

VILNIAUS UNIVERSITETAS  
MATEMATIKOS IR INFORMATIKOS FAKULTETAS  
MATEMATINĖS INFORMATIKOS KATEDRA

**Rugilė Lukšaitė**

Bioinformatikos studijų programa

Matematinės informatikos šaka

**Baltymų erdvinių struktūrų vizualizavimas**

Ataskaita

Vilnius, 2022

## Užduotis

- Nustatyti, kurioms „SCOP“ klasėms priklauso priskirtos baltymų struktūros;
- Atvaizduoti baltymų struktūras bei jų fragmentus naudojantis „Jmol“ [1] bei sukurtos „pdbx-draw-secondary“ programos pagalba.

## Duomenys

- Priskirtų baltymų struktūrų iš „Protein Data Bank“ [2] identifikatoriai: 6G45, 2C2A, 5A5T, 2DM1, 1WGX.
- Baltymų struktūrų URL nuorodos galima rasti trečioje nuorodoje ([3]).
- Duomenų formatas PDBx/mmCIF.
- Struktūros atsisiųstos iš „Protein Data Bank“ naudojantis „cURL“ [4] programos įrankiu.
  - Komandos pavyzdys:  
`curl -sSL --fail http://www.rcsb.org/pdb/files/6G45.cif > ./6G45.cif`
- Atsisiųstas baltymo struktūras galima rasti „data“ direktorijoje (kelias: 2as-darbas/working-copy/trunk/tests/data/[baltymo id].cif)

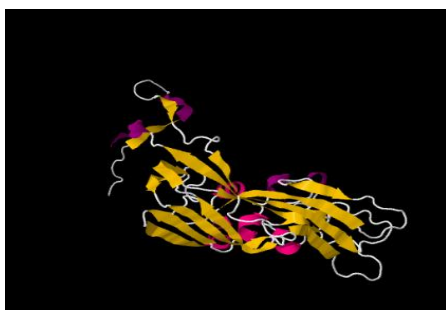
## „SCOP“ [5] ir „CATH“ [6] klasės pagal domenų

Baltymo identifikatorius pagal „PDB“	Pasirinkto domeno grandinė	Domeno liekanų ruožas	Domeno ID pagal „SCOP“ arba „ECOD“	„SCOP“ klasė	„CATH“ klasė
6G45	A	2 – 276	c6g45A2*	Nepateikta, manoma: $\alpha + \beta$	Nepateikta, Manoma: Alpha Beta
2C2A	A	321 – 481	8026739	$\alpha + \beta$	Alpha Beta
5A5T	C	446 – 770	8093045	Nepateikta, manoma: All alpha proteins	Nepateikta, manoma: Mainly Alpha
2DM1	A	8 – 67	8065200	All beta proteins	Mainly Beta
1WGX	A	8 – 67	8002303	All alpha proteins	Mainly Alpha

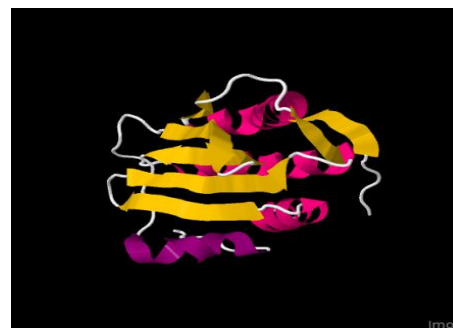
- \*- „SCOP“ bei „CATH“ tinklapiai nepateikė informacijos apie baltymo struktūra 6G45, todėl naudotas „ECOD“ [7] identifikatorius bei nurodytas domeno liekanų ruožas. Visos kitos struktūros nurodytos pagal „PDB“ informaciją.

## Baltymų domenų ir jų klasių vaizdavimas

Tretiniam baltymų vaizdų generavimui buvo naudojama „Jmol“ programa bei Jmol komandos „cartoon“ bei „color structure“. Šias nuotraukas galima rasti images direktorijoje (kelias: 2as-darbas/working-copy/trunk/images/domain\_[struktūros id].png). Šių nuotraukų generavimas aprašytas nurodytoje direktorijoje „images/“, faile „information-domain-generation.txt“.



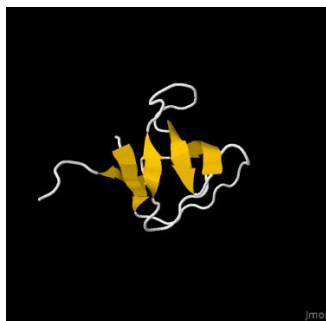
Pav. 1 Baltymo 6G45 struktūros domenas, 2-276 liekanų ruožas, grandinė A. Struktūroje matomi antilygiagretūs beta lakštai bei keletas alfa spiralių, todėl šio domeno „SCOP“ klasė priskirta kaip  $\alpha + \beta$ .



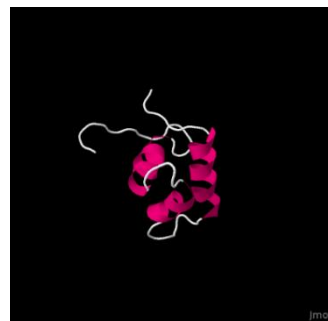
Pav. 2 Baltymo 2C2A struktūros domenas, 321-481 liekanų ruožas, grandinė A. Domene matomi daugiausiai antilygiagretūs beta lakštai ir alfa spirals. Spirals ir beta lakštai nepersipina, todėl priskirta „SCOP“ klasė  $\alpha + \beta$ .



Pav. 3 Baltymo 5A5T struktūros domenas, 446 – 770 liekanų ruožas, grandinė C. Domene matomos tik alfa spiralinės, todėl priskirta „SCOP“ klasė - All alpha proteins.



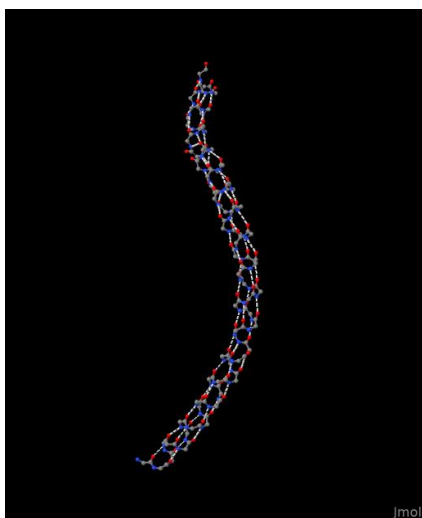
Pav. 4 Baltymo 2DM1 struktūros domenas, 8 – 67 liekanų ruožas, grandinė A. Struktūroje matomos antilygiagrečios beta statinės. Priskirta „SCOP“ klasė – All beta proteins.



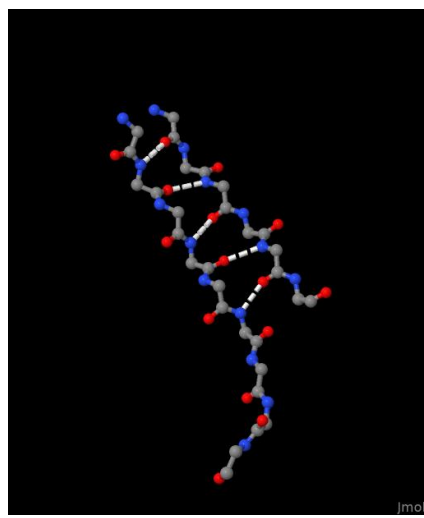
Pav. 5 Baltymo 1WGX struktūros domenas, 8-67 liekanų ruožas, grandinė A. Domene matomos tik alfa grandinės, todėl „SCOP“ klasė yra All alpha proteins.

### Alfa spiralių, lygiagrečių beta-lakštų ir antilygiagrečių beta-lakštų vaizdavimas

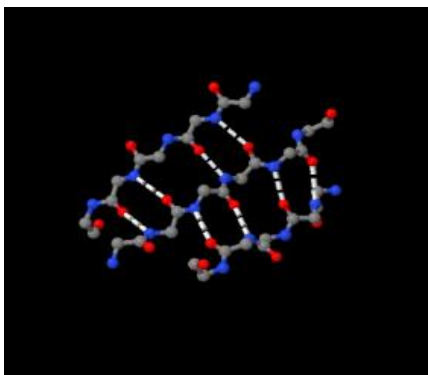
Šiems vaizdams generuoti buvo naudojama sukurta programa „pdbx-draw-secondary“, naudojanti „here-document“ kodo bloką, perduodantį komandas programai „Jmol“. Nuotraukos generuojamos naudojantis „makefile“ sistema. Šias nuotraukas galima rasti „outputs“ direktorijoje (kelias: 2as-darbas/working-copy/trunk/tests/outputs/\*.png). Kitos nuotraukos, nepavaizduotos šiame raporte, yra saugomos toje pačioje direktorijoje, jų generavimui naudojamas identiškas principas kaip ir vaizduojamų nuotraukų. Visas šiai daliai skirtas nuotraukas generuoja „case“ skriptai (nuo pdbx-draw-secondary\_15.sh iki pdbx-draw-secondary\_25.sh) esantys tests/cases/ direktorijoje.



Pav. 6 Baltymo 5A5T struktūros alfa spiralė. Liekanų ruožas 540-602, grandinė A. Apačioje turime N – terminus, viršuje C – terminus. Nuotrauką generuojantis „case“ skriptas - pdbx-draw-secondary\_022.sh.



Pav. 7 Baltymo 2C2A struktūros lygiagretus beta lakštas. Liekanų ruožas 350-355 ir 393-401, grandinė A. Pagal N – terminus ir C – terminus grandinėse, matoma, jog abiejų grandinių kryptis rodo į tą pačią pusę. Nuotrauką generuojantis „case“ skriptas - pdbx-draw-secondary\_016.sh



Pav. 8 Baltymo 6G45 struktūros antilygiagretus beta lakštas. Liekanų ruožas 484-488 ir 474-479 ir 281 – 285, grandinė B. Pagal C – terminus ir N – terminus matoma, jog vandenilniais ryšiais jungiasi grandinės, kurių kryptys eina į skirtingas puses, todėl šis lakštas vaizduoja antilygiagretų beta lakštą. Nuotrauką generuojantis „case“ skriptas - pdbx-draw-secondary\_021.sh

## Išvados

Baltymai gali būti daugiadomeniai arba turėti tik vieną domeną. Kiekvienas domenas gali būti priskirtas įvairioms „SCOP“ klasėms, šiame raporte yra analizuojamos - All alpha proteins, All beta proteins,  $\alpha + \beta$ ,  $\alpha / \beta$ . Pagal domeno struktūrą galima nuspėti jų funkcijas bei tarpusavio sąveiką. Pagal antrinę baltymų struktūrą, baltymų grandinės galima skirstyti į beta lakštus, alfa spirales, beta statines ir kitas antrines struktūras. Beta lakštus galima skirstyti į antilygiagrečius beta lakštus bei lygiagrečius beta lakštus. Siekiant atskirti koks beta lakštas yra vaizduojamas, galima numatyti pagal vandenilinių jungčių išsidėstymą bei C-terminus ir N-terminus pozicijas. Taip pat, antrinėse baltymų struktūrose galima pastebėti maišytų beta lakštų – turinčių tiek antilygiagrečius beta lakštus, tiek lygiagrečius beta lakštus.

## Literatūra

1. Jmol development team, 2016. Jmol, URL: <http://jmol.sourceforge.net/>. [Peržiūrėta: 2022-03-08 17:15 EEST]
2. H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne. (2000) The Protein Data Bank Nucleic Acids Research, 28: 235-242. URL: <https://www.rcsb.org/> [Peržiūrėta: 2022-03-10 12:43 EEST]
3. 6G45 baltymo struktūros URL: <https://www.rcsb.org/structure/6G45>  
2C2A baltymo struktūros URL: <https://www.rcsb.org/structure/2C2A>  
5A5T baltymo struktūros URL: <https://www.rcsb.org/structure/5A5T>  
2DM1 baltymo struktūros URL: <https://www.rcsb.org/structure/2DM1>  
1WGX baltymo struktūros URL: <https://www.rcsb.org/structure/1WGX>  
[Peržiūrėta: 2022-03-10 13:05 EEST]
4. Daniel Stenberg, cURL, 1996, URL: <https://curl.se/docs/manpage.html> [Peržiūrėta: 2022-03-08 20:08 EEST]
5. Murzin AG, Brenner SE, Hubbard TJP, Chothia C. 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. Journal of Molecular Biology 247:536-540. URL: <https://scop.mrc-lmb.cam.ac.uk/> [Peržiūrėta: 2022-03-15 12:42 EEST]
6. CATH: increased structural coverage of functional space. Sillitoe I, Bordin N, Dawson N, Waman VP, Ashford P, Scholes HM, Pang CSM, Woodridge L, Rauer C, Sen N, Abbasian M, Le Cornu S, Lam SD, Berka K, Varekova IH, Svobodova R, Lees J, Orengo CA. Nucleic Acids Res. 2021 Jan Pubmed: 33237325 doi: 10.1093/nar/gkaa1079. URL: <https://www.cathdb.info/> [Peržiūrėta: 2022-03-15 14:22 EEST]
7. H. Cheng, R. D. Schaeffer, Y. Liao, L. N. Kinch, J. Pei, S. Shi, B. H. Kim, N. V. Grishin. (2014) ECOD: An evolutionary classification of protein domains. PLoS Comput Biol 10(12): e1003926. URL: <http://prodata.swmed.edu/ecod/> [Peržiūrėta: 2022-03-25 18:29 EEST]