

SZZ – Bioinformatika / bakalářské studium

Zkušební otázky pro předmět KBC/SZZB4 Základy bioinformatiky

- 1) Definice bioinformatiky, její historický vztah ke genomice. Národní středisko pro biotechnologickou informaci, Evropská laboratoř molekulární biologie, Švýcarský ústav bioinformatiky. Cíle bioinformatiky a její postupy.
- 2) Sekvence proteinů a nukleových kyselin. Nukleotidy a báze, aminokyseliny a jejich dělení. Metody sekvenční analýzy pro nukleové kyseliny a proteiny s důrazem na moderní metody, které se užívají po roce 2000.
- 3) Formáty pro zápis aminokyselinové sekvence. Výhody formátu FASTA. Veřejně dostupné databáze sekvencí. Základní rozdělení těchto databází (primární, sekundární, kompozitní).
- 4) Vyhledávání v databázích. Nukleotidové a genomové databáze. Databáze aminokyselinových sekvencí proteinů. The Human Protein Atlas.
- 5) Možnosti získání informací ze sekvence. Konsenzuální sekvence. Vzorce formátu PROSITE. Databáze PROSITE, Interpro a jiné databáze sekvenčních motivů.
- 6) Transport proteinů v buňce. Predikce buněčné lokalizace proteinů, posttranslačních modifikací a sekundárních struktur.
- 7) Genová ontologie. Čtyři prvky datového modelu ontologie. Domény genové ontologie. Webové prohlížeče genové ontologie. Interpretace významu sady genů pomocí obohacení GO pojmů, webové nástroje.
- 8) Vztahy mezi sekvencemi, sekvenční homologie. Globální a lokální přiřazení sekvencí. Párové a mnohočetné přiřazení. Diagonála a dot plot. Substituční matice. Formáty mnohočetného přiřazení. Příklady softwaru pro přiřazení sekvencí. Sekvenční logo.
- 9) Prohledávání databází podle podobnosti se známou sekvencí. Vysvětlení algoritmu FASTA. Vysvětlení algoritmu BLAST. Vyhledávání PSI-BLAST.
- 10) Fylogenetické stromy (dendrogramy) a jejich znázornění. Konstrukce dendrogramů: znakové metody a metody vzdálenostních matic. Hodnocení kvality („bootstrapping“). Software pro konstrukci fylogenetických stromů.
- 11) Prostorové struktury proteinů. Ramachandranův diagram. Rentgenová krystalografie. Databáze Protein Data Bank, PDB formát. Databáze SCOP a SCOP2 a rozdíl mezi nimi. Prohledávání sekvenčních databází strukturním dotazem

- 12) Software pro molekulovou grafiku proteinů. Predikce proteinové struktury. Automatické modelování proteinových struktur.
- 13) Predikce struktury proteinu z transkriptu podléhajícímu sestřihu. Programy GENSCAN, Genebuilder, server NetGene2. Význam pojmu jednonukleotidový polymorfismus. Nomenklatura a predikce SNP. Databáze SNP.
- 14) Prostorová struktura RNA. Smyčky, výdutě a nepravé uzly. Databáze NDB. Význam miRNA a siRNA pro regulaci genové exprese.
- 15) Glykoproteiny. Typy glykosylace proteinů a dílčí klasifikace. Informace získávané při analýze glykoproteinů. Analýza glykosylace hmotnostní spektrometrií. Predikční servery pro glykosylaci proteinů.
- 16) Datové formáty sekvenčních dat – způsoby záznamu, kódování, reprezentace a uložení sekvenčních dat. Metody, algoritmy a nástroje informatiky při zpracování sekvenčních dat.
- 17) Datové formáty strukturních dat – způsoby záznamu, kódování, reprezentace a uložení strukturních dat. Metody, algoritmy a nástroje informatiky při zpracování sekvenčních dat.
- 18) Srovnání různých způsobů uložení dat – druhy datových souborů/úložišť, způsob přístupu k datům a jejich zpracování, výhody a nevýhody jednotlivých typů.
- 19) Limity současné výpočetní techniky. Složitost a řešitelnost algoritmů, vztah mezi náročností a objemem dat. Způsoby zpracování, výhody a omezení při paralelizaci výpočetních procesů.