

# Baltymų struktūrų modeliavimas naudojant HHpred ir SWISS-MODEL

## Laboratorinis darbas

Justas Dapkūnas

2017

## 1 Įvadas

Šio darbo tikslas yra praktiškai išbandyti baltymų struktūrų modeliavimą, naudojant paprastus ir viešai prieinamus internetinius įrankius HHpred iš MPI Bioinformatics Toolkit [1] ir SWISS-MODEL [2]. Darbo metu reikia sukurti keletą pateikto baltymo struktūrinių modelių ir naudojant įvairius modelių įvertinimo metodus iš jų išrinkti geriausią. Išrinkto geriausio modelio struktūrą reikia pavaizduoti keliais būdais, išryškinant įvairius jos aspektus.

## 2 Darbo eiga

Kiekvienam yra pateikiama baltymo seka, kuriai reikia sumodeliuoti struktūrą. Siūloma tokia tolimesnė veiksmų eiga:

1. Paieškoma informacijos apie pateiktą baltymą UniProt ar kitose baltymų duomenų bazėse, naudojant BLAST paiešką;
2. Paleidžiamas automatinis baltymo struktūros modeliavimas SWISS-MODEL serveryje;
3. Paleidžiama šablonų paieška HHpred serveryje;
4. Iš SWISS-MODEL sumodeliuotų modelių išrenkamas geriausias ir parsiunčiamas jo failą;
5. Iš HHpred rastų baltymų struktūrų pasirenkami šablonai struktūros modeliavimui ir paleidžiamas MODELLER (turi būti sukurti bent 2 modeliai);

6. Parsiunčiami naudojant HHpred ir MODELLER sukurti baltymų struktūrų modeliai ir taip pat išsaugomi HHpred serveryje pateikiami jų pirminiai įvertinimai;
7. (Neprivaloma papildoma užduotis) Baltymo struktūra sumodeliuojama naudojant kitus struktūrų modeliavimo serverius;
8. (Neprivaloma papildoma užduotis) Baltymų struktūriniai modeliai tikslinami naudojant locPREFMD ar FG-MD serverius;
9. Sukurtų modelių kokybė įvertinama naudojant ProSA, QMEAN, VoronQA ar kitus baltymų struktūrų modelių vertinimo metodus ir išrenkamas geriausias modelis;
10. Pasirinktas geriausias modelis vizualizuojamas ir analizuojamas naudojant baltymų struktūrų vaizdavimo programas.

### **3 Informacijos apie baltymą paieška**

Gavus baltymo seką visų pirma reikėtų pabandyti išsiaiškinti, koks tai yra baltymas. Informacijos radimui rekomenduojama pasinaudoti BLAST paieška UniProt (<http://www.uniprot.org/>), NCBI (<https://www.ncbi.nlm.nih.gov/>) ar kitose duomenų bazėse. Taip galima sužinoti, iš kokio organizmo yra baltymas, o jeigu jis yra iširtas, pasidomėti ir apie jo funkcijas.

### **4 Baltymo struktūros modeliavimas naudojant SWISS-MODEL**

SWISS-MODEL yra automatinis baltymų struktūrų modeliavimo serveris, sukurtas Šveicarijos Bioinformatikos institute [2]. Jis yra pasiekiamas adresu <https://swissmodel.expasy.org/>. Naudojant šį serverį galima ieškoti šablonų, o taip pat automatiškai pagal juos sumodeliuoti struktūrą. Norint modeliuoti struktūrą SWISS-MODEL užtenka pateikti baltymo seką, bet taip pat galima pateikti ir papildomų duomenų.

SWISS-MODEL pateikia automatiškai sukurtus modelius, taip pat informaciją apie modeliavimui naudotus šablonus ir kai kuriuos modelių kokybės įverčius. Visus pateikiamus duomenis galima žiūrėti interneto tinklalapyje arba parsiųsti į savo kompiuterį. Šis serveris, kaip ir dauguma kitų bioinformatikos serverių, duomenis saugo tik trumpą laiką, dėl to rekomenduojama rezultatus išsisaugoti.

## Pradinių duomenų įkėlimo langas:

### Start a New Modelling Project

**Target Sequence:**  
(Format must be Fasta, Clustal, Promod, plain string, or a valid UniProtKB AC)

Paste your target sequence here

+ Upload Target Sequence File...

**Supported Inputs**

- Sequence ▾
- Uniprot AC ▾
- Target-Template Alignment ▾
- User Template ▾
- Deepview Project ▾
- Hetero Project BETA ▾

**Project Title:**

**Email:**

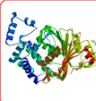
Search For Templates
Build Model

*By using the SWISS-MODEL server, you agree to comply with the following [terms of use](#) and to cite the corresponding [articles](#).*

## Rezultatai:

### Model Results

Order by: GMQE ▾





Model 01

**Oligo-State**  **Ligands**  
MONOMER None


**Global Quality**

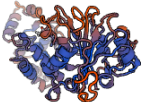
QMEAN	-3.87
Cβ	-1.99
All Atom	-1.84
Solvation	-0.45
Torsion	-3.43

**Local Quality** 

**Comparison** 

**Template** **Seq Identity** **Coverage** **Description**

**3kd.1.A** 42.14%  Type I restriction-modification system methyltransferase subunit



1 509

**Model-Template Alignment**

Model	Seq	Identity	Coverage	Description
Model_01	MSTKPTIQSENNIYVKACDIFRGMIDPSQYKDYILTLHFKVYSD			
31kd.1.A	MSETTQTCSSTYDAWDSADQLBSKMDCHDYKYLTLNFKYKSDKHLLEFVALD			
Model_01	-----VWKDKQNVVMEKYNQDEIRVD RALRNE RFQVPASCTFEHLFENRNE			
31kd.1.A	MEETEESDEATLVYKTYDDEETGDELLAVYIDEYSYALHDDTETALTEVYN			
Model_01	ANLG ELINTALEDELDANRSKLERVFRNIDFNSEANLQQT KORNRLKLLVD			
31kd.1.A	GSFQCELAGGFDTEQSDE-LYENLFEDIDLYSCTLGATGQKNTVVAVYKKE			
Model_01	FSALDLQPSHLAGNDITGDAYEYLIERFASDAGKAGEFYTPSGVSTLLAKLI			
31kd.1.A	SLDLYVAGH--AGDGDAYEYLIDFA DGGKAGEFYTPGVATLITQDAD			
Model_01	DPQPGNRIIDPTCGSGSLIKTAREVGSNHFSLY GQEIINGSTWALARMMEL			
31kd.1.A	GREDKQGETDATTGSGSLCTARSRQPQTVYVYQQLNCTSLARMMEL			
Model_01	EMDNANIENGDTLNN-----PKLLEGDSLKKAHIVANPPFSLDKMGADNA			
31kd.1.A	HGVPCENFIDDTLDEDWPTQEPINSG-----VFNPPYSA-KWASAG			
Model_01	AAD QENRFHRCIPPKSGDYAFVTHMIESTYEDIGKGVILPHGALFRGSSEG			
31kd.1.A	DDPRFCTGSK-LAPKSKADAFLLHDTYKQDNGTATLPHGTLFRGTEG			
Model_01	KIRKQITEENLLEAVIGLPAFLFPGTIPACINIFNRAKGENRDLIFDANKGYD			
31kd.1.A	KIRKQLTEGADIVIGLPANIFFNTSIPITVYKKNR-TNRDTEIDASKEFD			
Model_01	SGKNQRLSDTHIKKIVNTYQSFNRDAPLTATEGKVLLEEKYAYRATIQLLEENEY			
31kd.1.A	SGKNQNSITDHIKIKINYSREDG-----DKFAAACEETENIDY			
Model_01	LNLYPRVYDTFEEEEPVDTITATQTEINRLKGLSDVKEQMNLYKDLGFG			
31kd.1.A	NLNLPRVYDTFEEVEPLTEIYAINQTN---ATIEEQIASLLDNLG-			

**Oligo-State**  **Ligands**  
MONOMER None


**Global Quality**

QMEAN	-5.50
-------	-------

**Local Quality**

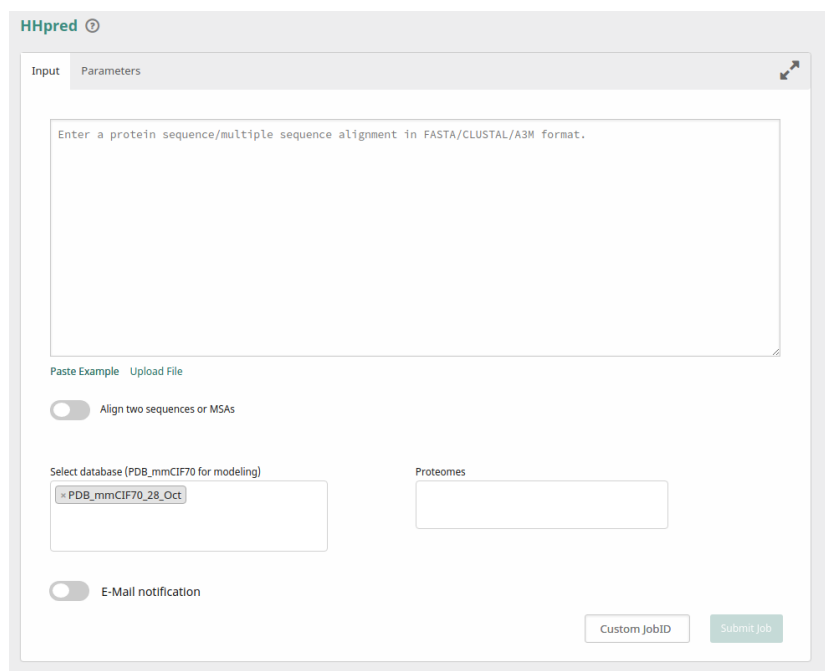
**Comparison**

**Template** **Seq Identity** **Coverage** **Description**

**2ar0.1.A** 22.69%  Type I restriction enzyme EcoKI M protein

## 5 Baltymo struktūros modeliavimas naudojant HHpred

HHpred yra tolimos baltymų sekų homologijos paieškos metodas, kuris remiasi profilio-profilio (HMM-HMM, paslėptųjų Markovo modelių) palyginimu [3]. Paprasčiausias būdas jį išbandyti yra pasinaudoti internetine versija, esančia MPI Bioinformatics Toolkit (<https://toolkit.tuebingen.mpg.de/hhpred>) [1]. Šiam serveriui taip pat reikia pateikti savo tiriamo baltymo seką, o prieš pradėdant paiešką galima patikslinti įvairius paieškos parametrus.



The screenshot displays the HHpred web interface. At the top, there are two tabs: 'Input' and 'Parameters'. The 'Input' tab is active, showing a large text area for entering a protein sequence or multiple sequence alignment in FASTA/CLUSTAL/A3M format. Below the text area, there are links for 'Paste Example' and 'Upload File'. A toggle switch is set to 'Off' for 'Align two sequences or MSAs'. Under 'Select database (PDB\_mmCIF70 for modeling)', a dropdown menu shows 'PDB\_mmCIF70\_28\_Oct'. To the right, there is an empty 'Proteomes' text field. At the bottom, there is another toggle switch for 'E-Mail notification' set to 'Off'. Two buttons are visible: 'Custom JobID' and 'Submit Job'.

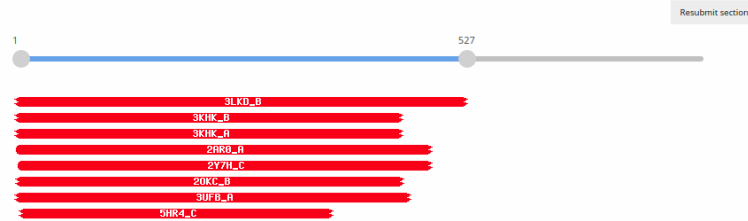
HHpred vartotojo sąsaja yra kažkiek panaši į NCBI BLAST. Serveris schematiškai pateikia rastus baltymus, taip pat lentelės pavidalo santrauką, o toliau galima detaliau nagrinėti sekų palyginius ir kitus duomenis. Naudojantis lentele ir nagrinėjant sekų palyginius galima pasirinkti šablonus struktūros modeliavimui:

Vis Hits Aln | Select all Forward Forward Query A3M Model using selection HHR file

Number of hits: 250

We have detected coiled-coil segments in your query protein!

### Visualization

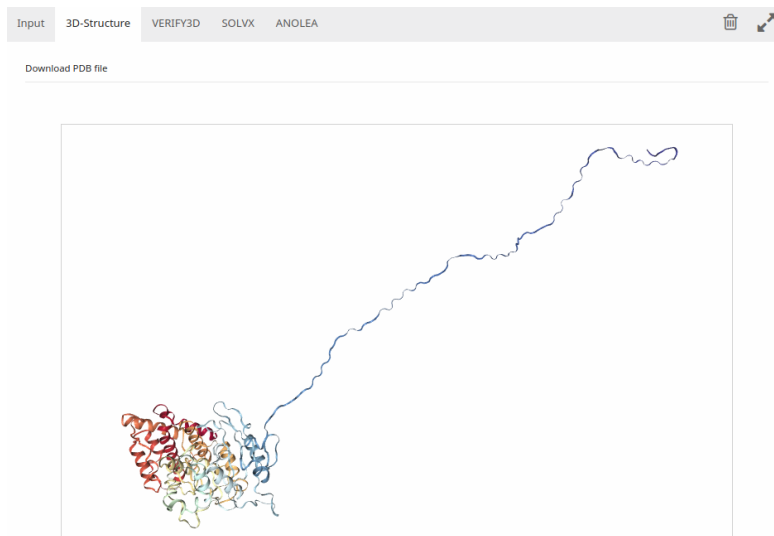


### Hittlist

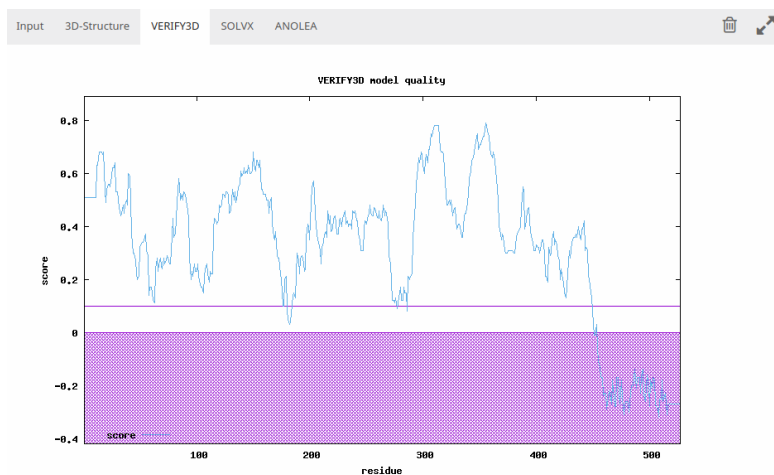
Show 25 entries

Nr	Hit	Name	Probability	E-value	SS	Cols	Target Length
<input type="checkbox"/>	1	3LKD_B Type I restriction-modification system methyltransferase; Q5M500_STRT2, methyltransferase, stu0711, NESG, SuR80; 2.25A {Streptococcus thermophilus}	100	9.1e-49	49.9	483	542
<input checked="" type="checkbox"/>	2	3KHK_B Type I restriction-modification system methylation; structural genomics, Type I restriction-modification; HET: SO4; 2.55A {Methanosarcina mazei}	100	2.2e-47	42.9	434	544
<input type="checkbox"/>	3	3KHK_A Type I restriction-modification system methylation; structural genomics, Type I restriction-modification; HET: SO4; 2.55A {Methanosarcina mazei}	100	2.2e-47	42.9	434	544
<input checked="" type="checkbox"/>	4	2AR0_A Type I restriction enzyme EcoKI; Structural Genomics Protein Structure; HET: UNX; 2.8A {Escherichia coli} SCOP: c.66.1.45	100	4.2e-44	43.6	445	541
<input type="checkbox"/>	5	2V7H_C TYPE-1 RESTRICTION ENZYME ECOKI SPECIFICITY; TRANSFERASE-DNA COMPLEX; HET: SAM; 18.0A {ESCHERICHIA COLI}	100	4.9e-43	43.1	443	529
<input checked="" type="checkbox"/>	6	2OKC_B Type I restriction enzyme StySI; NP_813429.1, N-6 DNA Methylase, Type; HET: SAM; 2.2A {Bacteroides thetaiotaomicron VPI-5482} SCOP: c.66.1.45	100	1.7e-41	40.3	411	445
<input type="checkbox"/>	7	3UFB_A Type I restriction-modification system methyltransferase; Methyltransferase activity, TRANSFERASE; 1.8A {Vibrio vulnificus}	100	1.9e-37	37.2	428	530
<input checked="" type="checkbox"/>	8	5HR4_C MmeI/DNA Complex; DNA-protein complex, Restriction-Modification enzyme, Hydrolase-DNA; HET: SFG; 2.5964A {Methylophilus methylotrophus}	99.86	1.6e-23	25.5	328	919





Taip pat galima naršyklėje panagrinti trimatę struktūrą arba atlikti pradinį jos įvertinimą keliais paprastais metodais, pvz., Verify3D.



Tolimesniam darbui sumodeliuota struktūra ir pradinio vertinimo duomenys parsiončiami į savo kompiuterį.

Iš viso naudojant HHpred serverį reikėtų sukurti 2 modelius. Tai gali būti modeliai, paremti skirtingais šablonais, arba vienas iš modelių gali būti sukurtas naudojant kelis šablonus. Be to, galima paleisti HHpred paiešką, naudojant skirtingus jos nustatymus (įvertinant arba ne antrinę struktūrą, skirtingais metodais atliekant sekų lygiavimą ir t.t.). Kaip ir SWISS-MODEL, HHpred taip pat gali parinkti šabloną automatiškai.

## 6 Kiti baltymų struktūrų modeliavimo metodai (papildoma užduotis)

Išbandyti HHpred ir SWISS-MODEL serveriai modeliuoja baltymų struktūras lyginamojo (homologinio) modeliavimo metodais, remdamiesi šablonais. Kai kuriais atvejais gali būti sudėtinga rasti šablonų. Tuomet norint sumodeliuoti struktūrą galima išbandyti laisvojo modeliavimo metodus. Vienas iš prieinamų pavyzdžių būtų I-TASSER serveris <http://zhanglab.ccmb.med.umich.edu/I-TASSER/>. Taip pat galima naudoti QUARK <http://zhanglab.ccmb.med.umich.edu/QUARK/> arba Robetta <http://rosetta.bakerlab.org/> serverius. Norint pasinaudoti šiais serveriais yra būtina registracija, taip pat jie struktūrą modeliuoja gana ilgai, gali užtrukti ir kelias dienas ar savaites.

## 7 Baltymų struktūrinių modelių tikslinimas (papildoma užduotis)

Baltymų struktūrų modelių tikslinimą galima atlikti naudojant locPREFMD (<https://feig.bch.msu.edu/web/services/locprefmd/>) [5] arba FG-MD (<http://zhanglab.ccmb.med.umich.edu/FG-MD/>) [6] serverius. Struktūrų tikslinimas gali užtrukti kelias valandas. Patikslintas modelis nebūtinai bus geresnės kokybės už pradinį, dėl to jį būtina paanalizuoti, lyginant su pradiniu modeliu.

## 8 Baltymų struktūrinių modelių kokybės vertinimas

SWISS-MODEL ir HHpred serveriai pateikia tam tikrus modelių kokybės vertinimus, tačiau norint išrinkti patikimiausią modelį jų neužtenka, nes serveriuose pateikiami skirtingi įverčiai. Norint palyginti skirtingus modelius tarpusavyje, reikia įvertinti juos naudojant tą patį struktūrų vertinimo metodą. Taip pat rekomenduojama įvertinti sukurtus modelius keliais metodais, kad būtų galima patikimiau išsirinkti patikimiausią modelį.

Šiame darbe visų pirma siūloma greitai įvertinti modelius naudojant ProSA (<https://prosa.services.came.sbg.ac.at/prosa.php>), o kaip alternatyvų metodą pasitelkti QMEAN (<https://swissmodel.expasy.org/qmean/>) ar VoromQA (<http://bioinformatics.ibt.lt/wtsam/voromqa/>). ProSA yra ganėtinai senas metodas [7], tačiau rezultatai gaunami labai greitai. QMEAN



serveris yra šiuolaikiškesnis, taip pat jame naudojamas struktūrų vertinimo metodas atsižvelgia į daugiau aspektų [8], tačiau skaičiavimai gali užtrukti ilgiau nei ProSA. VoromQA yra naujas, neseniai publikuotas baltymų struktūrų vertinimo metodas, kuris paremtas kontaktų tarp atomų plotų analize [9]. Galima naudoti ir kitus baltymų struktūrų vertinimo metodus: ProQ2 (<http://bioinfo.ifm.liu.se/ProQ2/>), ProQ3 (<http://proq3.bioinfo.se/>), ir pan., tačiau jų skaičiavimo laikas yra gerokai ilgesnis.

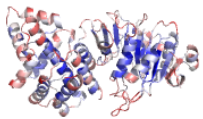
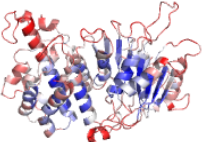
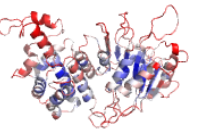

Visų šių serverių vartotojo sąsajos principas panašus: įkeliamas baltymo struktūros failas, kuriam programa suskaičiuoja kokybės įvertį. Patys modelių kokybės įverčiai, žinoma, skiriasi gana stipriai, nes jie remiasi skirtingais metodais ir principais.

Geriausio modelio iš kelių pasirinkimas yra kūrybiškas procesas, kuriame daug ką gali lemti modeliuotojo patirtis ir intuicija. Vertinant baltymų struktūrų modelius reikėtų atsižvelgti į keletą aspektų:

- Modelio struktūros įvertinimą reikia lyginti su naudoto šablono struktūros vertinimu, kadangi lyginamojo modeliavimo metodais gauto modelio kokybė priklauso nuo pasirinkto šablono kokybės. Jeigu šablono įverčiai yra nedideli, tai naudojant jį sukurtų modelių įverčiai bus dar žemesni dėl šablono pasirinkimo, o ne dėl modeliavimo netikslumų.
- Jeigu modelis yra sukurtas naudojant nepilną sekų palyginį, tai jis yra dalinis: sekos daliai, kurios nėra palyginyje, yra pateikiama atsitiktinė struktūra. Tokių nestruktūrizuotų „uodegų“ kokybės įverčiai visada yra maži ir sumažina viso modelio įverčius. Didelė struktūros dalis gali būti nusakyta pakankamai teisingai, tačiau to nepastebėsime žiūrėdami į globalų vertinimą. Dėl to patartina prieš vertinant modelį nukarpyti tas dalis, kurios nebuvo sumodeliuotos. Svarbu atsižvelgti ir į tai, kad SWISS-MODEL serveris tai padaro automatiškai, o HHpred serverio modeliai gali turėti ilgas nesumodeliuotas dalis. Tai ypač trukdo objektyviai palyginti HHpred ir SWISS-MODEL modelių kokybę, todėl prieš vertinant HHpred serverio modelius, būtina išskirti gerai sumodeliuotą jų dalį. Tai gali padėti padaryti Verify3D duomenys arba vizualinis modelio palyginimas su šablono struktūra.
- Tikėtina, kad gali būti ir atvirkščias variantas, kai modelis sukuriamas naudojant tik dalį šablono struktūros. Šiuo atveju šablono struktūros „apkirpimas“ reikalingas vertinant, ar labai homologinio modelio kokybė skiriasi nuo šablono kokybės.
- Struktūrinių modelių vertinimo programos paprastai pateikia grafiką, kaip modelis atrodo realių baltymų kontekste. Verta jį paanalizuoti.

- Visada reikia ne tik remtis įvairiais baltymų modelių mokybės įverčiais, bet ir pasižiūrėti, kaip atrodo sumodeliuota baltymo struktūra ir ją vizualiai įvertinti.

Lentelėje pateikti modelių įvertinimo rezultatai, apkerpant nesumodeliuotą dalį iš HHpred modelio. Reikia atkreipti dėmesį į tai, kad kai kurių įverčių didesnės reikšmės reiškia kokybiškesnius modelius, o kitais atvejais atvirkščiai, kuo mažesnis skaičius, tuo geresnis įvertinimas.

Modelis	QMEAN	ProSA	VoroMQA
Šablonas	-1.25	-10.47	0.56
			
SWISS-MODEL	-4.67	-6.52	0.45
			
HHpred (trumpintas)	-6.55	-5.82	0.39
			
HHpred (pilnas)	-6.37	-4.17	0.33
			

## 9 Baltymų struktūrinių modelių vizualizavimas

Baltymų struktūrų vizualizavimą rekomenduojama atlikti naudojant PyMOL, tačiau nedraudžiama naudotis ir kitomis programomis. Turi būti pateikti bent 3-4 vaizdai, kuriuose pasirinktas struktūros modelis pavaizduotas skirtingais būdais, spalvomis išryškinant norimus pabrėžti struktūros aspektus. Viename iš paveikslų turi būti panaudoti keli vizualizacijos būdai ir kelios spalvos, parodant gebėjimą naudotis struktūros dalių žymėjimo įrankiais.

Taip pat rekomenduojama viename iš paveikslų išsamiau patyrinti struktūrą: išmatuoti atstumą tarp jos galų arba funkciškai svarbių regionų, palyginti šablono ir modelio struktūras, palyginti geriausią modelio struktūrą su kitais modeliais ar pan.

## 10 Reikalavimai darbo aprašymui

Laboratoriniam darbui kiekvienam studentui bus pateiktos 1 arba 2 užduotys. Molekulinės biologijos studentai laboratorinį darbą atlieka namuose, pirmąją pradėdami vykdyti paskaitos metu, o antroji skirta savarankiškam modeliavimui ir žinių įtvirtinimui. Biochemijos studentai paskaitos metu treniruojasi, o kitą paskaitą **kontrolinį darbą daro auditorijoje**. Dėl to biochemikai nedaro papildomų užduočių, kurios užtrunka ilgiau nei paskaitai skirtas laikas.

Modeliavimo procesai turėtų būti aprašyti, kiekvieną žingsnį dokumentuojant ekrano nuotraukomis. Rekomenduojama taip pat išsaugoti nuorodas į rezultatus tiems atvejams, jei kiltų klausimų. Aprašyme rezultatai turėtų būti pateikiami naudojant lenteles tais atvejais, kai lyginami skirtingi modeliai ir pan. Struktūrų vizualizacijos rezultatai turi būti išsaugomi ne kaip ekrano nuotraukos, o naudojant paveikslėlio išsaugojimo funkciją. Kartu su aprašymu pateikiami ir sukurtų modelių PDB failai. Jeigu daromos ir papildomos užduotys, jų rezultatų aprašymas turi būti integruojamas ir interpretuojamas kartu su privalomų užduočių rezultatais.

Vertinimo kriterijai:

Etapas	Vertinimo dalys	Balų skaičius
Baltymo anotacijos paieška		0,5
SWISS-MODEL modeliavimas	Programos naudojimas, modelio pasirinkimas	1,5
HHpred modeliavimas	Programos naudojimas, šablonų pasirinkimo motyvacija	3
Kiti modeliavimo ir tikslinimo metodai	<b>Neprivaloma dalis</b>	2
Modelių vertinimas	Geriausio modelio pasirinkimo motyvacija	2
Geriausio modelio vizualizavimas	Parodytas gebėjimas įvairiai naudotis vizualizavimo programa	3

## Literatūros šaltiniai

- [1] Alva V, Nam SZ, Söding J, Lupas AN. The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis. *Nucleic Acids Res.* 2016 Jul 8;44(W1):W410-5.
- [2] Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Gallo Cassarino T, Bertoni M, Bordoli L, Schwede T. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* 2014 Jul;42(Web Server issue):W252-8.
- [3] Söding J. Protein homology detection by HMM-HMM comparison. *Bioinformatics.* 2005 Apr 1;21(7):951-60.
- [4] Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol.* 1993 Dec 5;234(3):779-815.
- [5] Feig M. Local Protein Structure Refinement via Molecular Dynamics Simulations with locPREFMD. *J Chem Inf Model.* 2016 Jul 25;56(7):1304-12.
- [6] Zhang J, Liang Y, Zhang Y. Atomic-level protein structure refinement using fragment-guided molecular dynamics conformation sampling. *Structure.* 2011 Dec 7;19(12):1784-95.

- [7] Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* 2007 Jul;35(Web Server issue):W407-10.
- [8] Benkert P, Künzli M, Schwede T. QMEAN server for protein model quality estimation. *Nucleic Acids Res.* 2009 Jul;37(Web Server issue):W510-4.
- [9] Olechnovič K, Venclovas Č. VoroMQA: Assessment of protein structure quality using interatomic contact areas. *Proteins.* 2017 Jun;85(6):1131-1145.