

Okruhy ke státním závěrečným zkouškám v navazujících magisterských studijních programech FCHT

Studijní program Bioinformatika a chemická informatika (N106)

Zkouška se skládá ze 4 povinných okruhů

1. **Výpočetní genomika**
(vychází z předmětů: Genomika – algoritmy a analýza, Analýza genové exprese, Fylogenetika a aplikovaná genomika)
2. **Strukturní bioinformatika a modelování**
(vychází z předmětů: Strukturní bioinformatika, Molekulové modelování)
3. **Chemoinformatika**
(vychází z předmětů: Počítačový návrh léčiv)
4. **Metody aplikované statistiky a výpočetní inteligence**
(vychází z předmětů: Statistická analýza dat, Metody výpočetní inteligence)

Výpočetní genomika

1. Sekvenační techniky a jejich porovnání: Sangerova metoda, vysoce výkonné sekvenování a jednomolekulové sekvenování . Využití pro funkční genomiku a zpracování dat.
2. Metodika kvantitativního PCR, metodiky RNA-seq, ChIP-seq, scRNA-seq, prostorově rozlišená transkriptomika: využití pro funkční genomiku. Zpracování dat.
3. Statistické metody pro funkčně genomická data: redukce dimenzionality, lineární modely, problém mnohonásobnosti testů. Návrh experimentu: využití randomizace, replikace.
4. Běžné bioinformatické datové formáty a vizualizace dat: genomové prohlížeče a databáze expresních dat. Online nástroje pro analýzu genomických dat.
5. Biologická interpretace: Genové ontologie, databáze signálních drah (KEGG). Analýza nabohacení v genových skupinách (GSEA).
6. Principy zarovnání (lokální a globální alignment; Needleman-Wunsch a Smith-Waterman algoritmus) a vícenásobného zarovnání biologických sekvencí: běžné programy (BLAST, FastA, ClustalOmega).
7. Metodiky skládání genomových sekvencí: de novo (de Bruijnovy grafy, overlap layout consensus); podle reference, mapování (Burrows-Wheeler aligner).
8. Anotace genomu: predikce genů u prokaryot a eukaryot, kódující sekvence, strukturní RNA, nekódující RNA, přiřazení biologické funkce. Projekt lidského genomu, funkční anotace v projektu ENCODE.
9. Metagenomika: ampikonové sekvenování, shotgun přístupy, anotace genů.
10. Molekulární fylogenetická analýza: molekulární evoluce; evoluce genomů; molekulární hodiny.
11. Fylogenetické stromy: evoluční modely. Metody konstrukce fylogenetických stromů (UPGMA, WPGMA, neighbour joining). Vyhodnocování kvality fylogenetických stromů (bootstrap).

12. Lidský genom, lidská populace a její genetická variabilita. Míry genetické variability, Hardy-Weinbergova rovnováha. Genetický drift, migrace a izolace, výpočetní metody měření selekce.
13. Krátké genomové varianty a jejich hledání. Genetická onemocnění, mutace v zárodečné a somatické linii. Strukturní varianty. Anotace a interpretace genetických variant.

Strukturní bioinformatika a modelování

1. Zápis molekulární struktury (kartézské a vnitřní souřadnice, valenční a torzní úhly, počítačové formáty zápisu struktur molekul)
2. Molekulová mechanika (popis meziatomových interakcí, silové pole, geometrická optimalizace, problém lokálních minim)
3. Analýza kolektivních pohybů molekul, normálové módy, vibrace, RMSD, RMSF
4. Principy molekulové dynamiky (systém, pohybové rovnice a jejich numerické řešení, periodické okrajové podmínky, Ewaldova sumace)
5. Systém v kontaktu s tepelným rezervoárem, význam teploty, soubory statistické mechaniky a jejich realizace
6. Výpočet volné energie pomocí molekulárních simulací (alchymistické simulace, metody využívající umělý potenciál nebo umělou sílu, paralelní temperování)
7. Stavební bloky proteinů – aminokyseliny, rozdělení aminokyselin, chemická struktura, fyz.-chem vlastnosti, zastoupení v proteinech, peptidová vazba – vlastnosti, konformační stavy hlavního a vedlejšího řetězce, Ramachandran plot
8. Vlastnosti proteinů odvozené od primární sekvence, základní charakteristiky, rozpustnost, hydrofobicita, charakteristika globulárních, fibrilních, membránových a nestrukturovaných proteinů
9. Proteinové databáze strukturní, proteinové databáze sekvenční, proteinové rodiny a jejich určování, databáze proteinových motivů (fold), definice proteinového motivu, přiřazení sekvence - struktura
10. Predikce sekundární struktury proteinů, vývoj predikčních metod a jejich srovnání, predikce terciární struktury proteinů - metody homologního modelování, metoda threading, ab initio predikce, principy metod a jejich srovnání, přesnost predikce
11. Sbalování proteinů – termodynamické a kinetické hledisko, misfolding, tranzitní stavy, agregace a amyloidizace, interakce protein-protein a metody pro jejich mapování a predikci.
12. Struktura nukleových kyselin (kvantitativní popis struktury DNA, motivy RNA, strukturní databáze)

Chemoinformatika

1. Farmakodynamika - molekulové síly, molekulové cíle, agonista a antagonist, kvantitativní vyjádření biologické aktivity
2. Farmakokinetika - ADME a Tox vlastnosti, metabolismus
3. Drug discovery and development - fáze, výběr molekulového cíle, výběr leadu, HTS, klinické fáze

4. Fragment based drug discovery - principy, fragmenty a jejich kombinování, efektivita ligandu, experimentální přístupy, srovnání s HTS
5. Molekuly jako grafy - konektivitní tabulka, formát MOL a SDF, Morganův algoritmus, podstrukturní vyhledávání a backtracking
6. Lineární notace zápisu chemických struktur (SMILES, InChI, InChIKey).
7. Příprava dat pro virtuální screening - structure curation and filtering
8. Bitová vektorová reprezentace struktur - molekulové klíče a hashované fingerprinty, ECFP a Morganovy fingerprinty, path fingerprinty
9. Podobnostní vyhledávání - koeficienty podobnosti, měření kvality podobnostního vyhledávání, zpřesňování výsledků podobnostního vyhledávání pomocí data fusion, výběr diverzní množiny struktur
10. Molekulové deskriptory - klasifikace, 0D, 1D, 2D a 3D deskriptory
11. Farmakoforové modelování - principy, bioisostery, scaffold hopping, farmakoforové klíče
12. Kvantitativní vztah mezi strukturou a biologickou aktivitou látky (QSAR)
13. Molekulové dokování

Metody aplikované statistiky a výpočetní inteligence

1. Popisná statistika - grafické vyjádření dat, míry centrality a variability dat
2. Rozdělení náhodné veličiny - normální a standardní normální rozdělení, Z-tabulky, výběrové rozdělení a centrální limitní teorém
3. Interval spolehlivosti
4. Základní principy testování statistických hypotéz
5. Testování hypotéz o průměru - t-test a Z-test
6. Analýza rozptylu (ANOVA)
7. Korelace a lineární regrese.
8. Neuronové sítě a gradientní učení.
9. Rekurentní neuronové sítě.
10. Konvoluční neuronové sítě.
11. Generativní sítě (GANs).
12. Autoencodery.
13. Transformery.