

2. Programy a postupy při zpracování experimentálních dat

2.1 Základní organizace práce, typy dat

Základním návykem během experimentální práce by mělo být 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.

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 podobách, **číselné, strukturní (včetně sekvenční) a obrazové**.

- **číselná data** - mohou představovat buď **surová data**, která naměříme na našem materiálu s využitím určitého přístroje či pomůcky (spektrofotometr, pravítko, průtokový cytometr) nebo to jsou do číselné podoby **převedená data** (např. densitometrie).

- **strukturní a sekvenční data** - jedná se o **analytická data** v podobě identifikace struktury 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**.

- **obrazová data** - v dnešní experimentální biologii je většina výstupů, které v laboratoři získáváme v podstatě obrazového charakteru. V tom také tkví jisté úskalí má-li být zachována **exaktnost výsledků**. Technologie rekombinantní DNA, RNA, proteinová spektra, mikroskopické metody, tyto všechny přístupu poskytují obrazová data. Je proto vhodné a nenajvýš účelné vždy zařadit k obdrženému obrazovému materiálu určitou **kvantifikaci**, např. densitometrie pruhů na proteinovém gelu či kvantifikace distribuce signálu v mikroskopických pozorováních.

K orientaci v takovém různorodém souboru dat se nejlépe hodí, jak jinak, počítač. Následující přehled není zdaleka úplný, ale může snad sloužit k základní orientaci ve spektru používaného programového vybavení dnešních experimentálních biologických laboratoří. Vzhledem k tomu, že produkty firmy Microsoft jsou v ČR na univerzitách a výzkumných ústavech oficiálně rozšířeny nejvíce (vzhledem k prodejní politice Microsoftu), setkání s nimi je nejpravděpodobnější. Samozřejmě, že existují také tzv. **Open Source** programy (operační systém Linux, kancelářský balík Open Office, atd.), jejich užívání však často nese nepříjemné problémy s kompatibilitou.

2.1 Průzkumníci souborů, správné ukládání dat

Dnes nejrozšířenější operační systém Windows XP nabízí poměrně dobrou správu souborů v podobě programu **Windows Explorer**, kde můžeme třídit soubory dle potřeby. Alternativní náhledy do adresářové struktury Windows jsou možné pomocí programů jako je oblíbený **Windows Commander**. 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. Vhodné množství dat je pak možné kopírovat do složky pro přípravu publikace apod. 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 nezalohování 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.

2.2 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 náročnějších statistických zpracování dat je lépe využít specializovaný software jako je **Statgraphics, NCSS**, apod.. Na vlastní editaci grafů se také vyplatí využívat program **Sigma Plot**, vzhledem k tomu, že Excel nemá editaci grafů dobře uživatelsky implementovanou.

Výhodou využívání **Excelu** 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.

Sigma Plot je pokročilým editorem grafů. 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 spreadsheet Excelu.

2.3 Zpracování strukturních a sekvenčních dat

Zatímco zpracování experimentálně získaných strukturních dat např. z rentgenové krystalografie je spíše náročnější matematicko-fyzikální aplikací, práce s jednoduššími strukturními sekvencemi 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).

2.4 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 scény a tato kvantifikace se dá automatizovat a hodnotit tak poměrně rozsáhlé série obrazů. Jeden z celosvětově nejrozšířenějších obrazových analyzátorů je volně šiřitelný software **ImageJ (NIH)**. V ČR je to pak zejména program **Lucia (či NIS Elements)**. Analyzátoři obrazu jsou v mnoha podobách a vybavení každého solidnějšího software obsluhujícího konfokální mikroskopy (např. **Leica, Zeiss**).

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 přiznat 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. Konečně, prezentační programy typu **Microsoft Power Point** umožňují účelným způsobem referovat o vlastních výsledcích a závěrech.