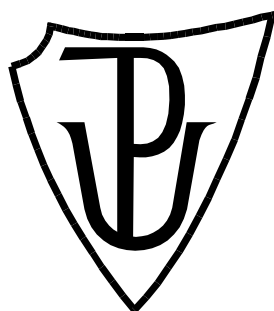


UNIVERZITA PALACKÉHO V OLOMOUCI

Přírodovědecká fakulta

Katedra fyzikální chemie



Příprava a validace rozsáhlých chemických knihoven

BAKALÁŘSKÁ PRÁCE

Autor: David Řepka

Studijní program: B1407 Chemie

Studijní obor: Aplikovaná chemie

Forma studia: Prezenční

Vedoucí práce: Ing. Mgr. Václav Bazgier, Ph.D.

Rok: 2019

Prohlašuji, že jsem bakalářskou práci vypracoval samostatně s použitím odborné literatury a zdrojů, uvedených v seznamu použité literatury.

V Olomouci dne podpis

Rád bych tímto poděkoval vedoucímu mé bakalářské práce panu Ing. Mgr. Václavu Bazgierovi, Ph.D. za odborné vedení a cenné připomínky. Také bych rád poděkoval panu Bc. Filipu Řepkovi za neocenitelné technické rady při vývoji webové aplikace.

Bibliografická identifikace

Jméno a příjmení autora:	David Řepka
Název práce:	Příprava a validace rozsáhlých chemických knihoven
Typ práce:	Bakalářská práce
Pracoviště:	Katedra fyzikální chemie
Vedoucí práce:	Ing. Mgr. Václav Bazgier, Ph.D.
Rok obhajoby práce:	2019
Klíčová slova	Chemická knihovna, databáze, fyzikálně-chemické vlastnosti, Qikprop
Počet stran	41
Počet příloh	4
Jazyk	český

Abstrakt

Pro vytvoření nových léčiv je potřeba dostatek informací o daných sloučeninách. Dnešní databáze obsahují jen omezený počet fyzikálně-chemických vlastností. Knihovna vytvořená v rámci bakalářské práce si klade za úkol tento nedostatek odstranit. Při predikci a výpočtu vlastností látek byl použit program Qikprop. Následná validace dat je založena na experimentálních hodnotách rozpustnosti ve vodě, rozdělovacím koeficientu oktanol/voda a molekulové hmotnosti. Knihovna je dostupná na adrese <http://pokusnice.upol.cz/bc/>.

Bibliographical identification

Autor's first name and surname:	David Řepka
Title	Design and Validation of Large Chemical Libraries
Type of thesis	Bachelor thesis
Department	Department of Physical Chemistry
Supervisor	Ing. Mgr. Václav Bazgier, Ph.D.
The year of presentation	2019
Keywords	Chemical library, database, physical-chemical properties, Qikprop
Number of pages	41
Number of appendices	4
Language	Czech

Abstract

Plenty of information is needed for the purposes of new drug development. Current databases contain only limited number of physical-chemical properties. Goal of the library created in this bachelor thesis is to eliminate the lack of properties. Qikprop software was used for prediction and calculation of physical-chemical properties. Following data validation is based on experimentally obtained data of water solutability, octanol/water partition coefficient and molecular weight. Library is freely accessible at <http://pokusnice.upol.cz/bc/>.

Obsah

1. Úvod	10
2. Teoretická část	11
2.1 Chemické knihovny	11
2.2 Kombinatorní chemie.....	11
2.3 Vysokokapacitní screening	12
2.3.1 Miniaturizace a automatizace metody	13
2.4 Virtuální screening	13
2.4.1 Ligandové metody	16
2.4.1.1 Podobnostní metoda	16
2.4.1.2 QSAR – kvantitativní vztah struktura-účinnost	16
2.4.2 Strukturní metody	16
2.4.2.1 Molekulové dokování	16
2.5 Program Qikprop	18
2.5.1 Popis klíčových vlastností	18
2.5.1.1 Molekulová hmotnost	18
2.5.1.2 Rozpustnost	19
2.5.1.3 Rozdělovací koeficient oktanol/voda	20
3. Praktická část	21
3.1 Použité databáze	21
3.1.1 Databáze Drugbank	21
3.2 Výpočet vlastností pomocí qikprop	21
3.2.1 Kritéria potřebná pro výpočet	21
3.2.2 Qikprop metodologie	24
3.2.3 zlepšení log S a log P vlastností	25
3.2.4 Problém přesnosti výpočtu vlastností	25
3.3 Databáze.....	26

3.3.1 Relační databáze	26
3.4 Webová aplikace	26
3.5 Návod k aplikaci	27
3.6 Experimentální data z databází	28
3.7 Vyhodnocení	29
3.7.1 Rozpustnost	29
3.7.2 Rozdělovací koeficient oktanol/voda	30
4. Závěr	31
5. Summary	32
6. Použitá literatura	33

Seznam zkratek

ADMET	absorpce, distribuce, metabolismus, vylučování a toxicita léčiva
API	Aktivní farmaceutická složka
AUC	Plocha pod křivkou
CSS	Kaskádové styly
CSV	Hodnoty oddělené čárkou
DMSO	Dimethylsulfoxid
FDA	Úřad pro kontrolu potravin a léčiv (z angl. Food and Drug Administration)
HTML	Hypertextový značkovací jazyk (z angl. Hypertext Markup Language)
kDa	Kilodalton
Log BB	Rozdělovací koeficient pro mozek/krev
LogP	Rozdělovací koeficient
Log PMDCK	Logaritmus propustnosti Madin-Darby psích ledvinových buněk
LogP o/w	Rozdělovací koeficient pro oktanol/vodu
Log S	Rozpustnost ve vodě
MSDS	Bezpečnostní list (z angl. Material Safety Data Sheet)
PCaco	Předpovězená propustnost Caco 2 buněk
PDB	Program data base
PHP	PHP: hypertextový preprocesor

QP	Qikprop
QSAR	Kvalitativní vztah struktura-účinnost (z angl. Quantitative Structure-Activity Relationship)
R^2	Koeficient determinace
SASA	Celkový přístupný povrch pro rozpouštědlo
SQL	standardizovaný strukturovaný dotazovací jazyk (z angl. Structured Query Language)
VKS	Vysokokapacitní screening
μ VKS	Mikro vysokokapacitní screening
Å	Ångström

1. Úvod

Hledání léčiv a jejich následná výroba byl vždy a stále ještě je zdlouhavý proces. Vývoj nového léčiva od začátku po jeho uvedení na trh zabere 12-15 let a zhruba 1 miliardu dolarů (Hughes *et al.*, 2011). K novému léčivu může vést buď náhoda, jak se stalo v případě objevení penicilinu nebo cílená snaha vědců jak ze státních institucí nebo v mnoha případech ze soukromých farmakologických firem. Dnešní společnost se nemůže spoléhat na náhodu a tak je vývoj a výzkum nových léčiv vědním odvětvím s velmi vysokými finančními obraty. Je to způsobeno velkými náklady na samotný vznik léčiva s dlouhou dobou vrácení původní částky investory. Metody vyvinuté na přelomu 80. a 90. let minulého století přinesly revoluci ve způsobu hledání nových léčiv. Jedná se především o dnes už samotné vědní odvětví kombinatorní chemie a z něj odvozené způsoby testování látek na aktivitu. Jeden z těchto způsobů se označuje jako vysokokapacitní screening (VKS) (z angl. High-Throughput screening) (Potáček, 2012). Druhý způsob je virtuální screening. Právě tyto metody využívají chemické knihovny ať experimentálního nebo virtuálního typu. Najít relevantní zdroj dat je pro chemika jedna ze zásadních věcí v chemoinformaticce. Mnohé chemické databáze mají zavádějící informace o chemikáliích, které se neslučují s realitou.

Ve své práci se zaměřím na problematiku dnešních volně dostupných chemických knihoven, a to především na jejich nedostatky. Výsledkem mé práce bude samostatně fungující online databáze chemických látek s napočítanými nebo předpovězenými vlastnostmi, které mnohdy chybí u dnes dostupných chemických databází. Vybrané napočítané vlastnosti také porovnáám s experimentálně zjištěnými daty pro dané látky. V teoretické části bakalářské práce nastíním možné využití chemické knihovny. V experimentální části se zaměřím na popis výpočtu a predikce vlastností. Nemalá část bude zaměřena také na vznik databáze a webové aplikace včetně validace dat.

2. Teoretická část

2.1 Chemické knihovny

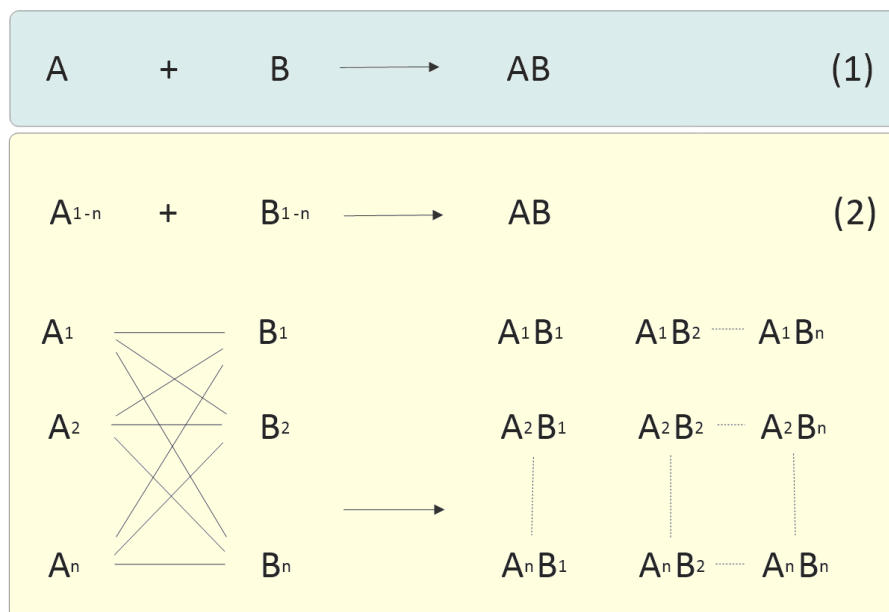
Chemické knihovny, často také označované jako chemické databáze, jsou sbírky sloučenin, které jsou uloženy buď fyzicky ve speciálních skladech nebo naopak virtuálně ve formě databází. Tyto virtuální databáze se dále mohou dělit na ty, které obsahují látky již syntetizované, nebo na ty, které obsahují virtuální chemické sloučeniny tzn. takové, které ještě nebyli připraveny v laboratoři. Právě fyzické chemické knihovny se podílí na výzkumu a vývoji nových léčiv pomocí vysokokapacitního screeningu (VKS). Typická virtuální knihovna měla ještě před pár lety několik milionů sloučenin. Dnešní knihovny se dostávají již k hranici 1 miliardy sloučenin (Irwin a Shoichet, 2005; Walters, 2018). Toto množství je ale stále zlomek v porovnání s předpokládaným množstvím možných sloučenin, které lze ještě vytvořit. Odhaduje se, že možných sloučenin je okolo 10^6 (*QikProp 3.5 User Manual*, 2012; Walters, 2018).

Jeden z největších problémů u chemických knihoven je objevování tzv. falešně pozitivních a falešně negativních chyb. Tyto chyby znamenají značné zkreslení výsledků. Jako falešně pozitivní označujeme takové molekuly, které jsou neaktivní, ale byla u nich předpovězena aktivita. Naopak falešně negativní molekuly jsou označeny za neaktivní i přes to, že ve skutečnosti aktivitu vykazují. V případě, kdy by měla knihovna 1 milion molekul a tyto chyby by se vyskytovaly v 1 % případů, což je optimistický odhad u dnešních používaných metod stanovení, pořád by se jednalo o 10 000 chybně stanovených molekul. Falešně pozitivní chyby mají z pohledu vědy menší váhu. Jelikož se takovéto látky ukážou jako aktivní na biologický cíl, prochází do fáze experimentální a musí se na ně vynaložit peníze na nákup, experimenty a čas potřebný k samotné analýze. Na druhou stranu látky, které vykazují falešnou negativitu, tedy počítačový screening je vyhodnotí jako neaktivní na biologický cíl, jsou vyřazeny a nedostanou se do experimentální fáze. Takové látky mohou být navždy ztraceny ve velkém objemu dat.

2.2 Kombinatorní chemie

Jedná se o spojení syntézy chemických sloučenin. Základním principem kombinatorní chemie je účinná paralelní syntéza mnoha látek za potřeby méně reakčních kroků. Charakteristika tohoto odvětví je taková, že při syntéze mnoha různých látek, které jsou

systematicky vytvářeny ve stejný čas při stejných reakčních podmínkách, dochází ideálně ke vzniku produktů všech možných kombinací začínajících látek ve stejnou dobu.



Obrázek 1: Rozdíl kombinatorní a organické syntézy. Číslo 1 označuje organickou syntézu, kde reaguje látka A s činidlem B za vzniku nové sloučeniny AB, tedy klasická syntéza. Kombinatorní syntéza je označena číslem 2. Látka A_1 až A_n reaguje simultánně s činidlem B_1 až B_n , tedy podle platných kombinatorních principů [převzato z cit. (Jung, 1999)].

Sbírka takto syntetizovaných látek se označuje jako kombinatorní knihovna. Takto vytvořené látky mohou být dále použity pro hledání biologické aktivity ve vysokokapacitním screeningu.

2.3 Vysokokapacitní screening

Jedno z využití chemických knihoven, v tomto případě převážně fyzického typu, je vysokokapacitní screening. Jedná se o jednu z nejpoužívanějších metod výzkumu a vývoje nových léčiv. Metoda je založená na skenování velkého množství látek a jejich bioaktivity k různým biologickým cílům podle toho, na co se hledá dané léčivo. Pomocí VKS se skenují látky většinou v kapalně formě s přidavkem dimethylsulfoxidu (DMSO) na aktivitu vůči biologickému materiálu. Látka, která vykazuje požadovanou aktivitu se označuje jako tzv. „hit“ (Potáček, 2012). Nejedná se o látky, které by hned šlo použít jako léčivo, ale jde pouze o nějaké ukazatele, kterou cestou je možné se vydat. Tyto hity pak putují dále do laboratorního experimentování, kde se k látce přidávají různé funkční skupiny podle požadovaných vlastností. Ať už se jedná o potlačení nějakého nežádoucího účinku nebo naopak o jeho zvýraznění. Máme-li dostatek informací o cílovém biologickém materiálu, jeho úloze při vývoji určité nemoci a jeho prostorové uspořádání je možno VSK nahradit

screeningem „*in silico*“. U tohoto druhu screeningu se hledají pomocí softwaru takové sloučeniny, které mají odpovídající tvar a funkční skupiny, aby se mohli navázat do alosterického nebo aktivního místa v cílové molekule. Podrobnější popis bude vysvětlený v kap. 2.4.

Ze začátku 90. let bylo možné pomocí VKS denně prověřit tisíce sloučenin. Metoda se postupem času, a hlavně díky zlepšení metodiky práce a lepší automatizaci posunula a bylo s ní možné prověřit až 100 000 látek za den (Hertzberg a Pope, 2000). Metoda se označila μ VKS. Díky tomuto vylepšení bylo nutné provést úpravy v množství použitého činidla tak, aby metoda byla dále ekonomicky únosná.

2.3.1 Miniaturizace a automatizace metody

U metod VKS a hlavně u μ VKS jsou používány mikrotitrační destičky o velikosti 96, 384 nebo 1 536 jamek. Většina firem využívá velikost s 384 jamkami (*High Throughput Screening*). Automatizace je důležitý proces pro zrychlení a zpřesnění práce. Největší riziko je při manipulaci s kapalinami a jejich následný transport z mikrotitračních destiček. Tento transport je potřebný např. pro analýzu hmotnostní spektrometrickou metodou výsledných produktů. U moderních metod se využívá ultrazvuk pro vyražení kapek z požadované jamky přímo do analytického zařízení (Sinclair *et al.*, 2016). Tato manipulace se sloučeninami je daleko přesnější a rychlejší než pomocí mikropipet. Navíc takovýto přenos je zcela bezdotykový, což odstraňuje možné chyby a odpad při manipulaci. Díky tomuto systému můžeme také odvozovat koncentraci v jamkách na základě známe rychlosti zvuku skrz vodu a DMSO (*Automating Compound Management Systems*, 2007).

2.4 Virtuální screening

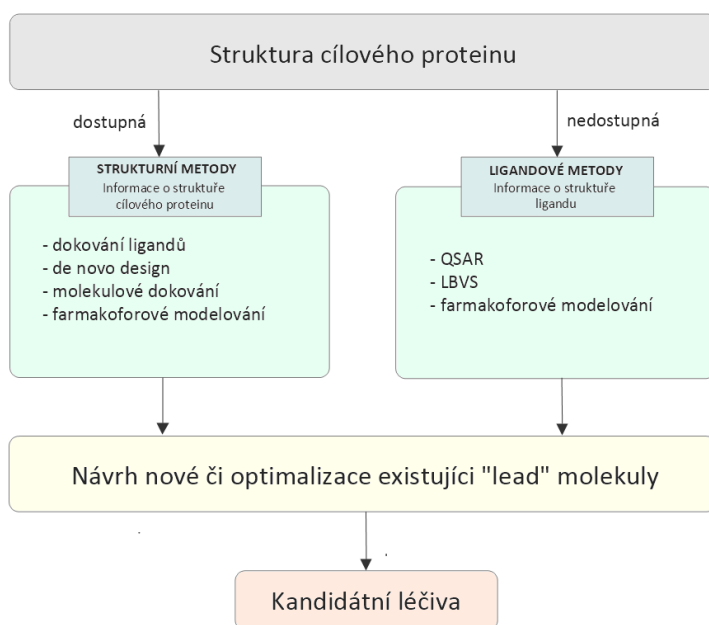
Virtuální screening je moderní výpočetní metoda pro vývin léčiv založená na práci s chemickými knihovnami obsahující malé molekuly. Při virtuálním screeningu se hledá látka, která je schopná vázat se na cíl léčiva. Většinou se jedná o proteinový receptor nebo enzym. Existuje několik přístupů k sestavení rozumně velkých knihoven pro virtuální screening. První a hlavní přístup je na základě předpokladu, že knihovny s největší možnou rozdílností látek nejsou rozumné pro praxi. Pokud vezmeme v potaz, že v tzv. chemickém prostoru je 10^{63} možných sloučenin a z toho jsme schopni syntetizovat pouze 10^6 látek, tak nemá smysl vytvářet knihovny s co největším možným počtem látek (Walters, Stahl a Murcko, 1998; Walters, 2018). Proto je potřeba knihovny, které jsou více zaměřené na

specifický problém, na látky, které je možné syntetizovat a látky podobné již známým léčivům. Specifickým problémem může být například vliv určitého druhu inhibitoru. Látkami, které lze syntetizovat se myslí takové, které vykazují vysoký výtěžek s reakčními činidly, lze je snadno připravit, popřípadě levně zakoupit. Poslední ze zmíněných jsou látky, které si jsou vlastnostmi podobné s léčivy, například, aby nebyly moc velké nebo moc lipofilní (Walters, Stahl a Murcko, 1998).

Množství sloučenin v knihovně, která by odpovídala výše popsaným předpokladům lze zjednodušeně určit. V chemické literatuře je zhruba 10 000 základních koster látek, které je možno docela jednoduše syntetizovat. Existuje jich mnohem více, ale ne všechny je tak jednoduché připravit. Pro porovnání, ve známých léčivech se nachází asi 2 500 různých základních koster. Po prozkoumání těchto koster se ukázalo, že průměrně 3 boční řetězce se vážou na základní kostru sloučeniny. Těchto bočních řetězců je přibližně 1 000 u známých léčiv. Výpočet sloučenin v knihovně je pak jednoduchý (Walters, Stahl a Murcko, 1998).

$$10\ 000\ \text{základních koster} \cdot (1000\ \text{bočních řetězců})^3 = 10^{13}\ \text{sloučenin}\ (1)$$

Existuje mnoho metod pro virtuální screening. Tyto metody jsou rozděleny podle požadovaných informací a času potřebného na výpočty. Základní dělení je podle dostupných informací o molekulárním cíli na ligandovou metodu a strukturní metodu.



Obrázek 2: Rozdělení virtuálního screeningu na základě struktury cílového proteinu [převzato z cit. (Koščová a Provazník, 2016)]

Na obrázku č. 2 je znázorněno dělení virtuálního screeningu na metodu ligandovou a strukturní podle dostupných informací. Pokud máme informace o podobě struktury cílové molekuly, jedná se o strukturní metodu. Zde se například řadí dokování ligandů, de novo desing, molekulové dokování a farmakoforové modelování. Naopak u ligandových metod neznáme strukturu cílové molekuly, ale známe strukturu ligandu. Zástupci ligandových metod jsou např. QSAR (kvalitativní vztah struktura-účinnost, z angl. Quantitative Structure-Activity Relationship), LBVS (z angl. Ligand Based Virtual Screening, zde podobnostní metoda) a farmakoforové modelování. Obě tyto metody, tedy ligandová a strukturní vedou k návrhu či optimalizaci molekul a následnému vzniku kandidátních léčiv.

Výsledkem virtuálního screeningu je nalezení bioaktivních sloučenin. Nalezení „hitu“ neznamená nalezení léčiva. Taková sloučenina má nespecifikovanou aktivitu vůči zkoumanému biologickému materiálu. Mimo jiné se pomocí počítačových programů vypočítávají ADMET vlastnosti, což je farmakologické označení pro absorpci, distribuci, metabolismus, vylučování a toxicitu léčiva v organismu. Vyznačují se specifitou k cíli tzn., že neovlivňují okolní cíle a nemají nežádoucí účinky (Koščová a Provazník, 2016).

Mimo této podmínky musí sloučenina mít i dobré výsledky v poločasu rozpadu v krevním řečišti. Do různých filtrů ve virtuálním screeningu se řadí i jedno z nejznámějších pravidel, a to Lipinského pravidlo pěti. Tímto pravidlem se posuzuje tzv. drug-likeness. Pokud látka splňuje všechny 4 Lipinského pravidla, znamená to větší šanci stát se léčivem, které vstupuje do organismu orálně. Tyto pravidla jsou:

1. Maximálně 5 donorů vodíkových vazeb (N-H, O-H)
2. Maximálně 10 akceptorů vodíkových vazeb (atomy dusíku a kyslíku)
3. Molekulová hmotnost < 500
4. Rozdělovací koeficient oktanol/voda $\log P < 5$ (Lipinski *et al.*, 2001).

Výjimky z pravidla jsou třídy sloučenin, které jsou substráty pro biologické transportéry. Lipinského pravidlo pěti vzniklo na základě porovnání několik tisíc léčiv. Název vznikl na základě toho, že hodnoty jednotlivých pravidel jsou násobky pěti. Toto pravidlo slouží převážně jako jednoduchý ukazatel na farmakokinetické vlastnosti látek při screeningu, protože se na něj nedá úplně spoléhat.

Pokud látka nespĺňuje alespoň jedno z pravidel, je považována ještě za možné orální léčivo. V případě, že poruší alespoň dvě pravidla, většinou se vyřazuje. Ne vždy, ale takové třídění platí. Zhruba 16 % léčiv porušují jedno z pravidel a 6 % porušuje dvě z pravidel, a i

přes to jsou vhodné jako orální léčiva (Bickerton *et al.*, 2012). Další rozšíření Lipinského pravidla pěti je Veberovo pravidlo. Veber studoval strukturní vlastnosti, které zvyšovaly orální bio-dostupnost u krys (Veber *et al.*, 2002). Jako důležité vlastnosti stanovil flexibilitu, polární povrch a počet vodíkových vazeb. Rozšířená pravidla jsou:

1. Počet rotačních vazeb ≤ 10
2. Polární povrch $\leq 140 \text{ \AA}^2$ nebo počet všech vodíkových vazeb ≤ 12

2.4.1 Ligandové metody

Jedná se o metody, kdy se zkoumají informace o aktivních nebo neaktivních látkách a jejich aktivitě na molekulový cíl. Mezi tyto metody se řadí např. podobnostní metoda nebo modelování kvantitativního vztahu mezi strukturou a aktivitou látky (QSAR, angl. Quantitative Structure-Activity Relationship) (Koščová a Provazník, 2016).

2.4.1.1 Podobnostní metoda

Tato metoda vyhledává látky z databáze, které jsou si podobné s již aktivními látkami. Základní předpoklad je, že takovéto látky budou vykazovat aktivitu již na základě podobnosti struktury, s čímž je spjata i podobnost fyzikálně-chemických vlastností. Pro fungování této metody je zapotřebí přítomnost alespoň nějakých aktivních látek (Svozil, 2017).

2.4.1.2 QSAR – kvantitativní vztah struktura-účinnost

Pomocí známé molekulární struktury se předpovídají vlastnosti jako např. toxicita. Předem určený parametr neboli vlastnost látky, se prostuduje u všech podobných látek v QSAR modelu. Na základě takto zjištěných informací se popíše vztah struktury podobných látek se zkoumanou vlastností na jejich aktivitě (Lionta *et al.*, 2014; Koščová a Provazník, 2016).

2.4.2 Strukturní metody

Jsou to metody, u kterých je známá 3D struktura molekulového cíle, nejčastěji se jedná o protein. Strukturní metody navrhují látky, které se co nejsilněji vážou do molekulárního cíle a mají lepší ADMET vlastnosti (Lionta *et al.*, 2014; Koščová a Provazník, 2016).

2.4.2.1 Molekulové dokování

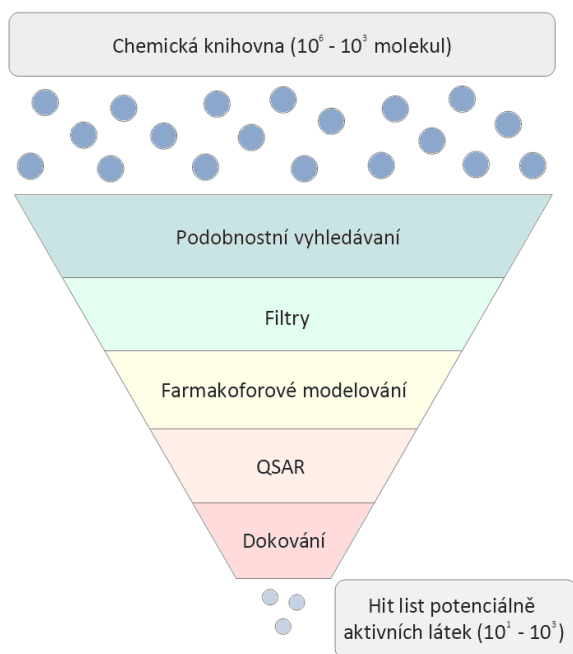
Jedna z neznámějších strukturních metod je molekulové dokování. Základem dokování je předpověď struktury komplexu ligand-receptor pomocí výpočetních metod (Meng *et al.*, 2011). Na základě známé struktury molekulového cíle se hledají látky z databáze, které jsou

schopné vázat se do tohoto cílového místa s dostatečnou specifitou a silou. Látky navázané v cílovém místě vyvolávají interakci s receptorem. Tento proces probíhá přirozeně v buňkách, mezi nejznámější můžeme řadit například reakci enzymu se substrátem za vzniku komplexu enzym-substrát.

Mezi obě tyto metody, tedy strukturní a ligandovou, lze zařadit farmakoforové modelování. Dle definice IUPAC je farmakofor „soubor sterických a elektronických vlastností, které jsou nezbytné k zajištění optimálních supramolekulárních interakcí se specifickým biologickým cílem a k ovlivnění jeho biologické odpovědi.“ (Koščová a Provazník, 2016).

U ligandové metody spočívá využití farmakoforového modelu v generování sloučenin na základně stejných vlastností, které jsou důležité pro biologickou aktivitu. Naopak u strukturní metody se farmakofory využívají ke studiu interakce mezi cílovým proteinem a ligandy.

Všechny výše zmíněné metody nemají samy o sobě dostatečně velkou přesnost a jelikož dnešní databáze jsou poměrně rozsáhlé, musí se látky protřídit, aby se snížily celkové náklady a zvýšila rychlost výzkumu. Další důvod, proč se metody řetězí, je snaha o odstranění co největšího počtu falešně pozitivních a převážně falešně negativních látek. Proto se využívá jejich propojení při prohledávání databáze. Výpočetně nejnáročnější kroky se uplatňují až ke konci vyhledávání. Postupným vyřazováním látek dochází k zisku „hitů“ (Berka a Bazgier, 2015).



Obrázek 3: Sekvenční uspořádání pro virtuální screening [převzato z cit.(Svozil, 2017)].

Pokud sloučenina typu „hit“ splňuje všechny podmínky, mezi které patří např. i nízká toxicita, označí se jako tzv. „lead“ sloučenina neboli vůdčí. Tyto vůdčí látky postupují do experimentální fáze, kdy se již upravují funkční skupiny na molekule za účelem zlepšení vlastností, popřípadě zamezení některým nežádoucím efektům. Toto upravování se provádí na základě předpokladu, že sloučenina, která vykazuje aktivitu, bude mít do určité míry podobné fyzikálně-chemické vlastnosti s různými postranními řetězci, neboť základní skelet zůstává zachován. Typickou úpravou je přidávání velkých hydrofobních skupin, což zapříčiní jak nárůst bioaktivity, ale také molekulové hmotnosti zhruba o 200 kDa a logP o 0,5-5 jednotky. Z přítomnosti hydrofobních skupin logicky vyplývá i zhoršení rozpustnosti ve vodě (Svozil, 2017).

2.5 Program Qikprop

Qikprop je software od společnosti Schrödinger pro rychlou předpověď ADME. Program byl navrhnutý profesorem William L. Jorensen. Qikprop předpovídá fyzikálně důležité ukazatele a lékařsky významné vlastnosti organických molekul. Qikprop dovoluje porovnávat vlastnosti určité molekuly s vlastnostmi 95 % známých léčiv. Také dovoluje označit 30 typů funkčních skupin, které mohou způsobovat falešně pozitivní výsledky ve VKS. Qikprop je schopný vypočítat nebo předpovědět na 51 různých vlastností a popisů látky. Všechny tyto vlastnosti jsou vypsány v tabulce č. 1 v obrazové příloze spolu s doporučenými hodnotami nebo jejich rozmezí (*QikProp 3.5 User Manual*, 2012).

2.5.1 Popis klíčových vlastností

Látku lze popsat různými fyzikálně-chemickými vlastnostmi. Mezi nejdůležitější tyto vlastnosti patří mimo jiné rozpustnost ve vodě, molekulová hmotnost a rozdělovací koeficient oktanol/voda. Optimální léčivo by mělo mít dostatečnou rozpustnost ve vodě, protože se musí rozpustit v krvi/plasmě a ostatních tělních tekutinách. Zároveň musí mít dostatečnou lipofilitu, aby látky mohli prostupovat biologickými membránami (*Experimental physicochemistry characterisation*).

2.5.1.1 Molekulová hmotnost

Molekulová hmotnost u léčiv může sloužit k filtrování databází. Většina léčiv má molekulovou hmotnost v rozmezí 100 – 1 000. Samozřejmě, že i zde se najdou výjimky, například trombolytické enzymy, které mají hmotnost přibližně 50 000. Obecně platí, že

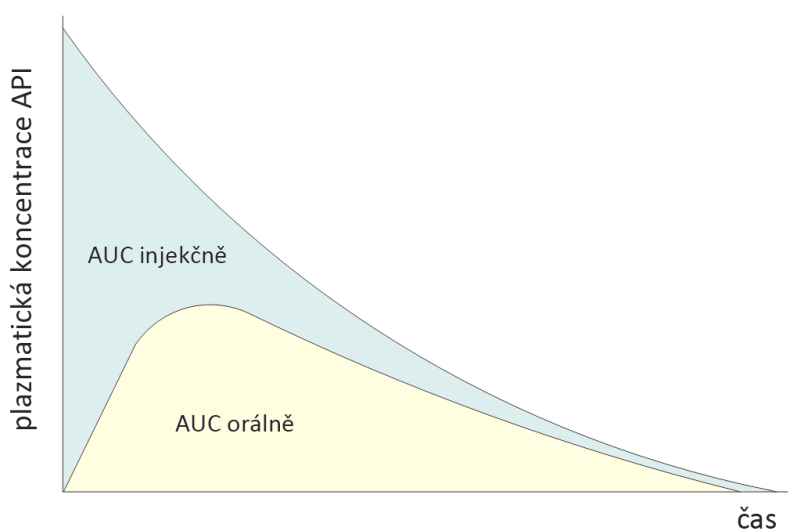
látky s nižší molekulovou hmotností mají větší šanci být vybrány jako léčivo, protože lépe prochází skrz membrány (*Stanovení farmakokinetiky léčiv*, 2013).

2.5.1.2 Rozpustnost

Rozpustnost ve vodě je jedna z klíčových vlastností pro určení bio-dostupnosti orálních léčiv. Tato vlastnost je součástí pojmů ADME, a to zvláště u absorpce a distribuce v organismu. Jedná se o jednu z hlavních fyzikálně-chemických vlastností při objevu léčiv. Od rozpustnosti se odvíjí i množství potřebného léčiva. Následná klasifikace rozpustnosti od FDA (angl. Food and Drug Administration) je udávána v objemu rozpouštědla v ml na rozpuštění 1 g léčiva.

- Velmi rozpustné < 1
- Volně rozpustné 1-10
- Rozpustné 10-30
- Mírně rozpustné 30-100
- Těžce rozpustné 100-1 000
- Velmi těžce rozpustné 1 000-10 000
- Prakticky nerozpustné > 10 000 (Jouyban, 2010).

Rozpustnost přispívá k bio-dostupnosti, která je závislá mimo jiné i na propustnosti. Jedná se o farmakokinetický parametr, který udává podíl aktivní farmaceutické složky v procentech (z angl. API – Active Pharmaceutical Ingredient), který se dostane do krve. Většinou se pohybuje mezi 10 a 80%.



Obrázek 4: plazmatická koncentrace léčiva v závislosti na čase u léčiv podaných injekčně nebo orálně. AUC je označení pro plochu pod křivkou (z angl. Area Under the Curve) [převzato z cit. (Kratochvíl, 2015)].

Orální biodostupnost v procentech se pak vypočítá jako:

$$\text{orální biodostupnost} = \frac{AUC \text{ orálně}}{AUC \text{ injekčně}} * 100 \quad (2)$$

2.5.1.3 Rozdělovací koeficient oktanol/voda

Lipofilita látky popisuje schopnost rozpouštět se v buněčné membráně. Je popsána jako rozložení sloučeniny mezi organickou a vodní fází. Rozdělovací koeficient pro systém oktanol/voda je definován jako podíl koncentrací rozpuštěné látky v systému dvou různě mísitelných rozpouštědel (Růžičková, 2007).

Rozdělovací koeficient je většinou vyjádřen jako logP kde:

$$P = \frac{c_{\text{oktanol}}}{c_{\text{voda}}} \quad (3)$$

c = koncentrace rozpouštědla oktanol nebo voda v mg/ml

Oktanol byl zvolen jako zástupce lipofilní vrstvy v buněčné membráně, protože má podobné fyzikálně-chemické vlastnosti jako tukové tkáně. Maximální hodnota logP, aby platilo Lipinského pravidlo pěti je 5, a to odpovídá rozdílu koncentrací mezi fázemi voda a oktanol 1:100 000.

3. Praktická část

3.1 Použité databáze

Při práci se využila hlavně databáze DrugBank, ve které látky mají zhruba kolem 20 vypočítaných nebo předpovězených fyzikálně-chemických vlastností (*DrugBank*). Za zmínku stojí i databáze Zinc a Pubchem. Obě zmíněné databáze mají omezený počet vlastností, které jsem se snažil rozšířit. Například databáze Zinc obsahuje zhruba 14 fyzikálně-chemických vlastností ke sloučenině Aspartamu a jeho dostupnost u různých obchodníků (*ZINC 15; PubChem*).

3.1.1 Databáze Drugbank

Látky pro naši databázi jsem čerpal z databáze DrugBank, volně dostupné na adrese <https://www.drugbank.ca/>. Jedná se o databázi, která uchovává informace o léčivech a jejich cílech. Mezi tyto informace patří například obecný popis daného léčiva, základní fyzikálně-chemické vlastnosti a popis farmakologických účinků. V práci jsem se převážně zabýval rozšířením předpovězených nebo vypočítaných fyzikálně-chemických vlastností.

3.2 Výpočet vlastnosti pomocí qikprop

K výpočtu všech vlastností ve vytvořené databázi jsem dospěl pomocí programu Qikprop. Samotný výpočet probíhal v prostředí Linux pomocí příkazového řádku přes software PuTTY. Jednotlivé látky před samotným výpočtem bylo nutné mít v PDB formátu. Qikprop je schopný vypočítat vlastnosti pouze v případě, že látka má 3D souřadnice heteroatomů v molekule. Další nutnou úpravou bylo odstranění všech informací o látce, kromě již zmíněných souřadnic heteroatomů. Tato úprava se provedla pomocí Bash skriptu viz. obr. 5 kde `grep HETATM $i > $filename` vyhledá a vypíše všechny řádky, které obsahují výraz *HETATM*. U tohoto výrazu jsou vypsány právě požadované souřadnice heteroatomů.

3.2.1 Kritéria potřebná pro výpočet

Před samotným výpočtem musí být každá molekula upravena podle následujících pravidel:

- Struktura musí být trojrozměrná.
- Atomy vodíku musí být explicitní.
- Každý soubor musí obsahovat pouze jednu molekulu.

- Povolené prvky jsou: H, C, N, O, F, Al, Si, P, S, Cl, Br a I. Pro zrychlený režim jsou povoleny všechny prvky, ale výsledky nemusí být platné.
- Počet atomů nesmí přesáhnout 1 000 ve zrychleném režimu nebo 100 nevodíkových a 150 vodíkových atomů v normálním režimu (*QikProp 3.5 User Manual*, 2012).

```

1 #!/bin/bash
2
3 SUFIX="pdb"
4
5 for i in /home/repka/prace/PDB/*.${SUFIX}
6 do
7
8     filename=$(basename $i)
9     fullfilename=$filename
10    filename=${filename%.*}
11    echo -n $i
12    grep HETATM $i > $filename
13
14
15 done

```

Obrázek 5: Zdrojový kód pro úpravu látek

HETATM	1	C1	UNK	900	3.600	5.403	0.351	1.00	0.00	C
HETATM	2	O1	UNK	900	2.659	4.376	0.088	1.00	0.00	O
HETATM	3	C2	UNK	900	2.803	3.178	0.679	1.00	0.00	C
HETATM	4	O2	UNK	900	3.704	2.877	1.464	1.00	0.00	O
HETATM	5	C3	UNK	900	1.669	2.235	0.255	1.00	0.00	C
HETATM	6	C4	UNK	900	0.309	2.626	0.875	1.00	0.00	C
HETATM	7	C5	UNK	900	-0.891	1.905	0.285	1.00	0.00	C
HETATM	8	C6	UNK	900	-1.412	2.307	-0.963	1.00	0.00	C
HETATM	9	C7	UNK	900	-2.514	1.630	-1.520	1.00	0.00	C
HETATM	10	C8	UNK	900	-3.099	0.549	-0.832	1.00	0.00	C
HETATM	11	C9	UNK	900	-2.582	0.146	0.415	1.00	0.00	C
HETATM	12	C10	UNK	900	-1.481	0.824	0.973	1.00	0.00	C
HETATM	13	N1	UNK	900	2.005	0.848	0.575	1.00	0.00	N
HETATM	14	C11	UNK	900	1.773	-0.214	-0.213	1.00	0.00	C
HETATM	15	O3	UNK	900	1.260	-0.120	-1.329	1.00	0.00	O
HETATM	16	C12	UNK	900	2.184	-1.581	0.356	1.00	0.00	C
HETATM	17	N2	UNK	900	1.388	-1.902	1.546	1.00	0.00	N
HETATM	18	C13	UNK	900	2.103	-2.709	-0.703	1.00	0.00	C
HETATM	19	C14	UNK	900	2.635	-4.057	-0.210	1.00	0.00	C
HETATM	20	O4	UNK	900	3.989	-4.052	-0.047	1.00	0.00	O
HETATM	21	O5	UNK	900	1.925	-5.035	0.011	1.00	0.00	O
HETATM	22	H1	UNK	900	1.577	2.320	-0.829	1.00	0.00	H
HETATM	23	H2	UNK	900	3.225	-1.487	0.671	1.00	0.00	H
HETATM	24	H3	UNK	900	4.603	5.099	0.050	1.00	0.00	H
HETATM	25	H4	UNK	900	3.334	6.301	-0.207	1.00	0.00	H
HETATM	26	H5	UNK	900	3.614	5.655	1.413	1.00	0.00	H
HETATM	27	H6	UNK	900	0.332	2.461	1.953	1.00	0.00	H
HETATM	28	H7	UNK	900	0.135	3.695	0.746	1.00	0.00	H
HETATM	29	H8	UNK	900	-0.963	3.130	-1.498	1.00	0.00	H
HETATM	30	H9	UNK	900	-2.909	1.937	-2.477	1.00	0.00	H
HETATM	31	H10	UNK	900	-3.943	0.029	-1.260	1.00	0.00	H
HETATM	32	H11	UNK	900	-3.028	-0.685	0.941	1.00	0.00	H
HETATM	33	H12	UNK	900	-1.085	0.503	1.926	1.00	0.00	H
HETATM	34	H13	UNK	900	2.450	0.688	1.467	1.00	0.00	H
HETATM	35	H14	UNK	900	1.615	-2.815	1.913	1.00	0.00	H
HETATM	36	H15	UNK	900	1.550	-1.245	2.295	1.00	0.00	H
HETATM	37	H16	UNK	900	1.070	-2.846	-1.029	1.00	0.00	H
HETATM	38	H17	UNK	900	2.671	-2.433	-1.593	1.00	0.00	H
HETATM	39	H18	UNK	900	4.315	-4.887	0.252	1.00	0.00	H

Obrázek 6: Výstup po úpravě látky Aspartam, vyznačené hodnoty jsou souřadnice atomu na ose xyz

Pro každou úspěšně vypočítanou látku byly vytvořeny 4 soubory a to:

1. název látky.qpsa
2. název látky.out
3. název látky.log
4. název látky.CSV(*QikProp 3.5 User Manual*, 2012).

První soubor obsahuje doplňující informace ke každé molekule. Mezi tyto informace patří: seznam kovalentních sousedů, přiřazení atomové typu, související parametry silového typu atomového pole a analýzu pro povrch přístupný rozpouštědlu (SASA). Na konci souboru je souhrn molekul, které byly zpracovány, popřípadě vynechány nebo přeskočeny. Také je zde informace o času potřebném pro výpočet. Pro příklad u molekuly Aspartamu je to 0,4 sekundy.

Druhý soubor, tedy soubor s příponou *.out* obsahuje předpovězené vlastnosti a porovnání těchto hodnot vlastností s hodnotami pro 95 % typických léčiv. Mimo jiné se zde nachází rozdělení QP. Tato sekce zahrnuje log Po/w, -log S, log BB a log PMDCK. Všechny tyto hodnoty byly vypočítány lineární kombinací základních molekulárních vlastností (*QikProp 3.5 User Manual*, 2012). Výčet všech těchto vlastností i s hodnotami pro vzorovou molekulu Aspartamu je znázorněn na obr. č.7.

```
QP Breakdown (< for descriptor over training max)
      log Po/w                -log S
H-bond Donor      -0.675      H-bond Donor      -0.904
H-bond Acceptor   -2.800      H-bond Acceptor   -3.011
Volume            6.365      SASA              10.798
Ac x Dn^.5/SASA  0.671      Ac x Dn^.5/SASA  1.513
FISA              -0.987      Rotor Bonds      -1.302
Non-con amines    -0.527      N Protonation     -1.220
Non-con amides    -0.626      Non-con amides    -0.560
WPSA & PISA       0.199      WPSA              0.000
Constant          -2.579      Constant          -3.118
Total             -0.959      Total             2.196

Aminoacid correction of -1.9 has been applied to log Po/w.

      log BB                log PMDCK
Hydrophilic SASA -2.077      Hydrophilic SASA -2.226
WPSA              0.000      WPSA              0.000
Rotor Bonds       -0.482      Rotor Bonds       0.000
N Protonation     0.399      Non-con amines    -0.608
FOSEA            0.000      COOH/SO3H acids  -0.540
Constant          0.564      Constant          3.771
Total            -1.597      Total             0.396
```

Obrázek 7: soubor základních molekulových vlastností pro výpočet Log Po/w, -Log S, Log BB a Log PMDCK

Třetí v pořadí jsou soubory s příponou *.log*. Zde je pouze výpis použitých nastavení pro výpočet dané látky. Jde například o tzv. Fast mode. Qikprop má možnosti výpočtů buď v normálním režimu nebo v již zmíněném Fast mode (zrychlený režim). V normálním režimu Qikprop dokáže vypočítat na 51 různých vlastností a popisů molekuly až u 10 000 molekul za hodinu. Na rozdíl ve zrychleném režimu se tento počet zvýší až na 300 000 molekul za hodinu, ale vypočítá se o 4 vlastnosti méně. Přeskočení těchto čtyř vlastností se podepíše na rozdílné hodnotě pro rozdělovací koeficient oktanol/plyn (*QikProp 3.5 User Manual*, 2012). V tomto zrychleném režimu Qikprop také přeskočí některé testy jinak obsažené v normálním režimu. To má za následek, že hodnoty vlastností nejsou platné, ale hodnoty popisující molekuly jsou. Pro tuto práci jsem ponechal normální režim.

Poslední z vytvořených souborů při výpočtu je s koncovkou *.CSV* (z angl. Comma Separated Value). V tomto souboru jsou vypsány názvy, popřípadě zkratky vlastností a popisu molekuly, a následné vypočítané hodnoty. Vše je odděleno čárkou, jak již název napovídá. Pro další práci s daty bylo nutné využít Microsoft Excel.

3.2.2 Qikprop metodologie

Samotný program Qikprop byl vytvořen pomocí programů BOSS a silového pole OPLS-AA. Tyto programy se využily pro statistickou simulaci organických rozpuštěných látek v opakujících se souborech explicitních molekul vody. Simulace proběhla metodou Monte Carlo. Výsledkem této simulace je konfigurační průměr pro několik deskriptorů, včetně počtu vodíkových vazeb a povrchu přístupného rozpouštědla (SASA). Byl získán vzájemný vztah deskriptorů k experimentálním hodnotám. Na základě tohoto vztahu vznikl algoritmus, který napodobuje kompletní simulaci metodou Monte Carlo a dosahuje srovnatelných výsledků. Pro tento vývoj byly použity experimentální data z více než 710 sloučenin zahrnujících okolo 500 léčiv a příbuzných heterocyklů.

Při provádění výpočtu, Qikprop rychle analyzuje typy atomů a náboje, počet rotorů, objem a povrch vzorku molekuly. Qikprop následně využije výsledky analýzy společně s fyzickými deskriptory, které byly vypočítány pomocí výše popsaného algoritmu v regresní rovnici. Výsledkem je předpověď farmakologicky podstatných vlastností (*QikProp 3.5 User Manual*, 2012).

3.2.3 zlepšení log S a log P vlastností

Pro zlepšení predikce klíčových vlastností molekuly, jakými jsou rozpustnost (log S) a rozdělovací koeficient (log P) se využívá podobnosti v chemickém prostoru molekulárních vlastností společně s větším množstvím experimentálních získaných dat a odpovídajícím předpovědím softwaru Qikprop. Pokud je dostatečně velká podobnost v chemickém prostoru molekulárních vlastností jsou předpovězené hodnoty upraveny za přispění experimentálních hodnot. Tato úprava probíhá na základě porovnání podobností u každé zpracované molekuly se známými experimentálními hodnotami. Podobnost je v rozmezí 0 (žádná podobnost) až 1 (zcela identické).

- Pokud je podobnost menší než 0,9, využije se předpověď pomocí Qikprop.
- Pokud je podobnost větší než nebo rovna 0,9, je předpovězená vlastnost (V_{Pred}) upravena podle následující rovnice.

$$V_{pred} = SV_{exp} + (1 - S)V_{QP} \quad (4)$$

Symbolem S se značí podobnost, V_{exp} nese označení pro experimentální vlastnosti a V_{QP} pro Qikpropem předpovězené vlastnosti nejpodobnější molekuly. Pokud tedy podobnost dosahuje hodnoty 1, budou takovéto předpovězené vlastnosti rovny jejich experimentálním hodnotám dané molekuly. V zájmu rychlosti je takto upravováno pouze prvních 10 000 položek (*QikProp 3.5 User Manual*, 2012).

3.2.4 Problém přesnosti výpočtu vlastností

Předpovědi vlastností jsou prováděny za specifických podmínek, kupříkladu jedna z těchto podmínek je pH 7. Porovnání výsledků s experimentem musí vzít v potaz tyto rozdíly. Následující problémy mohou ovlivnit přesnost porovnání. Největší nepřesnost je u předpovědi PCaco.

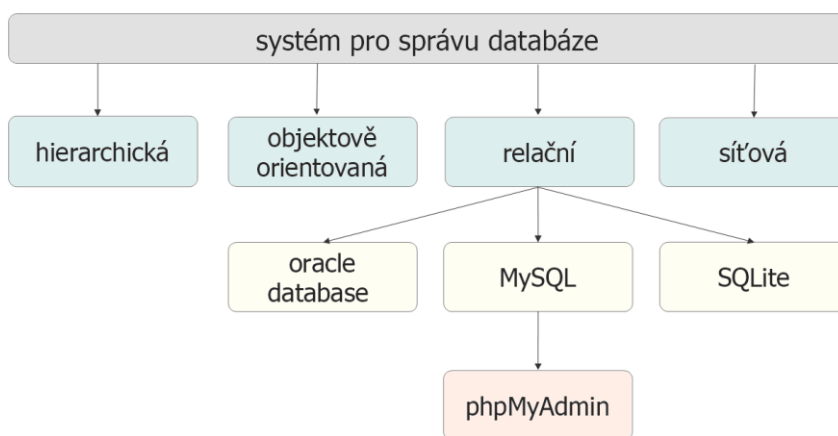
- Aktivní vs. pasivní transportní mechanismus.
- Mechanismus efflux pumpy.
- Menší počet sloučenin v trénovačích sadách.
- Různé experimentální podmínky než použité v regresi.
- Šum v experimentálních datech kvůli neodmyslitelné složitosti interakce chemického systému s biologickým (*QikProp 3.5 User Manual*, 2012).

3.3 Databáze

Databáze jako taková je strukturovaná sbírka dat. Na data se nevztahuje žádné omezení, záleží zcela na uživateli, co daná databáze bude obsahovat. Pro manipulaci s databází nám slouží tzv. systém pro správu databází (z angl. Database Management System). Tento systém umožňuje přístup do databáze, její manipulaci a pomoc při reprezentaci dat. Systém pro správu databází se může dělit na 4 druhy podle datového modelu, v naší databázi jsem využil relační typ (*Types And Classification Of Database Management System*, 2017).

3.3.1 Relační databáze

Mezi nepoužívanější systém pro správu databází patří relační. Tento systém ukládá data do tabulek. Využívá programovací jazyk SQL. Jde o označení pro standardizovaný strukturovaný dotazovací jazyk. Pomocí SQL se uživatel dotazuje na databázi a komunikuje s ní. Pod relační databáze spadá více systému jako např. SQLite, MySQL, PostgreSQL, Oracle Database. Mezi jednotlivými systémy můžou nastat malé rozdíly v použití SQL příkazů (*What is a Relational Database Management System?*). Na naší databázi byl použit systém MySQL. Databázi jsem upravoval přes rozhraní phpMyAdmin, které slouží jako webový nástroj pro úpravu databází.

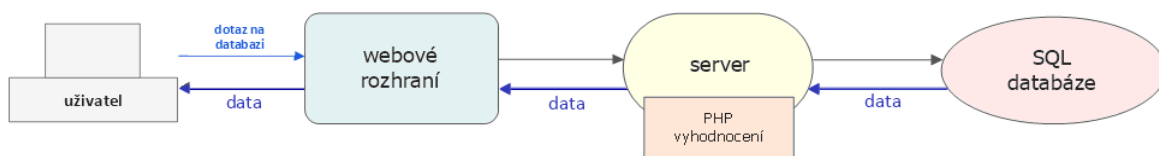


Obrázek 8: schéma rozdělení systému pro správu databáze

3.4 Webová aplikace

Součástí práce bylo také vytvoření uživatelsky přívětivého prostředí pro práci s databází, proto bylo vytvořeno webové rozhraní. Webová aplikace byla vytvořena pomocí PHP programovacího jazyka (z angl. PHP: Hypertext Preprocessor) a CSS neboli kaskádových stylů (z angl. Cascading Style Sheets) (*What is PHP?*).

PHP je programovací jazyk primárně používaný pro tvorbu dynamických a interaktivních webových aplikací a může být vnořen do HTML. PHP je serverově založený programovací jazyk, tzn. že nepracuje na straně uživatele, zde se jen pomocí HTML zobrazí výsledek. PHP může být integrováno s databázemi jako např. MySQL (*PHP 5 Tutorial; Differences Between HTML vs PHP*).



Obrázek 9: schéma způsobu dotazu na databázi

Kaskádové styly je označení pro programovací jazyk, který popisuje styl HTML dokumentu. Popisuje, jak by měl být HTML dokument vyobrazen (*CSS Tutorial*). Využívají se při složitějších a více rozvětvených webových stránkách, neboť dovoluují upravovat všechny složky stránek z jednoho CSS souboru.

3.5 Návod k aplikaci

Vytvořená webová aplikace propojující databázi s uživatelem je schopna vyhledávat sloučeniny podle jejich vypočtených fyzikálně-chemických vlastností nebo deskriptorů z programu Qikprop. Každou tuto informaci lze vyhledat samostatně. Na obrázku č. 10 je úvodní strana naší webové aplikace. Pro každou z vlastností je zobrazena nápověda po najetí kurzoru. Samotný dotaz na databázi a vyhledání výsledků se uskuteční zadáním požadovaných hodnot u vlastností a k tomu přiřazení požadovaného relačního operátoru <, > nebo =. Všechny vyplněné vlastnosti, tak aby každá platila je možné zahrnout do vyhledávání pomocí tlačítka „AND“, naopak pomocí tlačítka „OR“ je možné vyhledávat sloučeniny, které splňují alespoň jednu ze zadaných vlastností. Aplikace nabízí možnost vyhledat všechny látky buď použitím tlačítka *Vyhledat vše* nebo ponecháním prázdných políček a použitím tlačítka *Vyhledat*.

Databáze sloučenin

Molární hmotnost:	< ▾ <input type="text"/>	stars:	< ▾ <input type="text"/>	amin:	< ▾ <input type="text"/>
amidin:	< ▾ <input type="text"/>	počet karboxylový zbytků:	< ▾ <input type="text"/>	amid:	< ▾ <input type="text"/>
rotor:	< ▾ <input type="text"/>	rtvFG:	< ▾ <input type="text"/>	cns:	< ▾ <input type="text"/>
dipole:	< ▾ <input type="text"/>	SASA:	< ▾ <input type="text"/>	FOSA:	< ▾ <input type="text"/>
FISA:	< ▾ <input type="text"/>	PISA:	< ▾ <input type="text"/>	WPSA:	< ▾ <input type="text"/>
volume:	< ▾ <input type="text"/>	donorHB:	< ▾ <input type="text"/>	accptHB:	< ▾ <input type="text"/>
dip ² /V:	< ▾ <input type="text"/>	ACxDN ^{0.5} /SA:	< ▾ <input type="text"/>	glob:	< ▾ <input type="text"/>
QPpolrz:	< ▾ <input type="text"/>	QPlogPC16:	< ▾ <input type="text"/>	QPlogPoc:	< ▾ <input type="text"/>

Obrázek 10: Vytvořená webová aplikace, úvodní strana pro vyhledávání

Aspartame											
Molární hmotnost:	294.307	hvězdy:	0	amin:	1	amidin:	0	počet karboxylový zbytků:	1	amid:	1
rotor:	8	rtvFG:	1	cns:	-2	dipol:	5.232	SASA:	569.793	FOSA:	183.046
FISA:	217.192	PISA:	169.555	WPSA:	0.000	objem:	975.607	donor vodíkových vazeb:	2.250	akceptor vodíkových vazeb:	5.750
dip ² /V:	0.0280622	ACxDN ^{0.5} /SA:	0.0151371	glob:	0.8348931	QPpolrz:	29.372	QPlogPC16:	10.460	QPlogPoc:	16.393
QPlogPw:	12.168	QPlogPo/w:	-0.959	QPlogS:	-2.196	CIQPlogS:	-1.582	QPlogHERG:	-2.556	QPPCaco:	4.000
QPlogBB:	-1.597	QPPMDCK:	2.491	QPlogKp:	-6.220	ionizační potenciál:	9.400	elektronová afinita:	-0.226	metab:	6
QPlogKhsa:	-0.670	absorpce ústý:	2	Absorpce ústý v %:	32.105	SAfluorine:	0.000	SAamidO:	16.924	PSA:	143.284
NandO:	7	pravidlo pěti:	0	počet atomů v kruhu:	6	v34:	0	v56:	6	noncon:	0
nonHatm:	21	pravidlo tři:	1	Jm:	0.001	experimentální rozpustnost logS:		experimentální rozdělovací koeficient:	-0,1	zjištěná hodnota molární hmotnosti:	294.3031
Sloučenina v PDB formátu											
Astemizole											
Ataluren											

Obrázek 11: výsledek vyhledávání sloučenin

Na obrázku č. 11 se nachází výsledek vyhledávání všech látek a vzorová sloučenina Aspartamu je vyobrazena s jejími vypočítanými vlastnostmi a deskriptory. U každé látky se nachází odkaz pro stažení dané sloučeniny ve formátu *PDB*, který umožňuje zobrazení dané látky v její 3D podobě.

3.6 Experimentální data z databází

U částečného množství látek byla získána experimentální data pro rozpustnost ve vodě, rozdělovací koeficient oktanol/voda a hodnoty pro molární hmotnost z několika databází. Hodnoty pro rozpustnost jsou udávány v mol/dm³.

Tabulka 1: Zdroje experimentálních hodnot pro rozpustnost a rozdělovací koeficient

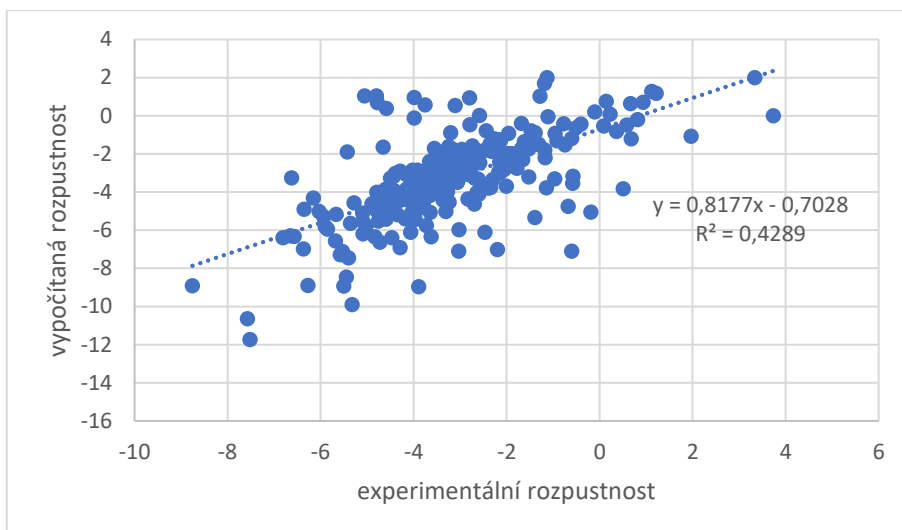
Název		Adresa
Drugbank	Bezpečnostní list	-
	FDA	https://www.fda.gov/
	PhysProp	https://www.srcinc.com/?id=386
	Wikipedie	https://en.wikipedia.org/wiki/Main_Page
Národní centrum biotechnologických informací		https://www.ncbi.nlm.nih.gov
Databáze lidských metabolomů		http://www.hmdb.ca
Chemspider		http://www.chemspider.com

3.7 Vyhodnocení

K popisu dat jsem využil model lineární regrese. Koeficient determinace, který slouží pro porovnání shody modelu s daty se označuje jako R^2 . Koeficient determinace může dosahovat hodnot v rozmezí intervalu $<0,1>$ (Zichová). Platí, že pokud hodnota R^2 je rovna 1, tak jsou vypočítané hodnoty odpovídající zjištěným a leží v přímce. Pokud hodnota R^2 je rovna 0, mezi daty není žádná statistická spojitost.

3.7.1 Rozpustnost

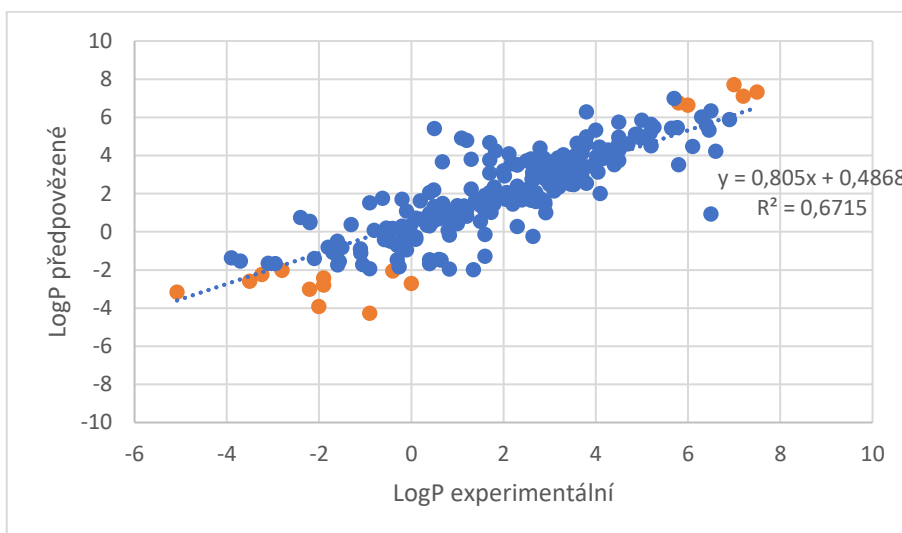
Všechny hodnoty pro rozpustnost byly získány v mg/ml a následně přepočítány na mol/dm³. Výsledná hodnota byla vypočítaná pomocí dekadického logaritmu hodnot v mol/dm³. Výsledky vypočítaných hodnot nepříliš odpovídají experimentálně zjištěným. To může být zapříčiněno tím, že Qikprop standardně vypočítává rozpustnost při pH 7. Mnou zjištěné hodnoty neměli ve většině případů napsané podmínky měření a zde může být jeden z problémů neodpovídajících výsledků. U rozpustnosti dosáhla hodnota $R^2 = 0,4289$. Experimentální hodnoty rozpustnosti se mi podařilo získat pro 60,5 % zjišťovaných látek.



graf 1: závislost experimentální rozpustnosti na vypočítané

3.7.2 Rozdělovací koeficient oktanol/voda

Při rozboru dat u rozdělovacího koeficientu oktanol/voda jsem získal hodnotu pro R^2 rovnu 0,731 pro všechny hodnoty. Na obrázku grafu č. 2 jsou zobrazeny modře body pro hodnoty, které jsou v rámci Qikpropem doporučených hodnot tedy, -2 až 6,5. Oranžové hodnoty jsou mimo tyto limity. Na grafu č. 2 je hodnota pro R^2 rovna 0,6715, která platí pouze pro hodnoty v rámci doporučených hodnot. Hodnoty jsou ověřeny pro 66,4 % zjišťovaných látek.



graf 2: závislost experimentálního rozdělovacího koeficientu na předpovězený

4. Závěr

Hlavním cílem mé bakalářské práce bylo vytvoření samostatné chemické knihovny. Aby takováto knihovna mohla konkurovat dnes dostupným databázím, bylo potřeba najít jejich nedostatky. Takovéto nedostatky se vyskytují například v rozsahu vypočítaných fyzikálně-chemických vlastností sloučenin obsažených v daných online databázích. Právě na různé fyzikálně-chemické vlastnosti jsem se zaměřil a snažil jsem se o jejich výpočet, popřípadě predikci. S tímto problémem nám pomohl program Qikprop, který byl schopný vypočítat na 51 vlastností.

Pro uživatelsky přívětivou databázi, bylo zvoleno vytvoření volně přístupné webové aplikace, která je dostupná na adrese <http://pokusnice.upol.cz/bc/>. Přes tuto webovou aplikaci je možné vyhledávat pomocí jednotlivých, popřípadě vícero hodnot sloučeniny odpovídající zadaným kritériím, u kterých jsem napočítal jejich vlastnosti. Pomocí relačních operátorů je uživatel schopen vybrat rozsah hledaných hodnot. Vyhledávání lze upravit ještě pomocí tlačítek „AND“ nebo „OR“.

Výsledkem dotazu na databázi je výčet názvů sloučenin, které splňují zadaná kritéria. Vedle vypsaných vypočítaných vlastností jsou vypsané experimentální data pro rozpustnost ve vodě, rozdělovací koeficient a ověřené molární hmotnosti. Každá látka je obohacena o odkaz na stažení 3D struktury ve formátu *PDB*.

Experimentální data nebyla nalezena pro všechny látky. Značné množství těchto dat bylo získáno skrze volně dostupné chemické databáze. Ostatní se nepodařilo z literatury ve vhodném časovém rozmezí získat. Pro experimentální hodnoty rozpustností nebyly ve většině případů dostupné reakční podmínky, jako je teplota nebo pH. Tyto podmínky mohou značně ovlivnit rozpustnost. Proto při vyhodnocení dat závislosti experimentálních hodnot na vypočítaných jsem bral v potaz pouze hodnoty s rozdílem 1. Právě tyto hodnoty by měly mít podobné reakční podmínky jako byly použity při výpočtu programem Qikprop. Z těchto důvodů byl koeficient determinace u rozpustnosti nejmenší a to tedy 0,4289. Další experimentální data se týkala rozdělovacího partičního koeficientu, u kterého koeficient determinace byl mírně vyšší, hodnota pro všechny data činila 0,731. Ve výsledku vyhledávání lze nalézt i ověřené molární hmotnosti. Tyto hodnoty sloužily pro ověření výpočtu.

5. Summary

The main goal of this bachelor thesis was to create a standalone chemical library. It was necessary to find imperfections of currently available libraries to create a new competitive one. Example of these imperfections is the lack of some calculated physical-chemical properties.

I focused mainly on exact calculation or prediction of these physical-chemical properties. A very helpful software Qikprop, which can calculate up to 51 different properties, was used for the purposes of the study.

To create a user-friendly database designed to search for chemical compounds by one or more values of properties, I used a free online application accessible at <http://pokusnice.upol.cz/bc/>. Using the relational operators, the user is able to select a desired range of values. Searching process can also be adjusted via “AND” or “OR” buttons.

A list of names of compounds fulfilling required values is received as a result of the database query. The list of calculated properties contains experimental data on solubility, partition coefficient and verified molecular weight. Also, a link to 3D structure PDB file download site is included with each of the displayed compounds.

Experimental data were not found for every compound. A large part of these data was obtained from freely accessible chemical databases. The rest of the data could not be obtained in a reasonable time interval. Furthermore, in most cases, solubility affecting reaction conditions as temperature or pH were not available. The resultant solubility value can be significantly biased by these conditions. That is the reason why only values with difference of 1 were taken in account. These values reaction conditions should match those used in Qikprop calculation. Because of that, solubility determination coefficient was the smallest (0,4289). Also, determination coefficient value of partition coefficient experimental data came out little higher (0,731). Verified molecular weight is included in the search results too. These values are used only for computational procedure validation.

6. Použitá literatura

- Automating Compound Management Systems* (2007). Dostupné z: <https://www.rdmag.com/article/2007/09/automating-compound-management-systems> (Viděno: 18. duben 2019).
- Berka, K. a Bazgier, V. (2015) *Racionální návrh léčiv pomocí in silico metod*.
- Bickerton, G. R. *et al.* (2012) „Quantifying the chemical beauty of drugs“, *Nature Chemistry*. Nature Publishing Group, 4(2), s. 90–98. doi: 10.1038/nchem.1243.
- CSS Tutorial*. Dostupné z: <https://www.w3schools.com/css/> (Viděno: 19. duben 2019).
- Differences Between HTML vs PHP*. Dostupné z: <https://www.educba.com/html-vs-php/> (Viděno: 18. duben 2019).
- DrugBank*. Dostupné z: <https://www.drugbank.ca/> (Viděno: 19. duben 2019).
- Experimental physicochemistry characterisation*. Dostupné z: <http://www.sbw.fi/lead-optimization/experimental-logp-logd-logs-pka-analysis/> (Viděno: 18. duben 2019).
- Hertzberg, R. P. a Pope, A. J. (2000) „High-throughput screening: new technology for the 21st century“, *Current Opinion in Chemical Biology*. Elsevier Current Trends, 4(4), s. 445–451. doi: 10.1016/S1367-5931(00)00110-1.
- High Throughput Screening*. Dostupné z: <https://www.tdi.ox.ac.uk/high-throughput-screening-2> (Viděno: 18. duben 2019).
- Hughes, J. P. *et al.* (2011) „Principles of early drug discovery Correspondence“. doi: 10.1111/j.1476-5381.2010.01127.x.
- Irwin, J. J. a Shoichet, B. K. (2005) „ZINC – A Free Database of Commercially Available Compounds for Virtual Screening ZINC - A Free Database of Commercially Available Compounds for Virtual Screening“, *J. Chem. Inf. Model*, 45(December 2004), s. 177–182. doi: 10.1021/ci049714.
- Jouyban, A. (2010) *Handbook of Solubility Data for Pharmaceuticals*.
- Jung, G. (1999) *Combinatorial chemistry : synthesis, analysis, screening*. Wiley-VCH. Dostupné z: https://books.google.cz/books?hl=cs&lr=&id=KWs14rVIvh4C&oi=fnd&pg=PP2&dq=Combinatorial+Chemistry:+Synthesis,+Analysis,+Screening&ots=Jer1dr62G1&sig=jtXEePcwBiLlj2D-0vIthete9Js&redir_esc=y#v=onepage&q=Combinatorial+Chemistry%3ASynthesis%2C+Analysis%2C+Screening&f=false (Viděno: 18. duben 2019).
- Koščová, P. a Provazník, I. (2016) „Racionální návrh léčiv s využitím farmkoforového modelování“. Dostupné z: http://www.chemicke-listy.cz/docs/full/2016_08_575-580.pdf (Viděno: 18. duben 2019).
- Kratochvíl, B. (2015) *Rozpustnost, rozpouštěcí rychlost, prostupnost (permeabilita), biodostupnost*. Dostupné z: http://uchpel.vscht.cz/files/uzel/0022888/Rozpustnost+2015.pdf?redirected&fbclid=IwAR3FybSZBusoH9CHuG1MEa8x7ezHiJW96_YDPZTT0PhMPdZXcsRtwxqN67c (Viděno: 19. duben 2019).
- Lionta, E. *et al.* (2014) „Structure-based virtual screening for drug discovery: principles, applications and recent advances.“, *Current topics in medicinal chemistry*. Bentham

Science Publishers, 14(16), s. 1923–38. doi: 10.2174/1568026614666140929124445.

Lipinski, C. A. *et al.* (2001) *Experimental and computational approaches to estimate solubility and permeability in drug discovery and development q settings*, *Advanced Drug Delivery Reviews*. Dostupné z: www.elsevier.com/locate/drugdeliv (Viděno: 18. duben 2019).

Meng, X.-Y. *et al.* (2011) „Molecular docking: a powerful approach for structure-based drug discovery.“, *Current computer-aided drug design*. NIH Public Access, 7(2), s. 146–57. Dostupné z: <http://www.ncbi.nlm.nih.gov/pubmed/21534921> (Viděno: 18. duben 2019).

PHP 5 Tutorial. Dostupné z: <https://www.w3schools.com/php/default.asp> (Viděno: 18. duben 2019).

Potáček, M. (2012) *Výzkum a vývoj nových léčiv*. Dostupné z: <https://is.muni.cz/el/1431/jaro2012/C8790/um/31610507/Farm03.pdf> (Viděno: 19. duben 2019).

PubChem. Dostupné z: <https://pubchem.ncbi.nlm.nih.gov/> (Viděno: 19. duben 2019).

QikProp 3.5 User Manual (2012). Dostupné z: http://gohom.win/ManualHom/Schrodinger/Schrodinger_2012_docs/qikprop/qikprop_user_manual.pdf (Viděno: 18. duben 2019).

Růžičková, P. (2007) *Experimentální studium enviromentálních procesů*. Dostupné z: https://is.muni.cz/th/p9qdf/Appendix_III_vysledky.pdf (Viděno: 18. duben 2019).

Sinclair, I. *et al.* (2016) „Novel Acoustic Loading of a Mass Spectrometer“, *Journal of Laboratory Automation*. SAGE PublicationsSage CA: Los Angeles, CA, 21(1), s. 19–26. doi: 10.1177/2211068215619124.

Stanovení farmakokinetiky léčiv (2013). Dostupné z: <http://ach.upol.cz/user-files/intranet/klch-farmakokinetika-1358888494.pdf> (Viděno: 18. duben 2019).

Svozil, D. (2017) *Virtuální screening*, *Chem. Listy*. Dostupné z: http://www.chemicke-listy.cz/docs/full/2017_11_738-746.pdf (Viděno: 18. duben 2019).

Types And Classification Of Database Management System (2017). Dostupné z: http://whatisdbms.com/types-and-classification-of-database-management-system/?fbclid=IwAR3ynqCcVgJ-tbxATP-P_XDz5_x_NGZP_d39yF0BtBWKJFMslugwNITCEv94 (Viděno: 19. duben 2019).

Veber, D. F. *et al.* (2002) „Molecular Properties That Influence the Oral Bioavailability of Drug Candidates“. American Chemical Society. doi: 10.1021/JM020017N.

Walters, M. W. P. (2018) „Virtual Chemical Libraries“. doi: 10.1021/acs.jmedchem.8b01048.

Walters, W. P., Stahl, M. T. a Murcko, M. A. (1998) „Virtual screening—an overview“, *Drug Discovery Today*. Elsevier Current Trends, 3(4), s. 160–178. doi: 10.1016/S1359-6446(97)01163-X.

What is a Relational Database Management System?. Dostupné z: <https://www.codecademy.com/articles/what-is-rdbms-sql> (Viděno: 19. duben 2019).

What is PHP?. Dostupné z: <https://www.php.net/manual/en/intro-what-is.php> (Viděno: 18.

duben 2019).

Zichová, J. *Lineární regresní model*. Dostupné z:
<http://www.karlin.mff.cuni.cz/~zichova/PRFUK/MatStat/LINEÁRNÍ REGRESNÍ MODEL.pdf> (Viděno: 19. duben 2019).

ZINC 15. Dostupné z: <http://pubs.acs.org/doi/10.1021/acs.jcim.5b00559> (Viděno: 19. duben 2019).

Přílohy

Tabulka 2: Výčet všech vlastností s jejich popisem a předepsanými hodnotami

vlastnosti nebo popis	popis	doporučená hodnota nebo rozmezí hodnot
<i>název molekuly</i>	název molekuly	-
<i>hvězdy</i>	počet vlastností nebo hodnot, které jsou mimo 95% hodnot pro známé léčiva. Čím větší číslo tím větší šance, že látka nebude vhodná pro použití jako léčivo. Pro určení počtu #hvězd jsou použity následující vlastnosti: molekulová hmotnost, dipol, IP, EA, SASA, FOSA, FISA, PISA, WPSA, PSA, objem, #rotor, donor a akceptor vodíkových vazeb, glob, QPpolrz, QPlogPC16, QPlogPoct, QPlogPw, QPlogPo/w, logS, QPLogKhsa, QPlogBB, #metabol	0 - 5
<i>amin</i>	počet nekonjugovaných aminových skupin	0 - 1
<i>amidin</i>	počet nekonjugovaných amidinových a guaninových skupin	0
<i>kyselina</i>	počet karboxylových skupin - COOH	0 - 1
<i>amid</i>	počet nekonjugovaných amidových skupin	0 - 1
<i>rotor</i>	počet nezákladních, neblokovaných rotačních vazeb	0 - 15
<i>rtvFG</i>	počet reaktivních funkčních skupin, přítomnost těchto skupin může vést k falešné pozitivnímu výsledku u VKS vzorkování a rozkladu, reakčním nebo toxickým problémům v <i>in vivo</i> . (viz výpis níže)	0 - 2
<i>CNS</i>	předpovězená aktivita centrálního nervového systému v rozmezí od -2 (neaktivní) do +2 (aktivní)	-2 (neaktivní), +2 (aktivní)
<i>mol_MW</i>	molekulární hmotnost	130,0 - 725,0
<i>dipol</i>	vypočítaný dipolový moment molekuly	1,0 - 12,5

vlastnosti nebo popis	popis	doporučená hodnota nebo rozmezí hodnot
<i>SASA</i>	celkový přístupný povrch pro rozpouštědlo ve čtverečných anströmech	300,0 - 1000,0
<i>FOSA</i>	hydrofobní složka SASA (nasycený uhlík a připojený vodík)	0,0 - 750,0
<i>FISA</i>	hydrofilní složka SASA	7,0 - 330,0
<i>PISA</i>	Π (uhlík a připojený vodík) složka SASA	0,0 - 450,0
<i>WPSA</i>	slabě polární složka SASA (halogeny, P, S)	0,0 - 175,0
<i>objem</i>	celkový objem přístupný rozpouštědlu ve čtverečných Ångströmech	500,0 - 2000,0
<i>Donor vodíkových vazeb</i>	odhadovaný počet vodíkových vazeb, které by byly darovány rozpouštědlem do molekul vody ve vodném roztoku	0,0 - 6,0
<i>Akceptor vodíkových vazeb</i>	odhadovaný počet vodíkových vazeb, které by byly přijaty rozpouštědlem z molekul vody ve vodném roztoku	2,0 - 20,0
<i>dip²/V</i>	odmocnina dipólového momentu dělená molekulovým objemem.	0,0 - 0,13
<i>ACxDN⁵/SA</i>	index přilnavosti v pevných látkách	0,0 - 0,05
<i>glob</i>	popis kulatosti, $\frac{4\Pi r^2}{SASA}$, r je poloměr koule s objemem rovným molekulárnímu objemu. Glob je 1 pro kulovitou molekulu	0,75 - 0,95
<i>Qppolrz</i>	předpovězená polarizabilita ve čtverečných Ångströmech	13,0 - 70,0
<i>QPlogPC16</i>	předpovězený rozdělovací koeficient hexadekan/plyn	4,0 - 18,0
<i>QPlogPoct</i>	předpovězený rozdělovací koeficient oktanol/plyn	8,0 - 35,0
<i>QPlogPw</i>	předpovězený rozdělovací koeficient voda/plyn	4,0 - 45,0
<i>QPlogPo/w</i>	předpovězený rozdělovací koeficient oktanol/voda	-2 - 6,5
<i>QPlogS</i>	předpovězená rozpustnost ve vodném roztoku	-6,5 - 0,5

vlastnosti nebo popis	popis	doporučená hodnota nebo rozmezí hodnot
<i>CIQPlogS</i>	předpovězená rozpustnost ve vodném roztoku nezávislá na konformaci	-6,5 - 0,5
<i>QPlogHERG</i>	předpovězená poloviční hodnota maximální inhibiční koncentrace pro blokování HERG K ⁺ kanálů	<-0,5
<i>QPPCaco</i>	předpovězená propustnost Caco-2 buněk v nm/sec. Caco-2 buňky jsou modelem pro bariéru střevokrev. Pouze pro neaktivní transport	<25 slabá, >500 skvělá
<i>QPlogBB</i>	předpovězený rozdělovací koeficient pro mozek/krev. Qikprop používá pouze orálně podávaná léčiva, pro příklad dopamin a serotonin nemají vliv na centrální nervový systém, protože jsou příliš polární na přechod bariéry krev/mozek.	-3,0 - 1,2
<i>QPPMDCK</i>	předpovězená propustnost MDCK buněk v nm/sec. MDCK buňky jsou dobrou napodobeninou bariéry krev/mozek. Qikprop předpovědi jsou pro neaktivní transport.	<25 slabá, >500 skvělá
<i>QPlogKp</i>	předpovězená propustnost kůží, log Kp	-8,0 - -1,0
<i>IP(ev)</i>	PM3 vypočítaný ionizační potenciál	7,9 - 10,5
<i>EA(eV)</i>	PM3 vypočítaná elektronová afinita	-0,9 - 1,7
<i>metab</i>	počet možných metabolických reakcí (viz výpis níže)	1 - 8
<i>QPlogKhsa</i>	předpověď vazby k lidskému seru albumin	-1,5 - 1,5
<i>Absorpce ústy</i>	předpovězená kvalitativní orální absorpce u člověka: 1, 2 nebo 3 pro nízkou, střední nebo vysokou	-
<i>Absorpce ústy v %</i>	předpovězená orální absorpce u člověka v rozmezí 0% - 100%. Předpověď je založena na kvantitativnímu několiknásobnému modelu lineární regrese	>80% vysoká, <25% nízká

vlastnosti nebo popis	popis	doporučená hodnota nebo rozmezí hodnot
<i>SAFluor</i>	přístupnost rozpouštědla k povrchu fluorových atomů	0,0 - 100,0
<i>SAamidO</i>	přístupnost rozpouštědla k povrchu kyslíku v amidové skupině	0,0 - 35,0
<i>PSA</i>	Van der Waals povrch polárních dusíkových a kyslíkových atomů	7,0 - 200,0
<i>NandO</i>	počet atomů dusíku a kyslíku	2 - 15
<i>Pravidlo pěti</i>	počet porušení Lipinského pravidla pěti. Pravidla jsou: mol_MW < 500, QPlogPo/w < 5, donor vodíkových vazeb <= 5, akceptor vodíkových vazeb <= 10. Látky, které splňují tyto pravidla se považují za možné léčiva. (číslo 5 odkazuje na limity, které jsou násobky pěti)	maximum 4
<i>Pravidlo tří</i>	počet porušení Jogensenova pravidla tří. Tři pravidla jsou: QPlogS > -5,7, QP Pcaco > 22 nm/s, hlavní metabolity < 7. Látky s méně (nejlépe žádným) porušením pravidel jsou více pravděpodobné být orálně dostupné.	maximum 3
<i>atomy v cyklu</i>	počet atomů v kruzích	-
<i>v34</i>	počet atomů ve třech nebo čtyř členných kruzích	-
<i>v56</i>	počet atomů v pěti nebo šesti členných kruzích	-
<i>noncon</i>	počet atomů v kruzích nemožné formovat konjugovaný aromatický systém	-
<i>nonHatm</i>	počet těžkých atomů (ne vodíkové atomy)	-
<i>Jm</i>	předpovězená maximální rychlost pro kožní transport	-

Výpis metabolických reakcí u *metab*

- Oxidace aromatického OH
- Enol oxidace
- Benzylový vodík → alkohol
- Allylový vodík → alkohol
- Sekundární alkohol → keton
- Primární alkohol → kyselina
- Terciální alkohol E1 nebo SN1
- Amin dealkylace
- Ether dealkylace
- Pyridin C2 hydroxylace
- Anilin NH → NOH nebo NCOR
- Nízká ionizační energie-lehce oxidované
- Alfa hydroxylace cyklického etheru
- Sulfoxid → sulfon
- Alfa hydroxylace karbonylu
- Alfa, beta dehydrogenace na karbonylu
- Thiol SH → SSR, SR
- Para hydroxylace arylu
- Aryl sulfid → S=O
- Redukce aryl nitro na amin
- Oxidativní deaminace primárního aminu

Výpis reaktivních funkčních skupin přispívající *rcvFG*

- Acylhalogenid
- Hetero-halogenová vazba
- NAS substrát
- Alkylhalogenid
- Halogen alfa k w-skupině
- Heteroatom v 3 uhlíkatém cyklu
- Aktivovaný cyklopropan
- Přítomný hliník (toxický)
- Přítomný křemík (toxický)
- Jednoduchá vazba hetero-hetero
- Azo, diazo nebo azid
- Akceptor karbonylu nebo derivatu
- Anhydrid nebo podobná
- Neomezený ester
- Sulfonát nebo příbuzná
- Fosfonát nebo příbuzná
- Acetal nebo podobná
- Uhličitan
- Thiol – možná oxidace
- Karbonyl v 3 uhlíkatém cyklu