

Univerzita Palackého v Olomouci

Bakalářská práce

Olomouc 2017

Štěpán Helmer

UNIVERZITA PALACKÉHO V OLMOUCI

Přírodovědecká fakulta

Katedra buněčné biologie a genetiky



Molekulové modelování UDP-glukuronosyltransferáz

BAKALÁŘSKÁ PRÁCE

Autor:	Štěpán Helmer
Studijní program:	B1406 Biologie
Studijní obor:	Molekulární a buněčná biologie
Forma studia:	Prezenční
Vedoucí práce:	doc. RNDr. Karel Berka, Ph.D.
Rok:	2017

Bibliografická identifikace

Autor: Štěpán Helmer

Název práce: Molekulové modelování UDP-glukuronosyltransferáz

Typ práce: Bakalářská

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

Rok obhajoby: 2017

Pracoviště: Katedra fyzikální chemie, UP v Olomouci

Abstrakt:

Náplní této bakalářské práce byla literární rešerše a zkoumání struktury UDP-glukuronosyltransferáz (UGT) *in silico* metodami. V teoretické části práce jsem shromáždil dosavadní informace o UGT, tedy o jejich významu při metabolismu xenobiotik, jejich názvosloví, o mechanismu glukuronizační reakce, kterou UGT zprostředkovávají, dále o struktuře UGT2B7, na kterou je tato práce zaměřena. V praktické části popisují programy, které jsem ve své práci použil. Ve výsledcích pak komentuji model UGT2B7 vytvořený programem I-TASSER. Ten jsem následně zanořil do membrány a ve formě monomeru i dimeru vložil do molekulové dynamiky. Následně jsem vyhledával dutiny vedoucí k aktivnímu místu. Do těchto dutin jsem poté pomocí molekulového dokování vkládal molekuly UDP-glukuronové kyseliny a ibuprofenu a počítal tak jejich afinitu k aktivnímu místu. V závěru jsem na základě výsledků vyvodil, jakým způsobem UGT dimerizují a proč.

Klíčová slova: UGT, homologní modelování, molekulárně dynamické simulace, molekulové dokování

Počet stran: 34

Jazyk: Český

Bibliographic identification

Author: Štěpán Helmer

Title: Molecular modeling of UDP-glucuronosyltransferases

Type of Thesis: Bachelor

Department: Department of Physical Chemistry, UP Olomouc

Supervisor: doc. RNDr. Karel Berka, Ph.D.

The Year of Presentation: 2017

Abstract:

The goal of this thesis is exploration of UDP-glucuronosyltransferase (UGT) structure by *in silico* methods. In theoretical part of thesis, I have assembled information about UGT, specifically about their importance in xenobiotic metabolism, their nomenclature, mechanism of glucuronization reaction and about structure of UGT2B7. In practical part, I have described applied programs. In result section, I have commented the model of UGT2B7, which has been generated by I-TASSER. I have embedded it to a membrane in form of monomer and dimer and analysed it with molecular dynamic simulations. Snapshots from the simulations were then used for active site cavity detection prior molecular docking. Lastly, I have used molecular docking of ibuprofen and UDP-glucuronic acid to detected cavities to measure their affinities toward active site. Finally, UGT dimerization is discussed.

Keywords: UGT, homology modelling, molecular dynamics simulations,
molecular docking

Number of Pages: 34

Language: Czech

Prohlašuji, že jsem bakalářskou práci vypracoval samostatně pod vedením doc. RNDr.
Karla Berky, Ph.D. s použitím uvedených literárních zdrojů.

V Olomouci dne.....

Štěpán Helmer

Poděkování:

Tímto bych chtěl poděkovat svému vedoucímu bakalářské práce doc. RNDr. Karlu Berkovi, za individuální přístup vedení práce. Dále bych chtěl poděkovat Mgr. Veronice Navrátilové za předávání zkušeností v oblasti molekulové dynamiky, dodávání učebních materiálů a trpělivost při konzultacích.

Obsah

Úvod	8
Teoretická část.....	9
Metabolismus xenobiobiotik	9
Fáze I 9	
Fáze II 9	
Fáze III 10	
UDP-glukuronosyltransferázy (UGT)	10
Funkce 10	
Reakce 10	
Rodiny UGT	12
Struktura UGT	12
Oligomerizace UGT	14
Důležité oblasti UGT	14
Metody.....	16
Databáze proteinů	16
Predikce membránového zanoření	16
Predikce proteinové struktury.....	17
Homologní modelování.....	17
Lokální pokus - EV fold + Modeller	17
Online pokus - I-TASSER.....	18
Molekulová dynamika	19
Dokování	20
Výsledky:.....	22
Predikce membránového zanoření	22
Predikce proteinové struktury.....	23
Molekulová dynamika	24
Dokování	27
Souhrn.....	31
Summary	32
Literatura	33

Úvod

UDP-glukuronosyltransferázy (UGT) jsou enzymy účastníci se metabolismu xenobiotik, tedy látek tělu cizích. Jsou to nejdůležitější enzymy druhé neboli konjugační fáze metabolismu léčiv. Přesto doposud nebyla odhalena jejich 3D struktura, protože se vyskytují problémy při jejich purifikaci a krystalizaci. V této práci jsem se proto věnoval konstrukci 3D modelu UGT2B7.

V teoretickém úvodu jsem se věnoval známým poznatkům o reakcích a struktuře rodiny UGT enzymů včetně důrazu na jejich membránovou orientaci, dimerizaci a lokalizaci významných strukturních motivů.

V praktické části jsem nejdříve určoval topologii membránového zanoření lidských UGT. Následně jsem vytvářel pomocí homologního modelování model monomeru UGT2B7, u něhož jsem zohlednil zjištěnou membránovou topologii. Poté jsem vytvořil model dimeru UGT2B7, kde jsem dal do blízké vzdálenosti dva monomery nasměřované k sobě oblastí, o které se předpokládá, že je zodpovědná za dimerizaci. Následně jsem provedl molekulárně dynamickou simulaci takto vytvořeného modelu monomeru i dimeru zanořeného v membráně obklopeného vodu s fyziologickou koncentrací sodných a chloridových iontů probíhající po dobu 500 ns. U monomeru i dimeru jsem poté vyhodnotil vznikání a přítomnost dutin uvnitř proteinu a do dutin následně vkládal molekuly UDP-glukuronové kyseliny a ibuprofenu pomocí molekulového dokování. Úspěšnost dokování jsem vyhodnocoval pomocí afinity a vzdálenosti funkčních skupin od katalytických aminokyselin. Pomocí tohoto pozorování jsem se pokusil vysvětlit, proč je potřebná dimerizace UGT2B7.

Teoretická část

Metabolismus xenobiobiotik

Jako xenobiotika označujeme pro organismus cizorodé látky, tedy látky, které nejsou tělu vlastní. Mohou to být kupříkladu léčiva či průmyslové látky. Aby se tyto cizorodé látky v organismu nehromadily, tak často procházejí biotransformací. Biotransformace, nebo také metabolismus xenobiobiotik je jedním ze základních farmakokinetických procesů, k nimž patří také absorpce, distribuce látky do jednotlivých buněk organismu a jeho konečná exkrece (pozměněné) látky z organismu ven. Při biotransformaci může dojít k bioaktivaci nebo naopak biodeaktivaci xenobiobiotika. Metabolismus xenobiobiotik probíhá nejčastěji v jaterních buňkách a dá se zhruba rozdělit do tří fází.

Fáze I

Při reakcích I. fáze se do chemické struktury xenobiobiotika zavádí funkční skupina, nebo je pozměněna funkční skupina léčiva, nejčastěji pomocí oxidačních, redukčních či hydrolytických procesů. Nejvýznamější enzymový systém I. fáze se nazývá monooxygenázový systém s hlavními enzymy - cytochromy P450. Monooxygenázový systém obsahuje NADPH-cytochrom P450-reduktázu, NADPH⁺, molekulární kyslík a cytochrom P450. V cytochromu P450 se nachází železitý ion vázaný v hemu. Po vazbě na léčivo (D) se změní konformace železitého iontu do vysokospinového stavu, čímž je umožněna následná redukce Fe³⁺ na Fe²⁺ z NADPH⁺ na cytochrom P450 pomocí elektronů dodaných NADPH-cytochrom P450 reduktázou. Za přítomnosti molekulárního kyslíku se pak vytvoří komplex Fe²⁺O₂DH, jenž přijímá proton a druhý elektron. Vytvoří se peroxidový komplex Fe²⁺OOH^{*}DH. Druhý proton rozštěpí tento komplex na vodu a komplex (FeO)³⁺DH. Vodíkový atom je pak z tohoto komplexu vytažen za vzniku volné radikálové molekuly léčiva D^{*}, jenž s navázaným komplexem (OH)⁻ tvoří hydroxylované léčivo, které se z proteinu uvolní za současné regenerace cytochromu P450.¹

Fáze II

II. fáze metabolismu je fáze konjugativní. Při konjugativní fázi metabolismu dochází k vytvoření kovalentní vazby mezi funkční skupinou xenobiobiotika a konjugačním činidlem. Enzymy katalyzující tuto reakci se nazývají transferázy a jsou pojmenovávány podle reakce, kterou způsobují. Mezi enzymy konjugační fáze se řadí sulfotransferázy, glykosyltransferázy, acetyltransferázy a metyltransferázy. Řadí se sem i UDP-glukuronosyl transferázy, které jsou předmětem této bakalářské práce.

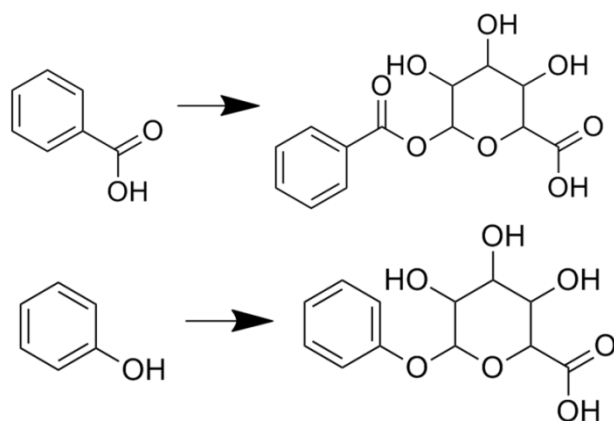
Fáze III

III. fáze metabolismu je fáze eliminační. Zprostředkovávají ji membránové transportéry skrze aktivní transport xenobiotik přes buněčnou membránu. Nejvýznamnější rodinou proteinů III. fáze metabolismu je ATP-binding cassette (ABC). Ty zprostředkovávají transport endogenních látek, jako jsou anorganické ionty, kovové ionty, aminokyseliny proteiny, ale i velké hydrofobní molekuly přes buněčné ale i intrabuněčné membrány za spotřeby ATP. Patří mezi ATP-pumpy. Jsou součástí buněk ledvin, jater, placenty a střev. Mezi ně patří i takzvané P-glykoproteiny (P-Gp), které jsou zodpovědné za rezistenci rakovinných buněk proti chemoterapeutikám.

UDP-glukuronosyltransferázy (UGT)

Funkce

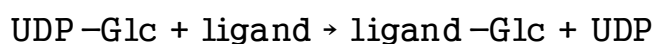
UDP-glukuronosyltransferáza (UGT) je enzym, nacházející se v endoplazmatickém retikulu obratlovců. Řadí se do skupiny enzymů zvaných glykosyltransferázy, které obecně připojují molekulu cukru k substrátu. UGT odpovídá za proces, zvaný glukuronidace, kdy nejprve dojde k rozštěpení UDP-glukuronové kyseliny na UDP a glukuronovou kyselinu, která se kovalentně naváže k substrátu (viz Obrázek 1). Substrát se připojením glukuronové kyseliny stane hydrofilnějším. Následně může být vyloučen ven z těla ve formě moči nebo žluči. Díky UGT tak dochází k detoxifikaci od mnohých odpadních látek organismu, jako je bilirubin, mastné kyseliny, ale i k detoxifikaci od xenobiotik, především léčiv.²



Obrázek 1: Ukázka výsledku glukuronidační reakce benzoové kyseliny, či fenolu.

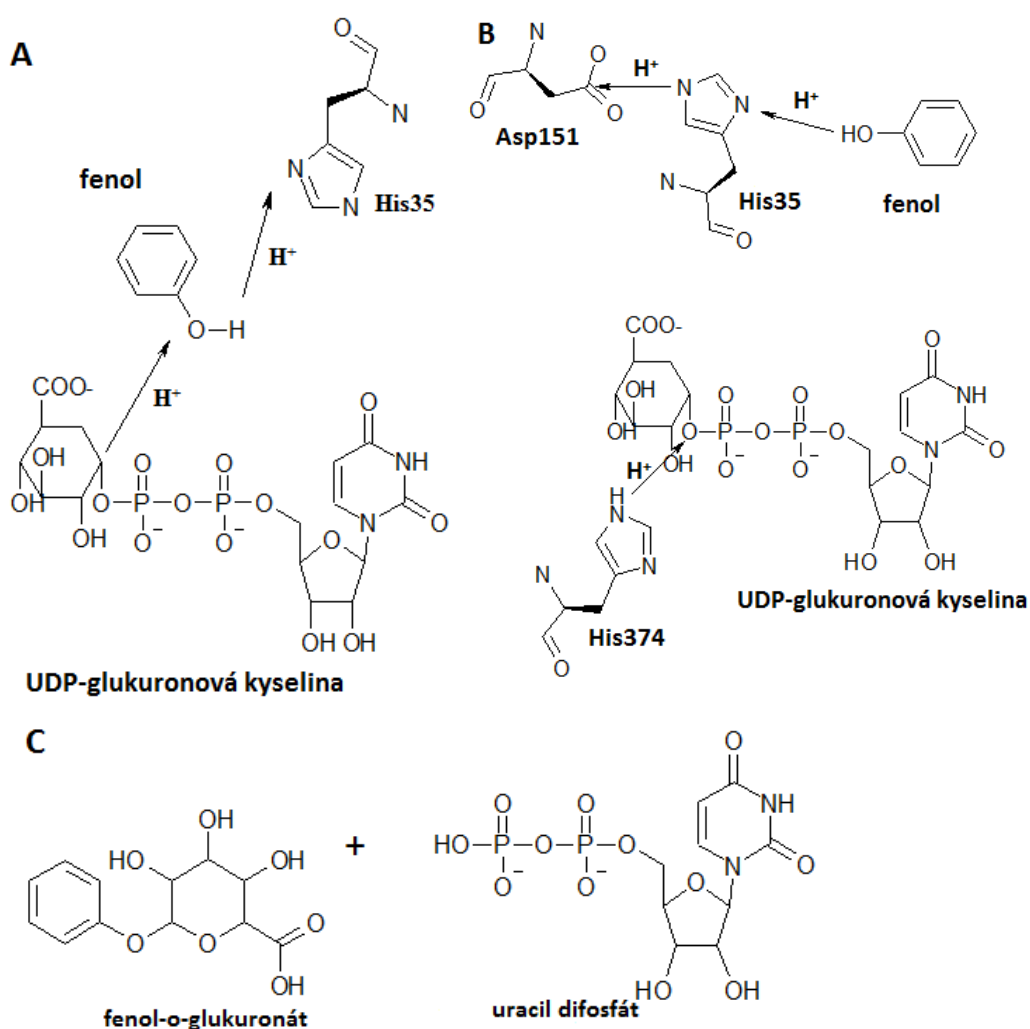
Reakce

Doposud byly navrženy dva reakční mechanismy (viz Obrázek 2) glukurodinace pomocí UGT dle souhrnné reakce³:



U prvního mechanismu dochází nejdříve k deprotonaci hydroxylové skupiny substrátu na aminokyselině His35 a následuje nukleofilní útok na UDP-glukuronovou kyselinu, čímž dojde k rozštěpení molekuly na UDP a glukuronovou kyselinu, která se naváže uhlíkem C1 na deprotonovaný kyslík substrátu.

Při druhém reakčním mechanismu dochází k interakci hydroxylové skupiny substrátu s aminokyselinou His35, které je odebrán vodík aminokyselinou Asp151. UDP-glukuronová kyselina je mezitím hydrogenizována aminokyselinou His374, čímž dojde k rozrušení vazby UDP a glukuronové kyseliny, která se následně kovalentně váže k substrátu (viz Obrázek 2).



Obrázek 2: Mechanismus glukurodinace-vlevo nahoře (A) 1. model glukurodinace, vpravo nahoře (B) 2. model glukurodinace, dole (C) výsledek glukurodinace

Rodiny UGT

UGT je řazeno do rodin a podrodin. Pokud mají UGT společných více než 45% procent shody v sekvenci aminokyselin, jsou součástí jedné rodiny. Pokud mají společných více než 65% procent shody, jsou součástí jedné podrodiny. Rodina UGT se značí číslicí za UGT (například UGT1). Podrodina se značí jako písmeno za číslovkou (např. UGT2B). Jednotlivé UGT jsou pak značeny číslem za písmenem podrodiny (například UGT1A1).⁴Lidské UGT má známých 16 různých UGT (v době sepsání knihy Introduction to a drug metabolism) 8 UGT enzymů je kódováno geny pro UGT1A a 7 s UGT2 geny. UGT1A se nachází v 2-q37 chromozómu.⁵

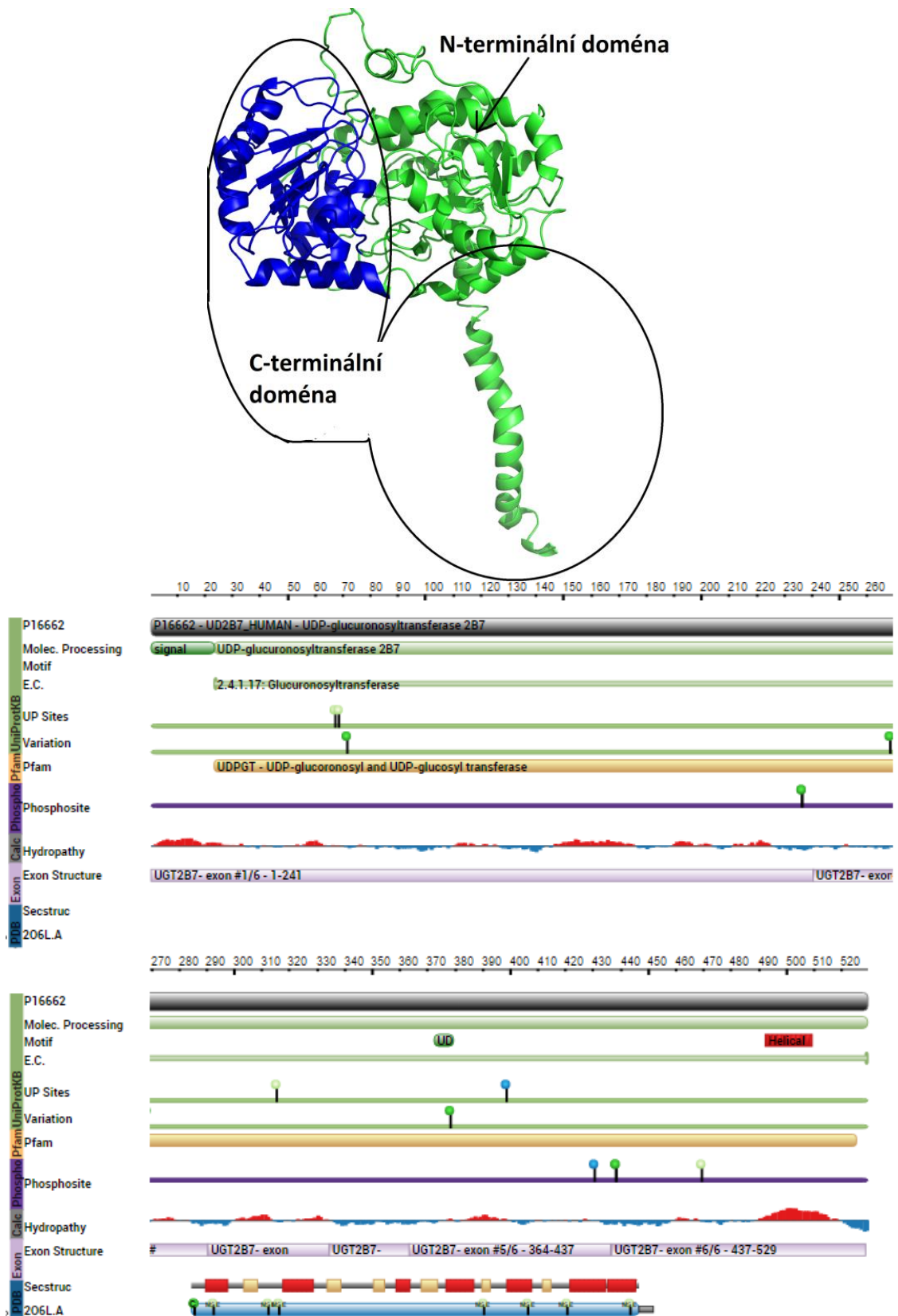
Jednotliví zástupci UGT se výrazně liší svou enzymovou aktivitou a specificitou vůči různým třídám ligandů (viz Tabulka 1).

Tabulka 1: Maximální enzymová aktivita jednotlivých lidských UGT v pmol/min/mg proteinu⁵

Substráty	1A1	1A3	1A4	1A6	1A8	2A1	2B4	2B7	2B15	2B17
Jednoduché fenoly	1900	239	30	2400	5300	735	0.4	5	167	38
Bilirubin	400	0	2	0	0	nd	0	0	0	0
Karboxylové kyseliny	0	121	0	nd	170	68	0	2	0	nd
Primární aminy	1	84	540	10600	1800	22	nd	3	0	nd
Opioidy	0	130	0	0	0	73	0	3462	0	nd

Struktura UGT

UGT patří do skupiny enzymů zvaných glykosyl transferázy (GT), konkrétně do GT-B nadrodiny glykosyl transferáz. O té je známo, že obsahuje dvě domény s Rossmanovým motivem (N-terminální a C-terminální domény). Domény s Rossmanovým motivem se skládají ze “sendviče” α -helix/ β -skládaný list/ α -helix a obecně se nachází u enzymů vázajících nukleotidy. Stejně tak i UGT se skládá ze dvou domén s Rossmanovým motivem. Aktivní místo GT-B se nachází na smyškách mezi těmito doménami. Doposud se nepodařilo experimentálně získat strukturu celého UGT. Důvodem je, že je příliš náročné UGT izolovat a následně purifikovat, jelikož se jedná o transmembránový protein. Přesto byl v roce 2007 uveřejněn článek o krystalové struktuře C-terminální domény UGT2B7.⁶ Z homologního modelování jiných výzkumných týmů vzešlo v souladu s našimi výsledky, že kolem celého proteinu UGT (přesněji UGT1A1 a UGT1A10) jsou dva obalové (envelope) α -helixy, nacházející se v sekvenci těsně před transmembránovým helixem. (viz Obrázek 3).^{7,8} Předpokládá se, že u C-konce se nachází vazebné místo pro UDP-glukoronovou kyselinu, zatímco N-doména je zodpovědná za vazbu substrátu.⁹



Obrázek 3: Navrhovaná struktura UGT s vyznačenými doménami. (Nahoře) Modře vyznačená část byla převzata z krystalové struktury ⁶ (PDB ID: 2O6L) a zbývající části byly dobudovány pomocí homologního modelování (viz výsledková část). (Dole) Pokrytí sekvence strukturálními a dalšími anotacemi (převzata z RCSB.org)

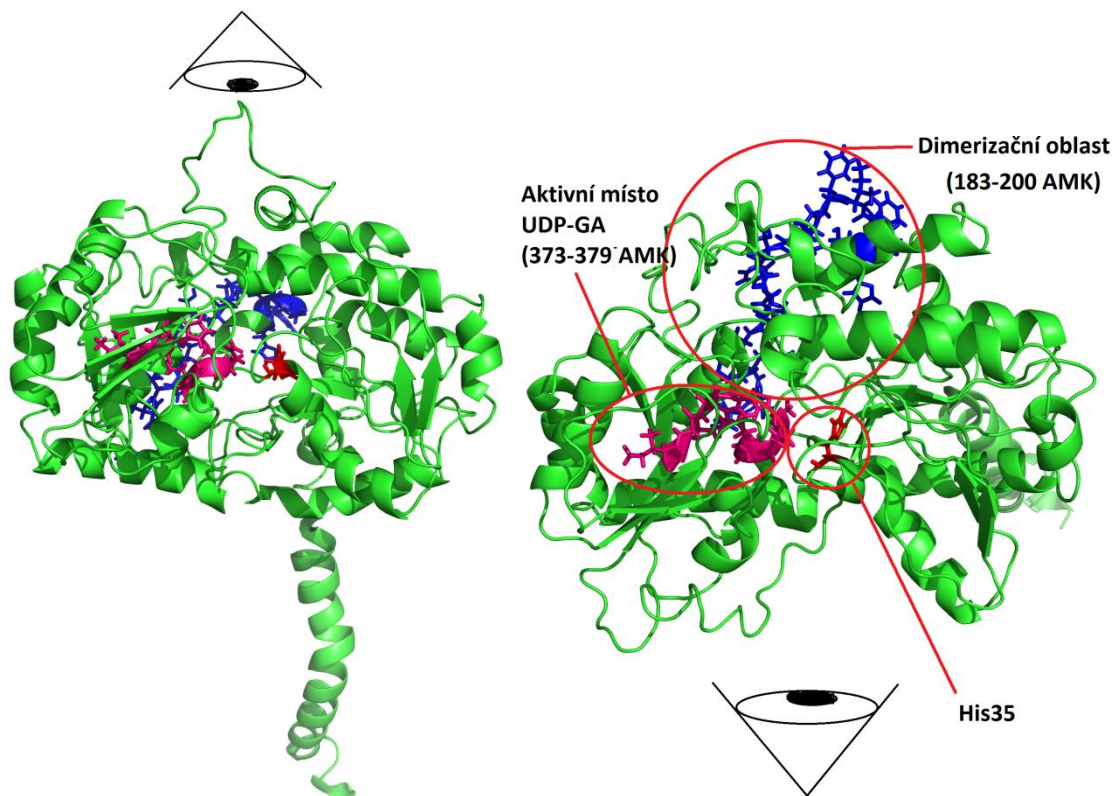
UGT je k membráně upevněno C-terminální α -helikální transmembránovou kotvou, přičemž katalytická doména se nachází v lumenu endoplazmatického retikula.¹⁰ Nejvyšší enzymovou aktivitu enzym vykazuje v membráně složené z dioleoylfosfatidylcholinu (DOPC),¹¹ proto jsme pro simulaci proteinu vygenerovali membránu složenou právě z tohoto fosfolipidu.

Oligomerizace UGT

Za pomoci metody FRET (frekvenčního resonančního energického transferu) bylo zjištěno, že UGT1A proteiny dimerizují (nebo vytvářejí tetramery), a to buďto v rámci stejného proteinu (homodimerizace) nebo tvoří dimery s proteiny ze stejné podrodiny proteinů (heterodimerizace). Dále byla prokázána homodimerizace proteinu UGT2B1, který dimerizuje i s proteiny z rodiny UGT1A. Bylo zjištěno, že za interakci v dimeru UGT2B1 je zodpovědná N-terminální doména¹². Předpokládá se, že pro dimerizaci je zásadní sekvence aminokyselin 183-200, jelikož je tato sekvence vysoce konzervativní u UGT1 i UGT2 rodin.¹³ Dále bylo zjištěno pomocí metody FRET, že pro dimerizaci jsou důležité hlavně aminokyseliny His268 a Ala71.¹⁴ Otázkou nicméně i nadále zůstává důvod, jakou funkci dimerizace UGT má.

Důležité oblasti UGT

Kromě dimerizační domény obsahuje UGT také aktivní místa pro substrát a pro kofaktor, tedy UDP-glukuronovou kyselinu. Aktivní místo pro UDP-glukuronovou kyselinu se na UGT2B7 nachází mezi aminokyselinami 373-379. Aminokyselina Gly382 váže ribosu-OH.⁶ Důležitá je také vzájemná pozice aminokyselin Arg338 a Gly379, které se vážou na uracilovou bázi a na fosfátovou skupinu (v tomto pořadí) a drží tak glukuronovou kyselinu v prostoru. V blízké vzdálenosti od sebe se nachází aminokyseliny His35 a Asp151 (na His35 dochází dle jednoho z reakčních mechanismů k deprotonaci). Aktivní místo pro fenolické ligandy se nachází na aminokyselinách 158-194. Podle Coffmana se opiáty (např. morfin) vážou v oblasti 84-118.¹⁵ Z posttranslačních úprav je u UGT2B7 známá pouze glykosylace aminokyselin Asn67, Asn68 a Asn315.



Obrázek 4: Vyznačené pozice důležitých aminokyselin zepředu a shora

Metody

Databáze proteinů

Pro práci s proteiny jsem se nejdříve musel naučit používat proteinové databáze. Jednou z nich je UNIPROT, kde shromážděny všechny sekvence proteinů, fragmenty proteinů, informace o zanoření proteinů v membráně, informace o posttranslačních modifikacích, o ligandech enzymů a názvy 3-D struktur, které byly zjištěny pomocí rentgenové krystalografie. (<http://www.uniprot.org/>).

Z této databáze jsme získali sekvenci proteinu UGT2B7, kterou jsme vkládali do programů uvedených dále k vytvoření modelu 3-D struktury proteinu (UNIPROT ID: P16662):

```

      10          20          30          40          50
MSVKWTSVIL LIQLSFCFSS GNCGKVLVWA AEYSHWMNIK TILDELIQRG
      60          70          80          90         100
HEVTVLASSA SILFDPNNS ALKIEIYPTS LTKTELENFI MQQIKRWSDL
      110         120         130         140         150
PKDTFWLYFS QVQEIMSIFG DITRKFCCKDV VSNKKFMKKV QESRFDVIFA
      160         170         180         190         200
DAIFPCSELL AELFNIPFVY SLSFSPGYTF EKHSGGFIFP PSYVPPVMSE
      210         220         230         240         250
LTDQMTFMER VKNMIYVLYF DFWEIFDMK KWDQFYSEVL GRPTTLSETM
      260         270         280         290         300
GKADVWLIRN SWNFQFPHPL LPNVDFVGGL HCKPAKPLPK EMEDFVQSSG
      310         320         330         340         350
ENGVVVFSLG SMVSNMTEER ANVIASALAQ IPQKVLWRFD GNKPDTLGLN
      360         370         380         390         400
TRLYKWIPQN DLLGHPKTRA FITHGGANGI YEAIYHGIPM VGIPLFADQP
      410         420         430         440         450
DNIAHMKARG AAVRVDFNTM SSTDLLNALK RVINDPSYKE NVMKLSRIQH
      460         470         480         490         500
DQPVKPLDRA VFWIEFVMRH KGAKHLRVAA HDLTFWQYHS LDVIGFLLVC
      510         520
VATVIFIVTK CCLFCFWKFA RKAKKGKND
```

Další databází je Protein data bank in Europe (<http://www.ebi.ac.uk/pdbe/>) (PDBe).¹⁶ Zde se nachází všechny experimentálně určené 3D struktury proteinů a jsou zde volně ke stažení. Základem tvorby modelu byla struktura C-domény UGT2B7 (PDB ID: 2O6L)

Predikce membránového zanoření

Ve své bakalářské práci jsem se nejdříve učil používat programy, pro odhalování membránového zanoření proteinů k určení topologie proteinu.

Prvním takovým programem pro určování polohy na membráně je TMHMM.¹⁷ TMHMM používá skrytý Markovovský model (HMM) pro predikci transmembránových oblastí. Zde se zadá sekvence proteinu a program vyhodnotí, s jakou pravděpodobností je sekvence na lumenální straně, bazální straně či jestli je zanořená v membráně. HMM je založený na vázaných markovovských pravděpodobnostech, které byly natrénovány na známých transmembránových sekvencích. TMHMM dokáže s vysokou pravděpodobností určit transmembránovou oblast i topologii proteinu.

Dalším predikčním programem pro určení zanořených segmentů proteinu do membrány je SOSUI.¹⁸ Program svou predikci zakládá na třech parametrech. Prvním je hydropatie aminokyselin (rozpuštěnost v polárních rozpouštědlech), vycházející z Kyle Doolittleho indexu. Poté je vypočtena distribuce elektrického náboje okolo helixu. Nakonec je do výpočtu zahrnuta délka sekvence. Výstupem programu je diagram, který ukazuje, kde se nachází predikovaný transmembránový segment.¹⁹

Predikce proteinové struktury

Homologní modelování

K predikci struktury proteinů se používají 3 metody. Jsou to homologní modelování, metoda navlékání (threading) a ab initio metody. V této bakalářské práci byly použity pouze metody homologního modelování.

Nejvýkonnější a nejpřesnější z výše zmíněných metod je metoda homologního modelování. Homologní modelování predikuje třírozměrné struktury proteinů na základě známých homologů z PDB. U homologního modelování závisí kvalita modelu na podobnosti sekvencí hledaného proteinu se sekvencemi templátových homologních proteinů, k nimž známe i strukturu.

Při homologním modelování se tak nejdříve srovná hledaná proteinová sekvence se známými sekvencemi strukturních templátů. Pomocí zarovnání se pak vyberou a nahradí segmenty (většinou smyčky) které musí být změněny kvůli rozdílům mezi hledanou sekvencí a templátovými sekvencemi. Poté se nahradí postranní řetězce a struktura se následně optimalizuje pomocí molekulové mechaniky.

Lokální pokus - EV fold + Modeller

EV fold²⁰ je predikční program, které k sobě přiřazuje sekvencně podobné proteiny na základě evolučně spjatých oblastí proteinů. Výstupem je seznam struktur, které mají se studovaným proteinem podobnou strukturu na základě geometrických kritérií.

Bohužel je veřejně přístupná verze EV fold (<http://evfold.org/evfold-web/evfold.do>) omezená maximální velikostí modelovaného proteinu na 300 aminokyselin. Proto jsem sekvenci rozdělil na tři části: první zahrnovala sekvenci

1-240, druhá 200-400 a třetí 350-529. Segmenty obsahovaly společné sekvence, aby mohlo dojít ke spojení jednotlivých částí proteinů na základě porovnávání (alignmentu).

Toto porovnání bylo provedeno v programu UCSF Chimera²¹ (<https://www.cgl.ucsf.edu/chimera/>) s pomocí nástroje Modeller 9.15. Modeller je nástroj pro komparativní modelování proteinové struktury. Výstupem je sada modelů. Tyto modely jsme zkontrolovali z pohledu přítomnosti predikovaných sekundárních struktur.

Online pokus - I-TASSER

I-TASSER²² (Iterative Threading ASSamble) je webový program (<http://zhanglab.ccmb.med.umich.edu/I-TASSER/>), založený na paradigmatu od sekvence-do struktury-do funkce. Výstupem programu je kompletní sekundární struktura a 5 modelů terciální struktury. U každé struktury je napsána i hodnota C-score, která vyjadřuje přesnost, s jakou byl model vytvořen. Dále jsou sepsány 3D struktury, které jsou svou sekvencí nejpodobnější zkoumané sekvenci, a ze kterých modely struktury vycházejí (viz tabulka 5).

Tabulka 4: Sekvence sloužící jako templát pro I-TASSER

Rank	PDB Hit	TM-score	RMSD ^a	IDEN ^a	Cov
1	2acwA	0.718	2.49	0.138	0.766
2	2vceA	0.703	2.71	0.170	0.762
3	3wc4A	0.695	3.03	0.148	0.767
4	3hbjA	0.690	2.95	0.153	0.766
5	2c9zA	0.689	2.94	0.142	0.760
6	2pq6A	0.668	2.83	0.188	0.730
7	3othA	0.645	3.27	0.153	0.728
8	2iyaA	0.616	3.70	0.172	0.709
9	1rrvA	0.611	4.08	0.161	0.720
10	3rscA	0.610	3.75	0.157	0.707

PDB Hit-proteinové struktury, použité k homolognímu modelování

TM-skóre- míra globální strukturní podobnosti templátové a zkoumané sekvence.

RMSD^a-RMSD (střední kvadratická odchylka atomových pozic) mezi zbytky srovnávaných struktur

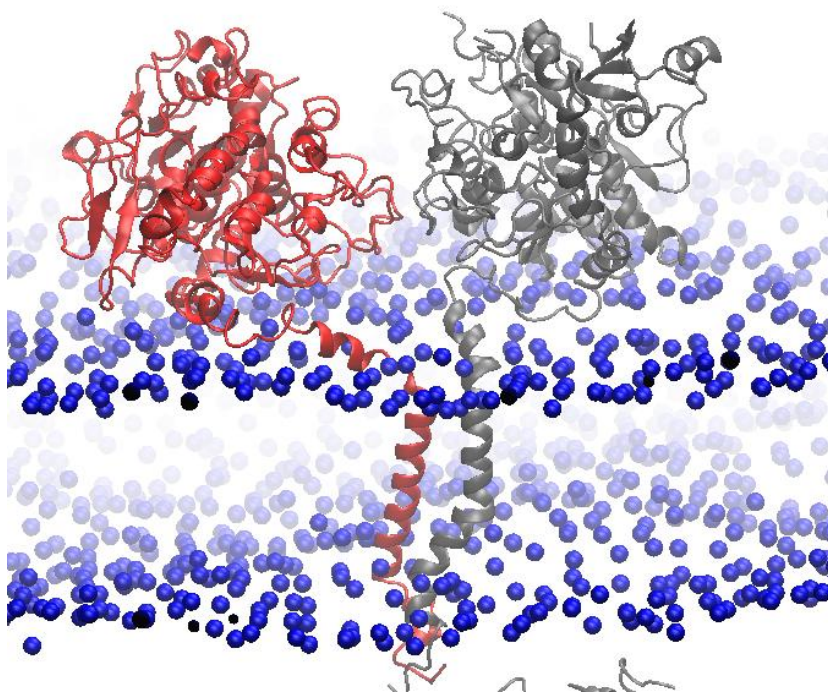
IDEN^a-procentuální sekvenční shoda templátové a zkoumané sekvence

Cov-poměr počtu strukturně přiřazených aminokyselin a celkové délky proteinu (v aminokyselinách)

Pomocí I-TASSERu jsme vygenerovali monomer UGT2B7. Zadali jsme jeho sekvenci a jako templát jsme zvolili strukturu 2o6l z proteinové databanky. Podobně jako v předchozím případě jsme zkontrolovali přítomnost predikovaných sekundárních struktur.

Problém byl pouze s transmembránovým helixem, který byl pokroucený a modelovaný jako náhodné klubko (random coil). Proto jsme vytvořili transmembránový helix s požadovanou sekvencí, tedy sekvencí 480-527 v programu PyMOL. A následně jsme oba modely (transmembránový helix i model z I-TASSERu) spojili do finálního modelu monomeru pomocí programu Modeller.

Následně bylo potřeba vytvořit i dimer UGT, jelikož jsem chtěl testovat chování monomeru a dimeru a pokusit se přijít na důvod, proč musí být UGT v interakci, aby byly enzymaticky aktivní. Dvě identické molekuly UGT jsem k sobě nasměroval tak, aby se vzájemně dotýkaly, a to již zmíněnými konzervativními sekvencemi mezi 183. a 200. aminokyselinou.



Obrazek 5: Dimer zanořený v membráně. Červeně je znázorněna sekvence 183-200, která podle dříve uvedených studií zodpovídá za dimerizaci proteinu

Molekulová dynamika

Poté jsem spouštěl simulace tohoto modelu pomocí simulačního balíku GROMACS 4.5.7 (Spoel, Lindahl, Hess, & Groenhof, 2005). Vytvářel jsem simulaci monomeru i dimeru zanořeného ve fosfolipidové dvouvrstvě. Ta byla složená z 500 molekul DOPC a byla získána z předchozích simulací na katedře fyzikální chemie. Monomer i dimer byl do membrány zasazen v programu PyMOL přibližně doprostřed této čtvercové plochy membrány tak, aby se transmembránový helix nacházel v membráně. Takto byl vytvořen soubor .pse.

Z toho byl vypreparován protein i membrána ve formátu .pdb, které byly následně samostatně převedeny do formátu .gro čitelném pro program GROMACS pomocí nástroje editconf v případě membrány a pdb2gmx v případě proteinu. Následně byly soubory .gro proteinu i membrány spojeny v jeden soubor .gro. Na konci souboru .gro se nachází rozměry simulačního boxu a tato velikost musela být změněna na 13,5 x 13,5 x 16,5 nm u monomeru a 13,5 x 13,5 x 16 nm u dimeru. Poté byl použit nástroj g_membed, který je používán na zanořování proteinů do membrány. Nástroj funguje tak, že se proteinu zmenšil v osách x a y na 0 a postupně se v těchto osách zvětšoval až do své původní velikosti a tím postupně vytlačoval, případně odstraňoval přecházející lipidy. Následně bylo volné místo v boxu, kde se nenacházela membrána ani protein doplněno molekulami vody nástrojem genbox. Poté bylo spočítáno množství molekul vody a část z nich byla pomocí nástroje genion nahrazena 173 ionty Na^+ a 177 Cl^- u monomeru a 152 ionty Na^+ a 160 Cl^- u dimeru. Molekul Na^+ a bylo o 4 méně u monomeru a o 8 méně u dimeru, aby byl celý obsah boxu celkově elektricky neutrální. Molekulově dynamická simulace se skládala z několika kroků. Nejprve byla provedena energetická minimalizace na 100 000 kroků pomocí konjugovaných gradientů tak, aby se zabránilo artefaktům při spuštění simulace. Následně byla provedena ekvilibrace po dobu 10 000 ns, ve které byl atomy páteře proteinu udržován na svých místech a nakonec byla provedena volná molekulárně dynamická simulace o celkové délce 500 ns.

Obecné parametry simulací byly následující - délka simulačního kroku byla 2 fs, semiizotropní tlak byl udržován pomocí metody Parrinello a Rahmana na 1 bar s časovou konstantou 10 ps a teplota byla udržována termostatem V-rescale na teplotě 310 K s časovou konstantou 0,5 ps.

Dokování

Dokování je metoda, pomocí níž vkládáme do modelu makromolekuly (v mém případě UGT2B7) molekulu ligandu (UDP-GA a ibuprofen či morfin). Pro dokování simulovaného proteinu je potřeba mít snímky simulace (v mém případě po každé ns, tedy 500 snímků monomeru a dimeru), ve kterých se nachází pouze protein.

Dále je potřeba vědět, kam se molekula ligandu váže. U enzymů se toto místo nachází v hydrofobní dutině proteinu. Lze je vizualizovat například pomocí PyMolu. Důležitým kritériem pro zhodnocení významnosti těchto dutin pro mne byla doba, po kterou byly tyto dutiny v simulaci přítomné. Dalším kritériem bylo, aby vedli k aktivnímu místu pro UDP-GA a zároveň procházeli v blízkosti aminokyseliny His35 (kde dochází k deprotonaci substrátu) a aby i toto aktivní místo bylo součástí této dutiny. Ke zjištění souřadnic takových dutin jsem používal Autodock Tools.

Pomocí programu Autodock Tools ²⁴ jsem zjistil oblast v boxu, kde se mají molekuly dokovat a také, při porovnávání velikosti gridovacího boxu s velikostí molekul zjistil, jak má být gridovací box velký. Autodock potřebuje pro zaměřování mít strukturu proteinu vypreparovanou od ostatních molekul. Pro zaměřování polohy využívá molekuly ve formátu pdbqt. Molekulu proteinu si může převést do formátu

pdbqt převést sám ze souboru formátu pdb, kdy si do struktury sám doplní vodíky a náboje v případě proteinu a navíc určí rotovatelné vazby v případě dokovaného ligandu.

Následně byla molekula UDP-GA a molekula ibuprofenu dokována pomocí programu Autodock Vina²⁵ do míst určených souřadnicemi z Autodock Tools. Vina, podobně jako Autodock Tools potřebuje, aby byla dokovaná molekula ligandu ve formátu pdbqt. Kromě dokovaných ligandů a makromolekul, do kterých jsou ligandy dokovány a jejich souřadnic (centrováno na pozici His35) je potřeba programu zadat počet modelů (v mém případě 10), které mají být výstupem programu a také velikost boxu (v mém případě 32x32x32), v rámci něhož dochází k hledání energeticky nejvýhodnější pozice ligandu. Při dokování jsou zkoušeny všechny konformace ligandu, ve kterých se může nacházet v rámci dutiny vymezené dokovacím boxem. Výstupem jsou výsledné pózy molekuly ligandu ve formátu pdbqt, které je možné zobrazit například v Pymolu současně s makromolekulou. Takto lze vizualizovat, jestli se ligand navázal do předpokládaného místa. Výstupem je také tabulka s hodnotami afinity jednotlivých póz udávaných v kcal/mol.

Výsledky:

Predikce membránového zanoření

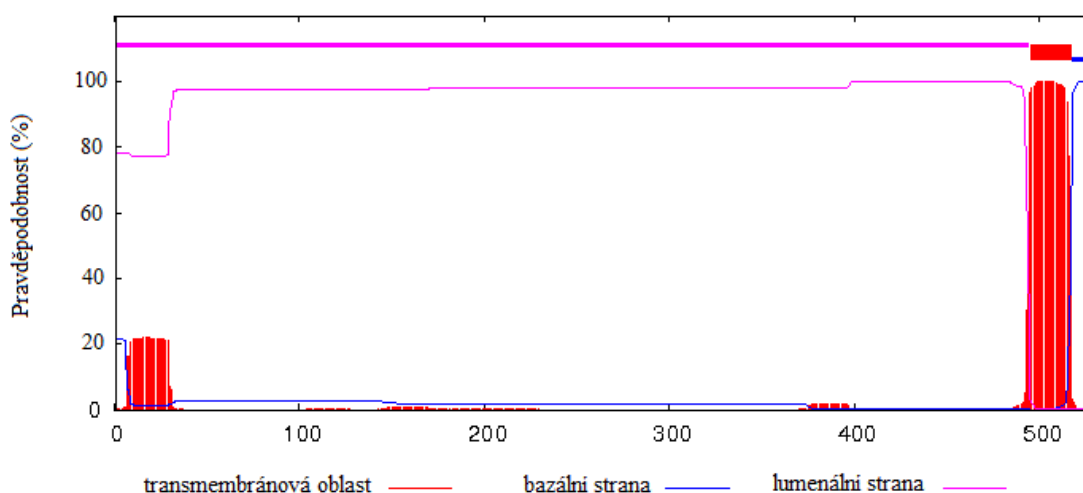
Pomocí SOSUI a TMHMM jsme zjišťovali, zda některé UGT nevytvářejí více transmembránových helixů, které by společně vytvářely membránový kanálek, kterým by putovali metabolity skrze membránu k UGT, které by je zpracovalo. Pro zhodnocení jsme zadali do SOSUI a TMHMM vždy jednu sekvenci z UGT1A, UGT1B, UGT2A a UGT2B. Výsledky jsou zaznamenány v tabulce (viz tabulka č. 5).

Tabulka 5: Transmembránové sekvence lidských UGT dle TMHMM a SOSUI

UGT	uniprot	Program	sekvence	helix	Sekvence TM sekvence
1A10	E9PXN7	TMHMM	488-510	_	IGFLLAIVLTVVFIVFKCCAYG
		SOSUI	486-508	Primary	LDVIGFLLAIVLTVVFIVFKCCA
			8-30	secondary	TSLPLCVCLLLASGLVQAGRLLV
1B2	D3XD61	TMHMM	496-518	_	IGFMLLIVLIVTLAMLKCCSLCW
		SOSUI	1-23	primary	MRLGVIFILVVGVCSSAAVDQKP
			159-181	secondary	PMGAILAYNLSVPAVYLLRGMAC
			494-516	primary	DVIGFMLLIVLIVTLAMLKCCSL
2A5	B8JMA2	TMHMM	6-26	_	GALLVCLLLCGVDVGWSGRVLVM
			211-233	_	MLLYIVHSIAFPLVATFTFDGYY
			491-513	_	AAFLLSITALITFLWVKACCFLE
		SOSUI	490-512	primary	VAAFLLSITALITFLWVKACCFLE
			2-24	primary	TSGALLVCLLLCGVDVGWSGRVL
2B7	P16662	TMHMM	495-517	_	GFLLCVATVIFIVTKCCLFCF
		SOSUI	1-23	secondary	MSVKWTSVILLIQLSFCFSSGNC
			493-515	primary	VIGFLLCVATVIFIVTKCCLFC

Obě predikční metody identifikovaly transmembránový segment v blízkosti C-konce, který svou délkou odpovídá pozici právě jednoho transmembránového helixu. Sekvence transmembránového helixu začíná obecně u lidských UGT mezi 486 a 496 aminokyselinou, končí mezi 508 a 518 aminokyselinou a je dlouhá 23 aminokyselin.

Oproti TMHMM vyhodnocovalo SOSUI častěji jako transmembránovou oblast i počáteční N-koncovou oblast, kde se nachází signální peptid. Ten tedy zodpovídá za vstupní zanoření proteinu do membrány endoplazmatického retikula a po převodu sekvence do lumenu ER je posléze odštěpen. Takže přestává být součástí proteinu a nemůže být součástí membránového kanálku.



Obrázek 6: Membránová topologie UGT2B7 dle TMHMM

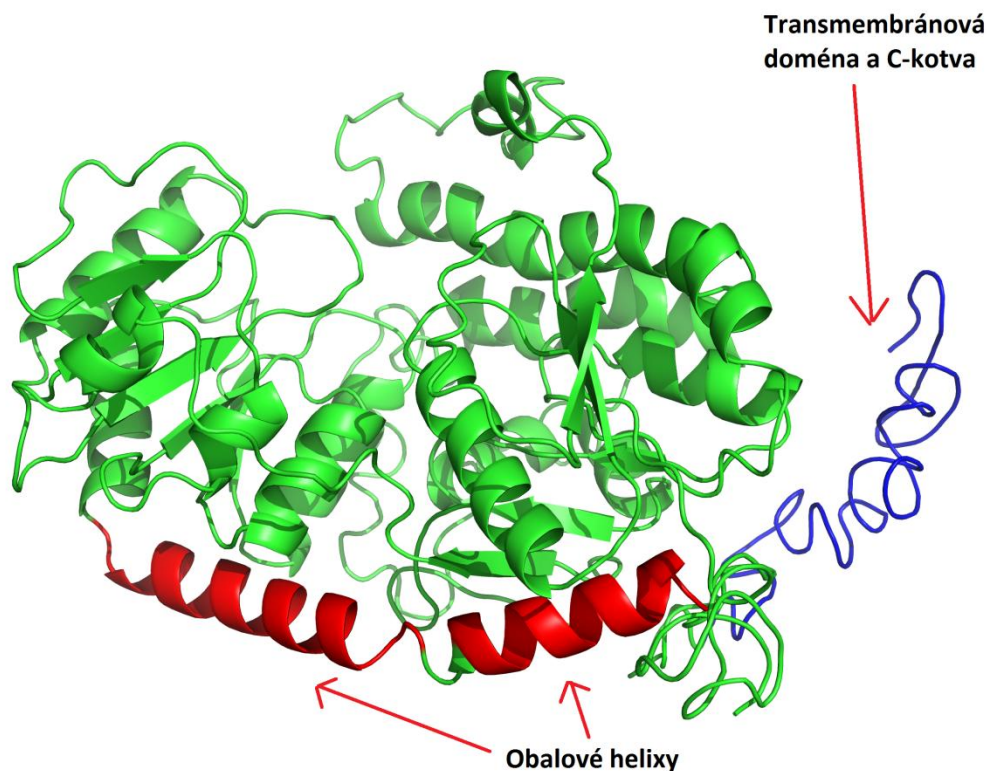
Predikce proteinové struktury

EV-fold + Modeller

Pomocí programu EV-fold vyhledal homologní sekvence proteinů, bohužel však nebylo nalezeno dostatečné množství struktur, které by odpovídaly svými sekvencemi N-terminální doméně. Proto ze struktur proteinů, které sloužily jako templát pro program Modeller nešlo odvodit strukturu, která by odpovídala struktuře glykosyltransferáz B a měla odpovídající sekundární struktury.

I-TASSER

I-TASSER se ukázal vhodnější, jelikož kromě homologního modelování využil i *ab initio* techniky a vygeneroval tak vhodnější strukturu, která lépe odpovídala generické struktuře příbuzných GT-B glykosyltransferáz, tedy je složený ze dvou domén s Rosmannovým motivem. Dále má dva envelope helixy. Na C-konci byla ale struktura nepodobná transmembránovému helixu (viz obrázek 7), kterou jsme proto domodelovali z ideálního helixu a spojili v programu Modeller (viz metody).



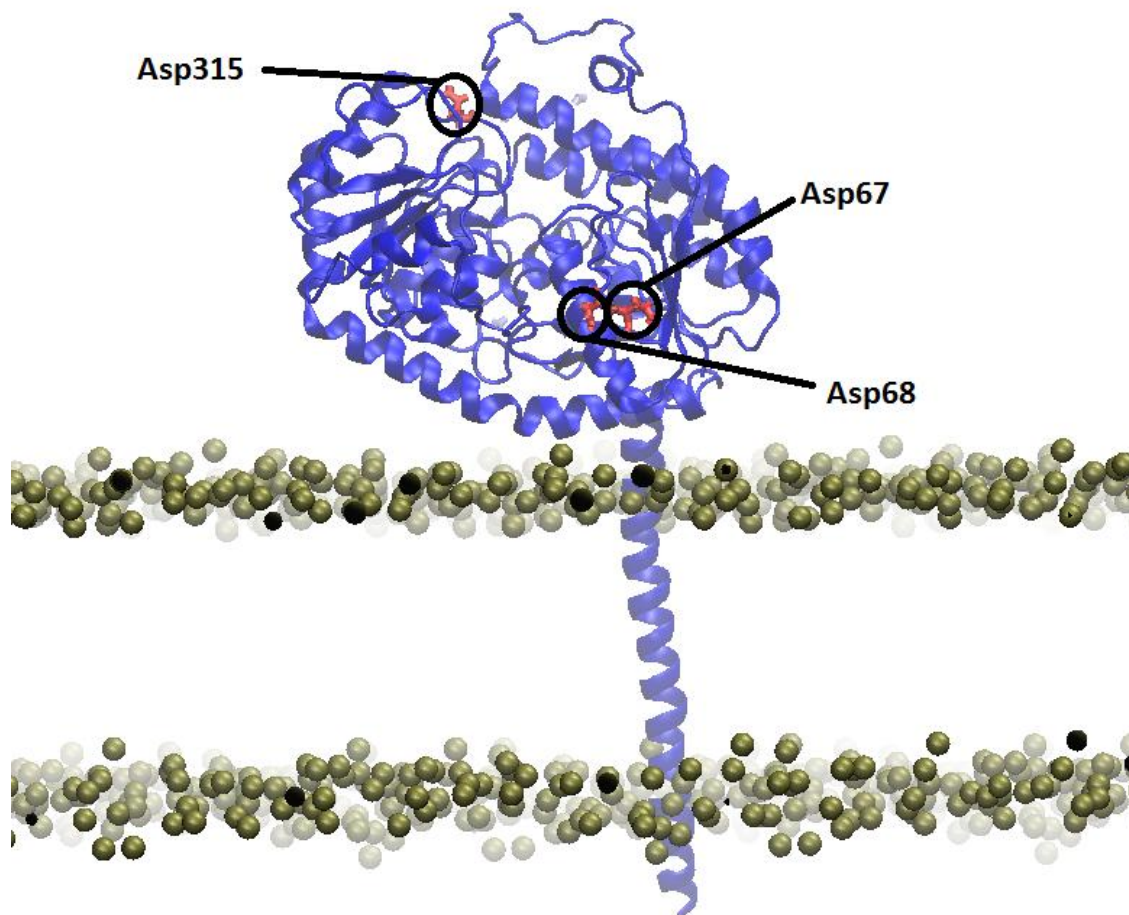
Obrázek 7: Původní model vygenerovaný pomocí programu I-TASSER

Aktivní místo pro UDP-glukuronovou kyselinu se nachází společně s aminokyselinou His35, která způsobuje deprotonaci substrátu, ve vzájemné blízkosti, což je zásadní pro samotnou glukurodinaci. Zároveň se tato místa nacházejí na smyčkách mezi doménami, což odpovídá umístění vazebných míst u GT-B glykosyltransferáz. Aminokyselina His35 se ve shodě s experimentálními daty nachází také v blízkosti aminokyseliny Asp151.⁶

Nasměrování na membránu jsme řídili podle pozice glykosylovaných aminokyselin. Ty jsou totiž hydrofilní, a proto by se glykosylované aminokyseliny Asn67, Asn68 a Asn315 měly nacházet na povrchu proteinu mimo membránu, což náš model proteinu UGT2B7 splňuje (viz Obrázek 8) a jeho strukturu jsme dále ekvilibrovali pomocí molekulové dynamiky.

Molekulová dynamika

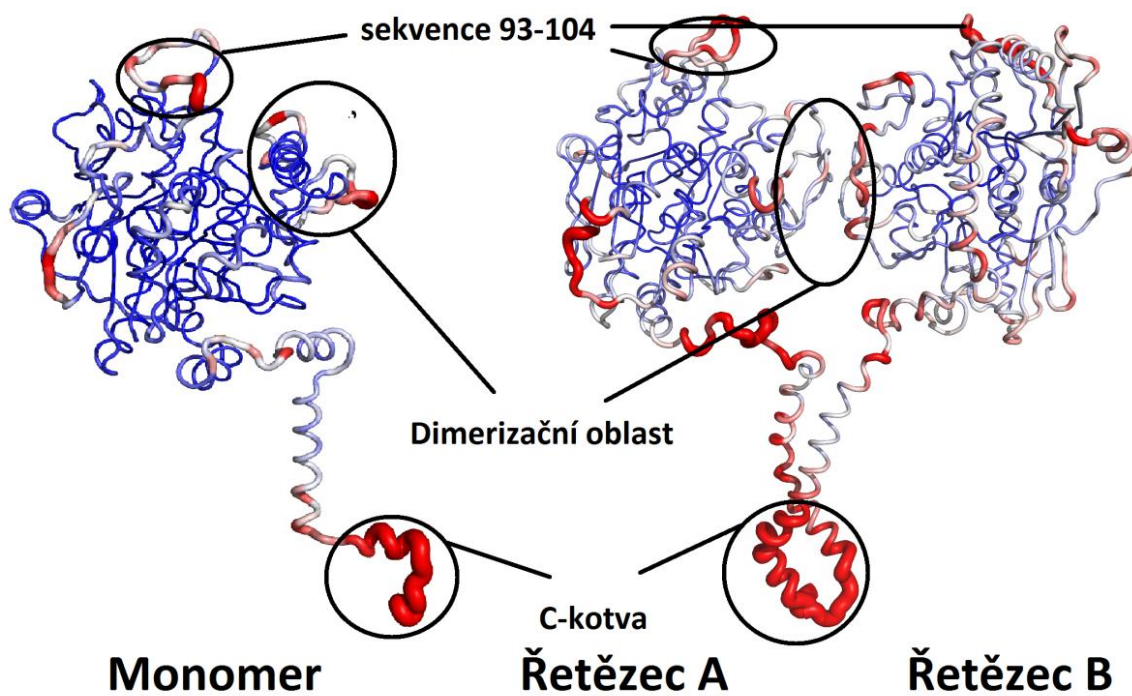
Monomer byl po dobu simulace většinou své struktury, nacházející se na lumenu endoplazmatického retikula poměrně stabilní. Domény se vůči sobě neposouvaly a ležely přibližně stejnou plochou na membráně a celkově byla sekundární i terciární struktura zachována, přičemž flexibilnější úseky byly smyčky kolem aminokyselin 93-104, či v dimerizační oblasti, dále pak transmembránový helix a nejvíc C-kotva (viz Obrázek 9).



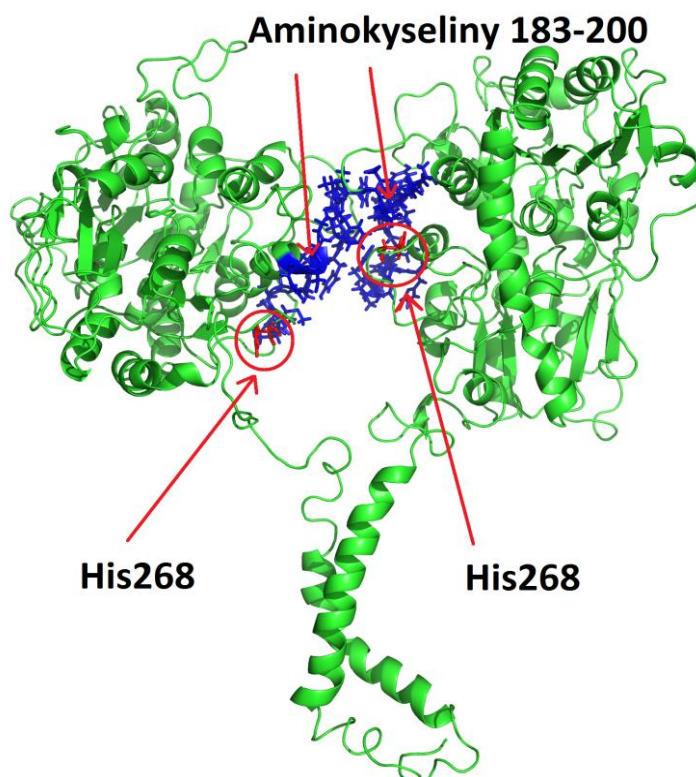
Obrázek 8: Monomer zanořený do membrány DOPC

U dimeru zůstala struktura proteinu nacházející se na lumenální straně endoplazmatického retikula taktéž zachována. Ale transmembránové helixy jsou v dimeru oproti monomeru více flexibilní (u řetězce A více než u řetězce B) což je zřejmě důsledek interakce obou helixů. C-kotvy jsou opět nejflexibilnější částí (stejně jako v případě monomeru). Naopak k drobnému potlačení flexibility došlo v dimerizační oblasti (viz obrázek 9).

Po simulaci se u dimeru projevilo, že kromě zmíněných aminokyselin 183-200 byla ve vzájemné interakci s dimerizační oblastí i aminokyselina His268, o které bylo experimentálně pomocí FRETu zjištěno, že je důležitá pro správnou funkci dimeru, a to u druhého řetězce.¹⁴ Přitom blízkost této aminokyseliny k místu dimerizace byla jinak zpočátku kontroverzní, protože v symetrickém modelu dimeru by byla od dimerizační oblasti příliš daleko. To mi bylo dostatečným důkazem, že jsou k sobě molekuly dimeru nasměřovány správně.

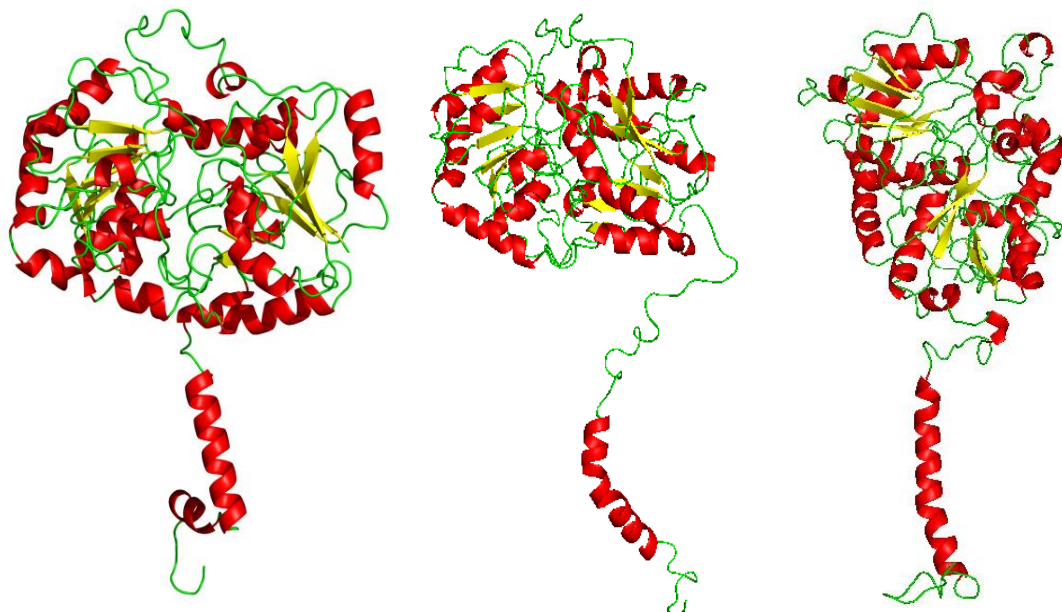


Obrázek 9: Znázornění flexibility proteinových zbytků. Vlevo je monomer a vpravo dimer. Modrá barva značí stabilní oblasti, bíle jsou středně stabilní oblasti a červeně nestabilní oblasti.



Obrázek 10: Dimer s označením aminokyselin 183-200 a His268

V průběhu simulace došlo nicméně k vychýlení jednoho řetězce, takže je jeden ze řetězců na konci simulace nasměrován téměř kolmo na membránu (viz obrázek 11).

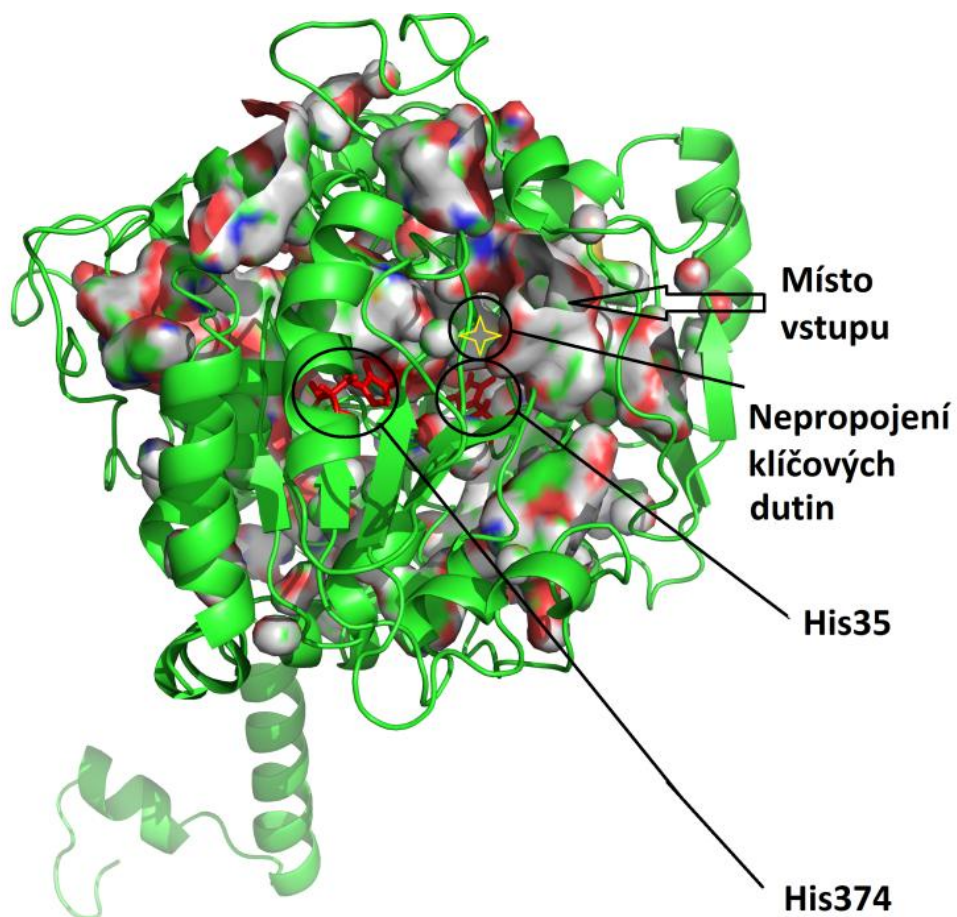


Obrázek 11: Výsledek molekulové dynamiky monomeru a dimeru po 500 ns. Nalevo monomer, uprostřed řetězec 1 a napravo řetězec 2, který je orientován oproti prvnímu řetězci kolmo na membránu.

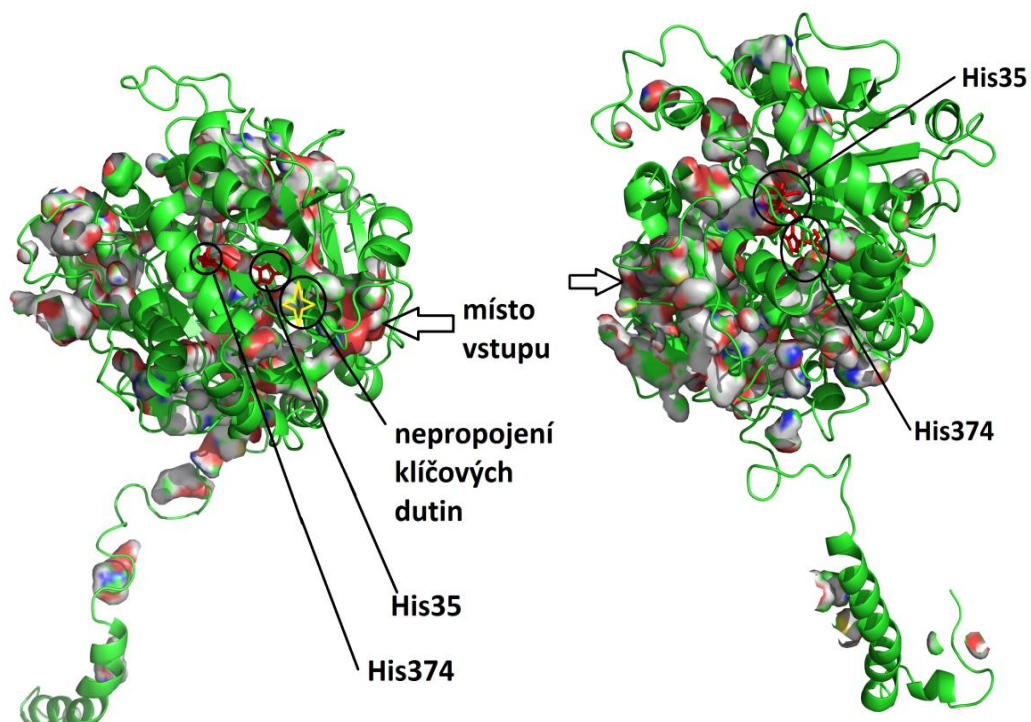
Dokování

Aktivní místo u monomeru je poměrně uzavřené a do tamní dutiny se nedaří vkládat žádnou látku. Dutiny vznikající u prvního řetězce dimeru se velmi podobají dutinám vznikajícím u monomeru. Dutiny se během dynamiky vytvářejí poměrně pravidelně na stejných místech a někdy se vytvářejí od povrchu proteinu až k aktivnímu místu pro UDP-GA, nicméně tato vznikající dutina je příliš úzká pro průchod UDP-GA či ligandu. Oproti tomu dutiny, vznikající u druhého řetězce dimeru jsou odlišné. V posledních 80 ns vzniká dutina vedoucí až k aktivnímu místu. Tato dutina je zároveň dostatečně široká pro průchod UDP-GA a umožňuje dokování. (Obrázek 12)

Při dokování do dutin bylo taky jediné úspěšné dokování právě do dutiny druhého řetězce, dokování monomeru a prvního řetězce dimeru probíhalo pouze na povrch proteinu a nikoliv do aktivního místa, i když jsem použil 500 snímků z molekulové dynamiky pro jednotlivé řetězce. Zde ukazují výsledky pro pózu nejlépe odpovídající navrhovaným mechanismům reakce.



Monomer

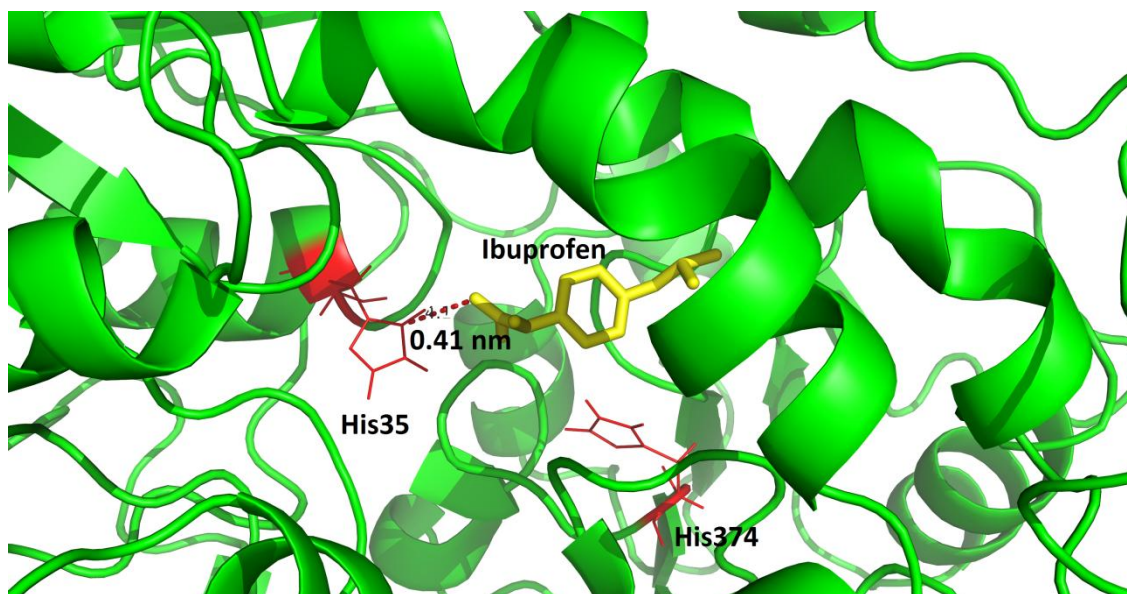


Dimer

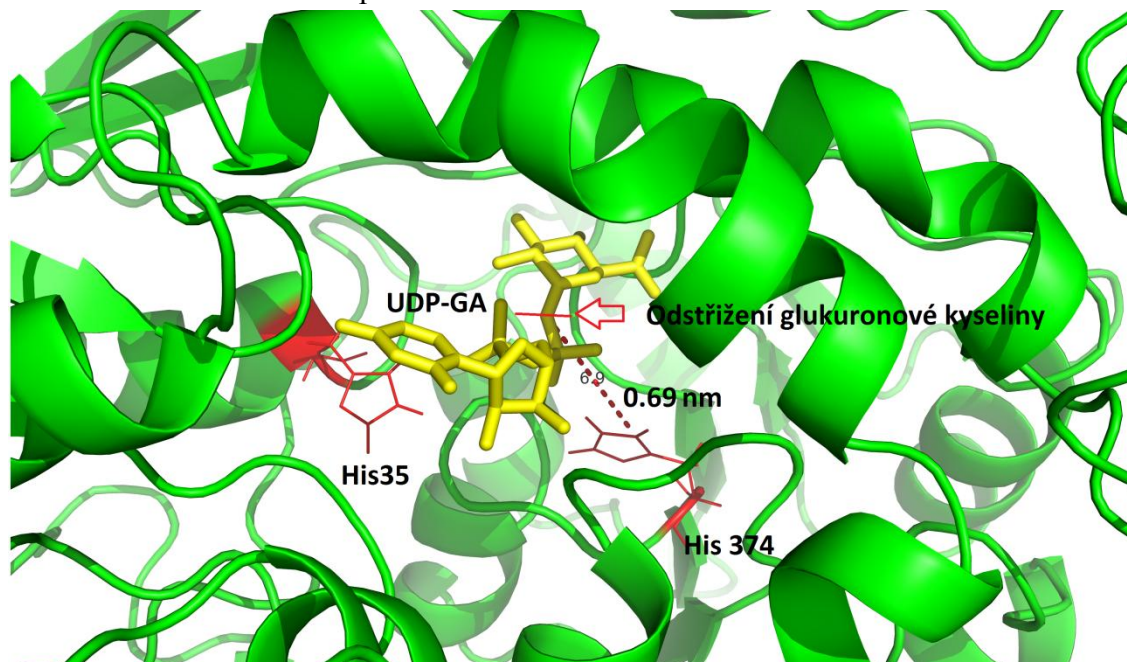
Obrázek 12: Dutiny a hydrofobní kapsy u monomeru a dimeru. Kruhem je označena pozice aktivního místa.

Afinita obou molekul k aktivnímu místu druhého řetězce byla vysoká. Pro dokování ibuprofenu jsem zobrazil 7. pozici (Obrázek 13), protože tato pozice měla dostatečně vysokou afinitu (-8,1 kcal/mol) a zároveň se nacházela pouhých 0,42 nm od aminokyseliny His35. Pro následné dokování UDP-GA a celkovému znázornění reakce jsem použil 2. pózu o afinitě -10,8 kcal/mol, neboť v této póze se UDP-GA nejvíce přiblížilo His374 (na vzdálenost 0,57 nm; Obrázek 14).

Přestože jsem nemohl dokovat současně UDP-GA a ibuprofen, lze z výsledků dokování usoudit, že by ke glukuronidaci substrátu (v tomto případě ibuprofenu) mohlo snadno docházet.



Obrázek 13: Dokování ibuprofenu do UGT



Obrázek 14: Dokování UDP-GA do UGT.

Úspěšným dokováním se tedy podařilo potvrdit, že aminokyselina His35 může navázat ligand - ibuprofen. Současně v našem modelu mezi Asp151 a His35 nedochází k dostatečné interakci, takže předpokládáme mechanismus, kdy nejdříve dojde k deprotonaci karboxylové kyseliny na substrátu - ibuprofenu - následovaný nukleofilním atakem na UDP-glukuronovou kyselinu, čímž dojde k rozštěpení molekuly na UDP a glukuronovou kyselinu, která se naváže uhlíkem C1 na deprotonovanou karboxylovou skupinu ibuprofenu.

Výsledkem této práce tedy je, že UGT2B7 dimerizuje, ale každý řetězec dimeru má jinou funkci. Zatímco druhý řetězec je natočený kolmo na membránu, aby mohla vznikat dutina potřebná pro glukurodinaci, druhý řetězec leží vodorovně na membráně a nevznikají u něj dutiny potřebné ke glukurodinaci a jeho úloha je podpůrná, tedy aby držel druhý řetězec v prostoru, který by bez prvního řetězce byl v nestabilní poloze. Kromě vznikajících dutin tomu nasvědčuje i úspěšné dokování a také skutečnost, že aminokyselina His268, která je podle analýz FRET důležitá pro dimerizaci i samotnou glukurodinaci je v bezprostřední blízkosti dimerizační oblasti právě u druhého řetězce

Souhrn

UDP-glukuronosyltransferázy (UGT) jsou jedněmi z nejdůležitějších biotransformačních enzymů v lidských buňkách. Provádějí biotransformace xenobiotických sloučenin jako například opioidů, ibuprofenu, nebo i endogenních látek jako například bilirubinu. UGT jsou lokalizovány na lumenální straně membrány endoplasmatického retikula jaterních a ledvinových buněk.

V této práci jsem připravil homologní membránový model nejdůležitějšího zástupce této třídy enzymů - UGT2B7. Následně jsem tento model otestoval a ekvilibroval v molekulárně dynamické simulaci jak v monomerní, tak v dimerní formě na modelové membráně. Dokováním ibuprofenu a UDP-glukuronové kyseliny do jednotlivých připravovaných struktur jsem zjistil, že funkční aktivní místo obsahuje pouze jeden z řetězců účastníci se dimeru. Zdá se tedy, že experimentálně potvrzená nutná dimerizace UGT by mohla být potřebná pro katalytickou funkci komplexu, kdy jeden proteinový řetězec v dimeru by mohl mít funkci stabilizujícího prvku lépe interagujícího s membránou a druhý řetězec by mohl mít funkci katalytickou.

Podobný případ v třídě GT-B glykosyltransferáz kdy jednotlivé řetězce mají v multimeru rozdílné funkce byl popsán například u cytosolické O-glykosyltransferázy organismu *Streptococcus pneumoniae*²⁶ a u proteinkinázy koronaviru SARS byl zaznamenán případ, kdy je jeden řetězec dimeru nemá katalytickou funkci, zatímco druhý ano.²⁷

Summary

UDP-glucuronosyltransferases (UGT) are one of the most important detoxifying enzymes of human cells. They metabolize xenobiotic compounds like opioids and ibuprofen or endogenous compounds like bilirubin. UGT are located in the luminal side of the endoplasmatic reticulum membrane of kidney and liver cells.

In this thesis, I have prepared the homology model of the most important enzyme in this family - UGT2B7 on the model membrane. The model was then tested and equilibrated in molecular dynamics simulations in a monomeric or dimeric forms embedded to model membrane. Molecular docking of ibuprofen and UDP-glucuronic acid into individual snapshots revealed that active site is opened just in the one of chains of the dimer. We can thus hypothesize that experimentally identified UGT dimerization could be needed for the function of catalytic complex. Within such dimeric complex, the first chain should have a membrane-stabilization function and the other one should serve as a catalytic centre.

Literatura

1. Lincová, D. & Farghali, H. *Základní a aplikovaná farmakologie*. (Galén, 2007).
2. Csala, M., Bánhegyi, G. & Benedetti, A. Endoplasmic reticulum: A metabolic compartment. *FEBS Lett.* **580**, 2160–2165 (2006).
3. Magdalou, J., Fournel-gigleux, S. & Ouzzine, M. Insights on membrane topology and structure / function of UDP-glucuronosyltransferases. **42**, 159–166 (2010).
4. Plant, N. *Molecular Toxicology*. (BIOS Scientific Publishers, 2003).
5. Gibson, G. & Skett, P. *Introduction to Drug Metabolism*. (Nelson Thornes, 2001).
6. Miley, M. J. *et al.* Crystal Structure of the Cofactor-Binding Domain of the Human Phase II Drug-Metabolism Enzyme. *J. Mol. Biol.* 498–511 (2007). doi:10.1016/j.jmb.2007.03.066
7. Taylor, P., Tripathi, S. P., Prajapati, R., Verma, N. & Sangamwar, A. T. Predicting substrate selectivity between UGT1A9 and UGT1A10 using molecular modelling and molecular dynamics approach. *Mol. Simul.* 37–41 (2015). doi:10.1080/08927022.2015.1044451
8. Laakkonen, L. & Finel, M. A Molecular Model of the Human UDP-Glucuronosyltransferase 1A1 , Its Membrane Orientation , and the Interactions between Different Parts of the Enzyme □. *Mol. Pharmacol.* 931–939 (2010). doi:10.1124/mol.109.063289.
9. Meech, R. & Mackenzie, P. I. Brief review structure and function of uridine diphosphate glucuronosyltransferases. *Clin. Exp. Pharmacol. Physiol.* 907–915 (1997).
10. Marquardt, H., Schäfer, S., McClellan, R. O. & Welsch, F. *Toxicology*. (Academic Press, 1999).
11. Zakimbs, D., Cantor, M. & Eiblb, H. Phospholipids and UDP-glucuronosyltransferase. *J. Biol. Chem.* (1988).
12. Meech, R. & Mackenzie, P. I. UDP-Glucuronosyltransferase , the Role of the Amino Terminus in Dimerization *. *J. Biol. Chem.* **272**, 26913–26917 (1997).
13. Lewis, B. C., MacKenzie, P. I. & Miners, J. O. Homodimerization of UDP-glucuronosyltransferase 2B7 (UGT2B7) and identification of a putative dimerization domain by protein homology modeling. *Biochem. Pharmacol.* **82**, 2016–2023 (2011).
14. Yuan, L., Qian, S., Xiao, Y., Sun, H. & Zeng, S. Homo- and hetero-dimerization of human and its allelic variants affect zidovudine glucuronidation activity. *Biochem. Pharmacol.* **95**, 58–70 (2015).
15. Coffman, B. L., Kearney, W. R., Goldsmith, S., Knosp, B. M. & Tephly, T. R. Opioids Bind to the Amino Acids 84 to 118 of UDP- Glucuronosyltransferase UGT2B7. *Mol. Pharmacol.* **63**, 283–288 (2003).
16. Best, C. *et al.* PDBe : Protein Data Bank in Europe. *Nucleic Acids Res.* **39**, 402–410 (2011).
17. Krogh, A., Larsson, È., Heijne, G. Von & Sonnhammer, E. L. L. Predicting Transmembrane Protein Topology with a Hidden Markov Model : Application to Complete Genomes. *J. Mol. Biol.* 567–580 (2001). doi:10.1006/jmbi.2000.4315

18. Hirokawa, T., Boonéchieng, S. & Mitaku, S. SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinforma. Appl. note* **14**, 378–379 (1998).
19. Baxevanis, A. D. & Oullette, F. B. F. *Bioinformatics A Practical Guide to the Analysis Genes and Proteins*. (John Wiley & Sons, Inc, 2005).
20. Marks, D. S., Hopf, T. A. & Sander, C. Protein structure prediction from sequence variation. *Nat. Biotechnology* **30**, 1072–1080 (2012).
21. Pettersen, E. F. *et al.* UCSF Chimera — A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* 1605–1612 (2004). doi:10.1002/jcc.20084
22. Ambrish, R., Alper, K. & Yang, Z. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5**, 725–738 (2011).
23. van der Spoel, D., Lindahl, E., Hess, B. & Groenhof, G. GROMACS: Fast, Flexible, and Free. *Wiley Intersci.* 1701–1718 (2005). doi:10.1002/jcc.20291
24. Morris, G. M. *et al.* AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **30**, 2785–2791 (2010).
25. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J. Comput. Chem.* **31**, 455–461 (2011).
26. Zhang, H. *et al.* New Helical Binding Domain Mediates a Glycosyltransferase Activity of a Bifunctional Protein. *J. Biol. Chem.* **291**, 22106–22117 (2016).
27. Chen, H. *et al.* Only One Protomer Is Active in the Dimer of SARS 3C-like. *J. Biol. Chem.* **281**, 13894–13898 (2006).