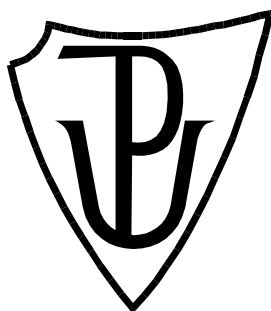


UNIVERZITA PALACKÉHO V OLOMOUCI

Přírodovědecká fakulta
Katedra fyzikální chemie



Databáze látek a jejich interakcí s membránami a membránovými proteiny

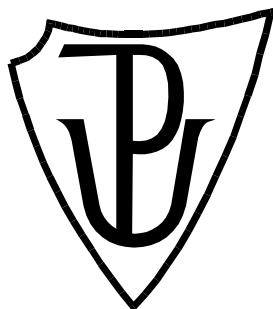
Diplomová práce

Autor:	Bc. Kateřina Storchmannová
Studijní program:	N1407 Chemie
Studijní obor:	Fyzikální chemie
Forma studia:	Prezenční
Vedoucí práce:	doc. RNDr. Karel Berka, Ph.D.
Rok:	2021

PALACKÝ UNIVERSITY OLOMOUČ

Faculty of Science

Department of Physical Chemistry



**Database of molecular interactions with membranes and
membrane proteins**

Diploma Thesis

Author: Bc. Kateřina Storchmannová
Study Programme: N1407 Chemistry
Field of Study: Physical Chemistry
Form of Study: Full-time
Supervisor: doc. RNDr. Karel Berka, Ph.D.
Year: 2021

Prohlašuji, že jsem diplomovou práci vypracovala samostatně s vyznačením všech použitých pramenů a spoluautorství. Souhlasím se zveřejněním diplomové práce podle zákona č. 111/1998 Sb., o vysokých školách, ve znění pozdějších předpisů. Byla jsem seznámena s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorský zákon, ve znění pozdějších předpisů.

V Olomouci dne

Poděkování

Na tomto místě bych ráda poděkovala doc. RNDr. Karlu Berkovi, Ph.D. za cenné rady a pomoc při vypracování této práce. Rovněž děkuji Mgr. Alžbětě Türkové, Ph.D. a Mag. Dr. Barbaře Zdražil, Priv. Doz bez jejíž přispění by tato práce nemohla vzniknout, stejně tak jako by nemohla vzniknout bez práce členů týmu databáze MolMeDB. V neposlední řadě mé poděkování patří také mé rodině za trpělivost a podporu.

Bibliografická identifikace

Jméno a příjmení autora	Bc. Kateřina Storchmannová
Název práce	Databáze látek a jejich interakcí s membránami a membránovými proteiny
Typ práce	Diplomová
Pracoviště	Katedra fyzikální chemie
Vedoucí práce	doc. RNDr. Karel Berka, Ph.D.
Rok obhajoby práce	2021
Abstrakt	Databáze jsou cennými zdroji nejrůznějších biologických a chemických informací s neustále rostoucím významem. Jednou z bioinformatických databází je také databáze MolMeDB. Tato databáze obsahuje informace o interakcích malých molekul s membránami. Cílem této práce je databázi MolMeDB rozšířit o interakce mezi malými molekulami a membránovými transportéry a poté obsah MolMeDB analyzovat.
Klíčová slova	Transportéry, biologické membrány, malé molekuly, bioinformatické databáze, MolMeDB, data mining
Počet stran	57
Počet příloh	7
Jazyk	Český

Bibliographical Identification

Author's name	Bc. Kateřina Storchmannová
Title	Database of molecular interactions with membranes and membrane proteins
Type of thesis	Diploma
Department	Department of Physical Chemistry
Supervisor	doc. RNDr. Karel Berka, Ph.D.
The year of presentation	2021
Abstract	Databases are valuable sources of various biological and chemical information ever-growing importance. One of the bioinformatics databases is MolMeDB. This database contained information about interactions among small molecules and membranes. The aim of this work is to extend the database MolMeDB with interactions of small molecules also with membrane transporters. Finally, the content of MolMeDB is analysed.
Keywords	Transporters, biological membranes, small molecules, bioinformatics databases, MolMeDB, data mining
Number of pages	57
Number of appendices	7
Language	Czech

Obsah

Teoretická část	2
1. Biologické membrány	2
2. Membránové transportní proteiny	7
3. Bioinformatické databáze	16
Metody	21
4. KNIME workflow pro data mining	21
5. Úprava dat a jejich ukládání do MolMeDB	27
6. Metody analýzy dat	28
Výsledky	30
7. Analýza dat v databázi MolMeDB	30
8. Fyzikálně chemické vlastnosti molekul v databázi MolMeDB.....	43
Závěr	52
Summary	53
Reference.....	54
Přílohy	X1

Úvod

Transport molekul přes biologické membrány představuje jeden ze základních kroků při cestě látek organismem a má naprosto klíčový význam pro organismus. Ve své nejjednodušší podobě je realizován jako prostá difúze látky. Prostá difúze ale není transportní mechanismus, který by byl vhodný pro všechny druhy molekul např. velké molekuly, nabitě molekuly nebo přímo ionty nejsou vesměs schopny procházet biologickými membránami tímto způsobem, a proto existují další transportní mechanismy pro jejich transport.

Jedním z takových to transportních mechanismů je transport pomocí membránových transportérů. Jedná se o širokou škálu velkého množství různých membránových proteinů, které jsou schopny realizovat transport molekul přes membránu. Typicky jsou transportéry schopny přenášet relativně úzkou skupinu molekul.

Z farmakologického hlediska jsou tyto transportéry velmi významné, protože mnohé z nich jsou spojeny s celou řadou genetických onemocnění, také jejich modulace je spojena s funkcí řady léčiv a v neposlední řadě mohou svojí činností působit jako nežádoucí bariéra při cestě léčiva k cíli.

S rozvojem vědy jsou neustále popisovány další a další transportéry, stejně tak přibývá informací o jejich interakcích a významu v organismech. Tyto informace jsou dnes uchovávány v celé řadě bioinformatických databází.

Cílem této diplomové práce bylo sesbírat data o interakcích mezi molekulami a membránovými transportéry pomocí *data miningu*. Následně tato data přidat do databáze MolMeDB, která do té doby obsahovala pouze interakce mezi malými molekulami a membránami. Posledním cílem bylo provést analýzu obsahu databáze MolMeDB a data o látkách se známými transportérovými interakcemi porovnat s daty o látkách s přímými interakcemi s membránami.

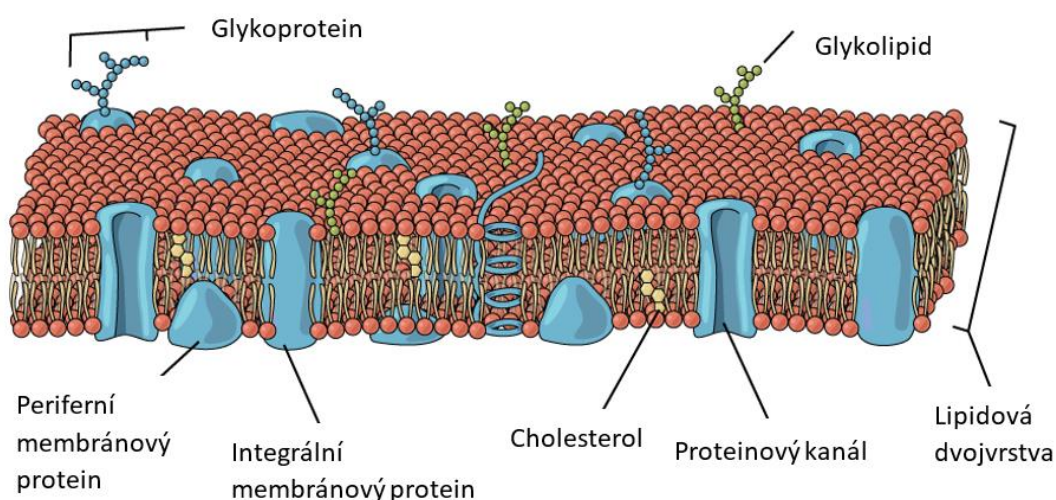
Teoretická část

1. Biologické membrány

Tato kapitola bude zaměřena na biologické membrány. Nejdříve bude věnována pozornost obecnému popisu biologických membrán a poté možnostem transportu látek přes ně.

1.1 Složení biologických membrán

Obecně lze říct, že biologické membrány jsou přirozenou hranicí buněk a některých buněčných organel. Biologická membrána bývá obvykle představována jako dvojvrstva^a tvořená lipidy, kterou znázorňuje tzv. model fluidní mosaiky.^{1,2}



Obrázek 1: Biologická membrána odpovídající modelu fluidní mosaiky (převzato a upraveno z:³)

Jak je vidět na obrázku 1, kromě lipidů biologická membrána obsahuje i řadu dalších komponent, které plní řadu důležitých funkcí v organismu. Jedná se o lipidy, proteiny a sacharidy.

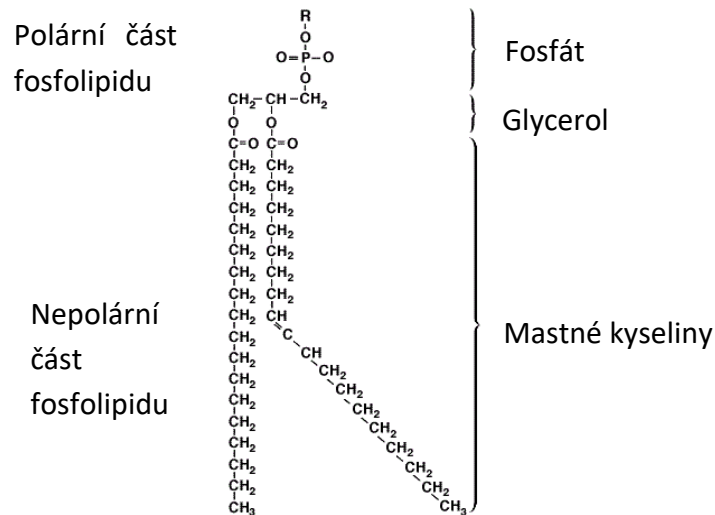
1.1.1 Lipidy biologických membrán

Jak již zaznělo v předchozím textu lipidy tvoří dominantní část biologické membrány. Z chemického hlediska rozlišujeme 3 typy lipidů biologických membrán: fosfolipidy, glykolipidy a steroly.¹

Fosfolipidy jsou složeny ze dvou mastných kyselin, které jsou připojené na glycerol.^b Glycerol je poté připojen na fosfát. Tyto tři složky tvoří základ fosfolipidů, který může být dále chemicky modifikovaný na fosfátové skupině, jak ukazuje obrázek 2, kde písmeno R vyjadřuje tuto modifikaci.

^a Existují výjimky např. membrána *strata cornea*.

^b Přesněji mastné kyseliny jsou připojeny na alkohol. Kromě glycerolu, zde může být také sfingosin. Podle alkoholu pak rozlišujeme glycerolfosfolipidy a sfingofosfolipidy. V textu práce nebude, ale sfingolipidům věnována pozornost.



Obrázek 2: Schéma chemické struktury molekuly fosfolipidu (převzato a upraveno z: <https://www.mrgscience.com/topic-13-membrane-structure.html> ze dne 3.3.2021)

Známe řadu derivátů tohoto základního skeletu např. fosfatidylcholin (PC), kde je na fosfát připojen cholin, fosfatidylserin (PS), fosfatidylentanolamin (PE), nebo fosfatidylinositol (PI).

Další velkou skupinou lipidů, které se vyskytují v biologických membránách jsou glykolipidy. Jedná se o molekuly, které jsou tvořeny opět mastnými kyselinami, glycerolem^c a cukernou částí např. glukózou.¹

Poslední skupinou lipidů membrán jsou steroly. Pokud hovoříme o sterolech živočišných membrán, pak hovoříme hlavně o cholesterolu. Cholesterol se ze strukturního hlediska od obou předchozích skupin lipidů nápadně odlišuje. Obsahuje polární hydroxylovou skupinu, která je součástí struktury sterolu, navíc cholesterol obsahuje typický cyklický steranový řetězec.¹

1.1.2 Proteiny biologických membrán

Proteiny biologických membrán plní rozmanité funkce. Některé z nich se podílí na přenosu informace, jiné mají zase transportní funkci. Jsou tedy velmi důležitou součástí buněčných membrán.¹

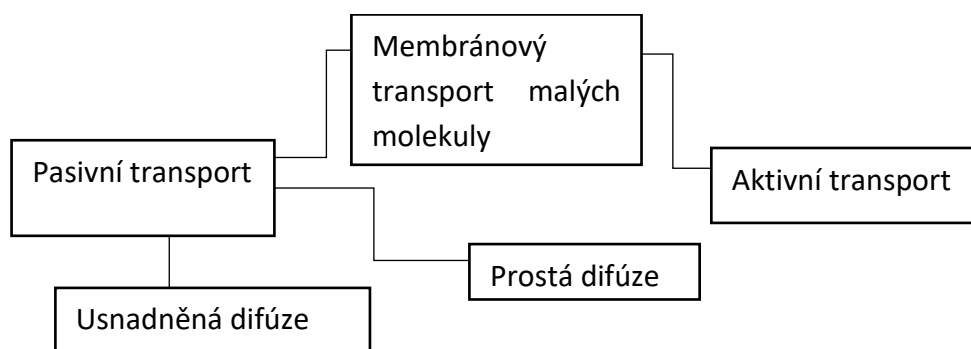
Obrázek 1 ukazuje různé druhy membránových proteinů. V nejhrubším dělení je lze rozdělit na dvě skupiny podle jejich interakce s membránou. První skupinou jsou tzv. periferní membránové proteiny, které neinteragují s hydrofobní částí lipidové dvojvrstvy, ale polárními částmi membrány. Druhou skupinou jsou tzv. integrální membránové proteiny, které jsou vnořené do membrány.⁴

^c Nebo sfingosinem

Periferní proteiny plní funkce např. strukturní jako třeba protein aktin nebo enzymatické což je případ protein kinázy C.⁴ Integrované proteiny jsou velmi často proteiny transportními, vzhledem k tomu, že transportní proteiny patří mezi stěžejní pojmy této práce bude jim věnována pozornost v samostatné kapitole (viz kapitola 2 Membránové transportní proteiny).

1.2 Transport látek přes biologické membrány

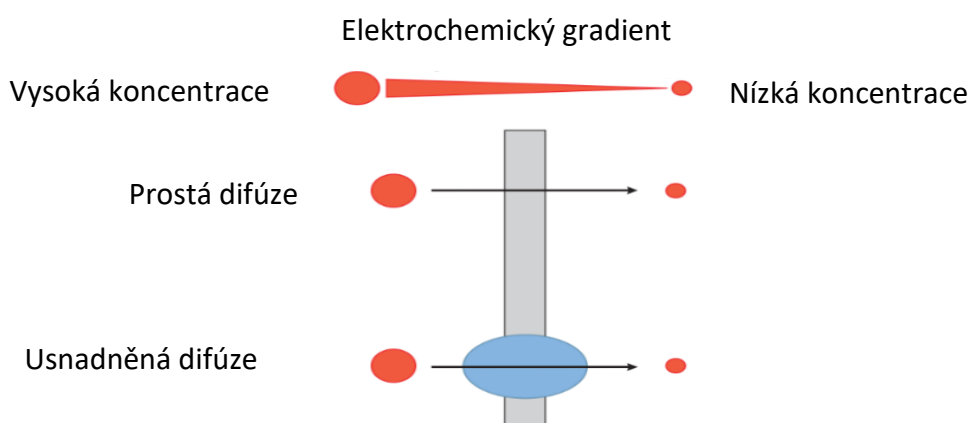
Pro molekuly existuje řada způsobů, jak buněčnou membránu zdolat. V tom nejhrubějším rozdělení můžeme transport látky přes membránu dělit na aktivní a pasivní transport,^d jak znázorňuje obrázek 3.



Obrázek 3: Schéma možností membránového transportu malých molekul (vytvořeno dle ref. ¹)

1.2.1 Pasivní transport přes biologické membrány

Za pasivní transport považujeme takový transport, který probíhá ve směru koncentračního gradientu (tedy z místa o vyšší koncentraci do místa s nižší koncentrací), rovněž pro něj platí, že nedochází ke spotřebě energie¹. Podle obrázku 3 za pasivní transport považujeme usnadněnou difuzi a prostou difuzi. Mechanismus prosté a usnadněné difúze zjednodušeně ukazuje obrázek 4.



Obrázek 4: Schématické znázornění pasivního transportu látek přes biologickou membránu (převzato a upraveno dle ref. ⁶)

^d Kromě aktivního a pasivního transportu látek přes membrány existuje ještě cytóza, kdy dochází k transportu z nebo do buňky pomocí lipidových „váčků“. Tento transportní mechanismus se uplatňuje při transportu velkých molekul např. polysacharidů, proteinů, polynukleotidů. Cytóza nebude v textu práce dále rozvedena.¹¹

Prostá difúze je nejjednodušší možností, jak může být molekula transportována přes membránu. Tento transport popisuje matematicky tzv. *Solubně-difuzní mechanismus*. Ten rozděluje celou prostou difúzi přes membránu na 3 kroky. Prvním krokem je transport molekuly z vodného prostředí do membrány, druhým je transport (difúze) látky přes samotnou membránu a třetím krokem je výstup malé molekuly z membrány opět do vody.⁵

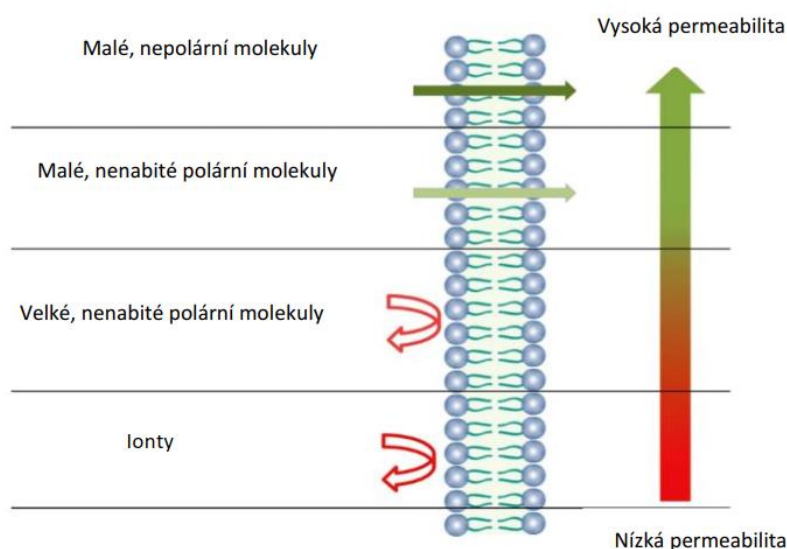
V případě, že budeme uvažovat, že membránou prochází neionizovaná molekula, můžeme vyjít z klasické fyziky a popsat prostou difúzi Fickovým prvním zákonem, který lze velmi jednoduše vyjádřit jako permeabilitu násobenou koncentračním gradientem sledované látky.⁶ Matematickými úpravami lze z tohoto zákona dostat rovnici 1, kde P_{erm} – permeabilitní koeficient látky, J – difúzní tok, Δc – rozdíl koncentrací na obou stranách membrány.

$$P_{erm} = \frac{J}{\Delta c} \quad (1)$$

Permeabilitní koeficient P_{erm} je veličina, která se udává nejčastěji v jednotkách cm/s. Je to důležitý experimentálně měřitelný údaj, který hovoří o schopnosti látky procházet biologickou membránou.⁵ Často se udává v logaritmických hodnotách $\log P_{erm}$.

Permeabilitní koeficient P_{erm} dobře koreluje s lipofilitou látky. K odhadu lipofility transportované látky, nám slouží partiční (rozdělovací) koeficient. Typicky pokud mluvíme o tomto údaji mluvíme o partičním koeficientu oktanol-voda. Partiční koeficient se typicky udává v logaritmické podobě ($\log P$), proto platí, že čím je jeho hodnota vyšší tím je látka lipofilnější.^{6,7}

Jen lipofilita látky by nebyla dostatečnou zárukou, že daná molekula bude membránou procházet. Dalšími důležitými parametry jsou také velikost a ionizace molekuly.



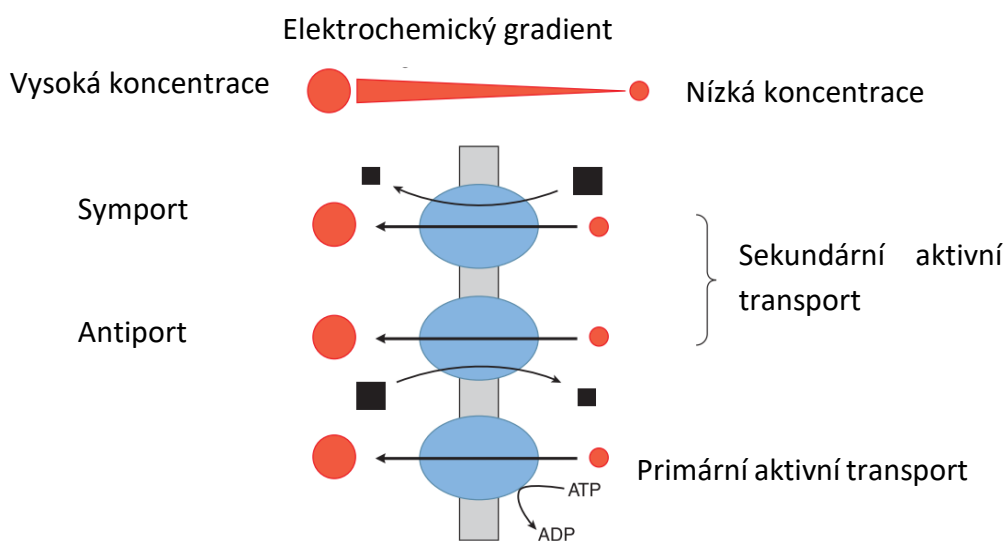
Obrázek 5: Permeabilita jednotlivých skupin látek (převzato a upraveno z: <https://mrrittner.weebly.com/unit-3-the-cell.html> ze dne 4.3.2021)

Obecně lze říct, že pasivní difuzí prochází nejlépe malé,^e neionizované, lipofilní molekuly.⁷ Obrázek 5 ukazuje, že celá řada skupin látek by nebyla schopna procházet biologickými membránami, pokud by prostá difúze byla jedinou možností jejich transportu.

Druhou možností, jak může pasivní transport přes membrány probíhat, je usnadněná difúze. Na rozdíl od prosté difúze, v tomto případě se transportu účastní proteiny, které transport molekul umožňují. V tomto případě platí, že transportní proteiny nepřenáší libovolné molekuly, ale jen úzkou skupinu molekul, pro kterou jsou určeny. Dalším rozdílem mezi prostou a usnadněnou difúzí je v tom, že usnadněná difúze umožňuje průchod ionizovaných i neionizovaných molekul.⁶

1.2.2 Aktivní transport látek přes biologické membrány

V případě, že hovoříme o tzv. aktivním transportu, tak hovoříme o způsobu přenosu transportovaných molekul za spotřeby energie. Dodat energii je v tomto případě zcela nezbytné, jelikož při aktivním transportu jsou látky přenášeny proti směru elektrochemického gradientu (z místa o menší koncentraci). Energie bývá často získávána hydrolýzou ATP.⁸



Obrázek 6: Nákres mechanismů aktivního transport, červený kruh představuje látku A, černý čtverec představuje látku B (převzato a upraveno dle ref.6)

Jak ukazuje obrázek 6, aktivní transport můžeme dále dělit na primární a sekundární aktivní transport. Primární aktivní transport je takový transport, který je přímo spojen s hydrolýzou ATP. Tento transport je spojen s přenosem jednoho druhu molekul. Zatímco při sekundárním aktivním transportu, jsou přítomny dva druhy molekul (A a B). V tomto případě vždy jeden druh molekul (A) putuje ve směru koncentračního spádu. Druhý druh (B) může putovat stejným směrem nebo opačným.

^e Myšleno, ne makromolekulární

Hnací silou sekundárního aktivního transportu je rozdíl koncentrací látky A. Pokud molekuly látky B putují stejným směrem jako molekuly látky A, pak hovoříme o symportu, pokud molekuly látky B putují opačným směrem, mluvíme o antiportu.⁶

2. Membránové transportní proteiny

Tato kapitola bude zaměřena na rozdělení membránových transportních proteinů (MTP) do jednotlivých skupin, bude detailněji popsán transport molekul pomocí MTP a dále budou uvedeni konkrétní zástupci pro každou skupinu. Pozornost bude věnována i jejich významu ve výzkumu léčiv.

2.1 Transport látek membránovými transportními proteiny

V této kapitole bude stručně popsána kinetika transportu látek pomocí transportérů a následně bude popsán rozdíl mezi transportem látek pomocí transportérů a prostou difúzí.

Transport látek pomocí MTP se řídí kinetikou Michaelise a Mentenové, kterou lze vyjádřit rovnicí 2, kde v – je, C – koncentrace substrátu, V_{max} – maximální rychlost transportu, K_M – je koncentrace substrátu, při které je rychlost toku rovna polovině maximální rychlosti.⁶

$$v = \frac{V_{max} \cdot C}{K_M + C} \quad (2)$$

K_M také zhruba odpovídá disociační konstantě komplexu transportér – substrát. Jeli C malé v porovnání s K_M pak rychlost průchodu látky přes transportér roste přímo úměrně s koncentrací transportované látky, jeli ale C větší, než K_M pak rychlost průchodu se stává konstantní a má hodnotu V_{max} .⁶

Zatímco za substrát transportéru považujeme látku, kterou je daný transportér schopen přenášet, inhibítorem je látka, která transportní schopnost MTP blokuje. Inhibitor v tomto případě nebývá substrátem daného transportéru, ale také může nastat situace, že si dva substráty téhož transportéru konkurují a tím jsou pro sebe kompetitivními inhibitory.⁹

Interakci transportér-substrát můžeme popsat pomocí již zmíněné konstanty K_M . Druhou možností, jak tuto interakci číselně vyjádřit je $EC50$ (polovina maximální efektivní koncentrace). Interakci inhibítora s transportérem naopak popisuje K_i (inhibiční konstanta) a $IC50$ (polovina maximální inhibiční koncentrace).

Jak vyplývá z předchozího textu transportéry⁸ mají řadu vlastností, které jsou typické pro proteiny – řídí se kinetikou dle Michaelise a Mentenové a jsou schopné podléhat inhibici, což jsou vlastnosti, které je zásadně odlišují od transportu prostou difúzí.¹⁰

^f Podobně jako pro enzymy i pro MTP rozlišujeme inhibici kompetitivní, nekompetitivní a akompetitivní.⁶

^g Pro účely následujícího textu budeme transport látek přes membrány dělit na prostou difúzi a transport pomocí membránových transportérů, přičemž transport pomocí transportérů bude zahrnovat usnadněnou difúzi i aktivní transport.

Další velkou odlišností mezi transportem látek transportéry a prostou difúzí je skutečnost, že transportéry přenáší pouze relativně úzkou skupinu látek např. ionty, sacharidy nebo třeba aminokyseliny, takže zatímco u prosté difúze byly pro transport rozhodující fyzikálně chemické vlastnosti jako je náboj, lipofilita nebo velikost molekuly. V případě transportu pomocí transportních proteinů je chemická struktura transportované látky zcela zásadní např. optická izomerie malých molekul se v případě tohoto transportu stává významnou.^{10,11}

V neposlední řadě je třeba si uvědomit, zejména při řešení reálných měření, že exprese konkrétních transportérů není ve všech tkáních stejná.¹⁰

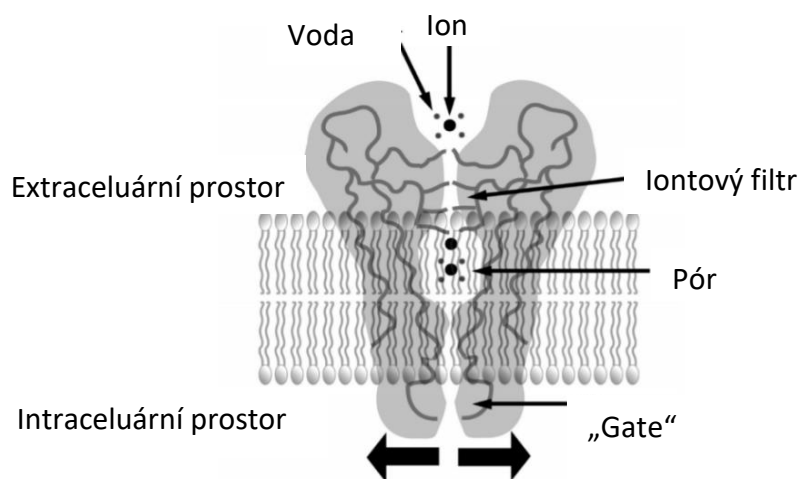
2.2 Typy membránových transportních proteinů

Obecně lze říct, jak již bylo i naznačeno v předchozím textu, že MTP jsou proteiny, které dokáží přenášet vybrané molekuly skrze biologickou membránu.¹¹

V následujícím textu práce budeme pracovat se třemi hlavními skupinami membránových transportních proteinů. (1) kanály, za něž budeme považovat transportní proteiny schopné pouze pasivního transportu. (2) přenašeče, za něž budeme považovat transportní proteiny schopné přenášet látky skrze biologickou membránu aktivně i pasivně. (3) pumpy, pumpami v textu práce budou transportní proteiny schopné aktivního transportu^{11,8}

2.3 Membránové kanály

Jak již zaznělo v předchozím textu, jsou to membránové transportní proteiny^h, které umožňují pasivní transport látek přes membránu. Schéma nejdůležitějších částí iontového kanálu ukazuje obrázek 7.



Obrázek 7: Schématické znázornění iontového kanálu (převzato a upraveno dle ref. ¹²)

^h Pojem membránový kanál je mnohem širší než iontový kanál, protože zahrnuje např. i aquaporiny.⁵⁷ V textu práce budou tyto pojmy mít stejný význam, jelikož pozornost bude věnována pouze iontovým kanálům.

Iontové kanály jsou složeny z proteinových domén, které mají uprostřed volný prostor, který je vyplněn vodou, vzniká tak silně hydrofilní cesta pro migrující ion. Stejně jako většina membránových transportních proteinů jsou i iontové kanály selektivní. Pro dosažení selektivity obsahují kanály tzv. iontový filtr, který reguluje, které ionty budou procházet.¹²

Iontové kanály mají schopnost být uzavřené nebo otevřené, tuto konformační změnu zajišťuje tzv. *gate*. Dle toho co změnu konformace kanálu vyvolává, je můžeme dále dělit¹ na napětově řízené kanály (*voltage-gated channels*), chemicky řízené kanály (*ligand-gated*) a mechanicky řízené iontové kanály (*mechanosensitive channels*).^{12,13}

Tabulka 1 ukazuje vybrané iontové kanály, se kterými bude v textu práce dále pracováno. S výjimkou epiteliálních sodných kanálů se jedná ve všech případech o kanály napětově řízené.

Napětově řízené iontové kanály se otvírají poté co dojde k depolarizaci buněčné membrány. Ligandem řízené napětově kanály přechází do otevřené konformace v okamžiku vazby konkrétního ligandu. Mechanicky řízené iontové kanály mění svoji konformaci na základě mechanického podnětu.¹³

Tabulka 1: Tabulka vybraných skupin kanálů (vytvořeno na základě ref. ¹⁴)

Přenášený ion	Název skupiny kanálů	Vybraní zástupci skupiny
K ⁺	Vápníkem a sodíkem aktivované draslíkové kanály	KCNMA1, KCNN1, KCNN3, KCNU1, ...
	Vnitřně usměrňující draslíkové kanály	KCNJ1, KCNJ2, KCNJ6, KCNJ12, ...
	Napětově řízené draselné kanály	KCNA1, KCNC3, KCND2, KCNQ5, ...
	Draselné kanály v doméně tandemových pórů	KCNK1, KCNK2, KCNK7, ...
Ca ²⁺	L-tyt	CACNB1, CACNB2, CACNB3, CACNB4, CACNA1C, CACNA1D, CACNA1F, CACNA1S
	N-tyt	CACNA1B
	P/Q-tyt	CACNA1A
	R-tyt	CACNA1E
	T-tyt	CACNA1G, CACNA1H, CACNA1I
Na ⁺	Napětově řízené sodné kanály	SCN1A, SCN3A, SCN5A, SCN8A, SCN10A ...
	Epiteliální sodné kanály	SCNN1A, SCNN1B, SCNN1D ...

¹ V literatuře je možno najít i další dělení a druhy membránových kanálů. V textu práce se omezíme jen na toto základní dělení, takže řada membránových kanálů nebude vůbec zmíněna.

^j V dalším textu práce bude věnována pozornost zejména těmto skupinám iontových kanálů.

Ligandem řízené kanály jsou velkou skupinou iontových kanálů. Ligandem může být celá řada látek např. GABA, některé aminokyseliny (např. glycin), inositoltrifosfát, acetylcholin, ATP či některé anorganické ionty.¹⁴

Napětové iontové kanály mohou být také děleny dle přenášených iontů, takže existují kanály: sodíkových iontů, draselných iontů, vápenatých iontů, chloridových iontů a mnohé další. Jejich seznam s klasifikací a odkazy do dalších databází lze najít např. ref. 14.

2.3.1 Role iontových kanálů ve farmakologii

Iontové kanály jsou důležitými regulátory buněčných funkcí. V lidském těle iontové kanály ovlivňují celou řadu procesů např. svalovou relaxaci a kontrakci, šíření nervových vzruchů, buněčnou proliferaci nebo hladinu krevního tlaku, díky čemuž jsou důležitými cíli řady léčiv.^{6,12}

V odborné literatuře je možno se setkat s pojmem „*Channelopathy*“, tento pojem označuje onemocnění způsobené chybnou činností iontového kanálu. Chybná činnost může být způsobena mutací v genu kódující daný kanál nebo v některém z regulačních proteinů. Mutace iontových kanálů způsobují např. změny v otevírání a zavírání kanálu. Dnes je takových „*Channelopathies*“ známo u člověka více než 60 a patří mezi ně např. epilepsie, osteoporóza nebo neonatální diabetes.^{12, 15}

Řada léčiv, která cílí na iontové kanály jsou jejich modulátory. Modulátory, jsou látky schopné ovlivňovat činnost kanálu. Skládají se ze dvou základních skupin látek – blokátorů (*channel blockers*) a aktivátorů (*channel openers/channels activators*). Vybrané blokátory a aktivátory ukazuje tabulka 2.

Tabulka 2: Léčiva cílící na iontové kanály (A) – aktivátor, (B) - blokátor (převzato a upraveno z ref. ¹²)

Skupina iontových kanálů	léčiva	B/A
Kanály sodných iontů	Karbamazepin	B
	Lidokain	B
	Flekainid	B
Kanály draselných iontů	Retigabin	A
	Flupirtin	A
	Sotalol	B
Kanály vápenatých iontů	Verapamil	B
	Diltiazem	B
	Amlodipin	B
	Nifedipin	B

Mezi látky ovlivňující činnost sodných iontových kanálů patří mimo jiné některá antikonvulziva např. karbamazepin, antiarytmika I. třídy např. flekainid či lokální anestetika např. lidokain. Léčiva ovlivňující draselné kanály jsou antiarytmika III. Třídy např. sotalol, některá antikonvulziva např. retigabin nebo flupirtin. Mezi léčiva ovlivňující vápenaté kanály antiarytmika IV. třídy jako jsou verapamil nebo diltiazem a léčiva na vysoký krevní tlak kam patří amlodipin a nifedipin.^{6,12}

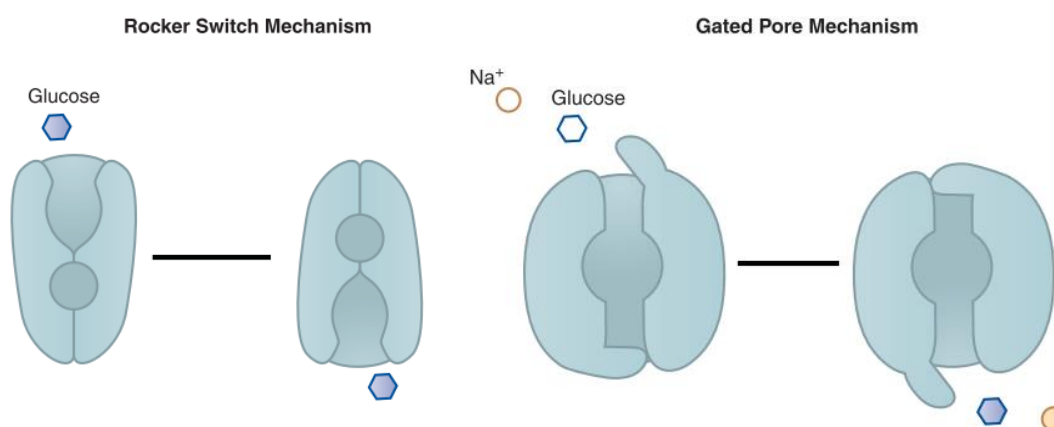
2.4 Membránové přenašeče

Jak již bylo naznačeno v předchozím textu za membránové přenašeče považujeme membránové transportní proteiny, které váží transportovaný solut. Transport solutu je spojen s konformačními změnami přenašeče.¹¹

Způsob, jakým přenáší membránový přenašeč molekuly přenášeného solutu se podobá reakci enzym – substrát. V případě transportu nedochází ke kovalentní modifikaci solutu během transportu. Každý typ přenašečových proteinů má jedno nebo více svých specifických míst, kam se solut váže.¹¹

Přenašeč se vyskytuje ve třech konformačních stavech. V prvním stavu je přenašeč otevřený na straně o vyšší koncentraci solutu. Ve stavu druhém obsahuje navázanou transportovanou látku a ve stavu třetím je přenašeč otevřen na druhé straně.^{11,16}

Obrázek 8 ukazuje dva známé mechanismy transportu pomocí přenašeče. Jsou to: *Rocker Switch* a *Gated-Pore* mechanismy^k. V případě mechanismu *Rocker Switch* se jedná o usnadněnou difúzi látky přes membránu, zatímco u mechanismu *Gated-pore* přenašeč koná sekundární aktivní transport, konkrétně symport dvou látek.



Obrázek 8: "Gated Pore" mechanismus a "Rocker Switch" mechanismus přenašečů (převzato a upraveno z ref. 6)

2.4.1 SLC přenašeče

Mezi membránové přenašeče patří tzv. SLC (*solute carrier*) transportéry. Jedná se o druhou největší skupinu membránových proteinů.¹⁷ Tato skupina (superrodina) zahrnuje ohromné množství členů (konkrétních transportérů). Názvy konkrétních členů obsahují informaci o jejich klasifikaci. SLC transportéry lze dělit do 65 tzv. rodin (SLC1 – SLC65), dále následuje písmeno, které určuje konkrétní podrodinu dané rodiny a po písmenu název obsahuje celé číslo, které reprezentuje konkrétního člena.¹⁸ Takže např. SLC1A1 je transportér, který je prvním členem podrodiny A rodiny SLC1.

^k Ještě existuje mechanismus Elevator.¹⁹

Konkrétní členy, roztržené do podrodin s informací o typu transportu, endogenních substrátech a tkáňové exprese lze najít např. <http://slc.bioparadigms.org/>.^{6, 18}

SLC transportéry jsou, podobně jako všechny přenašeče, schopné pasivního transportu a sekundárního aktivního transportu (symportu i antiportu) dokonce některé z těchto transportérů jsou i iontovými přenašeči.¹⁶

Skupiny endogenních substrátů jednotlivých SLC přenašečů ukazuje tabulka 3.

Tabulka 3: Přehled skupin endogenních substrátů pro konkrétní SLC rodiny (vytvořeno dle ref. ¹⁹)

Endogenní substráty	SLC podrodiny/ členové SLC podrodin
Aminokyseliny	SLC1, SLC7, SLC17A6, SLC17A7, SLC17A8, SLC36, SLC38, SLC43A1, SLC43A2
Anorganické kationty a anionty	SLC4, SLC8, SLC9, SLC11, SLC12, SLC20, SLC24, SLC30, SLC31, SLC34, SLC39, SLC40, SLC41
Cukry	SLC2, SLC5A1, SLC5A2, SLC37A4, SLC45, SLC50, SLC60
Lipidy	SLC5A8, SLC5A12, SLC27, SLC51
Neurotransmitery	SLC6A1, SLC6A2, SLC6A3, SLC6A4, SLC6A11, SLC6A12, SLC6A13, SLC18A1, SLC18A2, SLC18A3, SLC32
Kofaktory/Vitamíny	SLC19, SLC23A1, SLC23A2, SLC25A17, SLC25A19, SLC25A32, SLC46, SLC52
Nukleotidy, nukleosidy	SLC28, SLC29, SLC35
Peptidy	SLC15

Mezi SLC transportéry se řadí ještě tzv. SLCO transportéry, někdy se pro ně používá označení OATP („organic anion transporting polypeptides“). Tato skupina transportérů má celkem jedenáct členů organizovaných do 6 rodin¹⁸. Mezi jejich endogenní substráty patří žlučové kyseliny nebo jejich konjugáty. Konkrétními příklady endogenních substrátů jsou např. kyselina cholová, taurocholát, glykocholát atd.. Také jsou tyto transportéry schopny transportu řady hormonů a jejich metabolitů např. aldosteron, kortizol, estradiol-17 β -glucuronid, dijonodthyronin, trijonodthyronin, tyroxin. V neposlední řadě umožňují také transport leukotrienů.²⁰

2.4.2 Role SLC transportérů ve farmakologii

SLC transportéry jsou v lidském organismu hojně zastoupeny. Vyskytují se např. v ledvinách, játrech, v astrocytech nebo ve slinivce břišní. Díky širokému spektru endogenních substrátů (viz. Tabulka 3) jsou klíčové pro řadu důležitých fyziologických procesů.¹⁹

Tabulka 4 ukazuje vybraná onemocnění na jejíž řešení se podílí modulační funkce SLC transportérů. Z tabulky je patrné, že se SLC transportéry jsou asociovány běžné civilizační choroby i psychické nemoci.

Tabulka 4: Vybraná onemocnění s asociovanými SLC transportéry a vybranými léčivy (vytvořeno dle ref. ¹⁹)

SLC transportér	Onemocnění	léčivo ¹
SLC5A2	Diabetes II. typu	Canagliflozin, dapagliflozin, empagliflozin, ertugliflozin
SLC6A1	Epilepsie	tiagabine
SLC6A2, SLC6A3, SLC6A4	Úzkostná porucha, obsedantně- kompulzivní porucha, ADHD, narkolepsie	Citalopram, escitalopram, fluoxetine, fluvoxamine, desvenlafaxine, duloxetine, levomilnacipran, milnacipran, atomoxetine, benzotropine, ...
SLC9A3	Chronické renální selhání, hypertenze	tenapanor
SLC12A1, SLC12A2, SLC12A3,	Hypertenze	Furosemide, bumetanide, torasemide, ethacrynic acid, chlorothiazide, indapamide, hydrochlorothiazide, metolazone, ...
SLC2212	Dna	lesinurad
SLC25A4, SLC25A5, SLC25A6	Osteoporóza	clodronate
SLC65A2	Vysoká hladina cholesterolu v krvi	ezetimide

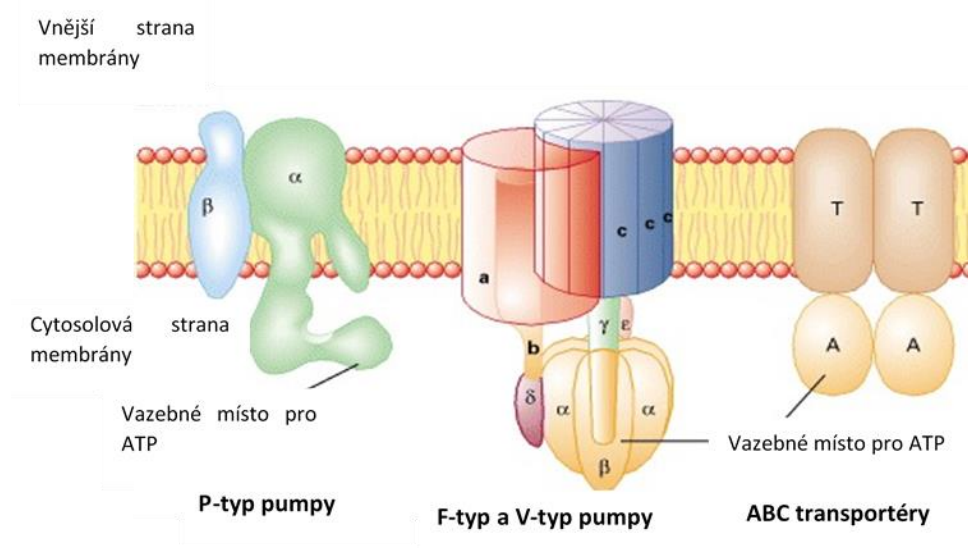
SLCO transportéry jsou rovněž z farmakologického hlediska zajímavé. Jsou schopny transportu např. statinů, antihistaminik, antihypertenziv a další. Obecně jsou tyto transportéry schopny transportovat amfipatické organické molekuly o molekulové hmotnosti větší než 300 Da, ale v některých případech transportují i neutrální molekuly či dokonce kationty.²⁰

¹ Názvy léčiv budou uváděny v anglickém jazyce.

2.5 Membránové pumpy

Za pumpy považujeme schopné přenášet molekuly z jedné strany biologické membrány na druhou pomocí primárního aktivního transportu, tedy za spotřeby energie. Energie v případě pump je získávána z ATP (ATP-řízené pumpy).¹¹

Tyto pumpy můžeme dále dělit na P-ty, F-ty, V-ty a ABC transportéry. Jejich struktury ukazuje obrázek 9. Typy F a P jsou si strukturně velmi podobné, a proto jsou znázorněny jedním obrázkem. Ve všech případech je přítomné místo pro vazbu ATP, které se označuje NBD (*nucleotide-binding domains*), na cytosolové straně proteinu.^{8,21}



Obrázek 9: Zjednodušené nákresy struktur ATP řízených membránových proteinů (převzato a upraveno z ref. ²²)

P-tyt, F-tyt a V-tyt pump můžeme nazývat shodně ATPázy. Všechny tyto ATPázy jsou schopny transportovat pouze ionty. P-tyt přenáší ionty: H^+ , K^+ , Na^+ a Ca^{2+} . F-tyt a V-tyt přenáší pouze ionty H^+ . ABC transportéry jsou aktivní transportéry, které umí přenášet kromě iontů také malé molekuly.²²

2.5.1 ATPázy

Jak již zaznělo F-tyt ATPáz transportuje pouze vodíkové protony, tato pumpa se nachází na vnitřní straně mitochondriální membrány, v chloroplastech případně v buněčných membránách bakterií. V mitochondriích je F-tyt ATPáz součástí procesu procesu oxidativní fosforylace.²³

V-tyt ATPás je přítomen např. v lysozomech, sekrečních granulách či osteroklastech nebo tkáni ledvin. Transport vodíkových iontů, které tyto ATPázy provádí vede k udržení správného pH a tím k správné funkci enzymů, které se nacházejí v cytosolických vesikulích, kde je tento typ přítomen. V případě buněk ledvin a osteroklasů, okyselují extracelulární prostor.²³

P-tyt ATPás je typ ATPás, který je tvořen několika významnými zástupci, kteří hrají velmi důležitou roli v organismu. Patří sem např. Na^+/K^+ -ATPáza, která se často nazývá sodno-draselná pumpa a patří mezi klíčové regulační proteiny koncentrace sodných a draselných iontů v organismu.²³

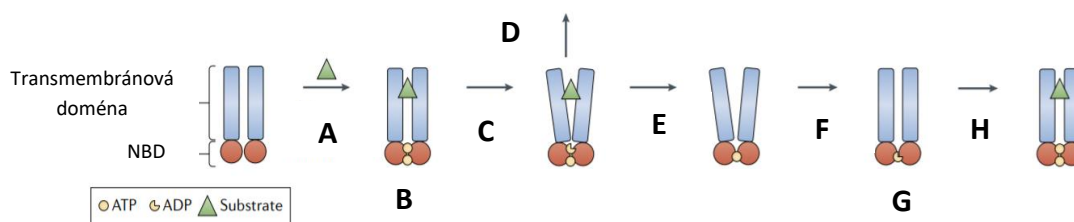
2.5.2 Role ATPáz ve farmalogii

V této kapitole zmíním pouze dvě farmakologicky významné ATPÁzy a těmi jsou již výše zmíněná Na^+/K^+ -ATPáza a H^+/K^+ -ATPáza. Na^+/K^+ -ATPáza je z farmakologického hlediska spojena s léčbou vysokého krevního tlaku, zatím co H^+/K^+ -ATPáza je spjata s léčivými pro regulaci kyseliny v žaludku.⁶

2.5.3 ABC transportéry

ABC transportéry (*ATP-binding cassette*) jsou superrodinou transportních proteinů, která dále rozpadá do sedmi podrodin (ABCA až ABCG). A celá rodina zahrnuje více než 100 členů, vyskytujících se napříč různými organismy.²²

Mechanismus transportu látek pomocí ABC transportéru ukazuje obrázek 10.



Obrázek 10: Schématické znázornění činnosti ABC transportéru (převzato a upraveno z ref. ²⁴) A – vazba substrátu do vazebného místa ABC transportéru, B – vazba dvou molekul ATP do na NBD, C – hydrolyza ATP, D – uvolnění substrátu z transportéru (eflux substrátu), tento krok je spojen s konformační změnou transportéru, E – hydrolyza jedné molekuly ADP, F – hydrolyza jedné molekuly ATP, G – návrat transportéru do původní konformace, H – opakování celého procesu znovu

Jak již bylo naznačeno v předchozím textu ABC transportéry přenáší ionty, ale také peptidy, aminokyseliny, sacharidy a řadu hydrofobních molekul. Mají tedy pro organismus velký význam.²¹ Konkrétnější přehled látek na jejichž transportu se ABC transportéry podílejí ukazuje tabulka 5.

Tabulka 5: Přehled látek na jejichž transportu se podílí ABC transportéry (převzato z ref. ²¹)

Transport	ABC transportér
steroly	ABCG5, ABCG8
cholesterol	ABCG1, ABCA7, ABCA1,
peptidy	ABCB2, ABCB3
anorganické ionty	ABCB6, ABCC7
organické anionty	ABCC2
fosfatidylcholin	ABCB4
žlučové kyseliny	ABCB11
nukleotidy	ABCC4, ABCC5
mastné kyseliny s velmi dlouhými řetězci	ABCD1

2.5.4 Role ABC transportérů ve farmalogii

Z pohledu výzkumu léčiv jsou ABC transportéry velmi zajímavé. Jsou totiž spojeny s celou řadou genetických onemocnění a také s lékovou rezistencí.

Z genetických onemocnění způsobených ABC transportéry lze vyjmenovat např. Taningerovu nemoc (ABCA1)^m, X-vázanou adrenoleukodystrofii (ABCD1, ABCD2) nebo cystickou fibrózu (ABCC7).²¹

Pokud chceme mluvit o roli ABC transportérů ve vývoji léčiv, nutně musíme narazit na pojem multilékový rezistentníⁿ protein (*Multidrug resistance protein*). Konkrétně se jedná o většinu proteinů z podrodiny ABCC a některé členy podrodin ABCB a ABCG jak ukazuje tabulka 6.

Tabulka 6: Multilékové rezistentní proteiny (převzato a upraveno z ref.^{25,6})

Značení	Gen	Značení	Gen
MDR1	ABCB1 ^o	MRP6	ABCC6
MDR3	ABCB4	MRP7	ABCC10
MRP1	ABCC1	MRP8	ABCC11
MRP2	ABCC2	MRP9	ABCC12
MRP3	ABCC3		
MRP4	ABCC4		
MRP5	ABCC5		

Tyto transportéry jsou schopné přenášet amfifilické molekuly s alespoň jedním kladným nábojem, což je případ MRP1, MRP2 a MRP3, nejrůznější analoga nukleotidů v případě MRP4, MRP5, MRP6, BCRP nebo větší neutrální či kladné molekuly jako v případě MDR1.⁶

3. Bioinformatické databáze

V následující kapitole bude věnována pozornost bioinformatickým databázím, protože ony jsou cenným zdrojem informací o interakcích malých molekul s membránami nebo membránovými transportními proteiny. Konkrétně se bude jednat o databáze: UniProt,²⁶ PubChem,²⁷ ChEMBL,²⁸ Metrabase,²⁹ IUPHAR,³⁰ a MolMeDB.³¹

3.1 UniProt

UniProt²⁶ je bioinformatická databáze uchovávající data o proteinových sekvencích i s jejich velmi obsáhlými anotacemi. V současné době tato databáze obsahuje více než 189 milionů sekvencí pro nejrůznější organismy (od virů a bakterií po člověka).²⁶ Aktuálně je na UniProt více než 564 000 manuálně anotovaných záznamů a více než 207 800 000 záznamů, které na anotaci teprve čekají.³²

^m V závorce za onemocněním je uveden ABC transportér, jehož mutace se na dané onemocnění podílí.

ⁿ Tato rezistence se týká např. léčby rakoviny. Kdy tato léčba je mařena effluxem podávaných léčiv pomocí některých transportérů z ABC superrodiny.²⁵

^o Pro tento protein je vžitý název P-glykoprotein, já v této práci budu upřednostňovat název ABCB1.

Informací, které lze pro konkrétní protein v UniProt najít je ohromné množství.²⁶ Dostupný bývá ze základních informací: popis funkce daného proteinu i s citacemi konkrétních vědeckých prací, informace o funkcích daného proteinu v organismu, exprese a lokace v buňce pro daný protein, atd. Navíc každému proteinu uloženému v databázi UniProt,²⁶ patří tzv. UniProt ID, které ho a jeho sekvenci v databázi jednoznačně identifikuje.

Z dalších informací bývají dostupná data o funkci, struktuře či topologii proteinu a odkazy na podobné proteiny v rámci této databáze a mnoho dalších informací, které jsou navíc doplněné odkazy do dalších databází.

3.2 PubChem

Pubchem patří k největším databázím v rámci biomedicínského výzkumu, bioinformatiky, chemické biologie i řady dalších odvětví.²⁷ V PubChemu se aktuálně nachází 111 M unikátních chemických struktur, 271 M záznamů o bioaktivitě z 1,2 M biologických testů.³³ Mezi přispěvatele PubChem patří cca 350 institucí např. laboratoře, univerzity, producenti chemikálií či farmaceutické společnosti. Většinu obsahu PubChem tvoří malé molekuly, lze v této databázi najít také informace o siRNA, miRNA, lipidech, peptidech či modifikovaných makromolekulách.³⁴

Databáze PubChem se skládá ze tří „vnitřních“ databází, které jsou spolu propojeny. Tyto „vnitřní“ databáze se nazývají *Compound*, *Substance* a *BioAssay*. V terminologii, kterou PubChem používá, *Substance* je databáze obsahující data, která poskytly přispěvatelé PubChem. *Compound* je databáze, které se dostávají data ze *Substance* poté co projdou standardizací PubChem.^{34,35}

V kontextu této práce bude nejdůležitější poslední z „vnitřních“ databází – databáze *BioAssay*. Tato databáze schraňuje informace z biomedicínského, farmaceutického nebo genomického výzkumu. Dají se zde najít informace o výsledcích testů bioaktivity. Pro každou malou molekulu v rámci *BioAssay* jsou dostupné výsledky všech testů a hodnoty bioaktivit^p, které jsou s ní spojeny a jsou na PubChem uloženy. Uživatelské přívětivosti velmi pomáhá, že u každého z testů je dostupný sloupec „Activity“, který jednoslovně shrnuje aktivitu dané látky. Sloupec „Activity“ často obsahuje výrazy jako „active“ nebo „inactive“.^{36,37}

3.3 ChEMBL

ChEMBL je databáze, která se zaměřuje na malé molekuly, ale lze v ní v malé míře najít také oligosacharidy, proteiny, enzymy, protilátky nebo oligonukleotidy.²⁸ Dohromady je v ChEMBL uloženo přes 2 M molekul pro více jak 14 000 cílů. V této databázi jsou k těmto molekulám dostupné nejen základní chemické informace, ale také informace o bioaktivitě nebo genomické informace.³⁸

^p Hodnoty bioaktivit jsou např. hodnoty K_i , IC50, EC50 atd.

Zdrojem dat pro ChEMBL jsou nejčastěji vybrané odborné časopisy z oblasti medicínální chemie, případně někteří přispěvatelé mohou do této databáze přispívat novými molekulami přímo, jedná se např. o farmaceutické společnosti.

Pro každou malou molekulu uloženou v této databázi je dostupná její „Compound Report Card“, ve které lze najít všechny informace o dané látce. Kromě zcela běžných informací, je zde dostupný i např. odkaz na klinické zkoušky v sekci „Clinical Data“, kde jsou odkazy do dalších databází, které se zaměřují na klinické informace. Sekce „Activity Chart“ se dále rozpadá na: „Bioactivity Summary“, „Assay Summary“ a „Target Summary“. „Bioactivity Summary“ lze najít hodnoty veličin jako jsou: $\log P$, K_d , IC_{50} , atd., které popisují interakci dané malé molekuly s membránou nebo s proteinem. „Assay Summary“, zde jsou uvedeny testy, které byly provedeny. „Target summary“ obsahuje informace o konkrétních cílech dané malé molekuly. Cíly mohou být např. enzymy, receptory, transkripční faktory atd.

Kromě samotného ChEMBL se jeho vývojáři také starají o další projekty. Jedním z nich je UniChem. UniChem je databáze, která umožňuje přes identifikátor látky např. InChI^q nebo InChIKey^r najít hledanou látku v dalších databázích. Získáme tedy odkazy do těchto dalších databází.³⁹

3.4 Metrabase

Na rozdíl od databází PubChem a ChEMBL je databáze Metrabase (Metabolism and Transport Database) podstatně menší a mnohem úžeji zaměřená.²⁹ Databáze obsahuje informace o interakcích malých molekul s transportéry a cytochromy P450. Informace o transportérech jsou v Metrabase dostupné pro 20 transportérů, jedná se o transportéry z řad rodin SLC a ABC.²⁹ Přes 60 % dat, která Metrabase obsahuje byla vyhledána v literatuře. Zbytek připadá na databáze: 20,5 % dat pochází z TP-Search⁴⁰, 10 % pochází z ChEMBL, 6,5 % z HITDB (Human Intestinal Transporter Database).⁴¹

Vyhledávání v rámci této databáze je možné dvojím způsobem. Buď jsou vyhledána všechny malé molekuly související s daným proteinem (*search by protein*) nebo jsou vyhledány všechny proteiny, které mají nějakou interakci s vybranou malou molekulou (*search by compound*). V případě *search by compound* je možné strukturu malé molekuly nakreslit nebo zadat její SMILES,^s SMART,^t InChI nebo MOL soubor.²⁹ Metrabase obsahuje nejen experimentálně stanovené hodnoty jako jsou IC_{50} , K_i , K_M a V_{max} , atd., ale také „Action Type“. „Action Type“ stručně popisuje vztah mezi malou molekulou a proteinem např. substrát, inhibitor, represor, stimulant, atd. Dále lze v Metrabase najít odkazy do původní literatury, bývá dostupné DOI a PubMed ID daného článku.

^q InChI – (IUPAC International Chemical Identifier) mezinárodně uznávaný identifikátor molekul

^r InChIKey – hashovaná verze InChI

^s SMILES – (simplified molecular-input line-entry system) jedná se o řádkovou formu zápisu struktury molekuly, která je čitelná pro počítače.

^t SMART – (SMILES arbitrary target specification) je rozšířením pojmu SMILES. Podrobnější vysvětlení tohoto pojmu je možno najít na:⁵⁸.

3.5 The IUPHAR/BPS Guide to PHARMACOLOGY

The IUPHAR/BPS Guide to PHARMACOLOGY (dále pouze IUPHAR) je manuálně anotovaná databáze, která se zabývá daty o interakci ligand – cíl (nejčastěji protein).³⁰ Z ligandů můžeme v této databázi najít např. léčiva (schválená i již nepoužívaná), metabolity, protilátky nebo třeba anorganické látky. Dohromady aktuálně obsahuje tato databáze 2 976 cílů a 10 659 ligandů. Tato databáze jako zdroj svých dat používá především literaturu zabývající se danou tematikou.⁴²

Vyhledávání v rámci této databáze je možné podle cílů, podle ligandů, onemocnění či případně je možné pokročilé vyhledávání. Cíle v rámci této databáze jsou rozděleny osmi skupin – receptory spřažené s G-proteiny, iontové kanály, jaderné receptory hormonů, kinázy, katalytické receptory, transportéry, enzymy, jiné proteinové cíle.

Pro každý proteinový cíl je možné si otevřít další jeho dělení založené na biologii. Následně ve vybrané skupině jsou pro každého člena uvedeny informace jako jsou: název, podjednotky, popis funkce, případně informace o inhibitech, blokátorech, aktivátorech atd. včetně hodnot jejich bioaktivit a odkazů na primární zdroje.

3.6 MolMeDB

MolMeDB³¹ (Molecules on Membranes Database) je otevřená databáze, která aktuálně obsahuje více než 95 000 interakcí pro téměř 457 000 molekul. Kromě interakcí malých molekul přímo s membránou byla v rámci této práce rozšířena informace o transportu látek přes membránu pomocí transportérů.

Ke každé látce, která je v MolMeDB uložena jsou dostupné obecné informace jako jsou SMILES, InChIKey, molekulová hmotnost nebo vypočtený $\log P$. Dále jsou dostupné odkazy na další databáze jako jsou PubChem²⁷, DrugBank⁴³, ChEBI, PDB⁴⁴ a ChEMBL.³¹ Sekce „Interactions“ a „Free Energy Graph“ obsahují informace o interakci dané malé molekuly s membránami, které jsou v MolMeDB³¹ dostupné. Údaje, které můžeme v tabulce „Interactions“ najít jsou: název membrány, název metody, náboj molekuly (Q), teplota (T), rozdělovací koeficient membrána – voda ($\log K_M$), permeabilitní koeficient ($\log Perm$), velikost penetrační bariéry (ΔG_{pen}) a míra afinity ligandu k membráně (ΔG_{wat}) a reference. V některých případech je možno zobrazit v sekci „Free Energy Graph“ profil volné energie.³¹

V případě, že je v MolMeDB³¹ uložena informace o průchodu látky membránou pomocí transportéru, informace o takové interakci je uložena v sekci „Transporters“. V této sekci lze najít název transportéru, jeho UniProt ID, informaci o vztahu mezi molekulou a transportérem, která je vyjádřena jednou ze čtyř možností – inhibitor, non-inhibitor, substrát, non-substrát. Tato informace je obsažena ve sloupci „Type“. Dále je v tabulce „Transporters“ uvedena numerická hodnota příslušné experimentálně stanovené bioaktivity – pK_i , pK_M , $pIC50$ nebo $pEC50$.^u Dále je zde informace o referenci a případně poznámka.

^u Přesněji jedná se o záporné dekadické logaritmy těchto hodnot.

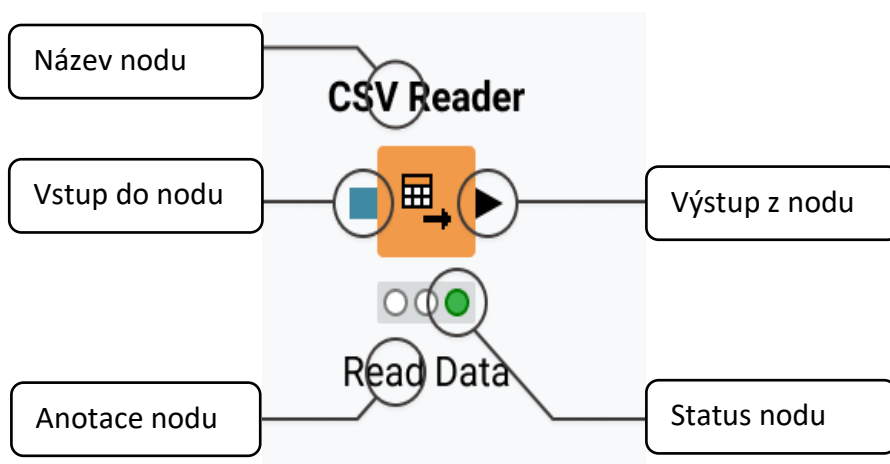
3.7 Metody získávání dat z bioinformatických databází

Jak již naznačuje předchozí text dat uložených v bioinformatických databázích je ohromné množství a není vždy v lidských silách všechna data procházet ručně a získávat z nich potřebné informace. Z tohoto důvodu se stávají stále populárnější a běžnější způsoby získávání dat pomocí počítačů. Tomuto počítačovému způsobu získávání dat se říká *data mining*.

Existuje celá řada programů, které *data mining* umožňují, ale v následujícím textu bude zmíněn jen software KNIME, který jsme použili.⁴⁵

3.7.1 KNIME software

KNIME je open-source program, který je vyvinutý pro datovou analýzu.⁴⁵ Prostředí tohoto programu umožňuje uživateli řešit jeho úlohy bez psaní kódů v programovacím jazyce, ale interaktivně pomocí tzv. nodů (uzlů). Node reprezentuje nějakou jednu úlohu, která má být vykonána. Úloh, které mohou různé nody řešit je velké množství např. čtení souborů, zapisování do souborů, transformace dat, vizualizace dat do tabulek či grafů atd. Vzhled a důležité části nodu ukazuje obrázek 11.⁴⁶



Obrázek 11: Node programu KNIME s popisem základních částí (převzato a upraveno z: <https://www.knime.com/getting-started-guide> dne 3.3.2021)

Každý node je opatřen názvem, který stručně popisuje jeho činnost, vstupem a/nebo výstupem a statusem. Status udává, zda je daný nod správně nakonfigurován a připraven ke spuštění (žlutá barva), není nakonfigurován (červená barva) či byl úspěšně spuštěný (zelená barva) nebo nastala nějaká chyba.⁴⁶

Nody jsou opatřeny vstupy a výstupy skrze které je možno je propojovat a vytvářet tak komplexní systémy řešení uživatelových úloh. Těmto vzájemně propojených nodů se říká workflow.⁴⁶

V případě, že je workflow složitější je výhodné vytvořit tzv. metanody, což jsou v podstatě subworkflows v rámci hlavního workflow.⁴⁶

Metody

V této sekci jsou postupně popsány metody, které byly využity k realizaci této diplomové práce. Nejdříve představím KNIME workflow, poté princip ukládání dat do databáze MolMeDB a následně analýzu výsledných dat.

4. KNIME workflow pro data mining

Data o průchodu látek biologickou membránou pomocí některého z membránových transportních proteinů byla získávána z bioinformatických databází. Konkrétně zdroji, které byly využity pro získání dat byly databáze: PubChem,²⁷ ChEMBL,²⁸ Metrabase,²⁹ a IUPHAR.³⁰

Data byla získávána pomocí workflow vytvořeného v programu KNIME.⁴⁵ Základní strukturu tohoto workflow ukazuje obrázek 12. Toto workflow je zaktualizovanou verzí workflow publikovaného v práci⁴⁷ a jeho autorkou je Mgr. Alžběta Türková, Ph.D.

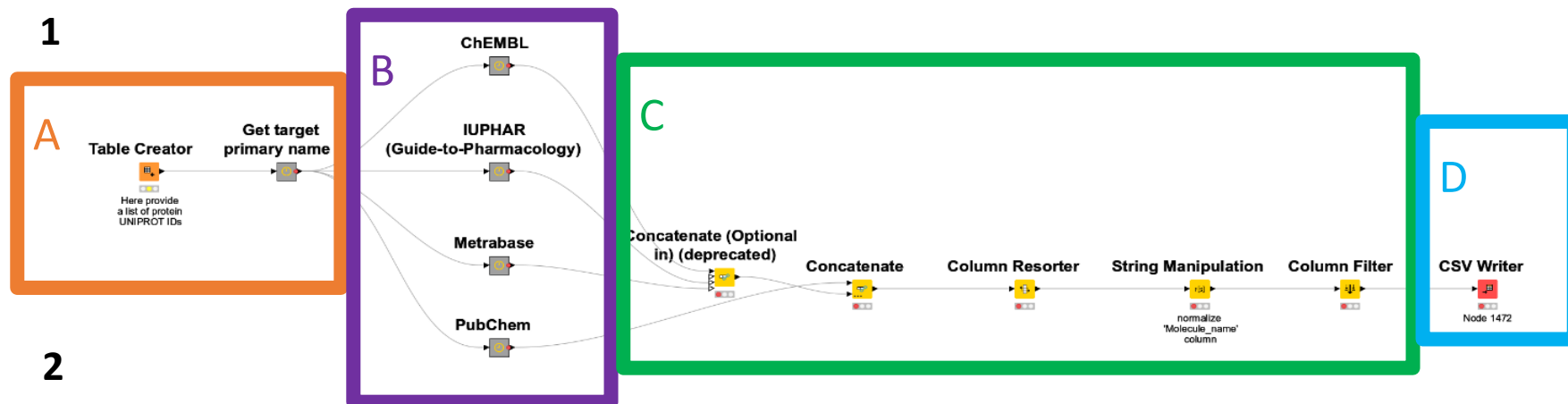
Node "Table Creator" slouží jako vstupní node do celého procesu *data miningu*. Do jeho tabulky se vkládají UniProt ID transportérů, pro která mají být data získána. Následuje metanode „Get Target Primary Name“, který přiřadí k příslušnému UniProt ID název transportéru např. UniProt ID P33527 přísluší název transportéru ABCC1, který je v tomto kroku získán. Část workflow zodpovědná za tuto činnost je označena na obrázku 12 jako oblast A. Oblast B na obrázku 12 označuje jednotlivé metanody pro získávání dat z výše zmíněných databází. Oblast B bude podrobněji popsána v textu níže. Oblast C slouží ke spojení dat získaných z různých databází (node „Concatenate“) a k finální úpravě dat (node „Column Resorter“ a node „string manipulation“). Poslední oblastí je oblast D, kde proběhne export výstupu z workflow do csv souboru pomocí nodu „CSV Writer“.

4.1 Získávání dat z konkrétních databází

Pro každou databázi jsme připravily subworkflow (metanode), které zajišťuje získávání dat z ní, jak ukazuje obrázek 12-1B. Tyto nody pracují až na jednu výjimku (node pro databázi Metrabase) s webovými servisí daných databází. Proto má každé subworkflow specifickou konfiguraci. Obrázek 12-2 ukazuje specifickou konfiguraci metanodu pro databázi PubChem.

Subworkflow pro získávání dat, z databází PubChem, ChEMBL a IUPHAR využívají webových servisů (API). Ve všech případech a začátek procesu je ve všech případech analogický.

Konec všech čtyř metanodů je také velmi analogický. Data jsou upravena do finální podoby, což v sobě obsahuje odstranění nepotřebných sloupců, přejmenování sloupců, tak aby sobě si odpovídající sloupce ze všech čtyř nodů měly stejné názvy a bylo možno je poté spojit do jednoho výstupu (nody „Concatenate“, obrázek 12C) a přidání sloupce „Source“, kde se objeví název konkrétní databáze. Zároveň na konci všech čtyř subworkflow probíhá standardizace, kterou popisuje kapitola 4.2.



Obrázek 12: Workflow vytvořené v programu KNIME pro získávání dat z vybraných databází (oblast 1 znázorňuje workflow jako celek, oblast 2 ukazuje v detailu metanode pro databázi PubChem)

4.1.1 PubChem subworkflow

Toto subworkflow nejdříve vytvoří URL adresu pro každý transportér, jehož UniProt ID bylo v nodu „Table Creator“ (obr. 12-1A). Pomocí nodu „String Manipulation“ se vytvoří nový sloupec „URL AID“. Z tohoto sloupce se pak získají příslušná AIDs z PubChem BioAssay. Následně jsou malé molekuly, které jsou v rámci daného záznamu v BioAssay zmíněny, vyextrahovány. Na základě svých identifikátorů CID (Pubchem Compound Identifier).

Dále jsou z tohoto záznamu získány informace o aktivitě a o publikaci, ze které data pochází (Pubmed ID).

Podle identifikátorů CID byl dále získán kanonický SMILES molekul a případně jejich obecné názvy. V případě, že obecný název molekuly nebyl obsažen v databázi PubChem, je jako název molekuly uloženo její ChEMBL ID. Pro případy, že je ChEMBL ID nalezeno jako název molekuly, spustí se další subworkflow, které je pro tyto případy vnořeno do tohoto a pokusí se název stáhnout z databáze ChEMBL.

Následně kromě standardizace nastupují finální úpravy, jak byly popsány výše.

4.1.2. ChEMBL subworkflow

V tomto subworkflow jsou nejdříve získány ChEMBL ID příslušných transportérů. Následně jsou získány informace o bioaktivitě a Pubmed ID. O tyto kroky se starají metanody „Get Bioactivities per Target“ a „Get Pubmed ID“

Informace o bioaktivitě jsou ještě profiltrovány, tak aby se ve finálním výstupu objevily jen výše definované informace. Tento proces běží zvlášť pro substráty a zvlášť pro inhibitory a zajistí, že ve výsledném výstupu budou pouze hodnoty bioaktivit odpovídající end-pointům, které si zadefinujeme pro substráty a inhibitory.

Nakonec jsou přidány ještě SMILES jednotlivých molekul a výstup je upraven do jednotného formátu. Takže opět proběhne proces standardizace a výše popsané finální úpravy.

4.1.3 IUPHAR subworkflow

Po získání IUPHAR ID daných cílů. Podle IUPHAR ID daných transportérů jsou následně získávány jejich ligandy a bioaktivita. V tomto případě ze souborů ve formátu JSON. O toto se starají metanody „Substrates“ a „Inhibitors“. Následně jsou informace doplněny o SMILES a opět proběhne standardizace a úprava dat do konečné podoby, jak bylo popsáno výše.

4.1.3. Metrabase subworkflow

V případě získávání dat z database Metrabase²⁹ je postup odlišný než u ostatních databází, zde není využit webový servis databáze, ale nody „HttpRetriever“ a „HtmlParser“.

Subworkflow má na svém počátku CSV soubor, ve kterém jsou uvedeny názvy všech transportérů, které jsou v databázi Metrabase a jejich UniProt ID. Do tohoto subworkflow rovněž vstupují data z subworkflow „Get target primary name“ (obrázek 12-1A). Podle UniProt ID jsou data ze vstupu přiřazena k informacím z CSV souboru. Následuje samotné získávání dat z Metrabase. Následně jsou data standardizována a upravena do finální podoby a vystupují z workflow.

4.2 Standardizace dat v rámci workflow

Standardizace SMILES, která byla použita v rámci workflow je inspirována prací.⁴⁸ Proces standardizace probíhá v každém subworkflow, které je na obrázku 12-1B. Tento proces lze stručně popsat následujícím způsobem.

Nejprve jsou ze SMILES jednotlivých molekul odstraněny symboly, které souvisí se stereoizomerií molekuly jako jsou @; \;/. Tento krok je nutný, aby došlo ke sjednocení SMILES, které pocházejí z různých zdrojů.

V případě solí dojde k odstranění příslušného iontu a ze SMILES soli se vytvoří SMILES elektroneutrální molekuly. Proces odstranění soli zajišťuje node „RDKit Salt Stripper“, zatímco proces neutralizace náboje má na starost node „RDKit Structure Normalizer“. Oba jsou součástí balíku nodů od společnosti RDKit,⁴⁹ které jsou v rámci programu KNIME volně dostupné. Dalším krokem je kontrola, jaké atomy jsou ve strukturách. V tomto kroku je zkontrolováno, zda SMILES obsahují pouze atomy H,C,N,O,F,Br,I,Cl,P,S.

Na konci procesu standardizace dostáváme kanonický SMILES molekuly, její InChI a InChIKey.

4.3 Výstup z workflow

O export získaných dat z workflow se stará node „CSV writer“, který je na obrázku 12- 1D. Po exportu je získán CSV soubor, který obsahuje jeden řádek pro každou získanou informaci. Informace jsou uspořádány do 11 sloupců, jak znázorňuje tabulka 7.

Tabulka 7: Informace získané pomocí data miningového workflow

Název sloupce	Popis
Molecule name	Obecný název molekuly
UniProt ID	Identifikátor daného transportéru v databázi UniProt
Target primary name	Název transportéru
Activity type	Obsahuje názvy veličin (EC50, IC50, K_i , K_M)
Activity pValue	Záporný dekadický logaritmus hodnoty veličiny ze sloupce „Activity type“
Activity label ^v	Sloupec obsahuje hodnotu 0 pro non-substráty a non-inhibitory a hodnoty 1 pro substráty a inhibitory
Pubmed ID	Identifikátor vědecké práce, ze které daná informace pochází
Source	Databáze, ze které byla daná informace získána
Molecule InChI	IUPAC identifikátor dané látky
Molecule InChIKey	Hashovaná verze InChI
Molecule canonical SMILES	SMILES dané molekuly

5. Úprava dat a jejich ukládání do MolMeDB

V této kapitole popisují přípravu dat pro ukládání a validaci pro MolMeDB.³¹

5.1 Úprava dat

Surová data získaná exportem z KNIME workflow se pro nahrání do MolMeDB nehodí a je třeba je upravit do finální podoby.

Prvním důležitým krokem bylo odstranění případných duplicit a následně výpočet aritmetického průměru hodnot „Activity pValue“, pokud některé dva a více řádků obsahovaly stejné hodnoty ve sloupcích: „Molecule Name“, „UniProt ID“, „Target Primary Name“, „Pubmed ID“, „Source“, „Activity Type“ a „Molecule Canonical SMILES“. V tomto případě byl kromě aritmetického průměru vypočtena také směrodatná odchylka a znovu přepočítána hodnota sloupce „Activity Label“

Na základě sloupců „Activity Type“ a „Activity label“ byl vytvořen nový sloupec „Type“, který obsahoval jednu z hodnot (INH, NIHN, SUB, NSUB)^w. Pokud sloupec „Activity Type“ obsahoval EC_{50} nebo K_M a sloupec „Activity Label“ obsahoval hodnotu 1, pak sloupec „Type“ obsahoval výraz SUB. Pro stejné hodnoty sloupce „Activity Type“, ale pro hodnotu 0 v sloupci „Activity label“ dostáváme hodnotu NSUB ve sloupci „Type“. Analogická úvaha platí i pro hodnoty INH a NIHN. Pokud sloupec „Activity Type“ obsahuje K_i nebo IC_{50} a sloupec „Activity label“ obsahuje 1, pak dostáváme ve sloupci „Type“ hodnotu INH. Je-li „Activity label“ 0 pak hodnota „Type“ je NINH.

^v Toto binární značení je založeno na nelogaritmované hodnotě sloupce „Activity pValue“. Kdy jako hraniční hodnota bylo zvoleno 10 $\mu\text{mol/l}$.

^w INH – inhibitor, SUB – substrát, NINH – non-inhibitor, NSUB – non-substrát

Další úpravou bylo vytvoření sloupců „ pK_M “, „ pK_i “, „ $pIC50$ “ a „ $pEC50$ “, do kterých se přesunuly příslušné hodnoty ze sloupců „Activity pValue“. Tyto čtyři sloupce byly vytvořeny podle sloupce „Activity Type“. Takto upravený dataset je možno nahrát do databáze MolMeDB.

5.2 Nahrávání dat do MolMeDB

MolMeDB má pro nahrávání dat o transportérech uploader. Tento uploader umí pracovat s následujícími daty, která ukazuje tabulka 8.

Tabulka 8: Sloupce uploaderu MolMeDB

Název sloupce	význam
Name	Název malé molekuly
Target	Název transportéru
SIMILES	SMILES dané molekuly
Uniprot ID	UniProt ID
pK_M / pK_i / $pIC50$ / $pEC50$	Sloupec s numerickými hodnotami těchto veličin
K_M_{acc} / K_i_{acc} / $IC50_{acc}$ / $EC50_{acc}$	Sloupec směrodatných odchylek průměrů hodnot
Type	Obsahuje hodnoty INH, NINH, SUB, NSUB
Primary reference	Pubmed ID
Secondary reference	Název databáze, ze které data pochází
Note	Poznámka (není povinná)

MolMeDB³¹ považuje dva řádky za duplicitní, pokud mají shodné SMILES, název molekul, transportér (sloupce „Target“ a „UniProt ID“), primární a sekundární referenci, či poznámku.

V případě, že dataset obsahuje místo názvu ChEMBL ID, pak je možno tento ChEMBL ID vložit do sloupce „Name“ a MolMeDB tento identifikátor sama přiřadí pomocí dotazu do PubChem pomocí veřejné API.

Po nahrání dat proběhne vygenerování InChIKey podle SMILES, po na základě vygenerovaného InChIKey jsou doplněny odkazy do dalších databází. Pro účely doplnění těchto externích odkazů je MolMeDB napojena na databázi UniChem.³⁹ Následně jsou nová data připravena na validaci.

6. Metody analýzy dat

Data byla pomocí workflow získána, upravena a následně nahrána do databáze MolMeDB. Když skončil proces validace dat, který v databázi je nastaven, tak jsem finální data stáhla a analyzovala.

Analýza dat probíhala v drtivé většině v programu RStudio,⁵⁰ který je velmi vhodný pro statistické zpracování dat. V tomto programu byla rovněž vytvořena většina grafů, s výjimkou teplotních map a bodových grafů, které byly vytvořeny pomocí knihovny Seaborn v Pythonu.

Při vytváření grafů jsem k datům přistupovala různými způsoby. Při tvorbě sloupcových grafů a koláčových diagramů jsem zpravidla počítala se všemi záznamy, které splňovaly danou podmínku např. daný typ membrány, metody nebo transportéru.

Při vytváření Vennových diagramů v RStudiosu jsem používala také funkci „intersect“, která umožňuje provádět množinovou operaci průnik v tomto programu. V případě těchto diagramů byl průnik realizován vždy podle sloupce se SMILES molekul. Tato funkce způsobí to, že se každá položka sloupce se pro každou množinu^x započítala jen jednou. Pokud se daný SMILES vyskytoval v množině vícekrát, byl do výsledného průniku v rámci diagramu započten jen jednou. Stejně tak byl každý SMILES započten jen jednou i v oblastech grafu mimo průniky, aby byly tyto diagramy konzistentní.

Cílem analýz v rámci mé práce bylo prostudovat data uložená v databázi MolMeDB. Data jsem studovala z hlediska metod, membrán, transportérů a jejich vybraných fyzikálně chemických vlastností.

^x Množinou jsou myšleny např. konkrétní databáze nebo typy interakcí.

Výsledky

Pomocí KNIME jsme získali 25 827 interakcí pro 14 747 molekul z databází PubChem, ChEMBL, Metrabase a IUPHAR. Data získaná pomocí KNIME workflow byla po úpravách do konečné podoby nahrána do databáze MolMeDB. Tato data jsme zanalyzovali nejprve vůči sobě a také vůči ostatním záznamům, které již v MolMeDB byly.

7. Analýza dat v databázi MolMeDB

Analýzu dat lze rozdělit na tři hlavní části. První část se týkala dat (molekul), která se již v MolMeDB nacházela před přidáním dat o interakcích s membránovými proteiny. Tato data popisují interakci malých molekul s membránou. Druhou oblast tvořila data z workflow, která se podařilo nahrát do databáze, a která úspěšně prošla procesem validace v databázi MolMeDB. Tato oblast obsahuje informace o interakci molekul s transportními proteiny membrán. Třetí oblastí bylo potom zkoumání překryvu dat z prvních dvou skupin.

7.1 Analýza dat interakce molekul s membránou

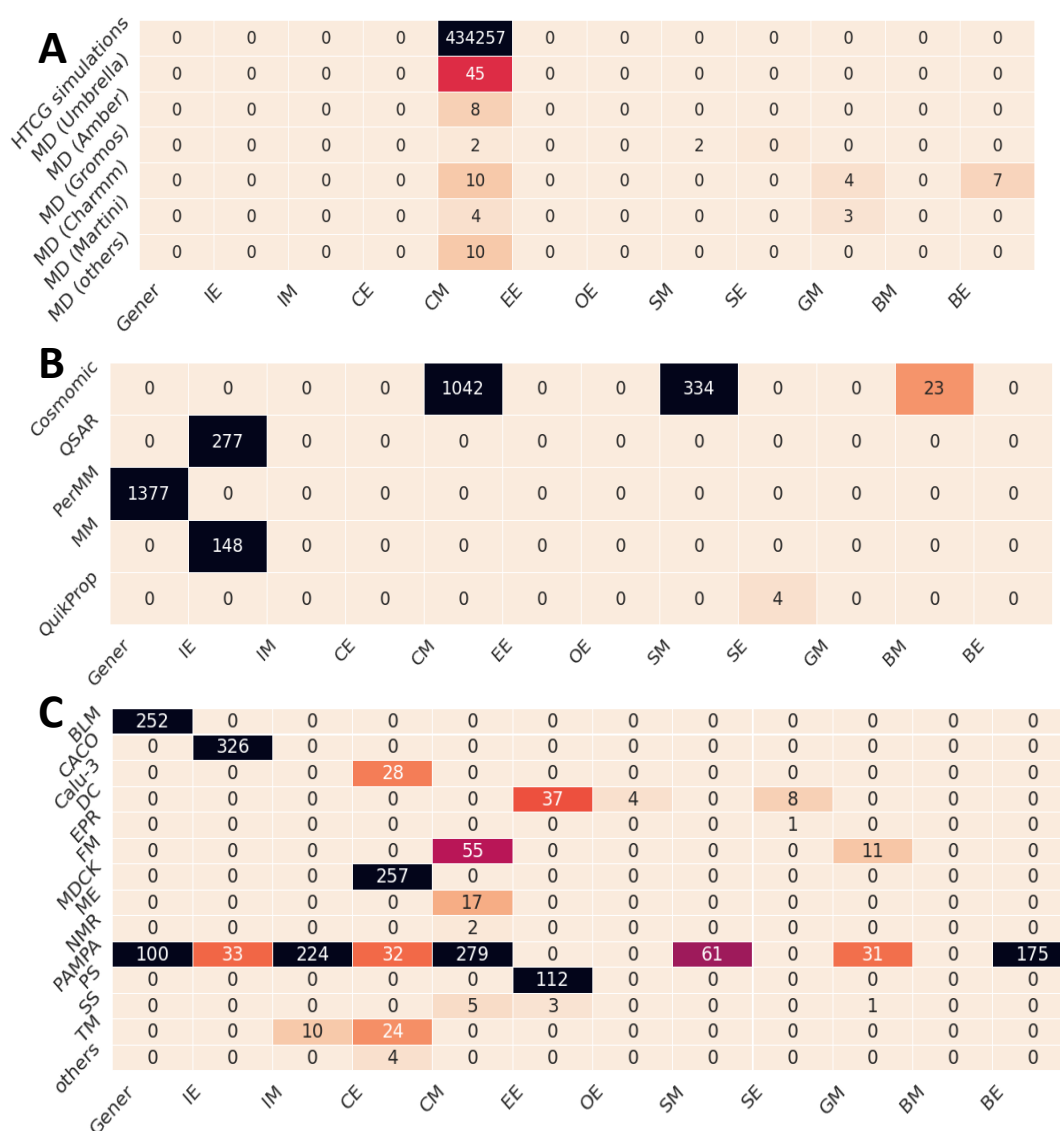
Jak již bylo naznačeno, tato část se týká analýzy informací, které na MolMeDB byly před přidáním dat z workflow. Tato data stále tvoří drtivou většinu obsahu databáze. Celkově se jedná o více než 435 000 interakcí, které byly určeny 59 metodami pro 43 membrán.

Tato sekce obsahuje opravdu velmi různorodá data. Z hlediska metod můžeme tato data primárně rozdělit na ta, která jsou experimentálního původu a na ta, která byla získána nějakou výpočetní metodou. Data získaná výpočetně můžeme dále dělit na data získaná simulacemi a vypočítaná. Z hlediska metod MolMeDB obsahuje ke dni 6.5.2021 nejvíce dat ze simulací (434 442 záznamů). Následují s odstupem data z metod výpočetních (3 205 záznamů) a z metod experimentálních (2 091 záznamů). Toto dělení je ale opravdu velmi hrubé, proto jsem si pro práci s těmito daty vytvořila menší skupiny metod, které budou dostatečně dobře reflektovat pestrost metod v MolMeDB a zároveň zjednoduší práci s daty a zlepší čitelnost výsledných grafů.

Membrány jsou v rámci MolMeDB děleny na modelové a buněčné. Dále jsou membrány v MolMeDB děleny podle „původu“ do 7 skupin („brain“, „cell“, „eye“, „intestine“, „oral“, „skin“ a „generic“). Podle tohoto dělení získáme 13 skupin membrán. Název každé skupiny, kterou v textu práce používám, se skládá ze dvou písmen. První písmeno je počáteční písmeno názvu „původu“ membrány a druhé písmeno odkazuje na to, zda je jedná o experimentální nebo modelovou membránu např. membrána DMPC je ve skupině CM, jedná se o modelovou membránu buňky. Pokud není membrána ve zdrojové publikaci popsána, aby bylo možno ji zařadit, tak pak je v MolMeDB označena jako GENER. Celkově tedy máme 14 skupin.

V příloze této práce jsou tabulky ukazující výše nastíněného dělení do skupin metod a membrán. Pro konkrétní metody i membrány tabulky obsahují počty záznamů v MolMeDB.

Obrázek 13 ukazuje rozložení dat v MolMeDB pro jednotlivé skupiny membrán a metod pomocí teplotních map.



Obrázek 13: Teplotní mapy dat v databázi MolMeDB (A - data ze simulačních metod, B - data z výpočetních metod, C - data z experimentálních metod) osa x : BE – brain experimental, BM – brain model, CE – cell experimental, EE – eye experimental, IE – intestine experimental, IM – intestine model, OE – oral experimental, SE – skin experimental, SM – skin model osa y: BLM – black lipid method, DC – diffusion cell methods, FM – fluorescence spectroscopy methods, HTCG simulations - high-throughput coarse-grained simulations MD – molecular dynamics, MM – mechanistic model, PS – perfusion system, SS - steady-state methods

Z obrázku 13 je vidět pestrost experimentálních metod, kterých je největší množství a také z této skupiny pochází data pro největší množství membrán. V případě simulačních dat je vidět masivní převaha jedné metody, která má ve výsledku dat nejvíc ze všech (HTCG, tzv. High-throughput Coarse Grain simulations – tedy vysokopropustné zhrubené simulace pomocí silového pole MARTINI).

7.2 Analýza dat interakce molekuly s transportním proteinem

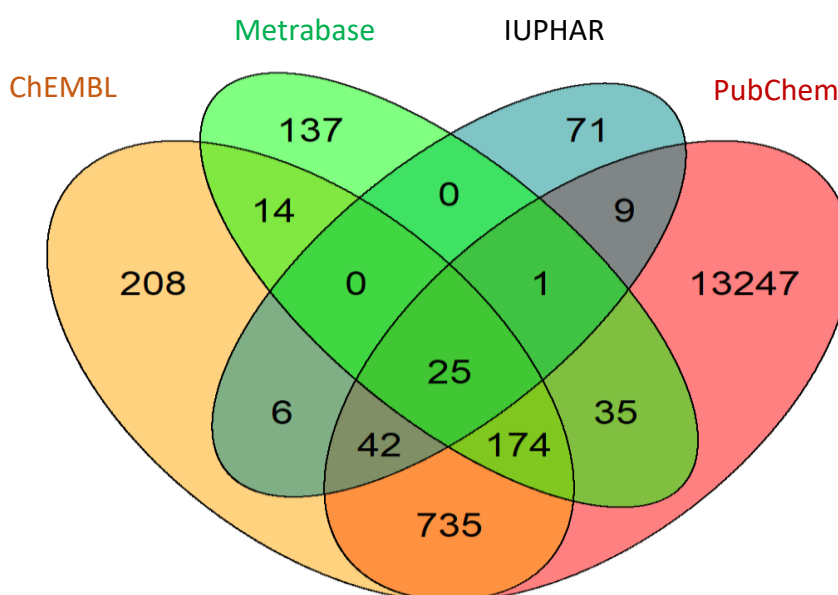
Jak již bylo naznačeno tato data pocházejí ze čtyř různých databází a byla získána pomocí KNIME workflow. Celkově tato data tvoří 3,2 % interakcí, které se v MolMeDB nacházejí. Data jsem analyzovala z hlediska jejich původu, interakce transportér – molekula a jednotlivých transportérů.

7.2.1 Analýza databází původu dat

První hledisko, ze kterého se lze na tato data dívat je jejich původ. Všechny řádky výstupu z workflow uvádějí dvě informace o původu. První je sloupec „Pubmed ID“, které identifikuje konkrétní vědecký článek a MolMeDB jej ukládá jako tzv. primární referenci a druhá je sloupec „Source“, kde je uvedena databáze, ze které workflow danou informaci získalo a MolMeDB tuto informaci ukládá do tzv. sekundární reference.

Z hlediska zdrojů byla data získána z celkem 1358 různých vědeckých textů a ze čtyř databází. Nejvíce dat pochází z databáze PubChem (21 588), dále pak následují databáze ChEMBL (3 706), Metrabase (721) a IUPHAR (210).

Z pohledu zdrojů dat jsem se rovněž dívala na to jaké je zastoupení jednotlivých SMILES v jednotlivých databázích. Jinak řečeno sledovala jsem překryvy dat za databází podle SMILES jednotlivých molekul. Výsledek této analýzy ukazuje obrázek 14.



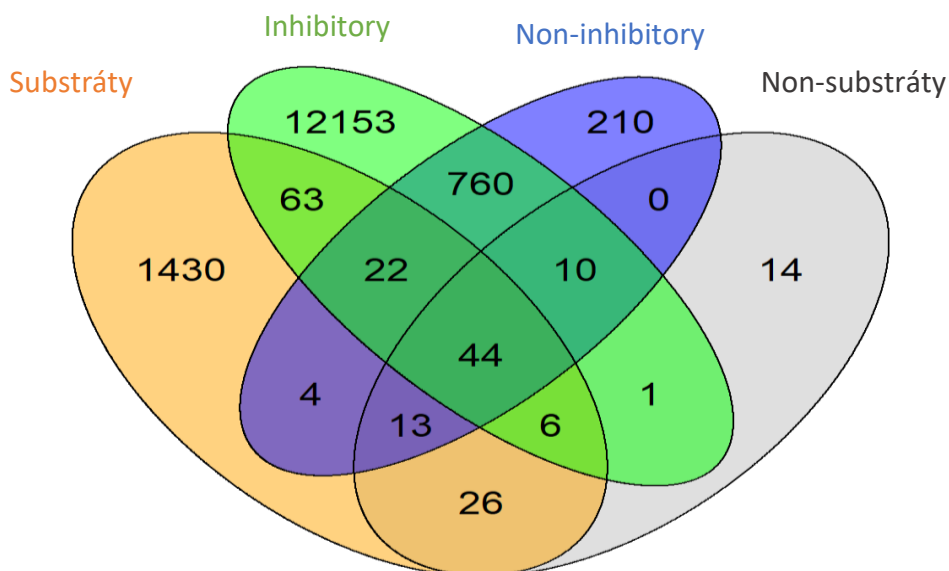
Obrázek 14: Překryvy dat podle databáze původu (vytvořeno pomocí programu RStudio)

Z Vennových diagramů na obrázku 14 jasně vidíme, že pouze ke 25 SMILES byla získána informace o transportu ze všech čtyř vytěžovaných databází. Také je vidět největší překryv mezi databázemi ChEMBL a PubChem. Tento překryv tvoří 735 SMILES. Většina interakcí se nicméně mezi databázemi nepřekrývá a největším zdrojem dat se ukázala databáze PubChem. To se díky její celkové velikosti nicméně dalo očekávat.

7.2.2 Analýza interakcí membránový transportní protein – molekula

Informaci o interakci mezi transportérem a molekulou máme v datech obsaženou dvěma způsoby. Prvním je sloupec, který MolMeDB nazývá „Type“ zde je podle pravidel, která jsem popsala výše v sekci metody uveden vždy jedna ze čtyř hodnot (INH, NINH, SUB, NSUB). Tato informace slouží jako rychlý popis, v jakém vztahu je daná molekula ke konkrétnímu proteinu. Druhým způsobem, jak je popsána tato interakce jsou číselné hodnoty příslušných veličin (K_M , K_i , IC_{50} a EC_{50}), které jsou vždy v databázi MolMeDB uvedeny.

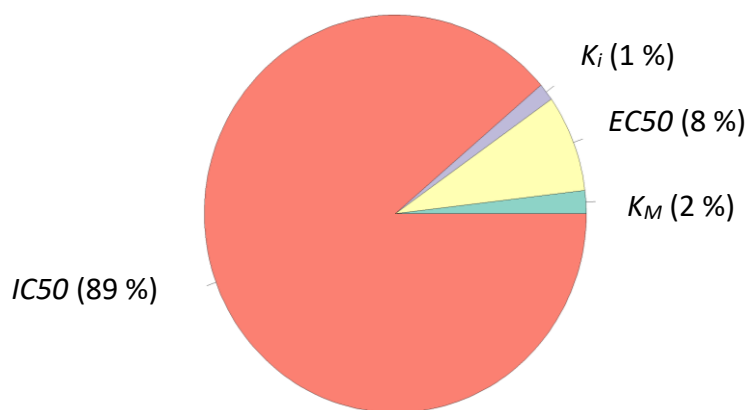
Drtivá většina látek v MolMeDB nahrány jsou inhibitory (13 059 SMILES). Ve výrazně menší míře jsou zastoupeny: non-inhibitory (1 063 SMILES), substráty (1 608 SMILES) a non-substráty (114 SMILES). Podobně jako tomu bylo při analýze dat z databází, i v tomto případě mě zajímaly vzájemné překryvy těchto čtyř typů. Výsledek opět zobrazuje Vennův diagram na obrázku 15.



Obrázek 15: Překryv dat podle jednotlivých typů (vytvořeno pomocí programu RStudio)

Obrázek 15 ukazuje, že 44 molekul má alespoň jeden záznam v každém ze čtyř typů. Velmi zajímavé jsou také počty molekul, pro které jsou v datech uvedeny pouze jako jeden z typů interakce. Celkem 1 430 ze 1 608 substrátů je označeno jen jako substráty, 12 153 ze 13 059 inhibitorů je označeno jen jako inhibitory. Pro oba typy se jedná o výrazně nadpoloviční podíl.

Jak již bylo zmíněno výše druhým typem informace, který o interakcích transportér – molekula máme jsou numerické hodnoty K_M , K_i , IC_{50} a EC_{50} . Zastoupení jednotlivých veličin ukazuje obrázek 16.



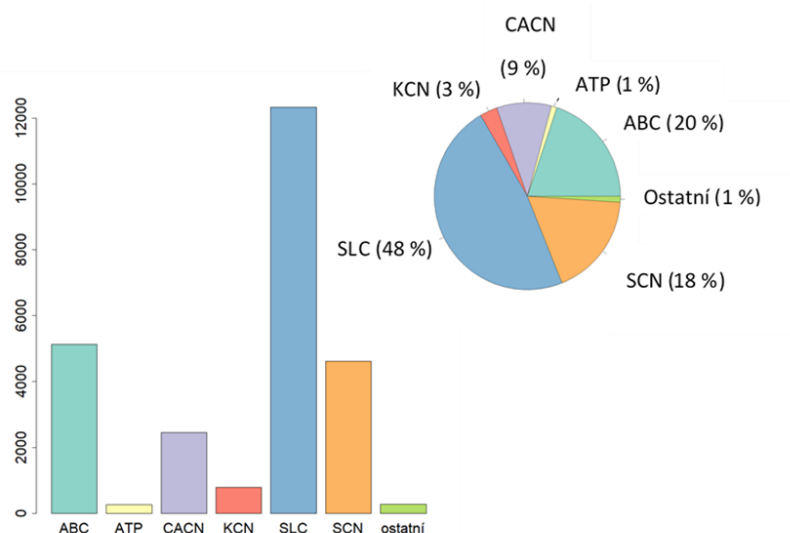
Obrázek 16: Koláčový diagram zastoupení veličin K_M , K_i , IC_{50} , EC_{50} v datech (vytvořeno pomocí programu RStudio)

Z obrázku 16 je jasně vidět, že nejvíce v datech dominuje vyjádření interakce pomocí IC_{50} , které tvoří drtivou většinu. Spolu s K_i , tvoří 90 %, což je v souladu se skutečností, že drtivá většina látek v datech jsou klasifikovány jako inhibitory, jelikož se v obou případech jedná o vyjádření inhibice.

7.2.3 Analýza transportních proteinů

Membránových transportních proteinů existuje ohromné množství, z tohoto důvodu jsem se rozhodla věnovat pozornost jen některým vybraným skupinám. Jedná se o skupiny SLC (SLC transportéry), ABC (ABC transportéry), ATP (ATPázové pumpy), CACN (napěťově řízené vápenaté kanály) KCN (draselné kanály), SCN (sodíkové kanály). Všechny transportéry, které se nedají zařadit ani do jedné z těchto skupin jsou ve skupině „ostatní“.

V MolMeDB je momentálně 184 transportérů. Zastoupení jednotlivých skupin ukazuje obrázek 17 v datech.



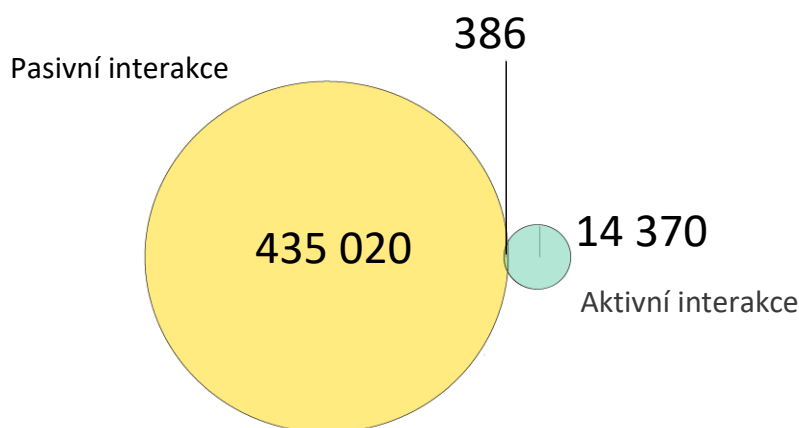
Obrázek 17: Zastoupení jednotlivých skupin transportérů v MolMeDB (vytvořeno pomocí programu RStudio)

Z obrázku 17 jasně vyplývá, že největší zastoupení v MolMeDB z hlediska transportérů mají přenašeče ze super rodiny SLC, na něž připadá 12 330 záznamů v této databázi. Tyto přenašeče tvoří téměř polovinu všech těchto dat. Druhé místo z hlediska zastoupení patří členům super rodiny ABC, jejíž zastoupení v datech činí 20 % a celkově se jedná o 5126 záznamů. Dále 18 % dat (4 616 záznamů) se vztahuje k sodným kanálům a 9 % (2 449 záznamů) dat k vápenatým kanálům a 3 % (792 záznamů) k draselným kanálům. 1% zastoupení shodně mají ATPázy (256 záznamů), které jsou poslední skupinou, která je v rámci dat ještě rozlišována a ostatní transportéry. Na ostatní připadá (267 záznamů) v MolMeDB. Další dělení v rámci těchto šesti definovaných skupin transportérů, včetně jejich kvantitativního zastoupení v datasetu a popisu lze najít v příloze práce v sekci Analýza konkrétních rodin transportérů.

7.3 Analýza překryvu dat

V této kapitole se zaměřím na překryv dat mezi původními daty interakcí s membránami a nově doplňovanými daty interakcí s membránovými transportními proteiny. Z důvodu zjednodušení a zpřehlednění následující části práce bude interakce molekuly s membránou, které se neúčastní protein nazývána „pasivní“, zatímco interakce molekuly s transportním proteinem bude označována jako „aktivní“. Jejich vzájemný překryv bude poté označován jen jako „překryv“.

Tento „překryv“ tvoří 386 molekul, pro které máme „aktivní interakci“ s 99 transportními proteiny a „pasivní interakci“ s 28 membránami, určenou 27 metodami. Tento „překryv“ graficky znázorňuje obrázek 18.



Obrázek 18: Překryv dat o molekulách se známými pasivními a aktivními interakcemi v rámci databáze MolMeDB (vytvořeno pomocí programu RStudio)

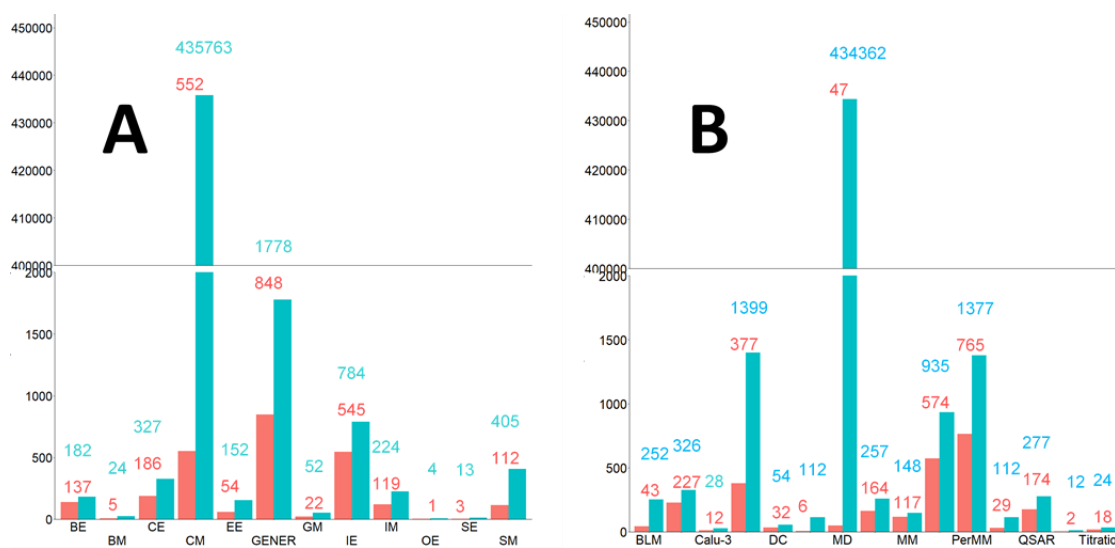
Jelikož se jedná o Vennův diagram do čísel, která vidíme na obrázku 18 se promítá každý SMILES pouze jednou. Lze si tak udělat představu o tom, kolik sloučenin konkrétní oblast zahrnuje.

Vzhledem k tomu, že v případě tohoto překryvu hovoříme o relativně malém množství molekul. Tak jsem tuto oblast zkoumala mnohem detailněji než obě oblasti předchozí.

Nejprve jsem pozornost metodám a membránám, které se v tomto překryvu vyskytují, poté transportérům a interakci transportér – molekula, nakonec jsem se zaměřila na samotné chemické struktury, které se v oblasti překryvu nacházejí.

7.3.1 Analýza membrán a metod v oblasti překryvu dat

V této části budou ukázány metody a membrány, které jsou v překryvu dat zastoupeny. Tyto informace graficky ilustruje obrázek 19, kde oblast A ukazuje membrány a oblast B ukazuje metody, které se vyskytují v rámci překryvu, včetně kvantitativního zastoupení jednotlivých skupin^y. Abychom popsali kvantitativní překryv interakcí, tak jsme na rozdíl od předchozího obrázku 19 tak se zde jednotlivé SMILES mohou počítávat vícekrát, pokud je pro ně známo více interakcí.



Obrázek 19: Graf zastoupení jednotlivých membrán a metod v oblasti překryvu dat (růžová barva) a jejich celkové zastoupení v MolMeDB (modrá barva),

A – zastoupení jednotlivých membrán (BE – brain experimental, BM – brain model, CE – cell experimental, EE – eye experimental, GE – GENE, IE – intestine experimental, IM – intestine model, OE – oral experimental, SE – skin experimental, SM – skin model),

B- zastoupení jednotlivých metod (BLM – black lipid method, DC – diffusion cell methods, FM – fluorescence spectroscopy methods, MD – molecular dynamics, MM – mechanistic model, PS – perfusion systems, SS - steady-state methods)

(vytvořeno pomocí programu RStudio)

Z obrázku 19A vidíme jasně, že nejvíce je v datech zastoupena generická membrána (GE). Téměř stejným dílem jsou zastoupeny modelové membrány buněk (CM) a experimentální membrány střeva (IE). Pokud ponecháme stranou generickou membránu, tak nejhojnějšími typy membrán v rámci překryvu jsou membrány buněčné modelové (CM) a buněčné experimentální (CE) dohromady čítají 738 záznamů. Po membránách buněčných následují membrány střeva (IE a IM), které dohromady mají 664 záznamů. Za zmínku stojí, že překryvy jsou až na CM dataset vesměs nadpoloviční, což otevírá možnosti srovnávání vlastností mezi aktivními a pasivními interakcemi na těchto membránách.

^y Seskupení metod a membrán bylo provedeno podle stejných pravidel jako v případě kapitoly 7.1 Analýza dat interakce molekul s membránou.

Na obrázku 19B vidíme jednotlivé metody z překryvu dat. Nejvýraznější podíl zde mají metody výpočetní, které dohromady činní 1 433 záznamů, přičemž absolutně největší podíl na tomto počtu má metoda PerMM,⁵¹ na kterou připadá 765 záznamů. Hned za metodami výpočetními následují metody experimentální, na něž připadá 1 107 záznamů. Z experimentální metod dominuje PAMPA s 574 záznamy. Nejmenší počet dat připadá na simulační metody, na které připadá pouhých 47 záznamů.

Zde vidíme obrovský rozdíl mezi daty týkající se pasivní interakce v rámci celé MolMeDB a tohoto velmi malého výseku dat. Zatímco MolMeDB obsahuje velké množství dat ze simulací (především z HTCG simulací), v tomto „překryvu“ je výsledků simulací jen malé množství. Naopak z pohledu ostatních metod, kromě nejhodněji zastoupených výpočetních metod, jsou zde také významně zastoupeny metody experimentální.

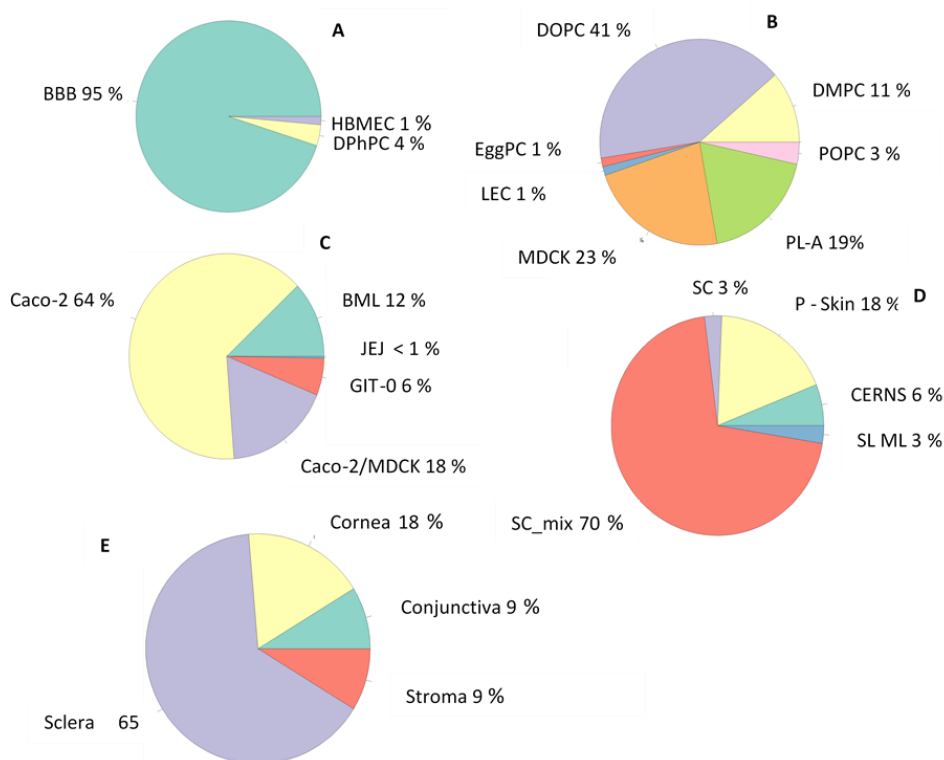
Tabulka 9 ukazuje množství záznamů pro kombinace metoda – membrána pro oblast „překryvu“. Zde je detailněji vidět to, co naznačoval již trochu obrázek 19. Největší množství aktivních dat máme pro kombinaci PerMM – GE.

Tabulka 9: Tabulka kombinací metod a membrán v překryvu dat (žlutá – membrány střeva, modrá – membrány buněk, oranžová – membrány kůže, zelená – membrány mozku, šedá – membrány oka)

Kombinace metoda – membrána	Počet záznamů	Kombinace metoda – membrána	Počet záznamů
PerMM – GE	765	PAMPA – IE	25
Cosmomic – CM	285	PAMPA – GM	22
CACO – IE	227	PAMPA – SM	21
PAMPA – CM	208	Titration – CM	12
QSAR – IE	174	Calu-3 - CE	12
MDCK – CE	164	FM – CM	6
PAMPA – BE	135	Titration – CE	6
PAMPA – IM	119	MD – SM	4
MM – IE	117	PAMPA – CE	4
Cosmomic – SM	87	DC – SE	3
BLM – GE	43	DC – IE	2
MD – CM	41	SS – EE	2
PAMPA – GE	40	MD – BE	2
PS – EE	29	DC – OE	1
DC – EE	26		

Z pohledu jen skupin membrán, pokud nebudeme počítat generickou membránu jsou nejvíce zastoupeny membrány buněk (modelové i experimentální). Pro membrány je v datech dohromady 783 záznamů. Následují je modelové a experimentální membrány střeva (664 záznamů). Zbylými hojněji zastoupenými skupinami jsou modelové a experimentální membrány mozku (137 záznamů) a kůže (115 záznamů).

Zastoupení konkrétních membrán v ve skupinách podle „původu“ ukazuje obrázek 20. Generické membrány a membrány orální obrázek 20 neukazuje, jelikož v obou případech je v datech zastoupen jen jedna konkrétní membrána. V případě orálních membrán se jedná o sublinguální sliznici (Sublingual mucous membrane) a v případě generické modelové membrány se jedná o n-hexadekan (n-Hex).



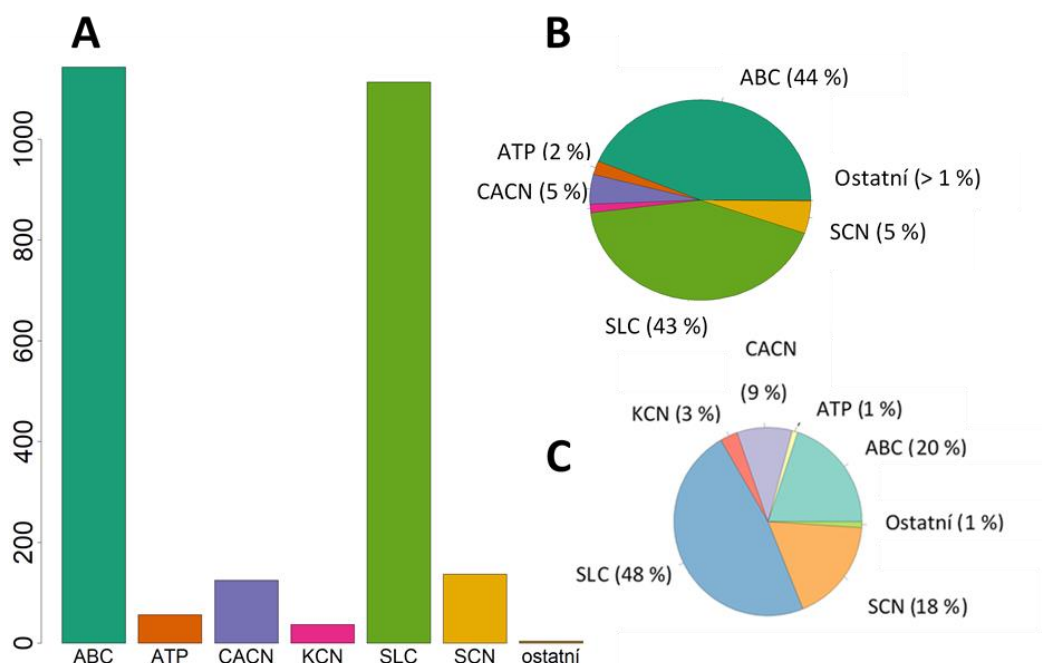
Obrázek 20: Koláčové diagramy zastoupení konkrétních membrán ve skupinách dle "původu" A – membrány mozku, B – buněčné membrány, C – membrány střeva, D – membrány kůže, E – membrány oka (vytvořeno pomocí programu RStudio)

Kvůli zjednodušení práce s daty a zřehlednění výsledků byly v rámci této kapitoly jednotlivé metody i membrány klastrovány do větších celků, ale uvědomíme-li si, podstatu jednotlivých konkrétních typů membrán, dospějeme k závěru, že jen některé z nich jsou schopny téměř „dokonalé“ pasivní interakce s molekulou např. membrána BLM, n-Hex, či DMPC. U ostatních lze očekávat možnou určitou míru aktivního transportu, jelikož víme, že tyto membrány obsahují také transportéry. Určitý podíl aktivního transportu lze očekávat např. membrány HBMEC⁵², MDCK¹⁰, JEJ⁵³, Caco-2,¹⁰ či cornea.⁵⁴

7.3.2 Analýza transportních proteinů v oblasti překryvu dat

V této části bude pozornost věnována transportním proteinům, které se v oblasti překryvu vyskytují.

Obrázek 21 ukazuje zastoupení jednotlivých rodin transportérů v překryvu a je tedy obdobou obrázku 17.



Obrázek 21: A - sloupcový graf zastoupení jednotlivých rodin transportérů v překryvu dat, B – koláčový diagram procentuálního zastoupení jednotlivých rodin transportérů v překryvu, C – koláčový diagram procentuálního zastoupení pro všechny aktivní interakce (vytvořeno pomocí programu RStudio)

Na obrázku 21 je vidět, že 87 % všech transportérů v překryvu jsou členové superrodin SLC a ABC. V případě ABC transportérů máme v této oblasti dat 1 143 záznamů, v případě SLC transportérů máme 1 113 záznamů. S velkým odstupem následují iontové kanály (CACN, KCN, SCN), které dohromady zaujímají 11 % všech dat v překryvu. Pro ATPázy je v překryvu 57 a zaujímají 2 % všech dat. Méně než 1 % připadá na ostatní transportéry, které jsou v překryvu zastoupeny pouze čtyřmi záznamy.

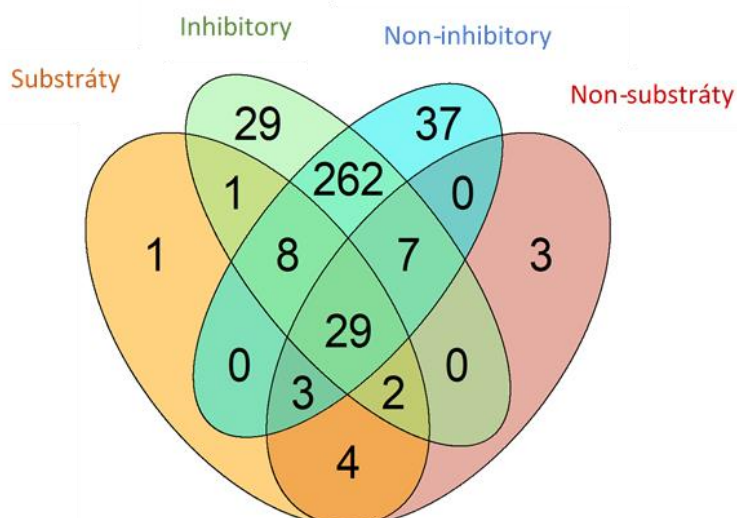
Porovnáním obrázků 17 a 21 můžeme zjistit, jak moc se liší zastoupení jednotlivých rodin transportérů v oblasti dat pro celou aktivní interakci a překryvu. V obou případech jsem použila stejná pravidla pro seskupení dat a stejný typ grafů pro jejich následnou vizualizaci, aby bylo možno výsledky mezi sebou kvantitativně porovnávat. Tabulka 10 ukazuje porovnání oblasti celkové počty záznamů pro oblast „aktivní interakce“ a „překryv“.

Tabulka 10: Tabulka ukazující počet záznamů pro jednotlivé transportérové rodiny v celé oblasti dat „aktivní interakce“, v „překryvu“ a procentuálně vyjádřený poměr

Transportérová rodina	Data „Aktivní interkace“	Data „překryv“	Poměr /%/
ABC	5 126	1 143	22,3
ATP	256	57	22,3
CACN	2 449	125	5,1
KCN	792	37	4,7
SCN	4 616	137	3,0
SLC	12 330	1 113	9,0
ostatní	267	4	1,5

Poslední sloupec tabulky vyjadřuje v procentech, na jakou hodnotu v procentech pokleslo množství záznamů v „překryvu“ oproti všem „aktivním interakcím“ pro danou skupinu. Vidíme, že z hlediska množství záznamů nejméně poklesly ABC transportéry a ATPásy, zatím co všechny ostatní skupiny jsou zastoupeny ani ne 10 % svého původního množství. Tento trend je patrný i z koláčových diagramů B a C na obrázku 21. Kde na obrázku 21 B je jasně vidět, že pokud budeme uvažovat pouze aktivní interakce v oblasti „překryvu“ překryvu bude velká část interakcí spojena hlavně s ABC a SLC transportéry. Všechny ostatní transportéry budou podstatně méně zastoupeny. Zatímco na obrázku 21 C vidíme, že co do množství ABC transportéry jsou zastoupené téměř stejně jako sodné kanály a od nejhojněji zastoupených SLC transportérů jsou co do počtu interakcí daleko.

Podobně jako v případě obrázku 15 jsem i v tomto případě sledovala překryvy mezi jednotlivými typy, což ukazuje obrázek 22. Na obrázku 22 vidíme významný překryv mezi inhibitory a non-inhibitory, který činí 262 SMILES. Počet SMILES, které se v datech vyskytují alespoň jednou jako každý ze čtyř typů činí nyní 29 SMILES. Na obrázku 16 to bylo 44 SMILES, v této oblasti nastal relativně malý pokles.



Obrázek 22: Překryv dat podle jednotlivých typů v oblasti tzv. „překryvu“ (vytvořeno pomocí programu RStudio)

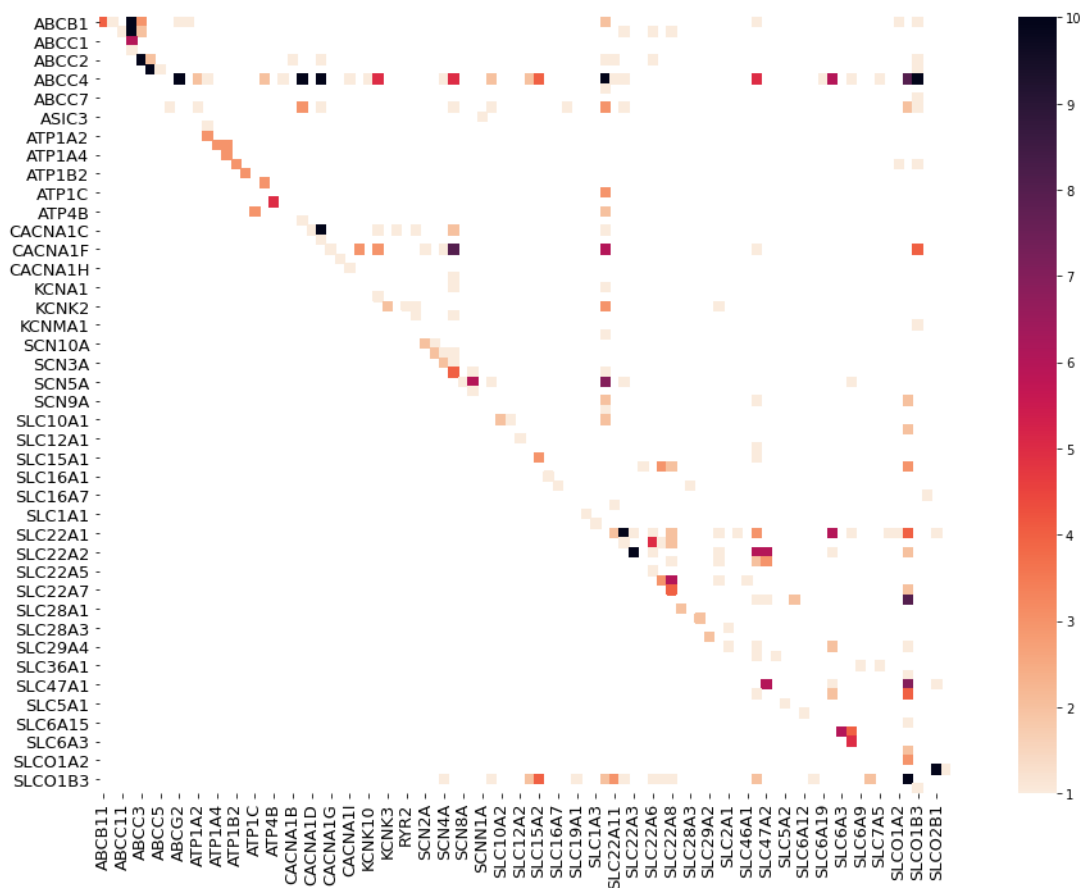
Molekuly, které se objevovaly v překryvu všech skupin z obrázku 22 jsou notoricky známé struktury. Do toho to překryvu patří např. antipsychotika (risperidone, chlorpromazine, clozapine, quetiapine a haloperidol), antivirotika (zidovudine, ganciclovir a acyclovir), léčiva ovlivňující činnost srdce (amiodarone, ouabain a digoxin). Celkový soupis těchto molekul ukazuje tabulka 11.

Tabulka 11: Názvy a použití molekul z oblasti překryvu všech typů z obrázku 23

Skupina	Zástupci
Antipsychotika	Chlorpromazine, Clozapine, Haloperidol, Risperidone, Quetiapine
Antineoplastika	Etoposide, Methotrexate, Tamoxifen
Léčiva na vysoký krevní tlak	Prazosin, Verapamil
Látky ovlivňující produkci kyseliny v žaludku	Famotidine, Ranitidin
Látky ovlivňující činnost srdce	Amiodarone, Digoxin, Ouabain
Antibiotika	Cefadroxil, Cefazolin, Tetracycline
Antivirotika	Acyclovir, Ganciclovir, Zidovudine
ostatní	Aminohippuric acid, Cimetidine, Loperamide, Mesalamine, Ochratoxin A, Taurocholic acid, Vinblastine

Kromě samotného množství dat a jejich rozložení do jednotlivých skupin, mě také zajímalo, jak jsou transportéry, které se nacházejí v překryvu propojeny. Propojení transportérů v rámci překryvu ilustruje obrázek 23. Obrázek 23 ukazuje kombinace dvojic transportérů. Dvojice jsem vytvořila seskupením dat podle SMILES a následně jsem pro seskupené transportéry pro každou takto identifikovanou molekulu určila kombinace transportérů. Čím je barva příslušného čtverce teplotní mapy tmavší, tím je daná dvojice transportérů v překryvu hojněji zastoupena.

Absolutně nejhojněji rozšířenou dvojicí transportérů byla dvojice ABCC3 - ABCC4, která je přítomná u 233 SMILES z celkových 386, druhou nejběžnější dvojicí byla dvojice ABCC2 - ABCC3. Tato dvojice byla přítomna celkem u 221 SMILES. Všechny ostatní dvojice následovali až s velkým odstupem. Následující další dvojice ABCB1 – ABCC2 se je přítomná u jen 35 SMILES.



Obrázek 23: Teplotní mapa dvojic transportérů

Z obrázku 23 je jasně vidět, že ABC transportéry, které jsou v překryvu přítomné se vyskytují často nejen v kombinaci se sebou, ale také i s velkým množstvím transportérů z jiných skupin. Tako skutečnost může být dána skladbou ABC transportérů, která v překryvu je. Šest z devíti zástupců, které obrázek 23 ukazuje, jsou ze skupiny MRP, tedy „multi-drug resistance protein“, které jsou známy širokou specificitou (viz. tabulka 6). Zbylí dva zástupci jsou ABCB11, což je exportér žlučových kyselin a ABCG2, který má velmi široké spektrum látek, které dokáže transportovat.

Opak je vidět v případě ATPáz, které tvoří jen malou část interakcí v překryvu. Z obrázku 23 je vidět, že většinou se vyskytují ve dvojici s jinou ATPázou. V několika případech tvoří dvojici s některým z členů SLC, SLCO nebo ABC transportérů. V případě iontových kanálů (CACN, KCN, SCN) situace vypadá podobně jako v případě ATPáz.

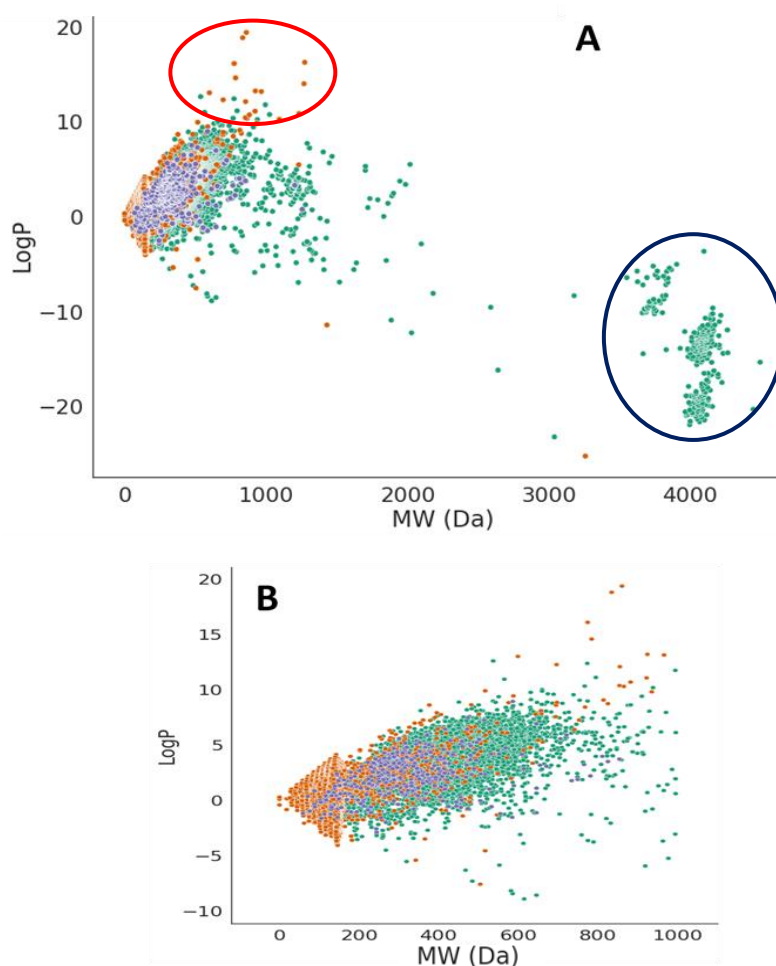
V případě SLC a SLCO transportérů, vidíme, že jsou často ve dvojicích spolu či z výše zmíněnými ABC transportéry, ale také s iontovými kanály.

8. Fyzikálně chemické vlastnosti molekul v databázi MolMeDB

V této kapitole se zaměřím na fyzikálně chemické vlastnosti molekul, které jsou dostupné v databázi MolMeDB. Pozornost bude věnována rozdělovacímu koeficientu oktanol-voda ($\text{Log}P$), molekulové hmotnosti (MW) a permeabilitnímu koeficientu (LogPerm). $\text{Log}P$ i molekulová hmotnost, se kterými jsem pracovala, byly získány výpočetními metodami. Obě tyto veličiny jsou dostupné pro drtivou většinu molekul, které jsou v MolMeDB. Hodnoty LogPerm , pochází z různých typů měření, výpočtů nebo simulací a je v databázi MolMeDB dostupný pro 435 405 molekul. Pro účely těchto analýz jsou data z MolMeDB rozdělena do tří skupin na „pasivní“, „aktivní“ a „překryv“, podobně jako tomu bylo v předchozím textu.

8.1 Analýza molekulových hmotností a hodnot $\text{log}P$ v databázi MolMeDB

V této kapitole budou diskutovány hodnoty MW a $\text{log}P$ látek, které se v MolMeDB nacházejí. Nejdříve bude věnována pozornost rozdílu mezi skupinami „aktivní“, „pasivní“ a překryv. Poté také bude věnována pozornost rozdílům z hlediska jednotlivých transportérů a typů.

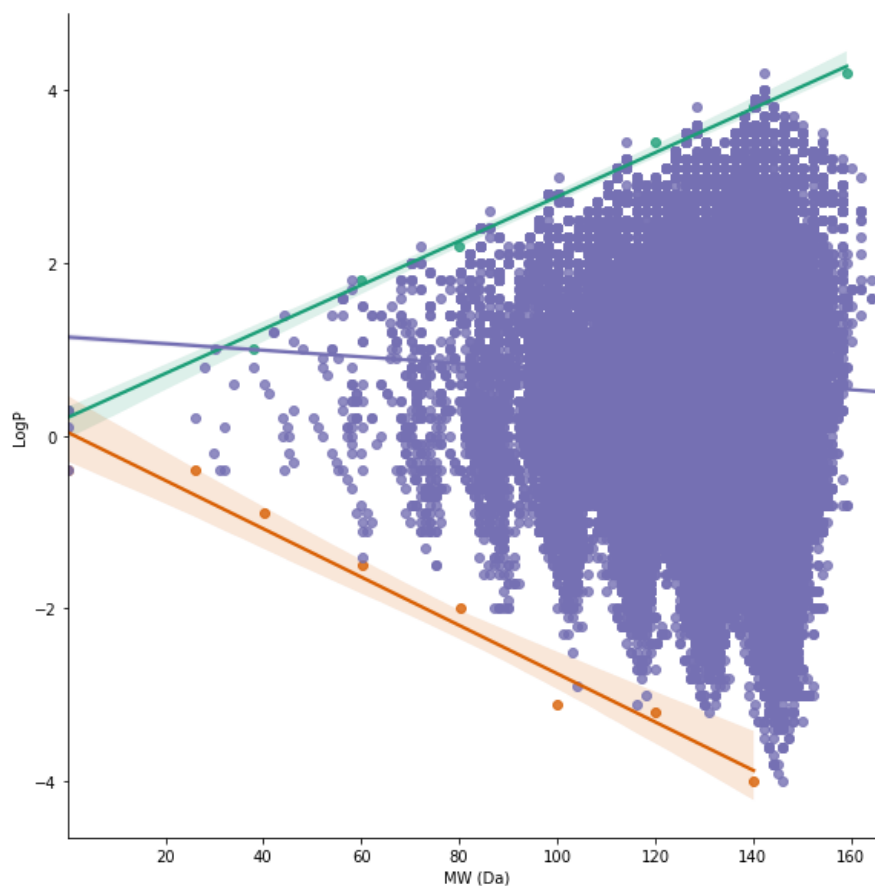


Obrázek 24: Bodový graf znázorňující molekulovou hmotnost (MW) a rozdělovací koeficient oktanol-voda ($\text{Log}P$) pro molekuly obsažené v databázi MolMeDB (zelená barva – molekuly s pouze aktivními interakcemi, oranžová barva – molekuly s pouze pasivními interakcemi, fialová barva – molekuly s alespoň jednou pasivní a aktivní interakcí), A – graf znázorňující všechna data, B – graf znázorňující pouze data s MW menší než 1000 Da

Z obrázku 24 je jasně patrné, že naprostá většina molekul v databázi má svoji molekulovou hmotnost do 1 000 Da, přesněji se jedná o 448 015 molekul z celkového počtu 449 766 molekul, se kterými jsem v této práci pracovala. Při detailnějším pohledu do dat je vidět, že většina molekul, která připadá do intervalu molekulových hmotností 0 – 1 000 Da, má svoji molekulovou hmotnost v interval 100 – 200 Da. Můžeme tedy říct, že většina molekul uložených v databázi MolMeDB vyhovuje, z hlediska své molekulové hmotnosti, definici malé molekuly. Byť většinu molekul v MolMeDB můžeme považovat za malé molekulu, z obrázku je jasné, že ne všechny budou malé molekuly. Molekulovou hmotnost větší než 1000 Da má 580 molekul v MolMeDB.

Na obrázku 24 je také vidět, jak se s rostoucí molekulovou hmotností mění zastoupení interakcí. V části grafu s nižšími molekulovými hmotnostmi jasně dominují molekuly s pasivními interakcemi (oranžová barva), následně přibývají interakce z oblastí překryvu (fialová barva) až nakonec vidíme že molekulových hmotnostech nad 1000 dominují jasně aktivní interakce (zelená barva). Tato skutečnost je dána tím, že pasivní difúze přes membránu jsou schopny jen malé molekuly a obecně lze říci, že pasivní interakce jsou více spjaty spíše s malými molekulami.

Pokud se podíváme na obrázek 24A z pohledu $\log P$ molekul, zjistíme, že velká část z nich se pohybuje zhruba v rozmezí hodnot -5 a 10 , což je patrné také z obrázku 24B. Také na obrázku 24 vidíme, že $\log P$ molekul s pouze pasivní interakcí a molekul z překryvu se v tomto intervalu v drtivé většině drží, zatímco molekuly s aktivní interakcí mají rozsah hodnot $\log P$ mnohem větší. Částečně je to způsobeno tím, že 4344257 molekul s pasivními interakcemi pochází ze studie Menichetti et al⁵⁵ a tyto molekuly byly pouze systematicky vygenerovány tak, aby popisovaly možné molekuly do velikosti cca 8 těžkých atomů (alias atomů C-O-N). Na těchto molekulách lze tak vidět rozsah $\log P$, který s růstem velikosti molekul souvisí a vytváří se tak trychtýř možných hodnot $\log P$ – maxima a minima rostou zhruba dle rovnic (4 a 5), které popisují přímky na obrázku 25.

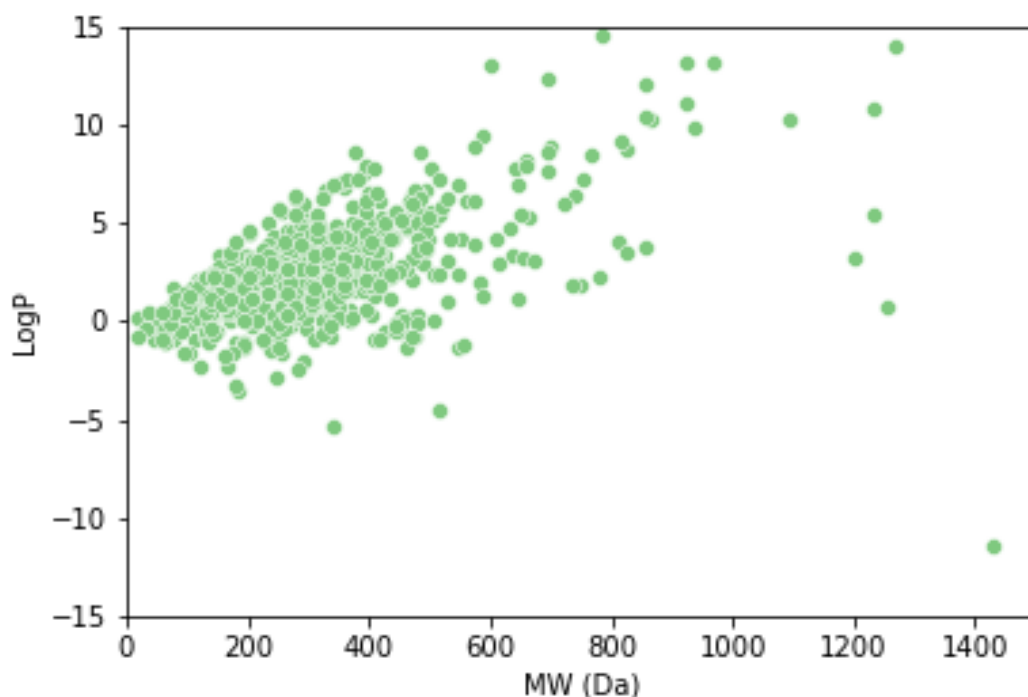


Obrázek 25: Bodový graf možných hodnot $\log P$ ze studie s lineární regresí (oranžová – přímka minim, zelená – přímka maxim, fialová – průměr ze všech bodů)

$$\log P_{Max} = 0,026 \cdot MW + 0,21 \quad (3)$$

$$\log P_{min} = -0,028 \cdot MW + 0,438 \quad (4)$$

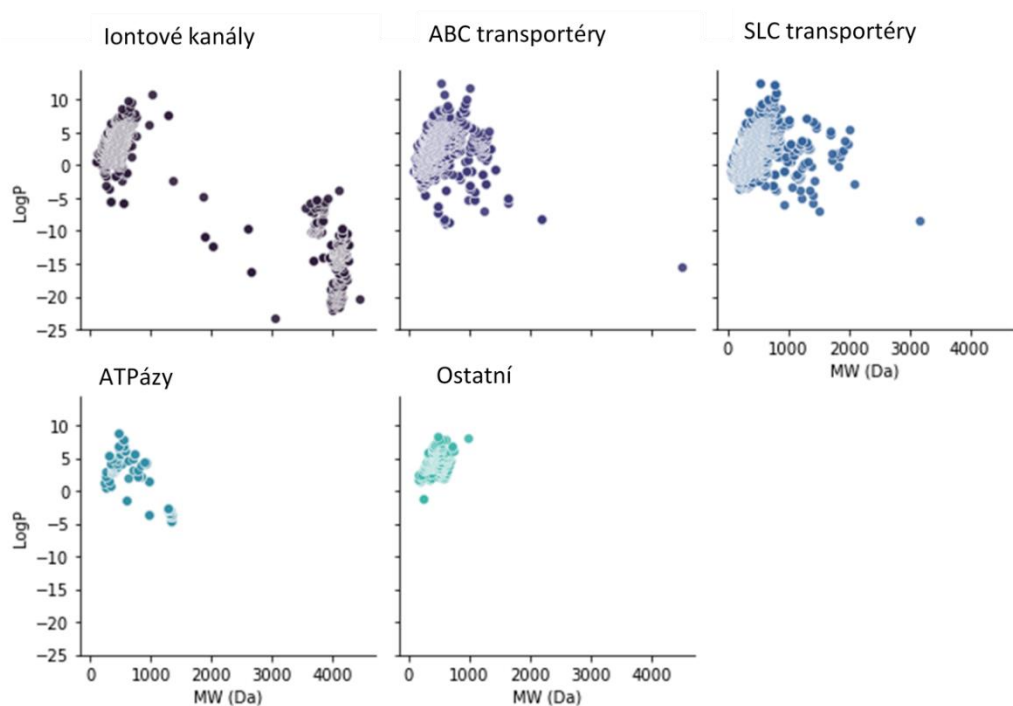
Nejvyšší hodnota $\log P$ v tomto datasetu činila hodnotu 4,15 zatím co nejnižší hodnota $\log P$ byla $-4,05$. Ale kromě těchto minimálních a maximálních hran se ukazuje, že průměrné $\log P$ se vůči velikosti příliš nemění. Obrázek 26 ukazující látky s dostupnými experimentálními daty, potvrzuje trend „nálevky“ pro maximální a minimální hodnoty $\log P$ u malých molekulových hmotností, ale u větších MW se rozvor „nálevky“ hlavně u polární oblasti výrazně zužuje.



Obrázek 26: Hodnoty $\log P$ a MW pro experimentální data z databáze MolMeDB

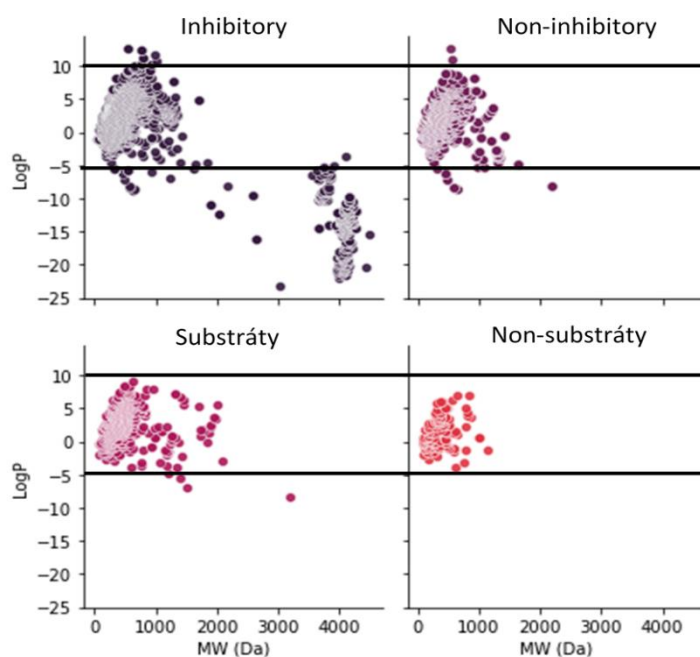
Když se znovu vrátíme k obrázku 24A, tak je v něm v červeném kruhu vyznačena oblast, kde se vyskytuje 19 molekul, pro které jsou dostupné pouze pasivní interakce s molekulovou hmotností kolem 1 000 Da a $\log P$ větším než 10, jsou to tedy již velmi lipofilní látky. V tomto případě se jedná o lipidické fluorescenční sondy, pomocí kterých se experimentálně sleduje chování membrán. Dále je pak v modrém kruhu vyznačen klastř 298 molekul velkých molekul s aktivními interakcemi, které mají molekulovou hmotnost kolem 4 000 Da případně i vyšší, avšak jejich $\log P$ je menší než nula a jsou to tedy velmi polární molekuly. V tomto případě se jedná o peptidy, v drtivé většině tzv. peptidické toxiny, schopné interakce s iontovými kanály (viz obrázky 27 a 28).

Obrázek 27 ukazuje pro molekuly s aktivními interakcemi jejich transportéry. Vzhledem k velkému množství zobrazovaných dat jsem všechny iontové kanály seskupila do jedné skupiny, ostatní skupiny jsem vytvářela podle stejných pravidel jako výše v textu. Na obrázku je jasně vidět, že molekuly interagující s ABC a SLC transportéry případně s ATPázami mají většinou $\log P$ větší než 0 a jsou tedy lipofilní případně některé z nich mají $\log P$ menší než 0 a jsou tedy hydrofilní. Zatímco v případě iontových kanálů, vidíme významné zastoupení molekul, které mají záporné hodnoty $\log P$ stejně tak jako molekul s velkou molekulovou hmotností.



Obrázek 27: Hodnoty $\log P$ a MW pro aktivní interace rozdělené dle transportéru

Dalším pohledem, jak se můžeme dívat na $\log P$ a MW u dat s dostupnou aktivní interakcí je přes jejich klasifikaci na substráty, non-substráty, inhibitory a non-inhibitory, jak ukazuje obrázek 28.



Obrázek 28: Hodnoty $\log P$ a MW pro aktivní interace rozdělené dle typu interakce

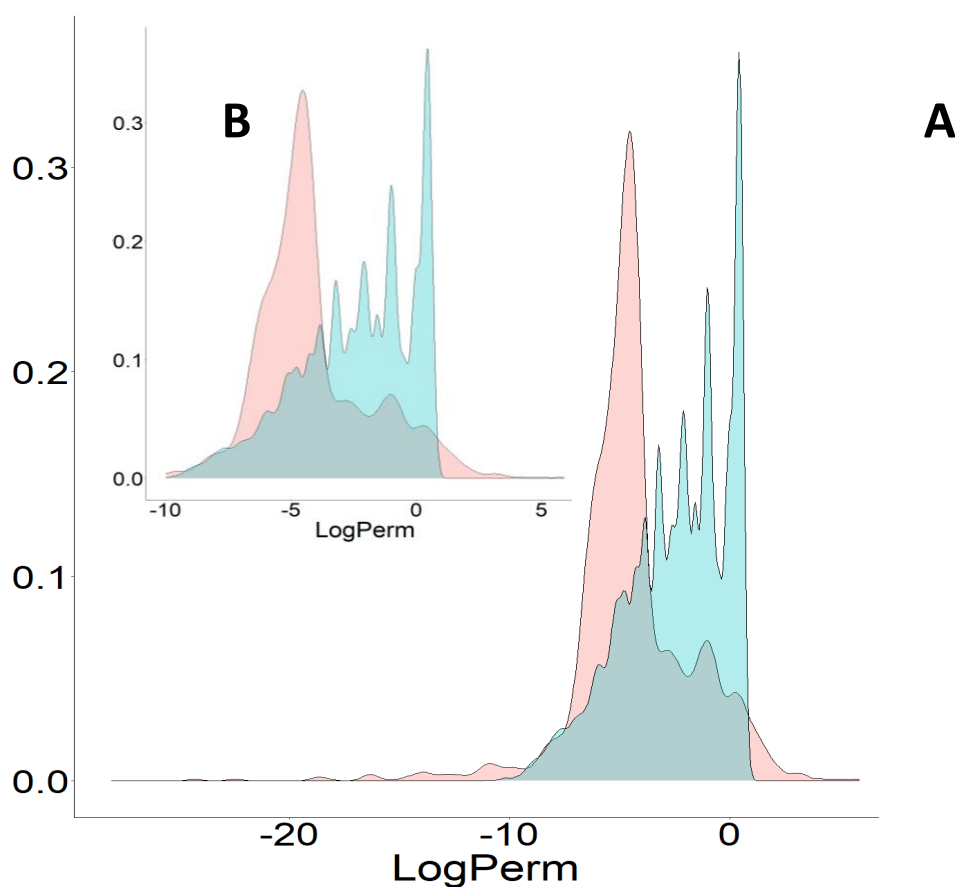
V případě substrátů a non-substrátů pozorujeme na obrázku 28, že většina dat se nachází mezi -5 a 10 jednotkami $\log P$ a jejich molekulová hmotnost v drtivé většině nepřekročí hranici 1000 Da. Pro inhibitory a non-inhibitory zastoupení i molekul s $\log P$ menším, než je -5, tento trend je velmi výrazný hlavně u inhibitorů, kde jsou rovněž také opět vidět peptidové toxiny.

Pokud dáme do souvislosti obrázky 27 a 28 zjistíme, že nejméně lipofilní molekuly v databázi – peptidy – jsou inhibitory iontových kanálů. Ve všech ostatních případech můžeme jen konstatovat, že bez ohledu na typ a na transportér se většinou molekuly s aktivními interakcemi vyskytují mezi hodnotami $\log P$ od -5 do 10 .

8.2 Analýza hodnot $\log Perm$ databázi MolMeDB

Zatímco $\log P$ popisuje míru lipofility dané molekuly, $\log Perm$ popisuje schopnost molekuly membránou procházet. V případě této kapitoly budu pracovat s daty z „překryvu“ a s daty „pasivními“.

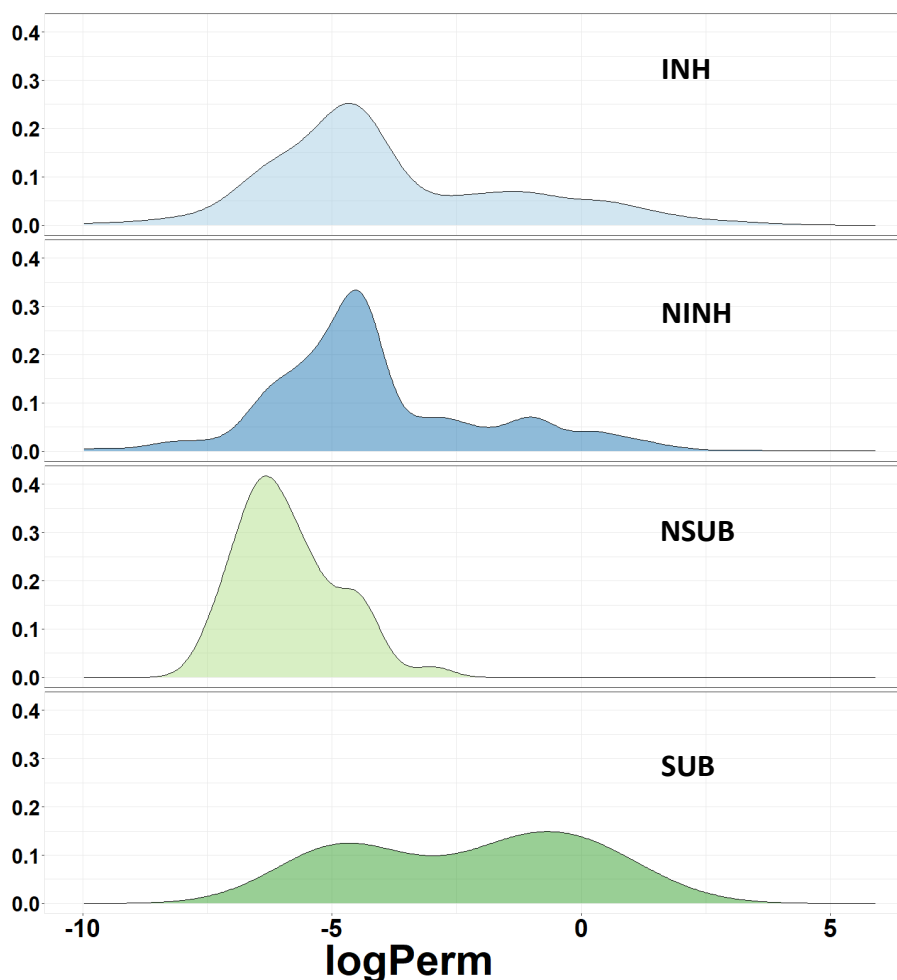
Na obrázku 29 jsou vyobrazeny hustotní funkce pro „překryv“ (růžová) a pro všechny „pasivní“ interakce (světle modrá). Přerušované čáry ukazují pozici aritmetického průměru hodnot $\log Perm$. Na obrázku 29 vidíme, že v oblasti od -10 do 0 je lokalizovaná drtivá většina dat, pro které je $\log Perm$ dostupný, zatímco pro hodnoty menší než -10 logaritmických jednotek roste podíl dat z překryvu. Aritmetický průměr této veličiny pro „překryv“ je $-4,35$ (tedy jde o $10^{-4,35}$ cm/s) zatímco pro všechny pasivní interakce činí pouze $-2,41$ ($10^{-2,41}$ cm/s). Můžeme tedy říci, že data z překryvu mají nižší hodnoty permeabilit než data „pasivní“ zhruba o dva řády.



Obrázek 29: Hustotní funkce pro hodnoty $\log Perm$ (růžová - pro překryv, modrá - všechny pasivní interakce kromě interakcí z překryvu), (vytvořeno programem RStudio) A – celková hustotní funkce pro hodnoty $\log Perm$, B – detail hustotní funkce pro hodnoty $\log Perm$ větší než -5 (vytvořeno programem Rstudio)

Tento rozdíl se nicméně poměrně výrazně změní, pokud si dekomponujeme data na jednotlivé skupiny látek. V následujícím grafu jsme se snažili rozebrat, jak se od sebe liší jednotlivé skupiny – u dat z překryvu můžeme porovnat jednotlivé typy, u pasivních dat pak data odpovídající různým membránám/metodám.

Obrázek 30 ukazuje hustotní funkce hodnoty $\log Perm$ větší než -10 z překryvu rozdělené dle jejich kategorií na substráty, non-substráty, non-inhibitory a inhibitory.

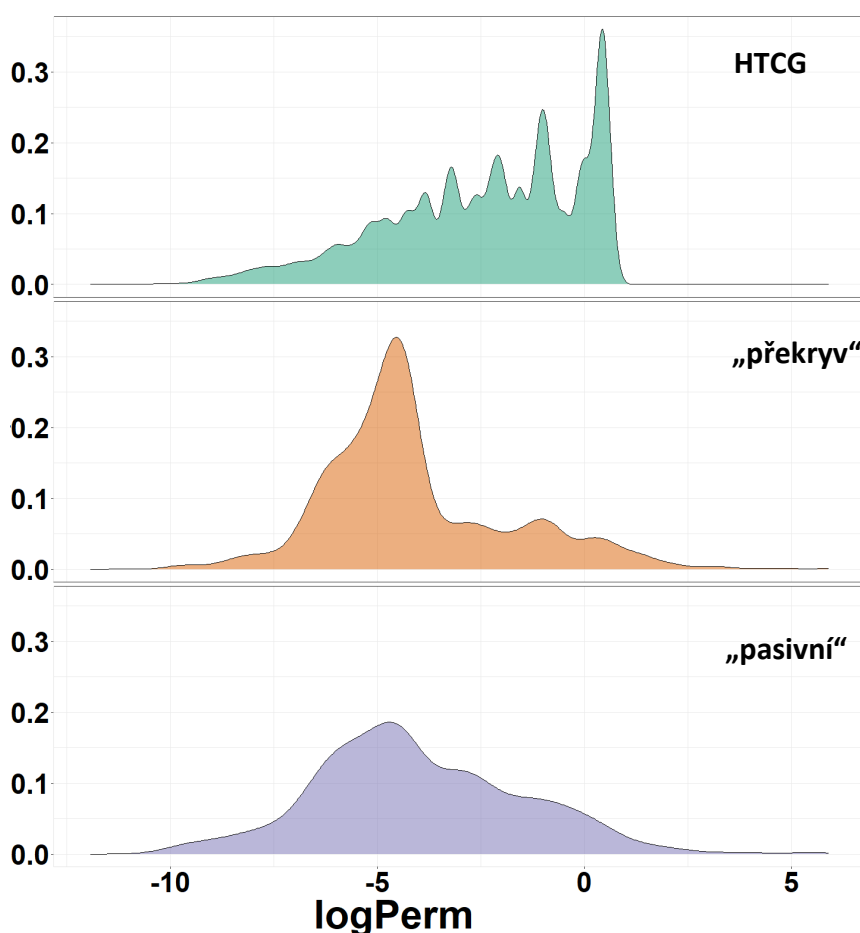


Obrázek 30: Hustotní funkce $\log Perm$ jednotlivých typů z oblasti překryvu (INH – inhibitory, NINH – non-inhibitory, NSUB – non-substráty, SUB – substráty) (vytvořeno programem RStudio)

Průměrná hodnota $\log Perm$ pro inhibitory činila -3,87 ($10^{-3,87}$ cm/s), pro non-inhibitory - 4,11 ($10^{-4,11}$ cm/s), pro non-substráty -5,88 ($10^{-5,88}$ cm/s) a pro substráty - 2,28 ($10^{-2,88}$ cm/s). Na základě těchto hodnot můžeme říci, že látky klasifikované jako non-substráty mají nejmenší hodnoty permeabilit a budou tedy nejhůře procházet membránami prostou difúzí. Největší hodnotu naopak mají molekuly klasifikované jako substráty.

Dále si můžeme povšimnout, že průměrné hodnoty non-inhibitorů a inhibitorů jsou velmi blízké, a i tvar jejich hustotních funkcí je velmi podobný. Důvodem pro to tak je, může být vysoká míra překryvu mezi inhibitory a non-inhibitory, kterou pro oblast „překryvu“ ukazuje obrázek 22. Z obrázku 22 rovněž vyplývá, že velká část molekul se v datech nachází klasifikovaná jako více než jeden typ a tato skutečnost se rovněž promítá do tvarů křivek na obrázku 30.

Z dalších trendů můžeme vybrat nejpočetnější skupinu HTCG simulací – na jejich datech vidíme jasné rozkategorizování do skupin způsobených zhrubeným popisem molekuly – hodně molekul pak je reprezentováno stejnou kombinací (dvojití) zhrubených „korálků“ (tzv. coarse-grain beads), s diskrétními interakcemi. Výsledkem pak je, že mnoho molekul má pak velmi podobné predikované $\log Perm$, které je udáno právě kombinací zhrubených korálků. HTCG simulace také obsahují jen malé molekuly do 8 těžkých atomů tvořených hlavně uhlíkem, a proto se tento dataset jeví jako velmi membránově propustný (viz obrázek 31).



Obrázek 31: : Hustotní funkce $\log Perm$ z oblasti „překryvu“ a „pasivních“ interakcí (HTCG - hodnoty $\log Perm$ z HTCG simulací ze studie Menichetti et al⁵⁵, „překryv“ – hodnoty $\log Perm$ z oblasti „překryvu“, „pasivní“ – hodnoty $\log Perm$ pro „pasivní“ interakce očištěné o data z HTCG simulací) (vytvřeno programem RStudio)

V článku Balouch et al.⁵⁶ pro prostup látek z liposomu autoři stanovili hranice pro propustnost membrány v následujících rozmezích *logPerm*:

- $\logPerm > -4$ – látky snadno prostupné membránou
- $-4 > \logPerm > -8$ – látky prostupné membránou
- $\logPerm < -8$ – látky neprostupné membránou

Pokud použijeme tyto hranice pro naše datasety, tak u látek z překryvu pozorujeme větší podíl látek spíše neprostupných membránou (snadno prostupné 31%, prostupné 64%, neprostupné 5,1%) než u datasetu s pasivními interakcemi i po vynechání HTCG simulací^z (42,7%, 53,1%, 4,2%). Potvrzuje se tak alespoň částečně předpoklad, že látky se známými aktivními interakcemi je mají proto, že jsou hůře pasivně propustné přes membrány.

U zbytku látek s aktivními interakcemi by bylo záhodno pasivní permeace získat a zvětšit tak doposud malý překryv mezi oběma částmi databáze MolMeDB.

^z Kdybychom je nevynechaly, tak by byly výsledky pro pasivní látky ještě propustnější, neboť HTCG data by díky svému počtu zcela dominovala.

Závěr

Cílem této práce bylo získat data o interakcích mezi molekulami a membránovými transportéry z několika otevřených databází, poté tato data doplnit do databáze MolMeDB, která do té doby obsahovala jen data o interakci mezi molekulami a membránami, a následně celý obsah této databáze podrobit analýze.

Pro získávání dat bylo použito data mining workflow vytvořené v programu KNIME. Toto workflow vzniklo v rámci spolupráce s Mgr. Alžbětou Türkovou, Ph.D. a Mag. Dr. Barbarou Zdrzil, Priv. Doz. z Univerzity ve Vídni. Toto workflow je schopno získávat data ze čtyř bioinformatických databází - PubChem, ChEMBL, Metrabase a IUPHAR. Při získávání dat byla pozornost zaměřena především na vybrané napěťově řízené kanály, ABC transportéry, SLC transportéry a ATPázové pumpy.

Po úpravách a validaci dat se databáze MolMeDB velikostně i tematicky významně rozšířila. Pro každou molekulu v databázi, pro kterou byl získána data pro transportéry, je nyní možné najít v MolMeDB popisy těchto získaných „aktivních“ interakcí charakterizovaných pomocí typu (substrát, non-substrát, inhibitor a non-inhibitor) či pomocí numerických hodnot příslušných měřených veličin (K_M , K_i , IC_{50} a EC_{50}).

Data v databázi MolMeDB byla následně podrobena analýze. Pro účely analýz byla data rozdělena na „pasivní“, „aktivní“ a „překryv“. Přičemž nejdetaileji byla studována oblast „překryvu“. Všechny tyto tři oblasti byly zkoumány z hlediska kvantitativního a z hlediska fyzikálně chemických vlastností (MW , $logP$ a $logPerm$).

„Pasivní“ interakce jsou z hlediska kvantitativního zastoupení nejhodněji zastoupeny simulačními daty. Největší díl „aktivních“ interakcí připadá na SLC transportéry, pokud se ale omezíme jen na „překryv“ zjistíme, že zde lehce převažují ABC transportéry nad SLC transportéry. Analýzou typů „aktivních“ interakcí bylo zjištěno, že mnohdy je jedna molekula klasifikována více než jedním typem interakce a také se často vyskytují u kombinace transportérů (hlavně ABCC) a ty tak tvoří podstatnou část všech interakcí této oblasti.

Analýzou $logPerm$ „pasivních“ dat a z „překryvu“ bylo zjištěno, že permeabilita dat z „překryvu“ je nižší než permeabilita dat, pro které jsou dostupné jen „pasivní“ interakce, a to v průměru zhruba o dva řády. Nicméně valná část látek účastnících se aktivních interakcí je schopna se zachytávat v membránách. Bohužel pouze na základě sledovaných $logP$ a MW nelze rozeznat substráty od inhibitorů.

Postupně zvětšující se množství dat nám snad do budoucna dá odpovědi na otázky typu, jaké funkční skupiny, či skelety jsou schopny penetrace přes biologické membrány. Rovněž v budoucnu bychom mohly být schopni dalším studiem interakcí molekul a transportérů najít klíčové fyzikálně chemické anebo strukturní vlastnosti ovlivňující jejich interakci. MolMeDB by tak mohla být využita pro návrhy systémů pro doručování léčiv přes membrány.

Summary

This diploma thesis aimed to obtain data about interaction among molecules and membrane transporters from open databases; add them to the MolMeDB database; and subsequently analyse all content of this database.

Data mining workflow was created by KNIME software in collaboration with Mgr. Alžběta Tuerková, Ph.D. and Mag. Dr. Barbara Zdražil, Priv. Doz. (University of Vienna). Workflow can fetch data from four bioinformatics databases - PubChem, ChEMBL, Metrabase and IUPHAR. The attention was focused mainly on selected voltage-gated ion channels, ABC transporters, SLC transporters and ATPase pumps.

After inclusion and validation of the data, the MolMeDB database was expanded thematically and in size. For each obtained molecule it is now possible to find in MolMeDB among other things the expression of such „active“ interactions characterized by so-called type (substrate, non-substrate, inhibitor and non-inhibitor) and by numerical values of relevant quantities (K_M , K_i , IC_{50} and EC_{50}).

MolMeDB data were further analysed upon division into three sections - „passive“, „active“ and „overlap“. The section „overlap“ was studied in a detail. All three sections were studied in terms of quantity and physicochemical properties (MW , $logP$ a $logPerm$).

The „passive“ interactions are the most represented by simulation data. The largest part of the „active“ interaction is represented by data for SLC transporters, but if we limit our consideration to „overlap“, we find that ABC transporters on SLC slightly predominate here. In the analysis of types within the „active“ interactions, it was revealed that often one molecule is classified as more than one type and interacts with multiple transporters (mainly ABCC).

The analysis of $logPerm$ values of „passive“ data and „overlap“ was revealed that the molecules from „overlap“ region are less permeable than the molecules with "passive" interactions in general, averaging about two orders of magnitude. However, most of the substances involved in „active“ interactions can be trapped in membranes. Unfortunately, only on the basis of the physicochemical properties monitored here, we cannot distinguish substrates from inhibitors.

The gradually increasing amount of data will hopefully give us answers to questions such as which functional groups or skeletons can penetrate through biological membranes in the future. Also in the future, we could be able to further study the interactions of molecules and transporters to find key physicochemical or structural properties influencing their interaction. MolMeDB could thus be used to design systems for drug delivery across membranes.

Reference

1. Watson, H. Biological membranes. *Essays Biochem.* **59**, 43–69 (2015).
2. Nicolson, G. L. The Fluid - Mosaic Model of Membrane Structure: Still relevant to understanding the structure, function and dynamics of biological membranes after more than 40 years. *Biochim. Biophys. Acta - Biomembr.* **1838**, 1451–1466 (2014).
3. Cell Membranes and the Fluid Mosaic Model. Available at: <https://courses.lumenlearning.com/boundless-ap/chapter/cell-membranes-and-the-fluid-mosaic-model/>. (Accessed: 11th February 2021)
4. Lodish, H., Berk, A. & Zipursky, S. *Molecular Cell Biology*. *Molecular Cell Biology* (W. H. Freeman, 2000).
5. Parisio, G., Stocchero, M. & Ferrarini, A. Passive Membrane Permeability: Beyond the Standard Solubility-Diffusion Model. *J. Chem. Theory Comput.* **9**, 5236–5246 (2013).
6. Thompson Coon, J. Goodman and Gilman's the Pharmacological Basis of Therapeutics. *Focus Altern. Complement. Ther.* **7**, 207–207 (2010).
7. Cocucci, E., Kim, J., Bai, Y. & Pabla, N. Role of Passive Diffusion, Transporters, and Membrane Trafficking-Mediated Processes in Cellular Drug Transport. *Clin. Pharmacol. Ther.* **101**, 121–129 (2017).
8. Robert K. Murray, David Bender, Kathleen M. Botham, Peter J. Kennelly, Victor W. Rodwell, P. A. W. Harper's Illustrated Biochemistry. in *Harper's Illustrated Biochemistry* (ed. McGraw-Hill Medical) 459–477 (McGraw-Hill Medical Publishing, 2012).
9. International Transporter Consortium *et al.* Membrane transporters in drug development. *Nat. Rev. Drug Discov.* **9**, 215–36 (2010).
10. Sugano, K. *et al.* Coexistence of passive and carrier-mediated processes in drug transport. *Nat. Rev. Drug Discov.* **9**, 597–614 (2010).
11. Alberts B, Johnson A, Lewis J, *et al.* *Molecular Biology of the Cell*. (2002).
12. Bagal, S. K. *et al.* Ion Channels as Therapeutic Targets: A Drug Discovery Perspective. *J. Med. Chem.* **56**, 593–624 (2013).
13. Demarche, S., Sugihara, K., Zambelli, T., Tiefenauer, L. & Vörös, J. Techniques for recording reconstituted ion channels. *Analyst* **136**, 1077 (2011).
14. The IUPHAR/BPS Guide to PHARMACOLOGY. Available at: <https://www.guidetopharmacology.org/GRAC/IonChannelListForward?class=VGIC>. (Accessed: 17th February 2021)

15. Kass, R. S. The channelopathies: novel insights into molecular and genetic mechanisms of human disease. *J. Clin. Invest.* **115**, 1986–1989 (2005).
16. Colas, C., Ung, P. M.-U. & Schlessinger, A. SLC transporters: structure, function, and drug discovery. *Medchemcomm* **7**, 1069–1081 (2016).
17. IUPHAR/BPS Guide to PHARMACOLOGY. SLC superfamily of solute carriers. Available at: <http://www.guidetopharmacology.org/GRAC/FamilyDisplayForward?familyId=863>. (Accessed: 7th February 2021)
18. NCCR TransCure. SLC Tables. Available at: <http://slc.bioparadigms.org/>. (Accessed: 7th February 2021)
19. Wang, W. W., Gallo, L., Jadhav, A., Hawkins, R. & Parker, C. G. The Druggability of Solute Carriers. *J. Med. Chem.* **63**, 3834–3867 (2020).
20. Hagenbuch, B. & Stieger, B. The SLCO (former SLC21) superfamily of transporters. *Mol. Aspects Med.* **34**, 396–412 (2013).
21. Vasiliou, V., Vasiliou, K. & Nebert, D. W. Human ATP-binding cassette (ABC) transporter family. *Hum. Genomics* **3**, 281 (2008).
22. Lodish H, Berk A, Zipursky SL, et al. *Molecular Cell Biology*. 4th edition. New York: W. H. Freeman; 2000. Section 15.5, Active Transport by ATP-Powered Pumps. Available at: <https://www.ncbi.nlm.nih.gov/books/NBK21481/>. (Accessed: 19th February 2021)
23. Baynes, J. W. & Dominiczak, M. H. *Medical Biochemistry*. (Mosby Elsevier, 2009).
24. Robey, R. W. *et al.* Revisiting the role of ABC transporters in multidrug-resistant cancer. *Nat. Rev. Cancer* **18**, 452–464 (2018).
25. Sodani, K., Patel, A., Kathawala, R. J. & Chen, Z.-S. Multidrug resistance associated proteins in multidrug resistance. *Chin. J. Cancer* **31**, 58–72 (2012).
26. Bateman, A. *et al.* UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).
27. Kim, S. *et al.* PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res.* **47**, D1102–D1109 (2019).
28. Mendez, D. *et al.* ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res.* **47**, D930–D940 (2019).
29. Mak, L. *et al.* Metrabase: A cheminformatics and bioinformatics database for small molecule transporter data analysis and (Q)SAR modeling. *J. Cheminform.* **7**, (2015).

30. Armstrong, J. F. *et al.* The IUPHAR/BPS Guide to PHARMACOLOGY in 2020: extending immunopharmacology content and introducing the IUPHAR/MMV Guide to MALARIA PHARMACOLOGY. *Nucleic Acids Res.* (2019). doi:10.1093/nar/gkz951
31. Juračka, J., Šrejber, M., Melíková, M., Bazgier, V. & Berka, K. MolMeDB: Molecules on Membranes Database. *Database* **2019**, (2019).
32. UniProt. Available at: <https://www.uniprot.org/>. (Accessed: 6th March 2021)
33. Kim, S. *et al.* PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res.* **49**, D1388–D1395 (2021).
34. Kim, S. *et al.* PubChem Substance and Compound databases. *Nucleic Acids Res.* **44**, D1202–D1213 (2016).
35. What is the difference between a substance and a compound in PubChem? Available at: <https://pubchemblog.ncbi.nlm.nih.gov/2014/06/19/what-is-the-difference-between-a-substance-and-a-compound-in-pubchem/>. (Accessed: 6th March 2021)
36. Wang, Y. *et al.* PubChem's BioAssay Database. *Nucleic Acids Res.* **40**, D400–D412 (2012).
37. PubChem BioAssays. Available at: <https://pubchemdocs.ncbi.nlm.nih.gov/bioassays>. (Accessed: 6th March 2021)
38. ChEMBL. Available at: <https://www.ebi.ac.uk/chembl/>. (Accessed: 3rd March 2021)
39. Chambers, J. *et al.* UniChem: a unified chemical structure cross-referencing and identifier tracking system. *J. Cheminform.* **5**, 3 (2013).
40. Ozawa, N. *et al.* Transporter Database, TP-Search: A Web-Accessible Comprehensive Database for Research in Pharmacokinetics of Drugs. *Pharm. Res.* **21**, 2133–2134 (2004).
41. Sedykh, A. *et al.* Human Intestinal Transporter Database: QSAR Modeling and Virtual Profiling of Drug Uptake, Efflux and Interactions. *Pharm. Res.* **30**, 996–1007 (2013).
42. The IUPHAR/BPS Guide to PHARMACOLOGY. Available at: <https://www.guidetopharmacology.org/about.jsp>. (Accessed: 1st March 2021)
43. Wishart, D. S. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **34**, D668–D672 (2006).
44. Burley, S. K. *et al.* Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Res.* **47**, D520–D528 (2019).

45. Berthold, M. R. *et al.* KNIME - the Konstanz information miner. *ACM SIGKDD Explor. Newsl.* **11**, 26–31 (2009).
46. KNIME. Getting Started Guide. Available at: <https://www.knime.com/getting-started-guide>. (Accessed: 3rd March 2021)
47. Türková, A., Jain, S. & Zdrzil, B. Integrative Data Mining, Scaffold Analysis, and Sequential Binary Classification Models for Exploring Ligand Profiles of Hepatic Organic Anion Transporting Polypeptides. *J. Chem. Inf. Model.* **59**, 1811–1825 (2019).
48. Gadaleta, D., Lombardo, A., Toma, C. & Benfenati, E. A new semi-automated workflow for chemical data retrieval and quality checking for modeling applications. *J. Cheminform.* **10**, 60 (2018).
49. Landrum, G. RDKit: Open-Source Cheminformatics Software. (2016).
50. RStudio team. RStudio: Intergrated Development Environment for R. (2016).
51. Lomize, A. L. & Pogozeva, I. D. Physics-Based Method for Modeling Passive Membrane Permeability and Translocation Pathways of Bioactive Molecules. *J. Chem. Inf. Model.* **59**, 3198–3213 (2019).
52. Kurosawa, T. *et al.* Expression and Functional Characterization of Drug Transporters in Brain Microvascular Endothelial Cells Derived from Human Induced Pluripotent Stem Cells. *Mol. Pharm.* **15**, 5546–5555 (2018).
53. Dahlgren, D. *et al.* Regional intestinal permeability of three model drugs in human. *Mol. Pharm.* **13**, 3013–3021 (2016).
54. Dahlin, A. *et al.* Gene Expression Profiling of Transporters in the Solute Carrier and ATP-Binding Cassette Superfamilies in Human Eye Substructures. *Mol. Pharm.* **10**, 650–663 (2013).
55. Menichetti, R., Kanekal, K. H. & Bereau, T. Drug–Membrane Permeability across Chemical Space. *ACS Cent. Sci.* **5**, 290–298 (2019).
56. Balouch, M. *et al.* In silico screening of drug candidates for thermoresponsive liposome formulations. *Mol. Syst. Des. Eng.* **6**, 368–380 (2021).
57. Verkman, A. S. Aquaporins in Clinical Medicine. *Annu. Rev. Med.* **63**, 303–316 (2012).
58. Daylight. 4. SMARTS - A Language for Describing Molecular Patterns. Available at: <https://www.daylight.com/dayhtml/doc/theory/theory.smarts.html>. (Accessed: 6th March 2021)

Přílohy

Tabulka 12: Dělení jednotlivých membrán do skupin. Čísla v kulatých závorkách ukazují počet záznamů pro každou konkrétní membránu (aktuální informace k 26.3.2021)

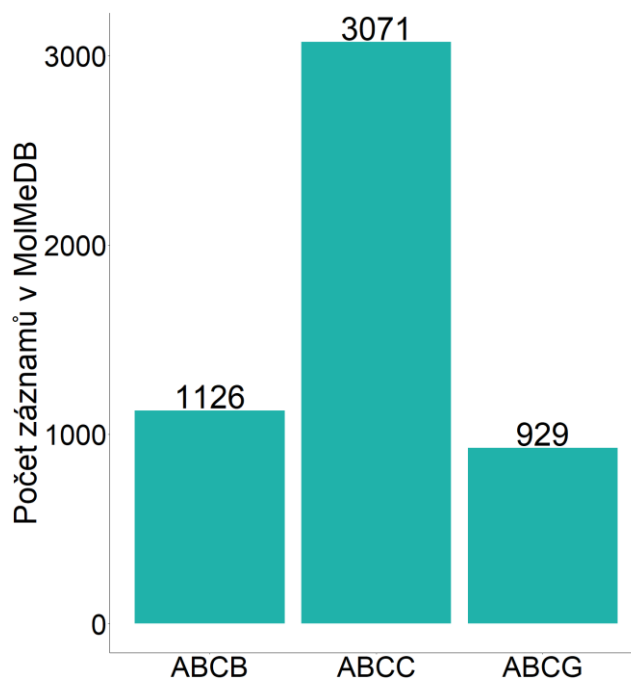
Skupina membrán podle původu	modelová/experimentální membrána (zkratka skupiny)	Konkrétní membrána
Mozek	Modelová (BM)	DPhPC (24)
	Experimentální (BE)	BBB (175), HBMEC (7)
Buňka	Modelová (CM)	DMPC (337), DOPC (435146), DPPC (13), POPC (69), PL-A (182), DOPC_DOPE_CHOL (1), SM_CHOL (1), LEC (14)
	Experimentální (CE)	EggPC (42), MDCK (257), CALU3 (28)
Oko	Modelová (EM)	-
	Experimentální (EE)	Cornea (107), Sclera (19), Conjunctiva (13), Stroma of cornea (7), Corneal endothelium (6)
Generická membrána	Modelová (GM)	CHCl ₃ (10), TOL (2), n-Hex (31), POPE (5), POPC:POPE_1:1 (1), OCT, DLPC (1), DOPS (1), DLPE (1)
	Experimentální (GE)	GENER (1778)
Střevo	Modelová (IM)	GIT-0 (124), BML (100)
	Experimentální (IE)	CACO-2 (636), JEJ (5), Caco2/MDCK (148)
Orální	Modelová (OM)	-
	Experimentální (OE)	Sublingual mucous membrane (4)
Kůže	Modelová (SM)	SC mix (325), CERNS (11), SC ML (8), PSKIN (61)
	Experimentální (SE)	EPD, SC (13)

Tabulka 13: Dělení jednotlivých metod do skupin. Názvy konkrétních metod jsou shodné s názvy, které jsou použity v rámci MolMeDB. Čísla v kulatých závorkách ukazují počet záznamů pro každou konkrétní metodu (aktuální informace k 26.3.2021)

hlavní skupina metod	skupina metod	Konkrétní metoda
Výpočetní metody	COSMOmic	CCM15 (961), CCM18 (438)
	Mechanistický model	CPapp (148)
	QSAR	CVOLSURF (51), QSAR-HB (113), QSAR-PSA (113)
	PerMM	CPMM (1377)
	QikProp	CQikProp (4)
Simulační metody	MD(Amber)	MDA14A99 (7), MDSLA99SB (1),
	MD(CHARMM)	MDABFC36 (9), MDC36 (7), MDCH27 (5)
	MD(Gromos)	MDUAG (2), MDG43A1S3 (1), MDG87 (1)
	MD (Umbrella)	MDUSA4A9 (14), MDUSBG56 (18), MDUSC36 (12), MDUSG43A1S3 (2), MDUSSlip (7)
	MD(Martini)	MMAR (2)
	HTCG simulace	MDCG (434257)
	MD (ostatní)	MDD (4), MDDMM (6), MDSmC36 (2)
Experimentální metody	NMR metody	DC-DMR (1), ENMRD (1),
	EPR metody	EEPR (1)
	Fluorescenční metody	ECFM (73), EFALUV (8), EFCS (1), EPFQ (1), ETRFES (22), ELFA (7)
	Titrační metody	EPHT (26), ITC (8)
	MDCK	EMDCK (257)
	CACO	ECACO (326),
	Calu-3	ECALU3 (28)
	Black lipid membrane metoda	EBLM (252)
	PAMPA	EBAMP (121), EPAM (814)
	Metody s difúzní celou	EFDC (19), EUC (35)
	Mikroextrakční metody	ESPME (17)
	Perfúzní systémy	EPS (112)
	Steady-state metody	ESSM (12)
	Ostatní experimentální metody	MA-GUV (3), EGUV-CCD (1)

Analýza konkrétních rodin transportérů

ABC transportéry lze rozdělit do sedmi podrodin (ABCA až ABCG) jak bylo popsáno v teoretické části. Ve výsledcích se objevují data pro tři z těchto rodin, a to pro rodiny ABCB, ABCC a ABCG. Jejich zastoupení v datech ukazuje obrázek 32.

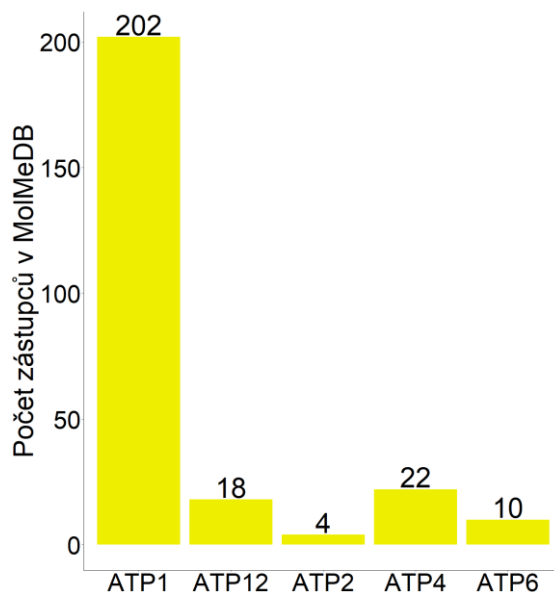


Obrázek 32: Zastoupení jednotlivých podrodin ABC transportérů v MolMeDB (vytvořeno pomocí programu RStudio)

Z tohoto obrázku je rovněž patrné, že absolutně nejvíce je zastoupena rodina ABCC a to 3071 jedna záznamy. Z čehož 223 záznamů připadá na transportér CFTR (neboli ABCC7), který je rovněž členem této rodiny. Nejhojnějším členem ABCC rodiny je v tomto případě je člen ABCC2, na který připadá 1240 záznamů, následují ho ABCC4 se 626 záznamy a ABCC3 s 586 záznamy. Druhé dvě skupiny, které se v datech vyskytují jsou ABCB, pro kterou je v datech 1126 záznamů a ABCG s 929 záznamy.

Data, která jsou pro ABC transportéry v MolMeDB aktuálně dostupná jsou z drtivé části klasifikována jako inhibitory případně non-inhibitory, celkem na tyto dvě skupiny připadá 4264 záznamů, zatím co substráty případně non – substráty jsou zastoupeny pouze 862 záznamy.

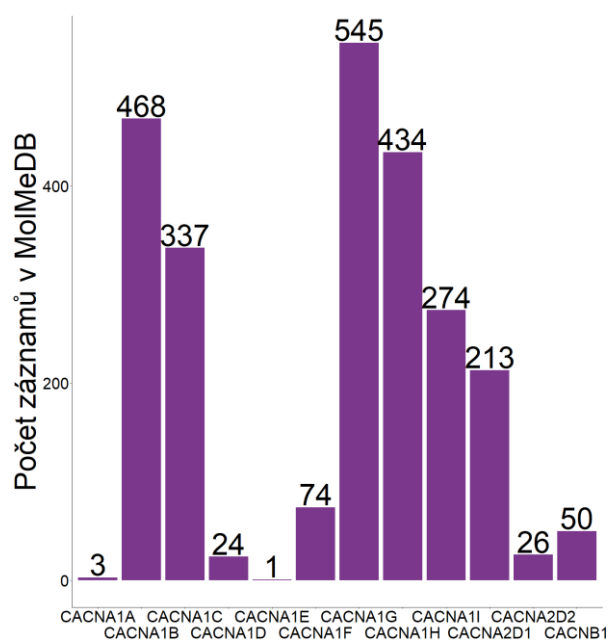
Existují tři druhy ATPáz (P-typ, V-typ a F-typ) v datech jsou zastoupeny všechny dva tyto typy. V-typ ATPáz je označen jako ATP6. ATP1, ATP2, ATP4 a ATP12 patří do kategorie P-ATPáz. Pro F-ATPázy žádný zástupce v MolMeDB není. Zastoupení jednotlivých skupin ATPáz v MolMeDB ukazuje obrázek 33.



Obrázek 33: Zastoupení jednotlivých podrodin ATPáz v MolMeDB (vytvořeno pomocí programu RStudio)

Pro ATPázy není v datech jediný záznam o substrátech případně non-substrátech. 112 záznamů je klasifikováno jako inhibitory a 144 jako non-inhibitory těchto transportních proteinů.

Vápenaté kanály tvoří velice širokou skupinu. V této práci je pozornost věnována primárně napětově řízeným vápenatým kanálům, ačkoliv několik málo zástupců ligandem řízených kanálů bude v kategorii „ostatní“. Obrázek 34 ukazuje jednotlivé zástupce napětově řízených vápenatých kanálů v MolMeDB.

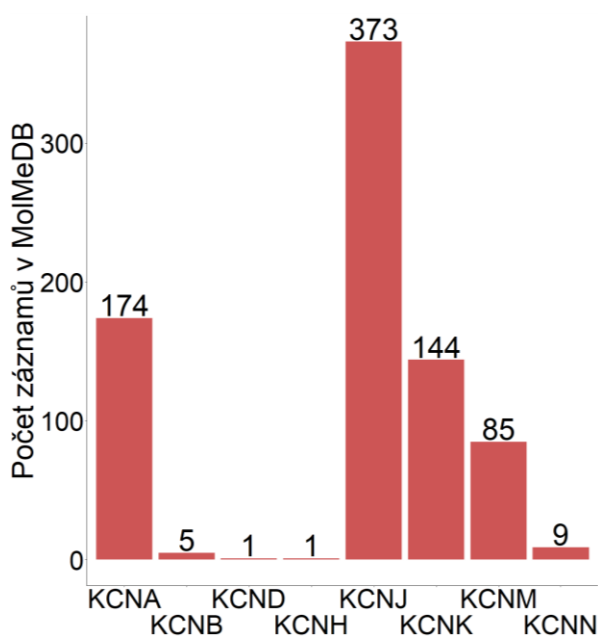


Obrázek 34: Zastoupení jednotlivých zástupců vápenatých kanálů v MolMeDB (vytvořeno pomocí programu RStudio)

Z obrázku 34 vidíme, že nejhojněji zastoupeným vápenatým kanálem je CACNA1G (545 záznamů), který patří spolu s CACNA1H (434 záznamů) a s kanálem CACNA1I (274 záznamů), do T-typu vápenatých kanálů. Dalším hojněji zastoupeným typem je N-typ, který zastupuje CACNA1B (446 záznamů). L-typ vápenatých kanálů v datech zastupují CACNA1C (337 záznamů), CACNA1D (24 záznamů) a CACNA1F (74 záznamů). Ostatní typy kanálů nejsou již tak hojně zastoupeny a počet záznamů pro něj je velmi malý.

Opět v této skupině z hlediska ligandů vedou inhibitory s non – inhibitory (2 404 záznamů). Na substráty a non-substráty připadá dohromady jen 45 záznamů.

Obrázek 35 ukazuje zastoupení jednotlivých skupin draselných kanálů. Podle tabulky 1 lze tyto skupiny klastrovat dále. Skupiny KCNA, KCNB, KCND, KCNH patří mezi tzv. napěťově řízené draselné kanály. Skupina označená jako KCNJ je tvořena zástupci tzv. vnitřně usměřujících draselných kanálů. Kanály skupiny KCNK lze označit jako draselné kanály v doméně tandemových pórů. KCNN a KCNM lze označit souhrnným názvem vápníkem a sodíkem aktivované draselné kanály.



Obrázek 35: Zastoupení jednotlivých skupin kanálů draselných iontů v MolMeDB (vytvořeno pomocí programu RStudio)

Vidíme, že zdaleka největší počet záznamů mám pro skupinu vnitřně usměřujících draselných kanálů (KCNJ), pro kterou máme 373 záznamů. Pro napěťově řízené draselné kanály (KCNA, KCNB, KCND, KCNH) máme dohromady 181 záznamů. Pro skupinu draselných kanálů v doméně tandemových pórů (KCNK) je 144 záznamů v datech. Vápníkem a sodíkem aktivované draselné kanály (KCNN, KCNM) dohromady mají 94 záznamů.

Opět z největší díl dat pro draselné kanály připadá na inhibitory a non-inhibitory, které dohromady tvoří 652 záznamů, zatím co na substráty případně non-substráty připadá 140 záznamů.

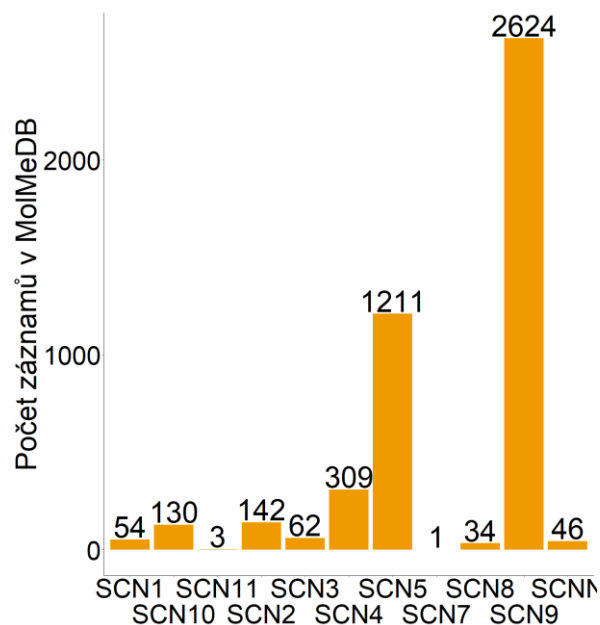
SLC transportéry tvoří téměř polovinu všech dat týkajících se transportérů v MolMeDB. V tomto případě byla získána data pro 30 rodin těchto transportérů. Tabulka 12 ukazuje množství záznamů, které pro tyto transportéry je v MolMeDB. Z hlediska kvantity vede rodina SLC6, pro kterou je v datech 6818 záznamů. Jedná se přenašeče, které transportují neurotransmitery, jak ukazuje tabulka 3. Další z nejhojněji zastoupených SLC rodin je SLC5 se 1357 záznamy. Členové této rodiny se podílejí na transport sacharidů, lipidů či vitamínů (viz. tabulka 3).

Tabulka 12: Zastoupení jednotlivých podrodin SLC transportérů v MolMeDB

SLC rodina	Počet záznamů v MolMeDB	SLC rodina	Počet záznamů v MolMeDB	SLC rodina	Počet záznamů v MolMeDB
SLC1	374	SLC13	13	SLC28	82
SLC2	181	SLC14	2	SLC29	159
SLC5	1357	SLC15	284	SLC32	1
SLC6	6818	SLC16	73	SLC36	37
SLC7	54	SLC17	1	SLC37	11
SLC8	2	SLC18	111	SLC40	38
SLC9	400	SLC19	22	SLC43	3
SLC10	228	SLC22	894	SLC46	13
SLC11	96	SLC25	2	SLC47	139
SLC12	15	SLC27	96	SLCO	824

I v tomto případě dominují inhibitory a non-inhibitory v datech, připadá na obě tyto skupiny 11191 záznamů na substráty a non-substráty poté připadá pouze 1139 záznamů, tedy téměř desetkrát méně než na inhibitory a non-inhibitory.

Kanály sodíkových iontů jsou třetí nejvíce zastoupenou skupinou v datech. V rámci MolMeDB je dělíme na napětově řízené sodíkové kanály a epiteliální sodíkové kanály, které nejsou napětově řízené. Do prvního typu sodíkových kanálů patří téměř všechny skupiny, které ukazuje obrázek 36, kromě skupiny SCNN, jejíž členové se řadí mezi epiteliální sodíkové kanály. Napětově řízené sodíkové kanály tvoří drtivou většinu dat i kvantitativně, dohromady pro ně je v MolMeDB 4570 záznamů, přičemž drtivá část těchto dat připadá na skupinu SCN9.



Obrázek 36: Zastoupení jednotlivých skupin kanálů sodných iontů v MolMeDB (vytvořeno pomocí programu RStudio)

I v tomto případě drtivou většinu látek, které pro ně byly data miningem získány, tvoří inhibitory a non-inhibitory (4559 záznamů) a pouze 57 z nich připadá na substráty případně non-substráty.

Ve skupině transportních proteinů, které budou v práci souhrnně označovány jako „ostatní“ se nachází velmi různorodé transportéry. Zahrnutí do skupiny „ostatní“ a skutečnost, že jim v textu práce není věnována taková pozornost je způsobeno tím, že jsem se rozhodla zvýšenou pozornost výše zmíněným šesti skupinám.

Tabulka 13: Zastoupení jednotlivých nezařazených transportní proteinů v MolMeDB

Transportní protein	Počet záznamů v MolMeDB	Transportní protein	Počet záznamů v MolMeDB
CLCN7	3	MCOLN1	3
ASIC3	63	MPC2	1
ANO1	29	NPC1L1	123
TRPV2	4	RYR2	2
TRPC6	24		