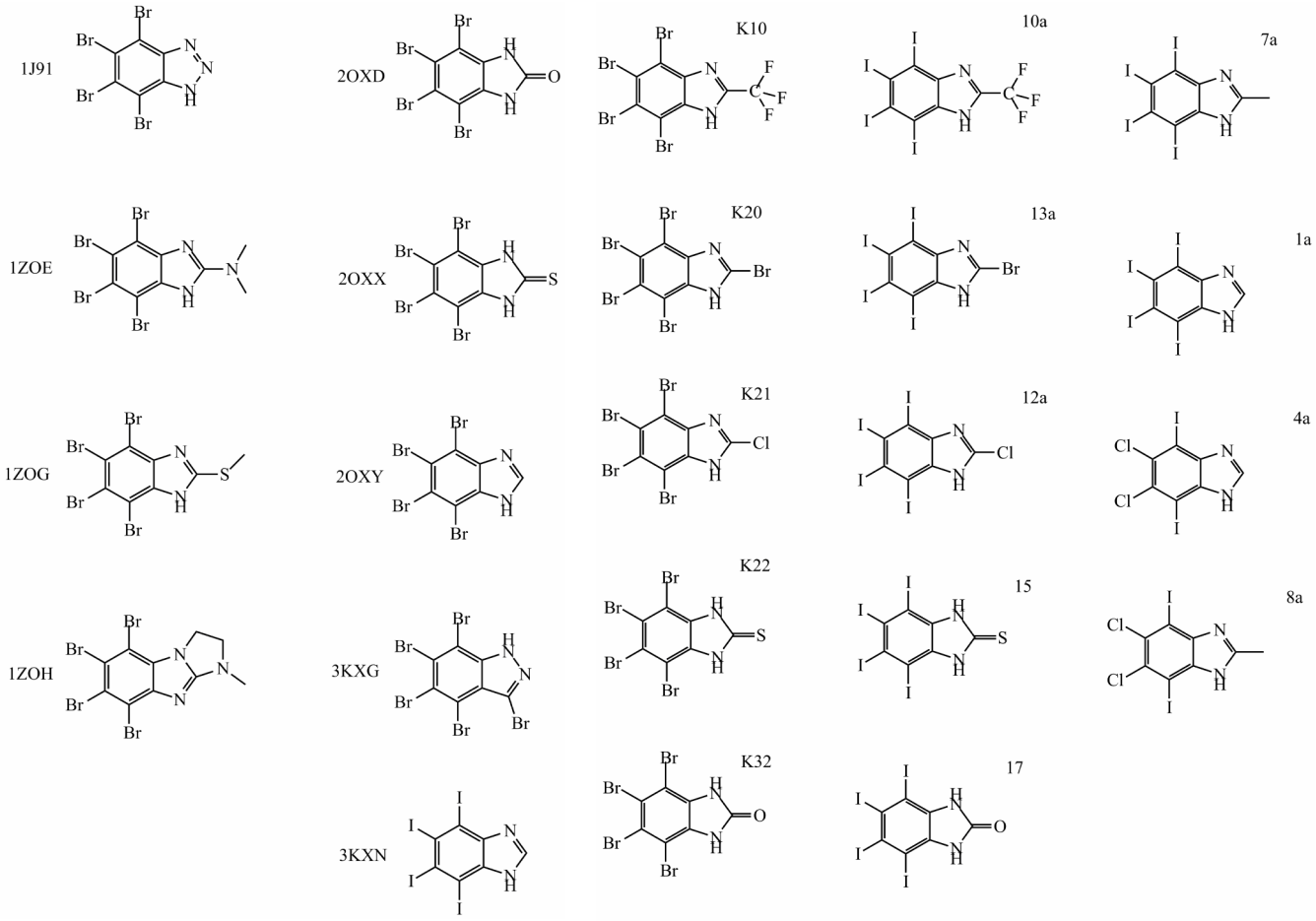


Obrázek 9 15 strukturně odlišných CDK2 inhibitorů s kódy PDB struktur jejich komplexu s CDK2

2.3 CK2 a inhibitory

Byly použity dvě skupiny struktur komplexu CK2 s inhibitorem. První skupina se skládala z devíti komplexů (osm s tetrabromoderiváty a jedním tetrajododerivátem) katalytické α-

podjednotky kukuřičné CK2, pro které byla strukturní data získána z PDB databáze struktur (Obr. 10 vlevo). Obecně zaujímají halogenované inhibitory CK2 dvě orientace, které se liší otočením asi o šedesát stupňů okolo osy kolmé k rovině inhibitoru. V komplexech s první orientací inhibitorů (PDB kódy 1ZOG, 1ZOH, 3KXG, 3KXN) se inhibitory váží přes Br5 a Br6 ke kyslíkům kyseliny glutamové 114 a kyslíkům páteře valinu 116, zatímco v komplexech s druhou orientací inhibitorů (PDB kódy 1ZOE, 2OXD, 2OXX, 2OXY) se inhibitory váží přes Br4 a Br5. V případě, kdy jsou obě orientace přítomny v krystalových strukturách (1ZOH a 3KXN) byla uvažována druhá orientace.



Obrázek 10 Dvě skupiny CK2 inhibitorů, první skupina z krystalových struktur kukuřičné CK2, druhá skupina z navržených struktur inhibitorů, které byly studovány v komplexu s lidskou CK2.

Druhá skupina struktur se skládala ze sady modifikovaných inhibitorů (navržených tak, aby měly maximální počet atraktivních interakcí s CK2) ovšem v komplexu s lidskou CK2. Struktury komplexů byly vytvořeny strukturním přeložením, kdy α-podjednotka lidské CK2 byla získána z krystalové struktury (PDB kód 1JWH – komplex CK2 s gama-imino-ATP (ANP, což je analog ATP)) a struktura inhibitoru byla získána z komplexu kukuřičné CK2 α-

podjednotky s inhibitorem (PDB kód 1ZOE). Následně byly modelovány různé modifikace inhibitorů (Obr. 10 vpravo)[32]. (Podrobnější popis modelování studovaných struktur je uveden v metodách článku přílohy 3).

3 Metody

3.1 Metody molekulové mechaniky

Molekulová mechanika popisuje molekuly pomocí zákonů klasické fyziky. Tento popis je vyjádřen pomocí tzv. silového pole, což je množina parametrů a matematických funkci, které popisují potenciální energii systému částic jako jsou atomy a molekuly v závislosti na jejich souřadnicích[41]. Je to jeden ze základních matematických modelů popisující fyzikálně chemickou strukturu molekul. Každý atom je popsán parciálním nábojem a atomovým typem. Podle atomových typů se přiřadí parametry silového pole (změny délky vazeb, ohýbání vazebných úhlů a deformace torzních úhlů). Empirické parametry používané ve funkci silového pole jsou odvozeny z experimentálních dat nebo získány pomocí výpočtů kvantové chemie. Tato funkce se skládá z vazebných a nevazebných členů. Vazebné členy popisují změny délky vazeb, ohýbání vazebných úhlů a deformace torzních úhlů. Nevazebné členy popisují interakce van der Waalsovy a elektrostatické interakce. Výhody a nevýhody silového pole vycházejí z jeho jednoduchého matematického popisu – jsou výpočetně nenáročné, ale na druhou stranu nejsou schopny popsat kvantové jevy jako přenos náboje, polarizační děje nebo existence σ –díry u halogenů. Jedním z nejpoužívanějších silových polí je AMBER[42].

3.2 Metody kvantové chemie

Pomocí kvantověchemických výpočtů je v principu možné přesně předpovídat fyzikálně-chemické vlastností molekul i molekulových klastrů. Je však třeba mít na paměti omezení, jež mají povahu jak matematickou tak i fyzikálně-chemickou. Metody kvantové chemie poskytují přesnější výsledky ve srovnání s metodami molekulové mechaniky, příslušné výpočty jsou však řádově náročnější, jak na výpočetní čas, tak na počítačový hardware. Základní metodou kvantové chemie je metoda Hartreeho a Focka (HF)[43]. Slouží k výpočtu základního stavu víceelektronové vlnové funkce při konkrétním uspořádání atomových jader (Born-Oppenheimerova aproximace), která je vyjádřena jako jednoduchý Slaterův determinant. Využívá variačního principu k výpočtu minimální energie systému. Výpočet se provádí řešením nelineárních rovnic, kdy každý elektron je popsán Fockovým operátorem v poli, které je tvořeno ostatními elektrony. Tato metoda bývá označována jako metoda SCF (angl. *self-*

consistent field). Příslušné rovnice jsou řešeny iterativní metodou k dosažení minima energie. Molekulové orbitaly jsou popsány jako lineární kombinace atomových orbitalů - bázových funkcí. 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 korelované metodu patří poruchová metoda druhého řádu podle Møllera a Plesseta (MP2), která poskytuje přiměřené zahrnutí elektronové korelace. Metoda spřažených klastrů CCSD(T) (zahrnující jednoelektronové a dvouelektronové excitace pomocí iterativního způsobu a tříelektronové excitace pomocí poruchového počtu)[44] poskytuje velmi přesné hodnoty energie a vlastnosti 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[45]. 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[46], což je aproximativní metoda DFT založená na principu těsné vazby. Po zahrnutí disperze empirickým členem vzniká metoda SCC-DFTB-D[47], která je schopna popsat disperzní interakce. Žádná z uvedených metod však není vhodná pro rozsáhlé systémy s tisíci a více atomy a tyto metody jsou proto nepoužitelné pro biomakromolekuly, např. proteiny. Jednou z možností řešení je použití 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.

Nejvhodnější z SQM metod pro použití na biomolekulové systémy je metoda PM6-DH2[48,49] 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, což je způsobeno nedostatečnou repulzí mezi halogenem a elektronegativním atomy (většinou O nebo N). Bylo proto nutné zahrnout dodatečný člen, který zvýší zmíněnou repulsi a tím dojde ke zvětšení vzdálenosti mezi halogenem a elektronegativním atomem. Takto opravená metoda (PM6-DH2X) je schopna popsat halogenovou vazbu srovnatelně s náročnými metodami kvantové chemie[50].

3.3 Skórovací funkce

Afinita inhibitoru k jeho receptoru souvisí s vazebnou volnou energii protein-inhibitorového komplexu ve vodě. Vazbu kompetitivního inhibitoru (I) s proteinem (P) lze vyjádřit pomocí inhibiční konstanty K_i definovanou pro rovnováhu (P + I → PI) nebo případně také pomocí koncentrace inhibitoru potřebné k snížení aktivity enzymu na polovinu, tzv. IC_{50} [51]. Hodnoty IC50 jsou ale silně závislé podmínkách měření. Inhibiční konstanta je proto vhodnější pro srovnání experimentu s vypočtenými výsledky. Inhibiční konstanta souvisí s vazebnou volnou energií pomocí vztahu $\Delta G'_b = RT\ln(K_i)$, kde R je molární plynová konstanta (R = 8,314 *J/·K/·mol*) a T je termodynamická teplota v Kelvinech. Odhad vazebné volné energie nazývaný skóre může být získán jako součet těchto členů: interakční entalpie, interakční entropie, deformační energie inhibitoru a proteinu a změna solvatace proteinu a inhibitoru při vzniku komplexu[52]. Klasické skórovací funkce jsou založené na použití empirického potenciálu. Navzdory zvyšování účinnosti a parametrizaci empirických skórovacích funkcí, jejich možnosti jsou v podstatě omezeny známými limitacemi empirických potenciálů (molekulová mechanika, silové pole). Toto převážně souvisí s neschopností zahrnutí popisu kvantových efektů, jako je např. přenos náboje mezi proteinem a inhibitorem[53,54] a existence σ –díry v halogenu, která umožňuje vznik halogenové vazby hrající významnou roli při vazbě halogenovaných inhibitorů[55]. Řešením je použití kvantověchemických metod. Jak však bylo uvedeno dříve, tyto metody jsou vzhledem k časové náročnosti pro účely návrhu léčiv nepoužitelné. S výhodu však lze použít pokročilou semiempirickou QM metodu PM6-DH2X. Tato metoda používá algoritmu MOZYME[56] založeném na lokalizovaných orbitalech, což vede k výraznému zrychlení výpočtu. V předložené práci se skóre počítá jako součet interakční entalpie ve vodě, entropického členu, deformační energie inhibitoru a korekce solvatační energie inhibitoru (rov. 1)

$$\Delta G'_w = \Delta H_w - T\Delta S_w + \Delta E_{def}(I) + \Delta E_{def}(P) + \Delta\Delta G_w(I) \qquad (1)$$

První člen rov. 1 popisující změnu entalpie při interakci inhibitoru s proteinem ve vodném prostředí patří mezi nejvýznamnější členy. Změna deformační energie inhibitoru (třetí člen rov. 1) se vyjadřuje pomocí následujícího vztahu:

$$\Delta E_{def}(I) = E(I)^{PI} - E(I)^{I} \qquad (2)$$

kde $E(\mathrm{I})^{\mathrm{PI}}$ je energie inhibitoru v geometrii komplexu a $E(\mathrm{I})^{\mathrm{I}}$ je energie izolovaného inhibitoru ve vodě. Deformační energie proteinu (čtvrtý člen v rov. 1) lze v případě zjišťování relativních vazebných energií série inhibitorů k jednomu proteinu zanedbat. Další člen v rov. 1 představuje korekci na solvataci inhibitoru:

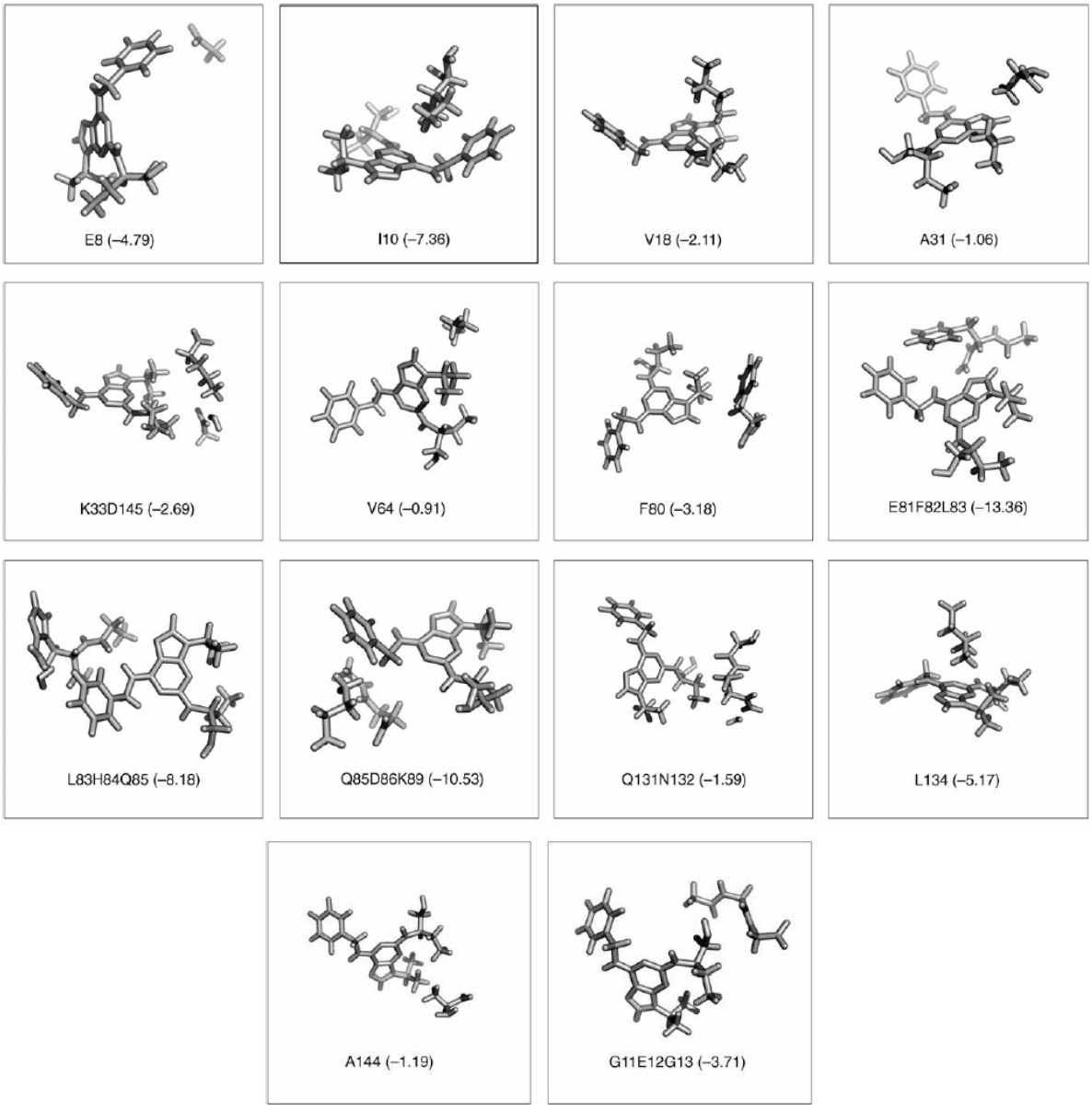
$$\Delta\Delta G_{\mathrm{w}}(\mathrm{I}) = \Delta G_{\mathrm{w}}{}^{\mathrm{COSMO}}(\mathrm{I})^{\mathrm{PI}} - \Delta G_{\mathrm{w}}{}^{\mathrm{SMD}}(\mathrm{I})^{\mathrm{PI}} \qquad (3)$$

kde člen $\Delta G_{\mathrm{w}}{}^{\mathrm{COSMO}}(\mathrm{I})^{\mathrm{PI}}$ a $\Delta G_{\mathrm{w}}{}^{\mathrm{SMD}}(\mathrm{I})^{\mathrm{PI}}$ představuje solvatační volnou energii spočítanou pomocí modelu COSMO v programu MOPAC[57] a pomocí modelu SMD v programu Gaussian 09 [58]. Entropický člen byl určen na základě silového pole AMBER[42] v implicitním solventu pomocí aproximace tuhého rotoru – harmonického oscilátoru. Je to proto, že výpočet druhých derivací energie podle všech koordinát je výpočetně velmi náročný a nelze jej provést pomocí kvantověchemických metod.

4 Výsledky a diskuse

4.1 CDK2 s inhibitorem roskovitinem

V publikaci u přílohy 1 byla studována interakce mezi CDK2 a inhibitorem roskovitinem. Pro výpočet interakční energie komplexu CDK2-roskovitin bylo nejprve nutné zmenšit popisovaný systém tak, aby jej bylo možné popsat kvantověchemickými metodami.



Obrázek 11 Fragmenty CDK2 kavity s inhibitorem roskovitinem a s MP2 interakčními energiemi (kcal/mol)

Z literatury je známo[40], že zahrnutí aminokyselin ležících do 5 Å od atomů inhibitoru je dostatečné k popsání interakčních energii mezi proteinem a inhibitorem. Proto jsme použili

toto okolí roskovitinu z komplexu s CDK2 a dále jsme jej rozdělili na menší části (Obr. 11), pro které bylo možno použít náročnější kvantověchemické metody. Jako referenční metoda byla použita metoda MP2/aug-cc-pVDZ která, jak bylo dříve ukázáno dává podobné výsledky jako velmi přesná metoda CCSD(T)/CBS[59]. Je to díky kompenzaci chyb (přeceňování (metoda MP2) a podceňování (poměrně malá baze) interakční energie[60]). 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). Výsledky výpočtů jsou uvedeny v tabulce 1. Hodnoty interakční energie byly spočteny ve vakuu a tak nezahrnují vliv proteinového okolí ani solventu. Metoda DFT selhala u fragmentu, kde interakční energie je disperzního charakteru (např. L134). Pro případy fragmentů s vodíkovou vazbou (E81F82L83) metoda DFT poskytuje srovnatelné hodnoty interakční energie s referenčními hodnotami. Podobných výsledků bylo dosaženo také pomocí metody SCC-DBTB, která nezahrnuje disperzní interakci. Po přidání empirického disperzního členu byly hodnoty interakčních energií (např. L134) blízké hodnotám referenční metody MP2. Srovnatelné hodnoty s referenčními daty také poskytlo silové pole ff99 AMBER, které zahrnuje empirický disperzní člen. Pomocí metody SCC-DFTB-D a silového pole AMBER (ff99) byla spočítána interakční energie roskovitinu s kavitou a bylo ukázáno, že dominantním atraktivním členem je disperzní člen. Obecně metody DFT nezahrnující disperzní korekci nejsou schopny popsat interakci CDK2 s roskovitinem, a proto nejsou vhodné pro další výpočty těchto a podobných komplexů.

fragment	MP2	DFT	SCC-DFTB	SCC-DFTB-D	AMBER
E8	-4,7	-3,8	-2,0	-2,4	-1,8
I10	-7,3	1,4	-1,0	-7,5	-7,1
V18	-2,1	2,1	0,1	-2,9	-2,8
A31	-1,0	2,9	0,3	-1,8	-1,6
K33D145	-2,6	1,9	0,1	-3,1	-1,8
V64	-0,9	0,4	0,0	-1,0	-1,0
F80	-3,1	0,9	0,0	-3,1	-3,1
E81F82L83	-13,3	-6,1	-4,3	-9,7	-9,0
L83H84Q85	-8,1	0,9	-2,5	-8,5	-8,5
Q85D86K89	-10,5	-5,0	-2,5	-12,0	-7,9
Q131N132	-1,5	1,6	0,6	-2,1	-0,9
L134	-5,1	2,6	-0,1	-5,2	-5,5
A144	-1,1	-0,9	-0,0	-2,0	-1,6
G11E12G13	-3,7	-1,5	-1,2	-3,3	-4,2
součet	-65,8	-0,5	-12,6	-65,3	57,3

Tabulka 1 Interakční energie fragmentu z CDK2 kavity s inhibitorem roskovitinem (kcal/mol)

Zde je třeba podotknout, že v poslední době bylo navrženo několik schémat, jak zahrnout empirickou disperzi i do DFT výpočtu[61,62] a tato metoda (DFT-D) se stala velmi oblíbenou pro popisy biomolekulových komplexu[63,64].

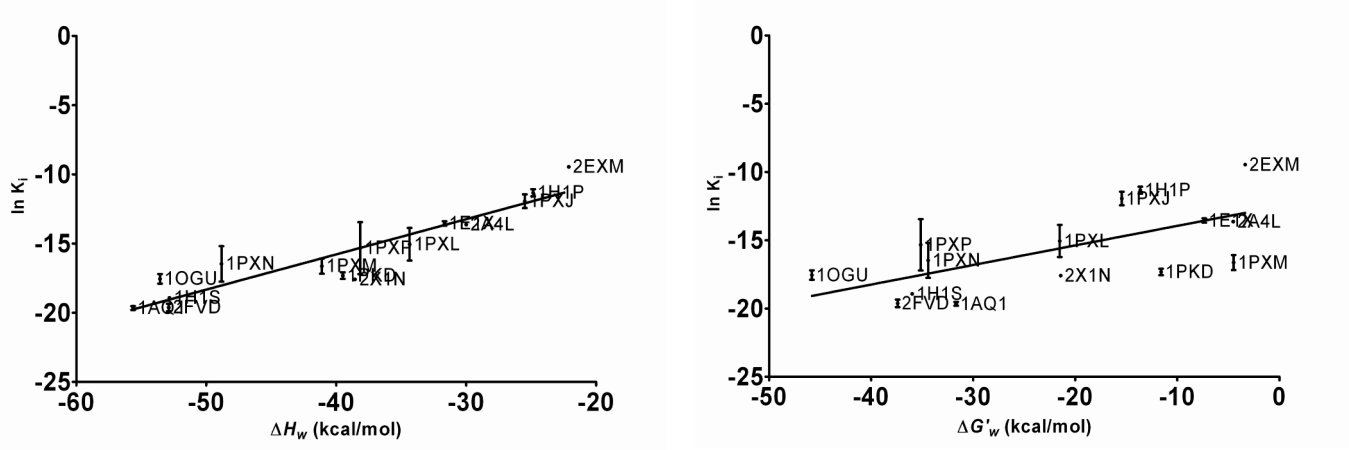
Dominantní vliv disperzních sil pro vazbu roskovitinu s CDK2 vyžaduje použití kvantověchemických metod zahrnujících disperzní interakci a to buď explicitně jako u metody MP2 nebo pomocí empirické korekce jako u metody SCC-DFTB-D. Dále je zřejmé, že metoda silového pole AMBER, která zahrnuje disperzní energii ve van der Waalsově členu, je též vhodná pro popis této interakce.

Dalším důležitým faktorem, který stabilizuje roskovitin v kavitě CDK2 jsou vodíkové vazby. Na rozdíl od disperzní energie mají vodíkové vazby směrový charakter a tvoří vazebný motiv, který je typický pro další inhibitory CDK2. Z uvedených hodnot lze také usoudit, že pouze omezená skupina aminokyselin má zásadní podíl na stabilizaci komplexu. Tyto informace dovoluji odhadnout, které části inhibitoru roskovitinu mohou být modifikovány tak, aby došlo ke zvýšení afinity inhibitoru, a také naopak změny kterých aminokyselinových reziduí při mutacích mohou způsobit destabilizaci vazby inhibitoru roskovitinu s CDK2.

4.2 CDK2 inhibitory

V druhé publikaci (příloha 2) byla studovaná korelace teoretických a experimentálních vazebných volných energií na strukturně různorodé skupině 15 inhibitorů CDK2, které se kompetitivně vážou do aktivního místa CDK2. Interakce mezi CDK2 a inhibitory v roztoku je experimentálně popsána pomocí inhibiční konstanty (K_i). Přirozený logaritmus inhibiční konstanty ($\ln K_i$) je úměrný vazebné volné energii ΔG_b dle rovnice $\Delta G_b = RT \ln K_i$. Volná energie vazby (skóre) byla spočtena dle rov. 1. Příslušná skórovací funkce je založena na metodě PM6-DH2, a tak se zásadně liší od skórovacích funkcí založených na empirických potenciálech. Tato metoda dává přesné výsledky pro různé typy nekovalentních interakcí jako jsou vodíkové vazby a disperzní interakce[65,49], které hrají důležitou roli pro stabilizaci komplexu CDK2 s roskovitinem (viz. předchozí kapitola). Nejprve byla uvažována korelace experimentálně změřených inhibičních konstant s vypočítanou interakční entalpií ve vodném 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) (Obr. 12 vlevo).



Obrázek 12 Korelace experimentálních a teoretických hodnot vazebných energií 15 inhibitoru s CDK2: vlevo ΔH_w vs. $\mathbf{lnK_i}$, vpravo ΔG_w vs. $\mathbf{lnK_i}$

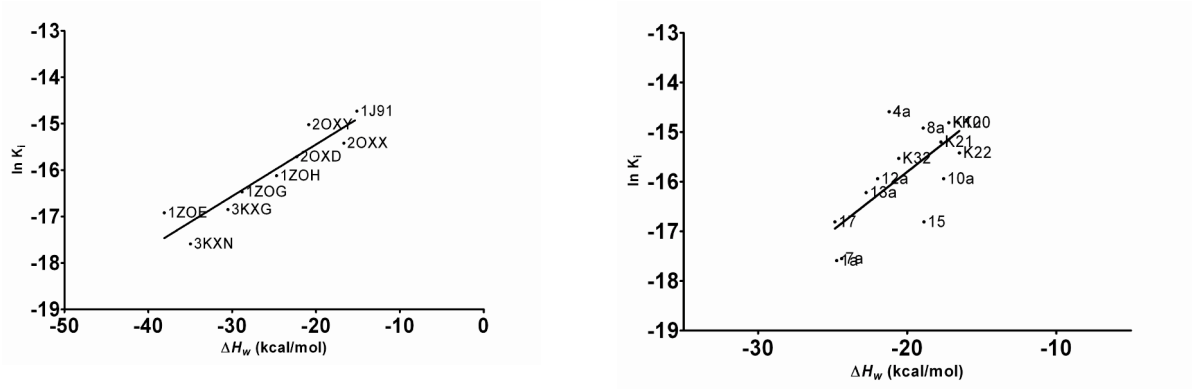
Přidáním korekce na deformaci a desolvataci inhibitoru se korelace mírně zhoršila (R^2=0,77) a konečně započtením entropického členu došlo k dalšímu zhoršení korelace (R^2=0,52) (Obr. 12 vpravo). Zde je nutno zmínit, že přesnost popisu entropického členu je často nižší než u entalpického členu[66]. Hlavním důvodem je ale skutečnost, že entropický člen byl určen pomocí silového pole AMBER (ff99). Na základě vypočtených dat lze konstatovat, že nejlepší korelace s inhibiční konstantou byla získána pro interakční entalpii určenou pomocí PM6-DH2 metody. Tato korelace byla silnější než všechny dříve publikované korelace pro inhibitory CDK2 založené na empirických a semiempirických skórovacích funkcích [67,68,69].

4.3 CK2 inhibitory

V roce 2004 byl publikován článek s přehledy struktur proteinů a nukleových kyselin v komplexu s nízkomolekulárními inhibitory, kde nekovalentní stabilizace měla příspěvek od halogenových vazeb[70]. V posledních letech byla zaměřena pozornost na možnost použití halogenové vazby v racionálním navrhování léčiv[71,55]. Velká skupina léčiv obsahující atom halogenu může totiž vytvářet halogenové vazby. Vznik halogenové vazby je způsoben kvantovým jevem, kdy dojde na vrcholu halogenu kovalentně vázaného s uhlíkem ke vzniku oblasti s kladným nábojem, tzv. σ-díry (angl. *σ-hole*). Empirické potenciály nemohou σ-díru popsat a tak tyto metody nejsou pro popis halogenové vazby vhodné. V uvedené studii byla

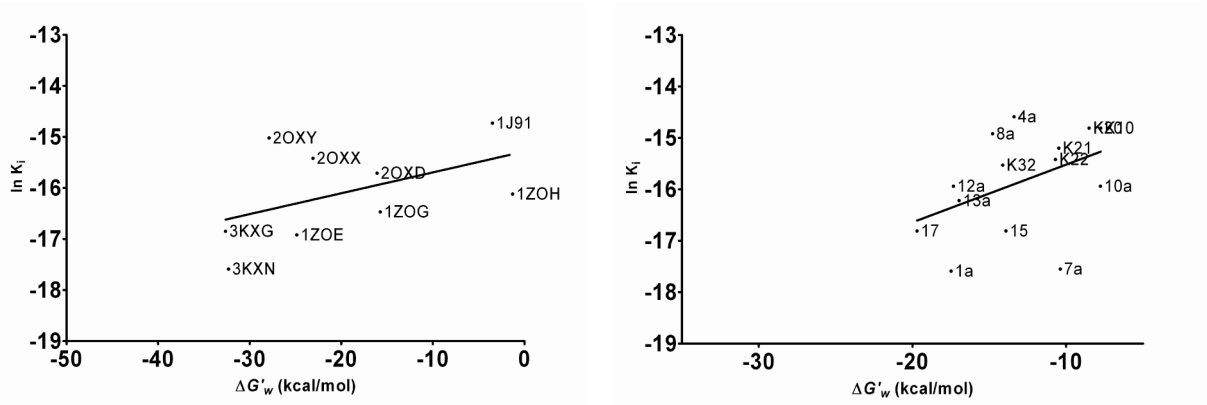
proto použita metoda PM6-DH2 rozšířena o korekci na halogenovou vazbu tj. metoda PM6-DH2X[50].

Skórovací funkce založená na metodě PM6-D2X byla aplikována na dvě skupiny halogenovaných inhibitorů s CK2, jak je uvedeno v publikaci (příloha 3). První skupina byla vytvořena podle krystalových struktur kukuřičné CK2 a druhá pomocí namodelovaných struktur inhibitorů s lidskou CK2.



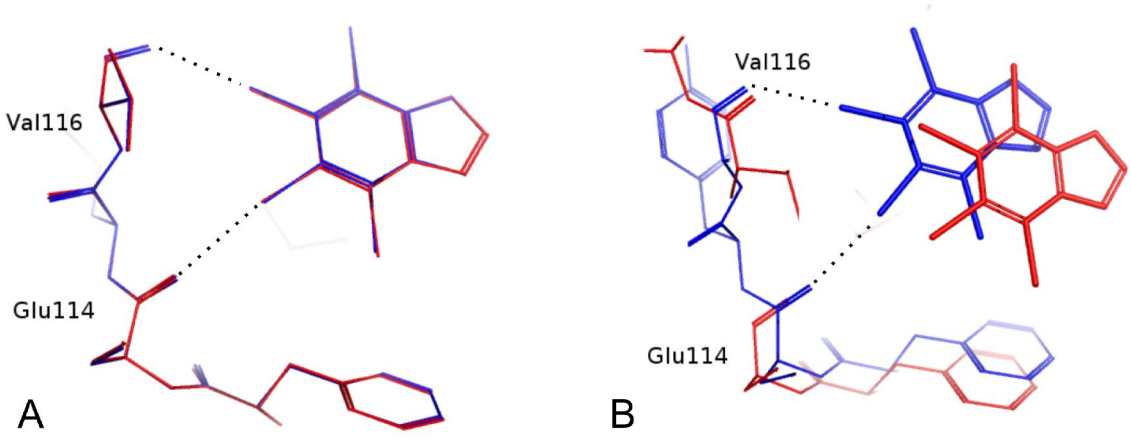
Obrázek 13 Korelace ΔH_w vs. $\mathbf{lnK_i}$ pro dvě skupiny CK2 v komplexu s inhibitory, vlevo skupina založená na krystalových strukturách a vpravo na strukturách s modifikacemi

Korelace mezi vypočítanou interakční entalpií ve vodě a 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) (Obr. 13). 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).



Obrazek 14 Korelace $\mathbf{\Delta G_w}$ vs. $\mathbf{lnK_i}$ pro dvě skupiny CK2 v komplexu s inhibitory, vlevo skupina založená na krystalových strukturách a vpravo na strukturách s modifikacemi

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) (Obr. 14), což ukazuje na nesprávný popis halogenové vazby pomocí silového pole. Všechny struktury byly optimalizované pomocí metody PM6-D2X v implicitním solventu popsaným COSMO modelem. Tyto struktury velmi dobře souhlasí s krystalovými strukturami. V případě uhlíků C_α je RMSD (angl. *root mean square deviation*) ~ 0,2 Å. Na druhé straně souhlas mezi optimalizovanými strukturami pomocí metody AMBER (ff99) a krystalovými strukturami byl znatelně horší (RMSD ~ 1 Å). Také u dalších strukturních vlastností charakterizujících halogenovou vazbu, tj. vzdálenosti mezi atomy halogenu inhibitoru a atomy kyslíků proteinu tvořících v krystalových strukturách první skupiny halogenovou vazbu, poskytla metoda PM6-D2X (na rozdíl od empirického potenciálu) velmi dobrou shodu s experimentem, což neplatí u empirického potenciálu AMBER. Nejmarkantněji je tento rozdíl patrný u komplexu 2OXY, u kterého došlo při optimalizaci empirickým potenciálem k posunu inhibitoru směrem ven z kavity (Obr. 15).



Obrázek 15 Struktury CK2 kavity s inhibitorem optimalizované pomocí PM6-D2X (nalevo) a AMBER (napravo), krystalové struktury jsou označeny modře a optimalizované struktury červeně

5 Závěr

1) Stabilizační energie komplexu CDK2 s roskovitinem je převáženě 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 pro tento a podobné komplexy doporučit. 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.

2) 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 určené pomocí semiempirické kvantověchemické metody PM6-DH2. Nejsilnější korelace s inhibiční konstantou bylo dosaženo pomocí interakční entalpie (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). Díky velmi dobré korelaci se jeví nadějné, že výpočty pomoci skórovací funkce založené na metodě SQM mohou být v budoucnu vzhledem k rychlému rozvoji v oblasti informačních technologií použitelné při návrhu léčiv.

3) 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 entalpii 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.

Literatura

1. Hunter, T., *Cell*, **1995**. 80: p. 225-236.

2. Johnson, L.N. and R.J. Lewis, *Chemical Reviews*, **2001**. 101(8): p. 2209-2242.

3. Manning, G., et al., *Science*, **2002**. 298: p. 1912-1934.

4. Lodish, *Molecular Cell Biology*. **2003**: 5th edition WH Freeman & Company..

5. Knockaert, M., P. Greengard, and L. Meijer, *Trends in Pharmacological Sciences*, **2002**. 23(9): p. 417-425.

6. Druker, B.J. and N.B. Lydon, *Journal of Clinical Investigation*, **2000**. 105(1): p. 3-8.

7. Bártová, I., et al., *Protein Science*, **2004**. 13(6): p. 1449-1457.

8. Echalier, A., J.A. Endicott, and M.E.M. Noble, *Biochimica et Biophysica Acta, Proteins and Proteomics*, **2010**. 1804(3): p. 511-519.

9. Morgan, D.O., *Annual Review of Cell and Developmental Biology*, **1997**. 13: p. 261-291.

10. Malumbres, M. and M. Barbacid, *Trends in Biochemical Sciences*, **2005**. 30(11): p. 630-641.

11. Johnson, L.N., M.E.M. Noble, and D.J. Owen, *Cell*, **1996**. 85(2): p. 149-158.

12. Malumbres, M. and M. Barbacid, *Nature Reviews Cancer*, **2009**. 9(3): p. 153-166.

13. Cohen, P., *Current Opinion in Chemical Biology*, **1999**. 3(4): p. 459-465.

14. Besson, A., S.F. Dowdy, and J.M. Roberts, *Developmental Cell*, **2008**. 14(2): p. 159-169.

15. Fischer, P.M. and D.P. Lane, *Current Medicinal Chemistry*, **2000**. 7: p. 1213-1245.

16. Meijer, L. and E. Raymond, *Accounts of Chemical Research*, **2003**. 36(6): p. 417-425.

17. Johnson, L.N., *Quarterly Reviews of Biophysics*, **2009**. 42(1): p. 1-40.

18. Tetsu, O. and F. McCormick, *Cancer Cell*, **2003**. 3(3): p. 233-245.

19. De Azevedo, W.F., et al., *European Journal of Biochemistry*, **1997**. 243(1-2): p. 518-526.

20. Gray, N.S., et al., *Science*, **1998**. 281(5376): p. 533.

21. Lawrie, A.M., et al., *Nature Structural Biology*, **1997**. 4(10): p. 796.

22. De Azevedo, W.F., et al., *Proceedings of the National Academy of Sciences of the United States of America*, **1996**. 93(7): p. 2735.

23. Havlíček, L., et al., *Journal of Medicinal Chemistry*, **1997**. 40(4): p. 408-412.

24. Azevedo, W.F., et al., *European Journal of Biochemistry*, **1997**. 243(1 2): p. 518-526.

25. Pinna, L.A. and J.E. Allende, *Cellular and Molecular Life Sciences.*, **2009**. 66(11): p. 1795-1799.

26. Filhol, O. and C. Cochet, *Cellular and Molecular Life Sciences*, **2009**. 66(11-12): p. 1830-9.

27. Ruzzene, M. and L.A. Pinna, *Biochimica et Biophysica Acta*, **2010**. 1804(3): p. 499-504.

28. Sarno, S. and L.A. Pinna, *Molecular BioSystems.*, **2008**. 4(9): p. 889-94.

29. Trembley, J.H., et al., *Cellular and Molecular Life Sciences.*, **2009**. 66(11-12): p. 1858-67.

30. Niefind, K., J. Raaf, and O.G. Issinger, *Cellular and Molecular Life Sciences*, **2009**. 66(11-12): p. 1800-16.

31. Pagano, M.A., et al., *Biochemical Journal*, **2008**. 415(3): p. 353-65.

32. Gianoncelli, A., et al., *Bioorganic & Medicinal Chemistry*, **2009**. 17(20): p. 7281-9.

33. Cozza, G., A. Bortolato, and S. Moro, *Medical Care Research and Review*, **2010**. 30(3): p. 419-462.

34. Battistutta, R., et al., *Chemistry & Biology*, **2005**. 12(11): p. 1211-9.

35. Wasik, R., et al., *Journal of Physical Chemistry B*, **2010**. 114(32): p. 10601-11.

36. Hobza, P. and K. Müller-Dethlefs, *Non-covalent interactions: theory and experiment*. **2009**: Royal Society of Chemistry.

37. Riley, K.E. and P. Hobza, *Wiley Interdisciplinary Reviews: Computational Molecular Science*. **2010**. 1(1): p. 3-17.

38. Arunan, E., et al., *Pure and Applied Chemistry*, **2010**

39. Riley, K.E. and P. Hobza, *Journal of Chemical Theory and Computation*, **2008**. 4(2): p. 232-242.

40. Hu, L.H., et al., *The Journal of Physical Chemistry A*, **2009**. 113(43): p. 11793-11800.

41. Leach, A.R., *Molecular modelling: principles and applications*. **2001**: Addison-Wesley Longman Ltd.

42. Case, D.A., et al., Journal of Computational Chemistry, **2005**. 26(16): p. 1668-88.

43. Levine, I.N., *Quantum chemistry*. Vol. 5. **2000**: Prentice Hall Upper Saddle River, NJ.

44. Helgaker, T., et al., *The Journal of Chemical Physics*, **1997**. 106(15): p. 6430.

45. Koch, W. and M.C. Holthausen, *A chemist's guide to density functional theory*. Vol. 293. **2001**: Wiley Online Library.

46. Elstner, M., et al., *Physical Review B*, **1998**. 58(11): p. 7260.

47. Elstner, M., et al., *The Journal of Chemical Physics*, **2001**. 114: p. 5149.

48. Řezáč, J., et al., *Journal of Chemical Theory and Computation*, **2009**. 5(7): p. 1749-1760.

49. Korth, M., et al., *Journal of Chemical Theory and Computation*, **2010**. 6(1): p. 344-352.

50. Řezáč, J. and P. Hobza, *Chemical Physics Letters*, **2011**. 506(4-6): p. 286-289.

51. Stryer, L., *Biochemistry*. **1995**: WH Freeman & Company.

52. Fanfrlík, J., et al., *Journal of Physical Chemistry B*, **2010**. 114(39): p. 12666-78.

53. Zhou, T., D. Huang, and A. Caflisch, *Journal of Medicinal Chemistry*, **2008**. 51(14): p. 4280-4288.

54. Raha, K. and K.M. Merz, *Journal of the American Chemical Society*, **2004**. 126(4): p. 1020-1021.

55. Lu, Y.X., et al., *Journal of Medicinal Chemistry*, **2009**. 52(9): p. 2854-2862.

56. Stewart, J.J.P., *Journal of Molecular Modeling*, **2009**. 15(7): p. 765-805.

57. Stewart, J.J.P., *Journal of Molecular Modeling*, **2007**. 13(12): p. 1173-1213.

58. Frisch, M.J., et al., *Inc.: Wallingford, CT*, **2009**.

59. Jurečka, P., et al., *Physical Chemistry Chemical Physics*, **2006**. 8(17): p. 1985-1993.

60. Riley, K.E., et al., *Chemical Reviews*, **2010**. 110(9): p. 5023-5063.

61. Grimme, S., *Journal of Computational Chemistry*, **2004**. 25(12): p. 1463-1473.

62. Jurečka, P., et al., *Journal of Computational Chemistry*, **2007**. 28(2): p. 555-569.

63. Černý, J., et al., *The Journal of Physical Chemistry A*, **2007**. 111(6): p. 1146-1154.

64. Otyepka, M., et al., *The Journal of Physical Chemistry B*, **2006**. 110(9): p. 4423-4429.

65. Řezáč, J., et al., *Journal of Chemical Theory and Computation*, **2009**. 5(7): p. 1749-1760.

66. Chang, C.A., W. Chen, and M.K. Gilson, *Proceedings of the National Academy of Sciences*, **2007**. 104(5): p. 1534.

67. Sato, H., L.M. Shewchuk, and J. Tang, *Journal of Chemical Information and Modeling*, **2006**. 46(6): p. 2552-2562.

68. Guimaraes, C.R.W. and M. Cardozo, *Journal of Chemical Information and Modeling*, **2008**. 48(5): p. 958-970.

69. Ferrara, P., et al., *Journal of Chemical Information and Modeling*, **2006**. 46(1): p. 254-263.

70. Auffinger, P., et al., *Proceedings of the National Academy of Sciences U S A*, **2004**. 101(48): p. 16789-94.

71. Hernandes, M.Z., et al., *Current Drug Targets*, **2010**. 11(3): p. 303-14.

Publikace autora nezahrnuté v této práci

Dobeš P, Kmuníček J, Mikeš V, Damborský J.: Binding of fatty acids to beta-cryptogein: quantitative structure-activity relationships and design of selective protein mutants. *Journal of Chemical Information and Modeling* **2004,** 44(6):2126-32.

Pohanka M, **Dobeš P**, Drtinová L, Kuča K.: Nerve Agents Assay Using Cholinesterase Based Biosensor. *Electroanalysis* **2009,** 21, 1177-1182

Trbušek M, Šmardová J, Malčíková J, Šebejová L, **Dobeš P**, Svitaková M, Vranová V , Mráz M, Skuhrová Francová H, Doubek M, Brychtová Y, Kuglík P, Pospšilová S, Mayer J,: Missense Mutations Located in Structural p53 DNA-binding Motifs are Associated with Extremely Poor Survival in Chronic Lymphocytic Leukemia. *Journal of Clinical Oncology* in press

Klusák V, **Dobeš** P, Černý J, Vondrášek J,: How to fragment a polypeptide? Ab Initio Computational Study of Pair Interactions between Amino Acids and Ligand -Amino Acid in Proteins. *Collection of Czechoslovak Chemical Communication.* under revision

Přílohy

Příloha 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. *Chemistry - A European Journal.* **2006**;12:4297-304.

Příloha 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. *Journal of Computer-Aided Molecular Design.* **2011**;25:223-35.

Příloha 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. *Journal of Physical Chemistry B*, under revision

DOI: 10.1002/chem.200501269

Interaction Energies for the Purine Inhibitor Roscovitine with Cyclin-Dependent Kinase 2: Correlated Ab Initio Quantum-Chemical, DFT and Empirical Calculations

Petr Dobeš,[a] Michal Otyepka,[b] Miroslav Strnad,[c] and Pavel Hobza*[a, b]

Abstract: The interaction between roscovitine and cyclin-dependent kinase 2 (cdk2) was investigated by performing correlated ab initio quantum-chemical calculations. The whole protein was fragmented into smaller systems consisting of one or a few amino acids, and the interaction energies of these fragments with roscovitine were determined by using the MP2 method with the extended aug-cc-pVDZ basis set. For selected complexes, the complete basis set limit MP2 interaction energies, as well as the coupled-cluster corrections with inclusion of single, double and noninteractive triples contributions [CCSD(T)], were also evaluated. The energies of interaction between roscovitine and small fragments and between roscovitine and substantial sections of protein (722 atoms) were also computed by using density-functional tight-binding methods covering dispersion energy (DFTB-D) and the Cornell empirical potential. Total stabilisation energy originates predominantly from dispersion energy and methods that do not account for the dispersion energy cannot, therefore, be recommended for the study of protein–inhibitor interactions. The Cornell empirical potential describes reasonably well the interaction between roscovitine and protein; therefore, this method can be applied in future thermodynamic calculations. A limited number of amino acid residues contribute significantly to the binding of roscovitine and cdk2, whereas a rather large number of amino acids make a negligible contribution.

Keywords: ab initio calculations · cyclin-dependent kinase · density functional calculations · molecular mechanics · roscovitine

Introduction

The transfer of information in a cell is mediated through various linked signalling pathways, leading finally to the control of diverse metabolic processes. Protein kinases play a crucial role in many of these pathways, for example, regulation of the cell cycle, differentiation, membrane transport and secretion of cellular proteins, such as growth hormones.[1,2] A deregulation of cyclin-dependent kinases (cdk) was demonstrated in human primary tumors and human tumor lines.[3] This discovery stimulated interest in cdk and their inhibitors because of their potential applications as anticancer drugs.[4] These cdk enzymes catalyse the transfer of the ATP phosphate to serine and threonine residues located on a protein substrate and, thus, activate other proteins. The activity of cdk2 is regulated by association with regulatory subunits (cyclins) and through phosphorylation by other kinases, which cause conformational changes in the protein so that the correct positioning of a substrate and catalytic residues is achieved.[5,2] The cdk2 associated with cyclin E can promote progress through the G1 phase and, consequently, it forms a complex with cyclin A, which is a necessary condition for entry of a cell into the S phase.

Cdk2 is comprised of 298 amino acid residues and possesses a typical kinase fold containing two lobes: a small N-terminal and a large C-terminal domain. The lobes are connected through a single polypeptide strand. The active site

[a] P. Dobeš, Prof. P. Hobza
Institute of Organic Chemistry and Biochemistry
Academy of Sciences of the Czech Republic and Center for Biomolecules and Complex Molecular Systems
166 10 Prague 6 (Czech Republic)
Fax: (+420) 220-410-320
E-mail: pavel.hobza@uochb.cas.cz

[b] Dr. M. Otyepka, Prof. P. Hobza
Department of Physical Chemistry and Center for Biomolecules and Complex Molecular Systems
Palacký University, tr. Svobody 26
771 46 Olomouc (Czech Republic)

[c] Prof. M. Strnad
Laboratory of Growth Regulators
Palacký University and Institute of Experimental Botany
Academy of Sciences of the Czech Republic
Šlechtitelu 11, 783 71 Olomouc (Czech Republic)

that binds ATP or purine-like inhibitors is situated between the lobes in a deep cleft, and serves as a primary target for drug design. It consists of three sections named according to the ATP moieties that it binds: the purine binding site, the phosphate-group pocket and the sugar pocket. The X-ray structures of cdk2 and its complexes with different inhibitors have been published[6–10] and the number of these continues to rise.

The inhibition of cdk2 plays a key role in drug development programmes addressing such different pathological conditions as inflammation, autoimmunity, cancer, and cardiovascular and neurodegenerative diseases.[4,11] Consequently, many pharmaceutical companies are interested in drugs that regulate the activity of specific eukaryotic protein kinases. Both experimental and theoretical approaches are used for this purpose. Various computational studies describing the thermodynamic properties of binding protein with inhibitor by using molecular mechanics-generalised Born solvation area (MM-GBSA) methodology,[12,13] the molecular simulation of this process[14] and the docking process exist.[15] All these studies were based on empirical potentials, which is understandable considering the size of the system under investigation. The applicability of empirical potential should, however, be carefully tested; our recent studies have shown its limitations in the context of interaction with DNA bases,[16] the hydrophobic core of proteins[17] and various types of interaction in proteins.[18]

The aim of the present study is to investigate the interaction of cdk2 with the inhibitor roscovitine. Our interest in roscovitine is due to its potential as a promising drug candidate in cancer therapy through blocking cell cycle transitions in cancerous cells by cdk1/2 inhibition. Interaction of cdk2 with roscovitine is very complex and is controlled by not only hydrogen bonding, but also by other energetic contributions, such as dispersion, electrostatic and charge-transfer. An accurate description of all these energetic contributions is difficult and requires high-level ab initio quantum-chemical correlated calculations. These calculations should be also applied in the design of a selective inhibitor with higher activity. An essential condition for such designs is a full understanding of the nature of the stabilisation of the ligand in the active site. Investigation will be focused on not only the physical nature of stabilisation (i.e., the role of various energy contributions, such as hydrogen bonding, stacking, etc.), but also on the steric nature (i.e., the role of various parts of the protein cavity in stabilising inhibitor). Thus, the main difference between previous studies and this present one concerns the non-empirical ab initio quantum-chemical approach, which properly describes the nature of the interaction between roscovitine and protein. We stress again that previous results were based on empirical potential and, therefore, might not be sufficiently accurate.

Strategy of calculations: Cdk2 interacts with roscovitine non-covalently. Of the non-covalent interactions, the most important role is played by hydrogen-bonding, electrostatic, charge-transfer and dispersion interactions. The role of hydrogen bonds is well recognised and their theoretical description is quite straightforward. In fact, almost any empirical, semiempirical and nonempirical methods can describe hydrogen bonding. Much less is known about the role of dispersion energy and originally it was thought to be of low importance. Only recently it was shown that the dispersion energy plays an important role in not only stabilising structures of DNA and proteins, but also in, for example, stabilising an intercalator in DNA. As well as dispersion interaction, charge-transfer also plays an important role in the interaction of protein and inhibitor. A theoretical description of this energy term is also difficult and empirical potential does not incorporate this term at all. Thus, it is necessary to evaluate the interaction of cdk2 with roscovitine at a high theoretical level by embracing all energy terms. Such calculations are impractical for large proteins consisting of several thousands of atoms. For this reason, the whole protein was fragmented into smaller systems consisting of one or a few amino acids. The high-level ab initio correlated calculations performed for these smaller complexes are used as benchmark data for testing the performance of the empirical potential. It is evident that to perform the molecular dynamics simulations of the present system (necessary for the description of dynamic properties) the empirical potential must be used.[19] Besides this, the calculated stabilisation energies allow us to estimate the importance of various sections of the protein cavity in the stabilisation of inhibitor. Here, the accuracy of the calculations is crucial, as stabilisation can originate from hydrogen bonding as well as from stacking, and the balanced description of various energy terms is extremely difficult.

Calculations of cdk2 with roscovitine were performed as follows. Firstly, empirical potentials that are used standardly for protein simulation were applied. This methodology is successful, but is associated with several problems. The potential is pairwise additive and the non-additivity plays a role if systems are polar or charged. Further, interaction energy is comprised of a sum of electrostatic, dispersion and repulsion contributions, which means that the charge-transfer term is completely missing. Thus, secondly, the fast approximative ab initio method that is free of the problems mentioned above is applied. In our laboratory we use the self-consistent charge density-functional tight-binding method augmented empirically for London dispersion energy (SCC-DFTB-D).[20] This method yields excellent results for DNA base pairs and DNA fragments, complexes of intercalator with DNA, as well as for amino acid pairs and protein fragments. The method is very fast and enables single-point calculations to be performed for systems containing several thousands of atoms, and for smaller systems, molecular dynamics simulations can be performed. Finally, the benchmark data for smaller model systems generated by accurate ab initio correlated calculations were used for testing empirical potential and SCC-DFT-D results.

Here, we focus on the interaction energy of the protein–ligand complex with the aim of understanding the nature of its stabilisation and quantification of the role of dispersion

contribution. Interaction energy is determined as a sum of interaction energies of representative fragments containing amino acid(s) and ligand, as well as directly for the whole protein–ligand complex.

Geometries, fragmentation: The crystal structure of protein cyclin-dependent kinase 2 with roscovitine (Figure 1) with 2.4 Å resolution was used.[9] Figure 2 shows a schematic

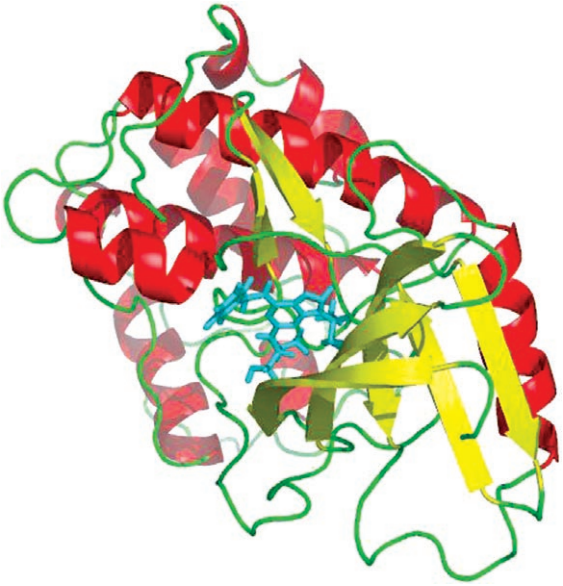


Figure 1. Secondary structure of cdk2 with roscovitine.

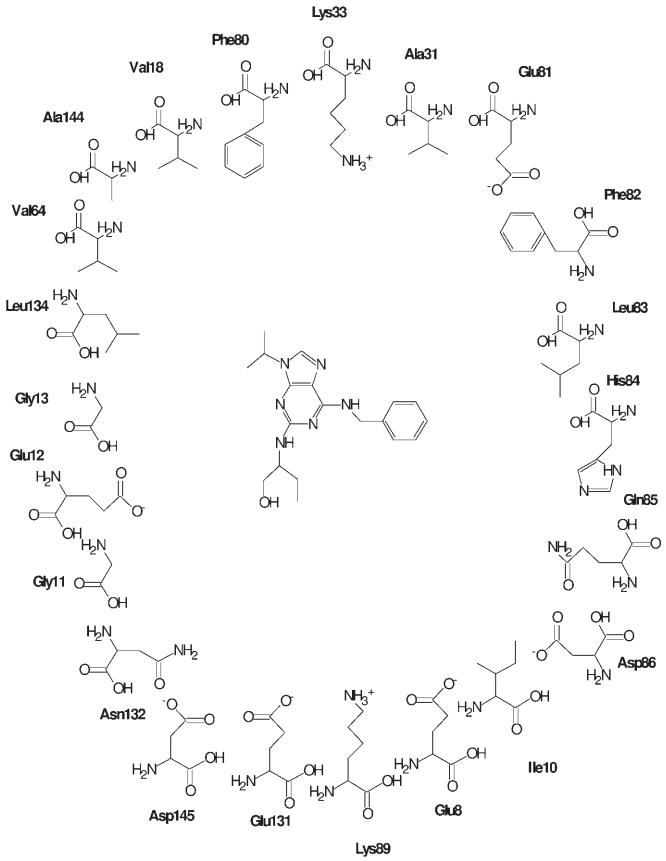


Figure 2. Schematic view of roscovitine (centre) and the key amino acids.

view of roscovitine and some of the key amino acids under consideration. The resolution of the crystal structure of roscovitine with cdk2 is rather high, but it is sufficient for the present computations. The system studied contains protein and ligand and is very complex. To understand the nature of stabilisation, high-level correlated ab initio calculations should be performed. These calculations are, however, impractical for the system considered, whose size should be reduced by, for example, fragmenting the protein into constituent amino acids. The method of fragmentation is not standard and several problems regarding biochemical relevancy should be overcome. In constructing representative fragments, we considered only complexes of specific amino acids directed towards roscovitine, as it is known that only the nearest groups have significant influence on the interaction energy. The amino acid charges were adjusted to pH 7, as the crystal structure considered was analysed under these conditions. A problem occurred for the interactions of roscovitine and the protein backbone, due to strong π–π interaction between peptide bonds and the purine aromatic rings of roscovitine. The protein backbone was, therefore, cut at the C^{α}–N bond and the peptide bond was maintained. During partitioning we considered only amino acids and crystal water molecules located within 5 Å from roscovitine. We created 14 fragmented complexes containing one or more amino acids and roscovitine: glutamic acid 8 (E8); isoleucine 10 (I10); valine 18 (V18); alanine 31 (A31); lysine 33 and aspartic acid 145 (K33D145); valine 64 (V64); phenylalanine 80 (F80); glutamic acid 81, phenylalanine 82 and leucine 83 (E81F82L83); leucine 83, histidine 84 and glutamine 85 (L83H84Q85); glutamine 85, aspartic acid 86 and lysine 86 (Q85D86K89); glutamine 131 and asparagine 132 (Q131N132); leucine 134 (L134); alanine 144 (A144); glycine 11, glutamic acid 12 and glycine 13 (G11E12G13). All complexes are shown in Figure 3 and are labelled according to the amino acid(s) included. Missing hydrogen atoms in these fragmented complexes were added by using InsightII and optimised by the gradient quantum-chemical optimisation at the B3LYP/6–31G** level; positions of heavy atoms were fixed. The molecules of water presented in the crystal structure were also included to improve the description of interaction between roscovitine and a protein fragment. Although all amino acid fragments are neutral, E8 bears a negative (−1) charge, and the substituents of the K33D145 fragment have +1 and −1 charges, respectively. The same is true for Q85D86K89, in which the D86 and K89 substituents have a −1 and +1 charge, respectively.

Because the partitioning of protein is associated with several uncertainties, it is recommended to calculate the interaction energy for as large a portion of the protein as possible. The model system consisting of 722 atoms was obtained

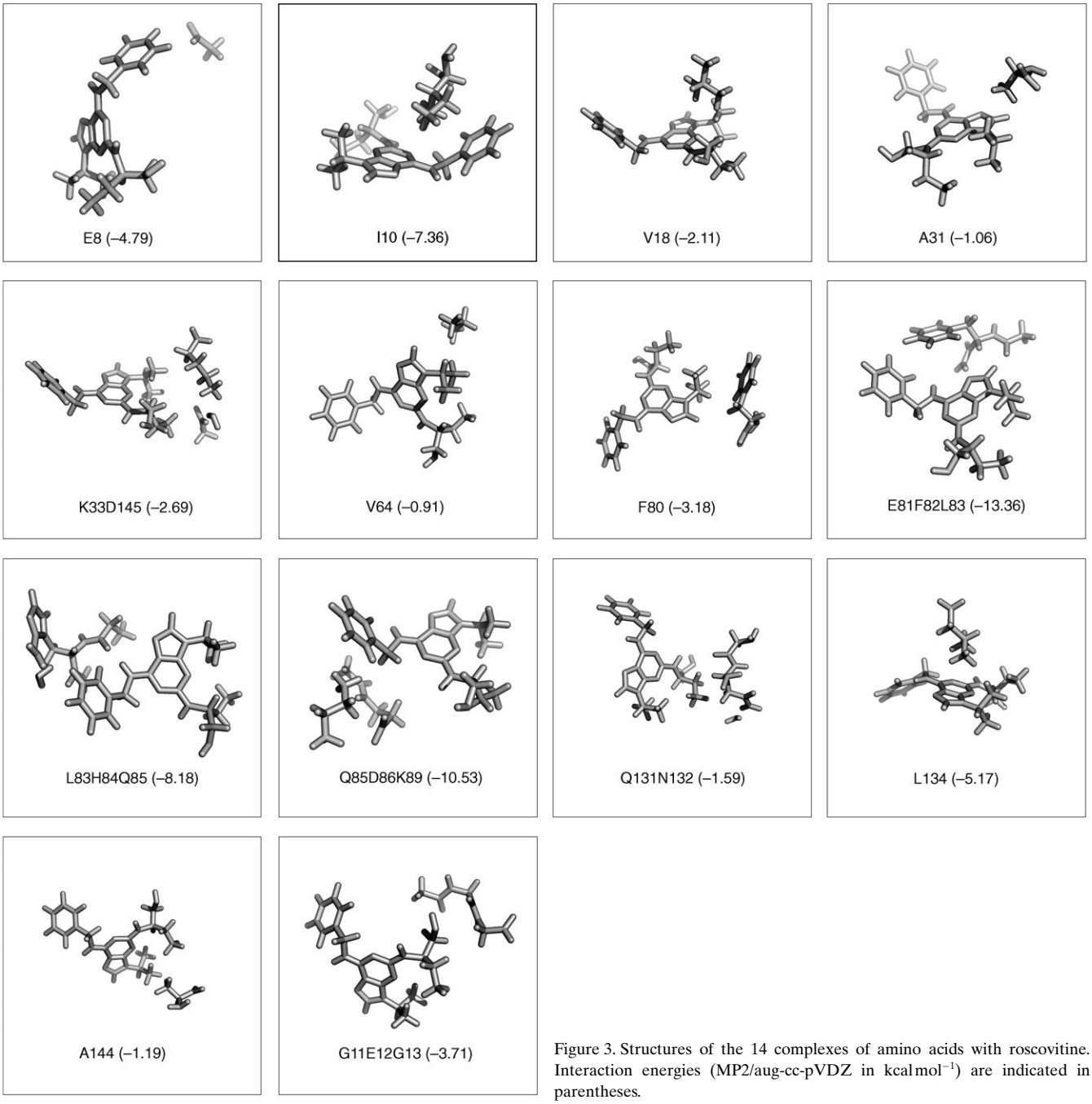


Figure 3. Structures of the 14 complexes of amino acids with roscovitine. Interaction energies (MP2/aug-cc-pVDZ in kcal mol^{-1}) are indicated in parentheses.

by considering all amino acids within a distance of 8 Å from roscovitine. The complex with roscovitine is shown in Figure 4. The positions of missing hydrogen atoms were optimised by using the semiempirical quantum-chemical PM3 method.

Computational Methods

The interaction energies between ligand and fragments of protein (or whole protein) include an important contribution from London dispersion energy. It is, therefore, appropriate to consider only those methods that cover the dispersion energy. We have shown recently[21] that accurate interaction energies of various types of intermolecular complexes (hydrogen-bonded, stacked, T-shaped) are obtained at the coupled-cluster singles and doubles theory with perturbational triples corrections [CCSD(T)] level by using the complete basis set (CBS) limit.

CBS interaction energies are obtained by extrapolating the total energies of supersystem and subsystems.[22] The first rational basis set is the aug-cc-pVDZ one and extrapolations are, thus, performed from the aug-cc-pVDZ and aug-cc-pVTZ energies. The coupled-cluster calculations [CCSD(T)] at the CBS limit (ΔE[CCSD(T)/CBS] for the present complexes is clearly impractical, therefore, we approximated their values by using Equation (1)

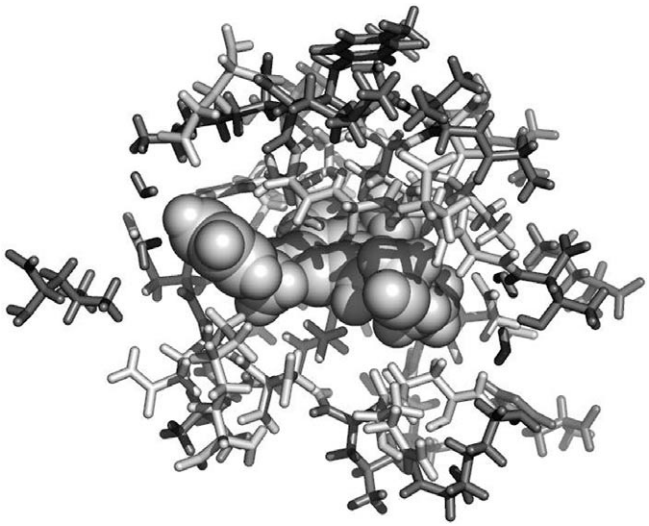


Figure 4. Structure of the protein cavity of cdk2 containing roscovitine. The roscovitine is represented in a ball-and-stick conformation.

$$\Delta E[\mathrm{CCSD(T)/CBS}] = \Delta E(\mathrm{MP2})/\mathrm{CBS} + \Delta E[\mathrm{CCSD(T)-MP2}]/(\mathrm{DZ+P}) \quad (1)$$

in which ΔE(MP2)/CBS stands for the MP2 CBS stabilisation energy and the latter term gives the difference between the CCSD(T) and MP2 stabilisation energies determined by using a medium basis set of the DZ+P quality. The evaluation of the ΔE-[CCSD(T)/CBS] limit by using this equation is based on the assumption that the difference between the CCSD(T) and MP2 interaction energies is less dependent on the quality of the basis set than the CCSD(T) and MP2 energies.[22]

The CCSD(T) and MP2 CBS calculations described are extremely time consuming (e.g., computations of fragment I10 with roscovitine at the MP2 and CCSD(T) levels of theory take about two weeks of CPU time by using a Pentium 4 3 GHz) and were, therefore, performed for only selected complexes. In all remaining cases, the interaction energies were determined at the MP2/aug-cc-pVDZ level. Interaction energies were systematically corrected for the basis set superposition error by using the function counterpoise method of Boys and Bernardi.[23] MP2 calculations with the present basis set would be very time consuming, so we replaced them by resolution of identity (RI)-MP2 calculations. It was shown recently[24] that absolute and relative RI-MP2 and MP2 energies differ marginally, although the former method is by one order of magnitude faster.

Density functional theory (DFT) methods are very popular as they are efficient and enable the study of extended complexes. However, the DFT methods fail to describe the dispersion energy and, thus, their use in biomolecular studies is limited. Here, we demonstrate this by performing DFT/B3LYP/6–31G** calculations. Furthermore, the approximative SCC-DFTB-D method with empirical dispersion term was used.

We used the Cornell empirical potential[19] by using the parm99 parameter set. For the 14 fragmented complexes the RESP HF/6–31G* charges were adopted, whereas for the large model cluster (consisting of 722 atoms) the standard atomic charges from the Cornell library were used.

Results and Discussion

Table 1 shows the interaction energies of the 14 complexes determined by using various theoretical procedures. The ab initio nonempirical results will be discussed first. Extrapolation to the MP2 CBS limit was performed for complexes I10, F80 and L134 and passing from aug-cc-pVDZ to aug-cc-pVTZ yields a non-negligible enlargement of stabilisation energy. The CBS limit stabilisation energies are 10, 9 and 9% larger than the respective aug-cc-pVDZ values. The CCSD(T) correction term was determined for slightly reduced complexes (F80 and L134), giving 0.41 and 0.70 kcal mol^{-1}, respectively, that is, it was slightly repulsive. From the values mentioned above it is clear that we can confidently use the RI-MP2/aug-cc-pVDZ stabilisation energies. Therefore, the remaining discussion will be based solely on these.

It is evident from the data in the first column of the Table 1 that by far the largest stabilisation energy was found for the E81F82L83 cluster. By investigating the structure of the complex (Figure 3) we found two short hydrogen-bonded contacts (2.35 Å): between the amino-group hydrogen of L83 and the N7 nitrogen of the purine ring of roscovitine, and between the carbonyl-group oxygen of the E81 peptide bond and the C8–H hydrogen of the purine ring of roscovitine (2.14 Å). Both contacts are undoubtedly connected by strong hydrogen bonds. Important stabilisation also comes from π–π interactions between the phenyl ring of roscovitine and the phenyl ring of F82. The stabilisation energy of the Q85D86K89 cluster is also large (>10 kcal mol^{-1}) and arises from the interaction between charged subunits D86 ($q=-1$) and K89 ($q=+1$) and neutral roscovitine. Figure 3 shows the existence of the close contact between the NH_3^+ group of K89 and the phenyl-ring group of roscovitine. The cluster L83H84Q85 is characterised by a moderately strong stabilisation energy (8.18 kcal mol^{-1}) that stems from a very short hydrogen bond (1.83 Å)

Table 1. Interaction energies [kcal mol^{-1}] for the 14 fragmented complexes (see Figure 3), determined at the MP2, B3LYP, DFTB, DFTB-D and Amber levels.

	Interaction energy				
Complex	MP2/aug-cc-pVDZ	B3LYP/6–31G**	DFTB	DFTB-D	Amber
E8	−4.79	−3.86	−2.03	−2.47	−1.88
I10	−7.36 (-7.76, −8.08)[a]	1.44	−1.09	−7.57	−7.14
V18	−2.11	2.1	0.18	−2.91	−2.8
A31	−1.06	2.9	0.37	−1.84	−1.6
K33D145	−2.69	1.94	0.14	−3.18	−1.82
V64	−0.91	0.41	0.03	−1.09	−1.02
F80	−3.18 (−3.35, −3.47, 0.41)[b]	0.9	0.07	−3.14	−3.19
E81F82L83	−13.36	−6.12	−4.39	−9.78	−9.04
L83H84Q85	−8.18	0.98	−2.53	−8.57	−8.54
Q85D86K89	−10.53	−5.02	−2.56	−12.02	−7.98
Q131N132	−1.59	1.69	0.67	−2.13	−0.98
L134	−5.17 (−5.42, −5.64, 0.7)[b]	2.68	−0.19	−5.28	−5.52
A144	−1.19	0.96	−0.05	−2.06	−1.6
G11E12G13	−3.71	−1.53	−1.25	−3.35	−4.26
sum	−65.83	−0.53	−12.63	−65.39	−57.36

[a] Numbers in parentheses correspond to MP2/aug-cc-pVTZ and MP2 CBS interaction energies, respectively. [b] Numbers in parentheses correspond to MP2/aug-cc-pVTZ and MP2 CBS interaction energies, and to the CCSD(T) correction term, respectively.

between the oxygen of the carboxyl group of L83 and the hydrogen of the N^6-amino group of roscovitine. Further stabilisation also originates in the π–π interactions between two peptide bonds of L83, H84 and Q85 and the phenyl ring of roscovitine. Stabilisation energies of the three complexes I10, L134 and F80 are 7.36, 5.17 and 3.18 kcal mol^{-1}, respectively, and are due partially to strong C–H–π interactions. In the case of I10, dispersion interaction between the side chain of I10 and the phenyl and purine aromatic rings of roscovitine also contribute to the stability of the complex. The E8 and G11E12G13 complexes have moderately high stabilisation energies (4.79 and 3.71 kcal mol^{-1}, respectively). Stabilisation of the anionic complex E8 is due to interaction between the aromatic phenyl ring of roscovitine and the negatively charged carboxylic group of glutamic acid 8. K33D145, V18, Q131N132 and A31 complexes possesses only weak stabilisation energies (2.69, 2.11, 1.59 and 1.06 kcal mol^{-1}, respectively). The remaining complex (V64) exhibits only negligible stabilisation, less than 1 kcal mol^{-1}.

The total stabilisation energy of roscovitine and the 14 neighbouring fragments is rather larger (~66 kcal mol^{-1}), however, the contributions of single amino acid–roscovitine interactions differ considerably. Five complexes (E81F82L83, Q85D86K89, L83H84Q84, I10 and L134) contribute greatly (about 68%) to total stabilisation. On the other hand, six fragments (V18, A31, K33D145, V64, Q131N132 and A144) contribute less than 14% to total stabilisation (the remaining contribution to stabilisation is from complexes E8, F80 and G11E12G13). This observation is slightly surprising and clearly indicates that selected sections of the protein cavity are much more significant for inhibition than others. Similarly, some regions of the cavity have almost no effect on inhibitor stabilisation. We can speculate about the importance of these findings in the light of the preparation of new (more active) inhibitors. Probably the most efficient way to increase stabilisation would be to increase the binding activity of those amino acid residues that contribute negligibly to the overall stability. Here, we investigated the interaction of cdk2 with inhibitor. In the future, we will study the interaction with modified roscovitine inhibitors. The mutations should be reflected in not only different interaction energies, but also in altered binding affinities modelled by the change in free energy of complexation. The resulting theoretical values could be then compared with existing experimental data, such as SAR values.

The segmentation of protein into amino acid fragments is justified by evaluating the interaction energy for the larger cluster. Figure 5 shows three partial clusters E81F82L83, L83H84Q85 and Q85D86 as well as the composed system E81F82L83H84Q85D86. To make the MP2/aug-cc-pVDZ calculation feasible we removed the side chains from all amino acids, as well as from the N^6 and N^2 roscovitine substituents. From Table 2 we can see that the sum of the interaction energies of the three partial clusters amounts to −20.51 kcal mol^{-1}, whereas the interaction energy of the whole system is −19.28 kcal mol^{-1}. The small difference between these values fully justifies the fragmentation of the protein.

Table 2. Interaction energies [kcal mol^{-1}] for three separate fragmented clusters and the composed system (see Figure 5), determined at the MP2 level.

Complex	MP2/aug-cc-pVDZ
E81F82L83	−11.01
L83H84Q85	−7.25
Q85D86	−2.26
sum of the three fragments	−20.51
E81F82L83H84Q85D86	−19.28

The MP2/aug-cc-pVDZ calculations for the present clusters are demanding and raise the question of whether a simpler quantum-chemical procedure can be applied. This concerns firstly the DFT methods, which are popular within the biochemical community. From the data shown in Table 1, however, we see that the B3LYP/6–31G** values are dramatically different from the correlated MP2/aug-cc-PVDZ data: the sum of all 14 MP2 interaction energies is −65.8 kcal mol^{-1}, whereas the sum of the B3LYP interaction energies is −0.53 kcal mol^{-1}. The huge difference of 65 kcal mol^{-1} is due to the lack of dispersion energy in the B3LYP treatment. Evidently, the B3LYP procedure fails completely and cannot be used for the study of predominantly noncovalent protein–inhibitor interactions. DFT/B3LYP calculations were performed with the 6–31G** basis

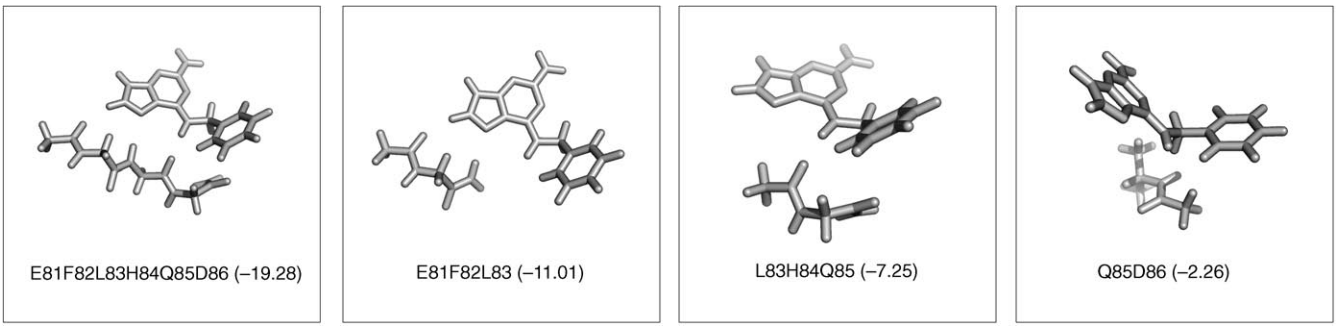


Figure 5. Structures of fragments of the protein backbone. Interaction energies (MP2/aug-cc-pVDZ in kcal mol^{-1}) are indicated in parentheses.

set. The DFT calculations are known to be less sensitive than the MP2 method to the quality of the basis set. We have shown recently[25] that the B3LYP procedure yields very similar stabilisation energies for amino acid pairs, for which the 6–31G** and cc-pVTZ basis set were applied.

Comparison of B3LYP and MP2 results can indicate the role of π–π interactions, which are governed by London dispersion energy. From the data in Table 1 it is evident that practically all of the complexes investigated are stabilised mainly by dispersion energy. The anionic complex E8 represents other example. As expected, the electrostatic contribution (which is properly described in the B3LYP procedure) is dominant and, consequently, the B3LYP and MP2 values differ by less than 1 kcal mol^{-1}. A similar situation occurs in the G11E12G13 cluster, for which the B3LYP and MP2 values basically agree. In all other cases, the differences between these energies were large. The largest difference was found for the L83H84Q84, L134, I10 and E81F82L83 clusters. In each of these clusters we can find an important π–π stacking motif between either a peptide bond and an aromatic ring, an aromatic ring and an aromatic ring or even an aliphatic chain and an aromatic ring.

The surprising role of dispersion energy, demonstrated in the previous paragraph by comparison of MP2 and B3LYP stabilisation energies, can be confirmed by investigating the stabilisation energies originating from the DFTB and DFTB-D procedures. These values differ by the empirical London dispersion energy alone: the sum of all 14 stabilisation energies is 12.6 and 65.4 kcal mol^{-1}, respectively. The latter value is practically identical with the MP2 value, whereas the former value is strongly underestimated and is close to the B3LYP value. The same is true for the comparison of stabilisation energies for single complexes. As for DNA base pairs, DNA base pair–intercalator, and amino acid pairs, in the present case the DFTB-D procedure mimics surprisingly well MP2 interaction energies, whereas B3LYP (and all other standardly used functionals) fails.[25]

After joining all 14 fragments into one "shell", the DFTB and DFTB-D stabilisation energies of the "shell" with the roscovitine amount to 8.8 and 56 kcal mol^{-1}, respectively. Evidently, amino acids in the shell are slightly repulsive and again, dispersion energy is dominant. By extending the model from the "shell" (consisting of 338 atoms) to a larger section of protein (consisting of 722 atoms, see Figure 4) we obtained a slight increase in stabilisation energy (from 56 to 61 kcal mol^{-1}, an increase of about 10%). This data indicates that "second shell" amino acids contribute negligibly to the stabilisation of roscovitine.

Let us finally investigate the performance of the Cornell empirical potential. The sum of the 14 stabilisation energies (57.4 kcal mol^{-1}) is close to the MP2 value. We also found good agreement in the stabilisation energies for single clusters. Evidently, the Cornell empirical potential[19] is well suited for the present type of protein–ligand complexes and its use can be recommended. Upon extending the model, we obtain (as in the previous case) only a slight increase in stabilisation energy (from 57 to 62 kcal mol^{-1}, an increase of about 5%). The Cornell empirical potential is also feasible for molecular mechanics-Poisson–Boltzmann (generalised Born) solvation area (MM-PB(GB)SA) analysis of large protein–ligand or protein–protein complexes. Finally, the calculated values of interaction energies agree well with the interaction energies averaged over molecular dynamics (MD) simulations.[14]

Conclusions

1) The majority of stabilisation energy between roscovitine and protein originates from dispersion energy.
2) Due to the lack of dispersion energy, DFT methods fail to describe the roscovitine–protein interactions and their use cannot be recommended for inhibitor–protein studies. If the DFT energy is augmented empirically by London dispersion energy, reliable stabilisation energies can be achieved.
3) The Cornell empirical potential describes reasonable well the interaction between roscovitine and protein. This supports the use of this potential for future free-energy calculations.
4) A limited number of amino acid residues contribute significantly to the binding of roscovitine to cdk2 and, conversely, a rather large number of amino acids contribute negligibly. Mutation of the former, as well as the latter residues, can, thus, have a dramatic influence on the binding of roscovitine to the cavity and, consequently, also on the biological activity of roscovitine.
5) Besides providing alterations to an inhibitor, mutation of roscovitine will play a key role in the design of potential drugs. In subsequent work we will study both the changes in protein and the mutation of roscovitine. This will not be limited to interaction energy only. Although we believe (and have evidence) that the interaction energy contributes dominantly to a change of free energy of complexation, we will determine explicitly, besides interaction energies, the change of free energy of complexation of cdk2 with the ligand roscovitine.
6) The E81F82L83 (EFL) fragment with a dominant contribution to the interaction energy is conserved for cdk2, cdk1 and cdk3, which are highly homologous and have similar affinities to roscovitine. Cdk9 and cdk5 are both sensitive to roscovitine and have DFC and EFC motifs, respectively, instead of the EFL motif of cdk2. Cdk4 and cdk6, which are less sensitive to roscovitine, possess the EHV motif, in which phenylalanine is mutated to histidine. This finding reveals the important role of π–π interactions of the phenylalanine ring in roscovitine selectivity. In all cases, a change in the structure of roscovitine at the point of contact with these residues influences the strength of binding to cdk2. This fact is well documented by a reduction in the interaction of cdk with C8-substituted roscovitine analogues, due to disruption of the hydrogen bond between C8–H and the E81 carbonyl group.[15,26,27]

Acknowledgements

This work was supported by grant nos. 2003/05/0009 (PH) and 301/05/0418 (MS) from the Grant Agency of the Czech Republic and LC512 (PH) and MSM6198959216 (MO) from the MSMT of the Czech Republic. It was also part of research project No. Z4 0550506 (PH).

[1] T. Hunter, *Cell* **1995**, *80*, 225–236.
[2] L. Johnson, R. Lewis, *Chem. Rev.* **2001**, *101*, 2209–2242.
[3] L. Johnson, M. Noble, D. Owen, *Cell* **1996**, *85*, 149–158.
[4] M. Knockaert, P. Greengard, L. Meijer, *Trends Pharmacol. Sci.* **2002**, *23*, 417–425.
[5] I. Bartova, M. Otyepka, Z. Kriz, J. Koca, *Protein Sci.* **2004**, *13*, 1449–1457.
[6] N. Gray, L. Wodicka, A. Thunnissen, T. Norman, S. Kwon, F. Espinoza, D. Morgan, G. Barnes, S. LeClerc, L. Meijer, S. Kim, D. Lockhart, P. Schultz, *Science* **1998**, *281*, 533–538.
[7] A. Lawrie, M. Noble, P. Tunnah, N. Brown, L. Johnson, J. Endicott, *Nat. Struct. Biol.* **1997**, *4*, 796–801.
[8] U. Schulzegahmen, J. Brandsen, H. Jones, D. Morgan, L. Meijer, J. Vesely, S. Kim, *Proteins: Struct., Funct., Genet.* **1995**, *22*, 378–391.
[9] W. DeAzevedo, S. Leclerc, L. Meijer, L. Havlicek, M. Strnad, S. Kim, *Eur. J. Biochem.* **1997**, *243*, 518–526.
[10] T. Sielecki, T. Johnson, J. Liu, J. Muckelbauer, R. Grafstrom, S. Cox, J. Boylan, C. Burton, H. Chen, A. Smallwood, C. Chang, M. Boisclair, P. Benfield, G. Trainor, S. Seitz, *Bioorg. Med. Chem. Lett.* **2001**, *11*, 1157–1160.
[11] P. Traxler, G. Bold, E. Buchdunger, G. Caravatti, P. Furet, P. Manley, T. O'Reilly, J. Wood, J. Zimmermann, *Med. Res. Rev.* **2001**, *21*, 499–512.
[12] M. Lepsik, Z. Kriz, Z. Havlas, *Proteins: Struct., Funct., Bioinf.* **2004**, *57*, 279–293.
[13] P. A. Sims, C. F. Wong, J. A. McCammon, *J. Med. Chem.* **2003**, *46*, 3314–3325.
[14] M. Otyepka, Z. Kriz, J. Koca, *J. Biomol. Struct. Dyn.* **2002**, *20*, 141–154.
[15] M. Otyepka, V. Krystof, L. Havlicek, V. Siglerova, M. Strnad, J. Koca, *J. Med. Chem.* **2000**, *43*, 2506–2513.
[16] P. Hobza, M. Kabelac, J. Sponer, P. Mejzlik, J. Vondrasek, *J. Comput. Chem.* **1997**, *18*, 1136–1150.
[17] J. Vondrasek, L. Bendova, V. Klusak, P. Hobza, *J. Am. Chem. Soc.* **2005**, *127*, 2615–2619.
[18] J. Vondrasek, P. Hobza, unpublished results.
[19] W. Cornell, P. Cieplak, C. Bayly, I. Gould, K. Merz, D. Ferguson, D. Spellmeyer, T. Fox, J. Caldwell, P. Kollman, *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
[20] M. Elstner, P. Hobza, T. Frauenheim, S. Suhai, E. Kaxiras, *J. Chem. Phys.* **2001**, *114*, 5149–5155.
[21] P. Jurecka, P. Hobza, *J. Am. Chem. Soc.* **2003**, *125*, 15608–15613.
[22] P. Jurecka, P. Hobza, *Chem. Phys. Lett.* **2002**, *365*, 89–94.
[23] S. F. Boys, F. Bernardi, *Mol. Phys.* **1970**, *19*, 553.
[24] P. Jurecka, P. Nachtigall, P. Hobza, *Phys. Chem. Chem. Phys.* **2001**, *3*, 4578–4582.
[25] J. Cerny, P. Hobza, *Phys. Chem. Chem. Phys.* **2005**, *7*, 1624–1626.
[26] J. Moravec, V. Krystof, J. Hanus, L. Havlicek, D. Moravcova, K. Fuksova, M. Kuzma, R. Lenobel, M. Otyepka, M. Strnad, *Bioorg. Med. Chem. Lett.* **2003**, *13*, 2993–2996.
[27] A. E. Gibson, C. E. Arris, J. Bentley, F. T. Boyle, N. J. Curtin, T. G. Davies, J. A. Endicott, B. T. Golding, S. Grant, R. J. Griffin, P. Jewsbury, L. N. Johnson, V. Mesguiche, D. R. Newell, M. E. M. Noble, J. A. Tucker, H. J. Whitfield, *J. Med. Chem.* **2002**, *45*, 3381–3393.

Received: October 12, 2005
Published online: March 31, 2006

J Comput Aided Mol Des (2011) 25:223–235
DOI 10.1007/s10822-011-9413-5

Transferable scoring function based on semiempirical quantum mechanical PM6-DH2 method: CDK2 with 15 structurally diverse inhibitors

Petr Dobeš · Jindřich Fanfrlík · Jan Řezáč · Michal Otyepka · Pavel Hobza

Received: 12 September 2010 / Accepted: 18 January 2011 / Published online: 1 February 2011

Abstract A semiempirical quantum mechanical PM6-DH2 method accurately covering the dispersion interaction and H-bonding was used to score fifteen structurally diverse CDK2 inhibitors. The geometries of all the complexes were taken from the X-ray structures and were reoptimised by the PM6-DH2 method in continuum water. The total scoring function was constructed as an estimate of the binding free energy, i.e., as a sum of the interaction enthalpy, interaction entropy and the corrections for the inhibitor desolvation and deformation energies. The applied scoring function contains a clear thermodynamical terms and does not involve any adjustable empirical parameter. The best correlations with the experimental inhibition constants (ln K_i) were found for bare interaction enthalpy ($r^2 = 0.87$) and interaction enthalpy corrected for ligand desolvation and deformation energies ($r^2 = 0.77$); when the entropic term was considered, however, the correlation becomes worse but still acceptable ($r^2 = 0.52$). The resulting correlation based on the PM6-DH2 scoring function is better than previously published function based on various docking/scoring, SAR studies or advanced QM/MM approach, however, the robustness is limited by number of available experimental data used in the correlation. Since a very similar correlation between the experimental and theoretical results was found also for a different system of the HIV-1 protease, the suggested scoring function based on the PM6-DH2 method seems to be applicable in drug design, even if diverse protein–ligand complexes have to be ranked.

Keywords CDK2 · Semiempirical quantum mechanical method · PM6-DH2 · Non-covalent interaction · Scoring function · Drug design

Abbreviations

CDK2	Cyclin-dependent kinase 2
MM	Molecular mechanics
QM	Quantum mechanics
SQM	Semiempirical quantum mechanics
PI	Protein (P)–inhibitor (I) complex

Introduction

Cyclin-dependent kinase 2 (CDK2) is one of the prominent cell cycle regulators [1, 2], which is dominantly active during the G1 phase and G1/S transition. The deregulation of cyclin-dependent kinases (CDKs) is known to be associated with many serious diseases, such as cancer [3, 4].

Electronic supplementary material The online version of this article (doi:10.1007/s10822-011-9413-5) contains supplementary material, which is available to authorized users.

M. Otyepka (✉) · P. Hobza (✉)
Regional Centre of Advanced Technologies and Materials, Department of Physical Chemistry, Faculty of Science, Palacký University, 17 listopadu 12, 771 46 Olomouc, Czech Republic
e-mail: michal.otyepka@upol.cz

P. Hobza
e-mail: pavel.hobza@uochb.cas.cz

P. Dobeš · J. Fanfrlík · J. Řezáč · P. Hobza
Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic and Center for Biomolecules and Complex Molecular Systems, Flemingovo nám. 2, 166 10 Prague 6, Czech Republic

P. Dobeš
Center of Molecular Biology and Gene Therapy, Department of Internal Medicine–Hematooncology, University Hospital Brno, 625 00 Brno, Czech Republic

This fact has attracted attention in the long term to the development of efficient inhibitors of CDKs [5–8]. A relentless effort in this field has succeeded in bringing some CDK inhibitors to clinical trials [9, 10]. Even though the targeting of CDK2 does not have to be an optimal strategy for cancer treatment because of the redundancy of CDKs in cell cycle regulation [11, 12], the CDK2 has remained a paradigm for rational drug design, because it is the best characterised cyclin-dependent kinase in terms of structure and biochemistry [13].

Competition between an inhibitor and the native ATP substrate is an effective strategy to inhibit CDK2. An inhibitor binds to a deep cleft between two CDK2 lobes [14] (Fig. 1a) and, despite the numerous structurally-varied CDK2 inhibitors known today, some common features can be identified. The inhibitors share a flat central ring and form H-bonds to the CDK2 hinge region (Phe80-Leu83 residues), which forms the back wall of the CDK2 active site. The discovery of the CDK2–inhibitor structure [14, 15] provided a useful starting point for the rational design of CDK2 inhibitors.

Since then, many studies exploiting various rational drug-design strategies have been carried out [16–19], among which a prominent place is taken by docking and scoring studies. The main aim of the molecular docking experiment [20, 21] is to find an optimal orientation of the ligand (e.g. inhibitor) in the active site of a target (e.g. an enzyme, in our case CDK2) and to evaluate the ligand activity (e.g. inhibition constant). Empirical scoring functions based on the empirical potentials are often used to evaluate the inhibition activity [22]. Despite the fine tuning and parameterisation of the empirical scoring functions, they are in principle restricted by the known limitations of the empirical potential (also known as molecular mechanics, MM, or force field), which mainly concerns its inability to incorporate quantum effects such as charge transfer between protein and ligand and the wide variation in atomic charges between different structures of the ligand and protein [23, 24]. Another important consequence of quantum effects is the existence of halogen bond that play an important role in the binding of halogenated ligands [25]. This limitation can be overcome e.g. by employing the nonempirical quantum mechanical (QM) ab initio calculation, but this approach is seriously limited by the computer demands since protein–inhibitor complexes possess several thousand atoms. The situation is further complicated since electron correlation (or London dispersion energy) frequently plays an important role in the protein–inhibitor binding, which makes the use of less demanding methods like Hartree–Fock or density functional theory (DFT) impractical [26]. It should be added here that the present case, i.e. inhibitor binding to CDK2, is known to be

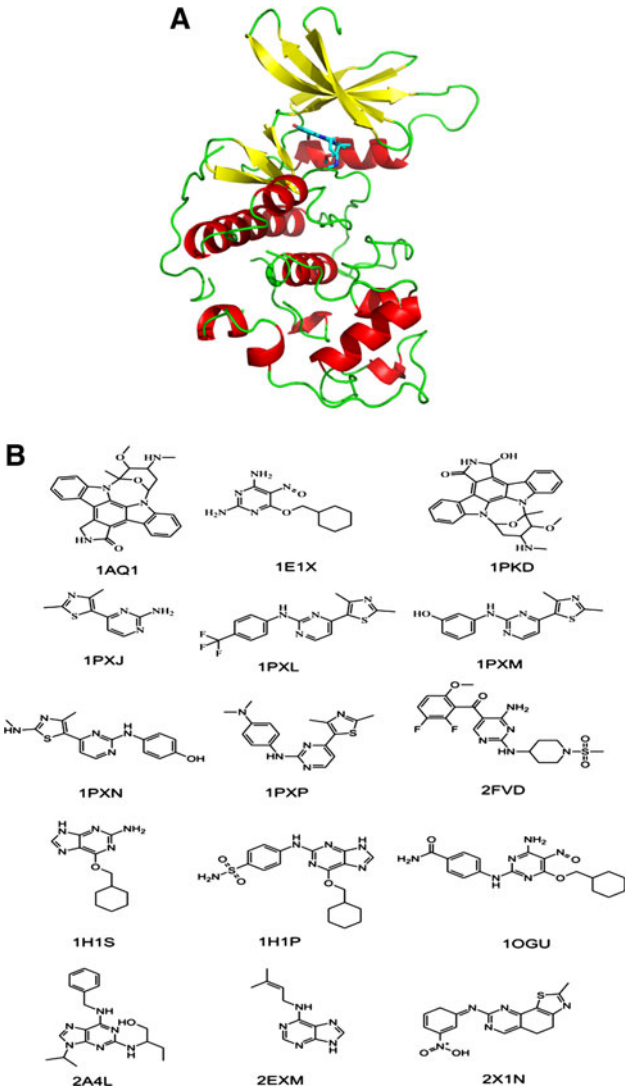


Fig. 1 **a** The ATP competitive inhibitor (here UCN-1 in sticks) binds to a deep cleft in the CDK2 structure (PDB ID code: 1PKD), which is shown in a cartoon model (the red curls represent α-helices and the yellow stripes β-strands, the N-terminal lobe is up and C-terminal lobes is down. **b** The structures of the fifteen inhibitors considered in this study

controlled by the electron correlation (dispersion) energy contributions [27].

The neglect of the quantum effects prevents the use of the computationally accessible MM methods while the size of complexes investigated prevents the use of the highly accurate nonempirical QM methods. There are two possibilities of how to solve the problem. The first is based on the use of the QM/MM approach while the other relies on the use of semiempirical QM (SQM) methods. Both of these cover quantum effects and have already been implemented in rational drug discovery [28–30]. The former enables the study of the protein active site-binding pocket at the QM level, with the rest of the protein–inhibitor complex (including water) being treated at the

MM level. The latter approach describes the whole complex directly using SQM methods. Both approaches obviously have their *pros* and *cons* [31–35], but we believe that the description of the whole complex consistently without any artificial division into QM and MM parts is better suited for drug-design purposes. The use of the SQM method for such large systems is certainly conditioned by the introduction of linear scaling techniques.

The use of the SQM technique in drug design was pioneered by Merz et al. (Refs. [28, 36] and the references therein), who used standard semiempirical AM1 or PM3 Hamiltonians. These methods are, however, known not to cover the London dispersion energy. Wollacott and Merz [36] solved the problem by constructing the binding free energy of a protein–inhibitor complex as a sum of the AM1 or PM3 heat of formation, solvation free energy, and the attractive term from the Lennard-Jones potential. There is a more straightforward way to solve the problem, namely by adding the missing dispersion term to a SQM method (see our recent review [26] for references).

The present SQM calculations are based on Stewart's PM6 semiempirical QM method [37]. Its implementation in the MOPAC package offered a linear scaling algorithm MOZYME [38]. MOZYME is a localized molecular orbital method which replaces the standard SCF procedure in the PM6 calculation. This method speeds us the SQM calculations significantly and unable us to consider systems with several thousands of atoms. For this SQM method, we have developed a dispersion correction as well as a correction for hydrogen bonds [39, 40]. The resulting method (using the second generation of corrections), named PM6-DH2, yields, to the best of our knowledge, the most accurate results for non-covalent interactions of all the SQM methods [26]. For small model non-covalent complexes, which allow for a comparison with the benchmark high-level QM calculations, the method reaches a chemical accuracy of $\sim$ 1 kcal/mol and the quality of the results is comparable to the much more computationally expensive wave-function theory (WFT) and DFT techniques [26].

The main limitation of present approach based on the use of SQM PM6-DH2 method is the computer time. Although the method is due to the use of the linear scaling procedure very efficient (see the end of "Discussion") it is still considerably slower than empirical potentials. In the present as well as in the previous paper we describe the PI complex fully by SQM method. In the case of larger proteins such calculations will be, however, time consuming and some acceleration procedure should be adopted. There are basically two possibilities, either the use parallelization (or even massive parallelization) or the use of QM/MM technique. Our approach will be, however, different from other QM/MM approaches in the literature. We believe that the QM part should be extended (several thousands of atoms) and, therefore, the SQM will be systematically used. This strategy was supported by Merz et al. [41] who showed that their SQM/MM method did not reproduce PL binding affinity as well as their full SQM method. However, the SQM/MM results were quite encourraging and were qualitatively competitive with full SQM results. These results show that SQM/MM approache can be considered for bigger complexes that would be too time demanding for the full SQM calculations.

The procedure described, which is based on the use PM6-DH2 scoring function, was introduced in our previous paper [42], where we studied the complexation of the HIV-1 protease with twenty-two ligands (eleven binders and eleven non-binders). The total score was constructed as an estimate of the binding free energy, i.e., as a sum of the interaction enthalpy, interaction entropy, ligand deformation energy and ligand desolvation free energy. The *in silico* predictions were finally cross-validated by the experimental data and very good agreement between the predictions and experimental data was achieved ($r^2 = 0.71$). All of the binders were either highly polar or charged, which made the desolvation free energy very large, sometimes comparable with the interaction enthalpy. In agreement with experiment, the interaction entropy plays an important role in the HIV-1 system. It was concluded that all the terms in the total score are significant and none of them can be neglected.

In the present study, we have used the same technique for another medicinally important target, CDK2. The PM6-DH2 method was used to score fifteen complexes of CDK2 with inhibitors (Fig. 1b) for which the experimental inhibition constants are known. Contrary to the previous study, most of the inhibitors are neutral, more rigid and less polar than in the case of the HIV-1 protease ligands. This point is important since it demonstrates the wide applicability of the scoring function suggested. Without any modification of the theoretical treatment for different target (CDK2 vs. HIV-1 protease in our previous paper) and different inhibitors (not only by their chemical constitution but also by their different charge, polarity and flexibility) we obtained comparable results. The theoretical predictions are again in a good agreement with the experimental inhibition constants.

In summary, the scoring function constructed as an estimate of the total binding free energy is divided into multiple separate components having clear physical interpretation. Each component is solved by the best performing method considering efficiency and today computer power. Since different procedures (based on different approximations) are adopted for each component the resulting scoring function cannot be directly equal to the change of the binding free energy. As all the components are physically clear terms the limitations stemming e.g. from approximations used in evaluation of each term are well known

which implies that the produres how to improve each term are in principle known. It is important to stress again that the assessed scoring function does not involve any adjustable empirical factor. As the scoring function does not involve any adjustable parameter it can be directly applied to a different target/ligand system without any modification and parameterization, which increases robustness and applicability of the scoring function. The promising results from both systems (HIV-1 and CDK2) with different natures of inhibitor binding document the transferability of the scoring function based on a robust and reliable PM6-DH2 method. However, more robust testing on various targets would be neccessary to find potential limitations of the scoring function.

Computational methods

The PM6-DH2 method provides accurate results for various types of noncovalent complexes, since it includes corrections to the dispersion and H-bonding energies. The dispersion correction is based on our previous work on the DFT-D methods [43] but has been adapted to the PM6 method, because it already contained a part of the dispersion term. The whole dispersion correction was scaled, and the overestimated dispersion contribution of the sp^3 hybridised carbon atoms was corrected. The H-bonding correction added a force field-like term to each pair of atoms which could possibly form an H-bond (as the proton donor and proton acceptor only oxygen and nitrogen were considered). This correction was directional—depending not only on the distance but also on the H-bond angle and other geometric parameters, which substantially improved the geometry of the hydrogen bonds. These corrections are of vital importance for predicting the correct structure of a protein–inhibitor complex.

Theoretical description of the protein–inhibitor binding

The formation of the protein (P)–inhibitor (I) complex from free (hydrated) subsystems represents a crucial step in the virtual docking scheme.

$$\mathrm{P} + \mathrm{I} \rightarrow \mathrm{P} - \mathrm{I} \tag{1}$$

The binding of a competitive inhibitor is evaluated by an inhibition constant (K_i). An inhibitor binding affinity is usually represented also by a half maximal inhibitory concentration (IC_{50}). However, IC_{50} depends on the experimental conditions more than K_i. Therefore, the inhibition (K_i) constant is a more appropriate (but not perfect, because the results from different laboratories may differ because of different experimental conditions) and thermodynamically well-defined term for measuring the inhibitor binding affinity. The inhibition constant is related to the change of the free energy of binding (binding free energy) by the following equation:

$$\Delta G_w = RT \ln(K_i). \tag{2}$$

The direct calculation of the inhibitor binding free energy using MD-based methods is nowadays possible [44–46]. However, these methods, although accurate within the limit of accuracy of empirical potential, are computationally demanding and time-consuming, which limits their application for certain stages of the drug development process, such as high-throughput screening. This quantity may be, however, determined undirectly using the thermodynamic cycle shown in our previous paper [42]. The ligand binding free energy is thus constructed as follows. First, the liquid-phase structure of the inhibitor is dehydrated and its structure is deformed to that the inhibitor possess in the PI complex. Similarly, the same is true about the protein, but the deformation term is presently omitted since we are considering the complexation of one protein with various inhibitors in this study (the desolvation energy of a protein is, however, properly taken into consideration). The deformation and often also the dehydration energies of inhibitor are positive, i.e. they oppose the binding. Second, the dehydrated and deformed inhibitor binds to the protein with an associated change of the interaction free energy. The latter term is determined as a sum of enthalpy and entropy terms. It would be desirable to evaluate all terms mentioned consistently at the same theoretical level (which is also our final goal). Unfortunately, this is unfeasible at present and several approximations have to be adopted in order to estimate the desired free energy of binding as accurately as possible (see later). Consequently, we have obtained only an estimate of the free energy of the inhibitor binding ΔG_w not involving some terms (e.g. the deformation energy of protein, protein desolvation etc.), which are expected to be constant for one target protein. This term is called the scoring function ($\Delta G_w'$) and it is calculated as follows (more detailed analysis is present in our previous paper [42]):

$$\Delta G_w' = \Delta H_w - T\Delta S_w + \Delta E_{def}(\mathrm{I}) + \Delta\Delta G_w(\mathrm{I}). \tag{3}$$

The individual terms represent the interaction enthalpy (ΔH_w), the interaction entropy (as $-T\Delta S_w$) and the corrections for inhibitor deformation energy $\Delta E_{def} = E(\mathrm{I})^{PI} - E(\mathrm{I})^{I}$ (with the upper indexes PI and I standing for the geometry of the inhibitor in a protein/inhibitor complex and the relaxed inhibitor geometry in water, respectively) and the corrections for the inhibitor hydration free energy $\Delta\Delta G_w(\mathrm{I}) = \Delta G_w^{MOPAC}(\mathrm{I})^{PI} - \Delta G_w^{G09}(\mathrm{I})^{PI}$ (where the lower index w stands for a water environment). The last term reflects the fact that structure

of the PI complex (in the water environment) is determined using the hydration model considering the electrostatic term only (MOPAC code) while accurate treatment requires the use of hydration model considering the nonelectrostatic term as well. Nonelectrostatic terms associated with formation of a cavity to accommodate the solute, and the van der Waals interaction between solute and solvent, are in the former approach neglected. The electrostatic approximation might work well for inhibitors of the similar shape but in the case of entirely different inhibitors it can cause a serious error. From this reason we adopted a more accurate calculation of the complete solvation free energy of the ligand with a SMD model based on IEFPCM calculation with radii and non-electrostatic terms by Truhlar and coworkers [47] as implemented in Gaussian 09 code.

The formation of a PI complex restricted motions of the ligand as well as the torsional motion of the protein. Furthermore, the vibrational entropy of a ligand was also restricted by the formation of a PI complex. The total entropy term was repulsive and, like the desolvation term, opposed the binding. The change of entropy accompanying the PI complex formation in a water environment was determined as the difference between entropies of a complex and a sum of the subsystem entropies. Ligand binding led also to the loss of the configurational entropy. The evaluation of the configuration entropy is notoriously difficult and this term was either neglected or was estimated on the basis of a reduced number of the accessible rotamers upon binding. In the present study we used the recently shown [48] evidence that vibrational entropy dominated configurational entropy. Consequently, it was recommended to evaluate the entropy term for the PI binding on the basis of the standard rigid rotor/harmonic oscillator approximation. The entropic contributions were determined using the rigid rotor/harmonic oscillator approximation based on Cornell et al. [49] empirical potential constants. The use of semiempirical Hamiltonian for the entropy calculations would be desirable but presently it is behind our possibilities. The structures of all of the systems were re-optimized using the empirical potential considering the continuous water, and the same method was used for the calculation of the second derivatives of the total energy.

Summarizing the procedure described we stress again that its main advantage is the fact that no any additional empirical parameter either for an individual component of the total score or for an individual PI complex was introduced.

Structures and inhibition constants

The structures of all the complexes (Table 1) were taken from the RSCB Protein Data Bank. Two different structural types were considered. In the first, the inhibitors are bound to the fully active form of CDK2 (Thr160-phosporylated CDK2 in complex with cyclin A3) while in the second type they are bound to the inactive form (monomeric CDK2). In the case of the Thr160-phosporylated CDK2/cyclin A3–inhibitor complexes, only one CDK2 monomer with removed cyclin A3 was used. A longer chain of CDK2 was considered for a complex with the 1OGU inhibitor. All of the crystal water molecules were removed. The hydrogen atoms were added to all of the structures considered, and their positions were minimised using AMBER [50]—ff03 [51] and gaff [52] force fields in Chimera software [53]. Partial charges of ligand atoms were calculated by RESP method from HF/6-31G* ESP charges for ff99 optimizations and from B3LYP/cc-pVTZ ESP charges for ff03 optimizations. In most cases, the crystallographically determined structures contained gaps in the chain (see Table 1), but they were situated far from the binding site of the inhibitor in complex, thus only a naturally charged group like COO^- and NH_3^+ was added to the amino acids on the gap termini to compensate for the charge of the whole structure. Staurosporines (1AQ1 and 1PKD) were considered N-protonated. The protonation of 1AQ1 ligand was deduced from the H-bonds network observed in the X-ray structure, however, the protonation state of 1PKD remained uncertain from the X-ray structure. The inhibition constants of all of the CDK2 inhibitors studied were taken from the literature and all of the references are listed in Table 1.

Strategy of calculations

The structures of all of the protein-inhibitor complexes were systematically reoptimised by the PM6-DH2 method in a continuum COSMO solvent model [54] as implemented in the MOPAC code (http://OpenMOPAC.net) according to MOPAC convergence criteria defined by exiting as soon as the gradient norm drop below 10.0 kcal/mol/Å. The respective interaction enthalpies ($\Delta H_w = \Delta H_w(\mathrm{PI}) - (\Delta H_w(\mathrm{P}) + \Delta H_w(\mathrm{I}))$) were determined at the same level ($T = 298$ K, $p = 1$ atm). In the second step, all of the complexes were reoptimised with AMBER ff99 or ff03 force fields (the missing ligand binding parameters were adopted from AMBER gaff force field) with a generalised Born solvent model [55] in the Nucleic Acid Builder (NAB, from the AMBER [50] package). The entropy term ($T = 298$ K, $p = 1$ atm.) was determined using the ideal gas/rigid rotor/harmonic oscillator approximation using the empirical potential, and the geometries of the systems were reoptimised at the empirical level with the L-BFGS TNCG algorithm to the gradient of 10^{-6} kcal/mol/Å and reoptimised by the Newton–Raphson method. This approach, assuming that the biomolecule occupies a single harmonic well, is known not to be fully adequate for biomolecules transversing many thermally accessible potential wells and

Table 1 A summary of the experimental data and calculated terms for the fifteen CDK2 inhibitors considered in this study

Structure	Res.[a]	Gaps[b]	ln(K_i)[c]	ln(err, K_i)[d]	ΔH_w^e	ΔE_{def}(I)[f]	$\Delta\Delta G_w$(I)[g]	$T\Delta S_w^h$	$\Delta G_w'$[i]
1AQ1 [73]	2.00	37–43, 150–160	−19.66 [74]	0.14	−55.63	−1.77	5.67	3.61	−55.34
1E1X [75]	1.85	37–43	−13.55 [75]	0.16	−31.66	0.75	−3.94	−23.45	−11.40
1PKD*	2.30	–	−17.32 [74]	0.24	−39.49	−0.40	7.87	−23.69	−8.33
1PXJ [76]	2.30	37–40	−11.94 [77]	0.50	−25.48	0.85	−2.87	−18.65	−8.85
1PXL [76]	2.50	37–43	−15.05 [77]	1.18	−34.35	0.55	−3.30	−17.82	−19.28
1PXM [76]	2.53	37–43	−16.63 [77]	0.55	−41.10	3.12	−1.45	−32.22	−7.21
1PXN [76]	2.50	37–40	−16.47 [77]	1.28	−48.83	3.52	−2.28	−27.61	−19.98
1PXP [76]	2.30	37–40	−15.33 [77]	1.88	−38.15	0.75	−3.99	−34.22	−7.17
2FVD [78]	1.85	38–49, 150–151	−19.62 [78]	0.27	−52.89	5.94	−7.03	−29.51	−24.47
1H1P [79]*	2.10	–	−11.33 [75]	0.26	−24.84	1.10	−1.38	−22.29	−2.83
1H1S [79]*	2.00	–	−18.93 [79]	0.08	−52.83	1.48	−2.62	−10.72	−43.25
1OGU [80]*	2.60	38–40	−17.55 [68]	0.35	−53.56	1.34	−1.87	−17.42	−36.67
2A4L [14]	2.40	36–47	−13.63 [81]	NA	−29.99	0.39	3.71	−13.97	−11.92
2EXM [82]	1.80	–	−9.46 [83]	NA	−22.10	0.19	−2.08	−15.97	−8.02
2X1N [84]*	2.75	–	−17.59 [84]	NA	−38.55	0.55	−4.52	−13.82	−28.70

NA not available

[a] Res. (in Å) is the experimental mean resolution of the respective X-ray structure, [b] missing residues in the X-ray structure, [c] ln(K_i) is the logarithm of inhibition constant (K_i in M), [d] ln(err. K_i) is the error of inhibition constant taken from literature, [e] ΔH_w is the interaction enthalpy, [f] ΔE_{def}(I) is correction for inhibitor deformation energy, [g] $\Delta\Delta G_w(I)$ is correction for ligand desolvation free energy, [h] $T\Delta S_w$ is the entropic contribution calculated by ff03, [i] $\Delta G_w'$ is the total score; all energies are in kcal/mol

* These structures contain the fully active form of CDK2, i.e. pT160-CDK2/Cyclin A3 system; only CDK2 with an inhibitor was considered here

may introduce some error. The interaction enthalpies (ΔH_w^{AMBER}) were also determined using the AMBER force fields ($T = 298$ K, $p = 1$ atm.) and generalised Born solvent model (Table 2).

The correction for the deformation energy of the inhibitor (ΔE_{def}(I)) was calculated as the difference between the energy of the inhibitor in the geometry taken from the protein/inhibitor complex ($E(I)^{PI}$) and the energy of the fully optimised inhibitor in water ($E(I)^{w}$). The PM6-DH2 optimizations with the COSMO solvation model were used for this calculation. The correction for the inhibitor deformation energy (ΔE_{def}^{AMBER}(I)) was also calculated at the empirical level using AMBER force fields with a generalised Born solvent model (Table 2).

The correction of the inhibitor desolvation free energy $\Delta\Delta G_w$(I) ($T = 298$ K, $p = 1$ atm.) was calculated on an inhibitor taken from the optimised CDK2–inhibitor complex. It was determined as the difference between the free energy of the inhibitor solvation calculated by the PM6-DH2/COSMO method ($\Delta G_w^{MOPAC}(I)^{PI}$) and the free energy of the inhibitor solvation calculated by the SMD model [47] (involving all of the non-electrostatic terms, i.e. the cavitation, dispersion and repulsion terms) at the HF/6-31G* level as implemented in Gaussian 09 [56] ($\Delta G_w^{G09}(I)^{PI}$) on PM6-DH2/COSMO geometry. We are aware that this energy difference is calculated with different hamiltonians (PM6 and DFT). However, better description of just the solvation of the ligand itself improves the final result. We believe that this procedure is also theoretically justified: the contribution from the protein is the difference between empty and filled active site, in our approach modelled consistently at PM6 level. In the case of the active site occupied by the ligand, only very small part of the surface of the ligand is exposed to the solvent, and therefore the contribution of the ligand to the $\Delta\Delta G_w$ of the protein is small. This separation of the second contribution, the solvation of the free ligand, allows us to use different method to calculate it more accurately.

We are aware of one important limitation of the model used which concerns the use of optimized crystal structure instead of averaged ones generated by MD simulations. The proteins are not rigid but moving in the solvent and, thus, the the single terms in the total score (Eq. 3) should be considered as an average quantity. In our papers as well as in majority of scoring function based studies this effect is neglected. The preliminary results obtained in our laboratory for different PI complexes indicate that effect of averaging is not critical. Nevertheless, these effects are under investigation in our laboratory and will be published as soon as possible.

Results

PM6-DH2 scoring of CDK2–inhibitor complexes

This study used fifteen structurally diverse CDK2 inhibitors with known inhibition constants (K_i) and the X-ray

Table 2 A summary of the experimental data and calculated terms for the fifteen CDK2 inhibitors by AMBER force fields ff99 and ff03

Structure	$\ln(K_i)$	ff99				ff03			
		ΔH_w^{MM} [a]	ΔE_{def}^{MM} [b]	$T\Delta S_w^c$	$\Delta Gw'^{MM}$ [d]	ΔH_w^{MM} [a]	ΔE_{def}^{MM} [b]	$T\Delta S_w^c$	$\Delta G_w'^{MM}$ [d]
1AQ1	−19.66	−39.61	6.58	−20.07	−12.97	−13.97	5.62	3.61	−11.96
1E1X	−13.55	−43.17	3.51	−29.00	−10.66	−36.55	3.08	−23.45	−10.02
1PKD	−17.32	−46.39	1.54	−20.42	−24.43	−41.14	2.12	−23.69	−15.33
1PXJ	−11.94	−12.48	1.75	−13.75	3.02	−17.56	3.64	−18.65	4.74
1PXL	−15.05	−41.09	1.24	−16.67	−23.18	−12.00	1.43	−17.82	7.24
1PXM	−16.63	−42.57	2.52	−41.19	1.13	−32.10	2.75	−32.22	2.87
1PXN	−16.47	−25.97	1.76	−20.22	−3.98	−39.33	3.11	−27.61	−8.61
1PXP	−15.33	−13.43	1.69	−7.75	−3.98	−53.10	1.62	−34.22	−17.26
2FVD	−19.62	−25.38	6.90	−28.47	10.00	−34.06	5.00	−29.51	0.44
1H1P	−11.33	−14.32	1.10	−13.70	0.48	−18.43	2.81	−22.29	6.66
1H1S	−18.93	−33.49	5.93	−20.95	−6.62	−31.72	5.12	−10.72	−15.87
1OGU	−17.55	−19.09	3.78	−10.97	−4.34	−29.97	2.84	−17.42	−9.71
2A4L	−13.63	−17.03	2.44	−22.15	7.55	−42.67	1.86	−13.97	−26.84
2EXM	−9.46	−19.58	1.67	−21.03	3.12	−36.92	1.26	−15.97	−19.69
2X1N	−17.59	−28.42	1.80	−22.20	−4.41	−20.62	2.14	−13.82	−4.65

[a] ΔH_w^{MM} stands for the interaction enthalpy calculated by AMBER ff99 or ff03 force fields, [b] ΔE_{def}^{MM} (I) corrections for inhibitor deformation energy, [c] $T\Delta S_w$ for the entropic contribution, [d] $\Delta G_w'^{MM}$ the total score; all energies are in kcal/mol

structures of CDK2–inhibitor complexes. The CDK2–inhibitor complexes were fully geometrically optimized (Table 3) by the PM6-DH2 method using the COSMO continuum solvent (involving the electrostatic terms only). The interaction enthalpy ΔH_w of the inhibitor with CDK2 was calculated at the same level as described in the Computational Details section. Figure 2a shows the correlation between the interaction enthalpy (ΔH_w) and the experimental inhibition constant ($\ln K_i$) values, where a very good correlation ($r^2 = 0.87$, $\ln K_i = (0.25 \pm 0.06)$ $\Delta H_w - (5.6 \pm 2.4)$, RMSE = 1.17, $n = 15$) was found. If the 1PKD is removed from the dataset (due to uncertain protonation state) the correlation slightly improves ($r^2 = 0.89$, $\ln K_i = (0.25 \pm 0.06)$ $\Delta H_w - (5.5 \pm 2.3)$, RMSE = 1.04, $n = 14$). In our previous study on HIV-1 protease, we demonstrated that the best correlation between the experimental binding free energy and theoretical data was not for the interaction enthalpy but for the total score containing also the interaction entropy, ligand deformation energy and ligand desolvation free energy. The good correlation obtained here between the interaction enthalpy and the inhibition constants indicates some compensation between the remaining terms. Our final goal is to estimate the free energy of the inhibitor binding (which is related to K_i, see Eq. 2), and, in order to achieve this, the interaction enthalpy ΔH_w has to be augmented by the interaction entropy term ($-T\Delta S_w$), inhibitor deformation energy (ΔE_{def}(I)) and correction for inhibitor desolvation. The solvation energy calculated by the COSMO model (implemented in MOPAC), which is involved in the ΔH_w, does not account for the non-electrostatic terms [57], and the estimate of the free energy of binding should be corrected also for this shortcoming. We have corrected the scoring function for the non-electrostatic terms of inhibitor solvation ($\Delta\Delta G_w$(I)). The correlation between the interaction enthalpy and inhibition constant is slightly worse when the corrections for inhibitor deformation ($\Delta E_{def}(I)$) and desolvation energies ($\Delta\Delta G_w(I)$) are also considered (Fig. 2b) with $r^2 = 0.77$ ($\ln K_i = (0.24 \pm 0.08)$ ($\Delta H_w + \Delta E_{def}(I) + \Delta\Delta G_w(I)$) $- (5.9 \pm 3.2)$, RMSE = 2.02, $n = 15$). In a multivariable linear fit, coefficients of both added variables ($\Delta E_{def}(I)$ and $\Delta\Delta G_w(I)$) to ΔH_w are statistically insignificant (at $\alpha = 0.05$) from zero ($\ln K_i = (0.25 \pm 0.07)$ $\Delta H_w + (0.00 \pm 0.56)$ $\Delta E_{def}(I) - (0.04 \pm 0.24)$ $\Delta\Delta G_w(I) - (5.7 \pm 2.7)$, $r^2 = 0.87$, RMSE = 1.15, $n = 15$). If the 1PKD is removed from the data set the correlation is significantly higher with $r^2 = 0.87$ ($\ln K_i = (0.26 \pm 0.06)$ ($\Delta H_w + \Delta E_{def}(I) + \Delta\Delta G_w(I)$) $- (5.0 \pm 2.6)$, RMSE = 1.18, $n = 14$). This indicates that corrections for ligand deformation and desolvation are not important in the case of CDK2. On the other hand, these corrections might become signifant in other systems as shown in the case of HIV-1 protease (see below). When, however, the interaction enthalpy is augmented only by the interaction entropy term ($-T\Delta S_w$, Fig. 2c) calculated at the empirical level (using AMBER ff03 force field), the correlation becomes worse ($r^2 = 0.56$, $\ln K_i = (0.14 \pm 0.08)$ ($\Delta H_w - T\Delta S_w$) $- (12.8 \pm 1.9)$, RMSE = 3.95, $n = 15$) but this value is still acceptable (the liner model and both variables are statistically significant at $\alpha = 0.05$, if the 1PKD structure is

Table 3 Summary of root-mean-square deviations (RMSD) of backbone Cα atoms and ligand atoms in PM6-DH (SQM) and AMBER (MM) optimized structures versus the X-ray structure

	SQM RMSD Cα	MM RMSD Cα	SQM RMSD lig	MM RMSD lig	X-ray E81 (backbone CO)	SQM	MM	X-ray L83 (backbone NH)	SQM	MM	X-ray L83 (backbone CO)	SQM	MM
1AQ1	0.63	1.63	0.24	0.27	2.75	2.89	2.80	2.63	2.84	2.84			
1E1X	0.52	1.14	0.39	0.35	2.64	2.85	2.75	2.97	3.17	3.09	2.49	2.71	2.82
1PKD	0.58	1.03	0.59	0.64	2.81	2.96	2.86	2.74	3.10	3.09			
1PXJ	0.57	1.46	0.23	0.62	2.86	2.95	2.81	3.30	3.62	3.23			
1PXL	0.49	1.20	0.46	0.79				2.92	3.25	3.02	2.58	2.79	3.04
1PXM	0.56	1.25	0.48	0.49				3.08	3.24	3.04	2.81	2.88	2.95
1PXN	0.59	1.35	0.39	0.82				2.85	3.35	2.93	2.58	2.94	2.94
1PXP	0.51	1.65	0.91	0.59				2.95	3.27	2.98	2.70	2.91	3.03
2FVD	0.49	1.23	0.49	0.46	2.74	2.79	2.82	3.68	3.32	3.19	2.76	2.87	2.86
1H1P	0.57	1.11	0.58	0.86	2.79	2.85	2.79	3.06	3.20	3.24	2.72	2.93	2.78
1H1S	0.60	1.04	0.28	0.32	2.83	2.85	2.79	3.35	3.24	3.37	2.82	2.81	2.82
1OGU	0.53	1.04	0.64	0.47	2.72	2.77	2.81	3.50	3.23	3.30	2.74	2.80	2.82
2A4L	0.58	1.18	0.61	0.80				3.38	3.36	3.37	2.82	2.89	2.88
2EXM	0.57	1.39	0.70	1.35	2.80	2.85	2.78	3.39	3.25	3.24			
2X1N	0.61	1.04	0.46	0.42				3.13	3.13	3.18	2.62	2.83	2.82

Values are given in Angstrom unit

Distances of key H-bonds between ligands and backbone atoms of CDK2 hinge residues (E81-L83) in X-ray, PM6-DH (SQM) and AMBER (MM) optimized structures

removed the correlation equals to $r^2 = 0.59$, ln $K_i = (0.15 \pm 0.08)$ $(\Delta H_w - T\Delta S_w) - (12.6 \pm 1.9)$, RMSE = 3.84, $n = 14$). In our previous study the inclusion of entropy term improved the correlation with experiment. This was fully in agreement with experimental finding showing that in the case of HIV-1 protease the entropy term is important. It is thus evident that entropy plays an important role (in case of HIV-1 protease) and it should not be neglected. Finally, when considering the complete scoring function $\Delta G_w'$, i.e. the sum of the interaction enthalpy, interaction entropy and both corrections for inhibitor deformation and solvation energies, the correlation with the experimental inhibition constants becomes $r^2 = 0.52$ (ln $K_i = (0.14 \pm 0.08)$ $\Delta G_w' - (12.8 \pm 2.0)$, RMSE = 4.31, $n = 15$), Fig. 2d; the removal of 1PKD does not improve the correlation significantly) and it is significantly worse than in the case of the interaction enthalpy (ΔH_w) itself. The decrease in the correlation between the calculated terms involving the entropy contribution and the inhibition constant is not surprising and documents that a reliable estimation of the entropic term cannot be based on the harmonic approximation, since a biomolecule can sample multiple thermally accessible minima, and also the empirical force field. On the other hand, the calculated entropy terms can play a significant role, as in the case of HIV-1 protease. For the sake of completeness we provide here also r^2 values for HIV-1 protease with 11 inhibitors: When all energy terms were included it equals to 0.52 and this correlation is comparable with r^2 value for HIV-1 protease [compare: $r^2 = 0.71$, ΔG_i (HIV-1 PR) = (0.11 ± 0.06) $\Delta G_w' - (14.5 \pm 0.9)$ vs. $r^2 = 0.52$, ln K_i (CDK2) = (0.17 ± 0.08) $\Delta G_w' - (11.8 \pm 2.2)$]. However, when only enthalpy term was included in the case of HIV-1 protease the correlation become statistically insignificant ($r^2 = 0.09$). The ligands of the HIV-1 protease are more flexible, therefore correction for ligand deformation and entropy play more important role in the case of HIV-1 protease.

The AMBER scoring of CDK2–inhibitor complexes

As in the previous case, the structures of all fifteen complexes of CDK2 with inhibitors were scored by AMBER force fields (both ff99 and ff03) with a generalised Born solvent model. Contrary to the previous case, where some single contributions to the total score were determined at different levels, in the present case all of the components of the total Amber score $\Delta G_w'^{AMBER}$, i.e. the sum of the interaction enthalpy ΔH_w^{AMBER}, interaction entropy $-T\Delta S_w$ and inhibitor deformation energy correction $\Delta E_{def}^{AMBER}(I)$, were determined consistently at the same empirical level. From Fig. 2e, it is evident that no statistically significant

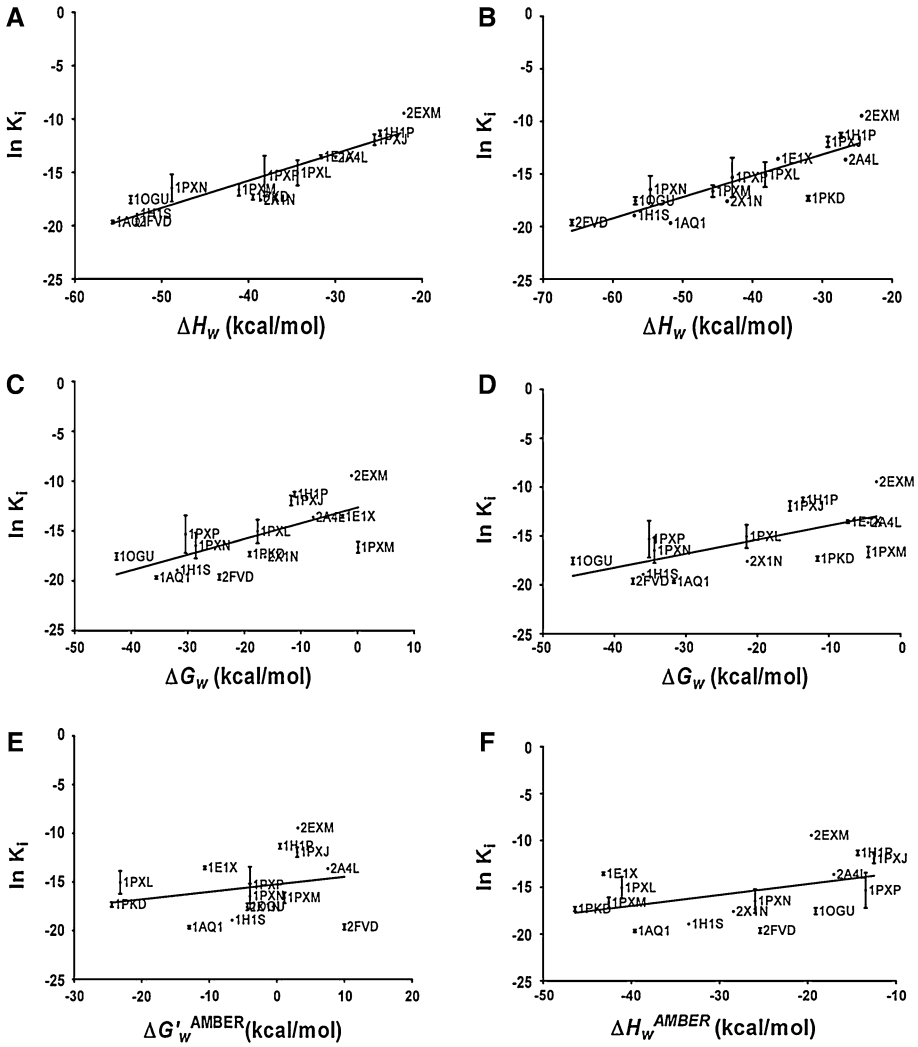


Fig. 2 The correlations between the inhibition constants (ln K_i) and **a** the interaction enthalpy ΔH_w (in kcal/mol) with $r^2 = 0.87$, **b** the sum of interaction enthalpy ΔH_w and the corrections for the inhibitor deformation ($\Delta E_{def}(I)$) and desolvation ($\Delta\Delta G_w(I)$) energies (in kcal/mol) with $r^2 = 0.77$, **c** the sum of the interaction enthalpy ΔH_w and interaction entropy $-T\Delta S_w$ (in kcal/mol) with $r^2 = 0.56$, **d** the total score $\Delta G_w'$ calculated as the sum of the interaction enthalpy (ΔH_w), the interaction entropy ($-T\Delta S_w$) and the corrections for the inhibitor deformation ($\Delta E_{def}(I)$) and desolvation ($\Delta\Delta G_w(I)$) energies (in kcal/mol) with $r^2 = 0.52$, **e** the total score $\Delta G_w'^{AMBER}$, where all of the terms were determined consistently using AMBER ff99 force field with $r^2 = 0.06$, **f** the interaction enthalpy ΔH_w^{AMBER} calculated using AMBER ff99 force field with $r^2 = 0.21$

correlation ($r^2 = 0.06$ for ff99 and $r^2 = 0.00$ for ff03) exists between the experimental inhibition constant (ln K_i) and the total AMBER score $\Delta G_w'^{AMBER}$, and the same applies also for the correlation ($r^2 = 0.21$ for ff99 and $r^2 = 0.00$ for ff03) with the interaction enthalpy ΔH_w^{AMBER} (Fig. 2f). This finding supports the idea that the interaction enthalpy should be determined as accurately as possible and that the empirical level used is clearly not adequate, at least in the case of CDK2.

Decomposition of the total score

Figure 3 shows the single contributions to the total PM6-DH2 score for four selected inhibitors. Following Eq. 3 the total score (the last bar) is constructed as a sum of the interaction enthalpy in solution (the first bar), the deformation energy of an inhibitor (the second bar), the correction for the inhibitor hydration free energy (the third bar) and the interaction entropy (the fourth bar). Evidently, the interaction enthalpy in solution (which varies from −34 to −52 kcal/mol) represents the dominant attractive term while deformation energy and entropy terms are repulsive and oppose the binding.

Optimized structures

Table 3 summarizes geometrical parameters of the optimized CDK2-inhibitor complexes (see Supplementary data for the optimized geometries). On average, the RMSD between X-ray and optimized structure is smaller for the PM6-DH method than for the empiricial calculation with ff99. According to a visual inspection, the PM6-DH and ff99 optimized structures are very similar, but 1PXP structure (and to some extent also 2FVD) optimized at the empirical level deviates from both X-ray and PM6-DH optimized structures. The key H-bond distances between inhibitor and CDK2 hinge residues are shorter in X-ray than in the optimized structures.

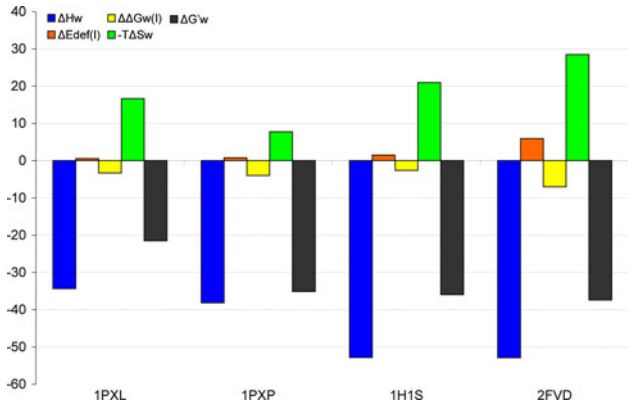


Fig. 3 The decomposition of the total score $\Delta G_w'$ into the interaction enthalpy in solution (ΔH_w), the deformation energy of an inhibitor (ΔE_{def}(I)), the correction for the inhibitor hydration free energy ($\Delta\Delta G_w$(I)) and the interaction entropy ($-T\Delta S_w$). All energies are in kcal/mol. The decomposition is shown for four inhibitors; **1PXP**: *N*-[4-(2,4-dimethyl-thiazol-5-yl)-pyrimidin-2-yl]-*n*′,*n*′-dimethyl-benzene-1,4-diamine, **1PXL**: 4-(2,4-dimethyl-1,3-thiazol-5-yl)-*n*-[4-(trifluoromethyl)phenyl]pyrimidin-2-amine[4-(2,4-dimethyl-thiazol-5-yl)-pyrimidin-2-yl]-(4-trifluoromethyl-phenyl)-amine], **2FVD**: (4-amino-2-([1-(methylsulfonyl)piperidin-4-yl]amino)pyrimidin-5-yl)(2,3-difluoro-6-methoxyphenyl)methanone, and **1H1S**: 4-[[6-(cyclohexylmethoxy)-9H-purin-2-yl]amino]benzenesulfonamide

Discussion

A sufficiently robust and reliable scoring function capable of scoring CDK2 inhibitors is an attractive but also challenging task. The attractiveness of this tasks arises from the fact that CDK2 is a member of a large family of protein kinases [58], which are involved in cellular signal transduction and the regulation of many cellular processes. Mutations and deregulations of protein kinases play also causal roles in many human diseases, and therefore great effort is invested in the development of protein kinase inhibitors (see Ref. [58] and the references therein). Despite this appeal, finding a reliable scoring function which could be highly predictive in the *in silico* drug design of CDK2 inhibitors still represents a challenging task. The first docking study, which was published 10 years ago, involved eight inhibitors, and the authors concluded that they had not found any statistically significant relationship between the activity (IC_{50}) and the interaction energy (using the DOCK score [59]) because of the low correlation coefficient ($r^2 = 0.15$) [60]. Later, many authors tried to find a more reliable function using the Gold score ($r^2 = 0.50$ between the Gold score and IC_{50}) [61], the Glide score ($r^2 = 0.61$ to IC_{50}) [62], the FlexX and LigandFit (the best score $r^2 = 0.60$ with IC_{50}) [63], GlideXP rescored by a molecular mechanics-generalised Born with a surface area (MM-GBSA) method ($r^2 = 0.69$ with IC_{50}) [64], scoring based on molecular dynamics simulations with a quantum-refined force field ($r^2 = 0.55$ with IC_{50}) [65], and the CDOCKER docked ligands rescored by the MM-GBSA method ($r^2 = 0.36$ after the removal of the outliers $r^2 = 0.63$, with pK_i) [66]. In general, the pharmacophore model named HypoRefire ($r^2 = 0.83$ between the score and IC_{50} for 302 molecules) [19] and 3D-QSAR CoMFA ($r^2 = 0.75$ to IC_{50}) [67] and CoMSIA ($r^2 = 0.81$ to IC_{50}) [67] gave better correlations than the scoring functions used by the docking software.

It is obvious that the best scoring function is able to explain only about 70% of the data variability and that the available scoring functions are not capable of describing all the physics behind the inhibitor binding to CDK2. Many authors have analysed the nature of CDK2–inhibitor binding using various methods and concluded that the dispersion interactions (part of van der Waals interactions) is important for inhibitor binding [17, 18, 27, 68] and may outweigh electrostatics [23]. Evidently, however, not only the dispersion but also the other terms should be properly covered. This is demonstrated by the fact that the total scoring with AMBER, which evaluates dispersion energy satisfactory [27, 69, 70], does not correlate well with the experimental data (see above). This further implies that also quantum chemical methods have to be chosen with caution, because any methods unable to cover the dispersion interaction (the HF method and the vast majority of the popular DFT functionals, e.g. B3LYP, and semiempirical methods, e.g. AM1 and PM6) will fail [26, 27]. In this respect, it is not surprising that the hybrid QM/MM approach did not provide any better correlation ($r^2 = 0.64$ with IC_{50} for 75 compounds) [71] than the empirical scoring functions, because it was based on the ONIOM model, where all of the layers (B3LYP/6-31G(d):HF/3-21G:PM3) failed in the description of the dispersion interaction. A large compensation of the gas-phase interaction energy by the solvation energy also has to be considered [18].

A reliable scoring function for CDK2 inhibitors has to involve a correct enough description of the dispersion and polarization interactions, but also the other physical processes have to be described carefully, as they might play an important role. Among them, the inhibitor deformation energy, changes in the inhibitor solvation (inhibitor desolvation) and the entropic contributions are likely candidates to be considered [64]. Here, we present a scoring function whose core is based on the semiempirical QM method accurately covering H-bonding and dispersion interaction (PM6-DH2) and further extended by corrections for the inhibitor deformation energy, the inhibitor desolvation and the entropic term. The last three terms are, however, based on empirical force field calculations. It must be stressed here that no empirical, adjustable parameter either for the single terms of the total score or for any individual inhibitor was introduced. This makes the

use of present technique for diverse protein–ligand complexes very promising. We have chosen fifteen structurally-diverse inhibitors of CDK2 with known X-ray structures and inhibition constants K_i. The correlation between pK_i and ΔH_w amounts to $r^2 = 0.87$, which is the best correlation found in the set of the above-discussed empirical scoring functions, pharmacological and 3D-QSAR models. When the corrections for the inhibitor deformation energy and inhibitor desolvation are involved, the correlation worsens to $r^2 = 0.77$. When full score is considered (i.e. also the entropic term is added), the correlation is significantly lower $r^2 = 0.52$. The decrease caused by the entropy term can be explained by the fact that a reliable calculation of the entropic changes based on the harmonic approximation and evaluated at empirical MM level is not adequate for biomolecules. The scoring function presented here yielded the best correlation with in K_i considering the empirical scoring function used in the docking experiments. There is also some room for improvement, as the correlation is not perfect, and specifically a better estimation of the entropic term would be useful. On the other hand, the presented scoring function represents significant progress in the field not only because it performs well but also because it does not require any additional empirical parameters, and therefore the availability of the protein–inhibitor complex (at least protein) structure is the only, albeit unavoidable, condition. The method is also capable of treating the complete CDK2–inhibitor complex, which typically contains about 5,000 atoms. The only known drawback is the computer time, because a single-point calculation (evaluation of one drug-receptor pose) typically takes $\sim$20 min (one core of Intel Core2 Quad 2.40 GHz processor) and the optimisation of the CDK2–inhibitor complex takes up to 3 weeks (1 core of Intel Core2 Quad 2.40 GHz processor). On the other hand, there is a large scope for acceleration, e.g. parallelisation, the usage of hybrid QM/MM models etc., and the increasing power of computers (Moor's law) works also for us. Considering all the *pros* and *cons*, the presented scoring function may represent a promising step in a development of a reliable scoring function for future drug design. This fact was supported by a very recent work by Zhou and Caflish, who showed that scoring functions based on quantum mechanics can be applied for high-throughput virtual screening [72].

Acknowledgments This work was a part of research project No. Z40550506 of the Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic and was supported by Grants No. LC512 and MSM6198959216 from the Ministry of Education, Youth and Sports of the Czech Republic. The support of Praemium Academiae, Academy of Sciences of the Czech Republic, awarded to P.H. in 2007, is also acknowledged. It was also supported by The Czech Science Foundation (P208/11/0295). This work was supported by the Operational Program Research and Development for Innovations—European Social Fund (CZ.1.05/2.1.00/03.0058).

References

1. Morgan DO (1997) Cyclin-dependent kinases: engines, clocks, and microprocessors. Annu Rev Cell Dev Biol 13:261–291
2. Malumbres M, Barbacid M (2005) Mammalian cyclin-dependent kinases. Trends Biochem Sci 30(11):630–641
3. Malumbres M, Barbacid M (2009) Cell cycle, cdks and cancer: a changing paradigm. Nat Rev Cancer 9(3):153–166
4. Child ES, Hendrychova T, McCague K, Futreal A, Otyepka M, Mann DJ (2010) A cancer-derived mutation in the pstaire helix of cyclin-dependent kinase 2 alters the stability of cyclin binding. Biochim Biophys Acta Mol Cell Res 1803(7):858–864
5. Cohen P (1999) The development and therapeutic potential of protein kinase inhibitors. Curr Opin Chem Biol 3(4):459–465
6. Besson A, Dowdy SF, Roberts JM (2008) Cdk inhibitors: cell cycle regulators and beyond. Dev Cell 14(2):159–169
7. Fischer PM, Lane DP (2000) Inhibitors of cyclin-dependent kinases as anticancer therapeutics. Curr Med Chem 7:1213–1245
8. Knockaert M, Greengard P, Meijer L (2002) Pharmacological inhibitors of cyclin-dependent kinases. Trends Pharmacol Sci 23(9):417–425
9. Meijer L, Raymond E (2003) Roscovitine and other purines as kinase inhibitors. From starfish oocytes to clinical trials. Acc Chem Res 36(6):417–425
10. Johnson LN (2009) Protein kinase inhibitors: contributions from structure to clinical compounds. Q Rev Biophys 42(1):1–40
11. Tetsu O, McCormick F (2003) Proliferation of cancer cells despite cdk2 inhibition. Cancer Cell 3(3):233–245
12. Murray AW (2004) Revisiting the cell cycle: cyclins revisited. Cell 116:221–234
13. Echalier A, Endicott JA, Noble MEM (2010) Recent developments in cyclin-dependent kinase biochemical and structural studies. Biochim Biophys Acta Proteins Proteomics 1804(3):511–519
14. De Azevedo WF, Leclerc S, Meijer L, Havlicek L, Strnad M, Kim SH (1997) Inhibition of cyclin-dependent kinases by purine analogues—crystal structure of human cdk2 complexed with roscovitine. Eur J Biochem 243(1–2):518–526
15. De Azevedo WF, Mueller-Dieckman HJ, Schulze-Gahmen U, Worland PJ, Sausville EA, Kim SH (1996) Structural basis for specificity and potency of a flavonoid inhibitor of human cdk2, a cell cycle kinase. Proc Natl Acad Sci USA 93:2735–2740
16. Ducrot P, Legraverend M, Grierson DS (2000) 3d-qsar comfa on cyclin-dependent kinase inhibitors. J Med Chem 43:4098–4108
17. Otyepka M, Krystof V, Havlicek L, Siglerova V, Strnad M, Koca J (2000) Docking-based development of purine-like inhibitors of cyclin-dependent kinase-2. J Med Chem 43:2506–2513
18. Otyepka M, Kriz Z, Koca J (2002) Dynamics and binding modes of free cdk2 and its two complexes with inhibitors studied by computer simulations. J Biomol Struct Dyn 20(2):141–154
19. Vadivelan S, Sinha BN, Irudayam SJ, Jagarlapudi SARP (2007) Virtual screening studies to design potent cdk2-cyclin a inhibitors. J Chem Inf Model 47(4):1526–1535
20. Halperin I, Ma BY, Wolfson H, Nussinov R (2002) Principles of docking: an overview of search algorithms and a guide to scoring functions. Proteins Struct Funct Genet 47(4):409–443
21. Brooijmans N, Kuntz ID (2003) Molecular recognition and docking algorithms. Annu Rev Biophys Biomol Struct 32:335–373
22. Leach AR, Shoichet BK, Peishoff CE (2006) Prediction of protein-ligand interactions. Docking and scoring: successes and gaps. J Med Chem 49(20):5851–5855

23. Zhou T, Huang D, Caflisch A (2008) Is quantum mechanics necessary for predicting binding free energy? J Med Chem 51(14):4280–4288
24. Raha K, Merz KM (2004) A quantum mechanics-based scoring function: study of zinc ion-mediated ligand binding. J Am Chem Soc 126(4):1020–1021
25. Lu YX, Shi T, Wang Y, Yang HY, Yan XH, Luo XM, Jiang HL, Zhu WL (2009) Halogen bonding-a novel interaction for rational drug design? J Med Chem 52(9):2854–2862
26. Riley KE, Jurecka P, Pitonak M, Hobza P (2010) Stabilization and structure calculations for noncovalent interactions in extended molecular systems based on wave function and density functional theories. Chem Rev 110:5023–5063
27. Dobes P, Otyepka M, Strnad M, Hobza P (2006) Interaction energies for the purine inhibitor roscovitine with cyclin-dependent kinase 2: correlated ab initio quantum-chemical, dft and empirical calculations. Chem Eur J 12(16):4297–4304
28. Raha K, Peters MB, Wang B, Yu N, Wollacott AM, Westerhoff LM, Merz KM (2007) The role of quantum mechanics in structure-based drug design. Drug Dis Today 12(17–18):725–731
29. Zhou T, Huang DZ, Caflisch A (2010) Quantum mechanical methods for drug design. Curr Top Med Chem 10(1):33–45
30. Menikarachchi LC, Gascon JA (2010) Qm/mm approaches in medicinal chemistry research. Curr Top Med Chem 10(1):46–54
31. Cavalli A, Carloni P, Recanatini M (2006) Target-related applications of first principles quantum chemical methods in drug design. Chem Rev 106(9):3497–3519
32. Peters MB, Raha K, Merz KM (2006) Quantum mechanics in structure-based drug design. Curr Opin Drug Discovery Dev 9(3):370–379
33. Raha K, Merz KM (2005) Large-scale validation of a quantum mechanics based scoring function: predicting the binding affinity and the binding mode of a diverse set of protein-ligand complexes. J Med Chem 48(14):4558–4575
34. Khandelwal A, Lukacova V, Comez D, Kroll DM, Raha S, Balaz S (2005) A combination of docking, qm/mm methods, and md simulation for binding affinity estimation of metalloprotein ligands. J Med Chem 48(17):5437–5447
35. Merz KM (2010) Limits of free energy computation for protein-ligand interactions. J Chem Theory Comput 6(5):1769–1776
36. Wollacott AM, Merz KM (2007) Assessment of semiempirical quantum mechanical methods for the evaluation of protein structures. J Chem Theory Comput 3(4):1609–1619
37. Stewart JJP (2007) Optimization of parameters for semiempirical methods v: modification of nddo approximations and application to 70 elements. J Mol Model 13(12):1173–1213
38. Stewart JJP (2009) Application of the pm6 method to modeling proteins. J Mol Model 15(7):765–805
39. Rezac J, Fanfrlik J, Salahub D, Hobza P (2009) Semiempirical quantum chemical pm6 method augmented by dispersion and h-bonding correction terms reliably describes various types of noncovalent complexes. J Chem Theory Comput 5(7):1749–1760
40. Korth M, Pitonak M, Rezac J, Hobza P (2010) A transferable h-bonding correction for semiempirical quantum-chemical methods. J Chem Theory Comput 6(1):344–352
41. Hayik SA, Dunbrack R, Merz KM (2010) Mixed quantum mechanics/molecular mechanics scoring function to predict protein-ligand binding affinity. J Chem Theory Comput 6(10): 3079–3091
42. Fanfrlik J, Bronowska AK, Rezac J, Prenosil O, Konvalinka J, Hobza P (2010) A reliable docking/scoring scheme based on the semiempirical quantum mechanical pm6-dh2 method accurately covering dispersion and h-bonding: Hiv-1 protease with 22 ligands. J Phys Chem B 114:12666–12678
43. Jurecka P, Cerny J, Hobza P, Salahub DR (2007) Density functional theory augmented with an empirical dispersion term. Interaction energies and geometries of 80 noncovalent complexes compared with ab initio quantum mechanics calculations. J Comput Chem 28(2):555–569
44. Deng YQ, Roux B (2006) Calculation of standard binding free energies: aromatic molecules in the t4 lysozyme l99a mutant. J Chem Theory Comput 2(5):1255–1273
45. Deng YQ, Roux B (2009) Computations of standard binding free energies with molecular dynamics simulations. J Phys Chem B 113(8):2234–2246
46. Wang JY, Deng YQ, Roux B (2006) Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. Biophys J 91(8):2798–2814
47. Marenich AV, Cramer CJ, Truhlar DG (2009) Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. J Phys Chem B 113(18):6378–6396
48. Chang CEA, Chen W, Gilson MK (2007) Ligand configurational entropy and protein binding. Proc Natl Acad Sci USA 104(5):1534–1539
49. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz JKM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) A 2nd generation force-field for simulation of proteins, nucleic-acids and organic-molecules. J Am Chem Soc 117:5179–5197
50. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods RJ (2005) The amber biomolecular simulation programs. J Comput Chem 26(16):1668–1688
51. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong GM, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang JM, Kollman P (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J Comput Chem 24(16): 1999–2012
52. Wang JM, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general amber force field. J Comput Chem 25(9):1157–1174
53. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) Ucsf chimera—a visualization system for exploratory research and analysis. J Comput Chem 25(13):1605–1612
54. Klamt A, Schuurmann G (1993) Cosmo—a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. J Chem Soc Perkin Trans 2(5):799–805
55. Jayaram B, Sprous D, Beveridge DL (1998) Solvation free energy of biomacromolecules: parameters for a modified generalized born model consistent with the amber force field. J Phys Chem B 102(47):9571–9576
56. Frisch MJ (2009) Gaussian 09, revision a.02. Gaussian Inc., Wallingford
57. Tomasi J, Mennucci B, Cammi R (2005) Quantum mechanical continuum solvation models. Chem Rev 105(8):2999–3093
58. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S (2002) The protein kinase complement of the human genome. Science 298:1912–1934
59. Ewing TJA, Kuntz ID (1997) Critical evaluation of search algorithms for automated molecular docking and database screening. J Comput Chem 18(9):1175–1189
60. Otyepka M, Bartova I, Kriz Z, Koca J (2006) Different mechanisms of cdk5 and cdk2 activation as revealed by cdk5/p25 and cdk2/cyclin a dynamics. J Biol Chem 281:7271–7281
61. Ibrahim DA, El-Metwally AM (2010) Design, synthesis, and biological evaluation of novel pyrimidine derivatives as cdk2 inhibitors. Eur J Med Chem 45(3):1158–1166

62. Duca JS, Madison VS, Voigt JH (2008) Cross-docking of inhibitors into cdk2 structures. 1. J Chem Inf Model 48(3): 659–668
63. Sato H, Shewchuk LM, Tang J (2006) Prediction of multiple binding modes of the cdk2 inhibitors, anilinopyrazoles, using the automated docking programs gold, flexx, and ligandfit: an evaluation of performance. J Chem Inf Model 46(6):2552–2562
64. Guimaraes CRW, Cardozo M (2008) Mm-gb/sa rescoring of docking poses in structure-based lead optimization. J Chem Inf Model 48(5):958–970
65. Ferrara P, Curioni A, Vangrevelinghe E, Meyer T, Mordasini T, Andreoni W, Acklin P, Jacoby E (2006) New scoring functions for virtual screening from molecular dynamics simulations with a quantum-refined force-field (qrff-md). Application to cyclin-dependent kinase 2. J Chem Inf Model 46(1):254–263
66. Pearce BC, Langley DR, Kang J, Huang HW, Kulkarni A (2009) E-novo: an automated workflow for efficient structure-based lead optimization. J Chem Inf Model 49(7):1797–1809
67. Dessalew N, Singh SK (2008) 3d-qsar comfa and comsia study on benzodipyrazoles as cyclin dependent kinase 2 inhibitors. Med Chem 4(4):313–321
68. Heady L, Fernandez-Serra M, Mancera RL, Joyce S, Venkitaraman AR, Artacho E, Skylaris CK, Ciacchi LC, Payne MC (2006) Novel structural features of cdk inhibition revealed by an ab initio computational method combined with dynamic simulations. J Med Chem 49(17):5141–5153
69. Hobza P, Muller-Dethlefs K (2010) Non-covalent interactions. RSC Publishing, Cambridge
70. Zgarbova M, Otyepka M, Sponer J, Hobza P, Jurecka P (2010) Large-scale compensation of errors in pairwise-additive empirical force fields: comparison of amber intermolecular terms with rigorous dft-sapt calculations. Phys Chem Chem Phys 12: 9611–9614
71. Alzate-Morales JH, Caballero J, Jague AV, Nilo FDG (2009) Insights into the structural basis of n2 and o6 substituted guanine derivatives as cyclin-dependent kinase 2 (cdk2) inhibitors: prediction of the binding modes and potency of the inhibitors by docking and oniom calculations. J Chem Inf Model 49(4): 886–899
72. Zhou T, Caflisch A (2010) High-throughput virtual screening using quantum mechanical probes: discovery of selective kinase inhibitors. Chem Med Chem 5(7):1007–1014
73. Lawrie AM, Noble MEM, Tunnah P, Brown NR, Johnson LN, Endicott JA (1997) Protein kinase inhibition by staurosporine revealed in details of the molecular interaction with cdk2. Nat Struct Biol 4:796–801
74. Zhao B, Bower MJ, McDevitt PJ, Zhao HZ, Davis ST, Johanson KO, Green SM, Concha NO, Zhou BBS (2002) Structural basis for chk1 inhibition by ucn-01. J Biol Chem 277(48):46609–46615
75. Arris CE, Boyle TF, Calvert AH, Curtin NJ, Endicott JA, Garman EF, Gibson AE, Golding BT, Grant S, Griffin RJ, Jewsbury P, Johnson LN, Lawrie AM, Newell DR, Noble MEM, Sausville EA, Schultz R, Yu W (2000) Identification of novel purine and pyrimidine cyclin-dependent kinase inhibitors with distinct molecular interactions and tumor cell growth inhibition profiles. J Med Chem 43:2797–2804
76. Wu SY, McNae I, Kontopidis G, McClue SJ, McInnes C, Stewart KJ, Wang SD, Zheleva DI, Marriage H, Lane DP, Taylor P, Fischer PM, Walkinshaw MD (2003) Discovery of a novel family of cdk inhibitors with the program lidaeus: structural basis for ligand-induced disordering of the activation loop. Structure 11(4):399–410
77. Wang SD, Meades C, Wood G, Osnowski A, Anderson S, Yuill R, Thomas M, Mezna M, Jackson W, Midgley C, Griffiths G, Fleming I, Green S, McNae I, Wu SY, McInnes C, Zheleva D, Walkinshaw MD, Fischer PM (2004) 2-anilino-4-(thiazol-5-yl)pyrimidine cdk inhibitors: synthesis, sar analysis, x-ray crystallography, and biological activity. J Med Chem 47(7):1662–1675
78. Chu XJ, DePinto W, Bartkovitz D, So SS, Vu BT, Packman K, Lukacs C, Ding QJ, Jiang N, Wang K, Goelzer P, Yin XF, Smith MA, Higgins BX, Chen YS, Xiang Q, Moliterni J, Kaplan G, Graves B, Lovey A, Fotouhi N (2006) Discovery of [4-amino-2-(1-methanesulfonylpiperidin-4-ylamino) pyrimidin-5-yl](2, 3-difluoro-6-methoxyphenyl)methanone (r547), a potent and selective cyclin-dependent kinase inhibitor with significant in vivo antitumor activity. J Med Chem 49(22):6549–6560
79. Davies TG, Bentley J, Arris CE, Boyle TF, Curtin NJ, Endicott JA, Gibson AE, Golding BT, Griffin RJ, Hardcastle IR, Jewsbury P, Johnson LN, Mesguiche V, Newell DR, Noble MEM, Tucker JA, Wang L, Whitfield HJ (2002) Structure-based design of a potent purine-based cyclin-dependent kinase inhibitor. Nat Struct Biol 9(10):745–749
80. Sayle KL, Bentley J, Boyle FT, Calvert AH, Cheng YZ, Curtin NJ, Endicott JA, Golding BT, Hardcastle IR, Jewsbury P, Mesguiche V, Newell DR, Noble MEM, Parsons RJ, Pratt DJ, Wang LZ, Griffin RJ (2003) Structure-based design of 2-arylamino-4-cyclohexylmethyl-5-nitroso-6-aminopyrimidine inhibitors of cyclin-dependent kinases 1 and 2. Bioorg Med Chem Lett 13(18):3079–3082
81. Meijer L, Bisagni E, Legraverend M (1998) Purine derivatives having, in particular, antiproliferative properties, and their biological uses USA Patent
82. Schulze-Gahmen U, Brandsen J, Jones HD, Morgan D, Meijer L, Veselý J, Kim S-H (1995) Multiple modes of ligand recognition: crystal structures of cyclin-dependent protein kinase 2 in complex with atp and two inhibitors, olomoucine and isopenthenyladenine. Proteins Struct Funct Genet 22:378–391
83. Veselý J, Havlíček L, Strnad M, Blow JJ, Donella-Deana A, Pinna L, Letham DS, Kato J-Y, Detivaud L, Leclerc S, Meijer L (1994) Inhibition of cyclin dependent kinases by purine analogues. Eur J Biochem 224:771–786
84. McIntyre NA, McInnes C, Griffiths G, Barnett AL, Kontopidis G, Slawin AMZ, Jackson W, Thomas M, Zheleva DI, Wang SD, Blake DG, Westwood NJ, Fischer PM (2010) Design, synthesis, and evaluation of 2-methyl- and 2-amino-n-aryl-4, 5-dihydrothiazolo[4, 5-h]quinazolin-8-amines as ring-constrained 2-anilino-4-(thiazol-5-yl)pyrimidine cyclin-dependent kinase inhibitors. J Med Chem 53(5):2136–2145

The Semiempirical QM Method PM6-DH2X Describes the Geometry and Energetics of CK2-Inhibitor Complexes Involving Halogen Bonds Well while the Empirical Potential Fails

Journal:	*The Journal of Physical Chemistry*
Manuscript ID:	jp-2011-02149z
Manuscript Type:	Article
Date Submitted by the Author:	07-Mar-2011
Complete List of Authors:	Hobza, Pavel; Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic Fanfrlik, Jindrich; UOCHB Řezáč, Jan; Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Center for Biomolecules and Complex Molecular Systems Otyepka, Michal; Palacky University Olomouc, Physical Chemistry Dobes, Petr; UOCHB

The Semiempirical QM Method PM6-DH2X Describes the Geometry and Energetics of CK2-Inhibitor Complexes Involving Halogen Bonds Well while the Empirical Potential Fails

Petr Dobeš,[1,3] Jan Řezáč,[1] Jindřich Fanfrlík,[1] Michal Otyepka,[2*] Pavel Hobza[1,2*]

[1]Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic and Center for Biomolecules and Complex Molecular Systems, 166 10 Prague, Czech Republic

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

[3]Center of Molecular Biology and Gene Therapy, Department of Internal Medicine – Hematooncology, University Hospital Brno, 625 00 Brno, Czech Republic

pavel.hobza@uochb.cas.cz; tel.: +420 220 410 311, fax: +420 220 410 320

Abstract

In the present study, we have investigated complexes of CK2 protein kinase with halogenated inhibitors by means of the advanced semiempirical quantum mechanical (SQM) PM6 method (called PM6-DH2X), which describes various types of non-covalent interactions including halogen bonding well. The PM6-DH2X method provides reliable geometries of those CK2 protein kinase-inhibitor complexes involving halogen bonds that agree well with the X-ray crystal structures. When the Amber empirical potential is applied, this agreement becomes considerably worse. Similarly, the binding free energies determined by the PM6-DH2X SQM method are much closer to the experimental inhibition constants than those based on the Amber empirical potential.

Introduction

Protein kinase CK2 (formerly known also as casein kinase-2) is a pleiotropic Ser/Thr protein kinase with hundreds of regulation targets, which have a variety of cellular functions. CK2

plays an important role in gene-expression regulation, the synthesis and degradation of proteins as well as the signaling and suppression of apoptosis[1,2]. CK2 is abnormally highly active in many tumor cells[3], thus the attenuation of CK2 activity can be a strategy for the treatment of different neoplastic diseases[4,5]. CK2 is composed of two catalytic subunits (alpha) and two regulatory subunits (beta)[6]. So far, a number of CK2 inhibitors have been described in micromolar and submicromolar ranges[7-9]. Halogenated benzoimidazoles belong to a widely and successfully tested family of CK2 inhibitors competing with ATP. There are a number of crystal structures available[9-11]. Of these, eight crystal structures contain aromatic tetrabromo-derivatives and one a tetraiodobenzimidazole inhibitor (cf. Table 1A)[11]. The inhibitors are bound to a CK2 α-catalytic subunit, specifically to a small hydrophobic cavity of the ATP binding site (Figure 1). The binding mode could explain the fairly narrow selectivity of those CK2 inhibitors. Further analysis highlights the major role of the hydrophobic effect for increasing potency within this inhibitor class[7,12,13]. It also shows that polar interactions are responsible more so for the orientation in the active site and, further, that the respective binding arises from the existence of two halogen bonds of the O...Br type[12]. Recently, it has been shown that novel tetraiodinated benzimidazoles[8] are more powerful inhibitors of CK2 than their tetrabrominated analogs[11]. Evidently, this class of CK2 inhibitors represents an ideal target for the study of halogen bonding in protein-ligand complexes.

A recently published survey of protein and nucleic acid structures has revealed the halogen bond as a stabilizing intermolecular interaction[14]. Lately, some papers have even been aimed at using halogen bonding in rational drug design[15-21], in crystal engineering[22,23] and directing macromolecular conformation[24]. The strength of the halogen bonding depends on its substitution[25] and solvent effects[26]. It has been shown that the origin of halogen bonding is in an electrostatic attraction between the positive σ-hole on the tip of the halogen atom (opposite the covalent bond between the halogen and carbon atoms) and the negative Lewis

base[27-29] and that it has a directional property similar to the hydrogen bond[20,27]. Moreover, the positive σ-hole is largest for the iodine atom and is not present for a fluorine atom, i.e. fluorinated analogs cannot make halogen bonds[23]. It should be noted that the interaction energy of a halogen bond can be comparable with the interaction energy of a hydrogen bond[30,31]; therefore, halogen bonds are very appealing in the process of rational drug design[18]. Recently, a halogen bond has been successfully exploited in the design of a selective CDK9 inhibitor[32]. The opening of the σ-hole is clearly a quantum effect and, consequently, halogen bonding cannot be described by current classical force fields using atom-centered partial charges. In addition, the description of the halogen bond might also be challenging for scoring functions, which are based on principles similar to empirical force fields. Without any doubt, to investigate the interaction of the inhibitor with a protein, where halogen bonds matter, it is necessary to employ quantum mechanical (QM) methods. Recently, we have developed a semiempirical QM (SQM) method (PM6-DH2)[33,34], which performs well in the description of various types of non-covalent complexes (see Methods for details). The PM6-DH2 method has also been successfully applied for scoring the inhibitors of the HIV protease[35] and CDK2 kinase[36].

The aim of the present study is to investigate complexes of CK2 protein kinase with different inhibitors, each having several halogens engaged in the halogen bonding. We will demonstrate that force-field methods completely fail to describe the structure as well as binding free energy of these complexes while the SQM method provides geometries in good agreement with the X-ray structures. Similarly, the scoring function based on the SQM method yields a good correlation between the binding energy and inhibition activity.

Methods

The formation of the protein (P)-inhibitor (I) complex from free (hydrated) subsystems is represented by the following equation:

$$P_w + I_w \rightarrow P\text{–}I_w \qquad (1)$$

The binding of a competitive inhibitor is expressed by an inhibition constant (K_i), which is related to the change of the free energy of binding (binding free energy) as follows:

$$\Delta G_w = RT \ln(K_i). \qquad (2)$$

The binding free energy was approximated by the total score ($\Delta G'_w$), expressed by the following equation (for details, see our previous paper)[35].

$$\Delta G'_w = \Delta H_w - T\Delta S_w + \Delta E_{def}(I) + \Delta\Delta G_w(I) + \Delta E_{def}(P) + \Delta\Delta G_w(P) \qquad (3)$$

The single terms designate the interaction enthalpy, interaction entropy, and deformation energy as well as the correction for the desolvation free energy of the inhibitor and protein, respectively (see below). The majority of the scoring functions use empirical potentials, also called molecular mechanics (MM) methods. The major drawback of all of the classical empirical potentials is their neglect of quantum effects. The most important of them are the polarization effects, proton and electron transfer and halogen bonding. All of these quantum effects are satisfactorily described by QM methods. The use of non-empirical QM methods is, however, limited to systems having no more than several hundred atoms. The P-I complexes are mostly much larger, with the smallest ones possessing several thousand atoms. The first possibility of how to solve the problem is to use methods like the *ab initio* fragment MO (FMO) method[37]. Two limitations of the FMO method should be taken into account, i.e. the applicability to charged protein – charged inhibitor complexes and the inability to perform geometry optimizations. The second possibility is the use of a hybrid QM/MM model[38]. When, however, the QM part is limited to a few hundred atoms, some artifacts arising from the communication between the layers can occur. The third possibility is the application of SQM methods[39]. The use of standard SQM methods is not straightforward owing to their poor treatment of dispersion energy and hydrogen bonding[40]. This problem has recently been solved in our laboratory by using Stewart's PM6[41] SQM method corrected for dispersion and

hydrogen bonding[33,34]. The developed PM6-DH2 method provides accurate results for various types of noncovalent complexes including hydrogen-bonded and dispersion-controlled complexes, which arises from the inclusion of corrections to the dispersion and hydrogen-bonding energies[33,34]. The PM6-DH2 method overestimates, however, the strength of the halogen bond and provides distances between halogen and electronegative atoms that are too short. This is because of an insufficient repulsion in the PM6. We have addressed this problem in a recent letter[42] by an additional repulsion term fitted specifically to halogen bonding. The corrected method, abbreviated as PM6-DH2X, is able to describe halogen bond with accuracy close to the high-level QM methods. For more details, see ref. [42].

The individual terms in $\Delta G'_{\mathrm{W}}$ (eq. 3) represent the interaction enthalpy (ΔH_{w}), the interaction-entropy (as $-T\Delta S_{\mathrm{w}}$), the inhibitor and the protein deformation energy $\Delta E_{\mathrm{def}} = E(\mathrm{I})_{\mathrm{w}}^{\mathrm{PI}} - E(\mathrm{I})_{\mathrm{w}}^{\mathrm{I}}$ (analogously for the protein P; the upper indexes PI and I standing for the geometry of the inhibitor (protein) in a protein/inhibitor complex and the relaxed inhibitor (protein) geometry in water, respectively) as well as the corrections for the inhibitor (protein) hydration free energy $\Delta\Delta G_{\mathrm{w}}(\mathrm{I}) = \Delta G_{\mathrm{W}}^{\mathrm{MOPAC}}(\mathrm{I})^{\mathrm{PI}} - \Delta G_{\mathrm{W}}^{\mathrm{G09}}(\mathrm{I})^{\mathrm{PI}}$ (analogously for protein P; the lower index w stands for a water environment). The last term reflects the fact that the structure of the PI complex (in a water environment) is determined using the hydration model considering the electrostatic term only (COSMO in the MOPAC code) while an accurate treatment requires the use of a hydration model considering the non-electrostatic term as well. Non-electrostatic terms associated with the formation of a cavity to accommodate the solute, and the van der Waals interaction between solute and solvent, are neglected in the former approach. This approximation might work well for inhibitors of a similar shape and the same charge, but in the case of entirely different inhibitors it can cause a serious error. For this reason, we have adopted a more accurate calculation of the complete solvation free energy of the ligand with a COSMO[43] model based on B3LYP calculations (Gaussian 09 code[44]).

The formation of a PI complex restricted the motions of the ligand as well as the torsional motion of the protein. Furthermore, the vibrational entropy of the ligand was also restricted by the formation of a PI complex. The entropic contributions were determined using the rigid rotor/harmonic oscillator approximation based on a Cornell et al.[45] AMBER force field. The structures of all of the systems were re-optimized using the empirical potential considering the continuous water (generalized Born model), and the same method was used for the calculation of the second derivatives of the total energy.

Structures preparation

Two sets of structural data were considered with appropriate inhibition constants. The first set of complexes (Figure 2A) involves nine crystal structures (eight tetrabromo-derivatives and one tetraiodobenzimidazole) of a maize CK2α subunit with an inhibitor for which both structural data as well as inhibition constants are available (Table 1A). The structures of all of the complexes were taken from the RSCB Protein Data Bank. When two CK2α units were present, only chain A was further considered. The hydrogen atoms were added to all of the structures, and their positions were minimized using AMBER[46] – ff03[45] and gaff[47] force fields in Chimera software[48]. All of the crystal waters were removed except for those in the 1ZOE complex, where the Cl- ion is located in an active site near an inhibitor. Consequently, nine waters located up to 5 Å from this inhibitor were considered; these waters are important for the shielding of the Cl- ion. Generally, two different conformations of an inhibitor exist; the only exception is the 1J91. Those conformations differ in rotation by about sixty degrees around the axis perpendicular to the plane of the inhibitors. In the first conformation (1ZOG, 1ZOH, 3KXG, 3KXN), Br5 and Br6 interact with an oxygen of Glu114 and Val116 through halogen bonds and in the second conformation (1ZOE, 2OXD, 2OXX, 2OXY) Br4 and Br5 interact with an oxygen of Glu114 and Val116 through halogen bonds. More specifically in the case of 1ZOH and

3KXN complexes, the inhibitor is found in two positions. In the present study, we have systematically considered only one position, which belongs to the first conformation. All of the inhibitors investigated were assumed to be electroneutral. The positions of the hydrogen atoms in the 1ZOE complex having nine waters connected to the Cl^- ion were optimized by means of PM6 (in the MOPAC code[49]) with the aim of obtaining an H-bonded network of waters, ion, CK2 and the inhibitor. The resulting structure as well as all other structures were further fully reoptimized using the PM6-D2X method (see later).

The second set was created by using a structural alignment, where one human CK2α subunit was taken from the 1JWH (a CK2complex with ANP, which is an ATP analog) and the structure of inhibitor was taken from the maize CK2α subunit (1ZOE). In the model structures, the inhibitors were modified according to their chemical structure to prepare complexes (Figure 2B) with known inhibition constants (Table 1B)[8]. Like in the previous case, hydrogen atoms were added to all of the structures considered, and their positions were minimized using AMBER[46] – ff03[50] and gaff[47] force fields in Chimera software[48]. All of the inhibitors in this group were considered to be electroneutral.

Strategy of calculations

The structures of all of the protein-inhibitor complexes were systematically optimized by the PM6-D2X method in a continuum COSMO solvent model as implemented in the MOPAC code[49] with the following optimization criteria (ΔE = 0.0300 kcal/mol, maxGrad =6.0 kcal/mol/Å and rmsGrad = 3 kcal/mol/Å). Note that the hydrogen bonding correction is not used for the optimization, because our model calculations suggest that the most accurate interaction energies are obtained using the H-bond-corrected PM6-DH2X energies on geometries obtained without the correction (PM6-D2X). The respective interaction enthalpies ($\Delta H_w = \Delta H_w(PI) - (\Delta H_w(P) + \Delta H_w(I))$) were determined using these optimized structures. In the second step, all of the complexes were reoptimized with AMBER ff99[51] and gaff force

fields[47]. The RESP charges[52] were computed at the HF/6-31G* level; in the case of iodine, the HF/SDD[53] level with the respective pseudopotentials as implemented in Gaussian 09 was considered. In all of the cases, a generalized Born solvent model in the Nucleic Acid Builder (NAB, from the AMBER[46] package) was utilized. The entropy term (T=298 K, p = 1 atm.) was determined using the ideal gas, rigid-rotor harmonic-oscillator approximation using the empirical potential; all of the geometries were reoptimized with the Newton-Raphson method and the L-BFGS TNCG algorithm to a gradient of 10^{-12} kcal/mol/Å. The interaction enthalpies (ΔH_w^{AMBER}) were also determined with the same empirical force field (T = 298 K, p = 1 atm.) and the generalized Born solvent model.

The inhibitor deformation energy (ΔE_{def}(I)) was calculated as the difference between the energy of the inhibitor in the geometry taken from the protein/inhibitor complex ($E(I)^{PI}$) and the energy of the fully optimized inhibitor in water ($E(I)^{w}$). It should be mentioned that the inhibitors investigated in the present paper are relatively rigid when compared to the floppy inhibitors considered in our previous studies[35,36]. For a further discussion of the inclusion of the deformation and desolvation energies of the ligand, see also another of our recent papers[54]. The PM6-D2X methods with the COSMO solvation model were used for these calculations. The correction for the inhibitor deformation energy (ΔE_{def}^{AMBER}(I)) was also calculated at the empirical level using AMBER force fields with a generalized Born solvent model.

The correction of the inhibitor desolvation free energy $\Delta\Delta G_w$(I) (T=298 K, p=1 atm.) was determined for an inhibitor structure taken from the optimized CK2-inhibitor complex. It was evaluated as the difference between the solvation free energies calculated by the PM6 /COSMO model ($\Delta G_W^{MOPAC}(I)^{PI}$), where only the electrostatic terms were taken into consideration, and the COSMO model (C-PCM from B3LYP/6-31G* wave function, in the case of iodine the B3LYP/SDD level with the respective pseudopotentials as implemented in

Gaussian 09 was taken into account), where also all of the non-electrostatic terms were considered.

The deformation energy $\Delta E_{def}(P)$ and correction for the desolvation free energy $\Delta\Delta G_w(P)$ of the protein were not considered since it is expected that these terms are changed only marginally for the structurally very similar inhibitors investigated in the present study.

Results and Discussion

Structural comparisons. Analyzing the structures of all of the complexes, we have found two basic positions of inhibitors with respect to the CK2 active site: i) as in 1ZOG, 1ZOH, 3KXG, 3KXN and ii) as in 1ZOE, 2OXD, 2OXX, 2OXY with the exception of 1J91. Tables 2A and 2B show a comparison of the geometrical parameters of the experimental (or model) structures with the optimized ones for all of the complexes of the first and second group when the PM6-D2X and AMBER methods were utilized. The PM6-D2X method provides much better RMSD of the Cα atoms from the X-ray than the AMBER force field. Specifically, the averaged RMSD for the nine complexes optimized by the PM6-D2X method of the first group amounts to 0.14 Å, when the AMBER method is used the RMSD value increases considerably to 1.14 Å. The halogen....oxygen distances in both of the halogen bonds between the inhibitor and the CK2 active site determined by the PM6-D2X agree well with the experimental data (Table 2A). On the other hand, when the AMBER force field is used, the optimized distances are 1–2 Å larger (Table 2A). The most apparent difference, found for the 2OXY complex, is shown in Figure 5. Again, the PM6-D2X optimized structure agrees well with the experimental X-ray structures while the AMBER-optimized structure differs considerably. It is evident that the inhibitor is shifted out of the cavity region in the AMBER optimized structure. The same applies for all of the inhibitors involved in the study.

All of these results clearly demonstrate that the SQM method reproduces the structure and geometry of CK2-inhibitor complexes well including the geometry of halogen bonds while the MM method fails. MM fails because of its inability to describe the halogen bond between the inhibitor and CK2 kinase. The reason is that the AMBER force field is not capable of describing the σ-hole on the halogen atom, because the halogen atom is represented as a homogenous Lennard-Jonnes sphere bearing a usually negative partial charge. Consequently, the interaction between the halogen atom and halogen-bond acceptor (also bearing a negative partial charge) is repulsive in the MM calculation. This fact explains why the inhibitors are shifted out of the cavity region in the MM optimized structures.

Binding free energy. In the present study, two sets of CK2 inhibitors with known inhibition constants (K_i) were considered. The X-ray structures of the maize CK2 were available for the first set while the model structures were derived for the second set. The primary sequence similarity between maize and human (or rat) CK2 is about 70%. Specifically, the ATP binding site is almost identical. The available inhibition constants from rat CK2 (Table 1B) were used for the maize CK2 structures in the case of the first group based on crystal data[55]. On the contrary, the second group created from a model based on human CK2 structures used inhibition constants measured on rat CK2[8]. We thus investigated the correlation between the experimental and computed binding free energy (inhibition constants) for the nine CK2-inhibitor complexes from the first group (*cf.* Figure 2A). In addition, the correlation between the experimental and computed binding free energy for the fourteen CK2-inhibitor complexes from the second group (*cf.* Figure 2B) was also evaluated. In both cases, the energy terms were calculated on fully optimized structures. This is important, because it makes the method independent of an exact knowledge of the experimental geometry.

Tables 1A and 1B summarize the single terms from the total score for both of the inhibitor groups calculated by the SMQ method. Figure 3A shows a tight correlation (R^2=0.86) between the PM6-DH2X interaction enthalpy (ΔH_w) and the experimental inhibition constants (ln K_i) for the first group of complexes (with known crystal structures). The results of the analogous calculations performed for the second group of complexes (where model structures were generated) are shown in Figure 3B, but the correlation is considerably worse (R^2=0.52). Despite the fact that the interaction enthalpies were calculated in a water environment, several important contributions to the total score (which estimates the free energy of binding) are still missing. Adding inhibitor deformation ($\Delta E_{def}(I)$) and correction for desolvation free energies ($\Delta\Delta G_w(I)$) to the interaction enthalpy does not change the correlation of the first group (Figure 3C) significantly (R^2 = 0.81) while it considerably improves the correlation of the second group (R^2=0.71, Figure 3D). When, however, the total score including also the entropy term is considered, the correlations for both of the groups of complexes (*cf.* Figures 3E and 3F) strongly deteriorate (R^2 = 0.24 and 0.19, respectively). It is worth noting that the entropy term is evaluated by the empirical potential (on structures minimized by the empirical potential), because the calculation of the second derivatives by the SQM method employed is still impractical for such large systems. The weak correlation is explained by the usage of the empirical potential for the evaluation of the entropy term, which fails in its description of halogen bonding, essential for the binding of the considered inhibitors to CK2. In other words, the empirical force field is not able to describe halogen bonding, as a result of which the optimized structures of inhibitor-CK2 complexes involving halogen bonds are incorrect, as shown in the previous paragraph.

For the sake of completeness, we have also analyzed the correlation of the inhibition constants (ln K_i) with the energy terms calculated consistently by the AMBER force field

on the AMBER optimized structures. For the first (Figure 4C, Table 1A) as well as the second group (Figure 4D, Table 1B), no statistically significant correlation between the inhibition constant and the total score was found (R^2 = 0.04 and 0.08, respectively). The same applies for the correlation of the interaction enthalpy with the inhibition constant (Figures 4A and 4B). The empirical potential fails in ranking the CK2 inhibitors involving halogen bond in binding. This finding is alarming, because the majority of the empirical scoring functions used in docking are based on the same principles as the AMBER empirical potential. The failure of the AMBER empirical potential in the geometry and energy of the protein-inhibitor complexes involving a halogen bond warns that the docking and scoring of the inhibitors bearing potential halogen-bond donors (Cl, Br, I atoms typically on the aromatic rings or neighboring the strong electron-acceptor groups) cannot be blind and the results should be carefully analyzed while keeping in mind the Achilles' heel of the empirical potential in the description of the halogen bond.

Conclusions

i) The PM6-D2X optimized structures of the eight tetrabromo-benzimidazoles and one tetraiodo-benzimidazole in a CK2α active site agreed well with the X-ray data, with the RMSD of the Cα atoms being equal to 0.14 Å. When the AMBER empirical potential was applied, the agreement became considerably worse with an average RMSD of 1.14 Å. The experimental distances between the halogen and oxygen in the halogen bonds were very well reproduced by the PM6-D2X SQM method (with the average difference being below 0.2 Å) while the AMBER force field showed significant differences (with the average difference being greater than 1.0 Å).

ii) The correlation between the PM6-DH2X interaction enthalpies and the inhibition constants for the CK2-inhibitor complexes for which experimental structural data exist

was high (R^2=0.86), and it was lower (R^2=0.52) for the modeled CK2-inhibitor complexes. When the interaction enthalpy was augmented by the inhibitor deformation energy and corrections for the inhibitor desolvation free energy, the correlation was practically the same in the first set of inhibitors (from r^2=0.86 to 0.81) and was significantly improved in the second set (from R^2=0.52 to 0.71). Adding the entropy term, which was evaluated at the empirical level, deteriorated the correlation strongly (to R^2=0.24 and 0.19, for the first and the second set, respectively). This shows that the entropy term determined for the AMBER optimized structures cannot be used in the construction of a total scoring function. The systematic use of the AMBER force field instead of the PM6-D2X SQM method does not lead to any improvement.

iii) The AMBER force field fails to describe halogen bonding and, consequently, also the structures of the CK2-inhibitor complexes. The opposite is true for the PM6-D2X SQM method. The AMBER force field also fails to predict the interaction enthalpies or the total score for the CK2-inhibitor complexes and, again, the PM6-DH2X method provides quite reliable values.

Acknowledgements

This work was a part of Research Project No. Z40550506 of the Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic and was funded by Grants No. LC512 and MSM6198959216 from the Ministry of Education, Youth and Sports of the Czech Republic. This work was also supported by grant GACR P208/11/0295 and by the Operational Program Research and Development for Innovations – European Social Fund (CZ.1.05/2.1.00/03.0058). The support of Praemium Academiae, Academy of Sciences of the Czech Republic, awarded to P.H. in 2007 is also acknowledged.

Table 1A The summary of the experimental data and calculated terms for the first set of CK2 inhibitors (*cf.* Figure 2A).[a]

System	K_i (μM)	ln K_i (M)	Res. (Å)	PM6-DH2X ΔH_w	PM6-DH2X $\Delta\Delta G_w(I)$	PM6-DH2X $\Delta E_{def}(I)$	PM6-DH2X $T\Delta S_w$	PM6-DH2X $\Delta G'_w$	AMBER ΔH_w	AMBER ΔE_{def}(I)	AMBER $T\Delta S_w$	AMBER $\Delta G'_w$
1J91[10]	0.400[12]	-14.73	2.22	-15.13	-4.70	0.67	-15.67	-3.49	-24.81	0.27	-15.67	-8.87
1ZOE[12]	0.045[12]	-16.92	1.77	-38.09	-7.16	0.87	-19.53	-24.85	-31.76	1.53	-19.53	-10.70
1ZOG[12]	0.070[12]	-16.47	2.30	-28.81	-6.80	1.04	-18.83	-15.74	-23.15	1.31	-18.83	-3.02
1ZOH[12]	0.100[12]	-16.12	1.81	-24.70	-7.13	2.58	-27.95	-1.30	-42.83	0.77	-27.95	-14.11
2OXD[56]	0.150[56]	-15.71	2.30	-22.29	-5.25	0.83	-10.62	-16.09	-8.24	0.68	-10.62	3.06
2OXX[56]	0.200[56]	-15.42	2.30	-16.68	-10.43	0.93	-3.09	-23.09	-18.99	0.32	-3.09	-15.58
2OXY[56]	0.300[56]	-15.02	1.81	-20.87	-5.87	0.17	1.31	-27.88	-14.15	0.14	1.31	-15.32
3KXG[11]	0.048[50]	-16.85	1.70	-30.50	-4.48	0.47	-1.87	-32.64	-11.33	0.31	-1.87	-9.15
3KXN[11]	0.023[8]	-17.59	2.00	-34.97	-4.90	1.02	-6.53	-32.32	-14.27	0.51	-6.53	-7.22

[a] The K_i inhibition constant (in μM); Res. is the experimental mean resolution of the respective X-ray structure (in Å); ΔH_w the interaction enthalpy; $\Delta\Delta G_w(I)$ the correction for the ligand desolvation free energy; ΔE_{def}(I) the inhibitor deformation energy; $T\Delta S_w$ the entropic contribution; $\Delta G'_w$, the total score (all of the energies are in kcal/mol)

Table 1B The summary of the experimental data and calculated terms for the second set of CK2 inhibitors (*cf.* Figure 2B).[a]

System	K_i μM	ln K_i M	PM6-DH2X ΔH_w	PM6-DH2X $\Delta\Delta G_w(I)$	PM6-DH2X $\Delta E_{def}(I)$	PM6-DH2X $T\Delta S_w$	PM6-DH2X $\Delta G'_w$	AMBER ΔH_w	AMBER ΔE_{def}(I)	AMBER $T\Delta S_w$	AMBER $\Delta G'_w$
10a	0.120	-15.94	-17.56	-4.87	0.42	-14.24	-7.77	-35.67	0.32	-14.24	-21.12
12a	0.120	-15.94	-21.99	-5.70	1.21	-9.14	-17.34	-27.12	0.17	-9.14	-17.81
13a	0.090	-16.22	-22.76	-5.99	0.25	-11.52	-16.97	-29.35	0.28	-11.52	-17.55
15	0.050	-16.81	-18.88	-9.21	0.89	-13.28	-13.92	-28.39	0.20	-13.28	-14.91
17	0.050	-16.81	-24.85	-4.85	0.27	-9.71	-19.70	-24.56	1.17	-9.71	-13.68
7a	0.024	-17.55	-24.40	-5.61	0.22	-19.42	-10.38	-33.87	0.30	-19.42	-14.15
K10	0.370	-14.81	-17.21	-4.23	0.11	-13.59	-7.75	-14.47	0.37	-13.59	-0.50
K20	0.370	-14.81	-16.59	-5.49	0.14	-13.43	-8.51	-25.85	0.17	-13.43	-12.24
K21	0.250	-15.20	-17.74	-5.11	1.82	-10.54	-10.48	-26.22	0.10	-10.54	-15.57
K22	0.200	-15.42	-16.50	-10.64	0.46	-15.98	-10.70	-39.35	0.14	-15.98	-23.23
K32	0.180	-15.53	-20.57	-5.10	0.16	-11.38	-14.13	-25.18	0.91	-11.38	-12.89
1a	0.023	-17.59	-24.73	-4.82	0.25	-11.81	-17.49	-28.45	0.12	-11.81	-16.53
4a	0.460	-14.59	-21.21	-2.57	0.11	-10.28	-13.39	-24.91	0.60	-10.28	-14.04
8a	0.330	-14.92	-18.93	-2.55	0.12	-6.58	-14.79	-20.94	0.77	-6.58	-13.59

[a] The K_i inhibition constant (in μM); ΔH_w the interaction enthalpy; $\Delta\Delta G_w(I)$ the correction for the ligand desolvation free energy; ΔE_{def}(I) the inhibitor deformation energy; $T\Delta S_w$ the entropic contribution; $\Delta G'_w$, the total score (all of the energies are in kcal/mol)

Table 2A A comparison of the structural X-ray data with the PM6-D2X- and AMBER-optimized geometries for CK2-inhibitor complexes (the first set, *cf.* Figure 2A).[a]

	RMSD		Halogen bonds					
System	PM6-D2X	AMBER	X-ray		PM6-D2X		AMBER	
			Arg47 N	Val44 O	Arg47 N	Val44 O	Arg47 N	Val47 O
1J91	0.10	1.30	2.99	3.99	3.12	4.2	4.41	3.58
			Glu114 O	Val116 O	Glu114 O	Val116 O	Glu114 O	Val116 O
1ZOE	0.11	1.12	3.43	3.24	3.41	3.22	4.33	4.18
1ZOG	0.14	1.30	3.23	2.98	3.21	3.06	4.77	4.04
1ZOH	0.14	1.33	3.38	3.23	3.40	3.05	3.61	3.63
2OXD	0.31	0.98	3.27	2.80	3.52	2.98	4.73	4.56
2OXX	0.11	0.87	3.73	2.72	3.84	2.86	4.25	3.79
2OXY	0.10	1.24	3.16	2.95	3.12	3.04	4.91	5.09
3KXG	0.13	1.03	3.57	2.85	3.24	2.96	4.26	3.84
3KXN	0.12	1.13	3.34	3.17	3.28	3.06	4.73	3.6

[a] The RMSD of the Cα atoms of the PM6-D2X- and AMBER-optimized structures from the X-ray structure; the fourth and the fifth columns show the distances between the oxygens of the CK2 amino acids and the halogen atoms of the inhibitor as found in the respective crystal structure, whereas the sixth and the seventh columns and the eighth and ninth columns show the same distances after the PM6-D2X and AMBER optimization (all of the distances are in Å).

Table 2B A comparison of the structural data of the PM6-D2X- and AMBER-optimized CK2-inhibitor complexes (the second set, *cf.* Figure 2B).[a]

	Halogen bonds			
System	PM6-D2X		AMBER	
	Glu114 O	Val116 O	Glu114 O	Val116 O
10a	3.06	2.89	4.47	4.06
12a	3.02	2.94	5.41	5.47
13a	3.01	2.96	5.4	5.19
15	3.03	2.94	5.25	5.25
17	3.01	2.97	4.94	3.76
7a	3.05	3.00	5.52	5.34
K10	3.06	3.07	5.75	5.8
K20	3.06	3.05	5.44	5.23
K21	3.10	3.03	5.54	5.44
K22	3.05	3.08	5.56	5.59
K32	3.04	3.07	4.82	3.68
1a	3.04	3.00	5.51	5.58
4a	3.04	3.04	4.67	3.29
8a	3.04	3.04	4.63	3.29

[a] The second and third columns and the fourth and fifth columns, show the distances between the oxygens of the CK2 amino acids and the halogen atoms of the inhibitor after PM6-D2X and AMBER optimization (in Å).

Figure 1 The ATP competitive inhibitor (here K25 in sticks) binds to a deep cleft in the CK2α subunit structure (PDB ID code: 1ZOE), which is shown in a cartoon model (the red curled strips represent α-helices and the yellow stripes β-strands)

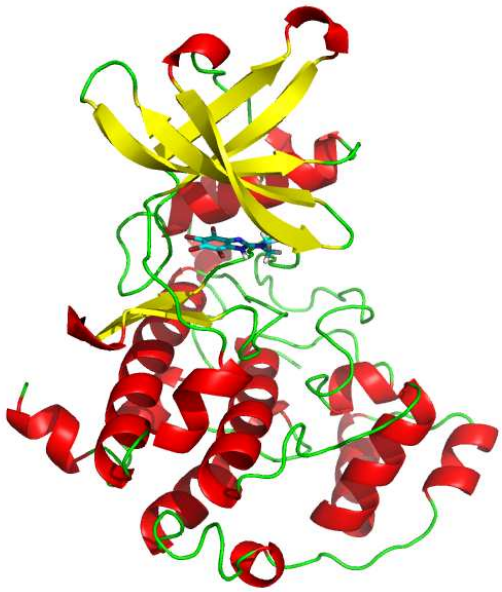


Figure 2A The structures of the nine inhibitors considered in the first set

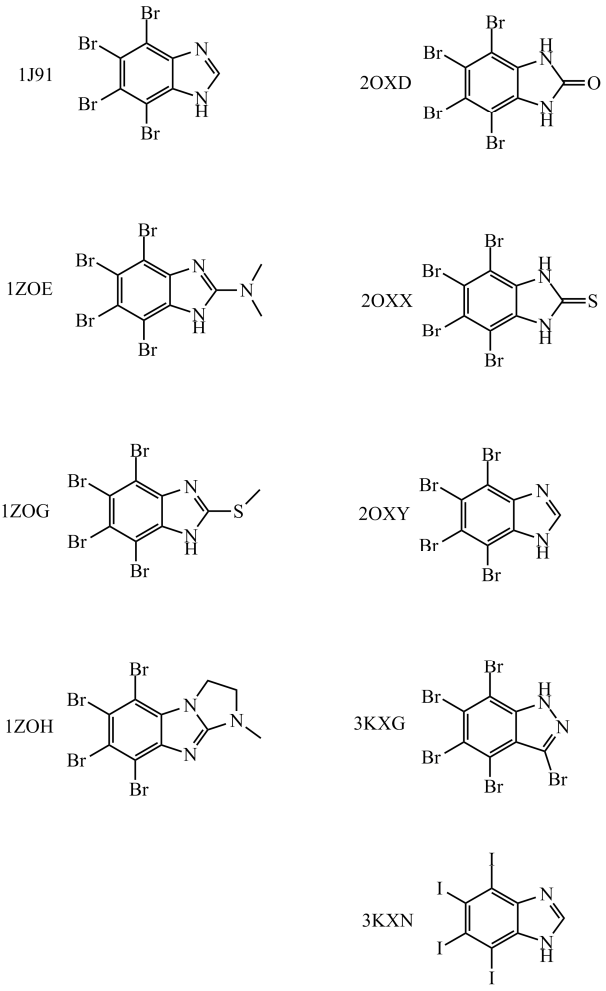


Figure 2B The structures of the fourteen inhibitors considered in second set

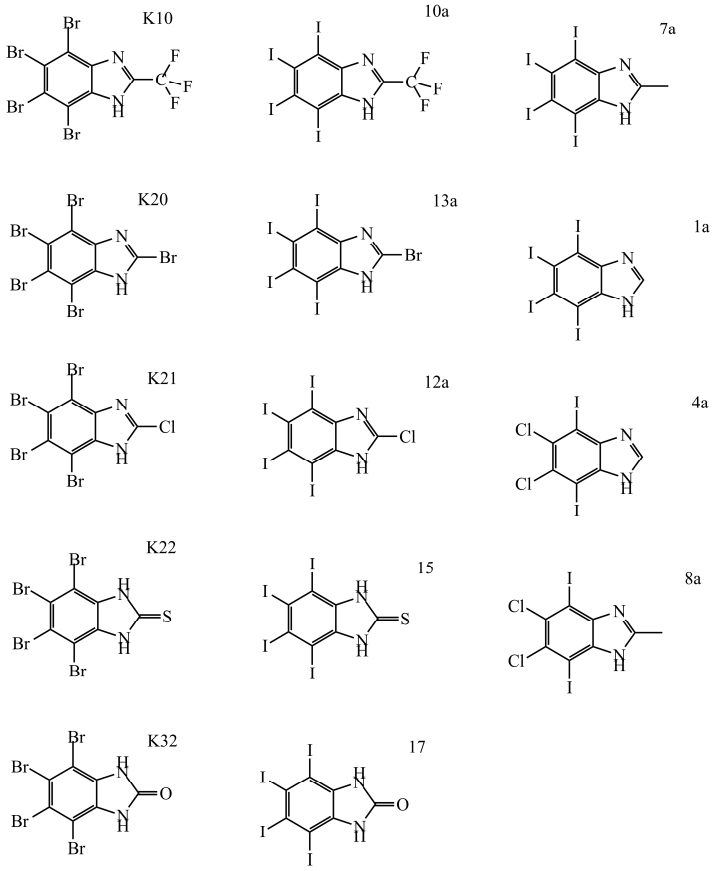


Figure 3 The correlations between the logarithm of the inhibition constants (ln K_i, in μM) and the energy terms (all of the energies are in kcal/mol) calculated by the PM6-DH2X method, i.e. with **A**: the interaction enthalpy (ΔH_w) for the first set (R^2 = 0.86), **B**: the interaction enthalpy (ΔH_w) for the second set (R^2 = 0.52), **C**: the sum of the interaction enthalpy, the inhibitor deformation and the corrections for the desolvation energies (ΔH_w + $\Delta E_{def}(I)$ + $\Delta\Delta G_w(I)$) for the first set (R^2 = 0.81), and **D**: the same as in C but for the second set (R^2 = 0.71), **E**: the total score ($\Delta G'_w$) calculated as the sum of the interaction enthalpy, the interaction entropy and the corrections for the inhibitor deformation and the desolvation energies ($\Delta G'_w = \Delta H_w - T\Delta S_w + \Delta E_{def}(I) + \Delta\Delta G_w(I)$) for the first set ($R^2$ = 0.24), **F**: the same as in E but for the second set (R^2 = 0.19).

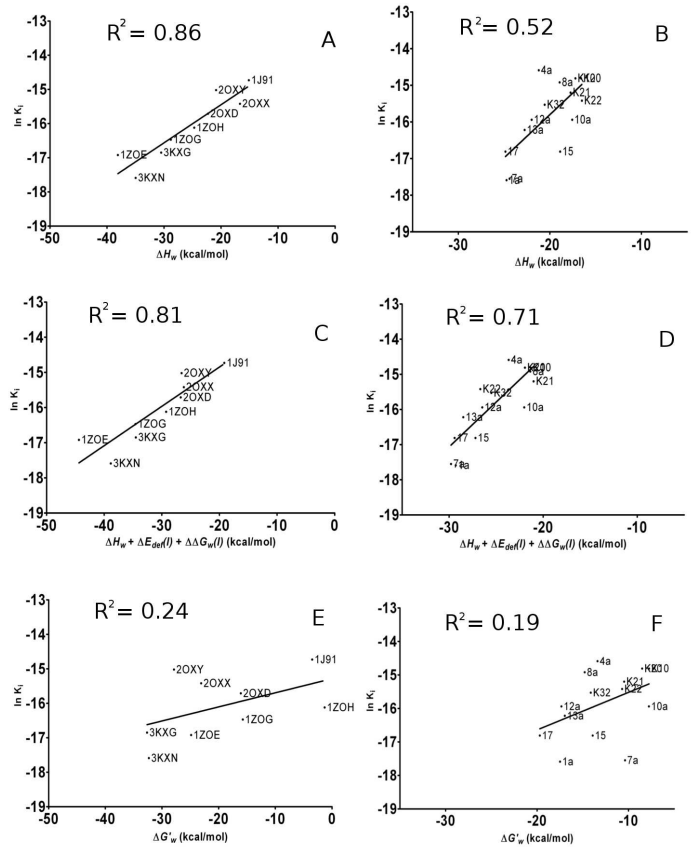


Figure 4 The correlations between the inhibition constants (ln K_i) and the AMBER characteristic calculated with **A**: the interaction enthalpy ΔH_w^{AMBER} for the first set (R^2 = 0.00), **B**: the interaction enthalpy ΔH_w^{AMBER} for the second set, (R^2 = 0.17) **C**: the total score $\Delta G'_w{}^{AMBER}$ for the first set (R^2 = 0.04), **D**: the total score $\Delta G'_w{}^{AMBER}$ for the second set (R^2 = 0.07).

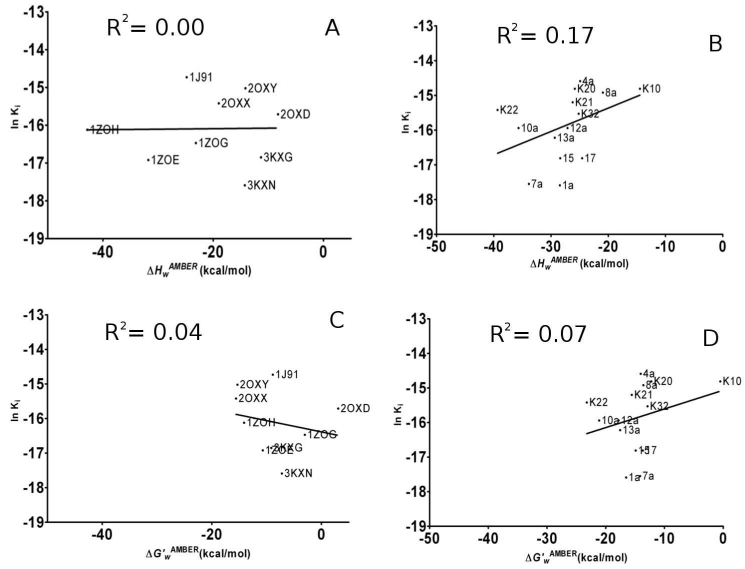
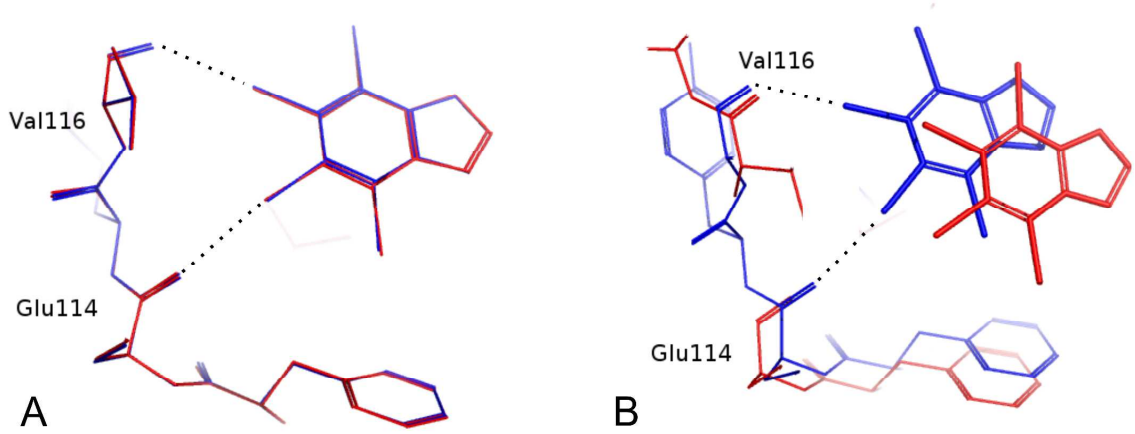


Figure 5 **A**: The PM6-D2X optimized structure (red lines) of the inhibitor bound to the CK2α active site agrees well with the 2OXY X-ray structure (blue lines), **B**: whereas the AMBER optimized structure (red lines) shows substantial differences from the X-ray geometry (blue lines). The failure of the AMBER empirical potential is caused by its inability to describe halogen bonds (shown by black dotted lines).



References

(1) Pinna, L. A.; Allende, J. E. *Cell. Mol. Life Sci.* **2009**, *66*, 1795-1799.

(2) Filhol, O.; Cochet, C. *Cell. Mol. Life Sci.* **2009**, *66*, 1830-1839.

(3) Ruzzene, M.; Pinna, L. A. *Biochim. Biophys. Acta* **2010**, *1804*, 499-504.

(4) Sarno, S.; Pinna, L. A. *Mol. Biosyst.* **2008**, *4*, 889-894.

(5) Trembley, J. H.; Wang, G.; Unger, G.; Slaton, J.; Ahmed, K. *Cell. Mol. Life Sci.* **2009**, *66*, 1858-1867.

(6) Niefind, K.; Raaf, J.; Issinger, O. G. *Cell. Mol. Life Sci.* **2009**, *66*, 1800-1816.

(7) Pagano, M. A.; Bain, J.; Kazimierczuk, Z.; Sarno, S.; Ruzzene, M.; Di Maira, G.; Elliott, M.; Orzeszko, A.; Cozza, G.; Meggio, F.; Pinna, L. A. *Biochem. J.* **2008**, *415*, 353-365.

(8) Gianoncelli, A.; Cozza, G.; Orzeszko, A.; Meggio, F.; Kazimierczuk, Z.; Pinna, L. A. *Bioorg. Med. Chem.* **2009**, *17*, 7281-7289.

(9) Cozza, G.; Bortolato, A.; Moro, S. *Med. Res. Rev.* **2010**, *30*, 419-462.

(10) De Moliner, E.; Brown, N. R.; Johnson, L. N. *Eur. J. Biochem.* **2003**, *270*, 3174-3181.

(11) Sarno, S.; Papinutto, E.; Franchin, C.; Bain, J.; Elliott, M.; Meggio, F.; Kazimierczuk, Z.; Orzeszko, A.; Zanotti, G.; Battistutta, R.; Pinna, L.A. *Curr. Topics Med. Chem., in press*.

(12) Battistutta, R.; Mazzorana, M.; Sarno, S.; Kazimierczuk, Z.; Zanotti, G.; Pinna, L. A. *Chem. Biol.* **2005**, *12*, 1211-1219.

(13) Wasik, R.; Lebska, M.; Felczak, K.; Poznanski, J.; Shugar, D. *J. Phys. Chem. B* **2010**, *114*, 10601-10611.

(14) Auffinger, P.; Hays, F. A.; Westhof, E.; Ho, P. S. *Proc. Natl. Acad. Sci. U S A* **2004**, *101*, 16789-16794.

(15) Hernandes, M. Z.; Cavalcanti, S. M.; Moreira, D. R.; de Azevedo Junior, W. F.; Leite, A. C. *Curr. Drug Targets* **2010**, *11*, 303-314.

(16) Lu, Y.; Shi, T.; Wang, Y.; Yang, H.; Yan, X.; Luo, X.; Jiang, H.; Zhu, W. *J. Med. Chem.* **2009**, *52*, 2854-2862.

(17) Voth, A. R.; Ho, P. S. *Curr. Top. Med. Chem.* **2007**, *7*, 1336-1348.

(18) Hardegger, L. A.; Kuhn, B.; Spinnler, B.; Anselm, L.; Ecabert, R.; Stihle, M.; Gsell, B.; Thoma, R.; Diez, J.; Benz, J.; Plancher, J. M.; Hartmann, G.; Banner, D. W.; Haap, W.; Diederich, F. *Angew. Chem. Int. Ed Engl.* **2010**, *50*, 314-318.

(19) Lu, Y.; Wang, Y.; Zhu, W. *Phys. Chem. Chem. Phys.* **2010**, *12*, 4543-4551.

(20) Metrangolo, P.; Neukirch, H.; Pilati, T.; Resnati, G. *Acc. Chem. Res.* **2005**, *38*, 386-395.

(21) Kortagere, S.; Ekins, S.; Welsh, W. J. *J. Mol. Graph. Model.* **2008**, *27*, 170-177.

(22) Metrangolo, P.; Meyer, F.; Pilati, T.; Resnati, G.; Terraneo, G. *Angew. Chem. Int. Ed Engl.* **2008**, *47*, 6114-6127.

(23) Politzer, P.; Murray, J. S.; Concha, M. C. *J. Mol. Model.* **2007**, *13*, 643-650.

(24) Voth, A. R.; Hays, F. A.; Ho, P. S. *Proc. Natl. Acad. Sci. U S A* **2007**, *104*, 6188-6193.

(25) Torii, H.; Yoshida, M. *J. Comput. Chem.* **2010**, *31*, 107-116.

(26) Sarwar, M. G.; Dragisic, B.; Salsberg, L. J.; Gouliaras, C.; Taylor, M. S. *J. Am. Chem. Soc.* **2010**, *132*, 1646-1653.

(27) Politzer, P.; Murray, J. S.; Clark, T. *Phys. Chem. Chem. Phys.* **2010**, *12*, 7748-7757.

(28) Politzer, P.; Lane, P.; Concha, M. C.; Ma, Y.; Murray, J. S. *J. Mol. Model.* **2007**, *13*, 305-311.

(29) Legon, A. C. *Phys. Chem. Chem. Phys.* **2010**, *12*, 7736-7747.

(30) Zierkiewicz, W.; Wieczorek, R.; Hobza, P.; Michalska, D. *Phys. Chem. Chem. Phys.* **2011**.

(31) Li, Q. Z.; Jing, B.; Li, R.; Liu, Z. B.; Li, W. Z.; Luan, F.; Cheng, J. B.; Gong, B. A.; Sun, J. Z. *Phys. Chem. Chem. Phys.* **2011**, *13*, 2266-2271.

(32) Baumli, S.; Endicott, J. A.; Johnson, L. N. *Chem. Biol.* **2010**, *17*, 931-936.

(33) Korth, M.; Pitonak, M.; Rezac, J.; Hobza, P. *J. Chem. Theory Comput.* **2010**, *6*, 344-352.

(34) Rezac, J.; Fanfrlik, J.; Salahub, D.; Hobza, P. *J. Chem. Theory Comput.* **2009**, *5*, 1749-1760.

(35) Fanfrlik, J.; Bronowska, A. K.; Rezac, J.; Prenosil, O.; Konvalinka, J.; Hobza, P. *J. Phys. Chem. B* **2010**, *114*, 12666-12678.

(36) Dobes, P.; Fanfrlik, J.; Rezac, J.; Otyepka, M.; Hobza, P. *J. Comput. Aided Mol. Des.* **2011,** *25*, 223-235.

(37) Mazanetz, M. P.; Ichihara, O.; Law, R. J.; Whittaker, M. *J. Cheminform.* **2011**, *3*, 2.

(38) Gleeson, M. P.; Gleeson, D. *J. Chem. Inf. Model.* **2009**, *49*, 670-677.

(39) Raha, K.; Peters, M. B.; Wang, B.; Yu, N.; Wollacott, A. M.; Westerhoff, L. M.; Merz, K. M., Jr. *Drug. Discov. Today* **2007**, *12*, 725-731.

(40) Dobes, P.; Otyepka, M.; Strnad, M.; Hobza, P. In *Chem. Eur. J.*; 2006/04/01 ed. 2006; Vol. 12, p 4297-4304.

(41) Stewart, J. J. *J. Mol. Model.* **2009**, *15*, 765-805.

(42) Rezac, J.; Hobza, P. *Chem. Phys. Lett., in press.*

(43) Klamt, A.; Eckert, F.; Hornig, M. *J. Comput. Aided Mol. Des.* **2001**, *15*, 355-365.

(44) Gaussian 09, Revision A.1, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H.

Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, . Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian, Inc., Wallingford CT, **2009**.

(45) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999-2012.

(46) Case, D. A.; Cheatham, T. E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668-1688.

(47) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157-1174.

(48) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, *25*, 1605-1612.

(49) Stewart, J. J. *J. Comput. Aided Mol. Des.* **1990**, *4*, 1-105.

(50) Duncan, J. S.; Litchfield, D. W. *Biochim. Biophys. Acta* **2008**, *1784*, 33-47.

(51) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049-1074.

(52) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269-10280.

(53) Dunning Jr, T. H.; In, P. J. H.; Plenum Press, New York: 1977.

(54) Kolar, M.; Fanfrlik, J.; Hobza, P. *in preparation.*

(55) Pagano, M. A.; Meggio, F.; Ruzzene, M.; Andrzejewska, M.; Kazimierczuk, Z.; Pinna, L. A. *Biochem. Biophys. Res. Commun.* **2004**, *321*, 1040-1044.

(56) Battistutta, R.; Mazzorana, M.; Cendron, L.; Bortolato, A.; Sarno, S.; Kazimierczuk, Z.; Zanotti, G.; Moro, S.; Pinna, L. A. *Chembiochem* **2007**, *8*, 1804-1809.

Graphical Abstract

The PM6-D2X optimized structure (red lines) of the inhibitor bound to the CK2α active site agrees well with the 2OXY X-ray structure (blue lines; A), whereas the AMBER optimized structure (red lines) shows substantial differences from the X-ray geometry (**B**).