



Vaistinių molekulių saveikos vietų identifikavimo ir vertinimo funkcijos kūrimas

Projekto „Mokslo pieva“ ataskaita

Tyrimą atliko:

Asta Lučiunaitė

Viktorija Rubytė

Paulina Amšiejūtė

Povilas Matusevičius

Mentoriai:

Vladas Oleinikovas

Vytautas Rafanavičius

Piotras Cimperman

Vilnius, 2016

TURINYS

ĮVADAS	3
Duomenų bazė	4
Parametrų atranka vaistinių molekulių sąveikos su taikiniu nustatymui.....	6
Programos, naudojamos baltymų kišenėms aptikti	10
Atrinktų parametrų apskaičiavimas	11
Kišenės paieška ir skaičiavimai	12
GHECOM programos struktūrinė analizė	14
Rezultatų apibendrinimas ir rekomendacijos	15
Literatūra	17

ĮVADAS

Šiuolaikiniam moksle vis rečiau matomi atsitiktinės sėkmės atvejai, o naujovės tai dažniausiai kryptingas daugelio metų darbas. Šie dėsningumai ypač galioja vaistų kūrimui, ligai turinčio įtakos baltymo ir jo slopiklių identifikavimas yra ilgas ir sudėtingas procesas. Dar sunkiau surasti tokias molekules (ligandus), kurie slopintų konkretų baltymą, tačiau nepaveiktų kitų organizme funkcionuojančių baltymų veiklos. Viską praktiškai išbandyti yra neįmanoma, todėl kuriami įvairūs metodai ir įrankiai, kaip ši procesą paspartinti.

Vienas iš jų – kompiuterinis įrankis, gebantis identifikuoti baltymo jungimosi vietą ir molekulę, galinčia prisijungti ir modifikuoti baltymo aktyvumą. Šios programos kuriamos ir nuolatos taikomos įvairiose farmacijos įmonėse ir institutuose, tačiau patirtis rodo, kad tik maža dalis, jų pateikiamų, molekulių sėkmingai prisijungia prie baltymų ir juos slopina. Šios programos turi trūkumų ir atsižvelgia ne į visus kriterijus, todėl pateikia neteisingus pasiūlymus. Tačiau tokio įrankio kūrimas nuolatos vykdomas, nes tai padėtų sutaupyti pinigų ir laiko farmacininkams bei galėtų išgydyti ar palengvinti įvairias ligas sergantiems.

Darbo tikslas:

Sukurti priemones įrankiui, kuris rastų vaistinių molekulių sąveikos su baltymais vietas.

Darbo uždaviniai:

1. Sukurti baltymų ir jų kompleksų su ligandais (vaistinėmis molekulėmis) bei juos apibūdinančių parametrų duomenų bazės prototipą.
2. Parinkti baltymo jungimosi vietą apibūdinančius parametrus.
3. Sukurti jungimosi vietą apibūdinančius parametrų skaičiavimo algoritmus.
4. Parinkti ir pritaikyti atviro kodo programas tinkamas ligando-taikinio jungimosi vietų aptikimui ir jų parametrų apskaičiavimui.

Duomenų bazė

Išsami, nuolatos atnaujinama ir nešališka baltymų ir ligandų duomenų bazė padeda identifikuoti fermentų funkcijas ir klasifikuoti baltymus (Lopez et al., 2007). Tačiau tai taip pat yra būtinas resursas, ieškant vaistinių molekulių modifikuojančių ligas sukeliančių fermentų aktyvumą (Langer and Hoffman, 2006). Tokios duomenų bazės sukūrimas, vartotojų prieigos tobulinimas ir priežiūra reikalauja daug pastangų. Su ta pačia problema susiduria dauguma duomenų bazių, jos aktyviai gyvuoja kelis metus, tada projektas iš lėto sustoja ir naujai identifikuoti ar patikslintos struktūros nebepasiekia duomenų bazių, o neteisingos struktūros nepanaikinamos. Pavyzdžiui, AH-DB buvo išleista 2012 m. balandį, paskutinį kartą atnaujinta 2013 m. kovą, BindingMOAD buvo išleista 2005 m. liepą, atnaujinta 2014 m. lapkritį. Tačiau net ir šie atnaujinimai daugiau skirti funkcionalumo tobulinimui negu duomenų atnaujinimui. Taigi projekto metu nuspręsta iškelti tris tikslus susijusius su duomenų baze:

1. Išanalizuoti esamų duomenų bazių duomenų kiekybę, parinkimo kriterijus ir patikimumą.
2. Sudaryti nedidelį duomenų bazės prototipą, pagal kurį būtų galima vykdyti kitus darbo uždavinius.
3. Sukurti įrankį gebantį paimti informacija iš bendros erdvinių struktūrų duomenų bazės – Protein Data Bank (PDB), ją apdoroti ir pateikti patogia forma mūsų sukurtų įrankių naudojimui.

Literatūros ir esamų duomenų bazių analizė parodė, kad nėra vieno šaltinio, kurį galėtume panaudoti, kaip pagrindą savo duomenų bazei. Visos jos pasižymėjo tam tikrais trūkumais – sudaryta iš 10-15 m. senumo duomenų, neturi informacijos apie ligandų slopinimo konstantas, neturi informacijos apie apo struktūras arba atmeta dalį perteklinių duomenų pagal savo kriterijus, kurių mes netaikome. Tačiau buvo rastas būdas, kaip gauti didžiausią ir šviežiausią duomenų intervalą. Amerikiečių sukurta PDB versija RCSB, leidžia lengvai atfiltruoti didžiąją dalį mums nereikalingų struktūrų ir duomenis išsisaugoti.

Norint paskubinti tolimesnius darbo etapus, buvo sudaryta prototipinė duomenų bazė (1 pav.). Kadangi mūsų pagrindinis tikslas identifikuoti fermentams vaistines molekules, todėl visos pasirinktos struktūros buvo žmogaus baltymai, o jų slopikliai – tai medžiagos naudojamos ar anksčiau naudotos vaistiniuose preparatuose. Taip pat buvo atsižvelgta ir į literatūros analizės duomenis, išrinktos tik geros rezoliucijos struktūros ir turinčios patikimas slopinimo konstantas.

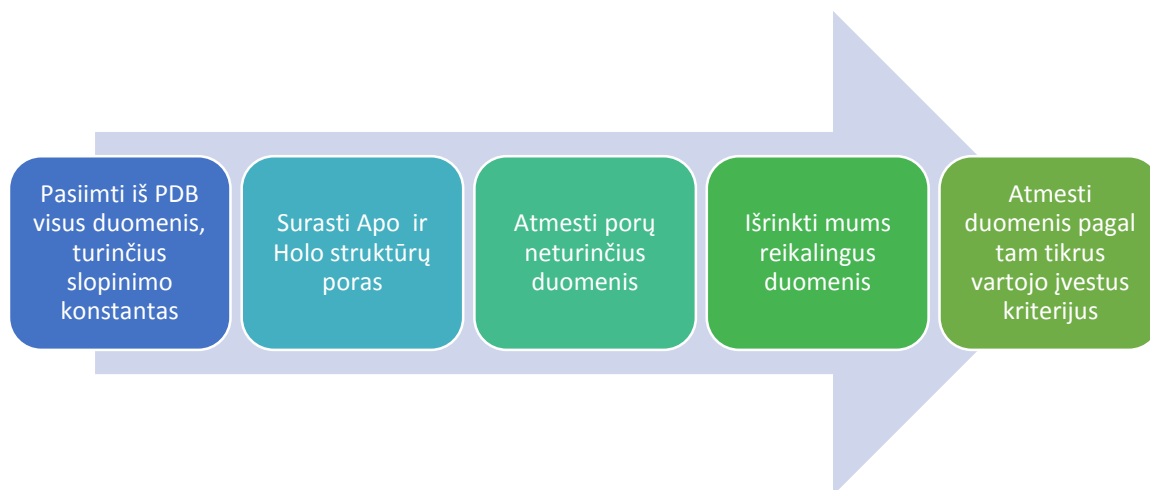
Nr	Pavadinimas	UniProt ID	Apo pav.	Apo rez.	Holo pav.	Holo rez.	Ligandas	Binding Affinity / nM	IC50, nM, iki	Ki, nM, iki	Kd, nM, iki	ECS0, nM, iki
1	Dihydrofolate Reductase	P00374	1PDB	2,2	1PD8	2,1	CO4	13	13			
2	Estrogen receptor	P03372	2B23	2,1	2QA8	1,85	GEN	990	990	370		956
3	HIV-1 reverse transcriptase	P04585	1RTJ	2,35	1DTQ	2,8	FPT	6	6			
4	Sterol 14alpha-demethylase	P9WPP9	2BZ9	2,21	2CIO	1,53	1CM	19500				19500
5	Acetylcholinesterase	P04058	2VA9	2,4	2V96	2,4	CFQ	29000		29000		
6	Peroxisome proliferator-activated receptor-gamma	P37231	1PRG	2,2	2PRG	2,3	BRL	4980	2000	440	4980	2880
7	Epidermal growth factor receptor	P00533	2EB2	2,5	3UG2	2,5	IRE	5,6	515	0,4	5,6	
8	Human cyclin-dependant kinase 2	P24941	1PW2	1,95	1PXI	1,95	CK1	17000	17000			
9	Human aldose reductase	P15121	1X96	1,4	1X98	1,3	FIS	11000	11000			
10	Phosphodiesterase-5	O76074	2H40	1,85	2H44	1,8	7CA	1700	1700			
11	Beta-lactamase	P00811	1KE4	1,72	1KDS	2,15	NPB	5000	5000	1700		
12	Angiotensin-converting enzyme	P12821	1O8A	2	1O86	2	LPR	0,27		0,27		
13	dUTPase	P06968	1EUW	1,05	2HRM	1,7	UCS	287500			287500	
14	Neuraminidase	P02477	2NWA	1,7	2NWC	1,6	DAN	280000	13000	280000		

1 pav. Mūsų sudaryta prototipinė duomenų bazė.

Sudarius prototipą, išanalizuota į kokius kriterijus atsižvelgė žinomų duomenų bazių kūrėjai. Daugelyje aprašymų kartojosi tam tikri pagrindiniai kriterijai:

1. Atmestos prastos rezoliucijos struktūros (>2,5 Å)
2. Atmestos struktūros su labai dideliu aktyvaus centro Cα-RMSD
3. Atmestos panašios struktūros pagal aminorūgščių sekų identiškumą.

Šie kriterijai labai svarbūs, norint sudaryti duomenų bazę, kurios duomenis programa galėtų interpretuoti. Pagrindinis dalykas, kuris paaiškėjo vykdant šią analizę, kad yra labai didelis skirtumas, kokia tvarka visus kriterijus pritaikysi. Todėl buvo sudarytas modelis, kurį mūsų duomenų bazės kūrimo programa turėtų įvykdyti.



2pav. Planas, kaip turėtų elgtis mūsų duomenų bazės sudarymo programa.

Projekto metu buvo įgyvendinti paveikslėlyje pateikti 1, 3, 4, 5 etapai, tik dalinai padarytas antrasis darbo etapas.

Darbai, kuriuos reikėtų vykdyti toliau kuriant įrankį duomenų bazės sudarymui:

1. Smulkių klaidų ištaisymas.

2. Antrojo etapo (2 pav.) pabaigimas.

Parametrų atranka vaistinių molekulių sąveikos su taikiniu nustatymui

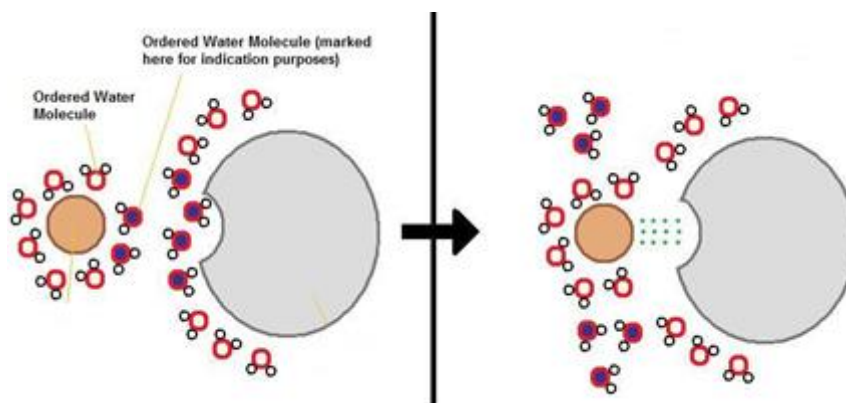
Biologinės molekulės, kaip antai baltymai, dalyvauja įvairiuose biologiniuose procesuose ir užtikrina normalią organizmo veiklą. Pavyzdžiui, fermentai pagreitina biochemines reakcijas, kraujo serumo antikūnai sąveikauja su patogeniniais mikroorganizmais ir taip juos neutralizuoja. Baltymų aktyvumas dažnai reguliuojamas kitomis organinėmis molekulėmis. Dėl to mokslininkai stengiasi surasti vaistus, gebančius jungtis prie šių biologinių molekulių ir jas slopinti ar tam tikrais atvejais aktyvinti. Taip pat ieškoma baltymo vietų, kurios galėtų sąveikauti su vaistinėmis molekulėmis. Suradimas šių taikinių bioinformatiniais metodais labai palengvintų jų tolimesnę paiešką atliekant biocheminius eksperimentus, kurių metu tikrinama baltymo ir vaistinės molekulės sąveika. Baltymai gali turėti keletą struktūrinių elementų, kišenių, tačiau tik prie tam tikrų jungiasi vaistinės molekulės. Šios kišenės vadinamos vaistinių molekulių taikiniais – vaistinių molekulių jungimosi vietos (JV). Jos pasižymi tam tikromis struktūrinėmis ir fizikinėmis savybėmis. Siekdami nustatyti tokias kišenes, pirmiausia atlikome išsamią parametrų, apibūdinančių vaistinių molekulių taikinius, literatūrinę analizę. Tada pasirinkome vienus tinkamiausių parametrų.

Galutinis tikslas ieškant JV parametrų yra iš jų sukurti funkciją, vadinamą *scoring function* (SF), kuri paskaičiuotų nustatytos kišenės įvertį. Šis įvertis pasako, kaip gerai vaistinė molekulė jungiasi prie taikinio. Tada jis palyginamas su eksperimentiškai nustatyta K_D – ligando ir jo taikio sąveikos stiprumo matu. Kuo K_D mažesnė vertė, tuo stipresnė sąveika. Tai reiškia, kad reikia mažiau vaistinių molekulių, kad įvyktų sąveika.

Savo darbe ieškojome JV parametrų tinkamų per burną vartojamų vaistinių taikinių nustatymui. Tokios vaistinės molekulės yra < 500 Da dydžio ir $< 140 \text{ \AA}^2$ paviršiaus ploto. Jos taip pat dažnai turi arba labai mažą krūvį arba jo išvis neturi, kadangi tik neutralios molekulės pasyviai pereina per ląstelės membraną. Pagal šias vaistinių molekulių savybes yra parenkamos JV savybės. Vaistinių molekulių taikiniai turi turėti prieinamą JV ir nekovalentiškai sąveikauti su vaistu (Kuntz et al. 1999).

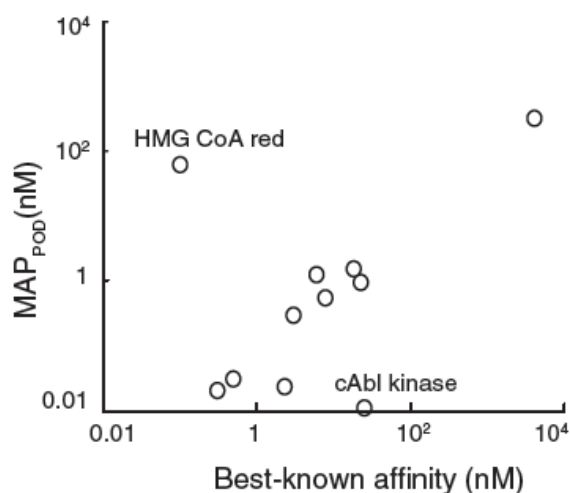
Vienas pirmųjų darbų pakankamai neblogai apibūdinusių vaistinių molekulių taikinių savybes, rėmėsi tik dvejais pagrindiniais parametrais: tirpikliui prieinamu paviršiaus plotu (SASA, ang. *Surface Accessible surface area*) ir kišenės paviršiaus įtempimu/kreivumu (Cheng et al. 2007). Jų idėja buvo tokia, kad baltymo ir ligando jungimosi energiją daugiausia lemia tirpiklio išstūmimas iš JV (3 pav.). Šį procesą geriausiai aprašo minėti parametrai. Vaistinių

molekulių taikinių SASA dažniausiai būna labai hidrofobiškas/nepolinis – sudaro molekulių elementai, neturintys krūvio. Dėl to šio darbo autoriai skaičiavo nepolinių molekulių užimamą plotą kišenėje. Tirpiklio išstūmimo energija priklauso nuo paviršiaus įtempimo, kurį apibūdina kitas parametras – kišenės kreivumas/nelygumai.



3 pav. Tirpiklio išstūmimas iš baltymo jungimosi su ligandu vietos.

Iš šių parametų buvo išvesta SF. Toliau atlikta transformacija ir gautas SF įvertis yra išreiškiamas kaip K_D . Nustatytų JV įverčiai buvo palyginti su eksperimentiškai nustatyta. Gauta koreliacija tarp autorių nustatytų įverčių ir eksperimentinių duomenų (4 pav.). Tai rodo modelio tinkamumą.



4 pav. Koreliacija tarp empiriškai nustatytų baltymų taikinių ir vaistinių molekulių sąveikos stiprumo ir eksperimentiškai gautų duomenų (Cheng et al. 2007).

Autorių nustatytas įvertis nesutapo dviejuose nagrinėtuose atvejuose. Vienu atveju svarbesnis buvo ne sąveikos stiprumas, o selektyvumas. Kitu atveju priežastis ta, kad jungimosi metu buvo svarbi ir elektrostatinė sąveika. Taigi, šis modelis nėra tobulas, tačiau tai vienas iš pirmųjų bandymų gerai ir svarbiausia paprastai apibūdinusių vaistinių molekulių jungimąsi su

JV. Dėl šių savybių šis darbas plačiai cituojamas ir tobulinamas. Mes taip pat rėmėmės juos pasirinkdami JV parametrus.

Atliktos kitos analizės rodo, kad yra labai daug JV parametrų. Vieni jų geriau apibūdina JV, gebėjimą prisijungti vaistines molekules, kiti silpniau. Nepaisant to, jų kombinacijos leidžia sukurti pakankamai tikslias SF.

Cheng et al. naudojo baltymų kristalines struktūras savo analizėje. Kristalinės gardelės duomenys dažniausiai naudojami tokioje analizėje. Kiti autoriai naudodami branduolių magnetinio rezonanso duomenis sukūrė SF (Hajduk, Huth, and Fesik 2005). Jų nustatyti SF parametrai: kišenės tūris, polinis ir nepolinis kišenės paviršiaus plotas, paviršiaus raukšlėtumas, apibūdinantis molekulės kompleksškumą, kišenės kompaktiškumas (kišenės tūris padalintas iš kišenės paviršiaus ploto), nepolinės aminorūgštys ir kiti parametrai apibūdinantys kišenės formą. Gauti SF parametrai koreliavo su eksperimentiniais duomenimis ($R^2 = 0,72$). Tai vienas iš pavyzdžių, kaip naudojant paprastus parametrus, apibūdinančius paviršiaus plotą, JV kompleksškumą ir kišenės formą, galima sukurti gerą SF.

Kitas svarbus parametras apibūdinantis JV savybę jungtis su vaistine molekule yra baltymo aminorūgščių sekos konservatyvumas. Panašias funkcijas atliekančių baltymų sekos yra panašios. Dažnai funkcinės molekulės dalys išlaiko savo konservatyvumą evoliucijos bėgyje. Pavyzdžiui, hemoglobinas išsaugojo hemo grupės seką evoliucijos bėgyje (Valdar 2002). Dėl to tikimasi, kad JV aminorūgščių sekos yra panašios ir pagal sekos konservatyvumą galima spręsti apie gebėjimą jungtis su vaistine molekule. Viename iš darbų buvo patvirtinta, kad šis jungimasis gerai koreliuoja su sekos konservatyvumu (Cimermancic et al. 2016).

Kiti autoriai kombinuodami panašius parametrus ir matematinius modelius sukūrė įvairių SF. Vienos iš jų geriau, kitos silpniau sutapo su eksperimentiniais duomenimis. Kombinuojant įvairius parametrus, sukuriama tobulesnė SF. Jų analizėse naudoti parametrai apibendrinti 1 lentelėje.

Tam tikri 1 lentelėje pateikti parametrai, kaip antai kišenės raukšlėtumas ir paviršiaus kreivumas, apibūdina panašias savybes. Pagrindinis skirtumas yra parametro apskaičiavimo metodas. Naudojant skirtingus skaičiavimo būdus, gaunami skirtingi rezultatai, tad ne visada galima sutapatinti skirtingų autorių naudojamus parametrus. Dažniausiai ieškoma paprasčiausio ir greičiausio būdo apskaičiuoti tam tikrą savybę. Apibendrinant, vaistinių molekulių taikinius apibūdina jo tūris, gylis, paviršiaus netolygumai, tirpikliui prieinamas paviršiaus plotas, krūvį turinčių ir neturinčių molekulių užimama dalis, aminorūgščių konservatyvumas.

1 lentelė. Vaistinių molekulių gebėjimą jungtis su taikiniu apibūdinantys parametrai

Parametras	Šaltinis
Tirpikliui prieinamas paviršiaus plotas (SASA, ang. <i>Surface Accessible surface area</i>)	(Cheng et al. 2007)
Kišenės kreivumas, susijęs su paviršiaus įtempimu	(Cheng et al. 2007)
Kišenės tūris	(Hajduk, Huth, and Fesik 2005); (Sheridan et al. 2010)
Nepolinis kišenės paviršiaus plotas	(Hajduk, Huth, and Fesik 2005)
Polinis kišenės paviršiaus plotas	(Hajduk, Huth, and Fesik 2005)
Paviršiaus raukšlėtumas, apibūdinantis molekulės kompleksiskumą	(Hajduk, Huth, and Fesik 2005)
Kišenės kompaktiškumas (kišenės tūris padalintas iš kišenės paviršiaus ploto)	(Hajduk, Huth, and Fesik 2005)
Nepolinės aminorūgštys	(Hajduk, Huth, and Fesik 2005)
Kišenės hidrofobiškumas (apibūdina krūvio neturinčias molekules)	(Schmidtke and Barril 2010); (Cimermancic et al. 2016); (Halgren 2009); (Nayal and Honig 2006); (Sheridan et al. 2010)
Aminorūgščių skaičius kišenėje	(Nayal and Honig 2006)
Atomų skaičius kišenėje	(Nayal and Honig 2006)
Kišenės gylis	(Nayal and Honig 2006); (Cimermancic et al. 2016)
Kišenės poliškumas (įvertinamas krūvį turinčių molekulių indėlis)	(Schmidtke and Barril 2010)
Kišenės uždarumas	(Halgren 2009)
Aminorūgščių sekos konservatyvumas	(Cimermancic et al. 2016)
B-faktorius, temperatūrinis faktorius, apibūdinantis atomų padėties molekulėje poslinkius	(Cimermancic et al. 2016)
RMSD – tai yra vidutinės atomų padėties skirtumas tarp baltymo molekulės prisijungusios vaistą ir ne	(Cimermancic et al. 2016)

Programos, naudojamos baltymų kišenėms aptikti

Darbo metu buvo analizuojamos *omnitools* programos, algoritmai, naudojami baltymo kišenių aptikimui. Šias programas būtų galima suskirstyti į 3 pagrindines kategorijas: geometriniais skaičiavimais pagrįsti algoritmai, molekulių energetinėmis sąvybėmis besiremiantys metodai bei fizikinėmis ir cheminėmis sąvybėmis pagrįsta analizė. Kiekviena programa yra savaip unikali, naudoja skirtingus metodus ir skirtingus parametrus siekiant kuo tiksliau aptikti ir įvertinti baltymo ir ligando (vaistinės molekulės sąveikos vietą). Vienos populiariausių yra: POCKET, SURFNET, LIGSITE, CASTp, SiteMap, JEDI, Q-Site Finder, fpocket, VISM-CFA, ghecom(grid-based HECOMi finder), ConCavity, Pocket-Finder,, DoGSite, VICE ir kitos. Jų palyginimas pateiktas 2 lentelėje.

2 lentelė. Programų, naudojančių įvairius algoritmus, baltymo ir ligandos prisijungimo vietos sėkmingo identifikavimo procentinis įvertinimas. Buvo analizuojamos baltymo struktūros su ligandu (holo) ir be ligando (apo). “Top 1”, jei didžiausia aptikta baltymo ertmė jungiasi su ligandu; “Top 3”. Jei viena iš trijų didžiausių ertmių jungiasi su ligandu. (*Proteins*. Author manuscript; available in PMC 2011 March 1.)

Method	Top 1		Top 3	
	Unbound	Bound	Unbound	Bound
VICE	83	85	90	94
Fpocket	69	83	94	92
PocketPicker	69	72	85	85
LIGSITE ^{cs}	60	69	77	87
CAST	58	67	75	83
PASS	60	63	71	81
SURFNET	52	54	75	78

Pasakyti, kuri yra tinkamiausia vaistinių molekulių identifikavimui labai sunku, programos yra nuolat tobulinamos keičiant ar pridėdant naujus parametrus ar tobulinant jų skaičiavimo metodus. Kai kuriais atvejais naudojamos kelios iš jų tam, kad gauti tikslesnius rezultatus, tačiau skaičiavimai gali užtrukti iki keletos dienų priklausomai tiek nuo programos metodo, tiek nuo naudojamos baltymų duomenų bazės dydžio, todėl renkatis tinkamiausią programą reikia atsisžvelgti į daugelį aspektų. Labai svarbu yra tai, kad jos ne tik apskaičiuoja baltymo ir ligando sąveikos galimybę, bet ir vizualizuoja baltymo 3D struktūrą. Remiantis šių programų analize bei kitais literatūriniais šaltiniais buvo apsvaistyti metodai atrinktų parametru,

tinkamų *scoring function* kūrimui, apskaičiavimui. Taigi, sekantis etapas buvo apskaičiuoti atrinktų parametrų vertes.

Atrinktų parametrų apskaičiavimas

Per projekto laikotarpį spėjome surasti būdą, kaip apskaičiuoti šiuos JV parametrus: SASA bei RMSD. Šiems parametrams apskaičiuoti buvo naudojama Python kalba, kuria ir buvo rašomas kodas. Šių parametrų apskaičiavimas susideda iš kelių žingsnių:

1. Pasirenkamos tiriamo baltymo struktūros (holo ir apo).
2. Sulyginamas apo ir holo struktūras sudarančių molekulių skaičius (struktūros sukarpomos).
3. Pasirenkamas atstumas tarp baltymo ir ligando atomų nedidesnis nei 0,45 Å. atmetant vandens molekules, kadangi jos supa visą baltymo paviršių ir taip pat sąveikauja su baltymu (5 pav.).

```
def get_binding_site_atoms(holo_traj, ligand_code):
    # select indices!
    atoms_holo_prot = holo.top.select("protein and not element hydrogen")
    atoms_holo_lig = holo.top.select("resname %s" % ligand_code)
    # make pairs of selections
    atom_pairs = holo.top.select_pairs(selection1=atoms_holo_prot, selection2=atoms_holo_lig)
    # calculate pair distances
    pair_dist = md.compute_distances(holo, atom_pairs, periodic=False)

    # select based on distance to ligands
    selection_mask = (pair_dist <= 0.45)[0] # select based on first frame, 0, here!

    # make unique list of PLA indices corresponding to holo structure!
    holo_PLA_ndx = []
    for atm_ndx_prot, atm_ndx_lig in atom_pairs[selection_mask]:
        if atm_ndx_prot not in holo_PLA_ndx:
            holo_PLA_ndx.append(atm_ndx_prot)

    return holo_PLA_ndx
```

5 pav. Ląstelėje vaizduojama komanda atrenkanti tik JV esančius atomus.

4. Gaunami JV esantys atomai ir amino rūgštys. Amino rūgščių sąrašas yra dar vienas svarbus parametras nustatant jų koncervatyvumą baltymo-ligando jungimosi vietose.
5. Patikrinama ar sutampa apo ir holo struktūrų atomų indeksai. (Indeksai turi sutapti, kitu atveju trūksta atomų ir skaičiavimai negali būti atlikti).

Atlikus šiuos etapus galime apskaičiuoti RMSD, kuris nurodo kaip stipriai skiriasi apo ir holo struktūros tarpusavyje. Apskaičiuojame tiek bendrą RMSD, tiek JV RMSD.

Sekantis parametras, kurį naudojome *scoring function* kūrimui yra SASA (ang. *solvent accessible surface area*). Šis parametras parodo baltymo plotą, kurį supa vandens molekulės, kadangi ligando prisijungimui prie baltymo vienas svarbiausių veiksnių yra solvatacija, tai yra

vandens išstūmimas, SASA parodo koks kišenėje esančių atomų plotas sąveikauja su vandeniu, tai svarbu vertinat energiją, kuri reikalinga vandeniui išstumti, kad ligandas galėtų patekti į baltymo kišenę. Šiam parametrui apskaičiuoti naudojami prieš tai paminėti žingsniai. SASA apskaičiuojamas holo ir apo struktūroms. 3 lentelėje pateiktos apskaičiuotos reikšmės.

3 lentelė. Baltymų, gautų iš PDB duomenų bazės, SASA ir RMSD reikšmės.

Apo ID	PDB ID	Ligand ID	UniProt ID	Binding Affinity / nM	Number of Binding site* atoms	Binding site atom RMSD / nm	BS atom SASA / nm²	SASA Difference to apo / nm²
1PDB	1PD8	CO4	P00374	13	54	0.09773199	4,78719	0,17544
2VA9	2V96	CFQ	P04058	29000	196	0.04455907	13,618	0,2531
1PW2	1PXI	CK1	P24941	17000	37	0.02644332	3,18387	0,21506
1X96	1X98	FIS	P15121	11000	14	0.01725101	1,66441	0,07332
2QWA	2QWC	DAN	P03472	280000	48	0.02154809	3,8166	-0,0636

Taigi, atlikome JV nustatymo programų literatūrinę analizę. Kiekviena programa pasižymi unikaliomis savybėmis. Jas taip pat jas galima panaudoti *scoring function* parametrų nustatymui. Per projekto laikotarpį suradome automatizuotą būdą, kaip apskaičiuoti vienus iš JV parametrų, SASA ir RMSD, naudojant *Python* programą. Atliktą teorinę ir praktinę analizę apie JV ir jų savybių, tinkamų vaistinių molekulių jungimuisi, nustatymą bus galima panaudoti ateityje.

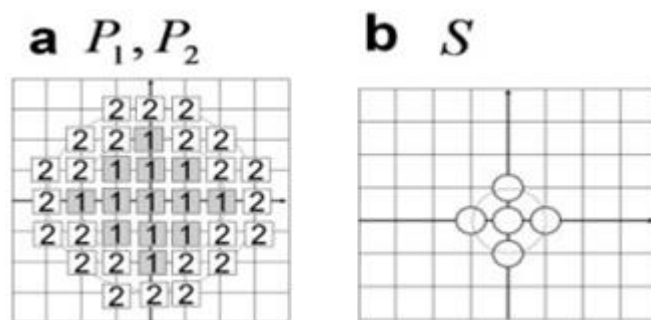
Kišenės paieška ir skaičiavimai

Išsikelti uždaviniai:

1. Kišenės tūrio ir gylio parametrų skaičiavimo pagal *GHECOM* programą analizė.
2. *GHECOM* programos detali kodo analizė ir optimizavimo sprendimai.

Internetinėje erdvėje yra pakankamai daug prieinamų atviro kodo programų atitinkančių ar dalinai atitinkančių mūsų tikslus. Dėl to buvo nuspręsta dalį tinkamo algoritmo sprendimų pritaikyti kuriant mūsų programą. Atlikus kelių programų funkcijų analizę, buvo nuspręsta, kokius naudingiausias parametrus iš visų pateiktų būtų galima panaudoti.

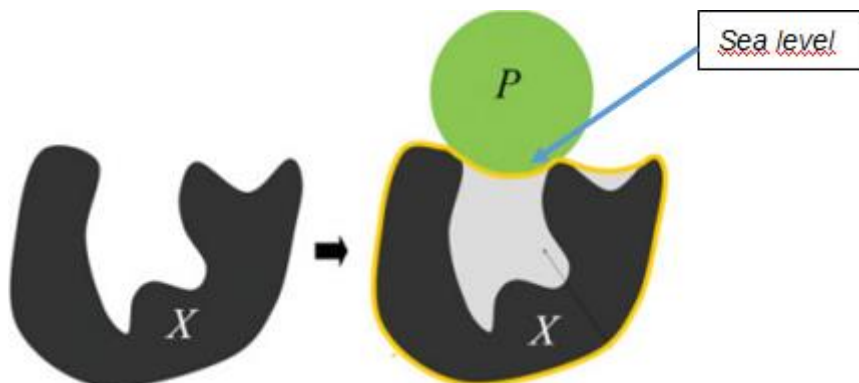
Mūsų uždavinius geriausiai atitiko vienu tyrėjų parašyta programa *GHECOM* (Kawabata 2010). Šios programos tikslas pagal struktūrinius baltymo parametrus (pagrindė paviršiaus reljefą) rasti susidariusias ertmes/kišenes. Ši programa ne tik randa kišenes, bet ir apskaičiuoja jų tūrį bei gylį, taip pat ir ligando panirimą į akceptorinį baltymą. Baltymo paviršius yra analizuojamas pasitelkiant, vadinamuosius programinius zondus, kurie yra sudaryti iš vokselių, be to skirtingo dydžio. Mažiausio zondo spindulys yra 1,87 Å šis dydis atitinka metilo grupės (-CH₃) dydį, kuris yra mažiausias galimas parametras nurodant ligando tūrį, plotą ir taip pat mažiausią kišenės tūrį (6 pav.).



6 pav. Naudojamo zondų struktūrinis vaizdas. Kiekvienas zondas sudarytas iš vokselių. Algoritme GHECOM buvo naudojami trijų dydžių zondai: (a) P_1 – 3-4 Å, P_2 – 5 Å, P_2 atvaizduotas kaip P_1 kiautas; (b) S – 1,87 Å.

Analizei naudojami zondai apeina tikrinamą baltymą tam tikra tvarka. Pirmiausia yra naudojamas didžiausias zondas, kuris aptinka didesnio ploto kišenes, platesnius įdubimus. Tolesniam vertinimui naudojamas vidutinio dydžio zondas, kuris gali patekti giliau į kišenės vidų nei pats didžiausias zondas. Galutinėje analizės stadijoje kišenės dugnas yra pripildomas 1,87 Å spindulio zondais, kur kiti zondai nebegali prieiti.

Tūrio skaičiavimai yra atliekami susumavus visų kišenėje esančių zondų tūrius (kurie yra žinomi). Kišenės pabaiga (riba tarp kišenės ir išorinės erdvės), vadinama jūros lygiu (ang. *sea level*) yra aptinkama didžiausio zondo (7 pav.).

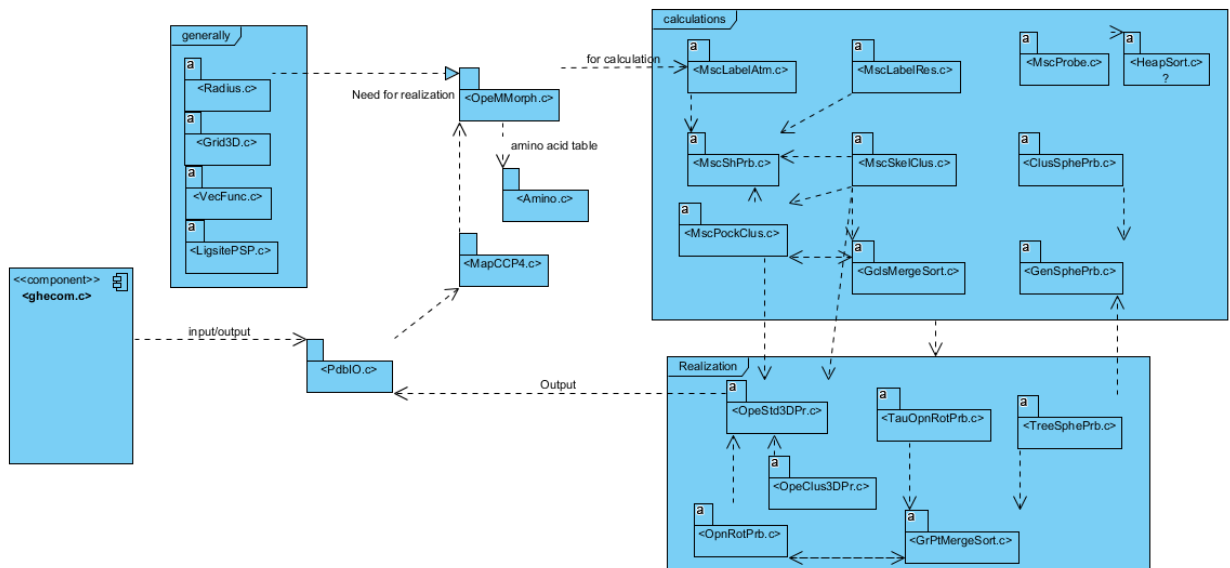


7 pav. *Sea level* aptikimas su didžiausiu zondų.

GHECOM programos struktūrinė analizė

Analizuojant *GHECOM* išeities kodą buvo sudaryta ryšių lentelė, kurioje yra matomos elementų priklausomybės (8 pav.). Tokia lentelė patogi ieškant struktūrų kurios ateityje bus naudingos, kurias reikės pataisyti ir kokių struktūrų dar reikės pridėti.

Aišku, kad reikės papildomai sukurti elementų kuriant papildomas funkcijas (skaitymas iš sukurtos duomenų bazės, rezultatų išvedimas į Python). *GHECOM* jau turi failus kurie („probe“, „radius“, „3Dgrid“) yra reikalingi etapui analizuojant baltymo paviršių. Taip pat *GHECOM* turi CCP4 žemėlapių keitimo funkciją kuri gali jį paversti į txt failą, kas gali būti patogu skaitant failus iš sukurtos duomenų bazės. Pagrindinis failas, kuris bus redaguojamas pagal mūsų poreikius yra „PdbIO“, nes PDB struktūras imsime iš savo sukurtos duomenų bazės.



8 pav. GHECOM programos kodo analizė aiškesniam kodo failų sąryšių supratimui.

Rekomendacijos tolimesnei veiklai

Atlikus *GHECOM* programos analizę yra aiškūs tolesni darbo su šia programa uždaviniai:

1. *PdbIO* kodo keitimas- programa turi nuskaityti failus iš sukurtos duomenų bazės.
2. Išskirtų parametrų – tūrio ir gylio – kodo pritaikymas *Python* programavimo kalbai.
3. Tūrio ir gylio skaičiavimo, bei kitų parametrų apjungimas į vieną programą.

Rezultatų apibendrinimas ir rekomendacijos

Baltymų jungimosi vietų su vaistinėmis molekulėmis nustatymas yra aktuali šių dienų problema. Prieš atliekant eksperimentus svarbu žinoti, ar tam tikri baltymai gali jungtis su tiriamomis vaistinėmis molekulėmis, su kuriomis iš jų jungiasi geriausiai. Šiame darbe siekėme sukurti priemones įrankiui, kuris rastų vaistinių molekulių sąveikos su baltymais vietas. Atlikome šiuos darbus:

1. Sudarėme žinomų struktūrų su ligandais duomenų bazės prototipą.
2. Atrinkome parametrus ligando sąveikai su taikiniui nustatymui. Šie parametrai apibūdina taikinio tūrį, gylį, paviršiaus netolygumus, tirpikliui prieinamą paviršiaus plotą, krūvį turinčių ir neturinčių molekulių užimamą dalį, aminorūgščių konservatyvumą.
3. Atlikome baltymų jungimosi su ligandais vietų nustatymo programų literatūrinę analizę. Jas taip pat galima panaudoti baltymų jungimosi vietų savybių nustatymui. Per projekto laikotarpį suradome ir optimizavome automatizuotą būdą, kaip apskaičiuoti vienus iš baltymų jungimosi vietų parametrų, SASA ir RMSD, naudojant *Python* programą.
4. Atlikta baltymų jungimosi vietų identifikavimo programos *GHECOM* analizė.
5. Atlikę *GHECOM* programos analizę atrinkome parametrus tinkamus baltymo jungimosi su ligandu vietų nustatymui. Atrinkti parametrus yra tūris ir gylis.

Tolimesni planai ir rekomendacijos:

1. Remiantis šiame darbe sudarytu teoriniu pagrindu surasti būdą, kaip apskaičiuoti likusius atrinktus baltymų jungimosi su ligandais vietų parametrus.
2. Nors padėtas pagrindas žinomų baltymų struktūrų ir jų kompleksų su ligandais duomenų bazės prototipo sukūrimui, įrankis automatizuojantis šį procesą nepabaigtas. Reikėtų jį pabaigti.
3. Darbai su *GHECOM* programa:
 - a. PdbIO kodo keitimas – programa turi nuskaityti failus iš sukurtos duomenų bazės.
 - b. Išskirtų parametrų – tūrio ir gylio – kodo pritaikymas *Python* programavimo kalbai.
 - c. Tūrio ir gylio skaičiavimo bei kitų parametrų apjungimas į vieną programą.

4. Atliktą teorinę ir praktinę analizę apie baltymų jungimosi su ligandu vietų ir jų savybių, tinkamų vaistinių molekulių jungimuisi, nustatymą bus galima panaudoti ateityje.

Literatūra

- Cheng, Alan C, Ryan G Coleman, Kathleen T Smyth, Qing Cao, Patricia Soulard, Daniel R Caffrey, Anna C Salzberg, and Enoch S Huang. 2007. "Structure-Based Maximal Affinity Model Predicts Small-Molecule Druggability." *Nature Biotechnology* 25 (1): 71–75. doi:10.1038/nbt1273.
- Cimermancic, Peter, Patrick Weinkam, T. Justin Rettenmaier, Leon Bichmann, Daniel A. Keedy, Rahel A. Woldeyes, Dina Schneidman-Duhovny, et al. 2016. "CryptoSite: Expanding the Druggable Proteome by Characterization and Prediction of Cryptic Binding Sites." *Journal of Molecular Biology* 428 (4): 709–19. doi:10.1016/j.jmb.2016.01.029.
- Hajduk, Philip J., Jeffrey R. Huth, and Stephen W. Fesik. 2005. "Druggability Indices for Protein Targets Derived from NMR-Based Screening Data." *Journal of Medicinal Chemistry* 48 (7): 2518–25. doi:10.1021/jm049131r.
- Halgren, Thomas A. 2009. "Identifying and Characterizing Binding Sites and Assessing Druggability." *Journal of Chemical Information and Modeling* 49 (2): 377–89. doi:10.1021/ci800324m.
- Kawabata, Takeshi. 2010. "Detection of Multiscale Pockets on Protein Surfaces Using Mathematical Morphology." *Proteins: Structure, Function, and Bioinformatics* 78 (5): 1195–1211. doi:10.1002/prot.22639.
- Kuntz, I. D., K. Chen, K. A. Sharp, and P. A. Kollman. 1999. "The Maximal Affinity of Ligands." *Proceedings of the National Academy of Sciences* 96 (18): 9997–10002.
- Nayal, Murad, and Barry Honig. 2006. "On the Nature of Cavities on Protein Surfaces: Application to the Identification of Drug-Binding Sites." *Proteins: Structure, Function, and Bioinformatics* 63 (4): 892–906. doi:10.1002/prot.20897.
- Schmidtke, Peter, and Xavier Barril. 2010. "Understanding and Predicting Druggability. A High-Throughput Method for Detection of Drug Binding Sites." *Journal of Medicinal Chemistry* 53 (15): 5858–67. doi:10.1021/jm100574m.
- Sheridan, Robert P., Vladimir N. Maiorov, M. Katharine Holloway, Wendy D. Cornell, and Ying-Duo Gao. 2010. "Drug-like Density: A Method of Quantifying the 'Bindability' of a Protein Target Based on a Very Large Set of Pockets and Drug-like Ligands from the Protein Data Bank." *Journal of Chemical Information and Modeling* 50 (11): 2029–40. doi:10.1021/ci100312t.

Valdar, William S.J. 2002. "Scoring Residue Conservation." *Proteins: Structure, Function, and Genetics* 48 (2): 227–41. doi:10.1002/prot.10146.