

**UNIVERZITA PALACKÉHO V OLOMOUCI**

Přírodovědecká fakulta

Katedra biochemie



Tvorba databáze interakcí malých molekul s biologickými  
membránami

**BAKALÁŘSKÁ PRÁCE**

Autor:	Jakub Juračka
Studijní program:	B1406 Biochemie
Studijní obor:	Bioinformatika
Forma studia:	Prezenční
Vedoucí práce:	doc. RNDr. Karel Berka, Ph.D.
Rok:	2018

Prohlašuji, že jsem bakalářskou práci vypracoval samostatně s vyznačením všech použitých pramenů a spoluautorství. Souhlasím se zveřejněním bakalářské práce podle zákona č. 111/1998 Sb., o vysokých školách, ve znění pozdějších předpisů. Byl jsem seznámen 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 7. května 2018

.....

### **Poděkování:**

Rád bych poděkoval vedoucímu mé bakalářské práce doc. RNDr. Karlu Berkovi, PhD. za cenné rady a připomínky v průběhu celého projektu. Dále bych rád poděkoval Mgr. Ing. Václavu Bazgierovi za pomoc při návrhu a uveřejnění databáze. Práce byla vypracována jako součást grantového projektu č. 17-21122S s názvem *MolMeDB - Ověřené predikce interakcí nízkomolekulárních látek s biologickými membránami*, který je financován Grantovou agenturou České republiky (GAČR).

## Bibliografická identifikace

Jméno a příjmení autora:	Jakub Juračka
Název práce:	Tvorba databáze interakcí malých molekul s biologickými membránami
Typ práce:	Bakalářská práce
Pracoviště:	Katedra fyzikální chemie
Vedoucí práce:	doc. RNDr. Karel Berka, Ph.D.
Rok obhajoby práce:	2018
Klíčová slova:	MolMeDB, biologická databáze, malé molekuly, interakce, biologické membrány, COSMOmic, MVC
Počet stran:	48
Počet příloh:	1 CD
Jazyk práce:	český

## Abstrakt

Většina léčiv dnes patří do skupiny malých molekul. Pro pochopení chování léčiv v těle je zásadní porozumění jejich průniku skrze biologické membrány, který je závislý na tom, jak daná molekula s membránou interaguje. Cílem této práce je návrh a následná tvorba online databáze, která bude shromažďovat informace o jednotlivých molekulách, modelových membránách, metodách a predikovaných a experimentálních hodnotách jednotlivých interakcí. Databáze je volně přístupná s propracovaným uživatelským rozhraním na adrese [molmedb.upol.cz](http://molmedb.upol.cz).

## **Bibliographical Identification**

Autor's first name and surname:	Jakub Juračka
Title:	Developing Database of Interactions of Small Molecules with Biological Membranes
Type of thesis:	Bachelor thesis
Department:	Department of Physical Chemistry
Supervisor:	doc. RNDr. Karel Berka, Ph.D.
Year of presentation:	2018
Keywords:	MolMeDB, biological databases, small molecules, interactions, biological membranes, COSMOmic, MVC
Number of pages:	48
Number of appendices:	1 CD
Thesis language:	Czech

## **Abstract**

Nowadays most drugs belong to the small molecules group. To understand the behavior of the drugs in human body, it is essential to thoroughly recognize the way they penetrate the biological membrane, which is dependent on the interaction between the molecule and the membrane. The aim of this thesis is the proposal and consequential creation of an online database, which will collect information about molecules, model membranes, methods and predicted and experimental values of individual interactions. The database is freely accessible with a sophisticated user interface at [molmedb.upol.cz](http://molmedb.upol.cz)

# Obsah

<b>1</b>	<b>Úvod</b>	<b>8</b>
<b>2</b>	<b>Současný stav řešené problematiky</b>	<b>9</b>
2.1	Biologické databáze . . . . .	9
2.2	Návrh biologické databáze . . . . .	10
2.3	Reprezentace molekulárních struktur v počítači . . . . .	12
2.3.1	Zápis molekul ve formě SMILES . . . . .	12
2.3.2	Další možnosti počítačového zápisu molekul . . . . .	14
2.4	Databáze interakcí . . . . .	16
2.5	Interakce molekul s biologickými membránami . . . . .	17
2.5.1	Biologické membrány . . . . .	17
2.5.2	Struktura a funkce membránových proteinů . . . . .	19
2.5.3	Malé molekuly . . . . .	21
2.5.4	Interakce malých molekul s biologickými membránami . . . . .	21
2.5.5	COSMOmic . . . . .	23
2.5.6	Profil volné energie podél normály membrány . . . . .	24
<b>3</b>	<b>Praktická část</b>	<b>25</b>
3.1	Použité technologie . . . . .	25
3.1.1	HTML5 a CSS . . . . .	25
3.1.2	PHP 7 . . . . .	25
3.1.3	JavaScript . . . . .	26
3.1.4	AJAX (Asynchronní JavaScript a XML) . . . . .	26
3.1.5	MVC architektura . . . . .	27
3.1.6	Databáze . . . . .	28
3.2	Návrh databáze MolMeDB . . . . .	29
3.3	Tvorba webového rozhraní . . . . .	31
3.3.1	MVC architektura databáze MolMeDB . . . . .	31
3.3.2	Souborová struktura aplikace . . . . .	32
<b>4</b>	<b>Výsledky a diskuze</b>	<b>34</b>
4.1	Vyhledávač . . . . .	34
4.2	Zobrazení detailu ligandu . . . . .	35
4.2.1	2D struktury . . . . .	35
4.2.2	3D struktury . . . . .	36
4.2.3	Grafické zobrazení profilů volných energií . . . . .	38
4.2.4	Tabulkové hodnoty interakcí . . . . .	40
4.3	Comparator - sekce pro vzájemné porovnávání molekul . . . . .	41
4.4	Administrátorská sekce . . . . .	42
4.4.1	Sekce pro nahrávání dat . . . . .	43
4.4.2	Sekce pro editace dat . . . . .	44
<b>5</b>	<b>Závěr</b>	<b>45</b>
<b>6</b>	<b>Literatura</b>	<b>46</b>

# **Cíle práce**

## **Teoretická část**

- Literární rešerše stavu poznání v oblasti databází interakcí malých molekul s biomembránami.
- Prostudování odborné literatury.

## **Praktická část**

- Návrh databáze.
- Implementace databáze a vytvoření webového rozhraní.
- Naplnění databáze.

# 1 Úvod

Hlavním tématem této práce je tvorba (biologické) databáze interakcí malých molekul s biologickými membránami. Studium interakcí malých molekul (léčiv) s biomembránami je důležité pro pochopení jejich chování v těle. Malé molekuly s membránami interagují mnoha způsoby – mají k nim různou afinitu, lokalizují se do nich, (ne)procházejí jimi a mohou membrány i rozpouštět. Hlavním nástrojem pro vyhodnocování interakcí je **profil volné energie** molekul na membránách, který nám nejen dává informace o afinitě látky k membráně, ale je taky důležitý pro zjišťování pozice energetického minima a penetrační bariéry.

Biologické databáze jsou obecně online knihovny, které obsahují informace o živých organismech získané nejen z vědeckých experimentů, ale také například analýzou publikované literatury, nebo využitím data-miningu za účelem spojení podobných databází. Takové databáze jsou nezbytným výzkumným nástrojem, neboť poskytují pohodlný a spolehlivý přístup k přechozím znalostem.

V oblastech interakcí dnes existuje několik databází; nejvíce se jich zaměřuje na interakce **protein/protein** a interakce **protein/DNA (RNA)**. V současnosti nicméně **neexistuje** žádná publikovaná databáze interakcí malých molekul s biologickými membránami i přesto, že jde o významnou část biologického osudu jednotlivých chemikálií.

**MolMeDB (Molecules on Membranes DataBase)** je aktuálně vyvíjená relační databáze (MySQL) s webovým rozhraním, která bude v budoucnu shromažďovat tisíce vypočtených profilů volných energií společně s experimentálními daty na základě prověřených studií. Databáze dále obsahuje podrobné informace o použitých metodách, membránách a molekulách, a pro pohodlný přístup k datům je k dispozici vyhledávač. Detaily interakcí je možné nejen prohlížet, ale také mezi sebou porovnávat díky sekci **Comparator**. Informace o molekulách a jejich interakcích jsou rozšířeny o 2D a 3D vizualizaci molekul a o interaktivní grafy zobrazující profily volných energií.



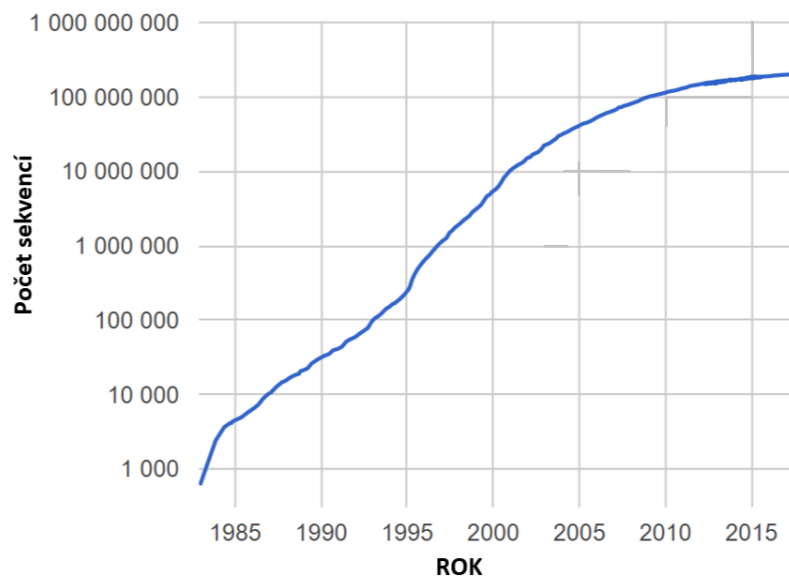
## 2 Současný stav řešené problematiky

### 2.1 Biologické databáze

V současné době je kladen velký důraz na systémy, které umožňují ukládání velkého množství dat, následné vybírání z těchto dat a finální zobrazení v uživatelsky čitelné formě. Tuto úlohu zprostředkovávají databázové systémy už od 50. let 20. století. Největší rozšíření databázových systémů do odvětví biologie přišlo v 70. letech s rozvojem bioinformatiky a nových bioinformatických metod, které měly za následek generování velkého množství biologických dat a s tím potřebnou nutnost tyto data nějakým způsobem ukládat a organizovat. Typickým příkladem je databáze **GenBank**, která vznikla (1982) jako odpověď na množství dat generované DNA sekvenčními technikami, které dnes rapidně roste kvůli snižování nákladů na sekvenování (Helmy *et al.*, 2016).

Biologické databáze jsou knihovny biologických dat, které byly získány z vědeckých experimentů, publikované literatury nebo různými počítačovými analýzami. Téměř všechny biologické databáze jsou dnes dostupné v online podobě a volně přístupné pro širokou veřejnost a obsahují strukturované informace o živých organismech z různých vědeckých oblastí; zejména se jedná o oblast genomiky a proteomiky. Tyto databáze obsahují informace o funkcích, struktuře a lokalizaci genů, proteinů a dalších látek, a tudíž jsou důležitým výzkumným nástrojem, neboť poskytují pohodlný a spolehlivý přístup k předchozím znalostem, který je nezbytný pro plánování a vyhodnocování budoucích výzkumů. Zvyšující se důležitost biologických databází má za následek neustálý nárůst jak jejich počtu, tak i množství dat v nich uložených (Stein, 2003).

Mezi nejobjemnější a rovněž nejstarší biologické databáze patří sekvenční databáze. Jedná se o databáze, které obsahují sbírku DNA nebo proteinových sekvencí. Největší a nejznámější jsou databáze **GenBank** a **ENA** (European Nucleotide Archive). GenBank obsahoval v roce 2007 přes 73 milionů genových sekvencí a dnes (03/2018) obsahuje více než 200 milionů sekvencí (přibližně 200 miliard bází) což odpovídá více než 200 petabajtům dat. Počet párů bází v databázi GenBank se už od roku 1982 každých 18 měsíců zdvojnásobuje (Obr. 1).



Obrázek 1: Grafické vyjádření nárůstu počtu genových sekvencí v databázi GenBank - převzato z oficiálních statistik GenBank (2018). *Povšimněte si logaritmického měřítka.*

## 2.2 Návrh biologické databáze

Na internetu existuje více než 2000 veřejně přístupných biologických databází. Některé z těchto databází jsou sofistikované a udržované profesionálními týmy, ale existují i databáze, které jsou nepřehledné, neobsahují kvalitní data a často se po nějaké době stávají zastaralými. Vytvoření vysoce kvalitní databáze představuje závazek, který vyžaduje sběr přesných a kvalitních dat a pravidelné kontroly obsahu a funkcionalit databáze (Wilkinson *et al.*, 2016).

Nejčastějším důvodem pro tvorbu nové databáze je poptávka vědecké obce po datech, která nejsou dostupná v elektronické podobě nikde jinde (**primární databáze**). Dalším důvodem může být snaha vytvořit databázi, která sbírá data stejného nebo podobného typu z různých databází, za účelem jejich nashromáždění na jednom místě kvůli lepším možnostem porovnávání (**sekundární databáze**). V každém případě má obsah dat a vytvořené funkce největší význam, pokud jsou originální a užitečné.

Před každým návrhem databáze je nutné si ověřit, zda už podobná databáze s danou funkcionalitou neexistuje. Pro ověření je možné použít několik internetových zdrojů, které se snaží udržovat biologické databáze organizované v kategoriích – například “Nucleic Acid Research Online Molecular Biology Database Collection”, která je dostupná na adrese <http://oxfordjournals.org/nar/database/c/> a v současné době obsahuje přes 1700 databází (Hemly *et al.*, 2016).

Ať už jde o tvorbu primární nebo sekundární databáze, je nutné provádět kontrolu kvality dat. Primární databáze jsou zodpovědné za kvalitu všech vstupních dat. Sekundární databáze musí zajišťovat, že nebude docházet ke vzniku chyb při procesech, které souvisí s jejich plněním (tzv. datové integrační procesy).

Při definování rozsahu databáze a použitých technologií je zásadní mít na paměti, jaký typ uživatelů bude databázi využívat. Uživatelé se dají rozdělit na 3 skupiny:

- **(a)** Biologové a další výzkumníci, kteří využívají manuálního dotazování na webu,
- **(b)** Výpočetní biologové, kteří se zajímají o stahování většího množství dat a dotazovací systémy,
- **(c)** Vývojáři software se znalostí API.

Většina uživatelů biologických databází má omezené schopnosti programování a tudíž je často nezbytné, vytvořit pro práci s daty přehledné uživatelské rozhraní (Schultheiss *et al.*, 2011).

Většina biologických databází je veřejně dostupná prostřednictvím webových stránek; ty by měly mít intuitivní grafický design, podporovat standardy webových prohlížečů a operačních systémů, měly by mít propracované grafické uživatelské prostředí (GUI) a v neposlední řadě by měly být bezpečné. V dnešní době existuje spousta populárních technologií, které velice usnadňují vytvoření takové webové stránky; neznámější jsou určitě PHP, HTML5, CSS3 a JavaScript. Navíc pro rychlou tvorbu příjemného grafického prostředí dnes existuje řada volně stažitelných nástrojů (např. Bootstrap), jež obsahují návrhářské šablony založené na HTML a CSS.

S tvorbou webového rozhraní prakticky vždy souvisí práce s JavaScriptem. Jedná se v dnešní době o nejrozšířenější způsob, jak webovou stránku udělat „živou“ a interaktivní, což může velice usnadnit a zpříjemnit uživateli práci s daty. Na druhou stranu, každou funkcionalitu je potřeba náležitě otestovat, hlavně z pohledu nezkušeného uživatele, a ošetřit chování programu před vznikem nejrůznějších chyb, které by mohly v některých případech způsobit i ztrátu uložených dat.

## 2.3 Reprezentace molekulárních struktur v počítači

Zápis a uchování molekul v počítačové podobě usnadňuje jejich analýzu a porovnávání, a tudíž je nezbytným nástrojem pro mnohé výzkumné týmy. Struktury molekul lze v počítači ukládat různými způsoby, z nichž nejjednodušší je forma obrázku. Ta nám ale kromě vizualizace neposkytuje žádné bližší informace o dané molekule. Problematika ukládání struktury molekul v počítačové reprezentaci se začala ve větším měřítku objevovat již od 80. let 20. století, následkem čehož vzniklo několik forem zápisu.

Téměř všechny počítačové zápisy chemických struktur vycházejí z teorie grafů; každý atom je reprezentován vrcholem, a tudíž celá struktura je zapsána pouze jako množina vrcholů a hran. Nejčastější metoda pro ukládání grafů do počítačové podoby je tzv. **linearizovaný zápis**. Hlavní myšlenkou linearizovaného zápisu je ukládání pouze kostry grafu. Nespornou výhodou je, že kostra je acyklický graf a ten umíme zapisovat jednoduše do řádky textu. Tento typ zápisu ale přinesl i určité problémy, které každá metoda řeší podle sebe. Bylo potřeba vyřešit, jak zapisovat cyklické vazby a vytvořit kódování pro jednotlivé typy vazeb, atomy, funkční skupiny a stereochemii. Mezi nejčastěji používané metody zápisu patří **SMILES** a **InChI** zápis.

### 2.3.1 Zápis molekul ve formě SMILES

SMILES zápis (Simplified Molecular Input Line Entry Specification) vznikl už v 80. letech, ale i přesto je jednou z často používaných metod, jak jednoduše ukládat chemické struktury v počítačové reprezentaci.

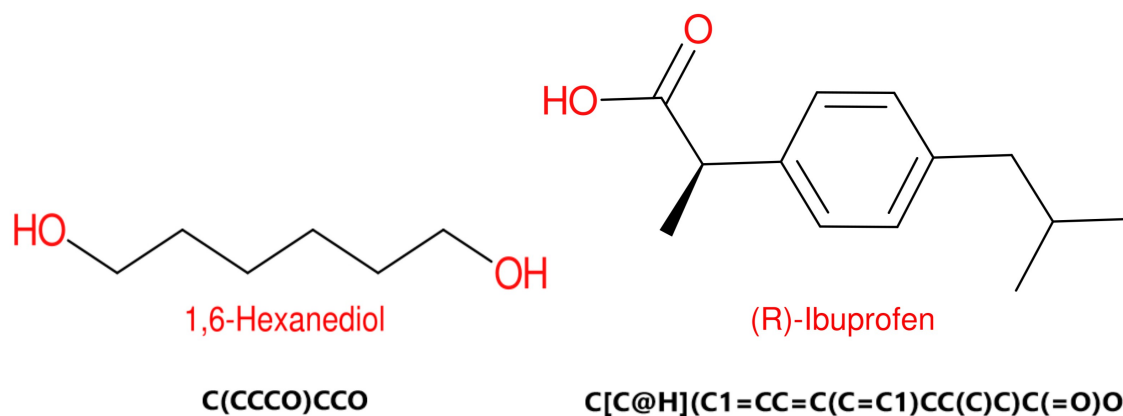
Atomy se zapisují pomocí jejich symbolů do hranatých závorek včetně počtu vodíků, specifikace náboje a izotopu (např. [CH<sub>4</sub>], [Zn<sup>++</sup>], [14CH<sub>3</sub>-]). Prvky, které se hojně vyskytují v organické chemii (tzv. organická podmnožina) se nemusí zapisovat do hranatých závorek a ani se u nich nemusí uvádět počet vodíků explicitně; mají implicitní vaznost. Jedná se o prvky B, C, N, O, P, S, F, Cl, Br a I. Jednotlivé vazby jsou reprezentovány následovně: *jednoduchá vazba* „-“, *dvojná vazba* „=“, *trojná vazba* „#“, *čtverná vazba* „\$“, *aromatická* „:“ a *nulová* „“. Další dohodnuté konvence:

- **a)** Pokud není uvedena žádná vazba, automaticky se předpokládá vazba jednoduchá.
- **b)** Kruhy se zapisují pomocí čísel za symbolem atomu (odpovídající čísla se spojí vazbou). Čísla lze recyklovat; lze využívat stejná čísla bezprostředně za sebou. Čísla

se používají pouze jednociferná. Dvouciferné číslo (C12) znamená dva kruhy s čísly 1 a 2.

- **c)** Aromatická vazba mezi atomy z organické podmnožiny je implikována, pokud jsou symboly zapsány malými písmeny (např. benzen: c1ccccc1), nebo pomocí střídání jednoduché a dvojné vazby (např. Benzen: C1=CC=CC=C1).
- **d)** Nulová vazba se používá například pro vyjádření solí nebo komplexů. Například [Na+].[Cl-] místo [Na][Cl], kde by chybějící symbol vazby značil kovalentní vazbu, což je nesmysl.
- **e)** Větvení se zapisuje pomocí jednoduchých závorek a lze různé větvení vnořovat do sebe.
- **f)** Pro vyjádření isomerie na dvojných vazbách se využívají odlišené symboly pro jednoduché vazby – „\“ značí směr dolů a „/“ směr nahoru. Např. C\C=C\C je zápis pro trans-but-2-en.
- **g)** Stereochemie na sp<sup>3</sup> uhlíku je zapisována průběžně při vytváření zápisu. Zapisujeme strukturu (graf) postupně a na chirálním centru zaznamenáme pořadí a smysl rotace substituentů tak, jak je „vidíme“. Rotaci označujeme symboly @ (proti směru hodinových ručiček) a @@ (po směru hodinových ručiček).

Zápis ve formě SMILES je silným nástrojem, jak struktury jednotlivých molekul zapisovat ve formě textu na řádky (Obr. 2), což je z hlediska počítačového zpracování a ukládání velice výhodné. Na druhou stranu má zápis v takové formě jednu velkou nevýhodu – každou strukturu lze zapsat více způsoby; to zcela znemožňuje použití SMILES například jako identifikátoru (Jirát a Svozil, 2017).



Obrázek 2: Struktura molekul 1,6-Hexanediolu a (R)-Ibuprofenu a k nim odpovídající SMILES zápisy.

### 2.3.2 Další možnosti počítačového zápisu molekul

Mezi další metody linearizovaného zápisu patří InChI (International Chemical Identifier), InChIKey a SMARTS (*SMiles ARbitrary Target Specification*). **InChI** je relativně mladý formát (2005), který vznikl ve spolupráci organizací IUPAC a NIST za účelem vytvoření standardního a dobře čitelného formátu pro kódování informací o molekulách. Na stránkách *The IUPAC International Chemical Identifier (InChI)* (2018) je uvedeno, že každý InChI zápis začíná klíčovým slovem „InChI=“ následovaný číslem verze, písmenem S (standardní InChI) a jednotlivými vrstvami (maximálně 6), které obsahují specifické informace a jsou odděleny znakem „/“ (Obr. 3). Nevýhodou InChI zápisu je jeho proměnlivá délka (může být velmi dlouhý). Proto byl vytvořen formát **InChIKey**, který je pouze hašovanou verzí původního InChI (použitím *SHA-256* algoritmu). Tento zápis má konstantní délku, ale pro uživatele ztrácí jakoukoliv čitelnost.

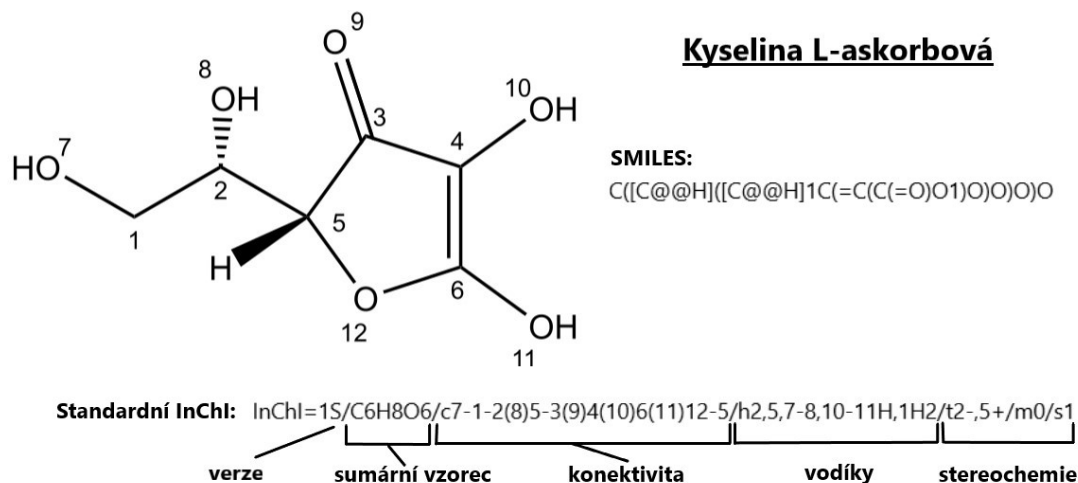
**SMARTS** byl vyvinut pro potřeby zápisu substruktur. Jedná se pouze o rozšíření zápisu SMILES; může obsahovat různé rozšiřující symboly. Navíc pokud bychom například vyhledávali molekulu fenolu ([OH]c1ccccc1) podle SMILES zápisu, tak najdeme pouze molekuly fenolu. Zatímco pokud bychom reprezentovali fenol jako SMARTS zápis, tak bychom našli všechny struktury, které obsahují molekulu fenolu kdekoli ve své struktuře (Jirát a Svozil, 2017).

Častou formou zápisu struktury molekul je zápis pozic atomů do textových souborů; nejčastěji se užívají formáty **MOL** a **PDB**.

**PDB** formát byl vytvořen tvůrci databáze PDB za účelem ukládání trojrozměrných proteinových struktur pomocí zápisu pozic jednotlivých atomů molekul v prostoru.

V hlavičce souboru jsou obsaženy základní údaje o molekule a autorovi. Samotné údaje o struktuře jsou zapsány po řádcích (jednotlivé atomy) a každý řádek je pevně rozdělen do několika sloupců, které obsahují po řadě tato data: číslo atomu, chemickou značku, typ aminokyselinového zbytku a jeho pořadí, XYZ souřadnice atomu v molekule, počet možných konformací a teplotní flexibilitu (Protein Data Bank Contents Guide, 2008).

**MOL** formát (**MDL Molfile**) má za úkol udržovat informace o atomech, vazbách a jejich souřadnicích v molekule.



Obrázek 3: Struktura kyseliny L-askorbové a počítačová reprezentace její struktury ve formě SMILES a InChI zápisu.

## 2.4 Databáze interakcí

Vzájemné interakce jednotlivých molekul hrají v organismech důležitou roli při mnoha biologických procesech. Jedny z nejdůležitějších interakcí v lidském těle jsou interakce nukleových kyselin s proteiny, které, mimo jiné, zajišťují regulaci genové exprese a tvorbu složitějších komplexů, bez kterých by se buňka neobešla.

V dnešní době existuje spousta biologických databází, které se zabývají různými interakcemi. Tyto databáze se liší podle toho, mezi čím je daná interakce studována:

- **protein/protein** - Nejvíce studovaná oblast interakcí. Mezi největší databáze interakcí patří BioGRID (The Biological General Repository for Interaction Datasets) databáze (Chatr-aryamontri *et al.*, 2017), která obsahuje informace o protein/protein interakcích, genetických a chemických interakcích a post-translačních modifikacích. V dnešní době (04/2018) obsahuje záznamy o více jak 1,5 milionu interakcích pro více než 30 modelových organismů. Další příklad databází protein/protein interakcí jsou EMPIAR (Iudin *et al.*, 2016), InterPRO (Apweiler *et al.*, 2001), STRING (Szklarczyk *et al.*, 2017), IntAct (Kerrien *et al.*, 2011) a iHOP (Fernández *et al.*, 2007).
- **protein/nukleová kyselina** - Slouží (nejen) ke studiu složitějších komplexů v buňce - příkladem je ribozom, který je tvořen kombinací rRNA a proteinů. Neznámější databáze těchto interakcí jsou ProTherm a ProNIT (Kumar *et al.*, 2006).
- **protein/membrána** - Databáze obsahují informace o strukturách, uložení a interakcích (trans)membránových a periferních proteinů - OPM (Orientations of Proteins in Membranes) (Lomize *et al.*, 2011), PDBTM (Protein Data Bank of Transmembrane Proteins) (Kozma *et al.*, 2012).
- **protein/malá molekula** - Jednou z databází, která se zabývá těmito interakcemi je databáze BindingDB (Gilson *et al.*, 2015).

Počet databází interakcí neustále narůstá, i když se převážně jedná o interakce protein-protein. Nicméně, v současnosti **neexistuje** žádná publikovaná databáze interakcí malých molekul s biologickými membránami i přes to, že jde o významnou část biologického osudu jednotlivých chemikálií.

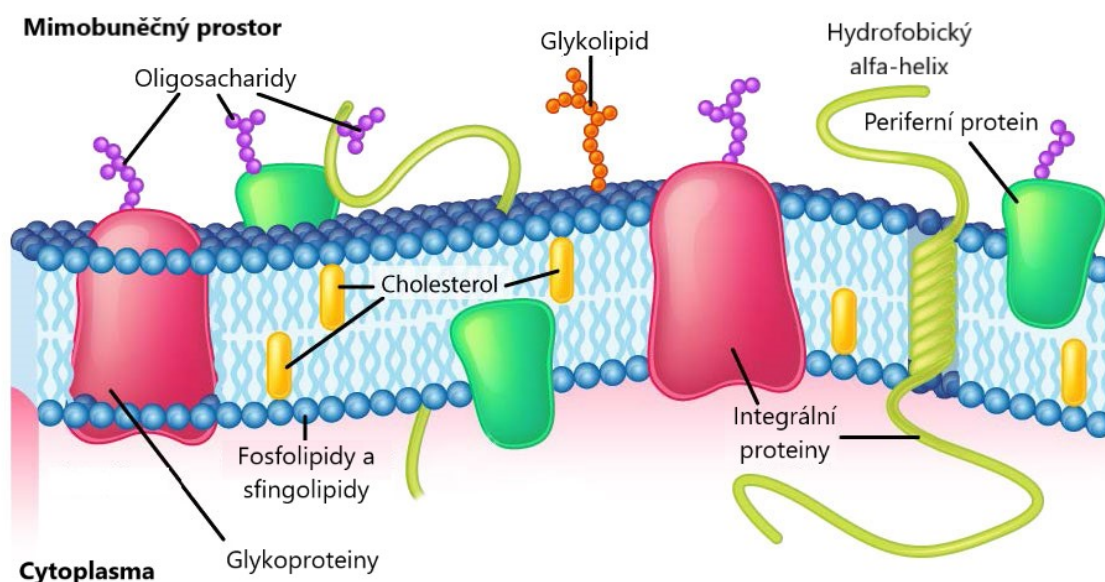


## 2.5 Interakce molekul s biologickými membránami

### 2.5.1 Biologické membrány

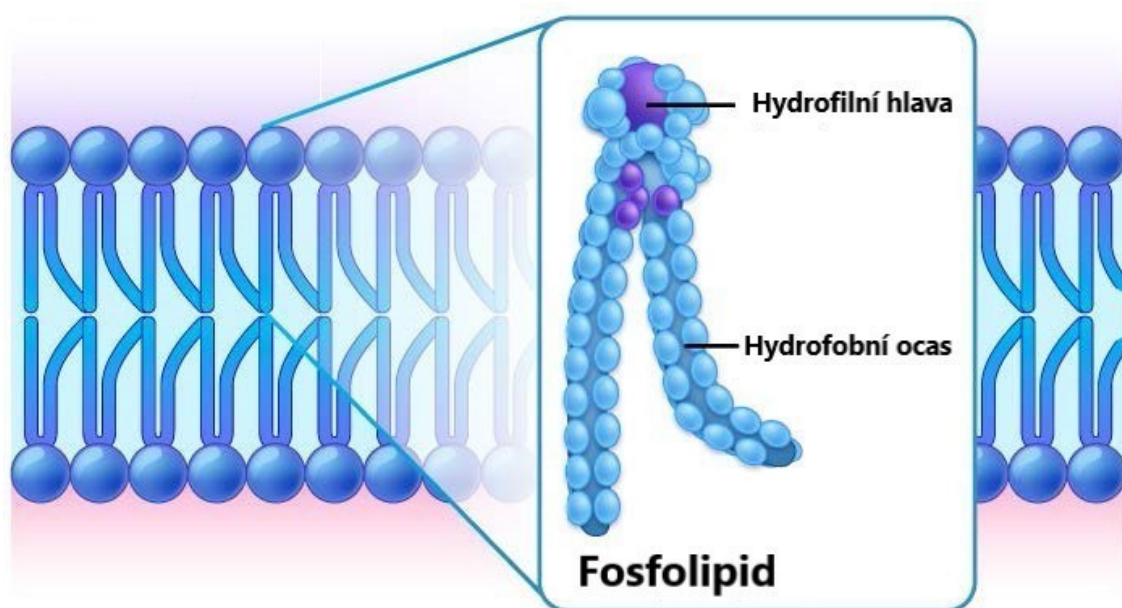
Membrány v lidském těle (a jiných organismech) umožňují život jako takový. Slouží nejen k formování buněk, ale také k oddělení prostředí s různými fyzikálně-chemickými vlastnostmi. Co je důležité, oddělují vnitřní prostředí buňky od mimobuněčného prostoru nebo od vnitřních prostředí buněčných organel. Dále umožňují průchod iontů a s tím spojenou tvorbu iontových gradientů, čímž buňkám zajišťují jednu z nejdůležitějších činností – tvorbu energie. Kromě toho řídí vzájemnou komunikaci mezi buňkami přijímáním, odesláním a zpracováním informací ve formě fyzikálních a chemických signálů (Lúcio *et al.*, 2010).

Membrány jsou 3 - 12 nm široké vrstvy složené z proteinů a lipidů (Obr. 4), přičemž hmotnostní zastoupení proteinů se pohybuje mezi 25 a 75 procenty. Zbytek hmoty tvoří hlavně lipidy, které jsou nejčastěji uspořádány do tzv. **lipidové dvojvrstvy** (Obr. 5). Membránové proteiny hrají klíčovou roli v biologických membránách. Pomáhají udržovat strukturální integritu a zajišťují aktivní transport látek přes membrány (Chen a Moy, 2000; Hine, 1999).

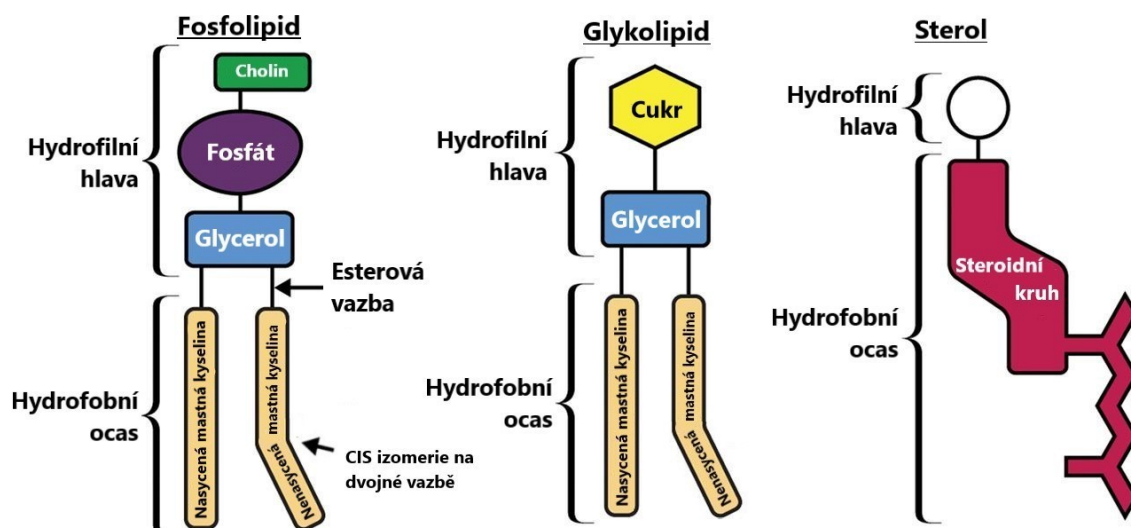


Obrázek 4: Model fluidní mozaiky lipidové dvojvrstvy a její molekularní organizace - vypracováno podle Liber Texts: Basic Concepts in Membranes (2018).

V membránách můžeme nalézt tři typy lipidů, jmenovitě fosfolipidy, glykolipidy a steroly. Všechny membránové lipidy jsou amfifatické, jinými slovy obsahují **hydrofilní** (hlava) i **hydrofobní** (ocas) část (Obr. 6). Fosfolipidy se skládají ze dvou mastných kyselin (ocas) připojených ke glycerolu a fosfátové skupině (hlava). Příkladem takového fosfolipidu je fosfatidylcholin (PC), který má molekulu cholinu připojenou k fosfátové skupině a má asi 50% zastoupení mezi fosfolipidy eukaryotických buněk (Van Meer *et al.*, 2008). Dalšími hojně zastoupenými fosfolipidy jsou fosfatidylserin (PS) nebo fosfatidylethanolamin (PE). Glykolipidy obsahují buď glycerol nebo sfingosin a vždy obsahují cukr místo fosfátu. Steroly jsou nepřítomné u většiny bakteriálních membrán, ale jsou důležitou součástí rostlinných (hlavně stigmasterol) a živočišných (hlavně cholesterol) membrán. Struktura sterolů se velice liší od struktury fosfolipidů a glykolipidů. Konkrétně cholesterol obsahuje hydroxylovou skupinu (hydrofilní hlava), steroidní skelet a krátký alifatický řetězec (hydrofobní ocas) (Lúcio *et al.*, 2010).



Obrázek 5: Struktura lipidové dvojvrstvy tvořené fosfolipidy - vypracováno podle Liber Texts: Basic Concepts in Membranes (2018).



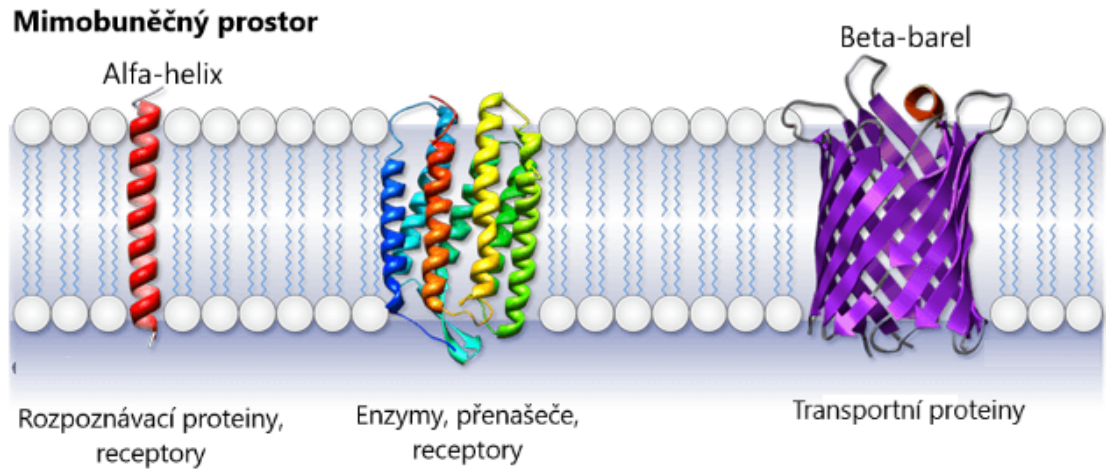
Obrázek 6: Schématické znázornění struktury fosfolipidu (konkrétně fosfatidylcholin), glykolipidu a sterolu - vypracováno podle Creative Biostructure (2018) .

## 2.5.2 Struktura a funkce membránových proteinů

Kromě lipidů se v membránách nachází také membránové proteiny. Membránové proteiny se liší strukturou (Obr. 7) i funkcí od rozpustných proteinů. Mohou být tvořeny alfa helixy nebo beta barely. Beta barely jsou velké beta listy, tvořené obvykle anti-paraletními řetězci, jenž se stáčejí do uzavřené formy, ve které je první řetězec spojen s posledním pomocí vodíkové vazby. Alfa-helikální membránové proteiny obsahují nejméně jeden alfa-helikální úsek s délkou kolem dvaceti reziduí. Pokud takové proteiny prochází skrz membránu více než jednou, obsahují více hydrofobních úseků (Cooper, 2000).

Rozhodující skupinou membránových proteinů jsou ty, jenž jsou zapojeny v aktivním nebo pasivním transportu látek přes buněčnou membránu, nebo přes membrány buněčných organel. Molekuly dokáží procházet membránami několika rozdílnými způsoby v závislosti na jejich koncentraci, velikosti a náboji. Některé molekuly, včetně vody, mohou procházet skrze membrány **pasivně** bez pomoci (tzv. difuzí), nicméně větší molekuly nebo nabitě molekuly nemohou. Nabité molekuly (například ionty), mohou procházet skrze (iontové) kanálky za pomoci elektrochemických gradientů. Pohyb bývá popisován jako přesun iontů z oblasti s vysokou koncentrací do oblasti s nízkou koncentrací. Při takovém pohybu jsou potřeba proteiny tvořící kanálky nebo samotné přenašečové proteiny, nicméně nedochází ke spotřebě energie. Proteinové kanály umožňují přístup molekul na základě náboje a velikosti, zatímco přenašečové proteiny interagují s přenašenou látkou, ta se váže do aktivního místa proteinu, mění jeho konformaci a je umožněn přesun přes membránu.

**Aktivní transport** může sloužit k přenášení molekul proti elektrochemickému gradientu za spotřeby energie získané z ATP, světla, nebo protipohybem jiné molekuly nebo iontu ve stejném transportním systému. Obvykle je zprostředkováván integrálními membránovými proteiny (Alberts *et al.*, 2006; Lodish *et al.*, 2000).



Obrázek 7: Uložení membránových proteinů v lipidové dvojvrstvě.

### 2.5.3 Malé molekuly

V oblastech molekulární biologie a farmakologie, jsou obvykle za malé molekuly považovány organické sloučeniny s hmotností do 900 Da a velikostí v řádech nanometrů. Do této skupiny patří většina léčiv, lipidy, monosacharidy, sekundární posli (second messengers) a další přírodní produkty a metabolity. Malé molekuly jsou zajímavé hlavně proto, že mohou procházet buněčnými membránami do buněk, kde mohou dále interagovat s dalšími molekulami a tím ovlivňovat jejich funkce.

Malé molekuly se *in vitro* široce využívají k regulaci mnoha biologických pochodů v buňce a část z nich se používá za tímto účelem i v klinických podmínkách. Mezi nejznámější patří ibuprofen, který se používá na tlumení bolestí, zánětů a horeček. Malé molekuly ale nacházejí uplatnění i v diferenciaci, produkci a udržování pluripotentních kmenových buněk. Mohou být použity *in vitro* k usnadnění produkce a rozšíření kmenových buněk, nebo k odstranění zbytkových pluripotentních kmenových buněk (Atala, 2016). Kombinace rozsáhlého popisu léčiv s jejich cílením je popsána v databázi **DrugBank**, která aktuálně obsahuje více jak 10 000 látek. V této databázi jsou rozdělena léčiva do několika skupin z nichž nejrozsáhlejší je právě skupina malých molekul (oficiální statistiky DrugBank, 2018). Tudíž hlavně v oblasti klinických studií má výzkum malých molekul velký význam.

Průchod molekul membránami je jedním z klíčových dějů při jejich vstřebávání v těle. Detailní znalost jejich průniku přes membrány by nám dala možnost takový transport předvídat a tím předvídat i samotné vstřebávání v těle.

Malé molekuly interagují s membránami mnoha způsoby – mají k nim různou afinitu, lokalizují se do nich, (ne)procházejí jimi a mohou membrány i rozpouštět. Hlavním nástrojem pro vyhodnocování interakcí je profil volné energie molekul na membránách, který nám dává nejen informace o afinitě látky k membráně, ale je také důležitý pro zjišťování pozice energetického minima, tedy lokalizaci látky v membráně, a penetrační bariéry.

### 2.5.4 Interakce malých molekul s biologickými membránami

I přes znalost důležitosti membránových proteinů v regulaci buněčných procesů se ve výzkumu a vývoji léčiv dříve uvažovala pouze interakce mezi ligandem (léčivo) a jeho cílem (např. enzym) (Seydel *et al.*, 1994). Podle toho, byla většina klinických léčiv vyvíjena na základě přímé interakce s proteiny a stále častěji docházelo k návrhu nových ligandů

pro klinické využití tak, aby modulovaly aktivitu membránových proteinů. Lipidová část byla často považována za inertní, strukturální prvek membrány, který podporuje proteiny a působí jako pasivní okolní prostředí. Tento pasivní pohled na membrány se však s postupem času měnil, a to především díky objevům, které prokazovaly, že membránové lipidy hrají ústřední roli při regulaci mnoha buněčných procesů, a tudíž nejsou pouze strukturální páteří membrán a zdrojem energie (Escriba *et al.*, 2006). Díky novým možnostem studia interakcí na molekulární úrovni je možné porozumět cílení léčiv, permeaci a průniku skrze buněčné membrány, interakci s membránovými receptory a využití liposomů při metodách řízeného uvolňování léčiv (Mouritsen *et al.*, 2001).

Ke studiu vlivu struktury lipidové dvojvrstvy na funkci proteinů, které jsou ať už integrovány přímo do membrány, nebo jsou připojeny na povrchu, je nutné vytvořit vhodný modelový systém. Modelové membrány s pouze několika komponentami mohou zajistit hodnotný vhled do vlastností více komplexních biologických membrán. Takové modely mohou být vytvořeny v různých formách: micely, liposomy, lipidové dvojvrstvy a lipidové jednovrstvy (Lúcio *et al.*, 2010). Na druhou stranu, schopnost malých molekul prostupovat biomembrány záleží z velké části na jejich lipofilitě, která je nejčastěji vyjadřována v podobě tzv. **partičního koeficientu**. Tento koeficient je definován jako poměr koncentrací látky ve dvou vzájemně nemísitelných fázích v rovnovážném stavu. Jinými slovy nám udává podíl rozpustnosti látky v jedné a druhé fázi. Ke studiu lipofility se nejčastěji používá partiční koeficient **oktanol/voda** (oktanol vykazuje podobné fyzikálně chemické vlastnosti jako rostlinné a živočišné tkáně). Jednotka koeficientu je bezrozměrná a zapisuje se obvykle ve formě logaritmu:  $\log K_{oktanol/voda}$  (zkráceně  $\log K_{o/w}$ , popř.  $\log P$ ).

$$K_{o/w} = c_o/c_w$$

$K_{o/w}$  je partiční koeficient oktanol/voda,  $c_o$  je koncentrace v oktanolu a  $c_w$  je koncentrace ve vodné fázi (Šimková, 2017).

K odhadu penetračních vlastností a partičního koeficientu malých molekul v závislosti na koncentraci, teplotě, tlaku, ad. lze použít experimentálních technik jako jsou ultracentrifugace, dialýza nebo mikroextrakce na pevné fázi. Tyto metody nám bohužel nedávají podrobný pohled na interakci na molekulární úrovni. Pro takový vhled můžeme využít predikčních metod, které v poslední době získávají stále větší uplatnění. Mohou být využity i v případech, kdy není možné provést experimentální měření (Paloncýová *et al.*, 2014).

Pro predikci chování molekul na membránách je nejprve nutné navrhnout vhodný **model** dané membrány. V aktuální verzi databáze **MolMeDB** jsou napočítány hodnoty interakcí pro modely membrány endoplasmatického retikula, na které se metabolizuje velká část léčiv, a svrchní části pokožky. Membrána endoplasmatického retikula obsahuje asi ze 40 % fosfatidylcholin a za její modely byly zvoleny membrány **DOPC** - dioleoylfosfatidylcholin a **DMPC** - dimyristoylfosfatidylcholin. Svrchní část pokožky (*epidermis*) a její nejsvrchnější vrstva *stratum corneum* (**SC**) je tvořena převážně z ceramidů, mastných kyselin a cholesterolu a znesnadňuje průchod molekul přes pokožku. Jako model byla zvolena membrána **SC MIX**, která je tvořena směsí lignoceroyl sfingosinu, cholesterolu a lignocerové kyseliny v poměru 1:1:1 (Šimková, 2017).

Pro predikci chování molekul na simulovaných membránách mohou být použity programy založené na teorii COSMO-RS (**CO**nductor like **S**creening **MO**del for **R**eal **S**olvents). Jedná se o termodynamickou metodu vytvořenou za účelem predikce elektrostatického potenciálu v kapalinách pomocí kvantové chemie. Jedním z použitých programů je COSMOmic (Klamt *et al.*, 2008).

### 2.5.5 COSMOmic

Jedná se o rozšiřující program, používající metodu COSMO-RS, pro anisotropický systém jako jsou lipidové membrány a micelární systémy, ve kterých dokáže předpovědět distribuci molekul i přes to, že se jedná o nehomogenní prostředí (Klamt *et al.*, 2008). Používá se pro výpočet profilů volné energie a pro predikci partičního koeficientu membrána/voda. Metody pro jeho výpočet obvykle vycházejí z výpočtu rozdělovacího koeficientu oktanol/voda, avšak COSMOmic umožňuje výpočet tohoto koeficientu přímo z profilu volné energie molekuly (Eckert *et al.*, 2014).

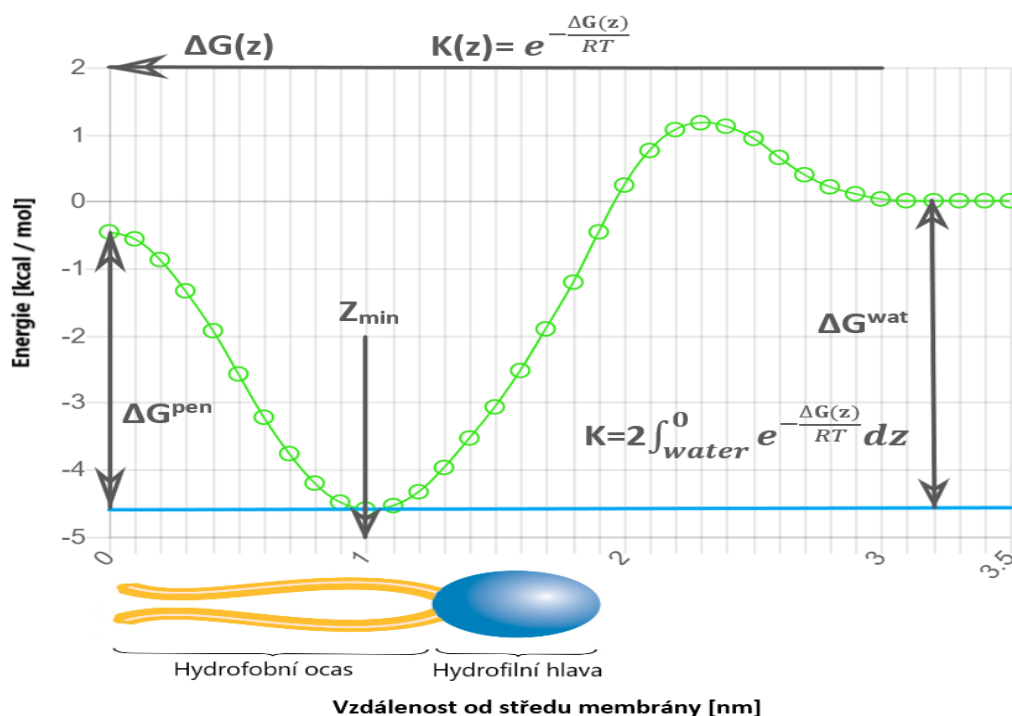
Pro predikci hodnot interakcí pomocí COSMOmic je nejprve nutné mít vhodný model membrány (DOPC, SC MIX, ad.). Základním předpokladem přístupu COSMOmic je uvažovat lipidovou dvojvrstvu nebo micelární systém jako vrstvenou kapalinu (50 vrstev). Centrální část molekuly se umístí postupně do každé vrstvy a v každé této vrstvě se uvažuje 162 různých směrových reprezentací molekuly. Pro každou orientaci v každé vrstvě COSMOmic počítá tzv.  $\sigma$ -profil (histogram hustoty náboje) z nichž je potom sestaven výsledný  $\sigma$ -profil. Výpočet je opět založen na geometrické orientaci molekuly a na identifikaci, jaká část molekuly je přítomna v jednotlivých vrstvách membrány. Z těchto



informací je následně vypočítána celková solvatační energie molekuly v každé vrstvě a je vytvořen **profil volné energie** (Klamt *et al.*, 2008).

### 2.5.6 Profil volné energie podél normály membrány

Profil volné energie je stěžejní nástroj pro predikci interakce molekuly s membránou. Membrána je heterogenní prostředí a profil volné energie k ní nám říká, jak se mění hodnoty interakcí v závislosti na poloze molekuly v membráně. Na obrázku 8 je znázorněn profil volné energie molekuly 1,3-decanediolu podél normály membrány DMPC s popisem veličin znázorňujících data, které lze z profilu volné energie získat.



Obrázek 8: Profil volné energie molekuly 1,3-decanediolu podél normály membrány DMPC (dimyristoylfosfatidylcholin) se znázorněním velikosti lipidové vrstvy dané membrány při zachovaném měřítku.  $G(z)$  znázorňuje hodnotu volné energie v závislosti na vzdálenosti molekuly od středu membrány,  $Z_{min}$  je pozice minima,  $G^{pen}$  je penetrační bariéra (hodnota energie potřebná k penetraci molekuly do středu membrány),  $G^{wat}$  udává energetickou bariéru lipid/voda (afinita molekuly k membráně).



## 3 Praktická část

### 3.1 Použité technologie

#### 3.1.1 HTML5 a CSS

HTML5 (zkratka pro HyperText Markup Language) je nástupcem HTML 4.01, který přišel s mnoha vylepšeními a stal se nedílnou součástí každého moderního webu. Stejně jako jeho předchůdci je HTML5 navrženo jako multiplatformní; to znamená, že nezáleží na tom, jestli uživatel používá Mac OS X, Windows, Linux, nebo jiný operační systém. Jediné, co je třeba ke správné funkčnosti, je moderní webový prohlížeč. Za takový prohlížeč jsou považovány téměř všechny (aktualizované) všeobecně rozšířené prohlížeče, včetně těch mobilních. HTML5 je značkovací jazyk sloužící ke strukturálnímu rozvržení stránky. Základem takového jazyka jsou značky (tagy), které definují jednotlivé části webu a dávají mu výslednou strukturu (Pilgrim, 2010).

CSS (zkratka pro Cascading Style Sheet) je jazyk určující vzhled celé webové stránky. Tento jazyk by neměl smysl bez samotného HTML. HTML udává stránce její strukturu a CSS jednotlivým značkám určuje styl (Sklyr, 2001).

#### 3.1.2 PHP 7

PHP (zkratka pro Personal Home Page) je programovací jazyk pracující na straně serveru, který se využívá pro tvorbu dynamických webových stránek a interaktivních webových aplikací. Často se o PHP uvažuje jako o hypertextovém preprocesoru. Vychází to z faktu, že na vstupu je PHP kód, ale na výstupu dostaneme HTML. To znamená, že uživatel nemusí instalovat žádný software k zobrazení webů generovaných PHP.

V porovnání s HTML jsou PHP stránky dynamicky generovány. V podstatě to znamená to, že namísto velké spousty HTML souborů pro zobrazování jednotlivých stránek vytvoříme pouze (pro stejné typy souborů z hlediska struktury) jeden templátový soubor a ten na základě požadavků uživatele plníme potřebnými daty. Takový přístup má nesporné výhody, ať už jde o potřebu editace jednoho typu záznamu (například zobrazení detailu nějakého produktu), nebo o maximální úsporu paměti a eliminaci redundancí kódu (Gutmans *et al.*, 2015).

### 3.1.3 JavaScript

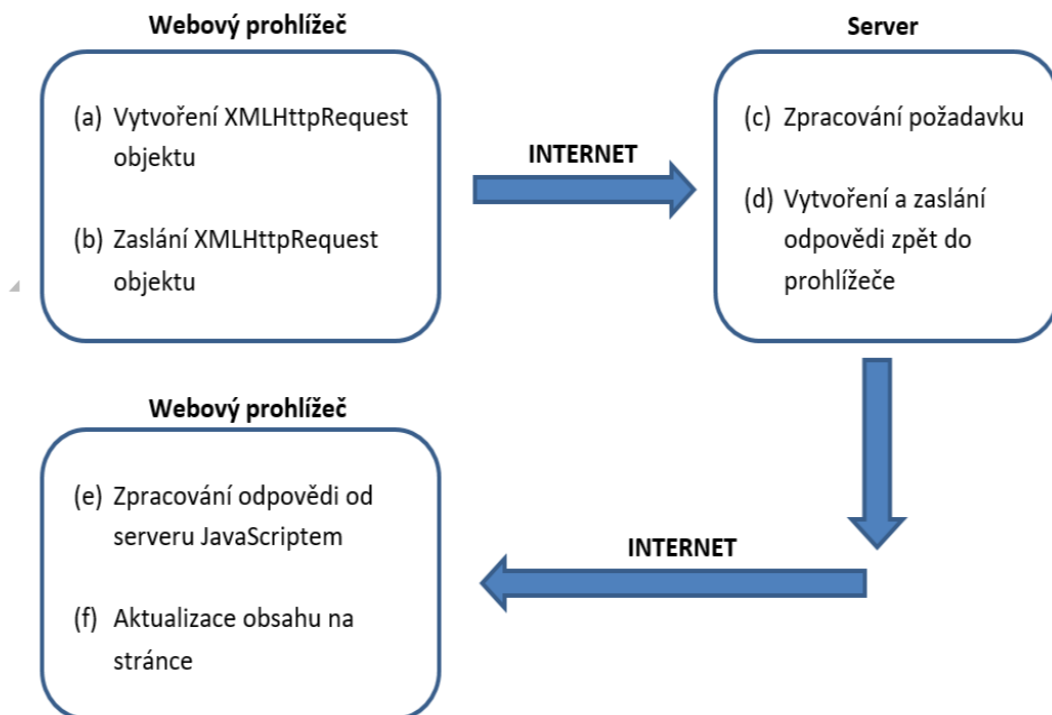
Syntaxe tohoto jazyka vychází z jazyka C, objektový funkcionální model je inspirován v jazyce Scheme a Self a jen některá standardní rozhraní se podobají Javě (proměnné, funkce, objekty). Sám jazyk je dynamicky typovaný (to znamená, že o typech proměnných se rozhoduje až za chodu programu). Velkým rozdílem oproti jiným jazykům je, že JavaScript staví na takzvané *Prototypové dědičnosti*, zatímco Java, PHP a mnoho dalších jazyků používá *Objektově orientovaný model*.

JavaScript byl od začátku navržen právě pro skriptování v prostředí webového prohlížeče. Důležitým aspektem je to, že jeho veškerá činnost je vykonávána na straně uživatele, a tudíž nezatěžuje server. Jeho hlavním úkolem je měnit stránky ze statických na dynamické. Ve zkratce zajišťuje to, že stránka dokáže reagovat na vstupy od uživatele (například ověřování správnosti zadaných údajů do formulářů, nebo provádění změn ve stylu jednotlivých entit - ve smyslu HTML entit). Je důležité mít na paměti, že JavaScript provádí ověřování pouze na straně uživatele a lze takové ověřování obejít, tudíž je neustále nutné provádět validaci i na straně serveru. (Žára, 2015)

### 3.1.4 AJAX (Asynchronní JavaScript a XML)

AJAX není programovací jazyk, ale je tvořen pouze kombinací JavaScriptu, HTML (pro zobrazení výsledných dat) a v prohlížeči zabudovaného XMLHttpRequest objektu, sloužícího k dotazování serveru na potřebná data. AJAX umožňuje webovým stránkám se asynchronně aktualizovat, neboli zajišťuje výměnu dat s webovým serverem, aniž by bylo nutné znovu celou stránku načítat. Typicky je využíván formát XML, ale je možné používat jakýkoliv jiný formát včetně HTML, prostého textu nebo JSON (JavaScript Object Notation) (Woychowsky, 2006). Funkčnost AJAXu se dá popsat v několika bodech (Obr. 9):

- **a)** Vytvoření požadavku (zavolání funkce) a následné vytvoření XMLHttpRequest objektu, který nám reprezentuje daný dotaz.
- **b)** Odeslání objektu na server.
- **c)** Zpracování serverem a následné zaslání odpovědi zpět do prohlížeče.
- **d)** Odpověď serveru je zpracována JavaScriptem.
- **e)** Následuje vrácení odpovědi pro uživatele; např. vypsání do textového pole.



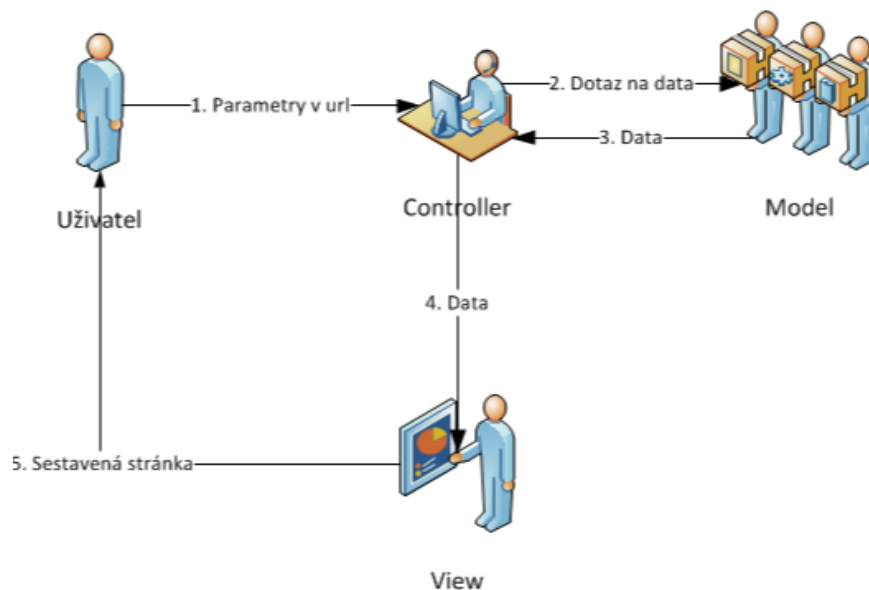
Obrázek 9: Schéma dotazování serveru pomocí AJAX.

### 3.1.5 MVC architektura

Jako základ webové aplikace byla zvolena architektura MVC (Model, View, Controller). Základní idea je taková, že aplikaci rozdělíme na tři základní části, které spolu spolupracují, ale každá z nich má svůj přesně daný úkol: Model, Pohled (View) a Kontroler (Controller). V praxi to znamená hlavně to, že pokud dojde k modifikaci některé z těchto částí, nebude to mít vliv na ostatní. Zvolená architektura umožňuje nejenom pohodlné oddělení samotného GUI (grafické rozhraní) od obecné aplikační logiky, ale také značně usnadňuje eliminaci zbytečných redundancí v kódu.

**Model** je přímo zodpovědný za samotnou práci s daty (například modelové komponenty, které přímo přistupují k datům v databázi MySQL), neboli obsahuje logiku aplikace a vše co do ní spadá. Může se jednat o výpočty, databázové dotazy, validace a podobně. Důležité je, že model vůbec neví o formě finálního zobrazení uživateli. Jeho úkolem je přijetí parametrů a následné vydání výstupu.

**Pohled** se stará o zobrazení výstupu uživateli. Většinou jde o nějakou *phtml* šablonu, která obsahuje HTML stránku a tagy zvoleného značkovacího jazyka (v našem případě PHP), který nám umožňuje do šablony vkládat proměnné, popřípadě provádět podmínky nebo cykly.



Obrázek 10: Schéma funkcionality MVC architektury (převzato z ITnetwork, 2018).

**Kontroler** je mezičlánek zajišťující správné propojení mezi pohledy a modely. Existuje mnoho přístupů, jak kontrolery navrhovat. V naší aplikaci jsme zvolili asi nejrozšířenější postup a to ten, že každá entita má jeden vlastní kontroler.

Výslednou činnost celého systému lze velice zjednodušeně prezentovat na životním cyklu stránky (Obr. 10). Ten začíná zavoláním adresy uživatelem (např. <http://molmedb.upol.cz/detail/14/Diethyleneglycol>). V prvním kroku je adresa předána hlavnímu kontroleru, který ji rozdělí na parametry. Podle prvního parametru zavolá správný kontroler (DetailController) a předá mu zbylé parametry (14, Diethyleneglycol) pomocí pole. Tento kontroler na základě parametrů vykoná správnou činnost, obvykle zavolá určitý model pro zajištění potřebných dat, a nakonec předá data do výsledného pohledu, který se vypíše (Pop *et al.*, 2014).

### 3.1.6 Databáze

Podstata bioinformatiky je práce s velkým množstvím biologických informací. Ať už jde o sekvenční data, microarray datové soubory, spektrometrická data (fingerprinty), katalog kmenů vzniklých z projektů vložené mutagenese nebo i velké množství PDF souborů. Je zřejmé, že taková data nemohou být dlouho ukládána jen v podobě souborů ve složkách. V tu chvíli přichází řada na použití databází.

Důležitým aspektem při návrhu databáze je volba typu databáze. Obecně se dají databáze rozdělit na dva základní typy: **Relační** (SQL databáze) a **Nerelační** („NoSQL“ databáze). SQL je standardizovaný strukturovaný dotazovací jazyk, který se využívá pro práci s daty v relačních databázích, neboli je využíván pro práci s daty, která jsou přirozeně tabulkově uchovávána a existují vzájemné vztahy mezi nimi. Pokud jsou data prezentována a uchovávána jako text, dokumenty nebo jiné netabulkové záznamy, pak bývá lepší použít některou z NoSQL databází. V každém případě je důležité si databáze pravidelně zálohovat, aby nedošlo ke zbytečné ztrátě dat (Helmy *et al.*, 2016).

Silným nástrojem při použití relačních databází je použití cizích klíčů (FK - Foreign Keys). Cizí klíče slouží k propojení dvou tabulek na základě zvoleného sloupce a tudíž nám zprostředkovávají samotné relace (kvůli tomu pojmenování *Relační databáze*). V databázi MolMeDB slouží například k propojení tabulek membrány, metody a molekuly s tabulkou jednotlivých interakcí (viz. dále).

### 3.2 Návrh databáze MolMeDB

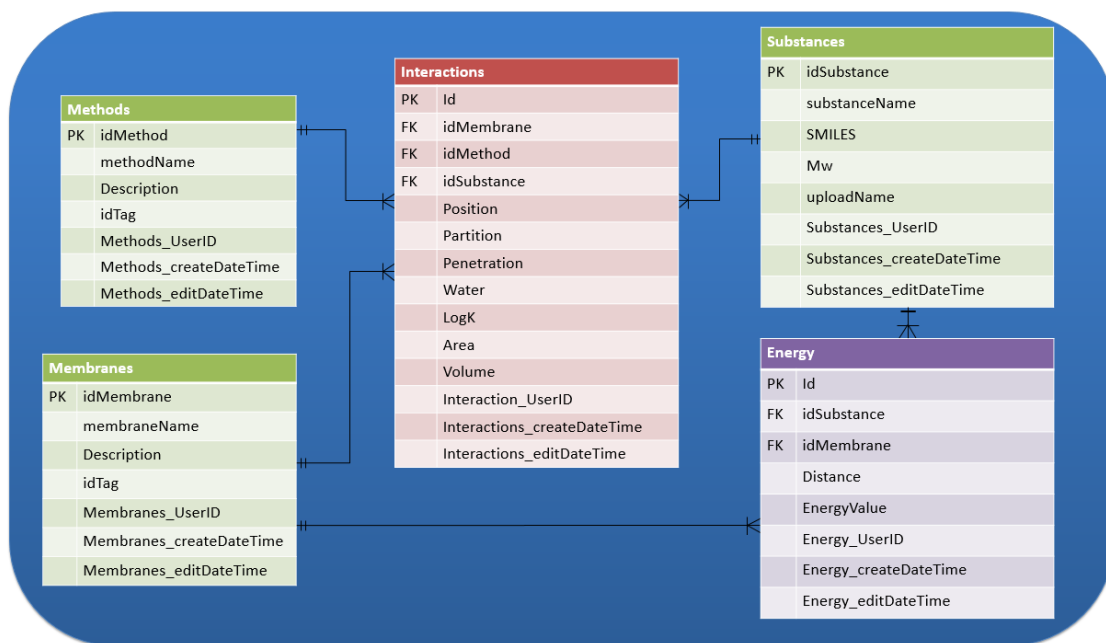
Jako model databáze MolMeDB byla zvolena databáze **MySQL**, která je jednou z nejpoužívanějších SQL databází. Tato volba byla podmíněna nejen tím, že data interakcí mají tabulkovou povahu dat, ale také nutností zachovávat vztahy mezi jednotlivými tabulkami. Jednotlivé tabulky nám zastupují stejné typy záznamů, jako jsou membrány, metody, ligandy, interakce, uživatelé a další. Vztahy se myslí relace mezi jednotlivými tabulkami: jako příklad nám poslouží tabulka interakce, která spojuje data z tabulek membrány, metody a ligandy doplněné o jednotlivé hodnoty daných interakcí. Tudíž každá interakce (každý záznam v tabulce interakce) je jednoznačně určena ligandem a membránou, které spolu interagují a také použitou metodou pro výpočet hodnot dané interakce.

Aktuální alfa verze databáze MolMeDB obsahuje několik hlavních tabulek pro ukládání dat (Obr. 11):

- **Ligandy** – Tabulka obsahující základní informace o jednotlivých molekulách, jako je jejich molekulová hmotnost, objem, povrch a SMILES. V naší aplikaci lze podle SMILES nejen vyhledávat molekuly, ale také slouží ke generování 2D a 3D struktur (viz. dále).
- **SMILES** – Tabulka shromažďující jednotlivé SMILES přidávané uživateli, ať už jed-

notlivě při editaci záznamů, nebo hromadným uploadem.

- **Membrány** – Obsahuje informace o jednotlivých membránách včetně popisu a případných referencí.
- **Metody** – Obsahuje informace o jednotlivých metodách včetně popisu a případných referencí.
- **Interakce** – Tabulka obsahující detailní záznamy jednotlivých interakcí, jež jsou jednoznačně určeny záznamy z tabulek Metody, Membrány a Ligandy pomocí cizích klíčů.
- **Energie** – Obsahuje hodnoty volných energií ligandů.



Obrázek 11: Schéma tabulek pro ukládání dat alfa verze databáze MolMeDB.

### 3.3 Tvorba webového rozhraní

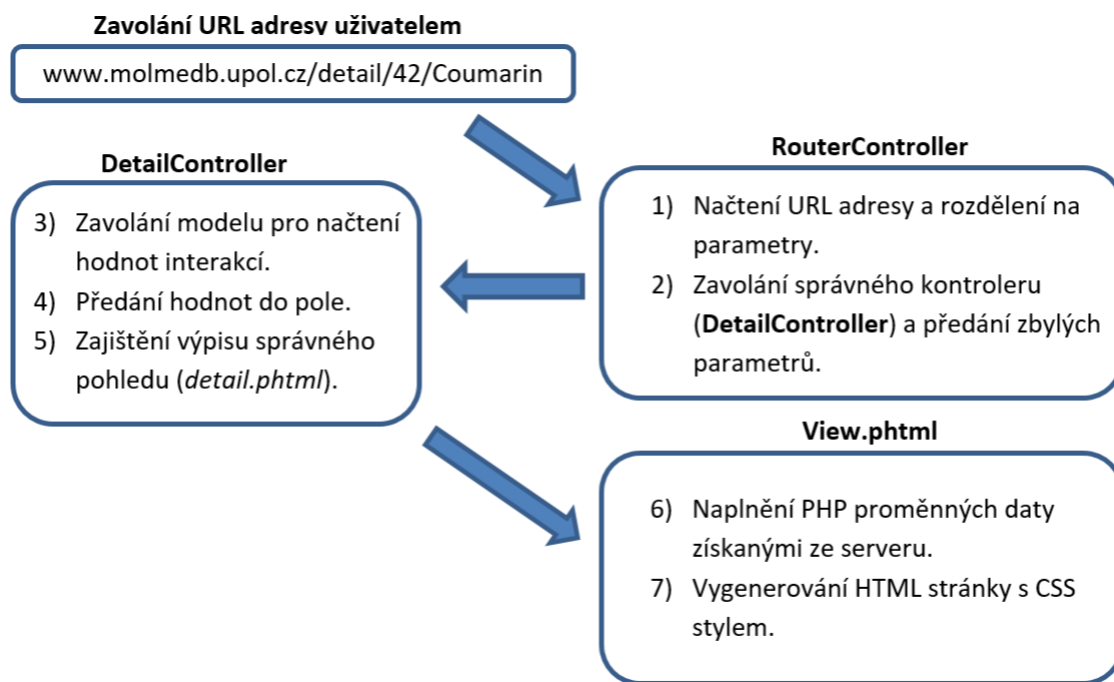
Pro tvorbu webového rozhraní byla zvolena kombinace programovacích jazyků PHP a JavaScript s vizualizací zajištěnou HTML značkovacím systémem s grafickou úpravou pomocí CSS. PHP a JavaScript se liší hlavně tím, že PHP pracuje na straně serveru, zatímco JavaScript pracuje na straně uživatele. To nám umožňuje například nechat běžet složitější výpočetní procesy u uživatele v prohlížeči a tím nepřetěžovat server, což by mohlo mít za následek jeho zpomalení. Jádro celé aplikace je postaveno na MVC architektuře.

#### 3.3.1 MVC architektura databáze MolMeDB

Databáze MolMeDB používá nejrozšířenější postup při návrhu MVC architektury; každá „entita“ (typ stránky) má vlastní kontroler. Jednotlivé kontrolery jsou volány speciálním kontrolerem, který je vždy volán jako první a zajišťuje správné chování stránky a volání kontrolerů na základě parametrů v URL adrese. Všechny kontrolery jsou napsány v jazyce PHP za využití objektově orientovaného modelu. Struktura volání jednotlivých kontrolerů (obr. 12):

- **(-)** Hlavní kontroler je vytvořen pouze jako abstraktní třída a obsahuje základní funkce pro přesměrování, přidávání informačních zpráv do záhlaví stránky, ověřování uživatelů, atd..
- **(a)** Dojde k zavolání kontroleru, který rozšiřuje hlavní kontroler a slouží k rozložení URL adresy na parametry a volání správných kontrolerů.
- **(b)** Zavolání kontroleru podle prvního parametru, který na základě dalších parametrů komunikuje s patřičnými modely za účelem získání potřebných dat z databáze.
- **(c)** Naplnění polí získanými daty, které se budou například vypisovat v pohledu a následně zajištění výpisu správného pohledu.
- **(d)** Výpis pohledu a zobrazení stránky uživateli.

Pohledy jsou založeny na podobném principu vnořování. Základním pohledem je pohled „*layout.phtml*“ obsahující základní HTML kód, který je stejný pro všechny pohledy a následně dochází pouze ke vnoření a výpisu správného pohledu (určeného v kontroleru) do těla („<body>“) základního pohledu.



Obrázek 12: Schéma funkcionality MVC architektury databáze MolMeDB při zobrazování detailu molekuly kumarinu.

Modely jsou samostatné třídy, které je možné volat z libovolného kontroleru a využívat jejich funkčnosti. Každý model má přesně daný název, který zároveň určuje s jakou částí databáze komunikuje.

### 3.3.2 Souborová struktura aplikace

Soubory uložené na serveru, které jsou (v omezené míře) přístupné uživateli, jsou poměrně standardně organizovány do složek podle jejich typu. Samotná aplikace běží na serveru Apache, a protože je postavena na MVC architektuře, bylo nutné, před samotným spuštěním, v nastavení serveru vypnout tzv. *Mod\_rewrite*. *Mod\_rewrite* je funkcionality serverů Apache, která zpracovává URL adresy. V naší aplikaci by nám ale způsobovala po zvolání adresy (např. [www.molmedb.upol.cz/search](http://www.molmedb.upol.cz/search)) hledání složky *search* ve struktuře souborů a volání by skončilo chybou. Ke konfiguraci *Mod\_rewrite* na Apache serveru se běžně využívá soubor *.htaccess* - jednoduše určuje co se má a nemá stát, když uživatel zavolá URL adresu. Nejprve jsme zajistili skrytí souborů ve složce na webu přidáním řádku č. 1 ve Zdrojovém kódu 1.

Dalším důležitým nastavením bylo vypnutí zpracování URL adres Apache serverem a zapnutí přesměrovacího enginu, který každou URL adresu přesměruje na **index.php**



```

1 Options - Indexes
2 RewriteEngine On
3 RewriteRule !\.(css|js|icon|zip|rar|png|jpg|gif|pdf)\$ index.php

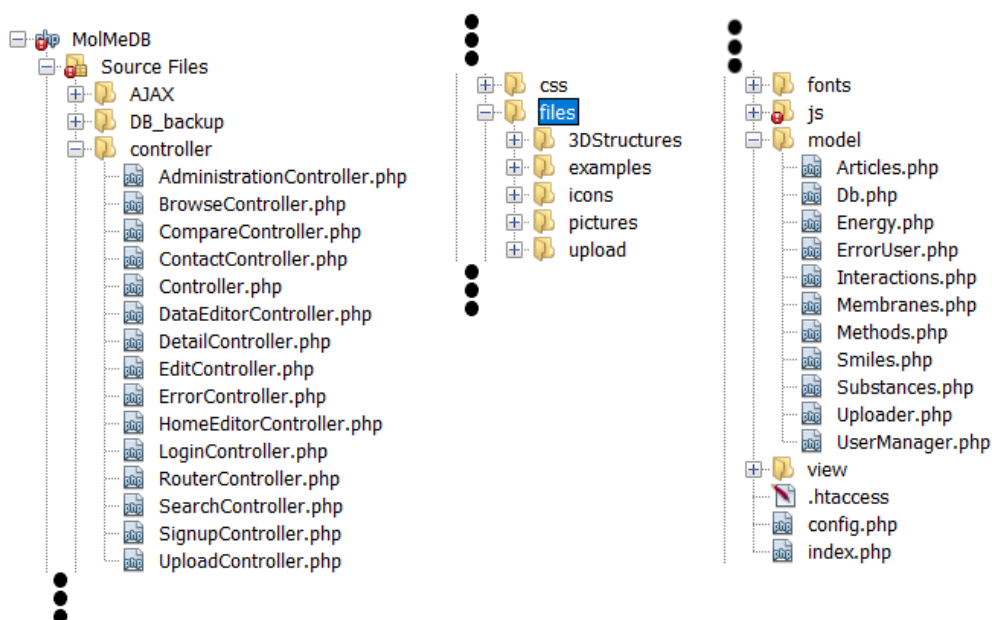
```

Zdrojový kód 1: Část obsahu souboru .htaccess pro nastavení serveru Apache.

(Zdrojový kód 1 - řádky 2 a 3). Tím zajistíme, že vždy dojde k zavolání index.php, který dále obsahuje funkci pro zavolání RouterControlleru a funkci pro výpis pohledu.

Na obrázku 13 je znázorněna struktura uložení souborů v aplikaci MolMeDB:

- **AJAX** - Obsahuje všechny soubory pro asynchronní dotazování serveru za běhu programu.
- **Controller, Model, View** - Obsahují všechny kontrolery využívané aplikací (controller), modely zajišťující nejen komunikaci s databází (model) a *.phtml* soubory, které obsahují HTML + PHP kód udržující celkovou strukturu zobrazení (view).
- **CSS, JS** - Obsahují strukturovaně uložené soubory pro vizualizi webu (css) a javascriptové soubory sloužící například ke generování grafů volných energií (js).
- **Files** - Obsahuje veškeré nahrané soubory - 3D struktury, vzorové soubory, . . .



Obrázek 13: Struktura uložení souborů aplikace MolMeDB na serveru.

## 4 Výsledky a diskuze

### 4.1 Vyhledávač

Vyhledávač je jednou z nejdůležitějších funkcionalit každé databáze. V aktuální alfa verzi databáze MolMeDB je k dispozici propracovaný vyhledávač, který nabízí možnost vyhledávat jednotlivé molekuly podle jejich názvu, případně podle SMILES zápisu. Další funkcionalitou je možnost si nechat vypsát pouze molekuly, které mají k dispozici záznamy o interakcích pro určitou membránu nebo metodu (Obr. 14b). Nechybí ani našeptávač, který při psaní dotazu využívá AJAX k dotazování serveru a nabízí průběžně nejlepší výsledky vyhledávání (Obr. 14a).

The image shows two screenshots of the MolMeDB search interface. Screenshot (a) shows the search results for the query 'Cort'. The search bar is set to 'Name' and the results are displayed as a list of suggestions: 'Cort', 'Cortexolone', 'Cortexone', and 'Hydrocortisone'. Screenshot (b) shows the search results for the query 'DOPC'. The search bar is set to 'Membrane' and the results are displayed as a list of membranes: 'DOPC', 'DMPC', and 'SC MIX'. Each membrane result is shown in a separate card with a description field and a 'Show compounds' button.

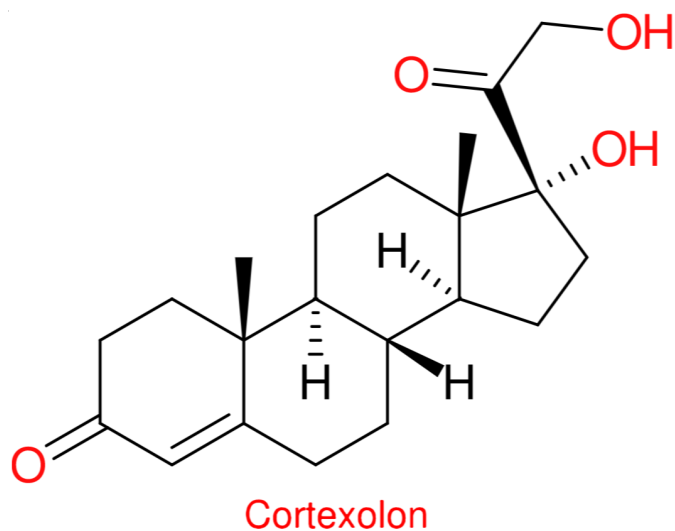
Obrázek 14: Aktuální verze vyhledávače v databázi MolMeDB. (a) Ukázka funkcionality našeptávače. (b) Ukázka zobrazení seznamu membrán s možností výpisu molekul, které pro danou membránu obsahují záznamy o interakcích.

## 4.2 Zobrazení detailu ligandu

V databázi MolMeDB má každý jednotlivý ligand svou stránku s detailním popisem. V aktuální alfa verzi databáze obsahují detaily jednak základní informace o dané molekule a hodnoty jednotlivých interakcí, ale také 2D a 3D struktury a grafové zobrazení profilů volných energií molekul na membránách.

### 4.2.1 2D struktury

Databáze MolMeDB nabízí pro každou v ní obsaženou molekulu vizualizaci její struktury, a to jak ve dvou, tak i ve třech rozměrech. Pro generování 2D struktur je používána aplikace CDK Depict (Steinbeck *et al.*, 2013). Jedná se o volně dostupnou aplikaci, jejíž nespornou výhodou je, že generuje vizualizace molekul přímo ze zápisu SMILES. Jelikož je SMILES zápis každé molekuly uložen v databázi, odpadá tímto nutnost ukládat jednotlivé obrázky molekul nebo textové soubory s rozmístěním atomů na server. Tvorba 2D struktury začíná ihned při zobrazování detailu molekuly a je načtena prakticky okamžitě po vyobrazení stránky. Dále nám aplikace umožňuje široké spektrum nastavení samotné vizualizace pro získání co nejlepšího výsledku (Obr. 15).



Obrázek 15: Grafické zobrazení molekuly Cortisolu z jejího SMILES zápisu pomocí aplikace CDK Depict.

### 4.2.2 3D struktury

Interaktivní 3D zobrazení zajišťuje aplikace LiteMol Suite. Samotná aplikace je vytvořena pro vizualizaci velkých makromolekulárních struktur a funguje na všech moderních webových prohlížečích a také na mobilních přístrojích. Aplikace zvládá generovat (i makromolekulární) struktury ve velice krátkém čase a v kombinaci s její dostupností, intuitivním ovládním, plynulým chodem a širokou škálou nastavení se stala silným nástrojem pro zobrazování trojrozměrných struktur. Samotné ovládním aplikace a možnosti zobrazení jednotlivých molekul je velice podobné široce rozšířenému programu PyMOL.

Aplikace je dále propojena s několika nejrozšířenějšími databázemi (makro)molekul z důvodu pohodlného načítání struktur. Jedná se zejména o databáze PDB, PDBe, EMDB a UniProt, ze kterých lze načítat data pomocí identifikátorů; například užití PDB id pro načtení struktury (např. 1TQN pro získání struktury cytochromu P450 3A4) (Sehnal *et al.*, 2017).

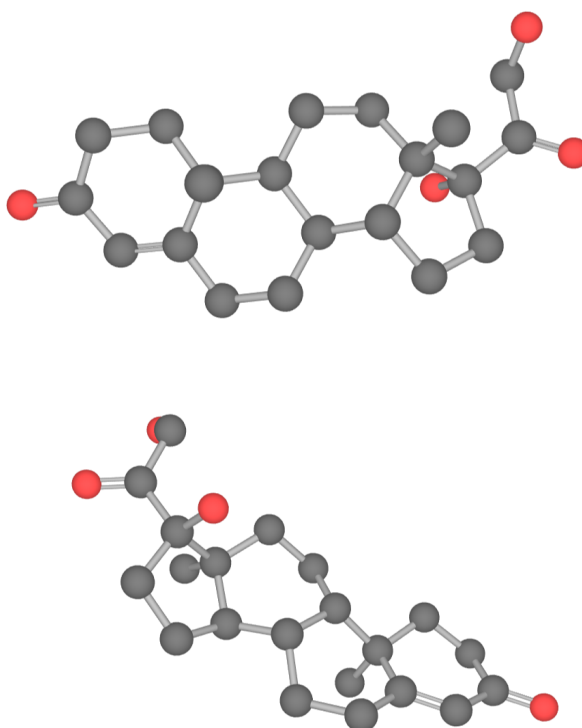
V databázi MolMeDB se ukládá pro každou molekulu zvlášť soubor („*example.mol*“), který obsahuje všechny potřebné informace k vizualizaci její struktury (obsah souboru pro vyobrazení struktury molekuly benzenu je znázorněna na obrázku 16); informace o atomech, vazbách, propojení a souřadnicích. Takový soubor je při zobrazení detailu pouze načten aplikací LiteMol a následně je vykreslena struktura (Obr. 17). Ukládání strukturních souborů na server není paměťově náročné (jedná se pouze o textové soubory) a navíc samotné načítání ze souboru a generování struktury je mnohem rychlejší než načítání z jiných serverů. K získávání strukturních souborů je využívána aplikace RDKit, která daný soubor vytváří převodem z kanonického SMILES. RDKit je sbírka chemoinformatických softwarů napsaných v jazycích C++ a Python.

```

Benzene
  RDKit      3D
6  6  0  0  0  0  0  0  0  0  0999 V2000
  1.1254 -0.8291 -0.0433 C  0  0  0  0  0  0  0  0  0  0  0
0  0
  -0.1553 -1.3899 -0.0079 C  0  0  0  0  0  0  0  0  0  0  0
0  0
  -1.2807 -0.5607  0.0355 C  0  0  0  0  0  0  0  0  0  0  0
0  0
  -1.1254  0.8291  0.0433 C  0  0  0  0  0  0  0  0  0  0  0
0  0
  0.1553  1.3899  0.0079 C  0  0  0  0  0  0  0  0  0  0  0
0  0
  1.2807  0.5607 -0.0355 C  0  0  0  0  0  0  0  0  0  0  0
0  0
  1  2  2  0
  2  3  1  0
  3  4  2  0
  4  5  1  0
  5  6  2  0
  6  1  1  0
M  END

```

Obrázek 16: Obsah souboru *benzene.mol* pro vizualizaci 3D struktury benzenu generovaného aplikací RDKit.



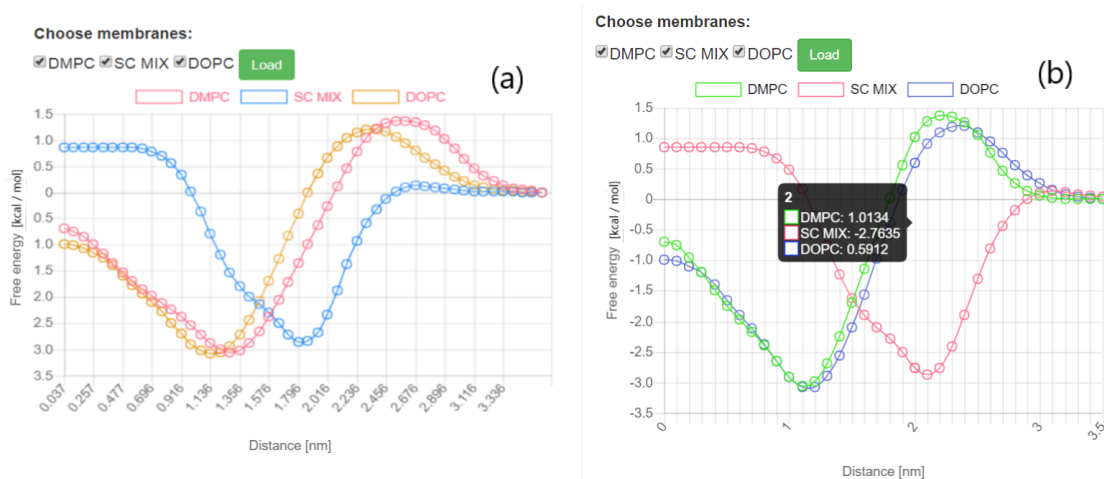
Obrázek 17: Zobrazení 3D struktury molekuly Cortisolonu ze dvou různých úhlů v programu LiteMol.

### 4.2.3 Grafické zobrazení profilů volných energií

Profil volné energie molekul na membránách je hlavním nástrojem pro vyhodnocování interakcí. Dává nám nejen informace o afinitě látky k membráně, ale je také důležitý pro zjišťování pozice energetického minima a penetrační bariéry. Jednotlivé profily volných energií jsou získávány různými ověřenými predikčními metodami. Výsledkem jsou tabulkové hodnoty, v nichž každý záznam je reprezentován daným ligandem, membránou a použitou metodou a pro každý takový záznam máme predikovaných 50 hodnot energií v závislosti na vzdálenosti ligandu od membrány. Databáze MolMeDB zobrazuje pro každý ligand všechny naměřené profily volných energií v grafické podobě.

K vizualizaci grafů je využívána aplikace Chart.js. Jedná se o volně dostupnou webovou aplikaci, která slouží k zobrazování dat osmi různými typy grafů; každý z nich je animovaný a volně přizpůsobitelný. Základním aspektem pro vizualizaci většího počtu grafů volných energií do jednoho grafu (za účelem porovnávání) je, že daná data musí mít stejné hodnoty na horizontální ose (Downie, 2015). V databázi MolMeDB jsou ukládány grafové hodnoty v tabulkové formě, ale pro různé membrány jsou naměřeny hodnoty v různých vzdálenostech. To při implementaci způsobilo nepřesné vyobrazení; došlo v některých částech grafu k nahromadění hodnot a také k různým posunům (Obr. 18a). Řešení tohoto problému (Obr. 18b) lze popsat ve dvou bodech:

- 1) Vytvoření ekvidistantní sítě uzlů na horizontální ose.
- 2) Následná interpolace uzlů hodnotami jednotlivých interakcí.



Obrázek 18: Porovnání vizualizace grafů volných energií molekuly 1,10-Decanediolu na 3 různých membránách (DMPC, SC MIX, DOPC) (a) bez použití interpolace hodnot a (b) s použitím interpolace včetně ukázky porovnávání hodnot ve vzdálenosti 2nm.

Vytvoření ekvidistantní sítě uzlů nám zajistí rovnoměrné rozložení hodnot v grafu s čímž souvisí jednak příjemnější vizualizace, ale také možnost porovnávat jednotlivé hodnoty daných interakcí ve stejných vzdálenostech pouhým pohybem myši (Obr. 18b). Následnou interpolací získáme jednotlivé hodnoty v daných bodech na základě zjištěných experimentálních hodnot. Pro interpolaci byl zvolen tzv. **Nevillův algoritmus** (*Iterovaná interpolace*). Samotná implementace (Zdrojový kód 2) je prováděna následovně:

- **1)** Vytvoření ekvidistantní sítě uzlů pro vzdálenosti od 0 nm do 3.5 nm s krokem 0.1 nm.
- **2)** Procházení jednotlivých bodů v cyklu:
  - **2.1)** Nalezení nejbližšího indexu pro daný bod v poli naměřených hodnot.
  - **2.2)** Načtení několika okolních bodů pro zjištěný index, které se následně použijí k interpolaci.
  - **2.3)** Vrácení výsledku interpolace a zaokrouhlení na 4 desetinná místa.

```

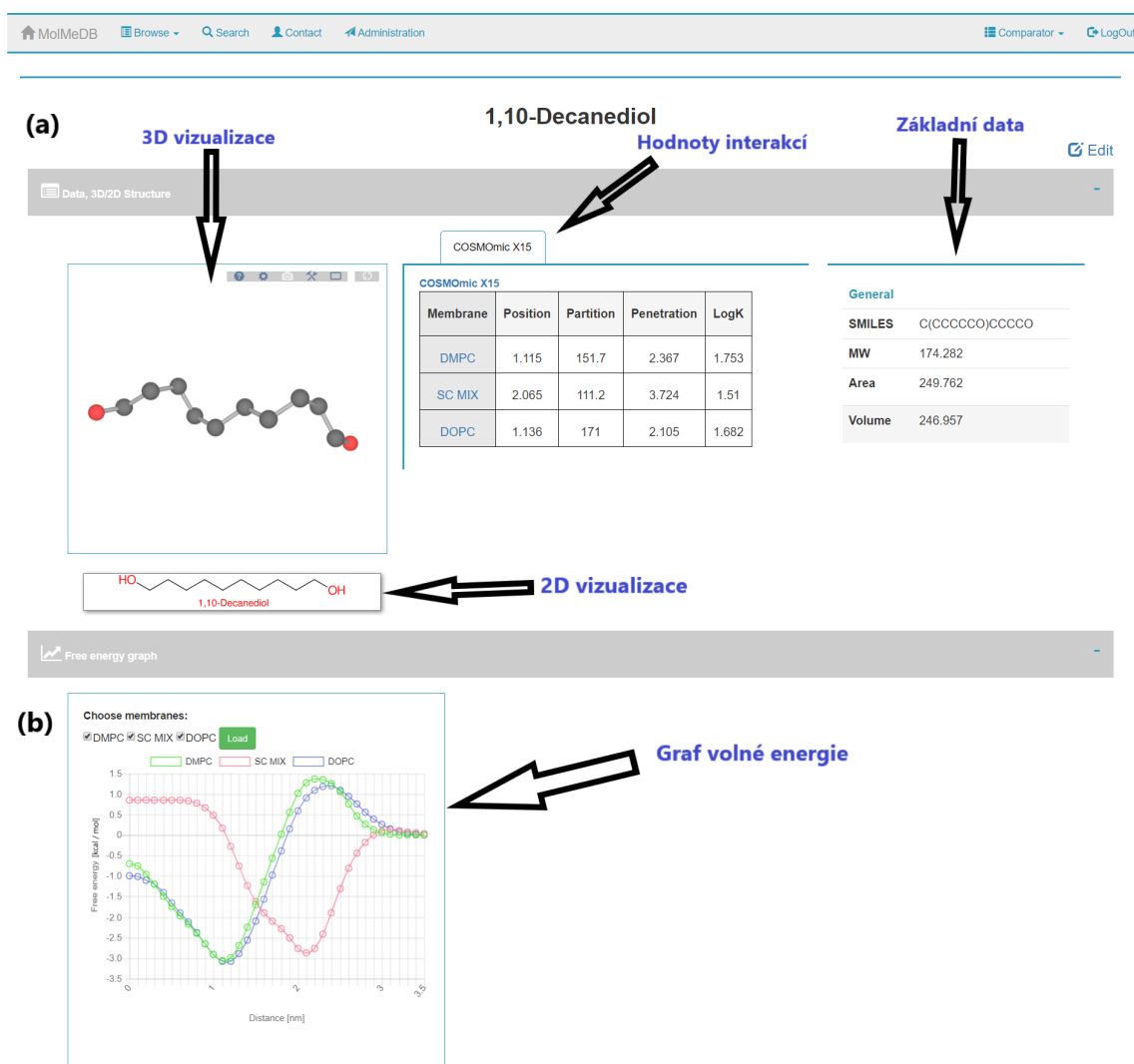
1 for (var j = 0; j < 36; j++) { //Interpolace po jednotlivých bodech
2   if (j === 0) {
3     values[j] = y_main[0];
4     continue;
5   }
6   index = findNearestIndex(x_main, labelsSet[j]); //Nalezení indexu
7   if (index === -1) //Podmínka ověřující jestli je bod v rozsahu pole
8     values[j] = 0;
9   else {
10    X = fillValues(x_main, index); //Získání hodnot X pro interpolaci
11    Y = fillValues(y_main, index); //Získání hodnot Y pro interpolaci
12    values[j] = nevillesIteratedInterpolation(labelsSet[j], X, Y);
13    values[j] = Math.round(values[j] * 10000) / 10000;
14  }
15 }
16 function nevillesIteratedInterpolation (x, X, Y) { //Nevillův alg.
17   var Q = [Y];
18   for (var i = 1; i < X.length; i++) {
19     Q.push([]);
20     for (var j = 1; j <= i; j++) {
21       Q[j][i] = ((x-X[i-j])*Q[j-1][i] - (x-X[i])*Q[j-1][i-1])/( X[i] - X[i-j] );
22     }
23   }
24   return Q[j-1][i-1];
25 }

```

Zdrojový kód 2: Interpolace ekvidistantních uzlů pro graf volné energie v jazyce JavaScript.

## 4.2.4 Tabulkové hodnoty interakcí

Samotné detaily jednotlivých ligandů a jejich interakcí s membránami jsou uživateli zobrazeny v podobě tabulek. Jedna část detailu obsahuje tabulku se základními vlastnostmi dané molekuly včetně SMILES zápisu (Obr. 19a) a druhá část obsahuje tabulky s hodnotami interakcí. Tyto tabulky jsou rozděleny podle metod použitých k měření hodnot jednotlivých interakcí a každá tabulka obsahuje po řádcích hodnoty jednotlivých interakcí příslušných k daným membránám (Obr. 19a).



Obrázek 19: Detail molekuly 1,10-Decanediolu. V části (a) zobrazeny sekce s 3D a 2D vizualizací, hodnotami interakcí pro různé membrány získané pomocí metody COSMOmic X15 a obecnými informacemi o molekule. V části (b) je vidět zobrazení grafu volné energie pro získané hodnoty.



### 4.3 Comparator - sekce pro vzájemné porovnávání molekul

Porovnávání jednotlivých molekul je důležitým aspektem každé databáze interakcí, neboť nabízí ucelený pohled na hodnoty jednotlivých interakcí. Databáze MolMeDB disponuje možností si navolit jakékoli z dostupných molekul a následně si je porovnat. K dispozici je jednak zobrazení hodnot v tabulkové formě a taky možnost porovnání grafů volných energií.

Prvním krokem pro použití comparatoru je volba molekul, které se budou porovnávat. To lze udělat jednoduše ve vyhledávači. Uživatel použije vyhledávač k nalezení určitého ligandu a tlačítkem (+) přidá látku do seznamu porovnávaných látek. Seznam zvolených ligandů je možné prohlížet v jakékoli sekci po rozkliknutí položky Comparator v hlavním menu. Odebírání látek ze seznamu funguje na stejném principu, ale navíc lze látky odebírat přímo v menu ze seznamu zvolených látek, popřípadě vyprázdnit celý seznam. Po volbě látek se lze přes seznam látek v hlavním menu přesunout do sekce Comparator, kde lze jednotlivé ligandy spolu porovnávat pomocí interaktivní tabulky.

Nad tabulkou se ještě nachází rozšíření (Advanced), ve kterém lze zvolit jaké metody a membrány uživatele zajímají (Obr. 20). Samotná tabulka je interaktivní a pomocí názvů sloupců ji lze třídit. Například kliknutím na sloupec s názvem „MW“ se látky seřídí vzestupně podle molekulové hmotnosti. Dalším kliknutím se seřídí sestupně. V posledním sloupci se nachází políčka, které lze zaškrtnout (pokud je dostupný profil volné energie) a následně si může uživatel nechat vygenerovat do jednoho grafu zvolené profily volných energií pro jejich grafické porovnání (Obr. 21).

### Comparator

**Advanced**

**Choose methods:**

 COSMOmic X15

Click on compound to remove from list:

Chlorzoxazone

2-Naphthylamine

2-Butoxyethanol

2-Isopropoxyethanol

1,2-Decanediol

1,10-Decanediol

**Choose membranes:**

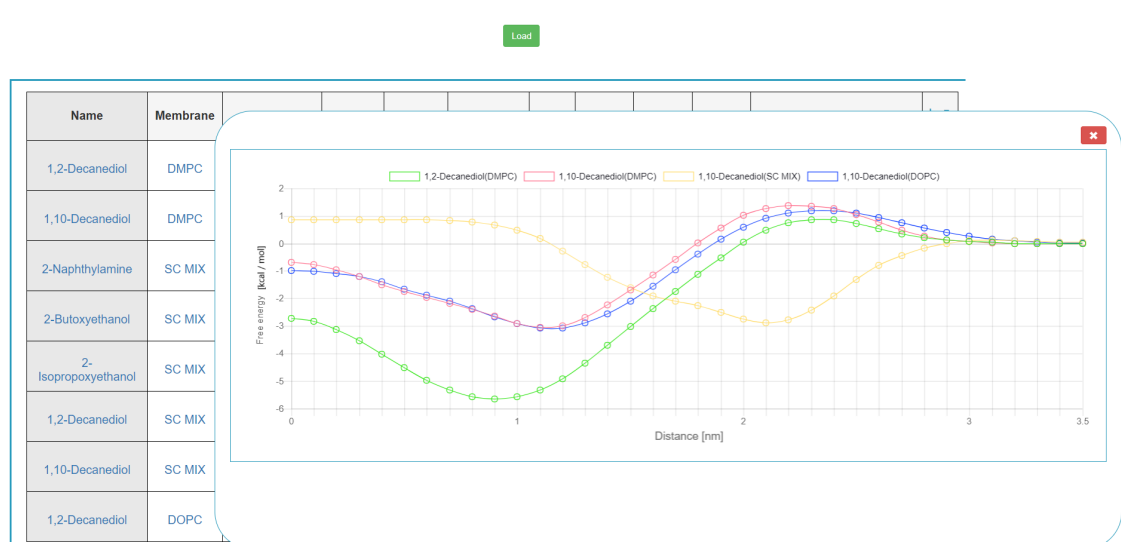
 DMPC  
 SC MIX  
 DOPC

**Tips:**

- Click on **Load** to generate the table
- Click on the name of column (e.g. **LogK**) to sort the table
- For visualising free energy graph check checkbox at last column and then Click on icon of graph in the header of table
- In **Advanced** are available selections of **compounds**, **membranes** and **methods**

Name	Membrane	Method	Position	Partition	Penetration	LogK	Area	Volume	MW	SMILES	
1,2-Decanediol	DMPC	COSMOmic X15	0.924	9786	2.914	3.566	246.818	246.617	174.282	CCCCCCCC(CO)O	<input type="checkbox"/>
1,10-Decanediol	DMPC	COSMOmic X15	1.115	151.7	2.367	1.753	249.762	246.957	174.282	C(CCCCCO)CCCCO	<input type="checkbox"/>
2-Naphthylamine	SC MIX	COSMOmic X15	1.897	884.1	1.285	2.415	182.332	182.762	143.186	NC1=CC=CC=C2C=C1	<input type="checkbox"/>

Obrázek 20: Ukázka vizualizace Comparatoru s otevřenou sekci Advanced pro rozšířené možnosti zobrazení tabulky.



Obrázek 21: Ukázka vizualizace porovnávání grafů volné energie u zvolených interakcí (1,2-Decanediol [DMPC], 1,10-Decanediol[DMPC,SC MIX, DOPC]) v sekci Comparator.

## 4.4 Administrátorská sekce

Administrátorská sekce je důležitou funkcionalitou, která nabízí možnost spravovat jednotlivá data v databázi v uživatelsky přijatelné formě. Přístup do ní je povolen pouze registrovaným uživatelům, kterým byla přidělena administrátorská práva. Administrátor poté dostane možnost přistupovat k různým sekcím pro správu dat jako jsou hlavně sekce pro nahrávání dat a editaci dat.

#### 4.4.1 Sekce pro nahrávání dat

Jednou z nejdůležitějších sekcí je sekce pro vkládání dat do databáze. V této sekci je aktuálně možné vkládat nové membrány a metody, záznamy SMILES, nové molekuly s hodnotami interakcí a také soubory s hodnotami profilů energií.

Vkládání metod a membrán je prováděno pomocí formuláře, ve kterém je potřeba pro vložení vyplnit základní informace o membráně (metodě) včetně popisu a případných referencí. Pro pohodlný zápis a formátování popisu a referencí byl do formuláře implementován webový textový editor, který nápadně vzhledem i funkcionalitou připomíná běžné desktopové textové editory (MS WORD, OpenOffice, ad.). Stejný editor je použit pro úpravu úvodní stránky.

Vkládání molekul s hodnotami interakcí, SMILES zápisů a hodnot volných energií probíhá rovněž přes formulář, ale tato činnost je už částečně automatizována, a to hlavně díky souborovému vkládání dat. Jako příklad poslouží vkládání souboru s novými molekulami a hodnotami jejich interakcí. Jedná se o soubor, který obsahuje po řádcích jednotlivé ligandy a po sloupcích naměřené hodnoty interakcí jednotlivých ligandů s danou membránou (Obr. 22). Pro nahrání souboru stačí zvolit metodu, kterou probíhalo měření hodnot interakcí, a membránu se kterou dané ligandy interagovali. Nejprve dojde k ověření správnosti souboru a následně po úspěšném nahrání dojde k automatickému přerozdělení hodnot a naplnění databáze.

Ligand	Position	Penetration	Water	Partition	LogK	LogK_COSMO	Area	Volume	MW	Resistance	Log_perm
1_10-decanediol	1.115	2.367	4.082	1.517e+02	2.181	1.753	249.762	246.957	174.282	6.053e+00	-0.782
1_2-decanediol	0.924	2.914	6.349	9.786e+03	3.991	3.566	246.818	246.611	174.282	2.765e+00	-0.442
1_2-hexanediol	1.115	3.620	2.406	2.801e+01	1.447	0.996	169.911	162.603	118.175	6.578e+00	-0.818
1_2-octanediol	0.987	3.114	4.598	7.474e+02	2.874	2.448	208.179	204.485	146.228	2.511e+00	-0.400
1_3-decanediol	1.051	4.131	5.483	1.753e+03	3.244	2.819	246.850	247.158	174.282	3.908e+00	-0.592
1_4-decanediol	1.051	3.703	5.268	1.227e+03	3.089	2.663	246.921	247.744	174.282	3.854e+00	-0.586
1_5-decanediol	1.051	3.526	5.040	8.070e+02	2.907	2.481	247.699	247.265	174.282	3.951e+00	-0.597
1_6-decanediol	1.051	3.303	4.731	5.096e+02	2.707	2.281	246.173	247.360	174.282	3.962e+00	-0.598
1_6-hexanediol	1.306	4.268	1.399	5.775e+00	0.762	0.204	171.171	162.630	118.175	7.229e+01	-1.859
1_7-decanediol	1.051	3.175	4.568	3.775e+02	2.577	2.151	246.971	247.281	174.282	4.366e+00	-0.640
1_8-decanediol	1.051	2.962	4.376	3.013e+02	2.479	2.052	245.033	247.408	174.282	4.378e+00	-0.641
1_8-octanediol	1.178	3.405	2.653	2.318e+01	1.365	0.915	212.128	205.374	146.228	7.553e+00	-0.878
1_9-decanediol	1.115	2.818	4.327	1.862e+02	2.270	1.843	248.064	247.525	174.282	6.677e+00	-0.824
diethyleneglycol	1.370	3.012	0.534	2.680e+00	0.428	-0.385	146.692	133.150	106.121	5.996e+01	-1.778

Obrázek 22: Ukázka vzorového souboru s ligandy a hodnotami jednotlivých interakcí, který je vhodný k použití pro plnění databáze MolMeDB.

## 4.4.2 Sekce pro editace dat

Žádná databáze se nevyhne občasné potřebě úprav hodnot jednotlivých dat. Může se třeba jednat jen o úpravu určitého typu dat (název molekuly) do uživatelsky lépe čitelné formy. Ve webovém rozhraní databáze MolMeDB je možné upravovat uložené informace o membránách a metodách přímo v administrátorské sekci a hodnoty jednotlivých interakcí lze upravovat v editační sekci (Obr. 23).

MolMeDB

Home MolMeDB Browse Search Contact Administration Comparator LogOut

### Editor Cortexone

<b>Substance name</b> Cortexone	<b>Membrane</b> SC MIX <b>Method</b> COSMOmic X15 <b>Load</b>
<b>SMILES</b>	<b>Position [nm]</b> 2.065
<chem>C[C@]12CC[C@@H]3[C@@H](CC4=CC(=O)CC[C@]34C)[C@@H]1CC[C@@H]2C(=O)</chem>	<b>Partition [1]</b> 664
<b>MW [Da]</b> 330.463	<b>Penetration [kcal / mol]</b> 2.817
<b>Volume [Å<sup>3</sup>]</b> 406.682	<b>Water [kcal / mol]</b> 4.635
<b>Area [Å<sup>2</sup>]</b> 332.129	<b>LogK [1]</b> 2.289
<b>Save</b>	<b>Save</b>

Obrázek 23: Ukázka editační sekce pro molekulu Cortexonu v databázi MolMeDB.

## 5 Závěr

V této práci jsme studovali problematiku interakcí malých molekul s biologickými membránami a tvorbu jejich databáze. V databázi **MolMeDB** je možné nalézt popis použité metody, kterou byly predikovány hodnoty interakcí léčiv s biomembránami; konkrétně se jedná o metodu COSMO-RS s nastavbou COSMOmic. Biomembrány se liší fyzikálně-chemickými vlastnostmi a pro predikci jejich interakcí je potřeba navrhnout vhodné modely studovaných membrán. V naší databázi jsou aktuálně dostupné 3 druhy membrán: 1) **DOPC**, 2) **DMPC** a 3) **SC MIX**. Membrány DOPC a DMPC slouží jako modelové membrány endoplazmatického retikula a membrána SC MIX imituje svrchní vrstvu pokožky.

Databáze **MolMeDB** byla vytvořena jako relační databáze, konkrétně MySQL, která shromažďuje tisíce profilů volných energií s napočítanými hodnotami interakcí léčiv s různými biomembránami. Pro pohodlné vyhledávání a zobrazování dat z databáze bylo vytvořeno webové rozhraní (dostupné na adrese [www.molmedb.upol.cz](http://www.molmedb.upol.cz)). Webová aplikace je postavena na MVC architektuře, která nám zajišťuje oddělení samotné aplikační logiky od GUI a díky tomu došlo k eliminaci zbytečných redundancí v kódu.

Díky grafickému rozhraní je možné v databázi vyhledávat podle názvu molekuly, SMILES zápisu nebo je možné si nechat vypsat ligandy, které obsahují záznamy o interakcích pro zvolenou membránu nebo metodu. Další funkcionalitou je možnost porovnávání molekul v sekci **Comparator**. Porovnávání je zprostředkováno jednak pomocí interaktivní tabulky, nebo je možné si nechat vygenerovat profily volných energií zvolených molekul.

Detaily jednotlivých ligandů obsahují tabulky s hodnotami jednotlivých interakcí rozdělené podle použitých metod a membrán. Další tabulka obsahuje základní informace o dané molekule. Detaily molekul jsou dále rozšířeny o **2D** a **3D vizualizace** a interaktivní **grafy volných energií**.

Do budoucna máme v plánu, kromě pokračujícího plnění databáze, rozšíření vyhledávače a administrátorské sekce - vylepšení automatického nahrávání a přidání možnosti ověřování vkládaných dat. Rádi bychom v budoucnu umožnili přidávání dat i jinými výzkumnými centry. Dále máme v plánu vytvoření tzv. *tag systému*, díky kterému bude možné jednoduše propojovat záznamy o metodách, membránách, molekulách a interakcích s jednotlivými vědeckými články, které o nich pojednávají.

## 6 Literatura

- Alberts B., Bray D., Johnson A., Lewis J., Raff M., Roberts K., Walter P. (2002): *Základy buněčné biologie - Úvod do molekulární biologie buňky*. Espero, Ústí nad Labem, ČR, 740 stran (373-391).
- Apweiler R., Attwood K. T., Bairoch A., Bateman A., Birney E., [...], Zdobnov E. M. (2001): The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acid Research*, Vol. 29, 1, 37-40. <https://doi.org/10.1093/nar/29.1.37>
- Atala A. (2016): *Small Molecules*. In: *In Situ Tissue Regeneration: Host Cell Recruitment and Biomaterial Design*. (Lee S. J., Atala A., Yoo J. J.), Wake Institute for Regenerative Medicine, Winston-Salem, NC, U.S.A., 87-110.
- Chatr-aryamontri A., Oughtred R., Boucher L., Rust J., Chang Ch., [...], Tyers M. (2017): The BioGRID interactions database: 2017 update. *Nucleic Acids Research*, Vol. 45, D1, 369-379. <https://doi.org/10.1093/nar/gkw1102>
- Chen A., Moy T. V. (2000): Cross-Linking of Cell Surface Receptors Enhances Cooperativity of Molecular Adhesion. *Biophysical Journal*, 2814-2820. [https://doi.org/10.1016/S0006-3495\(00\)76824-X](https://doi.org/10.1016/S0006-3495(00)76824-X)
- Creative Biostructure - Custom MemPro Membrane Protein Services*. [https://www.creative-biostructure.com/custom-mempro%E2%84%A2-membrane-protein-services\\_15.htm](https://www.creative-biostructure.com/custom-mempro%E2%84%A2-membrane-protein-services_15.htm) (23.3.2018).
- Escribá P.V (2006): Membrane-lipid therapy: a new approach in molecular medicine. *Molecular and Cellular Biomedicine* vol. 12, 33-34. <https://doi.org/10.1016/j.molmed.2005.11.004>
- Fernández M. J., Hoffmann R., Valencia A. (2007): iHOP web services. *Nucleic Acid Research*, Vol. 35, 21-26. <https://doi.org/10.1093/nar/gkm298>
- GenBank and WGS statistics*. <https://www.ncbi.nlm.nih.gov/genbank/statistics/> (20.3.2018).
- Gilson K. M., Liu T., Baitaluk M., Nicola G., Hwang L., Chong J. (2015): BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Research*, Vol. 44, D1, 1045-1053. <https://doi.org/10.1093/nar/gkv1072>
- Gutmans A., Bakken S., Rethans D. (2015): *PHP 5 Power Programming*. Prentice Hall, New Jersey, U.S.A.
- Helmy M., Crits-Christoph A., Bader G. D. (2016): Ten Simple Rules for Developing Public Biological Databases. *PLoS Computational Biology* 12(11): e1005128. <https://doi.org/10.1371/journal.pcbi.1005128>
- Hine R. (1999): *Membranes*. In: *The Facts on File Dictionary of Biology*, (Daintith J.), New York, Checkmark, U.S.A., 198.
- Iudin A., Korir P., Salavert-Torres J., Kleywegt J. G., Patwardhan A. (2016): EMPIAR: a public archive for raw electron microscopy image data. *Nature Methods* 13, 387-388. <https://doi.org/10.1038/nmeth.3806>
- ITnetwork - Popis MVC architektury*. <https://www.itnetwork.cz/php/mvc/objektovy-mvc-redakcni-system-v-php-popis-architektury> (12.3.2018).
- Jiráť J., Svozil D. (2017): Lineární reprezentace chemických struktur. *Chemické listy* 111, 710-715.
- Kerrien S., Aranda B., Breuza L., Bridge A., Broackes-Carter F., [...], Hermjakob H. (2011): The IntAct molecular interaction database in 2012. *Nucleic Acid Research*, Vol. 40, D1, 841-846. <https://doi.org/10.1093/nar/gkr1088>
- Klamt A., Huniar U., Spycher S., Kaldenich J. (2008): COSMOmic: A Mechanistic Approach

- to the Calculation of Membrane-water Partition Coefficients and Internal Distributions within Membranes and Micelles. *The Journal of Physical Chemistry B* 112, 38, 12148-12157. <https://doi.org/10.1021/jp801736k>
- Kozma D., Simon I., Tusnády E. G. (2012): PDBTM: Protein Data Bank of transmembrane proteins after 8 years. *Nucleic Acids Research*, Vol. 41, D1, 524-529. <https://doi.org/10.1093/nar/gks1169>
- Kumar S. D. M., Bava A. K., Gromiha M. M., Prakabaran P., Kitajima K., Uedaira H., Sarai A. (2006): ProTherm and ProNIT: thermodynamic databases for proteins and protein–nucleic acid interactions. *Nucleic Acids Research*, Vol. 34, 204-206. <https://doi.org/10.1093/nar/gkj103>
- Liber Texts: Basic Concepts in Membranes.* <https://bio.libretexts.org> (2018).
- Lodish H., Berk A., Zipursky SL. (2000): *Membrane Proteins*. In: *Molecular Cell Biology*, 4th edition, W. H. Freeman, New York, U.S.A. <https://www.ncbi.nlm.nih.gov/books/NBK21570/>
- Lomize A. M., Pogozheva D. I., Joo H., Mosberg I. H., Lomize L. A. (2011): OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Research*, Vol. 40, D1, 370-376. <https://doi.org/10.1093/nar/gkr703>
- Lúcio M., Lima J.L.F.C, Reis S. (2010): Drug-Membrane Interactions: Significance for Medicinal Chemistry. *Current Medicinal Chemistry* 17, 1795-1809. <https://doi.org/10.2174/092986710791111233>
- Mouritsen O. G., Andersen H. K., Andersen J. S., Davidsen J., Nielsen L. K., Jorgensen K. (2001): Structure of Liposomal Membranes in Relation to Permeation. In: *Pharmacokinetic Optimization in Drug Research: Biological, Physicochemical, and Computational Strategies*, Wiley-VCH, Weinheim, Germany, 33-34.
- Pilgrim M. (2010): *HTML5: Up and Running: Dive into the Future of Web Development*, O'Reilly Media, Inc., Sebastopol, California, U.S.A
- Palonciová M., DeVane R., Bruce M., Berka K., Otyepka M. (2014): Amphiphilic Drug-Like Molecules Accumulate in a Membrane below the Head Group Region. *The Journal of Physical Chemistry B* 118 (4), 1030-1039. <https://pubs.acs.org/doi/pdf/10.1021/jp4112052>
- Pop D. P., Altar A. (2014): Designing an MVC Model for Rapid Web Application Development. In: *Procedia Engineering* Vol. 69, 1172-1179. <https://doi.org/10.1016/j.proeng.2014.03.106>
- Protein Data Bank Contents Guide: Atomic Coordinate Entry Format Description.* [ftp://ftp.wwpdb.org/pub/pdb/doc/format\\_descriptions/Format\\_v33\\_A4.pdf](ftp://ftp.wwpdb.org/pub/pdb/doc/format_descriptions/Format_v33_A4.pdf) (10.4.2018).
- Qifang Xu, Dunbrack R.L. (2010): The protein common interface database (ProtCID) — a comprehensive database of interactions of homologous proteins in multiple crystal forms. *Nucleic Acids Research*, 761–770. <https://doi.org/10.1093/nar/gkq1059>
- Schultheiss J. S. (2011): Ten Simple Rules for Providing a Scientific Web Resource. *PLoS Comput Biol* 7(5): e1001126. <https://doi.org/10.1371/journal.pcbi.1001126>
- Sehnal D., Deshpande M., Vařeková R. S., Mir S., Berka K., Midlik A., Pravda L., Valenkar S., Koča J. (2017): LiteMol suite: interactive web-based visualization of large-scale macromolecular structure data. *Nature Methods* 14, 1121-1122. <http://dx.doi.org/10.1038/nmeth.4499>
- Seydel J.K., Coats E.A., Cordes H.P., Weise M. (1994): Drug Membrane Interactions and the Importance for Drug Transport, Distribution, Accumulation, Efficacy and Resistance. *Arch Pharm - Chemistry in Life Science*, 601-610 <https://doi.org/10.1002/ardp.19943271002>
- Szklarczyk D., Morris H. J., Cook H., Kuhn M., Wyder S., [...], von Mering Ch. (2016): The STRING database in 2017: quality-controlled protein–protein association networks, made

- broadly accessible. *Nucleic Acid Reserach*, Vol. 45, D1, 362-368.  
<https://doi.org/10.1093/nar/gkw937>
- Stein D. L. (2003): Integrating biological databases. *Nature Reviews Genetics* 4, 337-345.  
<https://doi.org/10.1038/nrg1065>
- Steinbeck Ch., Han Y., Kuhn S., Horlacher O., Luttmann E., Willighagen E. (2013), The Chemistry Development Kit (CDK): An Open-Source Java Library for Chemo- and Bioinformatics. *Journal of Chemical Information and Modeling* 43, 2, 493-500.  
<https://doi.org/10.1021/ci025584y>
- Šimková M. (2017): *Predikce interakce léčiv s modely biologických membrán*. Bakalářská práce, Univerzita Palackého v Olomouci, Olomouc, Česká Republika.
- The IUPAC International Chemical Identifier*: <https://iupac.org/who-we-are/divisions/division-details/inchi/> (10.4.2018).
- Van Meer G. Voelker D. R. Feigenson G. W. (2008): Membrane Lipids: where they are and how they behave. *Nature Reviews Molecular Cell Biology* 9, 112-124.  
<http://dx.doi.org/10.1038/nrm2330>
- Wilkinson M. D., Dumontier M., Aalbersberg J. I., Appleton G., [...], Mons B. (2016): The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3: 160018. <http://dx.doi.org/10.1038/sdata.2016.18>
- Woychowsky E. (2007): *AJAX: Creating Web Pages with Asynchronous JavaScript and XML*. Prentice Hall, New Jersey, U.S.A.
- Žára O. (2015): *JavaScript*. Computer Press, Brno, ČR.