

Univerzita Palackého v Olomouci

Přírodovědecká fakulta

Katedra botaniky



## ***In silico* návrh ligandů pro cyklin-dependentní kinázy**

**Bakalářská práce**

**Martina Suchánková**

Studijní program: Matematika

Studijní obor: Matematika – Biologie

Forma studia: prezenční

Olomouc 2013

Vedoucí práce: RNDr. Karel Berka, Ph. D.

## **Prohlášení**

Prohlašuji, že jsem tuto bakalářskou práci vypracovala pod vedením RNDr. Karla Berky, Ph.D a na základě uvedené literatury.“

V Olomouci dne .....

Podpis:.....

## **Poděkování**

Nejvíce bych chtěla poděkovat svému vedoucímu práce RNDr. Karlu Berkovi, Ph. D. za neuvěřitelnou trpělivost a pochopení. Také za velký přínos v oblasti znalostí a dovedností, které jsem se naučila v průběhu tvorby této práce. Dík patří také všem lidem, kteří mě podporovali.

## **Bibliografická identifikace:**

Jméno a příjmení autora: Martina Suchánková

Název práce: *In silico* návrh ligandů pro cyklin-dependentní kinázy

Typ práce: Bakalářská práce

Pracoviště: Katedra botaniky, PřF UP

Vedoucí práce: RNDr. Karel Berka, Ph.D.

Rok obhajoby práce: 2013

Abstrakt:

Téma cyklin-dependentních kináz (CDK) je velmi úzce spjato s farmaceutickými výzkumy. Vzhledem k funkci CDK při buněčném dělení jsou nejčastěji vyhledávanou skupinou ligandů jejich inhibitory, které se vědci stále snaží nalézat. Tvorba *in silico* databází CDK inhibitorů pak umožňuje nejen přehledné shromažďování dat, ale také hledání nejrůznějších vztahů mezi experimentálními daty a vlastnostmi ligandů. Cílem této práce bylo seznámení se s danou problematikou a vytvoření databází pro některé CDK na základě publikovaných dat a jejich vyhodnocení.

Klíčová slova: *in silico* metody, cyklin-dependentní kinázy, chemická databáze

Počet stran: 36

Počet příloh: 4

Jazyk: Čeština

## **Bibliographic identification:**

Autor's first name and surname: Martina Suchánková

Title: *In silico* drug design: cyclin-dependent kinases

Type of thesis: Bachelor Thesis

Department: Department of Botany, Faculty of Science, Palacký Univerzity in Olomouc

Supervisor: RNDr. Karel Berka, Ph.D.

The year of presentation: 2013

Abstract:

The theme of cyclin-dependent kinases (CDK) is very closely related to pharmaceutical research. CDK inhibitors are the most popular group of ligands due to the CDK's function in cell division. *In silico* databases of CDK inhibitors allow not only collection of experimental data, but also search a variety of relationships between the experimental data and the properties of ligands. The aim of this work was to learn about CDK inhibitors and to create a database for some CDK and evaluate it.

Keywords: *in silico* methods, cyclin-dependent kinases, chemical database

Number of pages: 36

Number of appendices: 4

Language: Czech

## Obsah

1. Cíle práce .....	- 8 -
2. Literární rešerše.....	- 9 -
2.1 <i>In silico</i> metody návrhu ligandů .....	- 10 -
2.1.1 Metody založené na znalosti struktury receptoru (structure-based approach)....	- 11 -
2.1.2 Metody založené na znalosti ligandu (ligand-based approach) .....	- 12 -
2.2 Cyklin-dependentní kinázy .....	- 13 -
2.2.1 Role CDK v buněčném cyklu .....	- 14 -
2.3 Inhibitory CDK.....	- 15 -
2.3.1 Znaky CDK inhibitorů .....	- 15 -
2.3.2 Užití CDK inhibitorů při léčbě chorob .....	- 16 -
2.3.3 Předchozí výzkum na CDK5/p25 .....	- 17 -
2.3.4 Předchozí výzkum na CDK2 .....	- 18 -
2.3.5 Předchozí výzkum na CDK1/cyklin B.....	- 20 -
3. Metody .....	- 21 -
3.1 Deskriptory .....	- 23 -
4. Výsledky .....	- 25 -
4.1 QSAR model pro CDK1 inhibitory .....	- 25 -
4.2 QSAR model pro CDK2 inhibitory .....	- 27 -
4.2.1 Rozpětí deskriptorů v databázi pro CDK2.....	- 30 -
5. Diskuze .....	- 31 -
6. Závěr .....	- 32 -

## Úvod

Objev cyklin-dependentních kináz jakožto regulátorů buněčného cyklu je poměrně nedávnou událostí. Bylo zjištěno, že zastavení funkce cyklin-dependentních kináz (CDK) vede také k zastavení buněčného cyklu, a tak zamezení dělení buněk. Ve spojení s onemocněními, kde zhoubné bujení buněk má fatální následky pro celý organismus, se tak zrodila myšlenka, jak uplatnit výzkum CDK přímo v medicíně. Inhibitory, které by zvládly soutěžit o vazebné místo na CDK s přirozeným substrátem adenosintrifosfátem (ATP) a tím zastavit funkci CDK, a tedy i buněčné dělení, by tak mohly být využity jako léčivo proti alespoň některým typům zhoubného bujení.

Hledání inhibitorů CDK mezi vhodnými molekulami vyžadovalo testování mnoha stovek až tisíců látek, u nichž se zjišťovala jejich inhibiční schopnost. Protože však *in vitro* a *in vivo* testování jsou relativně drahé, byly vyvinuty metody *in silico* využívající počítačové techniky pro prohledávání databází látek – možných inhibitorů a odhadování jejich inhibiční aktivity. Podstatnou výhodou *in silico* metod je finanční, materiální a také časová úspora v předvýběru vhodných látek pro biologické testování.

V učitelství lze poznatky z této práce využít například jako ukázkou, že i malé molekuly dokáží ovlivnit chod celé buňky, a to může mít dopad na celý organismus. Dále je také patrné z 3D zápisu, jak málo se od sebe mohou účinné a neúčinné látky lišit.

# 1. Cíle práce

Cílem této bakalářské práce bylo:

- provést literární rešerši týkající se cyklin-dependentních kináz a *in silico* metod se zaměřením na získání experimentálních dat,
- naučit se pracovat v programu MOE, umět kreslit molekuly a tvořit databáze,
- zhotovit databáze pro více cyklin-dependentních kináz a vytvořit QSAR model a analyzovat důležitost jednotlivých deskriptorů.



## 2. Literární rešerše

Ligand (z latinského *ligandum* – vazba, spojení) je molekula, která je schopna se navázat na určité aktivní místo biomakromolekuly (receptoru). K vazbě ligandu s biomolekulou dochází pomocí mezimolekulárních sil, jako jsou například elektrostatické interakce, van der Waalsovy síly a vodíkové můstky. Pevnější kovalentní vazba mezi ligandem a receptorem je v biologických systémech vzácná, ale může se využít při návrhu ligandů. Síla vazby se nazývá afinitou, která je určena nejen přímou interakcí ligandu s receptorem, ale také změnou interakce s rozpouštědlem.

Podle toho, jak se chovají, můžeme ligandy dělit na *substráty*, *aktivátory* a *inhibitory*.

**Substrátem** jsou takové ligandy, které je daná biomolekula (enzym, či ribozym) schopna přeměnit na produkty pomocí chemické reakce. Enzymy se rozlišují podle toho, jakou mají substrátovou specifitu a tedy například, zda jsou schopny zpracovávat pouze jeden substrát, nebo několik různých substrátů.

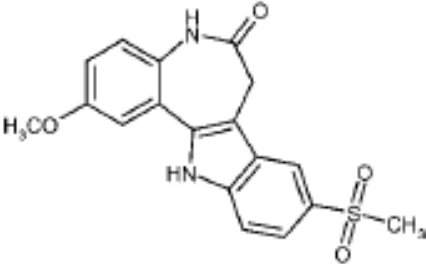
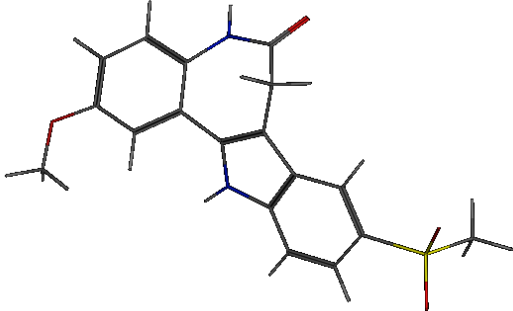
**Aktivátory** jsou pak látky, které ulehčí běh chemické reakce, aniž by byly samy měněny. Častokrát se jedná o látky, které se váží mimo aktivní místo, ale alosterickým působením změni jeho konformaci do účinnější.

**Inhibitorem** je látka, která brání přeměně substrátu na produkt. Podle typu blokáce rozpoznáváme inhibitory kompetitivní, akompetitivní a nekompetitivní. Při kompetitivní inhibici ligandy soupeří se substrátem o vazebné místo na enzymu. V případě akompetitivní inhibice se naváže inhibitor na komplex enzym-substrát a znemožní přeměnu substrátu na produkt. Nekompetitivní inhibice pak nastává, naváže-li se ligand na alosterické místo enzymu. Nebrání tedy navázání substrátu, ale zpomaluje rychlost jeho přeměny na produkt. Většina inhibitorů působí pomocí kompetitivní inhibice, kdy se inhibitor naváže do aktivního místa místo substrátu, a tak blokuje jeho přeměnu. Podle poměru afinit a koncentrací substrátu a inhibitoru pak dojde k menší či větší inhibici dotyčné chemické reakce. Inhibitory receptorů se dělí podle toho, zda vyvolají konformační změnu receptoru podobnou, jako by se navázal původní ligand (v tomto případě se také hovoří o efektoru), a tedy dojde k přenosu signálu dále po signální dráze – pak se jedná o tzv. *agonisty*; nebo zda navázáním inhibitoru nedojde ke kýžené změně konformace a signál se dál nepředá – v takovém případě mluvíme o *antagonistech*.

## 2.1 *In silico* metody návrhu ligandů

*In silico* je označení pro metody, které jsou založeny na využití počítačové techniky. Tento výraz vznikl jako zdůraznění důležitosti těchto metod spolu s původnějšími metodami *in vitro* a *in vivo* (Ekins, S. et al., 2007). Velkou výhodou *in silico* metod v porovnání s metodami *in vitro* (ve skle) je menší investice do materiálu, časová úspora a nezávislost na vnějších podmínkách.

Pomocí metod molekulového modelování není dnes problém navrhnout téměř libovolnou molekulu ligandu. Tu je třeba uložit v počítači tak, abychom s touto strukturou mohli i nadále pracovat. Vždy je zapotřebí vyjít z dostupných experimentálně zjištěných dat. Ligandy a nízkomolekulární látky obecně můžeme do počítače zapsat různým způsobem, podle toho jakou informaci potřebujeme posléze vyhledávat. Nejkratší zápis je tzv. 1D, kdy jde o způsob záznamu v řádce textu podobně, jako bychom zapisovali název molekuly ale ve strojově čitelném zápisu. 1D zápis bohužel často není příliš přehledný (obrázek č. 1). V chemické literatuře se nejčastěji používá způsob 2D zápisu, kdy se zakresluje schematicky tvar molekuly v rovině. Nevýhodou 2D zápisu bývá ne zcela jednoduchá orientace u složitějších prostorových molekul. Posledním typem je 3D zápis, kdy se zapisují pozice všech atomů v molekule do kartézské souřadné sítě. Nevýhodou tohoto zápisu bohužel je, že tak ke každému atomu musíme ukládat 4 bity informace – typ atomu a jeho XYZ souřadnice. Proto tento způsob zápisu zabírá nejvíce místa a v případě velkých databází se mnohdy za běhu generuje idealizovaná 3D struktura na základě 1D zápisu. Největší výhodou 3D zápisu pak je, že nejlépe odpovídá skutečné molekule a jejímu tvaru, a tedy je to nejvhodnější zápis pro *in silico* predikční metody.

Forma	Způsob zápisu
	1D
	2D
	3D

Obrázek č. 1: Porovnání způsobu zápisu 1D, 2D, 3D pro 2-methoxy-9-methylsulfonylpauillon.

*In silico* metody pro predikci ligandů pak můžeme rozdělit do dvou skupin v závislosti na tom, zda se k hledání nových látek využívá struktura receptoru (tzv. structure-based approach), nebo zda se používají pouze znalosti o známých ligandech (tzv. ligand-based approach).

### 2.1.1 Metody založené na znalosti struktury receptoru (structure-based approach)

Pro první přístup založený na struktuře je potřeba znát trojrozměrnou strukturu receptoru. Ta se většinou získá pomocí kombinace rentgenového záření a krystalografie, nukleární magnetické rezonance nebo tvorbou homologních modelů. Mezi těmito metodami najdeme například tzv. *dokování*.

Molekulární dokování je proces, který umožňuje hledat vhodné ligandy k daným biomakromolekulám, které jsou k nim komplementární (Ronald, M. A. et al., 1997). Dokování umožňuje najít nejvhodnější orientaci molekul z hlediska komplementarity při tvorbě komplexu ligand - receptor. Ve vazebných místech receptorů lze pak identifikovat i jednotlivé skupiny, podle kterých lze odhadnout vazbu s jinými molekulami.

Dřívější metody dokování pracovaly na principu rigidního zámku i klíče, později byl tento proces vylepšen tak, aby byly brány v potaz i konformační změny, zvláště u ligandů. Zohlednění této flexibility zpřesňuje odhady dokování (Daniel, L., 2009). Dokování v principu nevyžaduje znalost experimentálně ověřeného ligandu, ale omezená přesnost metod dokování vyžaduje mnohdy alespoň prvotní vstup experimentálních dat a návrh původních ligandů.

Dokování je velmi užitečným nástrojem pro vysvětlení funkcí proteinových komplexů. Jeho využití se dá uplatnit i v medicíně k výrobě léčiv. To spočívá v nalezení vhodných inhibitorů určitých enzymů, které budou co nejvíce selektivní, a tedy budou tvořit účinné léky. Jedním z příkladů využití dokování v medicíně je studium cyklin-dependentních kináz (CDK) a jejich inhibitorů.

### **2.1.2 Metody založené na znalosti ligandu (ligand-based approach)**

Druhou skupinou jsou metody, které pro hledání využívají již známých ligandů (ligand-based approach). Ligand, který se váže na daný receptor, slouží jako šablona pro další molekuly, které mají podobné fyzikální a chemické vlastnosti. Mezi tyto metody patří například QSAR a farmakofor. Tyto metody můžeme využít, i když neznáme trojrozměrnou strukturu receptoru (Daniel, L., 2011).

QSAR (quantitative structure activity relationship) – je metoda využívající matematický model korelující chemické či biologické vlastnosti molekuly podle její molekulární struktury pomocí statistických metod. Pokud pro sadu testovaných molekul najdeme korelaci s dostupnými biologickými daty, pak můžeme matematický model použít pro predikci biologických efektů dalších molekul. QSAR má ale bohužel své limity zvláště při hledání nových látek, které by se svou základní strukturou výrazně odlišovaly od látek, na jejichž základě byl QSAR model vybudován (Ekins, S., 2007).

Farmakofor je 3D uspořádání molekulárních skupin ve struktuře ligandu nezbytných pro biologickou aktivitu. Systém funguje na základě podobnosti ligandu s předlohou (Ekins, S., 2007).

Na závěr této podkapitoly je třeba upozornit, že pro otestování dat získaných *in silico* modelem je třeba znát experimentální data. Nejlépe dostupná je hodnota značená jako  $IC_{50}$ . Toto číslo udává koncentraci, při níž je inhibováno 50% množství dané látky.  $IC_{50}$  se běžně udává při výzkumech účinnosti antagonistických látek. Obměnou pak může být  $pIC_{50}$ , což

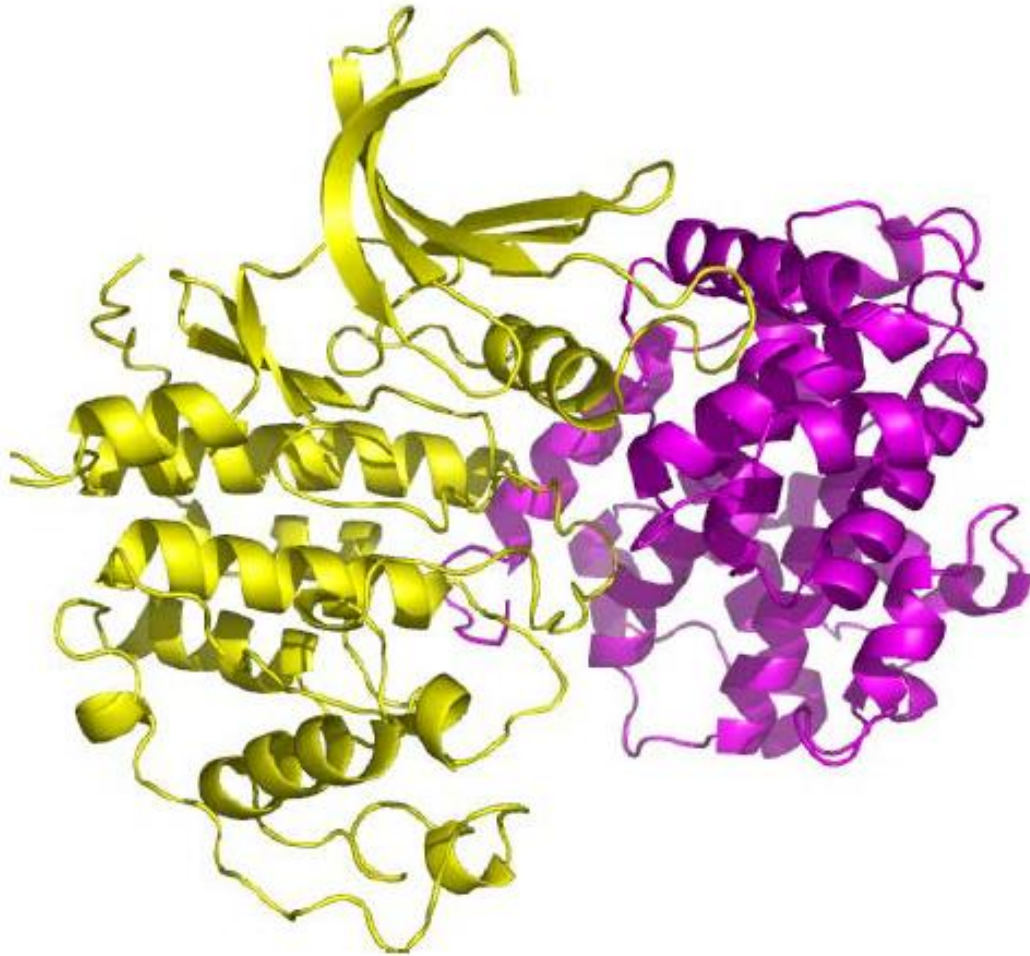
odpovídá hodnotě  $-\log IC_{50}$ , kdy s rostoucí schopností inhibovat roste i tato hodnota, její rozmezí je však v přijatelnějších číslech.

## 2.2 Cyklin-dependentní kinázy

Cyklin-dependentní kinázy patří mezi serin/threonin kinázy. Za objev těchto klíčových regulátorů buněčného cyklu byla pánům L. H. Hartwellovi, T. R. Huntovi a P. R. Nurseovi udělena roku 2001 Nobelova cena za medicínu. Funkce CDK má vliv nejen na průběh buněčného cyklu, ale také na apoptózu, diferenciaci a transkripci. Narušení funkcí CDK má tedy často vliv při onkologických onemocněních, ale i při poruchách nervového systému. Jak sám název napovídá, tyto kinázy tvoří komplex katalytické podjednotky (CDK) a podjednotky regulační pro jejich aktivaci, což je větší protein – cyklin. Pouze jako komplex mohou tyto regulátory buněčného růstu fungovat správně.

Na aktivitu CDK má vliv právě podjednotka cyklinu, dále CDK-aktivující kinázy (CAK) a také podléhání různým přeměnám, např. fosforylaci. Pro aktivitu CDK1 jsou rozhodující cykliny A a B, CDK2 je regulováno cykliny A a E, na CDK4 působí cyklin typu D, CDK7 se pojí s cyklinem H a CDK9 s cyklinem T. Různé kombinace CDK a cyklinů slouží k regulaci jiných kontrolních bodů v buněčném cyklu. Například CDK2 s cyklinem E hraje roli v G1/S fázi, zatímco v kombinaci s cyklinem A usnadňuje rozvoj S fáze (Dong, X., 2012; Otyepka, M., 2006). Inhibitory s menší selektivitou mohou negativně ovlivnit také transkripci, což sníží syntézu RNA a tedy i translaci a tvorbu proteinů, tedy zpomalí buněčný růst (Jašková, E., 2010).

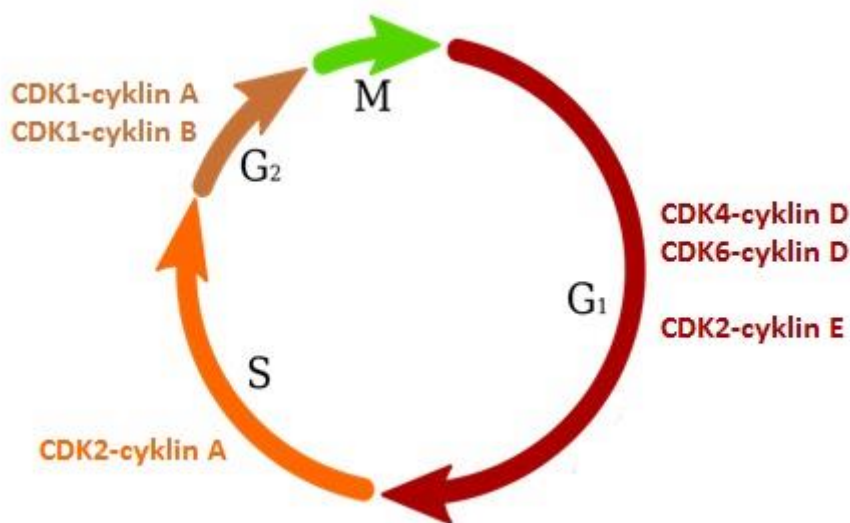
CDK se skládá z větší C-terminální domény, kterou tvoří  $\alpha$  šroubovice, a N-terminální domény tvořené  $\beta$  skládané listy, mezi nimiž je štěrbin. To je místo, kam se vážou ligandy, ať už jde o ATP jako přirozený substrát CDK nebo také další inhibitory. Přičemž vzájemná orientace domén souvisí právě s navázanou látkou. Na C-terminální doméně se pak nachází tzv. T-smyčka. Aktivace CDK pak může mít jednu nebo dvě fáze. První z nich je navázání regulační podjednotky – cyklinu (obrázek č. 2). Dalším krokem je fosforylace pomocí CAK v T-smyčce, tento krok napomáhá aktivaci CDK2, u jiných CDK však nemusí probíhat. Změny mohou nastat i při prvním kroku, kdy CDK místo spojení s cyklinem může být aktivována jinými proteiny jako je tomu u CDK5.



Obrázek č. 2: Lidská CDK2, žlutě vyznačený CDK podjednotka, fialově cyklin A, PDB ID:1FIN (Otyepka M., 2006).

### 2.2.1 Role CDK v buněčném cyklu

Buněčný cyklus je regulován mnoha faktory. Jak již bylo zmíněno, klíčovou roli zde hraje regulace pomocí komplexů CDK s cykliny. V G1 fázi působí CDK2 v komplexu s cyklinem E, CDK4 a CDK6 s cyklinem D. Tyto komplexy mají vliv i na přechod z G1 fáze do S fáze. Průběh S fáze a replikace DNA podporuje CDK2 ve spojení s cyklinem A. Tento komplex společně s CDK1 a cyklinem A a B zajišťují následný přechod z G2 fáze do fáze mitózy. Buněčné dělení je pak regulováno pomocí CDK1-cyklin B komplexu (obrázek č. 3). Další vliv CDK nastává také při transkripci, zde plní svou úlohu komplexy CDK7-cyklin H, CDK9-cyklin T a CDK8 s cyklinem C. Tyto komplexy umožní činnost RNA-polymerázy II a tím spustí transkripci DNA. CDK7-cyklin H komplex působí také zároveň jako CAK. (Jašková, E., 2010).



Obrázek č. 3: Buněčný cyklus. G<sub>1</sub> fáze (růst buněčné hmoty, tvorba RNA a proteinů), S fáze (replikace DNA), G<sub>2</sub> fáze (zdvojení organel, vznik mitotického vřeténka), M fáze (mitóza, karyokineze, cytokineze)

## 2.3 Inhibitory CDK

Vliv regulace buněčného cyklu pomocí CDK je významný zejména při výzkumu nádorových onemocnění, kdy dochází často k deregulaci CDK. Proto je snaha najít přirozené i umělé inhibitory CDK, které by mohly zabránit deregulaci a tak i nadměrnému buněčnému dělení. Největší potenciál pak mají nízkomolekulární inhibitory, které se jeví pro případnou léčbu nejvhodnější. Porucha funkcí CDK nenastává jen u zhoubného bujení, ale také u některých neurodegenerativních chorob (např. Alzheimerova nemoc), virových nákaz, revmatické artritidy, kardiovaskulární choroby a diabetických komplikací (Otyepka, M., 2006; Jašková, E., 2010).

Prvními významnými inhibitory, které se dostaly i do klinického testování, byly zejména přírodní inhibitory jako flavonoid, flavopiridol, staurosporin a jeho analogy, olomoucín, UCN-01, indirubiny, paullony a butyrolakton, které však měly nízkou selektivitu. Díky znalosti struktur CDK se však vyvinuly inhibitory s vyšší selektivitou, například EXEL-8647, BS-181, SNS-032, SU9516, AZ703, PD 0332991 a další. Poslední zmiňovaný inhibitor je kandidátem na specifický inhibitor pro CDK 4 a CDK6 (Jašková, E., 2010).

### 2.3.1 Znaky CDK inhibitorů

Přes celkovou rozmanitost mají CDK inhibitory některé společné vlastnosti: (1) mají nízkou molekulární hmotnost (<600); (2) jsou to ploché hydrofobní heterocykly; (3) jsou kompetitory

pro vazebné místo pro ATP; (4) váží se na kinázu především pomocí hydrofobních interakcí a vodíkových můstků (Knockaer, M., 2002).

CDK inhibitory můžeme rozdělit do tří skupin. Jsou to inhibitory, které nejsou selektivní pro žádnou specifickou CDK (např. flavopiridol a deschloroflavopiridol). Další skupinou jsou ty, které inhibují CDK1, 2, 5, případně 9 (např. olomoucín a (*R*)-roscovitin). Poslední skupinu tvoří inhibitory selektivní pro CDK4 a 6 (např. fascaplysin).

Tyto inhibitory však musí soutěžit o aktivní místo na CDK s ATP. Vzhledem k velkému množství ATP v buňkách a vysoké afinitě k proteinům, musí být inhibitor dosti zvýhodněn jinými vlastnostmi. Dalším problémem je také velká podobnost mezi jednotlivými typy CDK, která se jeví jako překážka pro selektivitu jednotlivých inhibitorů a také fakt, že kinázy zaujímají mnoho konformačních stavů. To ale do jisté míry nahrává právě rozvoji selektivních inhibitorů (Jašková, E., 2010). Pro použití CDK pro farmakologické účely je selektivita klíčovou otázkou, ale zdá se, že i kombinace působení inhibitoru může být terapeuticky účinná (Knockaert, M., 2002).

Význam inhibitoru však nespočívá pouze v jejich antiproliferačních účincích. U inhibitorů s nižší selektivitou dochází k působení na více typů CDK současně, což vede nejen k zastavení buněčného cyklu, ale může dojít až k apoptóze. Takovéto účinky vykazují například flavopiridol, roskovitin a SNS-032. Roskovitin podstoupil již klinická testování a jeho velkou výhodou je, že způsobuje apoptózu ve všech stádiích buněčného cyklu (Jašková, E., 2010).

### **2.3.2 Užití CDK inhibitorů při léčbě chorob**

#### ***Rakovina***

CDK jsou pro léčbu rakoviny atraktivní ze třech důvodů, a to pro jejich anti-proliferativní účinky, schopnost navodit apoptózu nebo může inhibice CDK přispět k buněčné diferenciaci. Působení CDK na buněčný cyklus a zároveň navození apoptózy jsou nezbytné pro efektivní chemoterapii (Knockaert, M., 2002). Inhibitory CDK uvažované v léčbě rakoviny se navíc jeví jako dobrá cesta, protože jsou všeobecně snášeny lépe než chemoterapie, která má mnoho vedlejších účinků (Jašková, E., 2010).

#### ***Poruchy nervového systému***

V nervové soustavě je pro fosforylaci mnohých substrátů uplatňována CDK5, která je za normálních okolností aktivována proteiny p35 a p39. Někdy ale může dojít k navázání



proteinů p25, respektive p29, které způsobí, že CDK5 hyperaktivně fosforyluje proteiny cytoskeletu, což vede k buněčné smrti. Případnou inhibicí CDK5 by se tedy mohlo zabránit některým neurodegenerativním onemocněním, například Alzheimerově chorobě (Knockaert, M., 2002).

### ***Virové infekce***

Během několika posledních let bylo zjištěno, že inhibice CDK zabraňuje replikaci virů *in vitro*. Mnoho virů vyžaduje aktivní CDK pro svou replikaci a některé typy virů mají zakódovány ve své struktuře vlastní cykliny, čímž regulují buněčný cyklus jejich hostitelské buňky. Nejlépe prostudovaný je případ herpes simplex viru (HSV). Roscovitin zde inhibuje transkripci okamžitých, časných a pozdních genů HSV a syntézu DNA (Knockaert, M., 2002).

### ***Jednobuněční parazité***

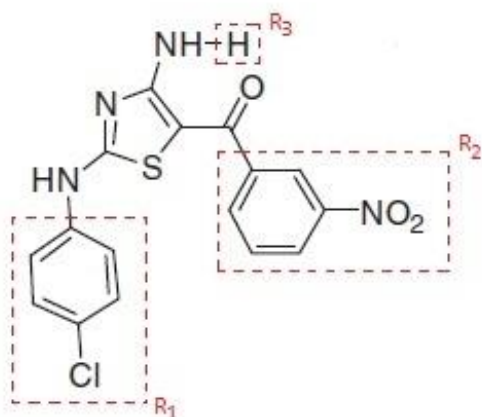
*Plasmodium falciparum* (původce malárie), *Leishmania major* (původce leishmaniózy), *Trypanosoma brucei* (spavá nemoc), *Trypanosoma cruzi* (původce Chagasovy choroby) a *Toxoplasma gondii* (původce toxoplazmózy) – tyto mikroorganismy mají specifické CDK. Tyto strukturní rozdíly mezi parazitickými a hostitelskými CDK mohou zapříčinit rozdílnou afinitu k inhibičním molekulám, a tak vést ke specifickým CDK inhibitorům proti parazitům (Knockaert, M., 2002).

### **2.3.3 Předchozí výzkum na CDK5/p25**

CDK5/p25 se objevil jako zásadní léčebný cíl pro mnohá akutní a chronická neurodegenerativní onemocnění včetně Alzheimerovy choroby. Na rozdíl od jiných CDK, které se aktivují spojením s cyklinem, aktivita CDK5 souvisí s navázáním proteinu p35. Tento protein může být proteolytickými procesy upraven až na p25, který stejně tak aktivuje CDK5. Při různých neuronálních úrazech může dojít k fosforylaci proteinu tau. Přílišná fosforylace tohoto proteinu (i jiných) při mnoha neurodegenerativních onemocněních zdůrazňuje potenciální roli CDK5/p25 komplexu, nejvíce při Alzheimerově chorobě. Další souvislost se zdá být i s chorobami jako je ischemie mozku, roztroušená skleróza, amyotrofická laterální skleróza, Huntingtonova nemoc a Parkinsonova nemoc.

Již dříve byla provedena kolorimetrická enzymoimunoanalýza (enzyme-linked immunosorbent assay = ELISA) a byly nalezeny dva inhibitory pro CDK5/p25, a to bellidin ( $IC_{50} = 0,2 \mu M$ ) a 2,4-diaminotiazol (**1**,  $IC_{50} = 2,0 \mu M$ ). Z této struktury se pak vycházelo a hledaly se inhibitory pro CDK5/p25 komplex s lepší inhibiční aktivitou (obrázek č. 4). Bylo

použito radiometrického testu. Aromatický dusík v původní sloučenině 1 nebyl nezbytný pro inhibici a bylo snahou jej nahradit (Laha, J. K. et al., 2011).

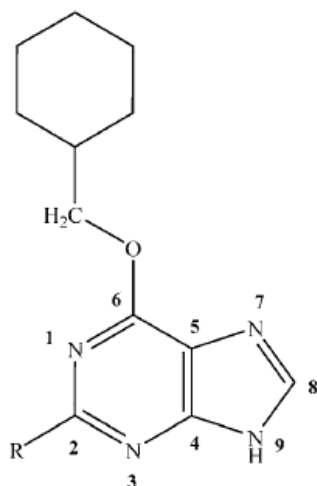


Obrázek č. 4: 2,4-diaminotiazol s označenými rezidui (upraveno z Laha, J. K. et al., 2011)

#### 2.3.4 Předchozí výzkum na CDK2

Pro data o CDK 2 jsme čerpali ze studie J. H. Alzate-Moralese a kol. (2007) využívající kombinace kvantové mechaniky a molekulární mechaniky (QM/MM) pro určení interakční energie pro vazbu protein-ligand, a to mezi CDK2 a pěti inhibitory se substituovaným N<sup>2</sup>-6-cyklohexylmethoxypurinem. Ukázalo se, že QM/MM interakční energie koreluje s biologickou aktivitou, a tedy může být využita alespoň v rámci určení substrátu. CDK2 ve vazbě s cyklinem E působí v S-fázi v takzvaném reskripčním bodu. CDK2 s cyklinem A pak hraje roli při ukončení S-fáze.

Aktivita CDK-cyklin komplexu může být redukována minimálně dvěma mechanismy. Buďto dojde k fosforylaci CDK podjednotky v inhibičním místě, nebo se CDK spojí se specializovanými proteinovými inhibitory, označovanými jako CKI nebo CDK inhibitory. Tyto inhibitory soupeří s ATP o aktivní místo na CDK. V některých rakovinných buňkách jsou CKI málo exprimované, a proto proběhly výzkumy pro nalezení syntetických inhibitorů. Mezi pozoruhodné pokusy o nalezení takových inhibitorů patří série sloučenin založených na O<sub>6</sub>-cyklohexylmethylpurinu nebo NU2058 (obrázek č. 5), které jsou inhibitory pro CDK1 i CDK2. Vykazují také dobrou selektivitu pro CDK4.



Obrázek č. 5: Struktury se substituovaným  $N_2$  na derivát  $O_6$  –cyklohexylmethylguaninu, se zobrazením číslování atomů. Sloučenina 2:  $R = 2$ -anilino derivát, sloučenina 3:  $R = 2$ -(4'-sulfamoylanilino), sloučenina 7:  $R = 2$ -methylamino, sloučenina 9:  $R = 2$ -(1-methylethylamino), sloučenina 25:  $R = 2$ -(4'-hydroxyanilino) (Alzate-Morales, J. H. et al., 2007).

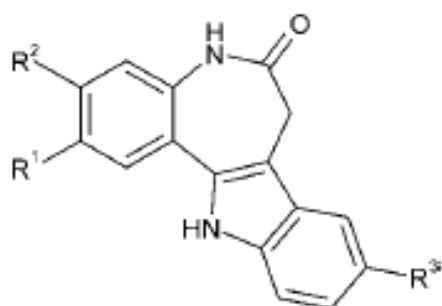
Bylo zjištěno, že je možné identifikovat tři druhy charakteristických interakcí pro skupinu sloučenin vázajících se na CDK bez znalosti aktivní části některé z CDK. První znak je přítomnost tripletu vodíkových můstků k páteři CDK. Tato vlastnost umožňuje jinou orientaci sloučenin uvnitř aktivního místa enzymu s ohledem na jiné inhibitory, jako je flavopiridol a olomoucín. Druhým znakem je optimální navázání látky s přiměřeně velkým alifatickým  $O_6$  substituentem. Třetí charakteristický znak interakce je spojen s tzv. speciálním povrchem, což jsou zbytky, které leží mimo velmi konzervativní místo pro navázání ATP.

Pro zhodnocení vazebné afinity ligandu je nezbytné uvažovat i flexibilitu proteinu. Zvláště kvůli konformační plasticitě katalytické podjednotky CDK. Nevyhnutelný je kompromis mezi přesností a výpočetními náklady. Metody molekulární dynamiky mohou být použity pro spočítání volné energie pro vazbu proteinu a ligandu s vysokou přesností se zachováním entropie, včetně flexibility receptoru. Problémem této metodiky je výpočetní úsilí potřebné k náležitě části z celého nakonfigurovaného prostoru a také parametrizace každého nového ligandu. Proto se využilo kombinace kvantových a molekulárních metod (QM/MM). Propojení těchto metod nabízí další výhody – zahrnuje polarizaci navázaného ligandu, dále díky QM nadefinování ligandu se dá snáze předpokládat jeho dostupná konfigurace (Alzate-Morales, J. H. et al., 2007).

### 2.3.5 Předchozí výzkum na CDK1/cyklin B

CDK1 a její aktivující podjednotka cyklin B jsou důležité v G2 a M fázi buněčného cyklu. Bylo zjištěno, že jejich přílišná aktivita se vyskytuje u karcinomu prsu, prostaty i kolorektálním karcinomu. Inhibice CDK1/cyklin B komplexu tedy nabízí také možnost terapeutického využití, zvláště u nádorových onemocnění (Pies, T. et al., 2004).

Za posledních deset let bylo nalezeno více malých molekul, které inhibovaly CDK1. Nejvýznamnější byl flavopiridol, který prošel i klinickým testováním jako protirakovinná látka. Další skupinou látek jsou paullony, které mají antiproliferativní účinky na kultivované buňky nádoru. Od nich byly odvozovány další molekuly obměnou reziduí R<sub>1</sub>, R<sub>2</sub> a R<sub>3</sub> (obrázek č.6).



Obrázek č. 6: Výchozí molekula pro další návrh inhibitorů CDK1 (Pies, T. et al., 2004)

V článku byla porovnávána data z dřívějších prací a data z QSAR analýzy. Z různých parametrů byla znatelná korelace nalezena pouze u elektronových parametrů. Dokonce byla nalezena lineární rovnice pro tento parametr, kde pIC<sub>50</sub>(CDK1/cyklin B) vystupovalo jako závisle proměnná (pIC<sub>50</sub> je označení pro  $-\log IC_{50}$ ). Zjištění, že lipofilní parametry nejsou relevantní pro biologickou aktivitu, může být vysvětleno faktem, že sloučeniny nemusí pronikat přes membránu při *in vitro* CDK1 inhibici. Pro kontrolu výsledků byly syntetizovány 4 molekuly, které byly dále testovány vzhledem k inhibici CDK1/cyklin B komplexu. Tyto sloučeniny se vyznačovaly síru obsahujícími substituenty. I pro ně byla aktivita uspokojivě předvídatelná podle QSAR modelu. Tyto 4 deriváty paullonu prokázaly potenciální inhibici CDK1/cyklin B komplexu a při nanomolárních koncentracích. (Pies, T. et al., 2004)

### 3. Metody

Chemická databáze je seznam molekul, ve kterém jsou zaznamenány i určité vlastnosti těchto molekul. Můžeme sem zanást experimentální data, provádět zde výpočty a predikovat chování obdobných molekul. Pro tuto databázi je nejvhodnější ukládat molekuly v jejich 3D struktuře, protože právě tak je popis dané molekuly nejpřesnější, i když datově nejnáročnější. Jednou z největších databází je Protein Data Bank (PDB). V této databázi jsou uloženy molekuly v jednotném formátu.

PDB vyvinula textový formát (.pdb), který slouží k ukládání trojrozměrných molekulárních struktur. (Richter, J., 2013).

PDB – Protein Data Bank je počítačová databáze pro makromolekulární struktury. Text zahrnutý v datech obsahuje kromě trojrozměrného záznamu vzájemné pozice všech atomů v molekule také informace o způsobu stanovení struktury, její základní analýzy a také primární citace (obrázek č. 7). Údržba databáze spadá pod Brookhaven National Laboratory. Všechna data uložená v bance jsou dostupná pro veřejnou distribuci pro Ameriku z Brookhaven na adrese [www.rcsb.org](http://www.rcsb.org). Pro Japonsko z Tokia a pro Evropu z Cambridge jsou pak tatáž data uložena v duplicitní kopii (Bernstein, F. C. et al., 1977).

```
REMARK 99
REMARK 99 MOE v2009.1 (Chemical Computing Group Inc.) Thu Jul 18 12:24:54
HETATM 1 C * 0 1.230 0.710 0.000 0.00 0.00
HETATM 2 C * 0 0.000 1.420 0.000 0.00 0.00
HETATM 3 O * 0 0.000 2.796 0.000 0.00 0.00
HETATM 4 C * 0 -1.230 0.710 0.000 0.00 0.00
HETATM 5 C * 0 -1.230 -0.710 0.000 0.00 0.00
HETATM 6 CL * 0 -2.720 -1.570 0.000 0.00 0.00
HETATM 7 C * 0 0.000 -1.420 0.000 0.00 0.00
HETATM 8 C * 0 1.230 -0.710 0.000 0.00 0.00
HETATM 9 H * 0 2.165 1.250 0.000 0.00 0.00
HETATM 10 H * 0 -0.882 3.207 0.000 0.00 0.00
HETATM 11 H * 0 -2.165 1.250 0.000 0.00 0.00
HETATM 12 H * 0 0.000 -2.500 0.000 0.00 0.00
HETATM 13 H * 0 2.165 -1.250 0.000 0.00 0.00
CONNECT 1 9 2 8
CONNECT 2 3 4
CONNECT 3 10
CONNECT 4 11 5
CONNECT 5 6 7
CONNECT 7 12 8
CONNECT 8 13
END
```

Obrázek č. 7: Ukázka struktury PDB souboru.

3D struktury molekul a jejich celé databáze se pak často ukládají jako Molfile (MDL) a SD formát (SDF). Struktura MDL souboru začíná hlavičkou, kde je název molekuly, na druhém řádku může být napsán uživatel, typ použitého programu, datum či podobné

informace. Třetí řádek hlavičky pak může obsahovat komentář. Za hlavičkou následuje tabulka atomů a vazeb. První řádek této části popisuje celkovou informaci, kolik máme atomů, vazeb apod. Dále už jsou popsány jednotlivé atomy, na řádku vždy jeden: jejich souřadnice  $x$ ,  $y$  a  $z$ , o jaký prvek jde a mohou zde být i další informace. Obdobně je to s vazbami, kde každý řádek popisuje jednu vazbu, číslo první molekuly ve vazbě, číslo druhé molekuly ve vazbě a typ této vazby. Opět i zde mohou být doplňující informace. Dále může být blok vlastností následovaný slovíčkem END značícím konec zápisu molekuly. Toto je charakter MDL souboru, který popisuje jednu molekulu. SDF soubory (obrázek č. 8) mají tu výhodu, že v sobě umí zahrnout i doplňující informace o molekulách, které chceme v databázi mít uloženy ve formě XML zápisu pomocí řádkové notace „<název vlastnosti>hodnota vlastnosti“ a navíc umí takto zapsaných molekul pojmout více (jednotlivé molekuly jsou odděleny symbolem \$\$\$\$).

```

molecule_1      2D
13 13 0 0 0 0 0 0 0 0999 v2000
  1.2300  0.7100  0.0000 C  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
  2.1650  1.2500  0.0000 H  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
  0.0000  1.4200  0.0000 C  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
  0.0000  2.7960  0.0000 O  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 -0.8820  3.2070  0.0000 H  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 -1.2300  0.7100  0.0000 C  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 -2.1650  1.2500  0.0000 H  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 -1.2300  -0.7100  0.0000 C  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 -2.7200  -1.5700  0.0000 Cl 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
  0.0000  -1.4200  0.0000 C  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
  0.0000  -2.5000  0.0000 H  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
  1.2300  -0.7100  0.0000 C  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
  2.1650  -1.2500  0.0000 H  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
  1  2  1  0  0  0  0
  1  3  1  0  0  0  0
  1 12  2  0  0  0  0
  3  4  1  0  0  0  0
  3  6  2  0  0  0  0
  4  5  1  0  0  0  0
  6  7  1  0  0  0  0
  6  8  1  0  0  0  0
  8  9  1  0  0  0  0
  8 10  2  0  0  0  0
 10 11  1  0  0  0  0
 10 12  1  0  0  0  0
 12 13  1  0  0  0  0
M END
> <Name>molecule_1
> <rgyr>2.1339495
> <weight>128.558
$$$$

```

```

MOE2009      2D
12 12 0 0 0 0 0 0 0 0999 v2000
  1.2300  0.7100  0.0000 C  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
  2.1650  1.2500  0.0000 H  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
  0.0000  1.4200  0.0000 C  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
  0.0000  2.7620  0.0000 F  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

```

Obrázek č. 8: Ukázka struktury SDF souboru (totožná molekula jako v případě ukázky PDB souboru – obrázek č. 7).

Před samotnou tvorbou databáze je nutné nejprve zanést do počítače 3D strukturu molekul. K tomu je nejjednodušší využít softwarů, pomocí nichž se dají molekuly kreslit. Ty jsou většinou dělány tak, že práce s nimi je velmi jednoduchá a stačí běžná znalost chemie. Můžeme zde tak spojovat prvky jednoduchými i násobnými vazbami. Stejně tak můžeme zadat široké spektrum jiných vlastností, náboj daného atomu či vazebný úhel. Protože však všechny údaje neznáme, je často zabudován také šikovný nástroj pro minimalizaci energie celé molekuly, což nám umožní najít konfiguraci, ve které se nejspíš běžně vyskytuje (díky své energeticky nejmenší náročnosti).

Pro tuto práci jsme si vybrali program MOE, ve kterém jsme dané molekuly kreslili s využitím minimalizace energie MMFF94X. Tento program umožňuje danou molekulu ihned zapsat do databáze s nejrůznějšími charakteristikami, jako je název molekuly, citace a také jsme zapisovali získaná experimentální data, buďto  $IC_{50}$ , které jsme následně pomocí kalkulátoru v editoru přepočítali na  $pIC_{50}$  pro sjednocení dat, a nebo jsme čerpali z článků právě hodnoty  $pIC_{50}$ . Pomocí deskriptorů jsme pak o molekulách získali přesnější data k vyhodnocení.

### 3.1 Deskriptory

Molekulární deskriptory se využívají v chemii, ve farmacii, v ochraně životního prostředí i při mnohých dalších výzkumech. Je to způsob, jak skutečné molekuly převést v čísla tak, aby charakterizovaly nějakým způsobem chemické vlastnosti dané molekuly a zároveň se daly využít pro matematické výpočty. Deskriptory se pak dají využít k předpovědi biologických, farmaceutických, chemických, fyzikálních vlastností z molekulární struktury, tedy při tvorbě QSAR modelů. Využitím v databázi pak slouží k rozpoznání důležitých vlastností, které ovlivňují, zda je daná molekula kupříkladu vhodným inhibitorem či jinak zvýhodněnou molekulou. Řešením QSAR problému je pak sestavení rovnice, která využívá jednotlivých hodnot deskriptorů, tak aby co nejlépe odpovídala výchozím datům a dala se využít pro predikci neznámých hodnot (Visco Jr., D. P. et al., 2002).

Jedním z důležitých deskriptorů je elektrický **dipólový moment** (v databázi pod kódem AM1\_dipole). Je to vektorová veličina popisující míru separace pozitivních a negativních elektrických nábojů (asymetrie elektrického náboje). Jednotkou je nejčastěji Debye – D. Pokud bychom jej chtěli převést na jednotky soustavy SI, pak platí, že  $1D = 3,33564 \cdot 10^{-30} \text{C} \cdot \text{m}$ .

Další dva deskriptory popisují plochu povrchu. Jeden z nich udává **velikost hydrofobní části povrchu** (kód ASA\_H), druhý udává **velikost polární plochy povrchu** (kód ASA\_P), oboje v jednotkách Å<sup>2</sup> (1Å = 0,1nm).

Využit by se dal také deskriptor pojmenovaný **Lipinski druglike test** (kód lip\_druglike), vysvětlení můžeme také najít pod názvem Lipinsky rule of five. Toto pravidlo znamená v praxi, že molekuly jsou špatně absorbovány nebo propouštěny, pokud

- 1) mají více než 5 „dárců“ vodíku,
- 2) molekulární hmotnost je větší než 500,
- 3) logP(o/w) je větší než 5,
- 4) součet atomů dusíku a kyslíku je větší než 10

V databázi je tedy možno najít v kolonce tohoto deskriptoru buďto 1 – molekula je absorbovatelná, a tedy nejspíše použitelná jako léčivo; nebo 0 – molekula nemůže být absorbována, tedy nejspíše nepůjde použít jako léčivo.

Následuje deskriptor značený jako **logP(o/w)** (kód také logP(o/w)), což je partiční koeficient oktanol-voda. Čím vyšší hodnotu tento koeficient má, tím větší je pravděpodobnost, že se molekula bude rozpouštět v oktanolu. Pravděpodobnost rozpustnosti ve vodě s rostoucím koeficientem klesá.

Dalším deskriptorem, který je dobré využít je, **gyrační poloměr** (kód rgyr). Tato veličina souvisí s velikostí objektu či skupiny bodů a jejich rozložením v prostoru. Tento poloměr vypočítáme jako střední kvadratickou vzdálenost částí objektu od těžiště, případně od dané osy.

Naším posledním důležitým deskriptorem je **molární hmotnost** (kód Weight), což je hmotnost jednoho molu dané látky.

Jako další můžeme uvést deskriptory, které uvádí počet různých atomů v molekule. Ať už se jedná o celkový počet atomů či počet atomů určitého prvku. Deskriptory však počítají i počet kyselých atomů nebo počet těžkých atomů. Možností je opravdu mnoho.



## 4. Výsledky

Vytvořili jsme celkem 4 databáze na základě rešerše (viz elektronická příloha), a to pro CDK1, CDK2, CDK4 a CDK5. Databáze pro CDK4 a CDK5 nejsou rozsáhlé a jejich výsledky jsou spíše doplňující. Pro CDK1 se podařilo vytvořit již rozsáhlejší databázi, se kterou jsme provedli QSAR analýzu.

### 4.1 QSAR model pro CDK1 inhibitory

Nejlepších výsledků jsme dosáhli zapojením téměř všech zvolených deskriptorů. Nedůležitý pro model byl deskriptor Lipinského pravidlo pěti, který nám vždy vyhodnotil molekuly vhodné jako léčivo. Dalším nepotřebným ukazatelem byl počet kyselých atomů, protože se v molekulách žádné kyselé atomy nevyskytovaly. Korelační koeficient  $r^2$  tak dosáhl hodnoty 0,655. Nejvíce se uplatnily deskriptory pro plochu hydrofobního a polárního povrchu, celkový počet atomů, gyrační poloměr a také počet akceptorů vazby s vodíkem (kód a\_acc). Rovnice pro tento lineární model je na obrázku číslo 9.

#### ESTIMATED LINEAR MODEL

```
pIC50 (CDK1) =  
  4.32772  
+0.15393 * AM1_dipole  
-0.03790 * ASA_H  
-0.04172 * ASA_P  
+0.36236 * logP(o/w)  
+3.00667 * rgyr  
-0.00189 * Weight  
+0.51399 * a_acc  
+0.06507 * a_aro  
-0.10259 * a_base  
+0.21120 * a_count  
-0.14506 * a_don
```

*Obrázek č. 9: Výsledná rovnice modelu pro CDK1.*

Relativní důležitost je znázorněna na obrázku číslo 10. Čím blíže je hodnota u daného deskriptoru hodnotě 1, tím více se uplatní tento deskriptor v QSAR modelu.

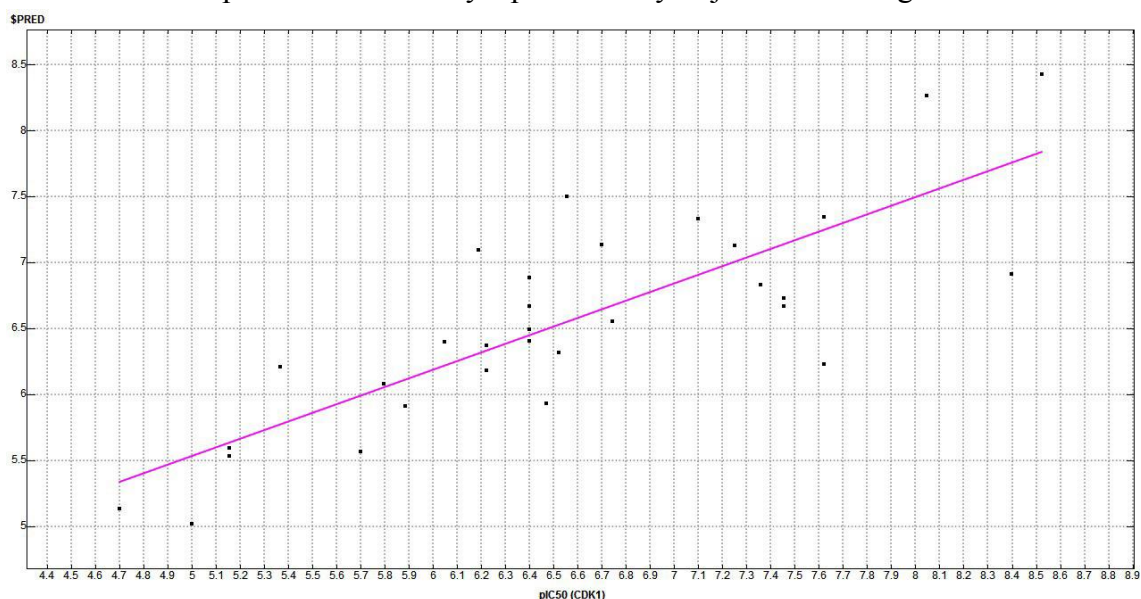
#### RELATIVE IMPORTANCE OF DESCRIPTORS

```

0.181811 AM1_dipole
1.000000 ASA_H
0.832137 ASA_P
0.140419 logP(o/w)
0.425597 rgyr
0.037657 Weight
0.275375 a_acc
0.034824 a_aro
0.014586 a_base
0.629358 a_count
0.039711 a_don
    
```

Obrázek č. 10: Relativní důležitost deskriptorů pro QSAR model pro CDK1 inhibitory.

Korelace mezi experimentálními daty a predikovanými je znázorněna grafem č. 1.



Graf č. 1: Korelace mezi experimentálními daty (osa x) a predikovanými daty (osa y) pro CDK1 inhibitory.

Protože výpočet všech deskriptorů může být obtížný, byl vytvořen model s šesti nejdůležitějšími deskriptory. Jeho korelační koeficient  $r^2$  byl o něco nižší, 0,603. Výsledný model a důležitost deskriptorů je na obrázku číslo 11.

#### ESTIMATED LINEAR MODEL

```

pIC50 (CDK1) =
  2.94101
-0.02678 * ASA_H
-0.02761 * ASA_P
+2.29353 * rgyr
+0.19814 * a_count
+0.14581 * AM1_dipole
+0.20938 * logP(o/w)
    
```

#### RELATIVE IMPORTANCE OF DESCRIPTORS

```

1.000000 ASA_H
0.779280 ASA_P
0.459431 rgyr
0.835574 a_count
0.243705 AM1_dipole
0.114819 logP(o/w)
    
```

Obrázek č. 11: QSAR model pro CDK1 inhibitory s šesti deskriptory; relativní důležitost těchto deskriptorů.

Pokud bychom chtěli model jen s pěti deskripty, odebrali bychom nejspíše partiční koeficient oktanol-voda, avšak korelační koeficient  $r^2$  by klesl na hodnotu 0,577. Tato hodnota je však již nízká v porovnání s počátečním korelačním koeficientem.

Nejúčinnějšími inhibitory pak byly Purvalanol A (Kryštof, V. & Strnad, M., 2001), Pies\_22, Pies\_8 a Pies\_9 (Pies, T. et al., 2004) a také Alsterpaullon (Kryštof, V. & Strnad, M., 2001). Celkový počet atomů kolísal, avšak aromatických atomů bylo vždy 15 (výjimka alsterpaullon – 12), ostatní deskripty se také pohybovaly v průměrných hodnotách v širším rozptylu.

## 4.2 QSAR model pro CDK2 inhibitory

Největší databáze pak byla pro CDK2, kde analýza dopadla velmi podobně. Deskriptor Lipinského pravidlo pěti zde opět nebyl zapotřebí (vždy hodnota 1), stejně tak počet kyselých atomů byl vždy 0. Ostatní deskripty zpřesňovaly korelační koeficient  $r^2$ , který dosáhl hodnoty 0,653, což je nepatrně nižší v porovnání s hodnotou v předešlém modelu. Nejužitečnější zde byl deskriptor pro molární hmotnost. Dále pak gyrační poloměr, celkový počet atomů, počet akceptorů vazby s vodíkem, plocha polárního povrchu a plocha hydrofobního povrchu. Výslednou rovnici tohoto modelu lze najít na obrázku číslo 12.

### ESTIMATED LINEAR MODEL

```
pIC50 (CDK2) =  
  1.54546  
+0.01993 * AM1_dipole  
-0.00350 * ASA_H  
-0.00517 * ASA_P  
+0.17222 * logP(o/w)  
+0.74795 * rgyr  
+0.01441 * Weight  
+0.30545 * a_acc  
-0.01873 * a_aro  
+0.38202 * a_base  
-0.05718 * a_count  
-0.00931 * a_don
```

*Obrázek č. 12: Výsledná rovnice modelu pro CDK2*

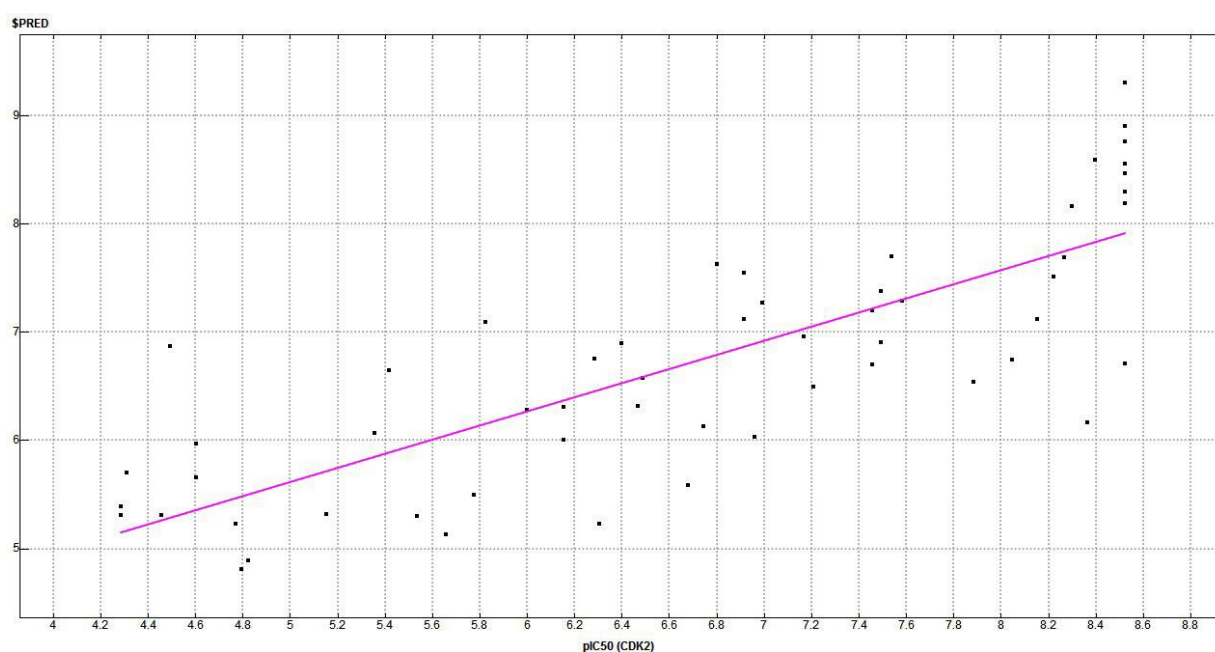
Důležitost jednotlivých deskriptorů je pak patrná na obrázku číslo 13.

#### RELATIVE IMPORTANCE OF DESCRIPTORS

0.121212	AM1_dipole
0.230037	ASA_H
0.246466	ASA_P
0.208735	logP(o/w)
0.440612	rgyr
1.000000	Weight
0.362771	a_acc
0.041725	a_aro
0.166892	a_base
0.433711	a_count
0.011911	a_don

Obrázek č. 13: Relativní důležitost deskriptorů pro QSAR model pro CDK2 inhibitory.

Graf číslo 2 zobrazuje korelaci mezi experimentálními daty pro CDK2 a predikcí.



Graf č. 2: Korelace mezi experimentálními daty (osa x) a predikovanými daty (osa y) pro CDK2 inhibitory.

I zde jsme vytvořili model, tentokrát s pěti deskriptory. Korelační koeficient  $r^2$  tak klesl na 0,631, což je ale uspokojivá hodnota v porovnání s původním korelačním koeficientem. Rovnice tohoto modelu a relativní důležitost deskriptorů je na obrázku č 14.

ESTIMATED LINEAR MODEL	RELATIVE IMPORTANCE OF DESCRIPTORS
pIC50 (CDK2) =	1.000000 Weight
1.32254	0.181514 rgyr
+0.01532 * Weight	0.429692 a_count
+0.32764 * rgyr	0.358102 a_acc
-0.06024 * a_count	0.080226 ASA_P
+0.32061 * a_acc	
-0.00179 * ASA_P	

*Obrázek č. 14: QSAR model pro CDK2 inhibitory s pěti deskriptory; relativní důležitost těchto deskriptorů.*

Velmi dobré výsledky pro inhibici CDK2 dosáhly molekuly Purvalon B, NU6102 a molekuly označené jako Babu\_10-18, Babu\_26, Babu\_27, Babu\_28, Babu\_37 (Babu, P. A. et al., 2007). Z ostatních zdrojů pak vynikala molekula označená Alzate-Morales\_2 (Alzate-Morales, J. H. et al., 2007).

#### 4.2.1 Rozpětí deskriptorů v databázi pro CDK2

Abychom provedli ještě lepší analýzu využití deskriptorů, vytvořili jsme tabulku pro extraúčinné inhibitory ( $pIC_{50} > 8$ , sub10nM), účinné inhibitory ( $pIC_{50} > 6$ , sub $\mu$ M) (tabulka č. 1) a porovnali jsme rozsah hodnot deskriptorů v těchto skupinách s hodnotami pro všechny inhibitory.

CDK2	pIC <sub>50</sub> > 8			pIC <sub>50</sub> > 6			všechny		
	min	med	max	min	med	max	min	med	max
AM1_dipole [D]	0	15	29	0	6	29	0	4	29
ASA_H [Å <sup>2</sup> ]	353	<b>485</b>	537	286	<b>282</b>	579	277	<b>431</b>	579
ASA_P [Å <sup>2</sup> ]	116	<b>236</b>	277	65	<b>185</b>	277	65	<b>166</b>	277
logP(o/w)	0,37	<b>1,55</b>	6,36	0,37	<b>2,79</b>	6,36	0,37	<b>2,41</b>	6,36
Rgyr [Å]	3,44	<b>3,84</b>	3,84	3,35	<b>4,38</b>	5,51	3,30	<b>3,96</b>	5,51
Weight [g·mol <sup>-1</sup> ]	323	<b>447</b>	467	261	<b>405</b>	504	220	<b>360</b>	504
a_acc	4	<b>6</b>	8	1	<b>5</b>	8	1	<b>5</b>	8
a_aro	15	<b>17</b>	20	9	<b>17</b>	20	9	<b>15</b>	20
a_base	0	<b>1</b>	2	0	<b>0</b>	2	0	<b>0</b>	2
a_count	40	<b>52</b>	61	31	<b>45</b>	61	30	<b>43</b>	61
a_don	2	<b>4</b>	7	1	<b>4</b>	7	0	<b>4</b>	7

Tabulka č. 1: Minimální, maximální hodnoty a medián deskriptorů pro extraúčinné ( $pIC_{50} > 8$ , sub10nM), účinné inhibitory ( $pIC_{50} > 6$ , sub $\mu$ M) a pro všechny testované látky.

U dipólového momentu, partičního koeficientu oktanol-voda a počtu bazických atomů (a\_base) se hodnoty deskriptorů u účinných nebo extraúčinných inhibitorů nijak výrazně neliší. Celková plocha hydrofobního a také polárního povrchu je však u extraúčinných látek vyšší. Naopak gyrační poloměr se u extraúčinných látek zmenšuje, což naznačuje, že jde o látky spíše kompaktní. Molární hmotnost se pohybuje ve vyšších hodnotách mezi přibližně 320 a 470 g·mol<sup>-1</sup>, což je v užším rozpětí v porovnání se všemi molekulami a opět jde spíše o větší molekuly, podobně jako tomu bylo u obou povrchových deskriptorů. Počet akceptorů vazby s vodíkem je u extraúčinných inhibitorů 4 a více, aromatických atomů je v této skupině také více v porovnání se všemi molekulami, a to od 15 do 20. Celkový počet atomů u extraúčinných inhibitorů je spíše vyšší, je zde i více donorů vazby s vodíkem (a\_don). Všechny tyto odlišné vlastnosti nám mohou pomoci při hledání nových potenciálních inhibitorů.

## 5. Diskuze

Cyklin-dependentní kinázy řídí buněčný cyklus a tím ovlivňují dělení buňky. Jejich nadměrné působení může mít následky pro celý organismus, ať už se jedná o různá onemocnění jako je rakovina, virová infekce či jiné neurodegenerativní choroby (Knockaert, M., 2002). Výzkum CDK je proto více než vhodnou cestou, jak tyto nemoci léčit. *In silico* metody se zdají být nejlevnějším a také nejrychlejším způsobem řešení. Jejich význam se zlepšující se technikou a dostupností těchto metod stoupá. Přestože *in silico* metody se dosti různí a mohou být postaveny jak na struktuře ligandů, tak i na stavbě receptoru, jejich cíl je v tomto případě stejný. Snahou je vždy nalézt co nejúčinnější inhibitory CDK a zjistit společné vlastnosti těchto inhibitorů pro hledání dalších účinných látek.

Právě pro srovnání různých vlastností inhibitorů je dobré vytvořit databáze, kde můžeme ligandy zaznamenávat pomocí 3D zápisu. Ten je nej přesnější a zachovává tak nejlepší informace o inhibitech. Využitím deskriptorů pak získáme přehled o společných, resp. různých vlastnostech inhibitorů.

Na základě rešerše a získání experimentálních dat byla připravena tedy databáze ligandů pro CDK1, CDK2, CDK4 a CDK5. Z QSAR analýz pro CDK1 a CDK2 byly určeny důležité deskriptory. V případě CDK1 šlo o celkovou plochu hydrofobního a polárního povrchu, celkový počet atomů v molekule, gyrační poloměr a o počet akceptorů vazby s vodíkem. Pro QSAR model pro CDK2 inhibitory byly důležité deskriptory s jinou prioritou, a to nejprve molární hmotnost, dále pak gyrační poloměr, celkový počet atomů v molekule, počet akceptorů vazby s vodíkem a částečně i celková plocha hydrofobního a polárního povrchu. To se nám podařilo potvrdit i podle rozpětí hodnot těchto deskriptorů u extraúčinných inhibitorů v porovnání se všemi molekulami.

Z důležitosti deskriptorů u jednotlivých CDK lze vidět, že nejsou důležité vždy stejné deskriptory. Tento fakt by mohl pomoci při hledání více selektivních inhibitorů.

## 6. Závěr

Byla provedena rešerše a získány experimentální data. Z těch jsme sestavili pomocí programu MOE databáze inhibitorů pro CDK1, CDK2, CDK4 a CDK5. K databázím pro CDK1 a CDK2 jsme dále vytvořili QSAR model s dobrými korelačními koeficienty. Pro naši největší databázi inhibitorů CDK2 jsme dále provedli analýzu rozpětí deskriptorů pro účinné a extraúčinné inhibitory. Tato analýza potvrdila relativní důležitost jednotlivých deskriptorů z QSAR analýzy. Databáze ligandů pro CDK1, CDK2, CDK4 a CDK5 mohou být použity pro následné další *in silico* zpracování a návrh dalších vhodných inhibitorů.



## Literatura

**Alzate-Morales, J. H. et al. (2007):** A Computational Study of the Protein-Ligand Interactions in CDK2 Inhibitors: Using Quantum Mechanics/Molecular Mechanics Interaction Energy as a Predictor of the Biological Activity. *Biological Journal* 92: 430-439.

**Babu, P. A. et al. (2007):** Pyridines, pyridazines and guanines as CDK2 inhibitors: a review. *ARKIVOC* 2: 247-265.

**Bernstein, F. C. et al. (1977):** The protein data bank: A computer-based archival file for macromolecular structures. *Journal of Molecular Biology* 112: 535-542.

**Daniel, L. (2009):** Studium substrátové specifity halogenalkan dehalogenáz metodou virtuálního sreeningu. Bakalářská práce. Masarykova univerzita, Brno, Přírodovědecká fakulta, Katedra biologie.

**Daniel, L. (2011):** In silico study of substrate specificity of halogenalkane dehalogenases. Diplomová práce. Masarykova univerzita, Brno, Přírodovědecká fakulta, Katedra biochemie.

**Dong, X. et al. (2012):** Pharmacophore identification, docking and „in silico“ screening for novel CDK1 inhibitors. *Journal of Molecular Graphics and Modelling* 37: 77-86.

**Ekins, S. et al. (2007):** In Silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. *British Journal of Pharmacology* 152: 9-20.

**Jašková, E. (2010):** Inhibitory cyklin dependentních kináz a transkripční stres. Diplomová práce. Univerzita Palackého v Olomouci, Přírodovědecká fakulta, Katedra buněčné biologie a genetiky.

**Knockaert, M. et al. (2002):** Pharmacological inhibitors of cyclin-dependent kinases. *TRENDS in Pharmacological Sciences* 23: 417-425.

**Kryštof, V. & Strnad, M. (2001):** Inhibitory cyklin-dependentních kinas. *Chemické listy* 95: 295-300.

**Laha, J. K. et al. (2011):** Structure-activity relationship study of 2,4-diaminothiazoles as CDK5/p25 kinase inhibitors. *Bioorganic & Medicinal Chemistry Letters* 21: 2098-2101.

**Otyepka, M.** (2006): Proteiny pod drobnohledem metod výpočetní a teoretické chemie. Habilitační práce z fyzikální chemie. Univerzita Palackého v Olomouci, Přírodovědecká fakulta, Katedra fyzikální chemie.

**Pies, T. et al** (2004): CDK1-Inhibitory Activity of Paullones Dependends on Electronic Properties of 9-Substituents. Arch. Pharm. Pharm. Med. Chem. 337: 486-492.

**Richter, J.** (2013): Interaktivní vizualizace 3D molekulárních motivů. Bakalářská práce. Masarykova univerzita, Brno, Fakulta informatiky.

**Ronald, M. A. et al.** (1997): Molecular docking to ensembles of Protein Structures. Journal of Molecular Biology 266: 424-440.

**Visco Jr., D. P. et al.** (2002): Developing a methodology for an inverse quantitative structure-activity relationship using the signature molecular descriptor. Journal of Molecular Graphics and Modelling 20: 429-438.

## **Elektronické přílohy**

Součástí práce jsou i elektronické přílohy na CD ROM, které obsahují databázi pro CDK1, CDK2, CDK4 a CDK5 ve formátu SDF.

## Seznam zkratek

CDK	cyklin-dependentní kináza/y
QSAR	quantitative structure activity relationship
RNA	ribonukleová kyselina
ATP	Adenosintrifosfát
HSV	herpes simplex virus
DNA	deoxyribonukleová kyselina
CAK	CDK-aktivující kinázy
IC <sub>50</sub>	inhibiční koncentrace pro 50% dané látky
QM/MM	kvantová mechanika / molekulární mechanika
CKI	inhibitor/y cyklin-dependentních kináz
pIC <sub>50</sub>	Odpovídá hodnotě $-\log IC_{50}$
PDB	Protein Data Bank
MOE	Molecular Operating Environment, software
MDL	Molecular Design Limited, zkratka molfile souboru
SDF	Structure-data format
XML	Extensible Markup Language, značkovací jazyk
AM1_dipole	dipólový moment
ASA_H	plocha hydrofobního povrchu
ASA_P	plocha polárního povrchu
lip-druglike	Lipinského pravidla pěti
logP(o/w)	partiční koeficient oktanol-voda
rgyr	gyrační poloměr
Weight	molární hmotnosti
a_acc	počtu akceptorů vazby s vodíkem
a_acid	počet kyselých atomů
a_aro	počet aromatických atomů
a_base	počet bazických atomů
a_count	Celkový počet atomů
a_don	počtu donorů vazby s vodíkem