

**Temos pavadinimas:**

Skaičiuojamųjų metodų kūrimas molekulių kompleksų struktūriniais modeliams analizuoti ir vertinti

**Title of the topic:**

Development of computational methods for the analysis and assessment of structural models of molecular complexes

**Aprašymas:**

Šio doktorantūros darbo tikslas - sukurti veiksmingus skaičiuojamuosius metodus, skirtus baltymų-nukleorūgščių ir baltymų-mažųjų molekulių kompleksų struktūriniais modeliams analizuoti ir vertinti, ypatingą dėmesį skiriant duomenų trūkumo problemai šiose sąveikų srityse. Pagrindinė metodologinė darbo idėja - perkeliama mokymas, leidžiantis išnaudoti didelius ir heterogeniškus struktūrinių duomenų išteklius, viršijančius ribotą eksperimentiškai nustatytų kompleksų struktūrų skaičių, perkeliant žinias iš duomenimis turtingų šaltinių, tokių kaip baltymų struktūrų ir mažųjų molekulių kristalų duomenų bazės. Tyrime bus derinami šiuolaikiniai mašininio mokymosi metodai su skaičiuojamosios geometrijos principais pagrįstais struktūros aprašais, siekiant identifikuoti perkeliamus struktūrinius ir fizikocheminius dėsningumus, būdingus tiek izoliuotoms molekulėms, tiek molekulių kompleksams. Ypatingas dėmesys bus skiriamas metodų atsparumui struktūriniam neapibrėžtumui, efektyviam prisitaikymui prie ribotai anotuotų kompleksų duomenų bei universalių, interpretuojamų ir greitų kompleksų vertinimo metodų kūrimui.

**Abstract:**

The aim of this PhD project is to develop effective computational methods for the analysis and assessment of structural models of protein-nucleic acid and protein-small molecule complexes, with particular emphasis on overcoming data scarcity in these interaction domains. The core methodological concept of the project is transfer learning, enabling the exploitation of large and heterogeneous structural data resources beyond the limited set of experimentally determined complex structures, by transferring knowledge from data-rich sources such as databases of protein-only structures and small-molecule crystal structures. The research will integrate modern machine learning techniques with computational geometry-based representations to identify transferable structural and physicochemical patterns shared between isolated molecules and interacting systems. Special attention will be given to robustness under structural uncertainty, efficient adaptation to sparsely labeled complex data, and the development of general, interpretable, and fast scoring and assessment methods applicable to protein-nucleic acid and protein-small molecule complexes.

**Mokslinis vadovas / supervisor:** Kliment Olechnovič

