

2. Praktické postupy zpracování experimentálních dat

2.1. Základní organizace práce, typy dat

Základním návykem během experimentální práce by mělo být především pečlivé a hlavně průběžné **zpracovávání výsledků**, které byly získány v rámci určitého experimentu. Není účelem nahromadit např. v rámci svěřeného diplomního tématu obrovské množství dat na dané téma, ale daleko spíše pravidelně se nad získanými daty kriticky zamýšlet a **průběžně je hodnotit**. Předejít tak lze i nechtěnému zabíhání do slepých uliček a plýtvání materiálem.

K takovému průběžnému získávání a hodnocení dat využívá vědec široké spektrum počítačových programů. K základnímu laboratornímu deníku ve formě sešitu tak přibývá často dosti různorodá směsice experimentálních dat v základních třech podobách: **číselné, obrazové a strukturní (včetně sekvenční)**.

Číselná data jsou představována buď **surovými daty**, která naměříme na našem materiálu s využitím určitého přístroje či pomůcky (spektrofotometr, pravítko, průtokový cytometr) nebo do číselné podoby **převedenými daty** (např. z densitometrie či mikroskopie).

Obrazová data představují v dnešní experimentální biologii velkou skupinu výstupů. V tom také tkví jisté úskalí má-li být zachována **exaktnost výsledků**. Technologie rekombinantní DNA, RNA, proteinová spektra a mikroskopické metody, tyto všechny přístupu poskytují obrazová data. Je proto vhodné a nanejvýš účelné vždy zařadit k obdrženému obrazovému materiálu určitou **kvantifikaci**, např. densitometrie pruhů na proteinovém gelu či kvantifikace distribuce signálu v mikroskopických pozorováních, jejich **obrazovou analýzu**.

Strukturní a sekvenční data jsou představována **analytickými daty** v podobě identifikace **struktury** či **povahy určité látky** (např. hormonu či proteinu) získané např. rentgenovou krystalografií či hmotnostní spektrometrií. Sekvenční data v podobě sekvencí nukleových kyselin (DNA, RNA) či aminokyselin (peptidy) představují obrovský zdroj informace a mohou sloužit jako základní soubor dat pro **bioinformatické analýzy**.

K orientaci v takovém různorodém souboru dat se nejlépe hodí počítač. Následující přehled není zdaleka úplný, může však sloužit k základní orientaci ve spektru používaného programového vybavení dnešních experimentálních biologických laboratoří. Vedle **specializovaných programů** řídících konkrétní zařízení a programů vyvíjených zejména **pro potřeby výzkumu** se lze v běžné laboratorní praxi setkat s **obecnějšími programy** typu textových a tabulkových editorů. Vzhledem k prodejní politice softwarové firmy Microsoft v ČR jsou její produkty na univerzitách a výzkumných ústavech oficiálně rozšířeny nejvíce, setkání s nimi v laboratoři je nejpravděpodobnější. Samozřejmě, že existují také tzv. **programy s otevřeným zdrojem (Open Source)** jako je oblíbený operační systém Linux, kancelářský balík Open Office, atd.. Jejich užívání v oblasti přenosu dat mezi uživateli však často přináší nepříjemné **problémy s kompatibilitou**. Je ovšem pravda, že tyto typy programů představují vynikající možnost ve specializovaných oblastech jako je analýza obrazu či práce se sekvenčními daty. Mohou být totiž vědeckou komunitou **dále upravovány** k dalším a dalším účelům. Poslední a stále rostoucí skupinou programů tvoří **programy přístupné online**, využívají se zejména v bioinformatice. Obecně platí, že v laboratoři by se neměli používat nelicencované

programy. Nákup software pohltí vždy nemalou část finančních prostředků laboratoře.

2.2. Průzkumníci souborů, správné ukládání dat

Dnes nejrozšířenější operační systém Windows (XP či Vista) nabízí poměrně dobrou správu souborů v podobě programu **Windows Explorer**, kde můžeme třídít soubory dle potřeby. Alternativní náhledy do adresářové struktury Windows jsou možné pomocí programů jako je oblíbený **Windows Commander**, jehož filosofie je bližší klasickým průzkumníkům s možností paralelního otevření dvou diskových prostorů na levé a pravé straně obrazovky.

Experimentátor by měl maximálně dbát na to, aby jeho data byla uložena spolehlivě a aby byla přehledně rozdělena. Výhodné je **protokolování podle data**, které koresponduje s datem v laboratorním deníku. Kvůli možnostem řazení adresářové struktury je přitom dobré dbát na **způsob zápisu data** v názvu příslušné složky. Nejvhodnější je **formát YYMMDD** (rok/měsíc/den tj. např. 081022). Takto zapsaná data lze třídít přehledně i v rámci více roků. Vhodné množství dat je pak možné pro účely zpracování kopírovat do další složky, která navíc obsahuje protokol o experimentu a jeho prezentační formu. Zásadně by však **neměl být** základní soubor dat získaný z experimentu **modifikován**, jeho data mazána či přesouvána. Nezbytné je však pravidelné **zálohování dat**. Není nutné vždy zálohovat všechna data, vhodné je např. experimenty z určitého období vypálit na několik nezávislých DVD, která je vhodné uložit na dvou místech. Cena záchrany dat z poškozených disků se pohybuje od stovek do deseti tisíců Kč. V případě nezálohování hrozí obrovská časová a finanční ztráta při opakování experimentů. Neustálé odkládání či dokonce ignorace zálohování hrozí problémem. Častou chybou bývá podcenění zapsání a uložení dat do počítače ihned po jejich získání a odložení tohoto na později. Experimentátor poté za čas zjišťuje, že mu v záznamech určitá data chybí a nechápe jak k tomu došlo. **Dokončit experiment** či jeho část s **perfektní dokumentací** patří ke cti správného experimentátora.

2.3. Tabulkové procesory, statistické programy, editory grafů

Velmi často je potřeba získaná experimentální data zpracovat hromadným způsobem. V tom případě je nejběžnější využít software, který bude zvládat jak přehledné uložení dat a jejich přepočty, tak základní statistické zhodnocení a vytvoření grafu. Všechny tyto nároky v základu splňuje nejrozšířenějším tabulkový procesor **Microsoft Excel**. V případě zpracování náročnějších statistických zpracování dat je lépe využít specializovaný software jako je **Statgraphics, NCSS**, apod.. Výhodou využívání tabulkového procesoru je v běžné praxi zejména možnost automatizovat stále se opakující procedury a výpočty. **Experimentální design a vlastně celé protokolování** je poté možno vytvářet přímo v jednotlivých listech Excelu. Samozřejmostí je vytvářet přehledné tabulky pro tisk a sestavy výsledků podle předem zvolených kritérií.

Ačkoliv v Excelu je možné vytvářet grafy, k tomuto účelu je lépe využít specializovaný software jako je např. **Sigma Plot**. Platí, že zdaleka nejrozšířenějšími grafy v laboratorní praxi jsou grafy **sloupcové, spojnicové a histogramy**. Sigma plot umožňuje vytvářet i další typy grafů, prakticky bez omezení. Výhodou je, že jako datová tabulka může sloužit i datová tabulka Excelu.

2.4. Zpracování strukturních a sekvenčních dat

Zatímco zpracování experimentálně získaných strukturních dat např. z rentgenové krystalografie je spíše náročnější matematicko-fyzikální aplikací, práce se sekvencemi DNA, RNA a proteinů je široce rozšířena. Pro práci se sekvencemi v podstatě stačí jednoduchý program typu **Notepad**, ve kterém můžeme provádět jednoduché úpravy zápisu sekvence a případně shromažďovat více sekvenčních dat. Pro náročnější nakládání se sekvenčními daty (alignment sekvencí, mapy vektorů, navrhování primerů) je vhodné využít některé z **online zdrojů** (bude o nich řeč podrobně v kapitole 4). Velmi doporučit lze také software **Vector NTI (Invitrogen)**, který může posloužit jako kompletní prostředí pro práci se sekvencemi. Jeho součástí je mnoho nástrojů **bioinformatické analýzy** (viz. kap. 4) jako jsou porovnávání sekvencí, restriční analýza, sledování evoluční příbuznosti sekvencí, klonování *in silico*, vytváření grafických map plasmidů a sekvencí, navrhování primerů pro PCR a mnoho dalších. tento software je pro akademickou obec dostupný zdarma.

2.5. Analyzátoři obrazu, grafické programy, prezentační programy

Obrazová data jsou nejčastěji získávána jako výsledek mikroskopických a makroskopických pozorování. Software sloužící k zachycení obrazu z kamery a jeho převedení do počítače souží v mnoha případech také jako tzv. **analyzátor obrazu**. Analýzou obrazu se rozumí určitá kvantifikace obrazové scény a tato kvantifikace se dá automatizovat pro hodnocení poměrně rozsáhlé série obrazů. Jeden z celosvětově nejrozšířenějších obrazových analyzátorů je volně šiřitelný software **ImageJ (NIH)**. V ČR je poměrně rozšířen obrazový analyzátor **NIS Elements** (Laboratory Imaging, Praha) vyvíjený pro mikroskopy Nikon. Analyzátoři obrazu jsou v mnoha podobách a vybavení každého solidnějšího software obsluhujícího fluorescenční či konfokální mikroskopy (např. **Leica, Zeiss, Olympus, Nikon**).

Grafické programy typu **Adobe Photoshop** či **Corel Photopaint** slouží k nezbytné úpravě fotodokumentace, neměly by však být využívány k dodatečným úpravám ve smyslu úpravy kontrastu jedné části scény na úkor druhé či umělého vytváření scén sestavených z více snímků. V případě gelů a blotů je pak nanejvýš vhodné zachovat ho celý a **příznat případné nedokonalosti**.

Programy pro sestavování výsledkových tabulí, jako jsou **Corel Draw** či **Adobe Illustrator**, jsou pomyslnou koncovkou před vlastní prezentací výsledků v původním článku. Představují však také ideální možnost jak si připravit výsledky pro vlastní orientaci v experimentování. Do této kategorie programů náleží také program **Adobe Acrobat a Acrobat Reader**. Tyto programy využívají vhodně komprimovaný formát **pdf**, který zachovává v rozumné kvalitě jak vektorovou grafiku, tak fotodokumentaci. Prezentační programy typu **Microsoft Power Point** umožňují účelným způsobem referovat o vlastních výsledcích a závěrech. Více o využívání grafických a prezentačních programů pro přípravu publikací v kap. 5.