

# ***Strukturní biochemie***

Skripta k přednášce C9530 Přírodovědecké fakulty Masarykovy univerzity

*Lukáš Židek*

*Na přípravě se podíleli Jiří Damborský, Eva Fadrná, Jaromír Marek,  
Markéta Munzarová, Pavel Plevka*

*1. prosince 2022*



# Obsah

<b>I</b>	<b>Procházka ZOO struktur</b>	<b>1</b>
<b>1</b>	<b>Struktura v chemii a biochemii</b>	<b>3</b>
1.1	Konfigurace a konformace . . . . .	3
1.2	Geometrický popis molekul . . . . .	4
1.3	Definice torzního úhlu . . . . .	7
<b>2</b>	<b>Struktura proteinů</b>	<b>9</b>
2.1	Proteiny jako příklad biomakromolekul . . . . .	9
2.2	Stavební jednotky . . . . .	9
2.3	Torzní úhly . . . . .	12
2.4	Primární struktura . . . . .	14
2.5	Sekundární struktura . . . . .	15
2.6	Struktura alfa . . . . .	15
2.7	Struktura beta . . . . .	17
2.8	Terciární struktura . . . . .	19
2.9	Spojovací prvky . . . . .	20
2.10	Hierarchie v rámci terciární struktury . . . . .	23
2.11	Alfa-proteiny . . . . .	23
2.12	Beta-proteiny . . . . .	28
2.13	Alfa+beta-proteiny . . . . .	38
2.14	Alfa/beta-proteiny . . . . .	38
2.15	Uspořádané proteiny bez sekundární struktury . . . . .	41
2.16	Kvarterní struktura a vyšší struktury . . . . .	41
<b>3</b>	<b>Struktura nukleových kyselin</b>	<b>45</b>
3.1	Stavební jednotky . . . . .	45
3.2	Torzní úhly . . . . .	47
3.3	Primární struktura . . . . .	48
3.4	Sekundární struktura . . . . .	48
3.5	Terciární struktura . . . . .	51
3.6	Kvarterní struktura a vyšší struktury . . . . .	56
<b>4</b>	<b>Struktura oligo- a polysacharidů</b>	<b>57</b>
4.1	Stavební jednotky . . . . .	57
4.2	Torzní úhly . . . . .	57
4.3	Primární struktura . . . . .	59
4.4	Sekundární struktura . . . . .	59
4.5	Terciární struktura . . . . .	59

<b>5</b>	<b>Struktura biologických membrán</b>	<b>63</b>
5.1	Membrány a biomakromolekuly . . . . .	63
5.2	Stavební jednotky . . . . .	63
5.3	Geometrie lipidových útvarů . . . . .	64
5.4	Tekutost membrán . . . . .	64
5.5	Vliv polární skupiny . . . . .	64
5.6	Vliv hydrofobního řetězce . . . . .	65
5.7	Vliv složení . . . . .	66
5.8	Vyšší struktury . . . . .	66
<b>6</b>	<b>Interakce určující strukturu biomakromolekul</b>	<b>67</b>
6.1	Energie a entropie . . . . .	67
6.2	Síly působící v biomakromolekulách . . . . .	68
6.3	Kovalentní vazby . . . . .	69
6.4	Van der Waalsovy interakce . . . . .	70
6.5	Elektrostatické interakce . . . . .	71
6.6	Vodíkové vazby . . . . .	71
6.7	Stacking . . . . .	72
6.8	Karbonylové interakce . . . . .	72
6.9	Voda jako rozpouštědlo . . . . .	72
6.10	Elektrostatické interakce ve vodě . . . . .	72
6.11	Vodíkové vazby ve vodě . . . . .	75
6.12	Hydrofobní efekt . . . . .	76
6.13	Denaturace . . . . .	78
6.14	Celková bilance . . . . .	78
6.15	Termodynamika a kinetika . . . . .	81
<b>II</b>	<b>Lov struktur</b>	<b>83</b>
<b>7</b>	<b>Úvod do metod strukturní biochemie</b>	<b>85</b>
<b>8</b>	<b>Příprava vzorku biomakromolekul</b>	<b>87</b>
8.1	Izolace z přírodního materiálu . . . . .	87
8.2	Metody molekulární biologie . . . . .	88
8.3	Expresní vektory . . . . .	89
8.4	Molekulové klonování . . . . .	90
8.5	Místně řízená mutagenese . . . . .	91
8.6	Bakteriální exprese . . . . .	91
8.7	Purifikace exprimovaných proteinů . . . . .	92
8.8	Chemická syntéza peptidů . . . . .	92
8.9	Chemická syntéza oligonukleotidů . . . . .	94
8.10	Chemická syntéza oligosacharidů . . . . .	94
8.11	Enzymová syntéza nukleových kyselin . . . . .	94
8.12	Enzymová syntéza proteinů . . . . .	95
8.13	Enzymová syntéza oligosacharidů . . . . .	95



<b>9</b>	<b>Metody určování sekvence</b>	<b>97</b>
9.1	Sekvenace proteinů . . . . .	97
9.2	Edmanovo odbourávání . . . . .	98
9.3	Sekvenace hmotnostní spektrometrií . . . . .	98
9.4	Sekvenace nukleových kyselin . . . . .	99
9.5	Chemická metoda Maxama a Gilberta . . . . .	99
9.6	Enzymová neboli dideoxymetoda (Sangerova) . . . . .	101
9.7	Sekvenace oligosacharidů . . . . .	101
<b>10</b>	<b>Výpočetní metody</b>	<b>107</b>
10.1	Cíl výpočetních metod ve strukturní biochemii . . . . .	107
10.2	Modely biomakromolekul . . . . .	108
10.3	Molekulová mechanika . . . . .	108
10.4	Hyperplocha potenciální energie . . . . .	110
10.5	Silové pole . . . . .	111
10.6	Implicitní a explicitní solvent . . . . .	111
10.7	Výpočet elektrostatických interakcí . . . . .	113
10.8	Hledání lokálního minima energie . . . . .	113
10.9	Prohledávání plochy potenciální energie . . . . .	114
10.10	Molekulová dynamika . . . . .	114
<b>11</b>	<b>Optické metody</b>	<b>117</b>
11.1	Elektromagnetické vlnění . . . . .	117
11.2	Popis elektromagnetického vlnění . . . . .	118
11.3	Světlo a molekula . . . . .	122
11.4	Světlo a krystal . . . . .	124
11.5	Světlo a molekuly v roztoku či skle . . . . .	128
11.6	Malouhlový rozptyl Röntgenova záření (SAXS) . . . . .	128
11.7	Absorpce . . . . .	130
11.8	Církulární dichroismus . . . . .	133
11.9	Vibrace jader . . . . .	135
11.10	Ramanova spektra . . . . .	139
<b>12</b>	<b>Rentgenová krystalografie</b>	<b>141</b>
12.1	Virtuální náhrada mikroskopu . . . . .	141
12.2	Braggovy roviny . . . . .	143
12.3	Difrakce na krystalu . . . . .	145
12.4	Strukturní faktor . . . . .	146
12.5	Krystalizace biomakromolekul . . . . .	147
12.6	Měření difrakce . . . . .	149
12.7	Vyhodnocení dat . . . . .	150
12.8	Řešení fázového problému . . . . .	151
12.9	Pattersonova mapa . . . . .	152
12.10	Metoda molekulárního přemístění . . . . .	152
12.11	Přímé metody . . . . .	153
12.12	Metody izomorfního nahrazení . . . . .	154
12.13	Metody využívající anomální rozptyl . . . . .	155
12.14	Určení struktury z difrakčních dat . . . . .	158
12.15	Kontrola správnosti . . . . .	160

<b>13 Elektronová mikroskopie</b>	<b>161</b>
13.1 Elektrony jako vlnění . . . . .	161
13.2 Konstrukce elektronového mikroskopu . . . . .	162
13.3 Výhody a nevýhody elektronové mikroskopie . . . . .	162
13.4 Optika a rozlišení mikroskopů . . . . .	164
13.5 Příprava vzorků pro elektronovou mikroskopii . . . . .	166
13.6 Typy vzorků v elektronovou mikroskopii . . . . .	168
13.7 Tvorba a zpracování digitálního obrazu . . . . .	169
13.8 Amplitudový a fázový kontrast . . . . .	170
13.9 Poškození vzorku . . . . .	174
13.10 Porovnávání dvojrozměrných obrazů . . . . .	174
13.11 Rekonstrukce trojrozměrné struktury biomakromolekul . . . . .	176
<b>14 Nukleární magnetická rezonance</b>	<b>181</b>
14.1 Magnetické chování jader . . . . .	181
14.2 Ovlivnění souborů jader vnějšími poli . . . . .	182
14.3 Excitace . . . . .	183
14.4 Interakce jader s nejbližším okolím . . . . .	185
14.5 Interakce s magnetickými poli párovaných elektronů . . . . .	185
14.6 Přímá interakce s magnetickými dipóly sousedních jader . . . . .	187
14.7 Vliv sousedních jader zprostředkovaný elektrony vazeb . . . . .	189
14.8 Spektrometr a experiment NMR . . . . .	191
14.9 Vícerozměrná spektroskopie a korelace frekvencí . . . . .	193
14.10 Postup strukturní analýzy proteinů pomocí NMR . . . . .	196
14.11 Sekvenční přiřazení . . . . .	196
14.12 Přiřazení postranních řetězců . . . . .	198
14.13 NMR spektra a struktura proteinů . . . . .	199
14.14 Sekundární struktura . . . . .	202
14.15 Výpočet konformace proteinu . . . . .	202
14.16 Struktura nukleových kyselin pomocí NMR . . . . .	204
14.17 Struktura oligosacharidů pomocí NMR . . . . .	208
14.18 Pohyby molekul v NMR . . . . .	208
14.19 Omezení metody a pokročilé aplikace . . . . .	209
<b>15 Bioinformatika</b>	<b>211</b>
15.1 Co to je bioinformatika . . . . .	211
15.2 Genomové projekty . . . . .	211
15.3 Biologické databáze . . . . .	212
15.4 Databáze sekvencí proteinů . . . . .	214
15.5 Sekundární databáze sekvencí proteinů . . . . .	214
15.6 Databáze sekvencí DNA . . . . .	215
15.7 Databáze proteinových struktur . . . . .	216
15.8 Párové přiložení (pairwise alignment) sekvencí . . . . .	217
15.9 Vícenásobné přiložení (multiple alignment) . . . . .	218
15.10 Předpovídání struktury ze sekvence . . . . .	220
15.11 Předpověď sekundární struktury . . . . .	220
15.12 Předpověď terciární struktury . . . . .	220

<b>III</b>	<b>Dodatky</b>	<b>225</b>
A	Energie různých interakcí . . . . .	227
A.1	Energie dvou nábojů . . . . .	227
A.2	Energie náboje a elektrického dipólu v jedné molekule . . . . .	227
A.3	Energie náboje a elektrického dipólu v různých molekulách . . . . .	229
A.4	Energie dvou elektrických dipólů v jedné molekule . . . . .	230
A.5	Energie dvou elektrických dipólů v různých molekulách . . . . .	234
A.6	Energie náboje a indukovaného elektrického dipólu . . . . .	236
A.7	Energie permanentního a indukovaného elektrického dipólu . . . . .	236
A.8	Energie dvou indukovaných elektrických dipólů . . . . .	237
B	Metody chemické syntézy na pevném nosiči . . . . .	239
B.1	Karbodiimidová metoda syntézy peptidů . . . . .	239
B.2	Fosforamiditová metoda syntézy nukleotidů . . . . .	239
C	Kvantové metody . . . . .	240
C.1	Kvantová mechanika atomů . . . . .	240
C.2	Molekulové orbitaly . . . . .	241
C.3	Metody <i>ab initio</i> . . . . .	243
C.4	Semiempirické metody . . . . .	246
D	Peptidová vazba jako příklad chromoforu . . . . .	247
D.1	Atomové orbitaly . . . . .	247
D.2	Hybridní orbitaly . . . . .	247
D.3	Molekulové orbitaly . . . . .	250
D.4	Energie stavů a vlnové délky přechodů . . . . .	252
D.5	Tranzitní dipóly a pravděpodobnosti přechodů . . . . .	253
E	Světlo a sekundární struktura . . . . .	255
F	Anomální rozptyl . . . . .	256
G	Skalární interakce a přiřazení spekter NMR nukleových kyselin . . . . .	259



Část I

**Procházka ZOO struktur**



# Kapitola 1

## Struktura v chemii a biochemii

### 1.1 Konfigurace a konformace

Chceme-li studovat strukturu biologicky zajímavých látek, musíme se nejdříve dohodnout na tom, co vlastně pojmem *struktura* myslíme. Chemické látky se skládají z atomů různých prvků. První informací, kterou o neznámé látce můžeme získat, je tedy údaj, kolik atomů jednotlivých prvků molekula látky obsahuje. Například můžeme zjistit, že molekula naší látky obsahuje čtyři atomy uhlíku, dva atomy kyslíku a osm atomů vodíku. Budeme znát *složení* látky<sup>1</sup>, ale stále ještě nebudeme znát strukturu látky.

Teprve když zjistíme, které atomy jsou navzájem spojeny kovalentní vazbou<sup>2</sup>, můžeme říci, zda zkoumanou látkou je kyselina máselná, 2-buten-1,4-diol, nebo třeba 3-hydroxybutanon.

Ani teď však nemusí být jednoznačně dáno, o jakou chemickou látku jde. Předpokládejme, že řekneme, že zkoumaná látka je 2-buten-1,4-diol. Tímto názvem lze ale označit dvě různé chemikálie s různými vlastnostmi. Tyto dvě látky se budou lišit uspořádáním atomů kolem dvojně vazby. Budou-li obě koncové vazby C–C na jedné straně dvojně vazby, půjde o izomer *cis* (chemicky úplný název látky bude (*Z*)-2-buten-1,4-diol), budou-li koncové C–C vazby směřovat na opačné strany, půjde o izomer *trans* ((*E*)-2-buten-1,4-diol).

Pokud by zkoumanou látkou byl 3-hydroxybutanon, měli bychom podobný problém. Museli bychom si vybrat ze dvou molekul, z nichž jedna je zrcadlovým obrazem druhé<sup>3</sup>. Úplný chemický popis by tedy byl (*R*)-3-hydroxybutanon nebo (*S*)-3-hydroxybutanon.

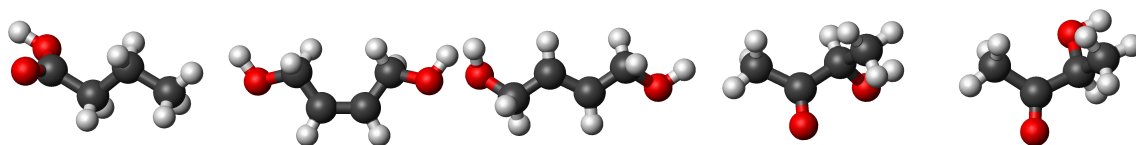
Způsob vzájemného uspořádání atomů propojených kovalentními vazbami, tak jak jsme jej probírali v předchozích odstavcích, se nazývá *konfigurace*. Zatímco kyselina máselná se vyskytuje v jediné konfiguraci, 2-buten-1,4-diol a 3-hydroxybutanon mají po dvou různých konfiguracích (obrázek 1.1).

Se znalostí konfigurace již umíme popsat molekulu například chemickým vzorečkem (i s vyznačením chiralit). Pokud budeme chtít z trubiček a kuliček sestavit model zkoumané látky (jako na obrázku 1.1), třeba kyseliny máselné, potřebujeme znát víc, než jen konfiguraci. Musíme se například rozhodnout, jestli uhlíkový řetězec ohneme tak, aby byla karboxylová kyselina co nejbliže koncovému methyly, nebo jestli bude uhlíkový řetězec připomínat klikový hřídel a karboxylová skupina bude od methylové co nejdále. Konkrétní rozmístění atomů v prostoru se nazývá *konformace*. Mezi konfigurací a konformací je důležitý rozdíl. Konformaci našich molekul můžeme měnit pouhým otáčením kolem kovalentních

<sup>1</sup>Chemici často mluví o *kompozici* látky.

<sup>2</sup>Informace udávající pouze to, které atomy jsou spojeny, se nazývá *konektivita*. Chemika ovšem zajímá i to zda jde o vazbu jednoduchou nebo dvojnou. Chemický popis atomů a vazeb v molekule se nazývá *konstituce*.

<sup>3</sup>Učeně řečeno, jde o *antipody*, tedy *enantiomery* s opačnou *chiralitou*. Takové látky se liší směrem, kterým stáčí rovinně polarizované světlo.



Obrázek 1.1: Molekuly kyseliny máselné, (*Z*)-2-buten-1,4-diolu, (*E*)-2-buten-1,4-diolu, (*R*)-3-hydroxybutanonu a (*S*)-3-hydroxybutanonu.

vazeb, zatímco ke změně konfigurace je třeba kovalentní vazbu přerušit a vytvořit novou.<sup>4</sup>

Konfigurace a konformace hrají důležitou roli v rozdílech chápání pojmu *určení struktury* v kontextu malých a velkých molekul. Pokud izolujeme z přírodního zdroje nějakou nízkomolekulární látku (tedy látku, jejíž molekuly se skládají z desítek atomů), například antibiotikum, musíme především určit konfiguraci této látky. Určování konformace je v takovém případě okrajová záležitost. U malých molekul, jako je kyselina máselná, jsou rotace kolem vazeb tak snadné a rychlé, že nemá většinou smyslu o jednotlivých konformerech mluvit. Složitější molekuly se sice vyskytují v jedné či více z několika jasně definovaných konformací, určení těchto konformací ale stále představuje jednodušší část určování struktury. Určováním struktury malých molekul se tedy myslí především určování konfigurace.

Situace je zcela odlišná, pokud izolujeme z přírodního zdroje makromolekulu, například protein. Biomakromolekuly se skládají z několika málo druhů monomerních jednotek, jejichž konfigurace je dobře známá. Stačí tedy určit pořadí (*sekvenci*) těchto stavebních jednotek, a budeme znát konfiguraci celé makromolekuly. Určování sekvence je dnes poměrně snadné, zvláště u nukleových kyselin a proteinů. U velkých molekul je ale velmi důležité určit konformaci. Otáčení kolem některých vazeb je totiž velmi ztíženo, ba téměř znemožněno, zbytkem makromolekuly. K tomu, abychom získali představu o tvaru molekuly, musíme zjistit, v které konformaci jsou uvězněny ty vazby, kolem kterých nemůžeme otáčet. Určováním struktury biomakromolekul se proto myslí především určování konformace. A to je úkol obtížný.

## 1.2 Geometrický popis molekul

Z matematického pohledu k určení propojení atomů stačí znát *topologii*<sup>5</sup> molekuly, zatímco určení konformace vyžaduje znalost přesné *geometrie*. Pokud jsme si v předchozím odstavci řekli, že určování struktury biomakromolekul je vlastně popis jejich konformace, musíme se nejprve seznámit s geometrickým popisem molekul.

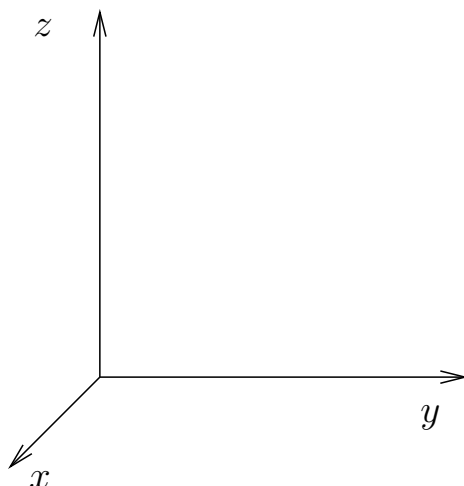
Geometrie nám umožňuje vytvořit zjednodušený obraz molekul, tedy jejich *model* (zde slovo „model“ znamená opravdu to, co si pod ním zpravidla představíme). Pro potřeby strukturní biochemie většinou stačí považovat atomy za body v prostoru a chemické vazby za úsečky, které body spojují. Uvědomme si, že k vytvoření modelu nestačí znát pouze polohy všech atomů, ale musíme také vědět, mezi kterými atomy je kovalentní vazba (musíme určit topologii molekuly).

Polohu atomů v prostoru můžeme určit několika způsoby. Prvním způsobem, který nás asi napadne,

<sup>4</sup>V některých případech není ke změně konfigurace třeba přerušit kovalentní vazbu, ale pouze deformovat vazebné úhly, tento detail není ale pro pochopení pojmu konformace zásadní.

<sup>5</sup>Topologie popisuje vlastnosti geometrických útvarů, které se nemění, když geometrický útvar vystavíme pružné deformaci (natažení nebo zkroucení), která je homeomorfní (každému bodu útvaru před deformací odpovídá jeden bod útvaru po deformaci). Zjednodušeně řečeno, topologie popisuje kvalitativní vlastnosti, zatímco geometrie udává kvantitativní míry. Oblíbeným příkladem rozdílu mezi topologií a geometrií je problém lajnování tenisového hřiště. Topologie nám pomůže vyřešit otázku, jestli lze hřiště nalajnovat jedním tahem, a pokud ne, kolikrát budeme muset lajnovačku přenést. Geometrie nám řekne, jak dlouhé musí jednotlivé čáry být a že spolu musí svírat úhel 90 °.





Obrázek 1.2: Kartézská souřadná soustava. Osa  $x$  míří k nám, osa  $y$  doprava, osa  $z$  nahoru.

je udat souřadnice  $x$ ,  $y$ ,  $z$  každého atomu v obvyklé kartézské souřadné soustavě<sup>6</sup> (obrázek 1.2). Tento způsob používá například databáze struktur PDB (Protein Data Bank), jak ukazuje příklad na obrázku 1.3. Pro popis modelu molekuly, která je tvořena  $N$  atomy, potřebujeme  $3N$  souřadnic.

ATOM	1	O	WAT	0	-0.267	0.269	0.096
ATOM	2	H1	WAT	0	0.021	0.014	-0.781
ATOM	3	H2	WAT	0	0.245	-0.283	0.686



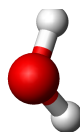
Obrázek 1.3: Soubor popisující polohy atomů vody ve formátu, který používá PDB (souřadnice jsou uvedeny v posledních třech sloupcích v jednotkách ångström, což je desetina nanometru).

Výhoda kartézských souřadnic je v tom, že je můžeme přímo a jednoduše použít k zobrazení modelu molekuly. Kartézské souřadnice mají ale také své nevýhody. Představme si, že molekulu definovanou na obrázku 1.3 otočíme o  $90^\circ$  kolem osy  $x$ . Polohy atomů pak budou udány souřadnicemi uvedenými na obrázku 1.4. Jak je vidět, hodnoty dvou souřadnic každého atomu se podstatně změnily – přitom jde stále o tutéž molekulu vody, která má naprosto stejný tvar. Proto se často polohy atomů udávají jiným způsobem. Pro molekulu skládající se z  $N$  atomů opět potřebujeme  $3N$  údajů. Z těchto  $3N$  hodnot budou tři udávat polohu těžiště molekuly, další tři natočení molekuly jako celku vzhledem k souřadné soustavě a zbývající údaje (kterých je pochopitelně  $3N - 6$ ) *vzájemnou* polohu atomů uvnitř molekuly<sup>7</sup>. Prvních šest údajů nás zpravidla nezajímá, neboť se mění, jak se molekula pohybuje Brownovým pohybem. Zajímá nás pouze posledních  $3N - 6$  údajů, které jsou pro danou molekulu stejné, ať už je její poloha v souřadné soustavě jakákoli. Tyto údaje se proto nazývají *vnitřní souřadnice* molekuly. Jako vnitřní souřadnice je zvykem udávat vzdálenosti mezi atomy, které jsou vázány kovalentní vazbou (*vazebné vzdálenosti*) a úhly, které kovalentní vazby svírají (*vazebné úhly*). Model molekuly vody, který jsme si uvedli jako příklad, je určen dvěma vazebnými vzdálenostmi  $|O-H1|$  a  $|O-H2|$ , které jsou obě rovny 95,7 pm a jedním vazebným úhlem  $\angle H1-O-H2$ , který je roven  $104,5^\circ$ .

<sup>6</sup>Soustava souřadná musí být definována tak, jak ukazuje obrázek 1.2. Kdybychom otočili směr jedné osy, popisovali bychom opačný enantiomer.

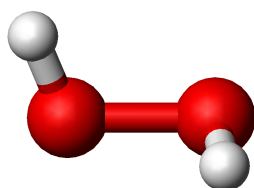
<sup>7</sup>Uvedené počty platí pouze pro molekulu, která není lineární. Lineární biomakromolekuly se ovšem nevyskytují, takže se jimi nemusíme zabývat.

ATOM	1	O	WAT	0	-0.267	-0.096	0.269
ATOM	2	H1	WAT	0	0.021	0.781	-0.014
ATOM	3	H2	WAT	0	0.245	-0.686	-0.283



Obrázek 1.4: Soubor popisující polohy atomů molekuly vody, která je otočena o  $90^\circ$  vzhledem k molekule definované v souboru na obrázku 1.3.

Vazebné vzdálenosti a vazebné úhly postačují k popisu tříatomových molekul, jejichž atomy leží v jedné rovině. Co se ale stane, má-li molekula atomů více? Zkusme se podívat na jednoduchý příklad – molekulu peroxidu vodíku (obrázek 1.5). Tato molekula se skládá ze čtyř atomů, pro její popis tedy potřebujeme 12 souřadnic. Šest z nich popisuje polohu molekuly jako celku, ty nás nezajímají. Z vnitřních souřadnic můžeme určit tři vazebné vzdálenosti,  $|O1-H1|$  a  $|O2-H2|$ , které jsou dlouhé 95,0 pm a  $|O1-O2|$ , která je dlouhá 147,5 pm. Dále můžeme určit dva vazebné úhly,  $\angle H1-O1-O2$  a  $\angle O1-O2-H2$ , oba jsou rovny  $94,8^\circ$ . Těmito pěti souřadnicemi není ještě poloha čtyř atomů peroxidu vodíku v prostoru jednoznačně dána, stále můžeme točit modelem kolem vazby O–O, aniž bychom měnili vazebné úhly a vzdálenosti. Šestou vnitřní souřadnicí, kterou musíme určit, je tedy *torzní úhel* (úhel otočení<sup>8</sup> kolem vazby O–O). Tato poslední souřadnice je pro definici struktury biomakromolekul nejdůležitější. Vazebné délky a vzdálenosti jsou totiž téměř stejné ve všech monomerních jednotkách, torzní úhly se ale mohou značně lišit. Můžeme tedy říci, že *konformace biomakromolekul je dána hodnotami torzních úhlů*.



Obrázek 1.5: Molekula peroxidu vodíku.

Vazebná délka jednoznačně popisuje dvojici atomů. Protože dvojice bodů musí ležet v jedné přímce, jde z principu o jednorozměrný případ. Podobně vazebný úhel popisuje trojici atomů, která vždy leží v jedné rovině (dvourozměrný případ). Konečně torzní úhel definuje čtveřici atomů a rozšiřuje popis na trojrozměrný prostor. Protože je naším cílem postavit trojrozměrný model molekuly, s žádným dalším typem vnitřních souřadnic se už nesetkáme, i když budeme dále zvětšovat počet atomů.

Stavbu modelu molekuly si tedy můžeme popsat následujícím způsobem:

- Polohu prvního atomu (označme si jej A) udávají tři souřadnice, které tímto definují polohu celé molekuly v souřadné soustavě. Jsou to souřadnice vnější, které nás většinou nezajímají, v případě jediného atomu nemá smysl mluvit o vnitřních souřadnicích molekuly.
- Polohu druhého atomu (B) udává jedna vnitřní souřadnice, vazebná délka  $|A-B|$ , a dvě vnější souřadnice, které definují natočení první vazby vůči souřadné soustavě.

<sup>8</sup>Jiná možnost, jak tento úhel popsat, je určit úhel, který svírají roviny, ve kterých leží atomy H1,O1,O2 a O1,O2,H2. Jako vnitřní souřadnici bychom mohli stejně dobře použít *dihedrální úhel*, tedy doslova úhel mezi dvěma stěnami.

- Polohu třetího atomu (C) udávají dvě vnitřní souřadnice, vazebná délka  $|B-C|$  a vazebný úhel  $\angle A-B-C$ , a jedna vnější souřadnice, která udává natočení druhé vazby vůči souřadné soustavě (teprve tím je definováno natočení celé molekuly v soustavě souřadnic).
- Polohu čtvrtého atomu (D) udávají tři vnitřní souřadnice. Pokud jej připojíme k atomu C, bude to vazebná délka  $|C-D|$ , vazebný úhel  $\angle B-C-D$ <sup>9</sup> a torzní úhel popisující rotaci kolem vazby B-C.
- Polohu všech dalších atomů udávají vždy tři vnitřní souřadnice, v případě nevětveného řetězce (lineárního nebo cyklického) délka vazby s předchozím atomem, úhel mezi dvěma předchozími vazbami a torzní úhel popisující rotaci kolem předposlední vazby.

### 1.3 Definice torzního úhlu

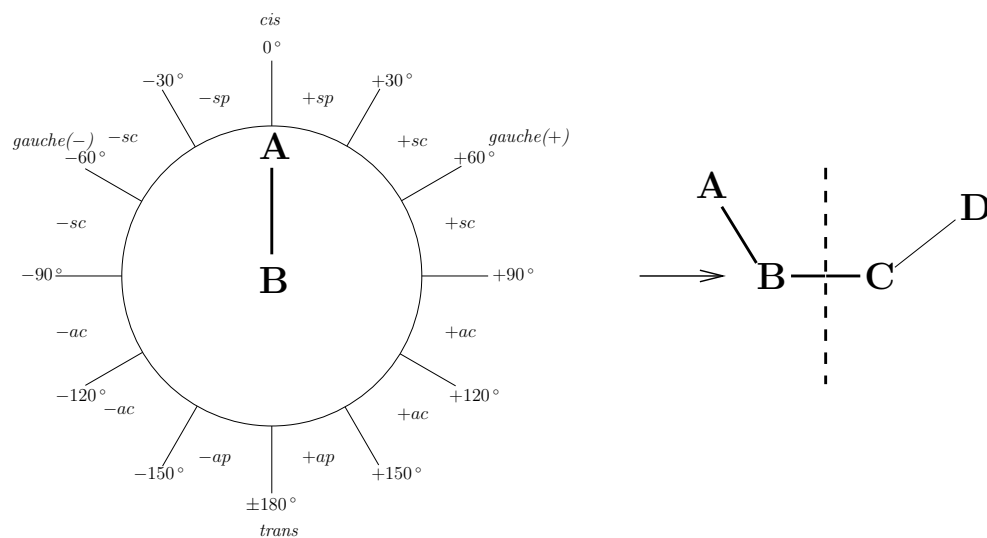
Zatímco definice vazebné délky a vazebného úhlu jsou asi každému zřejmé, definice torzního úhlu si zaslouží pečlivou pozornost. Pro každou kovalentní vazbu, která není koncová, definujeme torzní úhel (popisující rotaci kolem této vazby) následujícím způsobem:

1. Nalezneme řetězec atomů A-B-C-D, který definuje rotaci kolem vazby B-C. Pokud jsou na atom B nebo C (nebo oba) navázány více než dva atomy, určíme atomy A a D podle následujících pravidel:
  - Je-li vazba B-C součástí hlavního řetězce (páteře) makromolekuly, zvolíme atomy A a D tak, aby byly také součástí hlavního řetězce (páteře).
  - Je-li vazba B-C součástí postranního řetězce makromolekuly, zvolíme atomy A a D tak, aby měly nejvyšší prioritu podle pravidel používaných v organické chemii (zavedených Cahnem, Ingoldem a Prelogem).
2. Umístíme molekulu v prostoru tak, aby vazba B-C byla kolmá na rovinu papíru. Podíváme se na molekulu ve směru vazby B-C (atom B je blíže k nám a atom C je v zákrytu za ním) a promítneme vazby A-B a C-D do roviny papíru.
3. Změříme úhel mezi průměty vazeb A-B a C-D. Hodnoty úhlu udáváme v rozsahu od  $-180^\circ$  do  $+180^\circ$ , kladné hodnoty nabývá torzní úhel v případě, že vazba C-D je pootočena oproti vazbě A-B po směru hodinových ručiček (obrázek 1.6).

Kromě udání číselných hodnot lze přibližnou hodnotu torzního úhlu popsat slovně. Používají se k tomu různá názvosloví, která jsou také ukázána na obrázku 1.6. První z nich zavádí označení pro nejvýhodnější konformace. Tak jako v případě dvojnásobné vazby C=C jsou výrazně nejvýhodnější hodnoty torzního úhlu  $0$  a  $+180^\circ$  (izomery *cis* a *trans*), v případě jednoduché vazby C-C mezi dvěma  $sp^3$  uhlíky jsou nejvýhodnější tři torzní úhly, které se označují *gauche(+)* ( $+60^\circ$ ), *gauche(-)* ( $-60^\circ$ ) a *trans* ( $180^\circ$ ).

Druhé názvosloví vychází z rozdělení konformací na dvě skupiny. První skupina, označovaná *syn*, zahrnuje konformace, ve kterých jsou průměty vazeb A-B a C-D vzájemně pootočeny o méně než  $90^\circ$  jedním či druhým směrem (leží tedy ve stejné polovině kruhu). Naopak druhá skupina, označovaná *anti*, zahrnuje konformace, ve kterých jsou průměty vazeb A-B a C-D vzájemně pootočeny o více než  $90^\circ$

<sup>9</sup>Z pouhého geometrického popisu poloh atomů ale ještě nevyplývá, jestli atomy A, B, C, D tvoří lineární řetězec, nebo uzavřený čtyřčlenný kruh. Navíc můžeme geometricky popsat polohu atomů pomocí vzdáleností mezi atomy, které nejsou spojeny chemickou vazbou (a podobně také pomocí úhlů mezi úsečkami spojujícími atomy, které nejsou vázány a pomocí torzních úhlů popisujících rotaci kolem takových úseček. Je-li tomu tak, atomy mohou tvořit rozvětvenou molekulu nebo trojčlenný kruh s jedním bočním atomem. Pokud jsme si řekli, že k stavbě modelu nestačí topologie, ale musíme znát geometrii, musíme dodat, že geometrie sama také nedefinuje jednoznačně strukturu a vyžaduje také znalost topologie – tedy schema propojení atomů vazbami).



Obrázek 1.6: Definice torzních úhlů. Atom B je umístěn ve středu kruhu, atom C je v zákrytu za ním (vazba B–C je kolmá na rovinu kruhu). Průmět vazby A–B do roviny kruhu je vyznačen. Torzní úhel je definován podle toho, do které části kruhu dopadne průmět vazby C–D (žádný konkrétní příklad vyznačen není). Na obrázku jsou uvedeny číselné hodnoty i symbolická označení možných torzních úhlů. V pravé části obrázku je pohled kolmo na rovinu kruhu. Rovina kruhu je vyznačena čárkovaně, šipka ukazuje směr, kterým se díváme, když definujeme torzní úhel.

(leží tedy ve opačné polovině kruhu). Tyto dvě skupiny lze kombinovat o údaj popisující, zda čtveřice atomů leží spíše v jedné rovině (*periplanární* konformace) nebo ve dvou různých rovinách (*klinální* konformace). Podle tohoto názvosloví tedy hovoříme například o kladném synperiplanárním úhlu (0 až  $+30^\circ$ , zkratka  $+sp$ ), nebo o záporném antiklinálním úhlu ( $-150$  až  $-90^\circ$ , zkratka  $-ac$ ).

## Kapitola 2

# Struktura proteinů

### 2.1 Proteiny jako příklad biomakromolekul

Existují dva důvody, proč je vhodné probírat struktury proteinů (česky bílkovin) jako první ze struktur biomakromolekul. Prvním důvodem je velký význam proteinů a bezpočet funkcí, které v buňce hrají. Strukturou proteinů se proto zabývá nejvíce biochemiků. Druhý důvod je pedagogický. Na strukturách proteinů dobře vyniknou obecná pravidla, se kterými se budeme později setkávat i u jiných biomakromolekul.

Jedním z těchto pravidel je *stavební charakter* struktury. Z několika málo součástí si budeme moci sestavit základní strukturní motivy, jejichž další kombinací získáme i ty nejsložitější struktury. Jiným obecným rysem je *hierarchie* struktur. To znamená, že strukturní motivy, které můžeme sestavit ze základních dílů biochemické stavebnice, slouží jako stavební díly složitějších strukturních motivů. A tyto složitější struktury jsou zase stavebními kameny ještě složitějších strukturních útvarů a tak dále. Tato hierarchie je u proteinů nejlépe definovaná – biochemici rozlišují primární, sekundární, terciární strukturu i struktury vyšší.

V této chvíli asi není jasné, co je tak pozoruhodné na zmíněných rysech biomakromolekulárních struktur – stavební a hierarchické uspořádání. Následující stránky by měly ukázat, že tyto základní principy jsou klíčem k pochopení stavby biologicky zajímavých molekul. Složitost makromolekuly není dána složitostí základních stavebních jednotek, ale neuvěřitelným množstvím možností, jak můžeme poměrně jednoduché stavební jednotky kombinovat. Živí tvorové nejsou vyjímeční tím, z jakých atomů se skládají, ale vynikají způsobem, jak jsou uspořádány atomy v jejich molekulách, molekuly v buňkách, buňky v orgánech. O kráse katedrál se také dozvíme více z papírového modelu, který nese informaci o architektuře, aniž by říkal cokoli o materiálním složení, než z příslušných hromad kamene, písku a vápna, které naopak nenesou informaci o myšlence architekta.

Obecná pravidla stavby biomakromolekul bychom mohli dlouho teoreticky probírat. Mohli bychom se snažit uvěznit tato pravidla do co nejpřesnějších definic, což by se nám stejně nepodařilo. Pojdme se raději podívat na konkrétní příklad biomakromolekulární architektury – na molekuly proteinů.

### 2.2 Stavební jednotky

Základními stavebními jednotkami proteinů jsou  $\alpha$ -L-aminokyseliny. Ačkoli je skutečná syntéza proteinů v buňce velmi složitá, formálně můžeme molekulu proteinu získat spojováním karboxylových a aminových skupin *kondenzační* reakcí, při které se odštěpuje molekula vody. Vzniklý produkt se nazývá *peptid*. V peptidech jsou zbytky aminokyselin (aminokyselinová *rezidua*) spojeny amidovou vazbou, která se často označuje jako vazba *peptidová*. Při kondenzaci většího počtu aminokyselin vzniká lineární

molekula, jejíž *páteř* je tvořena atomy z kondenzovaných karboxylových a aminových skupin a zbytek molekuly aminokyseliny tvoří *postranní řetězec*. Označení *peptid* a *protein* jsou poněkud nejednoznačná a do jisté míry synonymní. V literatuře se můžeme setkat s několika způsoby rozlišování mezi peptidem a proteinem:

- Peptidy a proteiny se rozlišují podle počtu aminokyselinových zbytků. Molekuly skládající se ze dvou až deseti aminokyselinových zbytků se označují jako oligopeptidy, molekuly složené z 11 až 100 aminokyselin se označují jako polypeptidy a molekuly obsahující více aminokyselinových zbytků se označují jako proteiny. Přestože je toto dělení oblíbené autory učebnic, je zcela umělé a o struktuře nic nevyovídá.
- Jako peptidy se označují kratší řetězce, které snadno přechází z jedné z mnoha možných konformací do jiné a netvoří stabilní trojrozměrné struktury. Naopak jako proteiny se označují delší řetězce, které mají zřetelnou tendenci vyskytovat se v jedné nejvýhodnější konformaci (nebo v závislosti na vnějších podmínkách měnit konformaci v rámci malého počtu možných konformací) a mají tedy dobře definovanou prostorovou strukturu. Je zřejmé, že mezi takto definovanými peptidy a proteiny není jasně vymezená hranice, ale spíše plynulý přechod. Navíc může být i značně velká molekula proteinu málo uspořádaná, připomínající flexibilní peptid.
- Při studiu složitějších molekul, které kromě aminokyselinového řetězce obsahují chemicky odlišnou část, se jako polypeptid označuje samotný peptidový řetězec a jako protein celá složená molekula. Jiné názvosloví nazývá peptidovou část *apoprotein* a celou složenou molekulu *holoprotein*.

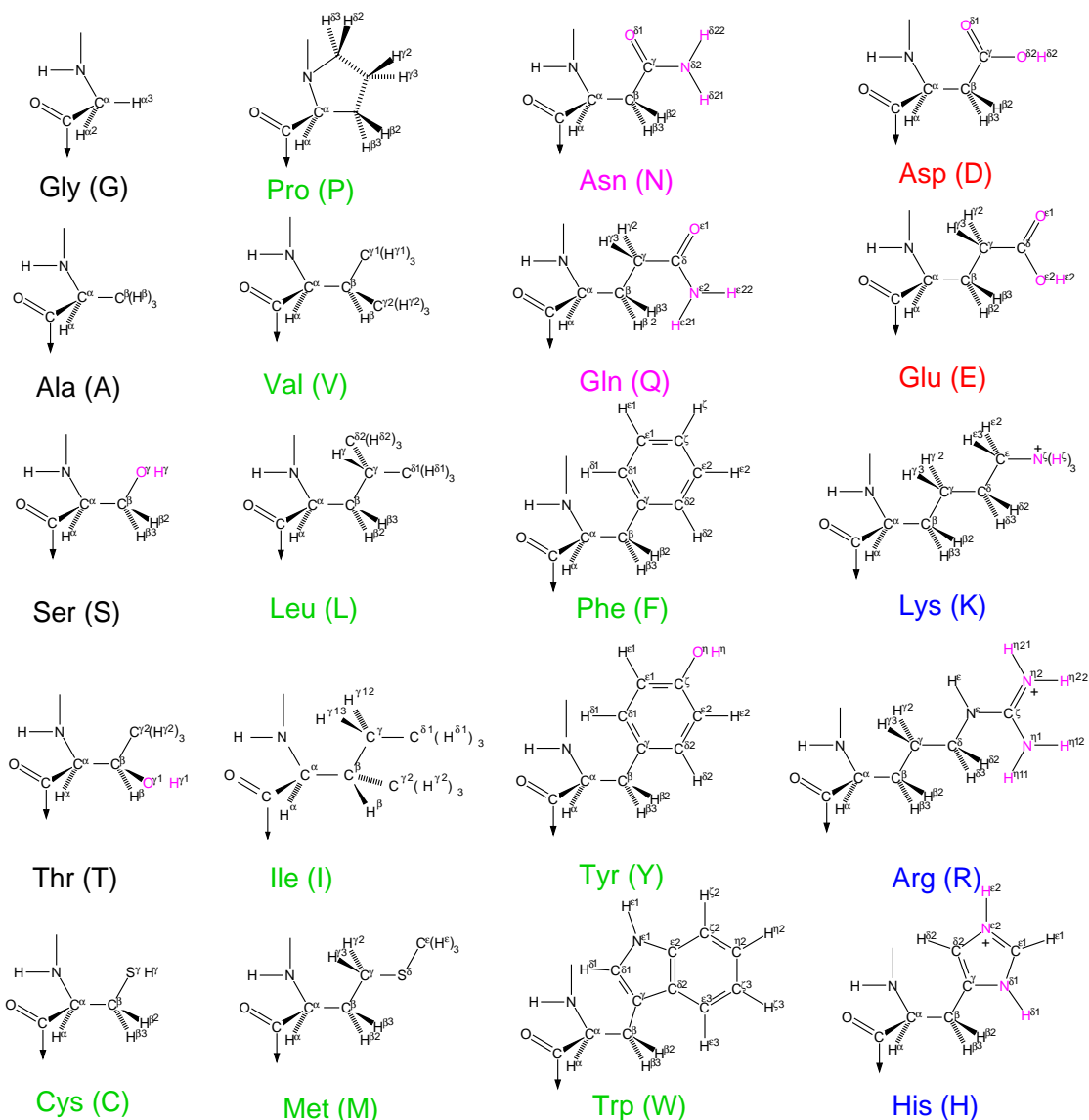
Přestože buňka obsahuje větší počet aminokyselin, proteinový řetězec se syntetizuje pouze z 20 geneticky kódovaných (tzv. *proteinogenních*) aminokyselin<sup>1</sup>. Pro jednoznačný popis struktury je nutné zavést jasně definované označení všech atomů aminokyselin. Bylo by samozřejmě možné použít systematického názvosloví organické chemie. Z částečně historických a především praktických důvodů se ale zavádí odlišné biochemické názvosloví, které se řídí pokyny Mezinárodní unie pro čistou a užitou chemii a Mezinárodní unie pro biochemii IUPAC-IUB. Jistě není nutné se všechny detaily názvosloví učit nazpaměť, je ale nezbytné se jimi ve všech případech řídit. Přehled značení atomů aminokyselin můžete najít na obrázku 2.1.

Při pohledu na vzorce 20 proteinogenních aminokyselin 2.1 si asi každý uvědomí chemickou různost postranních řetězců. Tato pestrost má zásadní význam pro strukturu i funkci proteinů a vysvětluje, proč jeden typ makromolekuly může mít tak odlišné chemické a fyzikální vlastnosti a hrát tak různé role v životě buňky. Z pohledu interakcí, které určují energetickou výhodnost konformace proteinu, si můžeme aminokyseliny rozdělit na několik typů.

- Aminokyseliny, které jsou schopny vytvářet iontovou vazbu, zahrnují arginin, lysin, histidin, tyrosin, cystein, kyselinu asparagovou a kyselinu glutamovou. Podmínky, za kterých jsou tyto aminokyseliny ionizovány, jsou dány především kyselostí (hodnotou pH) a elektrickou permitivitou jejich nejbližšího okolí. Arginin a lysin tvoří kationty v kyselém a neutrálním prostředí, histidin tvoří kationty v kyselém prostředí. Kyselina asparagová a glutamová tvoří anionty v neutrálním a zásaditém prostředí, cystein v zásaditém a tyrosin jen v silně zásaditém prostředí. Konkrétní hodnoty *disociační konstanty*<sup>2</sup> závisí na elektrické permitivitě okolí. V hydrofobním jádře proteinu budou postranní řetězce hůře ionizovat, takže kationty se budou tvořit až při nižším pH a anionty až při vyšším pH.

<sup>1</sup>V molekulách proteinů izolovaných z buněk nacházíme více než 20 různých aminokyselin. „Nestandardní“ aminokyselinové zbytky ale v naprosté většině případů vznikají chemickými úpravami již vzniklého polypeptidového řetězce. Tyto úpravy se nazývají *posttranslační modifikace* a zahrnují zavedení jednoduchých funkčních skupin jako hydroxyl nebo fosfát i navázání velkých molekul jako mastné kyseliny, porfyriny (hem, chlorofyl), oligosacharidy a další.

<sup>2</sup>Dekadický logaritmus disociační konstanty s opačným znaménkem je roven pH, při kterém je právě polovina molekul disociována.



Obrázek 2.1: Vzorce aminokyselin s atomy označením atomů podle doporučení IUPAC-IUB. Purpurově jsou zvýrazněny atomy, které se podílejí na tvorbě vodíkových vazeb. Barvy trojpísmenných a jednopísmenných zkratk aminokyselin rozlišují krátké postranní řetězce (černá), nepolární postranní řetězce (zelená), polární postranní řetězce (purpurová), kyselé (záporně nabitě) postranní řetězce (červená) a zásadité (kladně nabitě) postranní řetězce (modrá).

- Postranní řetězce schopné dipolárních interakcí jsou zároveň schopné tvořit vodíkové vazby<sup>3</sup>. Patří mezi ně všechny aminokyseliny schopné tvořit ionty a navíc serin, threonin, asparagin, glutamin a tryptofan.
- Silný hydrofobní efekt vykazují alifatické postranní řetězce prolinu, leucinu, isoleucinu, valinu a do jisté míry alaninu. Míra hydrofobních a lipofilních vlastností je dána délkou alifatického řetězce. Dále můžeme mezi hydrofobní postranní řetězce zařadit aromatické aminokyselinové zbytky tryptofanu, tyrosinu a fenylalaninu a sirté zbytky cysteinu a methioninu<sup>4</sup>.
- Glycin a do jisté míry i alanin, serin a threonin můžeme považovat za zvláštní skupinu „malých“ postranních řetězců. Vzhledem k jejich velikosti je totiž obtížné říci, jestli vykazují spíše hydrofilní či hydrofobní charakter. Tím, že zabírají malý prostor poskytují zvýšenou volnost okolním postranním řetězcům.
- Prolin má netypickou konfiguraci (cyklický postranní řetězec), která významně omezuje možné konformace proteinové páteře.
- Cystein je vyjímečný v tom, že může tvořit kovalentní disulfidovou vazbu s postranním řetězcem cysteinu v jiné části proteinu. To je velký zásah do topologie proteinové páteře, který má zásadní vliv na konformaci proteinu.

Stejně jako je důležitá různost postranních řetězců, tak je významná neměnnost uspořádání na prvních dvou uhlících. Všechny aminokyseliny můžeme považovat za deriváty alaninu, všechny zachovávají stejnou konfiguraci  $\alpha$ -uhlíku (přestože zrcadlová konfigurace D by byla z chemického pohledu stejně dobře možná<sup>5</sup>). Tato jednotnost základních stavebních dílů umožňuje stavebnicové uspořádání popsané v sekci 2.1. Stejná konfigurace  $\alpha$  uhlíku otvírá dveře téměř neomezenému kombinování aminokyselin při zachování pravidelnosti struktur<sup>6</sup>. Glycin, který je achirální, a prolin, který se liší konfigurací na dusíku, jsou výjimky, které potvrzují pravidlo a dávají ještě více vyniknout logice stavebnicového uspořádání.

## 2.3 Torzní úhly

Jak jsme diskutovali v sekci 1.2, konformaci biomakromolekuly jednoznačně určují torzní úhly. Podívejme se proto na torzní úhly, se kterými se setkáváme v proteinech (obrázek 2.2).

Konformace peptidové páteře aminokyselin je určena třemi torzními úhly,  $\phi$ ,  $\psi$  a  $\omega$ . Čtveřice atomů, které torzní úhel definují (A–B–C–D na stránce 7) jsou pro uvedené úhly následující. V reziduu číslo  $i$  je úhel  $\phi$  je definován atomy  $C'_{i-1}$ ,  $N_i$ ,  $C_i^\alpha$ ,  $C'_i$ , úhel  $\psi$  atomy  $N_i$ ,  $C_i^\alpha$ ,  $C'_i$ ,  $N_{i+1}$  a úhel  $\omega$  atomy  $C_i^\alpha$ ,  $C'_i$ ,  $N_{i+1}$ ,  $C_{i+1}^\alpha$ . Protože peptidová vazba se chová částečně jako dvojná vazba<sup>7</sup>, úhel  $\omega$  nabývá pro většinu aminokyselin hodnot blízkých  $180^\circ$  (konformace *trans*). Výjimkou je prolin, který se vyskytuje i v *cis* konformaci ( $0^\circ$ ). Konformace peptidové páteře je tak definována především hodnotami úhlů  $\phi$  a  $\psi$ . Peptidové konformery je proto možno graficky znázornit jako body v grafu, jehož osy tvoří hodnoty úhlů

<sup>3</sup>Vodíkové vazby je schopna tvořit každá peptidová vazba, zde je řeč pouze o postranních řetězcích.

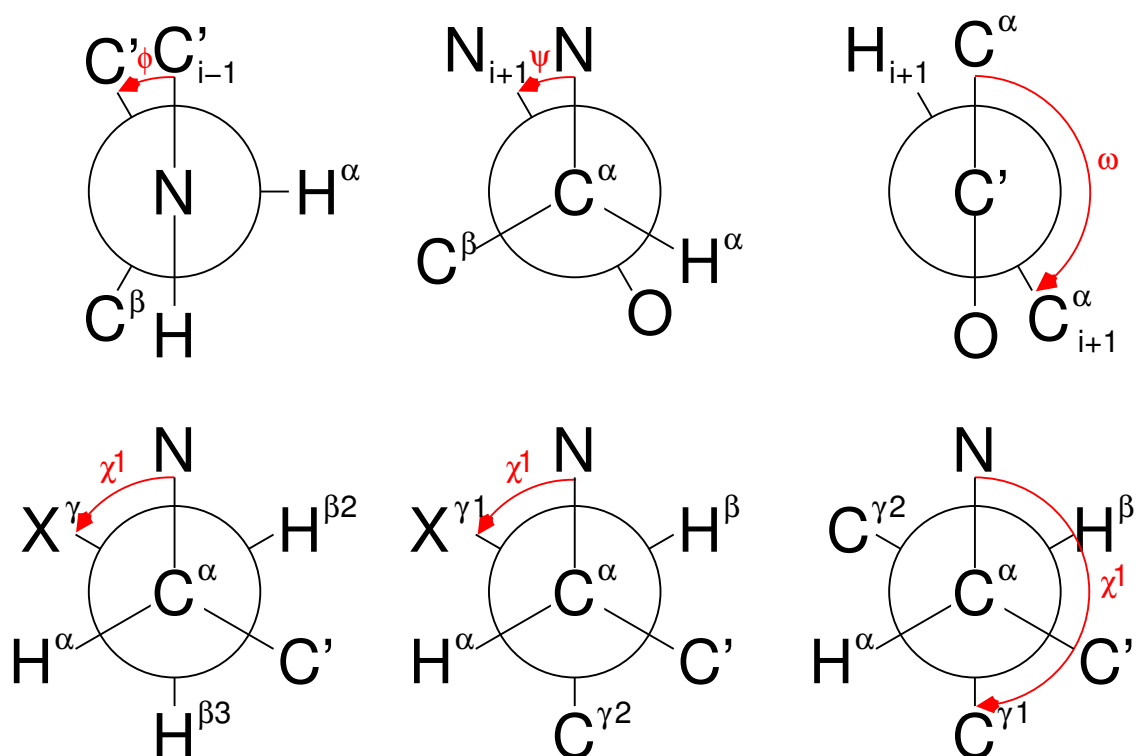
<sup>4</sup>Aromatické aminokyseliny, zařazené do skupiny hydrofobních postranních řetězců, mohou krom van der Waalsovských sil interagovat prostřednictvím elektronů v delokalizovaných  $\pi$ -orbitalech. Totéž platí o d-orbitalech síry. Jde o interakce svým popisem připomínající kovalentní vazby, ale mnohonásobně slabší. V našem zjednodušeném výčtu sil působících v biomakromolekulách jsme si je neuváděli.

<sup>5</sup>Na otázku proč se v proteinech vyskytují právě L-aminokyseliny dosud neznáme uspokojivou odpověď.

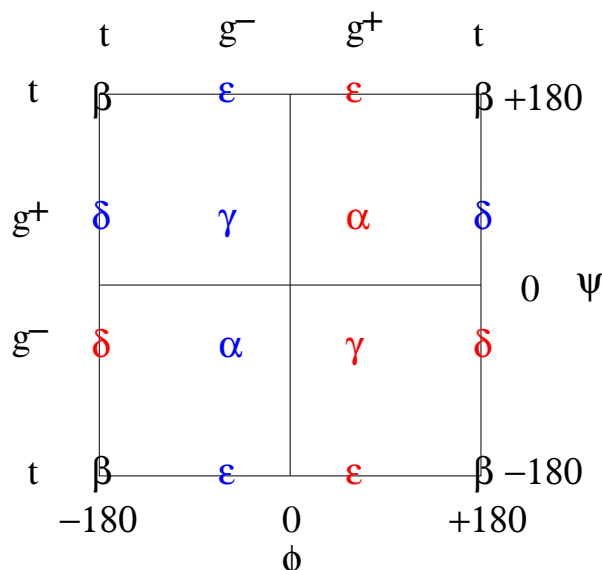
<sup>6</sup>Pokud stejnou strukturu můžeme postavit z většího počtu možných kombinací základních stavebních prvků, je více pravděpodobné, že se s takovou strukturou v buňce setkáme

<sup>7</sup>Vyjádřeno pojmy molekulových orbitalů, o kterých bude řeč v kapitole 10, pokud leží vodík a oba uhlíky vázané na dusík v jedné rovině (hybridizace  $sp^2$ ), mají orbitály  $p$  dusíku, kyslíku i uhlíku v karbonylu stejnou orientaci, ideální pro překryv a tvorbu delokalizované  $\pi$ -vazby.





Obrázek 2.2: Definice torzních úhlů hlavního řetězce (nahore) a torzního úhlu  $\chi^1$  postranních řetězců (dole) proteinů. Vlevo dole je definován  $\chi^1$  nevětvených řetězců, uprostřed threoninu a isoleucinu, vpravo valinu.



Obrázek 2.3: Ramachandranův diagram s vyznačeným názvoslovím peptidových konformací. Modře jsou označeny konformace výhodnější pro L aminokyseliny. Červeně jsou označeny konformace výhodnější pro D aminokyseliny, které se běžně v geneticky kódovaných proteinech nevyskytují.

$\phi$  a  $\psi$ . Takový graf se nazývá Ramachandranův diagram. Běžné konformace peptidové páteře budou diskutovány později. Na obrázku 2.3 je uvedena obecná konvence názvosloví peptidových konformerů. Pro chirální aminokyseliny (všechny běžné kromě glycinu) existují většinou dvojice (zrcadlové obrazy) konformerů, z nichž jeden (označený modře) je výhodný pro L aminokyseliny (ty se běžně vyskytují v proteinech), druhý (označený červeně) by byl výhodnější pro D aminokyseliny:

Konformace postranních řetězců je dána torzními úhly, které se značí písmenem  $\chi$ , a jejichž počet je pochopitelně různý pro různé aminokyseliny. Úhel  $\chi^1$ , definovaný atomy N,  $C^\alpha$ ,  $C^\beta$ , X se nachází nejčastěji v oblasti *gauche(-)* nebo *trans*. Konformace *gauche(+)* je nevýhodná stericky kvůli přiblížení peptidové vazbě, častěji se vyskytuje u serinu a threoninu, kde ze stejných důvodů umožňuje tvorbu vodíkové vazby mezi kyslíkem postranního řetězce a peptidovou vazbou. Valin a isoleucin, které mají dva  $\gamma$  uhlíky, se vyskytují převážně v konformaci, ve které není žádný  $\gamma$  uhlík orientován *gauche(+)* vzhledem k hlavnímu řetězci (pozor na zálužnost doporučeného názvosloví organické chemie: u isoleucinu se taková konformace označuje *gauche(-)*, ale u valinu *trans*).

## 2.4 Primární struktura

Poté, co jsme prošli základní definice, se již můžeme podívat na vlastní struktury. Jak jsme se již zmínili v sekci 2.1, zavedli biochemici hierarchickou posloupnost struktur. Nejnížší úrovní této hierarchie je *primární struktura*.

Primární struktura je pořadí (*sekvence*) aminokyselin v polypeptidu. Protože konfiguraci jednotlivých aminokyselin známe, poskytne nám primární struktura veškerou informaci o *konfiguraci* proteinu (s jednou důležitou výjimkou – z primární struktury stále nevíme, které cysteiny jsou spojeny disulfidovou vazbou). Naopak se z primární struktury nedovíme nic o konformaci proteinu.

Podle konvence začíná sekvence N-terminální aminokyselinou (obsahuje volnou aminoskupinu) a končí C-terminální aminokyselinou (obsahuje volnou karboxyskupinu). Tímto směrem také probíhá syntéza proteinů v buňce. Pořadí aminokyselin je dáno genetickým kódem, který je zapsán v molekule

deoxyribonukleové kyseliny (DNA). Každým třem monomerním jednotkám DNA (nukleotidům), které tvoří takzvaný *kodón*, odpovídá jedna monomerní jednotka proteinů (aminokyselina). Vlastní překlad z genetického kódu do „jazyka“ aminokyselin je složitý proces, ve kterém hrají nejdůležitější roli různé molekuly ribonukleových kyselin (RNA). Kódovací tabulka je společná všem živým tvorům<sup>8</sup>. Více si o biosyntéze proteinů řekneme v kapitole 8.

V laboratorní praxi je poměrně snadné určit sekvenci DNA a s využitím znalosti kódovací tabulky se tak ihned dozvědět sekvenci proteinu. Proto je primární sekvence proteinu zpravidla známá dříve, než je řešena jeho trojrozměrná struktura<sup>9</sup>. Ze sekvence je v principu možno předpovědět prostorové uspořádání proteinů, v současnosti je však spolehlivost takových předpovědí omezená.

## 2.5 Sekundární struktura

Zatímco definice primární struktury byla celkem jednoznačná, u vyšších struktur se již definice různých autorů často liší. My budeme mít pod pojmem sekundární struktura na mysli *pravidelný, libovolně opakovatelný konformační motiv polypeptidové páteře*<sup>10</sup>. U proteinů se setkáváme se dvěma základními typy sekundárních struktur.

## 2.6 Struktura alfa

Struktura  $\alpha$  je pravotočivá šroubovice neboli *helix* (obrázek 2.4, pokud čtete skripta v elektronické formě a jste připojeni na internet, můžete si model  $\alpha$  šroubovice prohlédnout na svém počítači). Všimněme si, že poskládáním aminokyselin do vyšší struktury vznikl nový tvar, který je sám o sobě chirální. Tato chiralita nijak nesouvisí s chiralitou aminokyselin. Kdybychom postavili  $\alpha$ -šroubovici ze samých glycinů, byla by stále chirální, tak jako je chirální každý šroub. Stejně bychom mohli postavit z glycinů levotočivou šroubovici, která by byla zrcadlovým obrazem  $\alpha$  šroubovice. Ale pozor, pokud sestavíme levotočivou šroubovici z jiných aminokyselin, zrcadlový obraz  $\alpha$  šroubovice nezískáme. Vzájemnými zrcadlovými obrazy jsou totiž pravotočivé šroubovice postavené z L aminokyselin ( $\alpha$ -šroubovice) a levotočivé šroubovice postavené z D aminokyselin.

Rozdíl mezi levotočivou a pravotočivou šroubovicí je v konformaci stavebních dílů, nikoli v jejich konfiguraci. Na úrovni sekundárních struktur můžeme ale šroubovice považovat za útvary s rozdílnou konfigurací<sup>11</sup>. Toto je typický příklad hierarchie struktur. Na úrovni sekundární struktury má smysl mluvit o šroubovici jako o základním stavebním prvku, jako bychom zapomněli, že jde o útvar daný konformací aminokyselin, ze kterých je postaven. To si můžeme dovolit, protože konformace aminokyselin jsou natolik uvězněny v určité poloze<sup>12</sup>, že se šroubovicí můžeme zacházet jako se stabilním a rigidním stavebním dílem. Vidíme, že jsme se na hony vzdálili od jasného rozlišení konfigurace a konformace podle rotace kolem vazby, které bylo tak pěkně použitelné na úrovni jednotlivých aminokyselin.

Úhly  $\phi$  i  $\psi$  odpovídají v  $\alpha$ -strukturách konformaci *gauche(-)* (ideální  $\alpha$  šroubovice je definována hodnotami  $\phi = -57^\circ$  a  $\psi = -47^\circ$ ).  $\alpha$ -Šroubovice je stabilizována vodíkovými vazbami, které jsou orien-

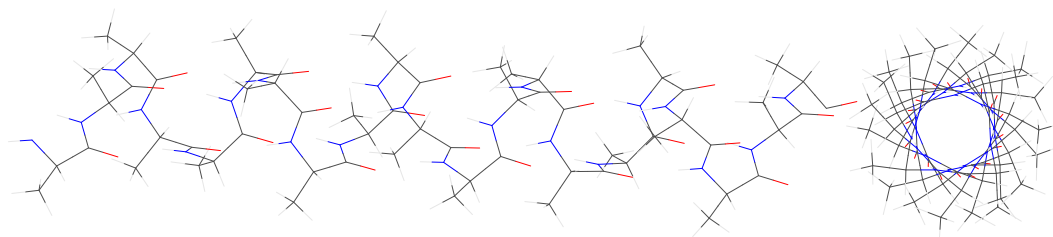
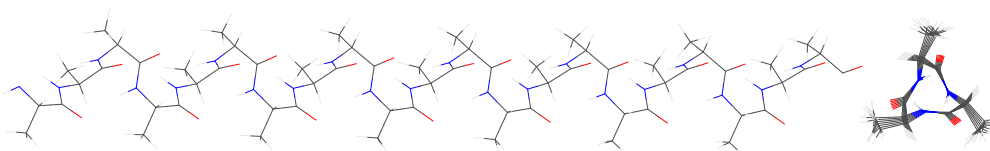
<sup>8</sup>Tato univerzalita genetického kódu má své výjimky. Poněkud jednodušší rozpoznávání nukleotidů například v mitochondriích vede k drobným odchylkám kódovacích pravidel. Jiným příkladem výjimky je využití některých kodónů pro zabudování selenu do molekuly cysteinu (místo síry).

<sup>9</sup>V současné době je již pro řadu organismů (včetně člověka) známa téměř veškerá genetická informace. U těchto organismů známe tedy sekvenci všech nukleotidů genové DNA. Stačí nám tedy určit, kde geny jednotlivých proteinů začínají a končí (a také, kde jsou přerušeny nekódujícími úseky) k tomu, abychom znali sekvenci všech proteinů, alespoň v okamžiku, kdy jsou v buňce syntetizovány (v mnoha případech se část proteinu odštěpí dříve, než začne protein plnit svoji úlohu).

<sup>10</sup>Sekundární struktury lze často (alespoň formálně matematicky) popsat jako šroubovice.

<sup>11</sup>Pravidla IUPAC-IUB definují nejen konfiguraci čtyřřetěnu, používanou na úrovni atomů, ale i konfiguraci šroubových útvarů. Pravotočivé útvary se označují symbolem *P*, levotočivé symbolem *M*.

<sup>12</sup>Stále samozřejmě mluvíme o konformaci peptidové páteře, nikoli postranních řetězců.

Obrázek 2.4: Model  $\alpha$ -šroubovice.Obrázek 2.5: Model šroubovice  $3_{10}$ .

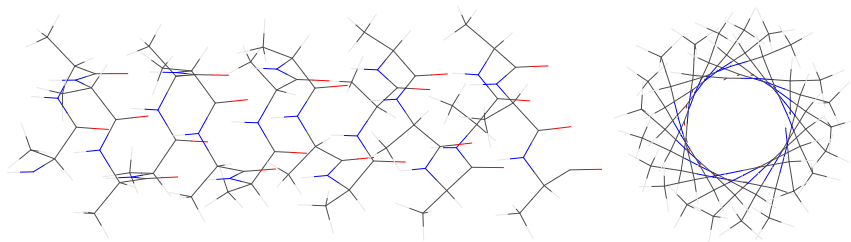
továny téměř rovnoběžně s osou helixu<sup>13</sup> a tedy i s jeho výsledným dipólovým momentem (nejvýhodnější orientace). Protože vodíkové vazby se tvoří mezi amidovými vodíky a karbonylovými kyslíky peptidových vazeb, jsou aminokyseliny v sousedních závitěch šroubovice navzájem posunuty zhruba o jednu třetinu. Na jeden závit tak připadá 3,6 (přibližně  $3\frac{2}{3}$ ) aminokyselin.

V proteinové chemii se používá speciální názvosloví šroubovic. Každá šroubovice je popsána dvěma čísly. První číslo udává počet aminokyselin na jeden závit, u  $\alpha$ -šroubovice tedy 3,6. Druhé číslo je definované složitěji. Je založeno na tom, že sousední vodíkové vazby uzavírají ve šroubovicích pravidelné smyčky atomů spojených kovalentními a vodíkovými vazbami. V  $\alpha$ -helixech je každá taková smyčka tvořena 13 atomy. Podle názvosloví odvozeného od počtu reziduí na závit a od počtu atomů ve smyčkách by se tedy  $\alpha$ -šroubovice označovala jako helix 3.6<sub>13</sub>.

Struktura  $\alpha$ -šroubovice je energeticky nejvýhodnější stočená forma peptidového řetězce. Těsnější vinutí šroubovice (o  $\frac{2}{3}$  rezidua) tvoří strukturu helixu  $3_{10}$  (3 aminokyseliny na závit, 10 atomů ve smyčkách, viz obrázek 2.6 a model). Tato šroubovice je méně výhodná z pohledu dipólových interakcí (vodíkové vazby jsou orientovány šikmo k ose helixu) a sterického bránění (postranní řetězce jsou v zákrytu). S touto strukturou se setkáváme zejména na C-koncích  $\alpha$ -helixů, kde již karbonyly nemají k dispozici amidové vodíky v dalším závitě, se kterými by mohly tvořit stabilizující vodíkové vazby. Je tedy výhodnější šroubovici „utáhnout“ do helixu  $3_{10}$ , který umožňuje vytvořit více vodíkových vazeb při stejném počtu aminokyselin.

Pokud je šroubovice vinuta naopak o  $\frac{2}{3}$  rezidua volněji než  $\alpha$ -helix, tvoří tzv.  $\pi$ -šroubovici (helix 4.4<sub>16</sub>, viz obrázek 2.6 a model). Vodíkové vazby jsou v  $\pi$ -helixu sice orientovány rovnoběžně s osou, energeticky je ale  $\pi$ -helix nevýhodný a ve strukturách proteinů velmi vzácný. Důvodem je tvorba dutiny

<sup>13</sup>Ačkoli je latinské slovo *helix* rodu ženského, v tomto textu používáme mezi biochemiky vžitě skloňování podle mužského vzoru *hrad*.

Obrázek 2.6: Model  $\pi$ -šroubovice.

uvnitř  $\pi$ -šroubovice (tím přicházíme o výhodné van der Waalsovy interakce<sup>14</sup>) a sterické pnutí vedoucí k rozevírání úhlu  $N-C^\alpha-C'$ .

Všechny uvedené šroubovice se nacházejí na Ramachandranově diagramu v oblasti  $\alpha$  konformací (*gauche(-), gauche(-)*). Na obrázku 2.6 je tato oblast  $\alpha$  povolena pro L-aminokyseliny označena černě. Vyjíměnou aminokyselinou je prolin<sup>15</sup>, který se uprostřed pravotočivé  $\alpha$ -šroubovice nevyskytuje ze sterických důvodů (kolize s karbonylem předchozí peptidové vazby).

Zrcadlový obraz pravotočivé  $\alpha$ -šroubovice odpovídá levotočivé šroubovici s konformací *gauche(+), gauche(+)* (červená oblast). V této konformaci se často nalézají achirální glyciny, občas v ní nalezneme asparagin, pro který je nevýhodnost této konformace kompenzována možností tvorby vodíkové vazby mezi páteří a postranním řetězcem.

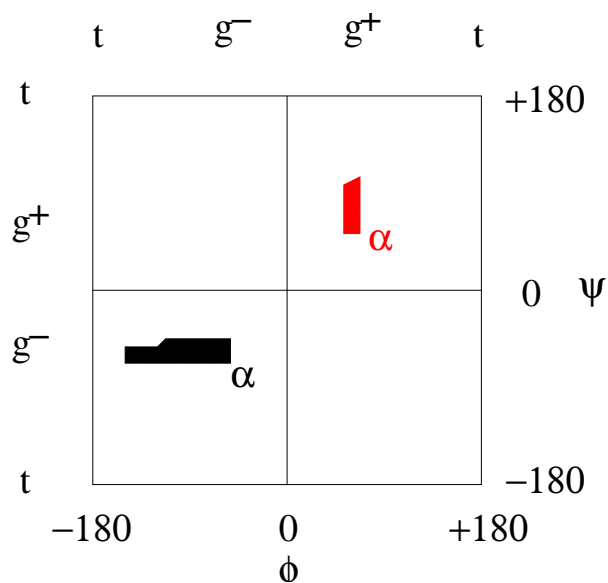
## 2.7 Struktura beta

Druhou základní sekundární strukturou proteinů je struktura  $\beta$ . Na rozdíl od  $\alpha$ -šroubovice odpovídá  $\beta$ -struktura maximálně rozvinuté konformaci peptidové páteře. Stejně jako  $\alpha$ -šroubovice je  $\beta$ -struktura stabilizována vodíkovými vazbami. Na rozdíl od  $\alpha$ -šroubovice, kdy vodíkové můstky spojovaly atomy aminokyselin téhož helixu,  $\beta$ -struktury jsou propojeny se sousedními  $\beta$ -vlákní sítí vodíkových vazeb přibližně kolmých k ose páteře. Spojena mohou být vlákna orientovaná opačným směrem (*antiparalelní*) nebo stejným směrem (*paralelní*). Vznikají tak přibližně ploché struktury listů, *sheets*, propojených sítí vodíkových vazeb (paralelní uspořádání vyžaduje větší odchylku vodíkových vazeb od kolmého směru). Základní strukturu paralelního a antiparalelního skládaného listu ukazuje obrázek 2.8 a modely antiparalelních a paralelních vláken. Reálnou ukázkou skládaných listů si můžete prohlédnout v modelu thioredoxinu.

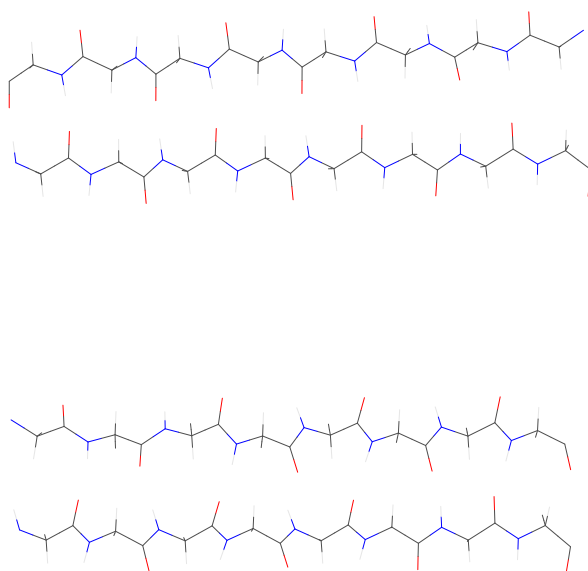
Zcela natažená *trans,trans* konformace peptidové páteře by odpovídala struktuře skládaného listu, ve které by následující aminokyseliny byly navzájem otočeny o  $180^\circ$  (formálně by se dala tato struktura považovat za šroubovici, ve které na jeden závit připadnou dvě rezidua). V proteinech se zcela natažená konformace nevyskytuje, hodnoty úhlů  $\phi$  a  $\psi$  jsou menší než  $180^\circ$  ( $\phi = -139^\circ$ ,  $\psi = +135^\circ$  v antiparalelních a  $\phi = -119^\circ$ ,  $\psi = +113^\circ$  v paralelních  $\beta$ -lístech). Peptidová páteř v  $\beta$ -strukturách je tedy mírně pravotočivá (i  $\beta$ -struktury jsou vlastně pravotočivé šroubovice) a v důsledku toho jsou  $\beta$ -listy

<sup>14</sup>Interakce, které jsou důležité pro stabilitu struktur, jsou více diskutovány v kapitole 6.

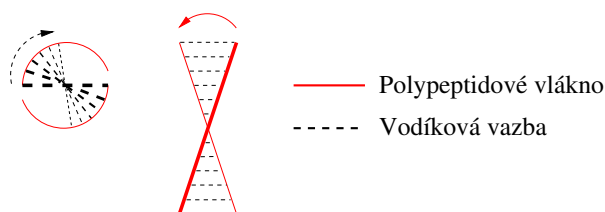
<sup>15</sup>Prolin, jehož konformace jsou omezeny cyklickým postranním řetězcem, a glycin, který je naopak velmi flexibilní, protože postranní řetězec postrádá, jsou konformačně výjimečné aminokyseliny. Polypeptidové řetězce složené z těchto kyselin tvoří pravidelné šroubovice, které ale podle naší definice neoznačujeme jako sekundární struktury (pojem sekundární struktura jsme vyhradili pravidelnému uspořádání, které nezávisí na postranním řetězci). Z těchto struktur je asi nejznámější šroubovice tvořená opakující se sekvencí Gly-Pro-Pro, která se dále splétá do trojšroubovice stabilizované vodíkovými vazbami. Tuto trojšroubovici, která připomíná lano a je mimořádně pevná v tahu, vytváří strukturální protein kolagen.



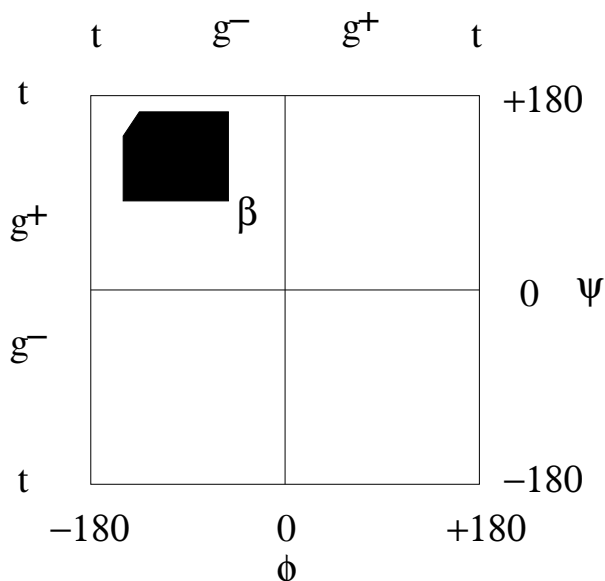
Obrázek 2.7: Ramachandranův diagram s černě vyznačenou oblastí konformací aminokyselin v šroubovicích proteinů (strukturách  $\alpha$ ). Červeně je vyznačena oblast levotočivých šroubovic, které by byly výhodnější pro D-aminokyseliny.



Obrázek 2.8: Model dvou vláken antiparalelního (nahore) a paralelního (dole)  $\beta$ -listu.



Obrázek 2.9: Schematické znázornění pravotočivé orientace vodíkových vazeb (vyznačeny černě a čárkovaně) a levotočivého zkrutu  $\beta$ -listů (peptidová páteř je vyznačena červeně). Vlevo pohled shora, vpravo pohled z boku.



Obrázek 2.10: Ramachandranův diagram s vyznačenou oblastí konformací aminokyselin v  $\beta$ -listech proteinů.

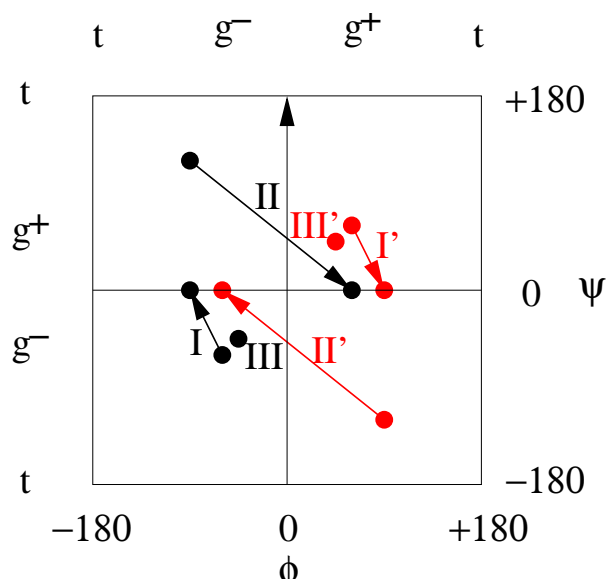
levotočivě zkrouceny. Skutečnost, že pravotočivost v rámci jednotlivého peptidového řetězce způsobuje levotočivost vzájemného uspořádání různých řetězců, jistě působí na první dojem záhadně. Vše je ale zřejmější, když si dobře prohlédneme schema na obrázku 2.9.

Oblast povolených  $\beta$ -konformací v Ramachandranově diagramu je poměrně rozsáhlá (obrázek 2.10). Stejně jako  $\alpha$ -helixy nejsou  $\beta$ -struktury výhodné pro prolin, kde cyklická struktura postranního řetězce vyžaduje úhel  $\phi \approx -60^\circ$ . Prolin (ale i jiné aminokyseliny) se naopak často vyskytuje v konformaci *polyprolin II*, s torzními úhly  $\phi \approx -80^\circ$  a  $\psi \approx 150^\circ$ , tedy poblíž oblasti  $\beta$  v Ramachandranově diagramu. Zastoupení konformace *polyprolin II* v takzvaných přirozeně neuspořádaných proteinech je zmíněno níže.

## 2.8 Terciární struktura

Molekuly některých proteinů jsou tvořeny jedinou monotónně se opakující sekundární strukturou (například keratiny<sup>16</sup>).

<sup>16</sup>Sekundární struktury se ve vláknitých (*fibrilárních*) proteinech nevyskytují izolovaně, ale splétají se do vyšších struktur, srovnatelných s kvartérními strukturami *globulárních* proteinů (viz sekci 2.16). Vytvoření takových vyšších struktur je možné pouze tehdy, když do sebe sekundární struktury přesně zapadají. Takové přesné uspořádání většinou vyžaduje určitou sekvenci aminokyselin, která se ve fibrilárních proteinech mnohokrát opakuje. Vidíme, že zde struktura není



Obrázek 2.11: Ramachandranův diagram s vyznačenými konformacemi reverzních ohybů. Šipka naznačuje směr od první ke druhé aminokyselině ohybu, v ohybu III a III' mají obě aminokyseliny stejnou konformaci.

Mnohé proteiny nebo jejich části nejsou přítomny v jedné dobře definované konformaci, ale vyskytují se ve směsi velkého množství rychle se přeměňujících konformací. Takové proteiny se obvykle označují jako *přirozeně neuspořádané* (*intrinsically disordered*). Krátkodobě v nich mohou vznikat segmenty sekundárních struktur nebo i terciární struktury, důležité pro jejich funkci. Přechodnými sekundárními strukturami mohou být  $\alpha$ -šroubovice, nebo více natažené struktury *polyprolin II*, tvořené i jinými aminokyselinami, než prolin.

Fibrilárními ani přirozeně neuspořádanými proteiny se více zabývat nebudeme. Zaměříme pozornost na proteiny, ve kterých se prvky sekundárních struktur stavebnicově skládají do přibližně globulárních molekul. Takové skládání polypeptidových řetězců se nazývá terciární strukturou. Aby byla molekulová stavebnice proteinů použitelná, musí krom pravidelných sekundárních struktur ( $\alpha$  a  $\beta$ ) obsahovat i spojovací díly.

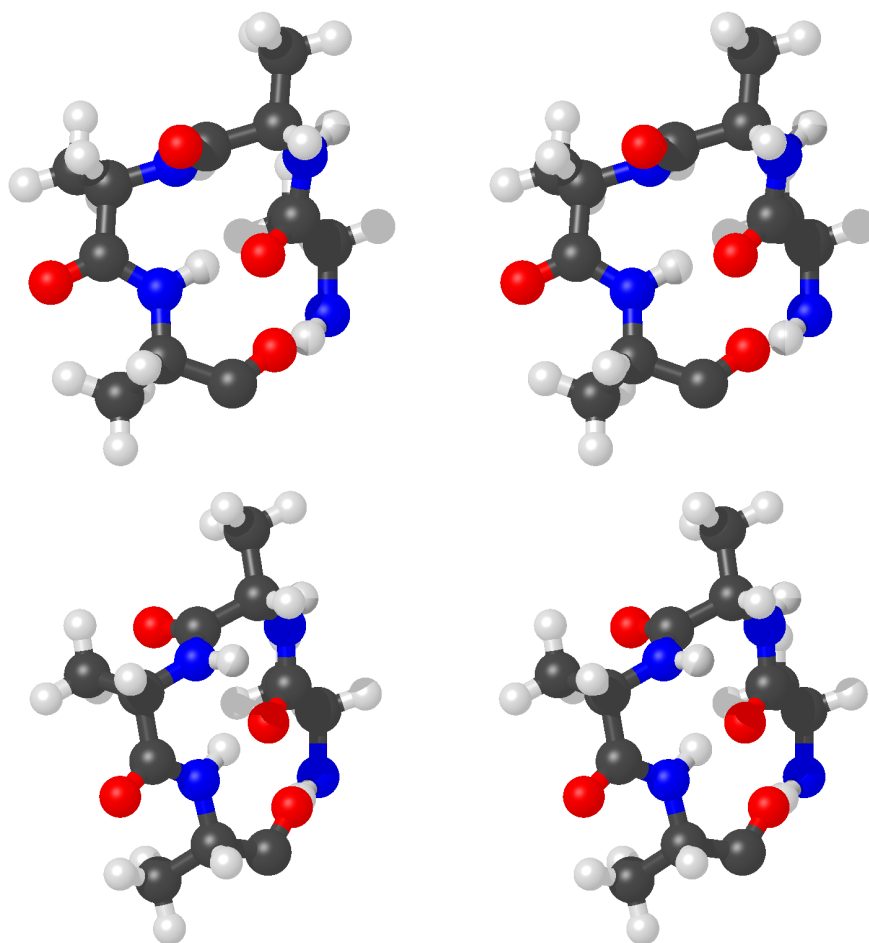
## 2.9 Spojovací prvky

Spojovacích prvků se ve strukturách proteinů vyskytuje celá řada. Nejjednodušší jsou takzvané *ohyby* (*turns*). Jediná aminokyselina v  $\gamma$  konformaci tvoří *inverzní  $\gamma$ -ohyb* („obyčejný“  $\gamma$ -ohyb je pro L-aminokyseliny stericky nevýhodný).

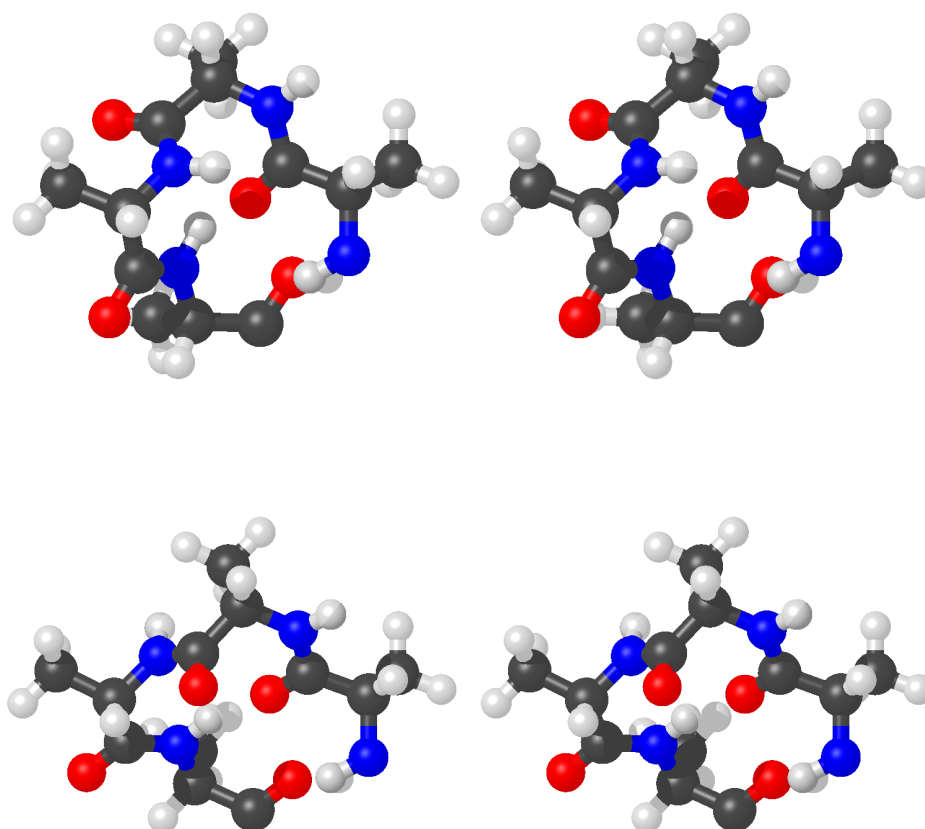
Takzvané *reverzní ohyby* (základní z nich jsou značené I a II) se skládají ze dvou aminokyselin a jsou stabilizované vodíkovým můstkem mezi vodíkem peptidové vazby s následující aminokyselinou a karbonylovým kyslíkem peptidové vazby s předcházející aminokyselinou. Ohyb III je vlastně výsek z výše popisované  $3_{10}$ -šroubovice. Později byla identifikována a vyššími římskými čísly označena řada dalších ohybů, kterými se ale zde zabývat nebudeme. Místo toho zmíníme ohyby značené I', II' a III'. Až na konfiguraci na uhlíku  $C^\alpha$  jde o zrcadlové obrazy reverzních ohybů I, II, III. Ramachandranovy diagramy zmíněných jednoduchých ohybů jsou uvedeny na obrázku 2.11. Struktury nejběžnějších reverzních ohybů si můžete prohlédnout na obrázcích 2.12 (ohyby I a II) a 2.13 (ohyby I' a II').

nezávislá na sekvenci, příčinu ale nenajdeme na úrovni struktur sekundárních, ale vyšších.

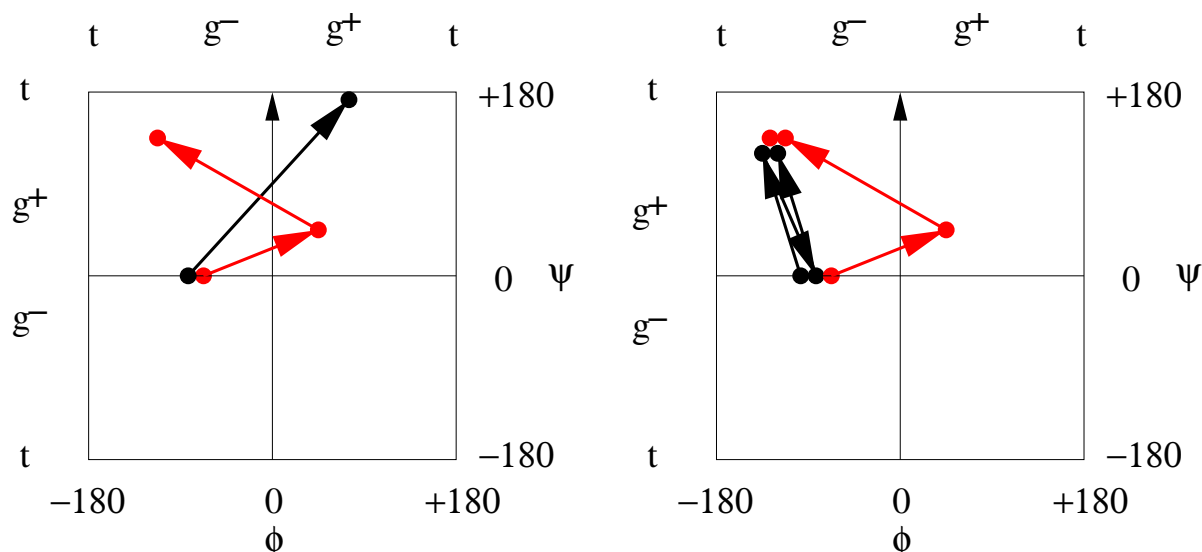




Obrázek 2.12: Stereo obrázky reverzních ohybů I (nahore) a II (dole).



Obrázek 2.13: Stereo obrázky reverzních ohybů I' (nahore) a II' (dole).



Obrázek 2.14: Ramachandranův diagram s vyznačenými konformacemi  $\alpha$ -ohybů tvořených dvěma (černě v levém diagramu), třemi (červeně v levém diagramu), nebo čtyřmi (dvě možnosti jsou uvedeny různými barvami v pravém diagramu) aminokyselinami. Šipky ukazují od první ke druhé a od druhé ke třetí aminokyselině.

## 2.10 Hierarchie v rámci terciární struktury

Základní hierarchie struktur (primární, sekundární, terciární) je pro detailní popis struktur proteinů příliš hrubá. Proto se v posledních letech objevily v literatuře jemnější stupně v rámci terciární struktury. Názvosloví takového jemnějšího dělení není ale dosud ustálené a často vychází z terminologie různých počítačových programů, které se snaží struktury proteinů analyzovat. Zkusme si tedy popsat vzrůstající složitost prostorového uspořádání atomů pomocí pojmů běžných ve strukturní biochemii, s tím, že se soustředíme spíše na podstatu věci, než na názvoslovná úskalí.

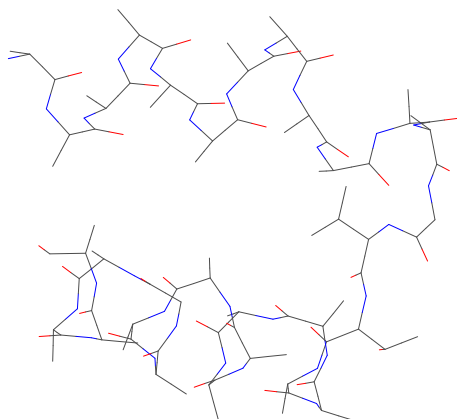
Ohyby (příklady některých byly zmíněny výše) spolu s úseky sekundárních struktur  $\alpha$  a  $\beta$  tvoří vyšší strukturní prvky, takzvané *supersekundární motivy*. Jejich kombinací vznikají *foldy*, které dále tvoří *domény* (struktury s funkční autonomií) a z domén je složena trojrozměrná struktura celého polypeptidového řetězce. Následující přehled *tříd proteinů* ilustruje tuto hierarchii proteinové architektury na strukturách skládaných z různých stavebnicových prvků.

## 2.11 Alfa-proteiny

Proteiny tvořené  $\alpha$ -šroubovicemi označujeme jako  *$\alpha$ -proteiny*. Nejjednodušším supersekundárním motivem  $\alpha$ -proteinů je dvojice antiparalelních  $\alpha$ -šroubovic spojených krátkým ohybem (v anglické literatuře označovaná *helix-turn-helix*). Typické jsou ohyby skládající se ze dvou, tří, nebo čtyř aminokyselin. Ramachandranovy diagramy typických  $\alpha$ -ohybů jsou uvedeny na obrázku 2.14.

Jiným příkladem supersekundárního motivu je „zlomený helix“ ( *$\alpha$ -corner*), ve kterém je pravidelná  $\alpha$ -šroubovice přerušena trojicí aminokyselin v neobvyklé konformaci (obrázek 2.15).

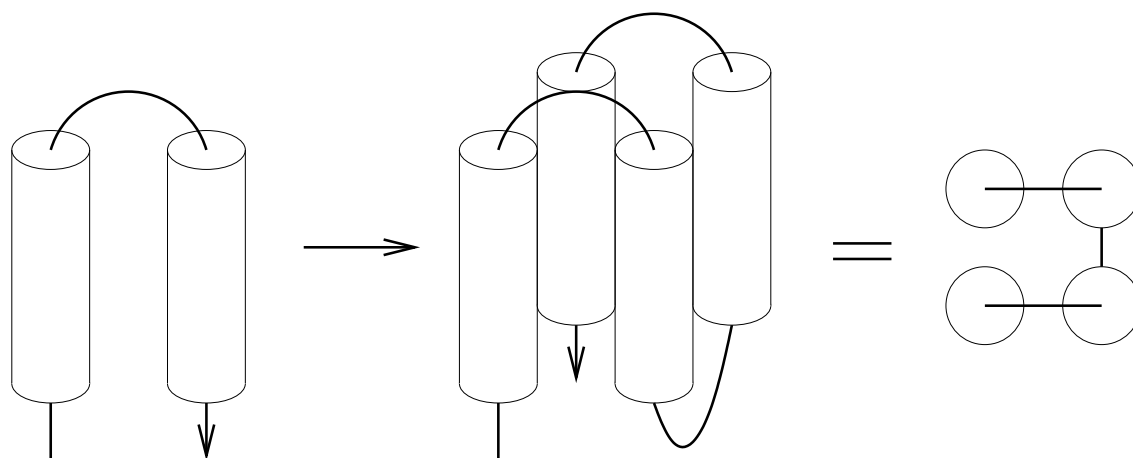
Termínem „fold“ se označují složitější útvary, které se skládají z několika úseků sekundární struktury a spojovacích prvků. Typickým příkladem je pro  $\alpha$ -proteiny svazek antiparalelních helixů (*helix bundle*). Obrázek 2.16 ukazuje jak je možné kombinací supersekundárního motivu helix-turn-helix sestavit svazek čtyř šroubovic. Existuje více možností, jak propojit pomocí ohybů několik šroubovic, které tvoří svazek. Tak jak jsme hovořili o topologii a konfiguraci jednotlivých atomů, můžeme hovořit o topologii i na

Obrázek 2.15: Model zlomeného helixu ( $\alpha$ -corner).

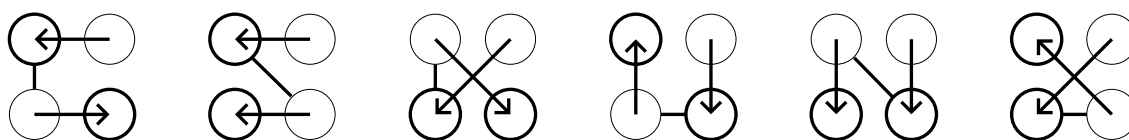
úrovni celých úseků sekundární struktury. Obrázek 2.17 ukazuje možné topologie svazků tvořených čtyřmi helixy. Jak vidíme, těchto možností je šest. Kdybychom podobně analyzovali svazky tvořené více šroubovicemi, dospěli bychom k velkému počtu možných topologií (propojení). Ne každá topologie je ale stejně běžná. Uspořádání, ve kterých se spojovací smyčky “kříží” vyžadují, aby se jedna smyčka druhé vyhnula, což je entropicky nevýhodné.

Jistě vás napadla otázka, jaké síly drží svazek šroubovic pohromadě. Odpověď můžeme hledat v rozložení postranních řetězců podél helixů. Když se podíváme na  $\alpha$ -šroubovici shora (levá část obrázku 2.18 nahoře), uvidíme takzvané *helikální kolo* (*helical wheel*). Postranní řetězce aminokyselin směřují do jedenácti různých směrů (jinými slovy, dvanáctý postranní řetězec míří stejným směrem, jako první). V helikálním kole se velmi často střídají polární a nepolární aminokyseliny tak, že nepolární postranní řetězce tvoří souvislý pruh podél délky helixu (například je nepolární první, pátá, devátá, dvanáctá atd. aminokyselina). Čtyři takové šroubovice se mohou výhodně orientovat tak, aby nepolární pruhy mířily dovnitř svazku a polární ven, do vodného prostředí. Spolu s dalšími interakcemi mezi postranními řetězci tak bude svazek stabilizován hydrofobním efektem. Důsledkem popsaného uspořádání je to, že dojde k vytvoření nepolární dutiny mezi helixy. Protein vlastně odděluje vodné a nepolární prostředí. To je důležité, pokud protein funguje jako katalyzátor chemické reakce, která musí probíhat v nevodném prostředí. Reakce uvnitř proteinu může probíhat i ve vodném prostředí cytoplasmy, protein vlastně tvoří jakousi zkumavku s nevodným rozpouštědlem. Jiným příkladem oddělení vodného a nevodného prostředí jsou proteiny, které procházejí lipidovou membránou. Takové proteiny tvoří opačně uspořádané svazky. Nepolární řetězce míří ven do nepolárního prostředí membrány, zatímco polární řetězce utvářejí polární dutinu, která spojuje vodná prostředí na obou stranách membrány (například cytoplasmu s mezibuněčnou tekutinou).

Ve většině proteinů nejsou helixy uspořádány rovnoběžně do svazků. Jednotlivé šroubovice navzájem svírají určitý úhel. Logiku takového uspořádání můžeme opět hledat v rozložení postranních řetězců podél helixu. V případě šroubovice  $3_{10}$  tvoří každý čtvrtý (například první, pátý, devátý, atd.) postranní řetězec jakýsi hřebínek na boku šroubovice. V případě  $\alpha$ -helixu není každá čtvrtá aminokyselina nad sebou, ale jsou pootočený vždy o  $33,7^\circ$ . V důsledku toho není spojnice postranních řetězců lišících se o tři aminokyseliny rovnoběžná s osou helixu (jak tomu bylo u šroubovice  $3_{10}$ ), ale svírá s ní úhel asi  $-26^\circ$ . Jednodušeji řečeno, náš hřebínek postranních řetězců se vine kolem šroubovice. Dvě šroubovice



Obrázek 2.16: Supersekundární motiv helix-turn-helix a fold svazku tvořeného čtyřmi šroubovicemi. Vpravo je propojení helixů ohyby znázorněno schematicky.



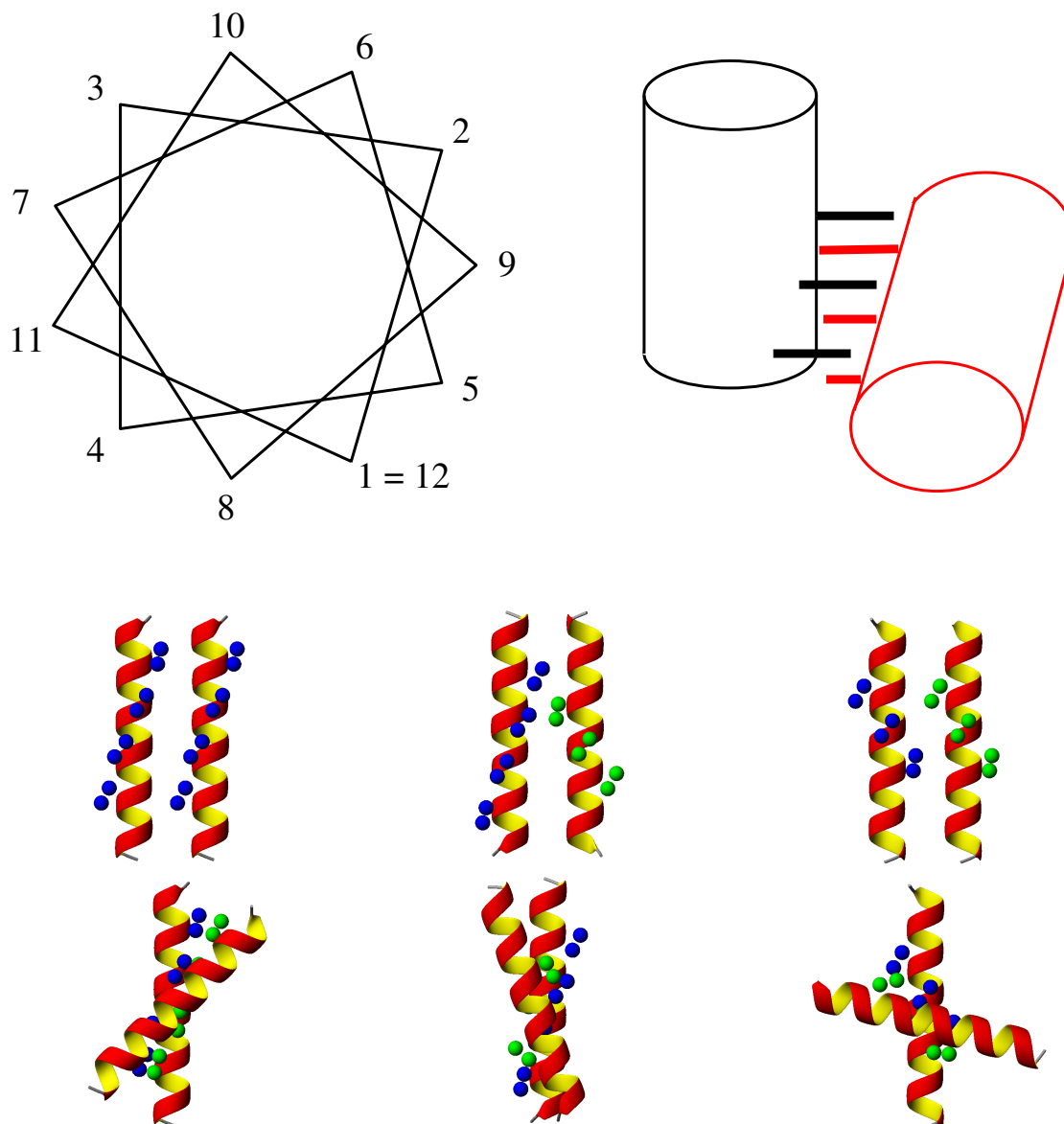
Obrázek 2.17: Možné topologie svazků čtyř helixů znázorněné schematicky podle obrázku 2.16. Šipky ukazují směr polypeptidového řetězce, tlustou čarou jsou kresleny kroužky natočené k nám *N*-koncem.

spolu často interagují tak, že do sebe hřebínky postranních řetězců zapadají jako dvě poloviny zdrhovadla (zipu). Protože se hřebínky vinou kolem šroubovic, musí být osy helixů vzájemně otočeny o  $-52^\circ$  (dvojnásobek zmíněných  $-26^\circ$ ), aby do sebe hřebínky ideálně zapadly (pravá část obrázku 2.18 nahoře). Proto velice často nacházíme ve strukturách proteinu šroubovice svírající spolu úhel kolem minus padesáti stupňů. Kromě hřebínku tvořeného každou čtvrtou aminokyselinou (říkejme mu  $i + 4$ -hřebínek) můžeme nalézt i pro aminokyseliny lišící se o tři rezidua ( $i + 3$ -hřebínek) nebo pro sousední aminokyseliny ( $i + 1$ -hřebínek). Pokud má  $i + 4$ -hřebínek zapadat do  $i + 3$ -hřebínku nebo  $i + 1$ -hřebínku, musí šroubovice svírat úhel  $+23^\circ$  respektive  $-105^\circ$ . Schematické modely uspořádání šroubovic s různými hřebínky jsou ukázány na obrázku 2.18 dole.

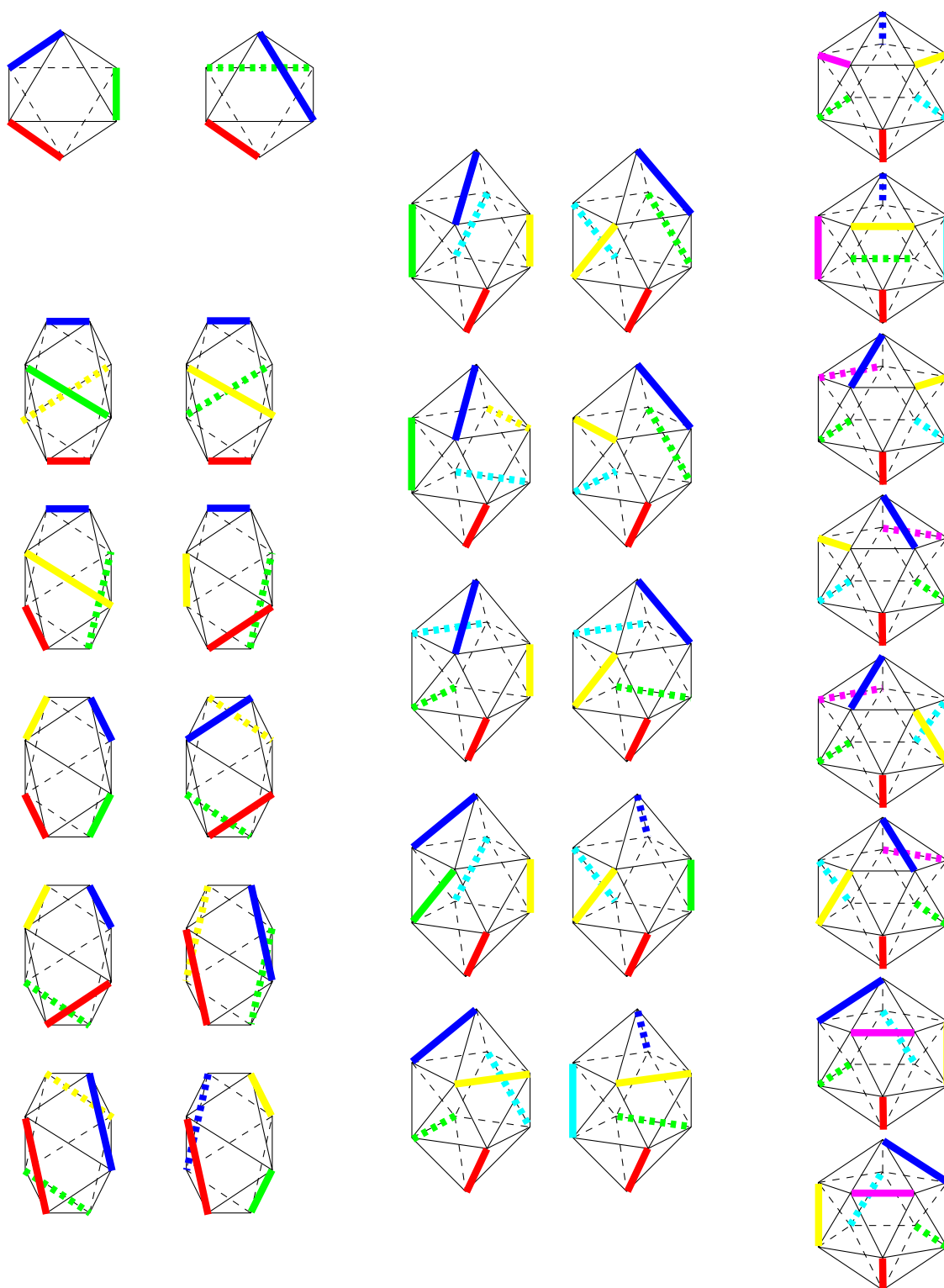
Jakým způsobem můžeme popsat foldy tvořené vzájemně pootočenými šroubovicemi? Analýza struktur proteinů ukázala, že helixy velmi často leží v hranách pomyslných mnohostěnů sestavených z rovnostranných trojúhelníků. Obrázek 2.19 ukazuje možnosti, jak uspořádat tři helixy v osmistěnu, čtyři helixy v dvanáctistěnu, pět helixů v šestnáctistěnu a šest helixů v dvacetistěnu. Když uvažíme i různé možnosti propojení helixů, vidíme, že možných foldů je velké, ale omezené množství. Pomyslné mnohostěny sestavené z šroubovic vlastně představují způsob, jak získat zhruba kulový tvar z rovných úseků sekundární struktury  $\alpha$ . Přibližně kulové makromolekuly proteinů opět umožňují oddělit nevodné prostředí uvnitř koule od vodného okolí. Podobně jako v případě svazků šroubovic směřují nepolární postranní řetězce většinou dovnitř molekuly a polární ven.

Je zajímavé, že foldy složené z více než šesti šroubovic už do pravidelného mnohostěnu uspořádány nejsou. V těchto foldech jsou jeden nebo dva helixy umístěny uvnitř mnohostěnu nebo svazku šroubovic. Takové centrální helixy obsahují především nepolární postranní řetězce.

Půjdeme-li v hierarchii struktur ještě výše, dospějeme na úroveň *domény*. Tímto termínem se označují uspořádané úseky polypeptidového řetězce, které mají nejen pravidelnou strukturu, ale často



Obrázek 2.18: Nahoře vlevo helikální kolo s pořadovými čísly aminokyselin, vpravo ilustrace helixů, které interagují prostřednictvím do sebe zapadajících hřebínků postranních řetězců. Dole schematické modely  $\alpha$ -šroubovic s hřebínky valinů vinoucími se kolem šroubovice (vlevo dva  $i+4$ -hřebínky, uprostřed jeden  $i+3$ -hřebínek a jeden  $i+4$ -hřebínek, vpravo dva  $i+3$ -hřebínky). Z postranních řetězců valinů jsou vždy ukázány pouze methylové skupiny, znázorněné jako modré a zelené kuličky. Pod rovnoběžně orientovanými šroubovicemi jsou vždy ukázány šroubovice svírající takový úhel, aby do sebe hřebínky nejlépe zapadaly.



Obrázek 2.19: Ideální uspořádání šroubovic v hranách pomyslných mnohostěnů. Umístění šroubovic jsou vyznačena barevně.

i svou dobře definovanou funkci (například schopnost specificky vázat jinou molekulu, nebo katalyzovat enzymovou reakci). Od podjednotek oligomerních proteinů, o kterých se zmíníme v sekci 2.16, se domény vlastně liší jen tím, že jsou součástí jediného peptidového řetězce. Příkladem  $\alpha$ -domény mohou být dva čtyřhelixové svazky.

Příklady proteinů obsahujících svazek  $\alpha$ -šroubovic a globulární struktury s  $\alpha$ -šroubovicemi tvořící hrany mnohostěnů jsou ukázány na stereo-obrázcích 2.20–2.24.

## 2.12 Beta-proteiny

Proteiny tvořené skládanými listy bývají nazývány  $\beta$ -proteiny. Obdobou supersekundárního motivu *helix-turn-helix*, který jsme si uvedli pro  $\alpha$ -proteiny, je u  $\beta$ -proteinů takzvaná  $\beta$ -vlásenka ( *$\beta$ -hairpin*), která je součástí struktury zobrazené na obrázku 2.25 (vlásenka je v pravé horní části struktury).

Název je inspirován tvarem motivu. V  $\beta$ -hairpinech jsou dvě antiparalelní  $\beta$ -vlákna spojena krátkým ohybem (takzvaný  *$\beta$ -turn*), který může být tvořen dvěma, třemi, nebo čtyřmi aminokyselinami. V případě dvouaminokyselinových  $\beta$ -ohybů jde nejčastěji o výše zmíněné typy I' a II'. Ramachandranovy diagramy tří- a čtyřaminokyselinových  $\beta$ -ohybů jsou uvedeny v levé části obrázku 2.26.

Spodní část struktury na obrázku 2.25 je  $\beta$ -obdobou „zlomeného helixu“ a nazývá se  *$\beta$ -corner*. Pravá část obrázku 2.26 znázorňuje konformace glycinu na vnitřním vlákně  *$\beta$ -corneru*, kde dochází ke zlomu a aminokyseliny, kterou je přerušeno párování  $\beta$ -listu na vnější straně zlomu.

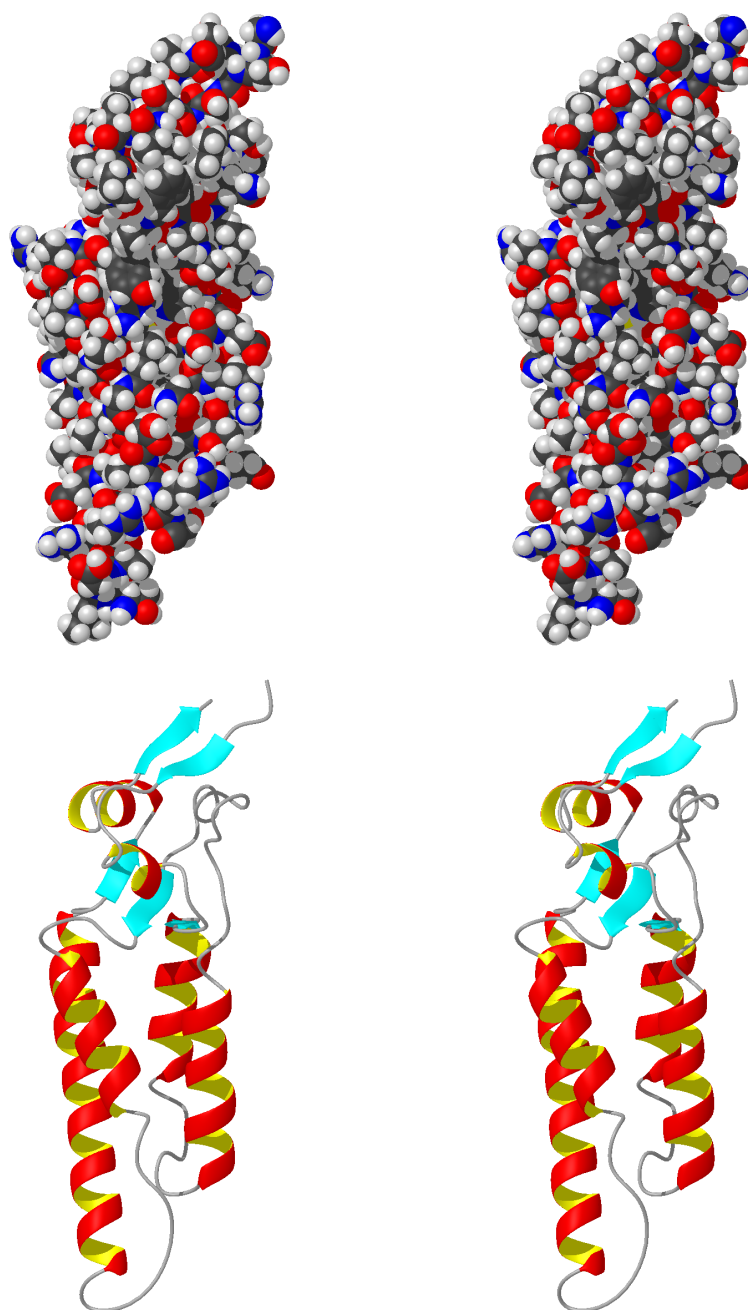
Kombinací supersekundárních struktur vznikají  $\beta$ -foldy, stejně jak jsme si to popsali u  $\alpha$ -proteinů. Skládáním většího počtu antiparalelních  $\beta$ -vláken vznikají  *$\beta$ -meandry* (horní část obrázku 2.27). Možných propojení  $\beta$ -vláken je samozřejmě více. Jako příklad  *$\beta$ -foldu* s odlišnou topologií můžeme uvést motivy řeckého klíče (*Greek key*) a „rolády“ (*jellyroll*), ukázané ve spodní části obrázku 2.27.

Na rozdíl od jasně definovaných šroubovic jsou  $\beta$ -listy mnohem volnější, jak ukazuje mnohem větší konformační volnost v Ramachandranově diagramu. Struktura  $\beta$ -proteinů je jen vyjimečně tvořena jediným skládaným listem. Často dva skládané listy leží proti sobě a tvoří takzvaný  *$\beta$ -sandwich*. Základní princip tvorby této struktury je zřejmý na jejím nejjednodušším příkladu, takzvaném motivu  *$\beta$ - $\beta$ - $\beta$*  (obrázek 2.28 vlevo). Všimněte si, že existují dvě možnosti propojení antiparalelních  $\beta$ -vláken. Ukázaná pravotočivá konformace je výhodnější a setkáváme se s ní v proteinech<sup>17</sup>. V  *$\beta$ -sandwichových* foldech jsou jednotlivá vlákna zkroucena obvyklým způsobem a protilehlá vlákna jsou vzájemně pootočena asi o 30°. Postranní řetězce mezi listy  *$\beta$ -sandwiche* do sebe nezapadají jako v případě interagujících šroubovic, ale směřují spíše proti sobě a vyplňují prostor mezi listy (obrázek 2.28 vpravo). Opět tak vzniká možnost utvořit nepolární prostředí mezi listy. Dva příklady  *$\beta$ -sandwichů* s různou topologií jsou ukázány na obrázku 2.29.

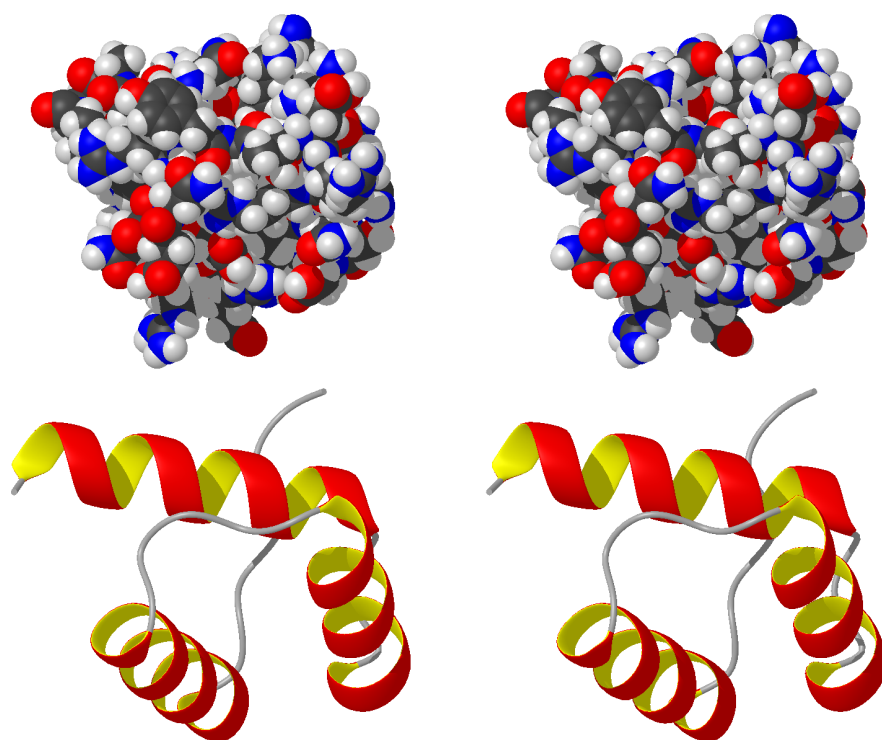
Jiným typickým foldem  $\beta$ -proteinů je takzvaný  *$\beta$ -soudek* ( *$\beta$ -barrel*), který umožňuje  $\beta$ -proteinům vytvořit přibližně kulový útvar, s nepolárními postranními řetězci orientovanými dovnitř a polárními ven (obrázek 2.30). Na rozdíl od  *$\beta$ -sandwiche* svírají protilehlá vlákna v  $\beta$ -soudcích téměř pravý úhel. Vlákna tvořící soudek neleží rovnoběžně s osou soudku, ale svírají s ní určitý úhel. Tento úhel je u různých struktur různý, stejně jako počet vláken tvořících soudek (od čtyř po dvacet). Jen některé kombinace ale umožňují vyplnit soudek postranními řetězci a vytvořit tak nepolární dutinu. Existují také proteiny tvořící  $\beta$ -soudky s hydrofobními aminokyselinami na povrchu a hydrofilními postranními řetězci mířícími dovnitř. Takové proteiny se zabudovávají do membrán jako póry, kterými mohou procházet malé hydrofilní molekuly.

<sup>17</sup>Výhodnost pravotočivého uspořádání souvisí s tím, že  $\beta$ -vlákna jsou sama zkroucena. Pokud by byla  $\beta$ -vlákna zcela natažena, musely by smyčky s prostředním  $\beta$ -vlákem uzavřít celý kruh, tedy umožnit otočení hlavního řetězce o 360°. Jsou-li  $\beta$ -vlákna zkroucena o 30°, stačí pravotočivém uspořádání stočit hlavní řetězec o 330°, zatímco levotočivé uspořádání by vyžadovalo přetočení o 390°. Čím je úhel otočení menší, tím více kombinací konformací ve spojovacích smyčkách umožňuje správné napojení řetězců. Výhodnost pravotočivé konformace je tedy entropická.

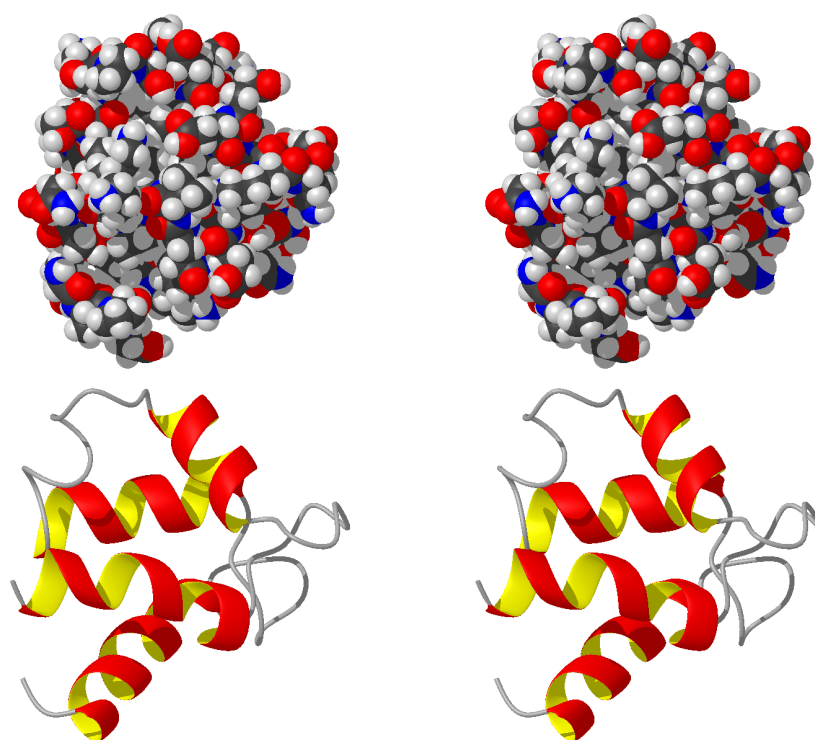




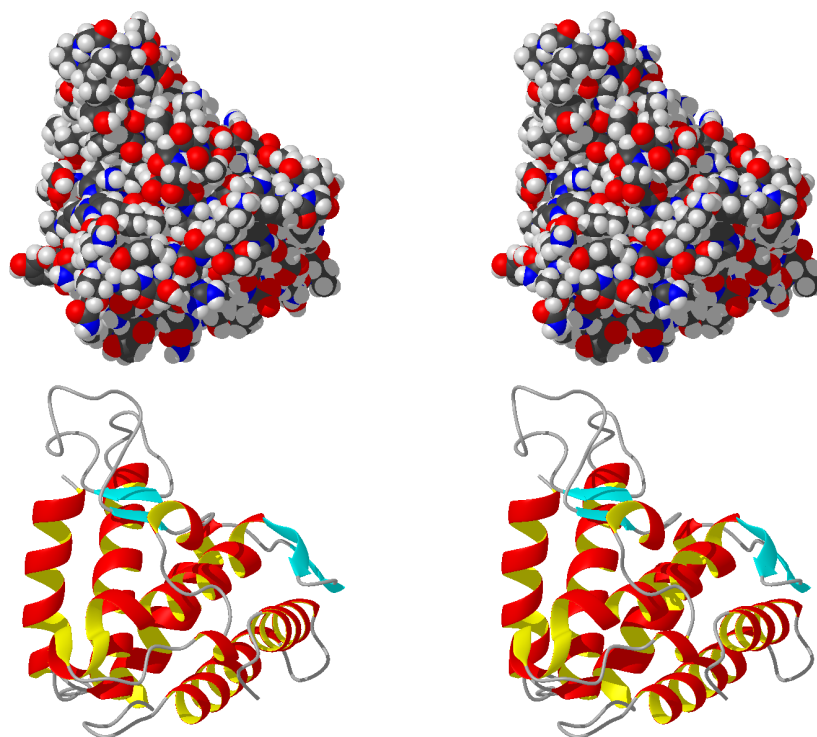
Obrázek 2.20: Protein tvořící svazek  $\alpha$ -šroubovic: peptid viru tabákové mozaiky (struktura 4GQH.pdb).



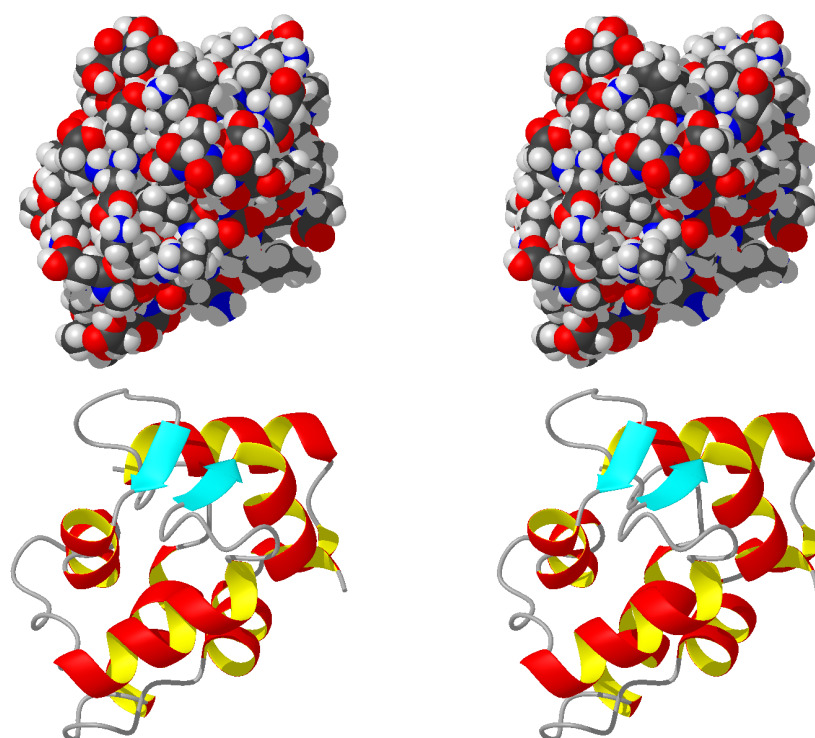
Obrázek 2.21: Protein s globulární strukturou složenou ze tří  $\alpha$ -šroubovic tvořících hrany osmistěnu: homeodoména Engrailed (struktura 1ENH.pdb).



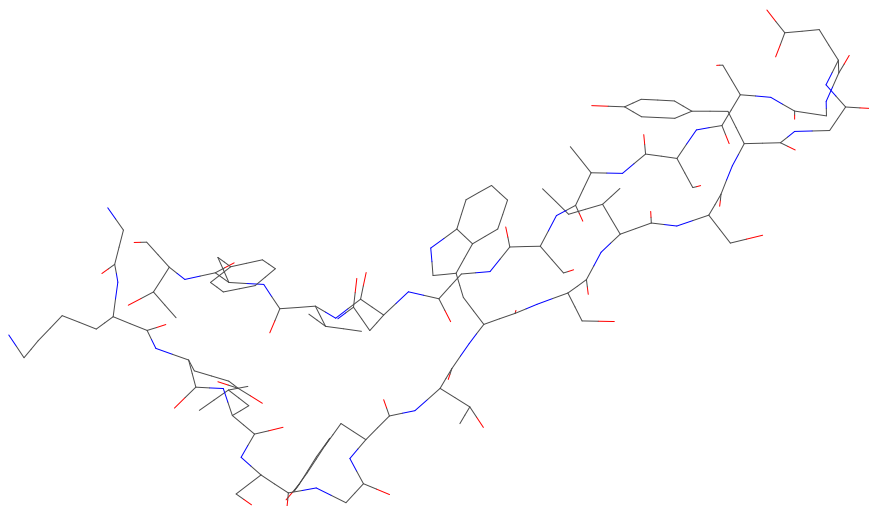
Obrázek 2.22: Protein s globulární strukturou složenou ze čtyř  $\alpha$ -šroubovic tvořících hrany dvanáctistěnu: střední protein vázající vápenaté ionty (struktura 3ICB.pdb).



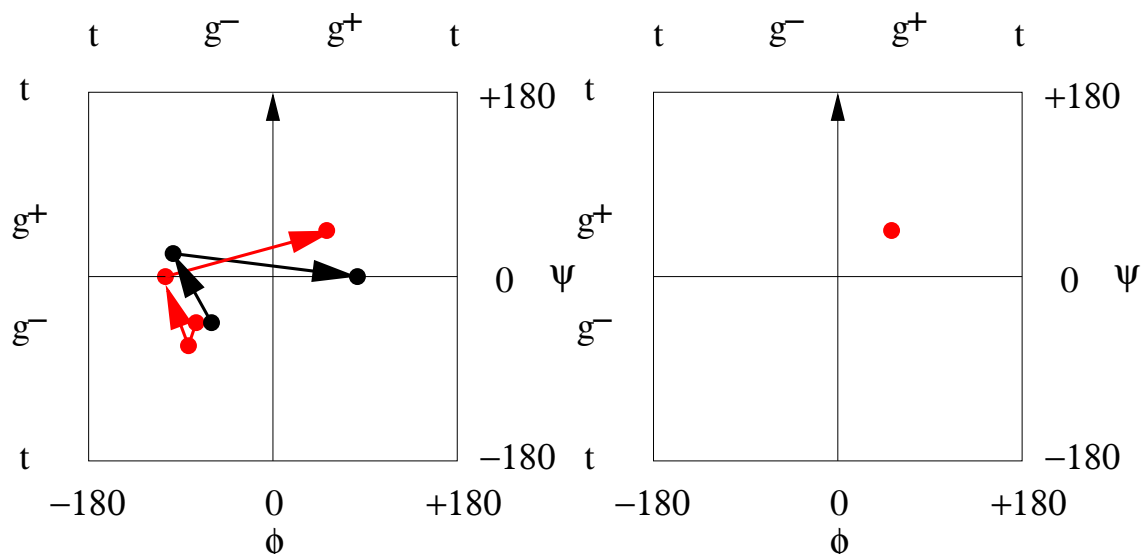
Obrázek 2.23: Protein, jehož C-koncová doména (levá část obrázku) má globulární strukturu složenou z pěti  $\alpha$ -šroubovic tvořících hrany šestnáctistěnu: thermolysin (struktura 1TMN.pdb).



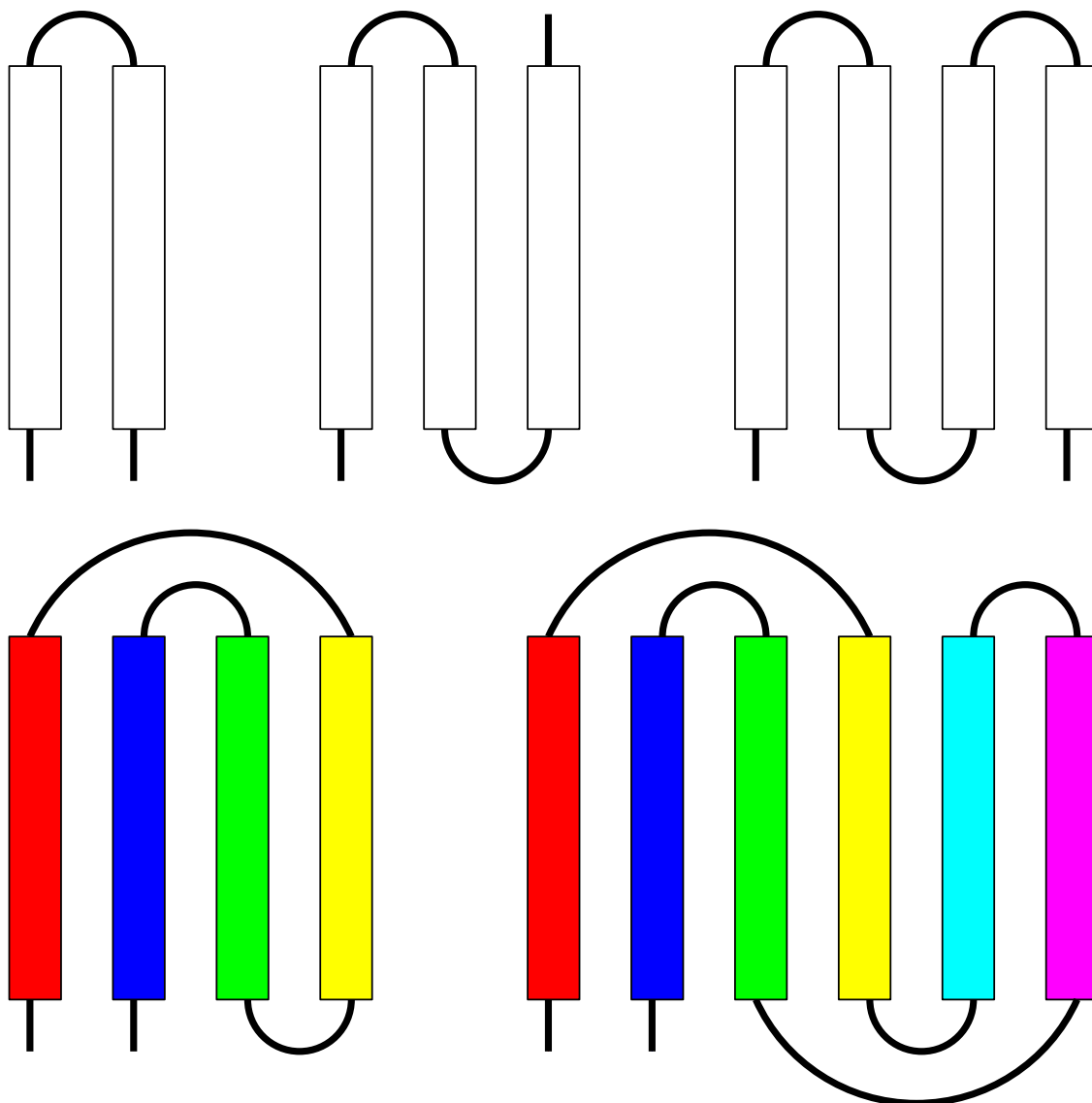
Obrázek 2.24: Protein s globulární strukturou složenou ze šesti  $\alpha$ -šroubovic tvořících hrany dvacetistěnu: parvalbumin (struktura 5CPV.pdb).



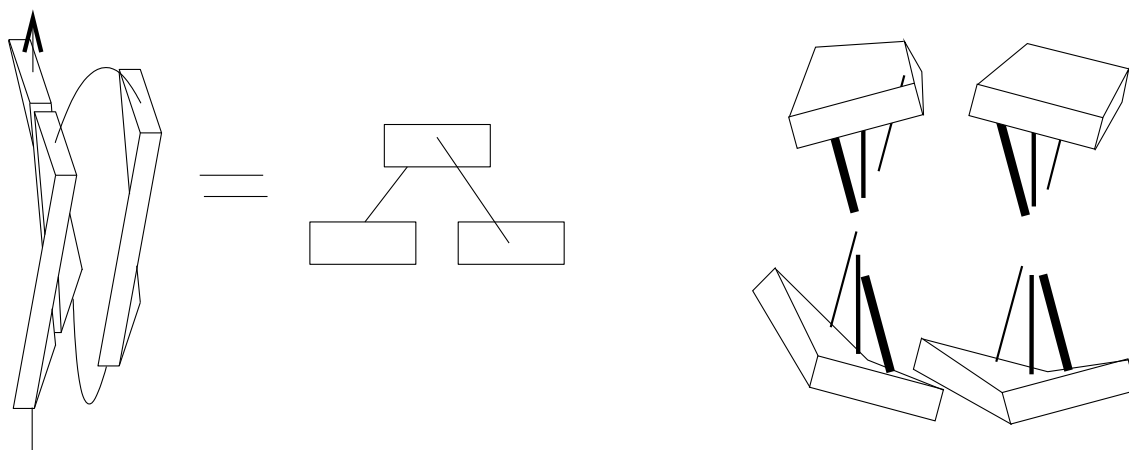
Obrázek 2.25: Model struktury obsahující  $\beta$ -vlásečku (vpravo nahoře) a  $\beta$ -corner (ve spodní části).



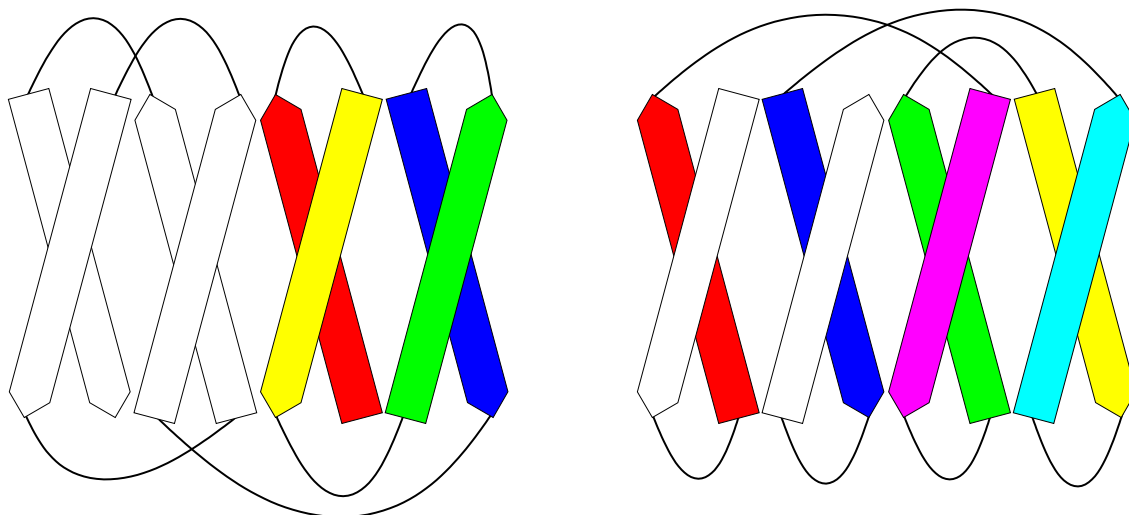
Obrázek 2.26: Ramachandranův diagram s vyznačením konformací tří nebo čtyř aminokyselin v ohybu  $\beta$ -vlásečky (vlevo) a konformací glycinu, na kterém dochází ke zlomu v  $\beta$ -corneru (vpravo). Význam šipek je podobný jako na obrázku 2.14.



Obrázek 2.27: Nahoře ukázka tvorby  $\beta$ -meandrů kombinací  $\beta$ -vlásenek, dole motivy řeckého klíče (vlevo) a „rolády“.

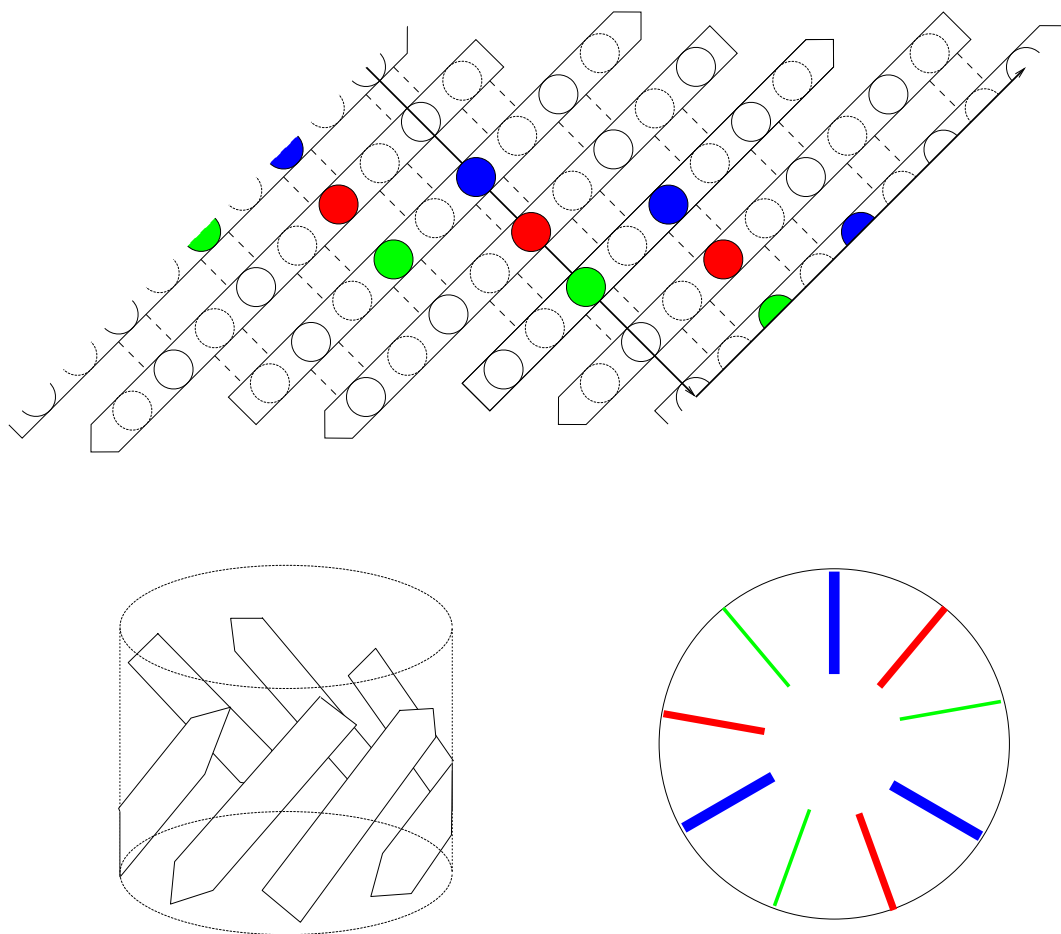


Obrázek 2.28: Vlevo pravotočivý  $\beta$ - $\beta$  motiv (se schematickým znázorněním topologie), vpravo uspořádání postranních řetězců mezi listy  $\beta$ -sandwichu (ukázána jsou jen dvě vlákna v každém listu).

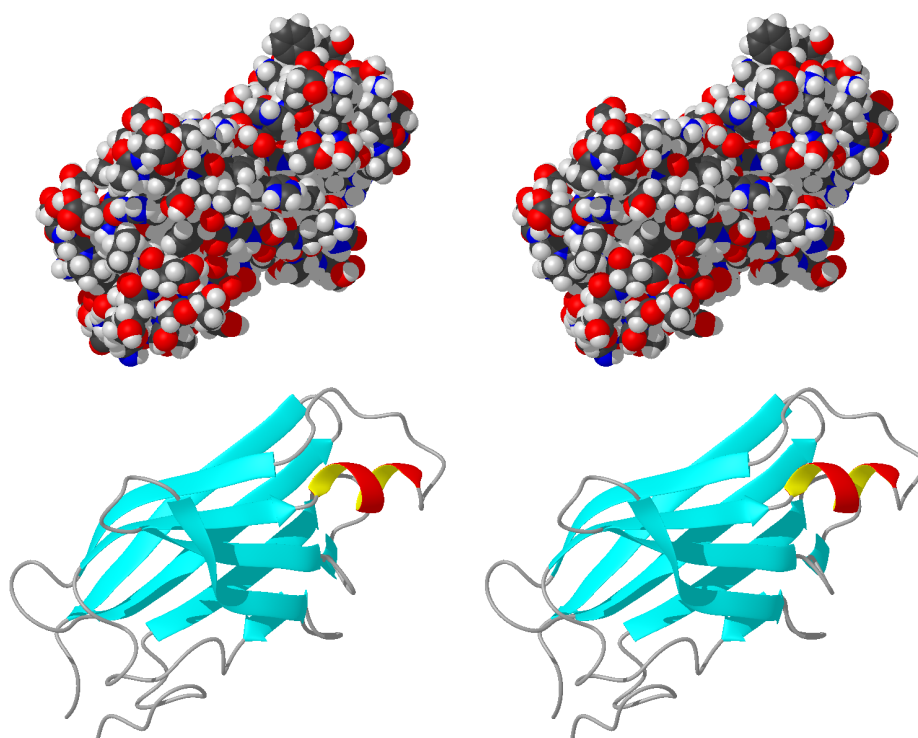


Obrázek 2.29: Příklady  $\beta$ -sandwichů. Příklad vlevo obsahuje čtyři vlákna s topologií řeckého klíče, příklad vpravo obsahuje šest vláken s topologií rolády. Pro lepší orientaci jsou vlákna obarvena stejně jako na obrázku 2.27.





Obrázek 2.30: Rozvinutý list tvořící  $\beta$ -sodek (nahore) a schematický nákras sodeku (vlevo dole). Pohled dovnitř sodeku (vpravo dole) ukazuje uspořádání nepolárních postranních řetězců uvnitř sodeku. Postranní řetězce směřující před rovinu a za rovinu nákresu (po svinutí dovnitř sodeku a ven) jsou odlišeny plnou respektive tečkovanou čarou obrysu kroužku. Vodíkové vazby jsou vyznačeny čárkovaně.



Obrázek 2.31: Protein se strukturou  $\beta$ -sandwiche: transthyretin (struktura 1ETA.pdb).

V bohatství foldů  $\beta$ -proteinů můžeme narazit i na další uspořádání  $\beta$ -listů. Jednotlivé listy mohou být uspořádány jako lopatky vrtule či turbíny, mluvíme pak o foldu  $\beta$ -propeller. Listy mohou tvořit stěny trojbokého hranolu ( $\beta$ -prism), s vlákny orientovanými kolmo k podstavě nebo rovnoběžně s podstavou. Struktura  $\beta$ -sandwiche se podobá takzvaný  $\beta$ -helix, ve kterém jsou protilehlá vlákna téměř rovnoběžná a velmi málo zkroucená. Celé listy pak jsou mírně šroubovicově zkrouceny (pravotočivě nebo levotočivě).

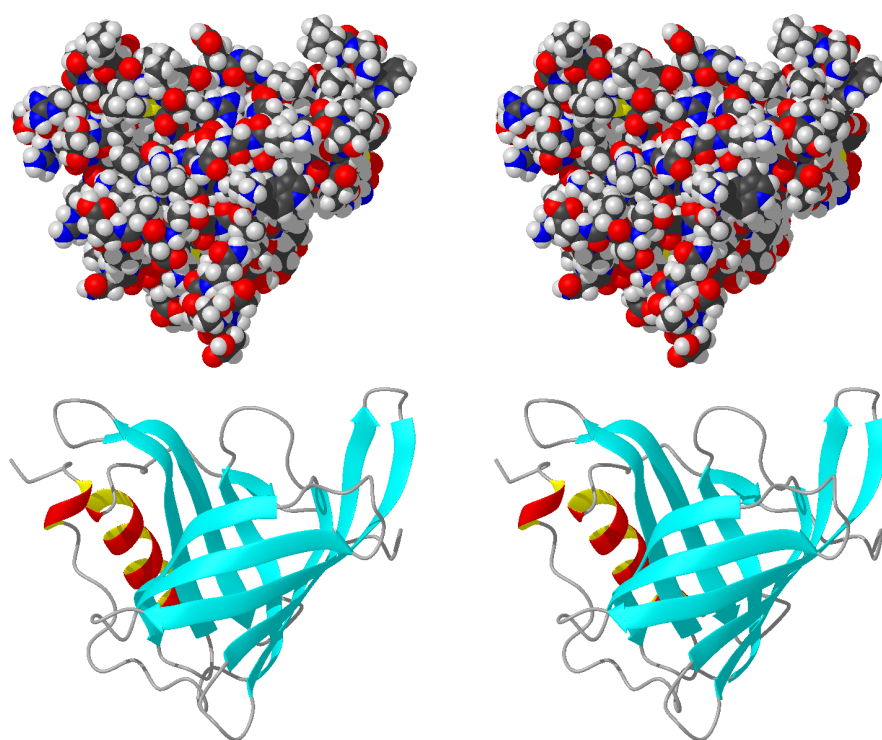
Foldy  $\beta$ -proteinů nám ukazují, jak lze s rozdílnou architekturou dosáhnout stejného cíle, kterým jsme se zabývali v případě  $\alpha$ -proteinů. Tímto cílem je vytvoření stabilní struktury s možností oddělení vodného a nevodného prostředí. I u  $\beta$ -proteinů jsou z foldů stavěny domény, které dále tvoří celkovou terciární strukturu. Příklady  $\beta$ -sandwiche a  $\beta$ -soudku jsou ukázány na obrázcích 2.31 a 2.36.

## 2.13 Alfa+beta-proteiny

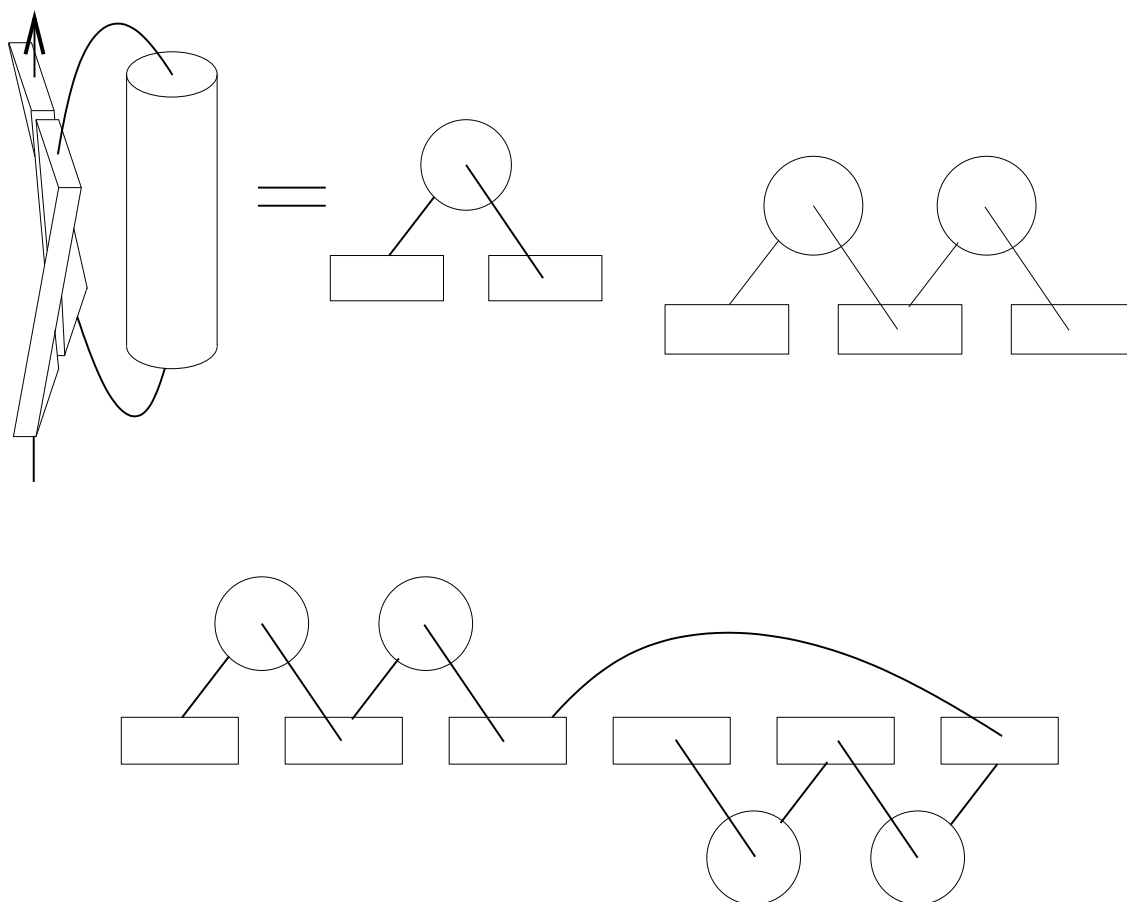
Mnohé proteiny obsahují  $\alpha$ - i  $\beta$ -foldy, které jsou spojeny až na úrovni vyšších strukturních celků. Takové proteiny, na jejichž stavbě se podílejí oddělené  $\alpha$ -motivy a  $\beta$ -motivy označujeme zkratkou  $\alpha + \beta$ .

## 2.14 Alfa/beta-proteiny

Kromě  $\alpha + \beta$ -proteinů s oddělenými částmi tvořenými buď  $\alpha$  nebo  $\beta$  strukturami se setkáváme s proteiny, v nichž se kombinují  $\alpha$  a  $\beta$  struktury již na úrovni supersekundárních motivů. Pro tyto proteiny se používá označení  $\alpha/\beta$ . Jako příklad supersekundárního motivu tvořeného oběma typy sekundárních struktur můžeme uvést  $\beta$ - $\alpha$ - $\beta$  motiv (obrázek 2.33 nahore). Jde o dvě paralelní  $\beta$ -vlákna, mezi která



Obrázek 2.32: Protein se strukturou  $\beta$ -soudku: protein vázající retinol (struktura 1RBP.pdb).



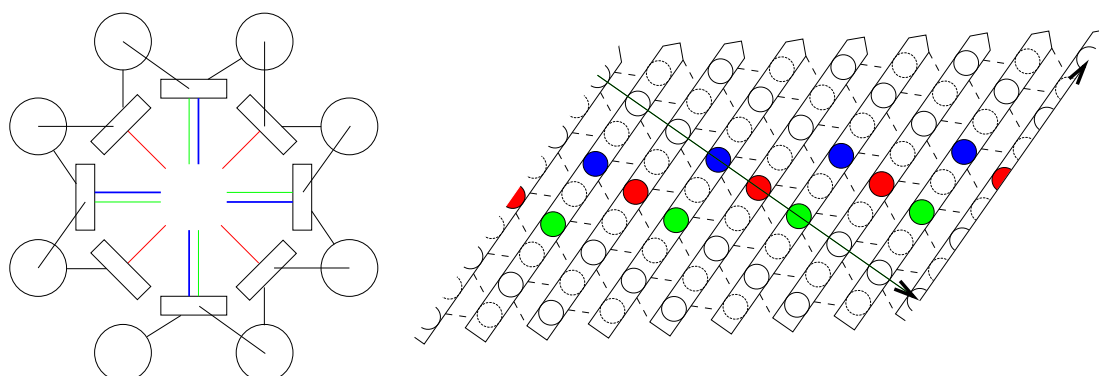
Obrázek 2.33: Vlevo nahoře pravotočivý  $\beta$ - $\alpha$ - $\beta$  motiv (se schematickým znázorněním topologie), vpravo nahoře schematické znázornění topologie Rossmanova foldu, dole schematické znázornění topologie domény tvořené dvěma Rossmanovými foldy.

je vložen opačně orientovaný  $\alpha$ -helix. Stejně jako v případě  $\beta$ - $\beta$ - $\beta$  motivu se v proteinech vyskytuje výhodnější pravotočivé uspořádání.

Pokračujeme-li v kombinování prvků sekundárních struktur  $\alpha$  a  $\beta$ , získáme  $\beta$ - $\alpha$ - $\beta$ - $\alpha$ - $\beta$  strukturu, která se jinak nazývá Rossmanův fold a tvoří vazebná místa pro nukleotidy (obrázek 2.33 nahoře).

Popsané  $\beta$ - $\alpha$ - $\beta$  motivy mohou být uspořádány do vyšších struktur podobně jako  $\beta$ -listy. Na obrázku 2.33 dole je příklad domény tvořené dvěma Rossmanovými foldy.

Jinou možností je sbalení do cyklických (nebo spíše válcových) struktur, obdoby  $\beta$ -soudků. Typickým příkladem je architektura poprvé nalezená u enzymu triosafosfatisomerasy, podle toho nazývaná *TIM-barrel*. V této struktuře tvoří  $\beta$ -listy uzavřený soudek, obklopený  $\alpha$  šroubovicemi (obrázek 2.34 vlevo). Na rozdíl od  $\beta$ -soudku jsou tentokrát všechna vlákna orientována jedním směrem, jde o paralelní skládaný list. Je pozoruhodné, že pro architekturu *TIM-barrelů* platí úplně stejná geometrická pravidla, jako pro  $\beta$ -soudky. Při tvorbě soudků zřejmě nezáleží na tom, zda je skládáme z paralelních nebo antiparalelních vláken. Klasický *TIM-barrel* je tvořen osmi  $\beta$ -vláknými a osmi helixy. Vnitřek je vyplněn třemi vrstvami postranních řetězců. Každá z vrstev je tvořena čtyřmi aminokyselinami, postranní řetězce vrchní a spodní vrstvy míří stejným směrem, zatímco prostřední vrstva je pootočená o  $45^\circ$ . Bylo by asi zbytečné opakovat, že vnitřek soudku může poskytnout nevodné prostředí.



Obrázek 2.34: Schematické znázornění topologie *TIM-barrelu* s vyznačenou polohou postranních řetězců vyplňujících soudek (vlevo) a rozvinutý list tvořící *TIM-barrel* (vpravo).

Příklady Rossmannova foldu a *TIM-barrelu* jsou ukázány na obrázcích 2.35 a ??.

## 2.15 Uspořádané proteiny bez sekundární struktury

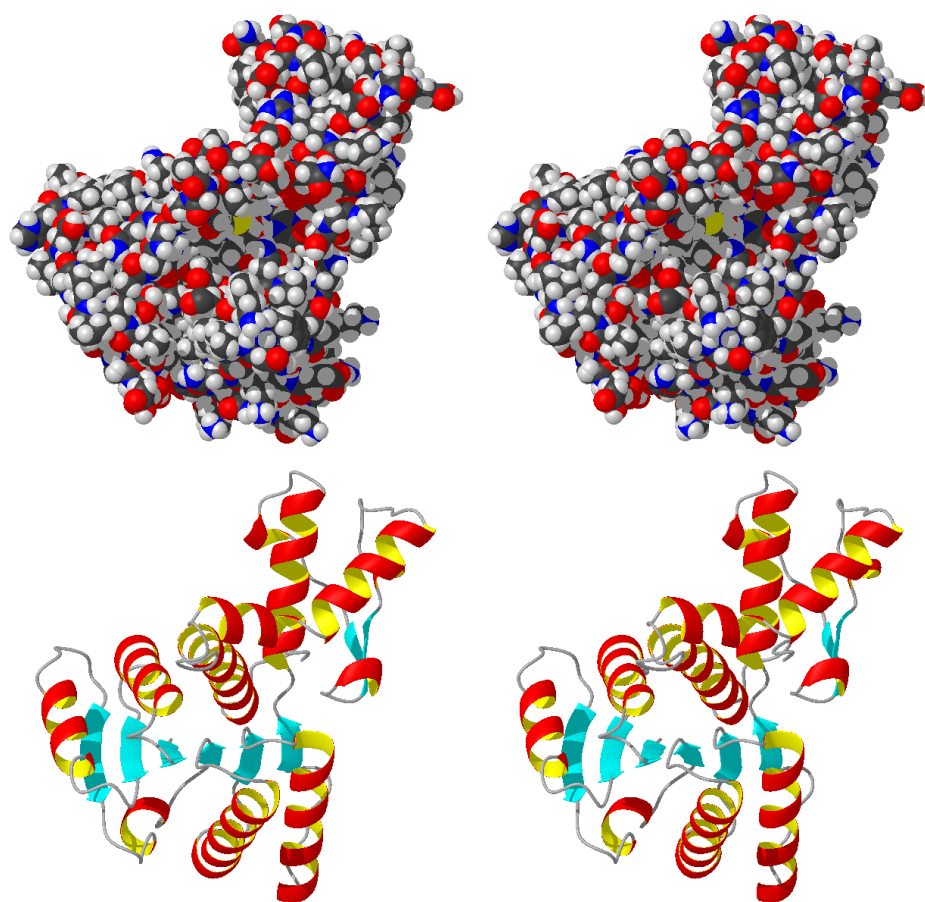
Ve velmi vzácných případech neobsahuje architektura proteinu žádné prvky pravidelné sekundární struktury. Jde o malé proteiny, jejichž struktury jsou stabilizovány disulfidovými vazbami (agglutinin) nebo koordinačními vazbami s navázaným kovovým iontem (například ferredoxin).

## 2.16 Kvarterní struktura a vyšší struktury

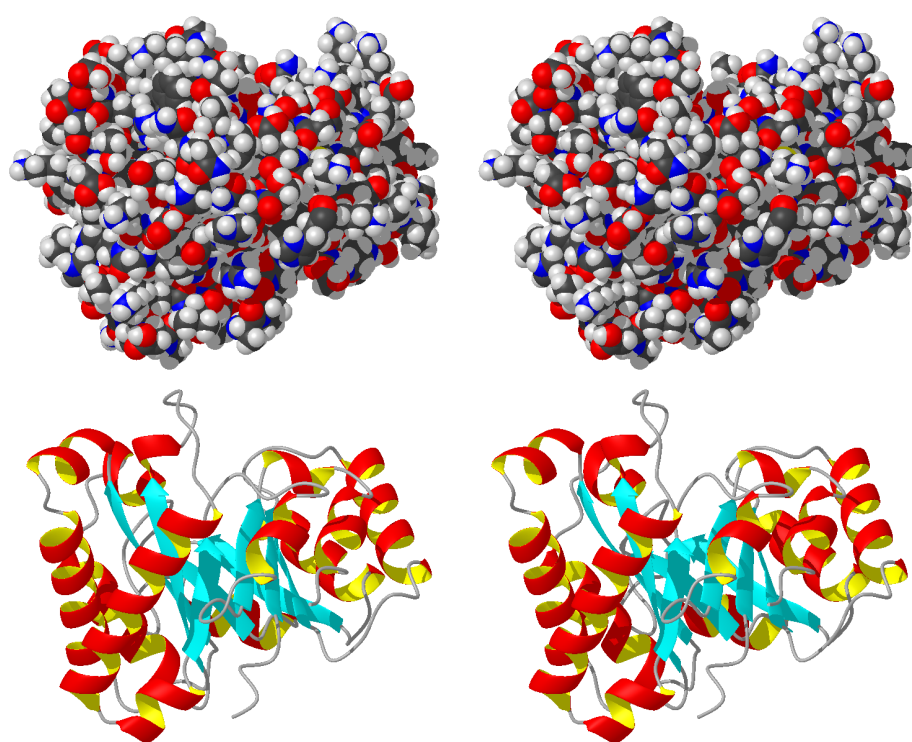
Polypeptidy tvořící terciární struktury mohou dále agregovat a tvořit oligomerní molekuly. Architektura takových komplexů se označuje jako kvarterní struktura. Oligomery se mohou skládat buď z několika stejných podjednotek (*homooligomery*) nebo z podjednotek odlišných (*heterooligomery*). Na rozdíl od domén tvoří podjednotky samostatné polypeptidové řetězce, které mohou být vzájemně vázány kovalentními disulfidovými vazbami mezi cysteiny jednotlivých podjednotek, nebo různými nekovalentními interakcemi, které jsme diskutovali v sekci 6.2. Je také časté, že funkce proteinů souvisí s opakovanou disociací a zpětnou asociací podjednotek.

Velká část proteinů obsahuje jako součást nepeptidovou molekulu. Může jít o jednoduché kovové ionty nebo o složité organické molekuly, jako například vitaminy, chlorofyl, hem, nukleotidy, oligosacharidy, mastné kyseliny. Nepeptidová část může být k peptidovému řetězci vázána kovalentně (v enzymologii se takto vázaná molekula nazývá *prostetická skupina*) nebo nekovalentně.

Proteiny se mohou dále shlukovat do nadmolekulárních útvarů (například ribozomů), ve kterých často interagují s molekulami většími a složitějšími, než jsou samy – jde zejména o oligosacharidy a nukleové kyseliny. Konečně se mohou proteiny podílet se na stavbě buněčných membrán, organel a jiných biologických útvarů, které svými rozměry přesahují hranice „nanosvěta“, ve kterém jsme zvyklí se pohybovat v chemii i biochemii.



Obrázek 2.35: Protein se strukturou Rossmanova foldu: malatdehydrogenasa (struktura 5KKA.pdb).



Obrázek 2.36: Protein se strukturou TIM-barrelu: triosafosfatisomerasa (struktura 5KKA.pdb).





## Kapitola 3

# Struktura nukleových kyselin

### 3.1 Stavební jednotky

V sekci 2.2 jsme viděli, že „abeceda proteinů“ má dvacet písmen, kterými je dvacet geneticky kódovaných proteinogenních aminokyselin. „Abeceda“ nukleových kyselin je mnohem jednodušší, tvoří ji pouze čtyři písmena genetického kódu, čtyři *nukleotidy*. Chemická struktura stavebních jednotek je ale složitější, než v případě proteinů. Každý nukleotid se skládá z pětiuhlíkatého monosacharidu, na který je navázána esterovou vazbou kyselina fosforečná a *N*-glykosidickou vazbou aromatický heterocyklus, zvaný *báze* (obrázek 3.1).

Atomy nukleotidů se číslují podle pokynů IUPAC-IUB, které se liší od standardního chemického názvosloví. Sacharidová a aromatická část nukleotidu se číslují zvlášť, proto se pro odlišení čísla atomů sacharidu označují čárkou.

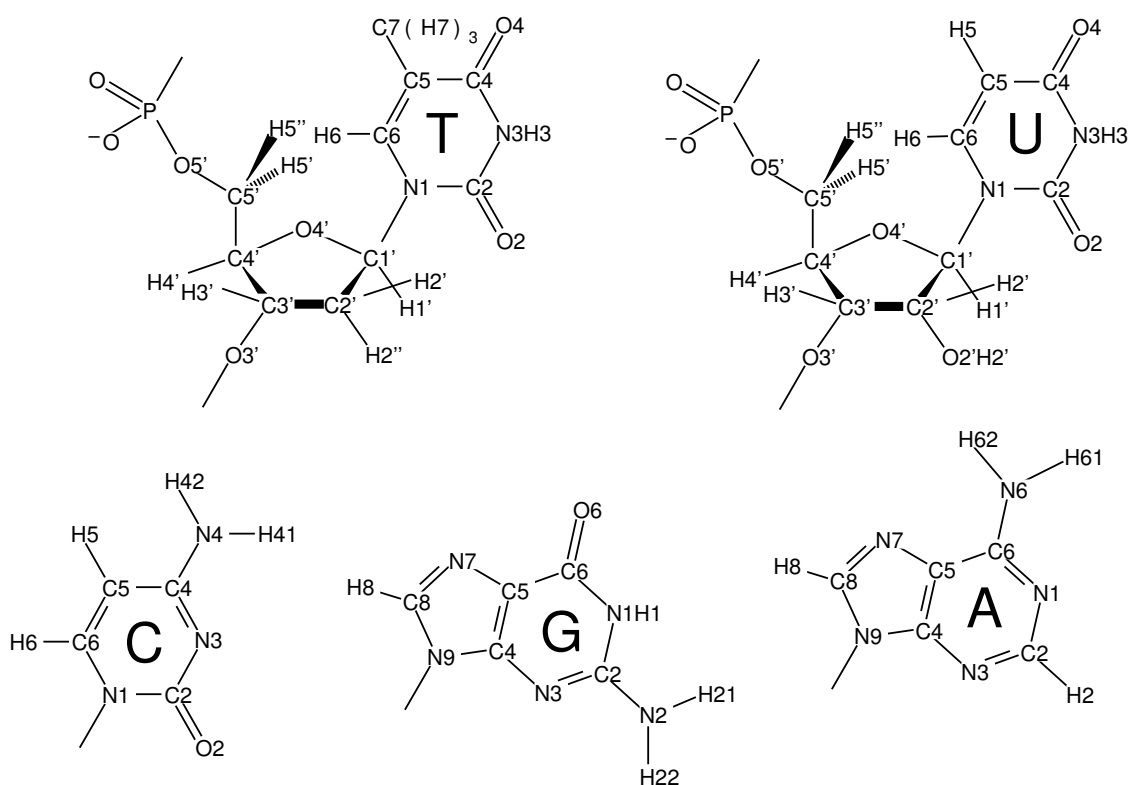
Struktury stavebních jednotek ribonukleové kyseliny (RNA) a deoxyribonukleové kyseliny (DNA) se liší ve dvou detailech. Jak již názvy napovídají, první odlišnost se týká sacharidu. V případě RNA je jím ribosa, zatímco v molekule DNA se vyskytuje deoxyribosa, která postrádá hydroxylovou skupinu na uhlíku C2'. Tato zdánlivá drobnost je příčinou velkých rozdílů ve struktuře a funkci RNA a DNA.

Zbytek kyseliny fosforečné je navázán v poloze 5' ribosy či deoxyribosy. Tak jako peptidový řetězec vzniká kondenzací karboxylové a aminové skupiny za vzniku amidové vazby, řetězec nukleové kyseliny vzniká kondenzací fosfátové skupiny vázané na uhlík C5' s hydroxyskupinou vázanou na C3' následujícího nukleotidu<sup>1</sup>. Kyselina fosforečná je tedy v polynukleotidovém řetězci vázána dvěma esterovými vazbami. Ribosy či deoxyribosy spojené přes fosfát tvoří hlavní řetězec nukleových kyselin (nazývaný *cukrfosfátová páteř*), který je obdobou peptidové páteře proteinů.

Roli postranních řetězců hrají v nukleových kyselinách báze. Zatímco konfigurace cukrfosfátové páteře je stejná v celé molekule, postranní řetězec může tvořit jedna ze čtyř bází. Je to tedy báze, co rozlišuje písmena genetické abecedy. Báze nukleových kyselin jsou amino- a hydroxyderiváty purinu a pyrimidinu (hydroxyskupina přechází tautomerií na oxo-formu).

Ve struktuře jedné z bází, pyrimidinu hydroxylovaného v polohách 2 a 4, je druhý chemický rozdíl mezi RNA a DNA. Zatímco v RNA se vyskytuje přímo 2,4-dihydroxypyrimidin (nazývá se uracil), v DNA nacházíme derivát této báze s methylovou skupinou na uhlíku C5 (označovaný jako thymin). Prvním písmenem genetické abecedy je tedy U (=uracil, v RNA) nebo T (=thymin, v DNA). Druhou pyrimidinovou bází je cytosin (C). Všimněme si, jak se liší struktury těchto bází vzhledem k možnosti tvorby vodíkových vazeb. Uracil a thymin mají donor v poloze 3 a akceptory v polohách 2 a 4. U cytosinu jsou v důsledku náhrady hydroxyly v poloze 4 aminoskupinou role donoru a akceptoru v polohách 3 a

<sup>1</sup>Uvědomme si, že v deoxyribose po vytvoření vazby nezbyvá žádný volný hydroxyl. Proto je DNA mnohem méně reaktivní než RNA, ve které je na ribose ještě hydroxyl v poloze C2'.



Obrázek 3.1: Struktura typické stavební jednotky DNA (vlevo nahoře) a RNA (vpravo nahoře). Ve stavebních jednotkách jsou zabudovány báze thymin a uracil, kterými se DNA a RNA liší. Dole jsou ukázány tři báze společné DNA i RNA (cytosin, guanin a adenin). Atomy jsou číslovány podle doporučení IUPAC-IUB.

4 vyměněny.

Jakýsi zrcadlový obraz vztahů, jaké jsme popsali u pyrimidinových bází, platí i pro další dvě písmena genetického kódu. Jsou to báze odvozené od většího heterocyklu purinu. Adenin (A) má v poloze 6 aminoskupinu (donor) a dusík N1 slouží jako akceptor. Guanin (G) má v poloze 6 hydroxyl a v poloze 2 aminoskupinu, takže ve výsledné tautomerní formě jsou donory v polohách 1 a 2 a akceptor v poloze 6.

Jako aromatické heterocykly jsou báze nukleových kyselin ploché. V molekule nukleové kyseliny mohou báze vzájemně interagovat buď plochami nebo hranami. Pokud jsou báze uspořádány nad sebou tak, že se k sobě blíží celou plochou (která je vymezena elektrony v delokalizovaných orbitalech, jež dodávají bázi aromatický charakter), dochází ke složité interakci zvané *stacking*. Pokud se naopak báze k sobě blíží hranami, jsou velmi důležité vodíkové vazby mezi atomy, které tvoří příslušné hrany.

Jeden z dusíků bází (N1 pyrimidinových a N9 purinových) je vázán glykosidickou vazbou k uhlíku C1' ribosy nebo deoxyribosy. Ribosa s navázanou bází (bez fosfátu) se nazývá *nukleosid*. Nukleosidy tvořené uracilem, thyminem, cytosinem, adeninem a guaninem se nazývají (ve stejném pořadí) uridin, thymidin, cytidin, adenosin a guanosin<sup>2</sup>.

## 3.2 Torzní úhly

Zatímco konformace peptidové páteře aminokyselin je určena třemi torzními úhly, k popisu cukrfofosátové páteře nukleových kyselin je třeba úhlů šest:  $\alpha$  (definovaný atomy  $O3'_{i-1}$ ,  $P_i$ ,  $O5'_i$ ,  $C5'_i$ ),  $\beta$  (definovaný atomy  $P_i$ ,  $O5'_i$ ,  $C5'_i$ ,  $C4'_i$ ),  $\gamma$  (definovaný atomy  $O5'_i$ ,  $C5'_i$ ,  $C4'_i$ ,  $C3'_i$ ),  $\delta$  (definovaný atomy  $C5'_i$ ,  $C4'_i$ ,  $C3'_i$ ,  $O3'_i$ ),  $\varepsilon$  (definovaný atomy  $C4'_i$ ,  $C3'_i$ ,  $O3'_i$ ,  $P_{i+1}$ ),  $\zeta$  (definovaný atomy  $C3'_i$ ,  $O3'_i$ ,  $P_{i+1}$ ,  $O5'_{i+1}$ ). K určení geometrie celého nukleotidu je dále třeba určit konformaci furanosového kruhu (deoxy)ribosy a glykosidický torzní úhel  $\chi$  (definovaný atomy  $O4'_i$ ,  $C1'_i$ ,  $N1_i$ ,  $C2_i$  u pyrimidinů a  $O4'_i$ ,  $C1'_i$ ,  $N9_i$ ,  $C4_i$  u purinů). Protože konformace pětičlenného cyklu (stejně jako jakéhokoli řetězce pěti atomů, viz sekci 1.2) je jednoznačně dána dvěma torzními úhly a protože jeden z těchto úhlů je již definován úhlem  $\delta$ , celkový počet torzních úhlů potřebných k popisu nukleotidu je osm.

Až dosud jsme se vždy zabývali torzními úhly v lineárních řetězcích. Ribosa představuje příklad cyklického řetězce.

Pentosový kruh má poměrně bohatý výběr různých konformací: může být buď v obálkové konformaci ( $E$ , *envelope*), kdy čtyři atomy cyklu leží v rovině, nebo v konformaci zkrácené položidličky ( $T$ , *twisted halfchair*), kdy tři atomy leží v rovině, jeden nad rovinou a jeden pod rovinou. V  $E$  konformaci může být každý z pěti atomů buď nad rovinou kruhu (*endo*, orientován „dovnitř“ nukleotidu) nebo pod ní (*exo*, orientován „ven“ z nukleotidu). Schematicky značíme takový atom horním (*endo*) nebo dolním (*exo*) indexem (například  ${}_3E$ ,  $C3'$ -*exo*). Podobně existuje deset zkrácených konformerů, které rozlišujeme obdobně definovanými indexy.

Úhly<sup>3</sup>  $\nu_i$  popisující konformaci cukerného kruhu<sup>4</sup> jsou definovány podél směru (deoxy)ribosového řetězce:  $\nu_1$  definovaný atomy  $O4'_i$ ,  $C1'_i$ ,  $C2'_i$ ,  $C3'_i$  až po  $\nu_0$  definovaný atomy  $C4'_i$ ,  $O4'_i$ ,  $C1'_i$ ,  $C2'_i$ . Vidíme, že v pětičlenném kruhu můžeme naměřit pět torzních úhlů  $\nu_i$ , z nichž k popisu konformace postačují libovolné dva. Je tedy otázkou, které z nich použít. K popisu furanosových konformací se často používá formalismus takzvané *pseudorotace*, který problém volby dvou torzních úhlů obchází definicí dvou nových proměnných, které nahrazují dvojici torzních úhlů. První z těchto proměnných je *pseudorotační fáze*  $P$ . Slouží jako strelka kompasu (obrázek 3.2), na němž je každý typ konformeru definován určitým azimutem. Druhým parametrem, nezbytným k jednoznačnému určení geometrie je *pseudorotační amplituda*  $\Psi_m$ , která udává, o kolik je příslušný atom vychýlen z roviny. Pseudorotační parametry lze na endocyklické úhly  $\nu_i$  přepočítat pomocí vztahu

<sup>2</sup>Pozor na odlišný význam koncovky *-osin* u cytosinu a purinových nukleosidů.

<sup>3</sup>Tyto torzní úhly, definované atomy, které přímo tvoří cyklický řetězec, se nazývají *endocyklické*.

<sup>4</sup>V anglické literatuře se pro tento druh konformace používá označení *pucker*.

$$\nu_i = \Psi_m \cos(P + 144^\circ(i - 2)) \quad (3.1)$$

Definice týkající se konformací a pseudorotace jsou opět dány pravidly IUPAC-IUB.

### 3.3 Primární struktura

Význam pojmu primární struktura je u nukleových kyselin stejný jako u proteinů. Pořadí (sekvence) nukleotidů se podle konvence čte od 5' konce do 3' konce vlákna nukleové kyseliny. Zejména v případě prezentace dvoušroubovicových sekvencí je vhodné 5' a 3' konce vláken označit. Jak jsme zmínili v sekci 2.4, je určení sekvence nukleových kyselin poměrně snadné a primární struktura (konfigurace) je vždy známá dříve, než je řešena trojrozměrná struktura (konformace).

### 3.4 Sekundární struktura

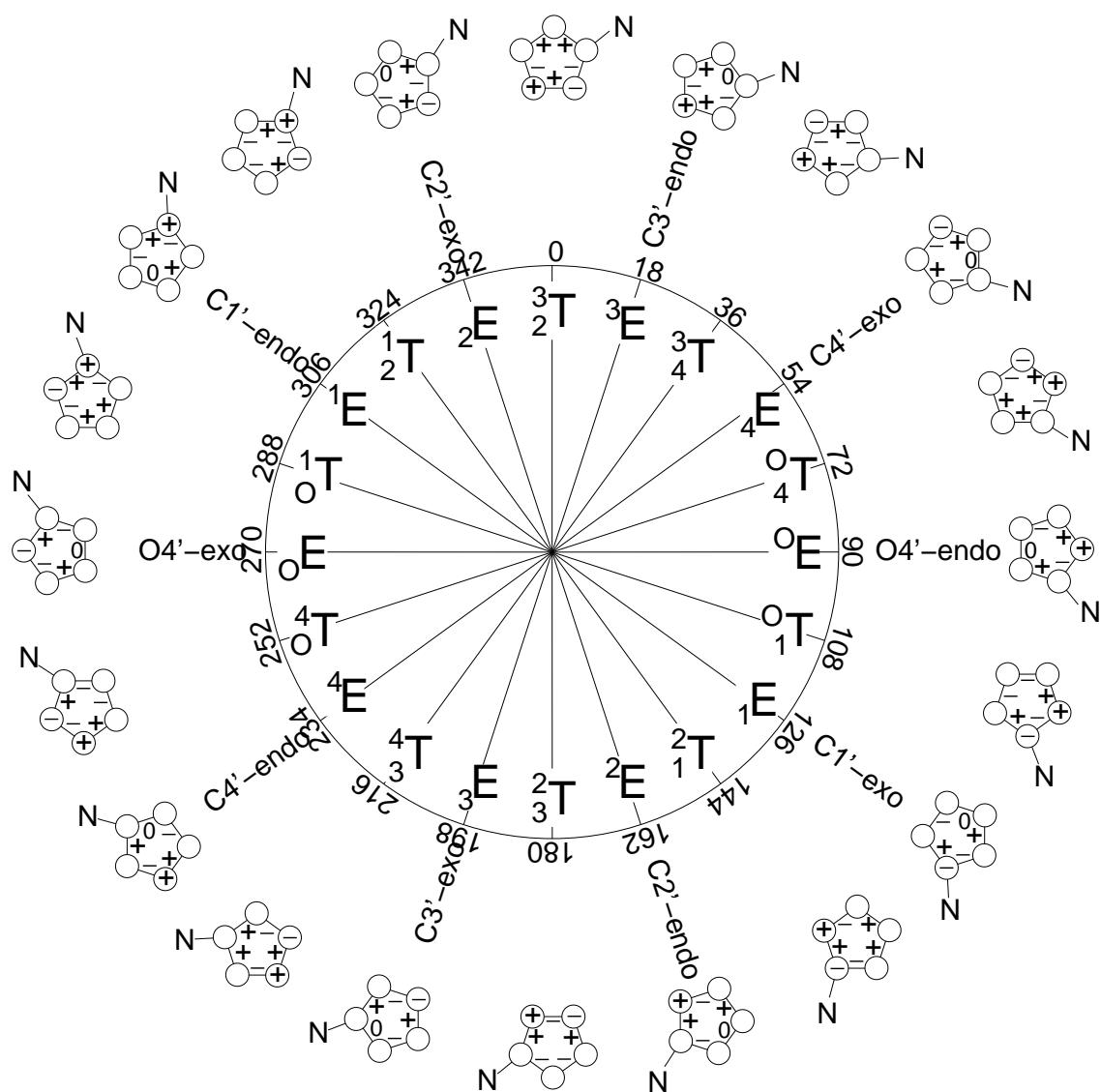
Na základě analogie s proteiny sekundární strukturou označujeme pravidelný, libovolně opakovatelný konformační motiv polynukleotidového řetězce. Základními sekundárními strukturami nukleových kyselin jsou dvouvláknové šroubovice stabilizované vodíkovými vazbami mezi bázemi. Všimněme si důležitého rozdílu mezi proteiny a nukleovými kyselinami: u sekundárních struktur proteinů jsme vždy hovořili o vodíkových vazbách mezi atomy páteře, zatímco dvojšroubovice nukleových kyselin jsou spojené sítí vodíkových vazeb mezi bázemi, které tvoří postranní řetězce polynukleotidu. Sekundární struktura není proto určena pouze konformací páteře, jako v případě proteinů. Je proto třeba definovat všech osm torzních úhlů určujících konformaci nukleotidu, aby byl strukturální motiv jednoznačně popsán.

Stejně jako proteiny, tvoří nukleové kyseliny dva základní typy sekundárních struktur, A a B. Na rozdíl od proteinových  $\alpha$  a  $\beta$  struktur se zmíněné sekundární struktury nukleových kyselin liší poměrně málo. V obou případech jde o pravotočivou dvojšroubovici, ve které jsou antiparalelní vlákna spojena bázemi párujícími se stejným způsobem.

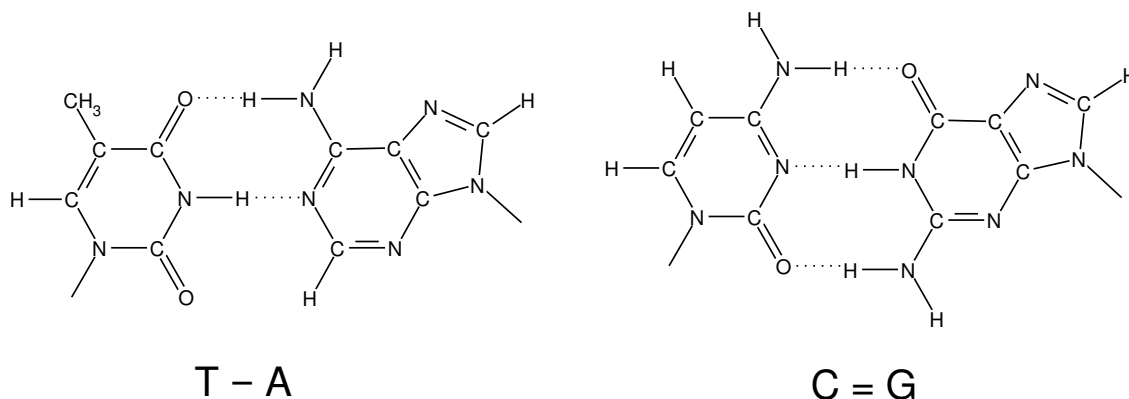
Způsob párování bází je klíčem k pochopení struktury dvojšroubovic nukleových kyselin. Vzpomeňme si na sekci 3.1, kde jsme si popisovali uspořádání donorů a akceptorů vodíkových vazeb v bázích. Zkusme si nakreslit jednu purinovou a jednu pyrimidinovou bázi. Začneme šestičlennými kruhy, natočenými „naležato“ tak aby dusík N1 purinu a dusík N3 pyrimidinu ležely proti sobě. Nakresleme vzorce bází tak, aby jedna glykosidická vazba směřovala vpravo dolů a druhá vlevo dolů (uspořádání *cis*). V uspořádání, které jsme si právě nakreslili (obrázek 3.3), se báze stýkají hranami, které obsahují nejvíce donorů a akceptorů vodíkových vazeb. Aby se vodíkové vazby mohly vytvořit, musí ležet vždy akceptor proti donoru. To je možné pouze tehdy, když guanin tvoří pár s cytosinem a adenin s uracilem nebo thyminem.

Všimněme si, že všechny páry popsané v předchozím odstavci mají téměř stejný tvar (jsou *izosterické*). Páruje se vždy menší pyrimidinová báze s větší purinovou a vcelku nezáleží na tom, která je vpravo a která vlevo. Tento způsob párování se nazývá podle Watsona a Cricka a zaručuje věrné předávání genetické informace. Máme-li jedno vlákno, které kóduje dědičnou informaci, může se s každou jeho bází párovat jen jedna určitá báze vlákna druhého. Pokud bude druhé vlákno dosyntetizováno tak, aby se všechny báze párovaly, ponese stejnou informaci. Ale pozor, toto druhé vlákno nebude kopií prvního. Vztah mezi nimi bude stejný jako vztah mezi formou a odliškem, teprve další syntézou vlákna, které by se párovalo s vláknem novým, bychom získali stejné pořadí bází.

Stejný tvar všech párů bází podle Watsona a Cricka umožňuje, aby byly součástí pravidelného útvaru – dvojšroubovice (viz obrázek 3.7). V ní musí mít správnou hodnotu také osm torzních úhlů, které popisují konformaci nukleotidu. V obou typech dvojšroubovic je konformace taková, že v páru nakresleném tak, jak jsme výše popsali, vystupuje hlavní řetězec báze nakreslené vlevo z roviny papíru k nám a hlavní řetězec báze nakreslené vpravo míří dolů pod rovinu papíru. Tabulka 3.1 udává typické



Obrázek 3.2: Pseudorotační kruh (deoxy)ribosy. Uvnitř kruhu je uvedeno symbolické značení konformací, vně kruhu pseudorotační fáze ve stupních a slovní označení konformací. Na obvodu diagramu jsou schematicky znázorněny struktury jednotlivých konformerů. Atomy jsou symbolizovány kroužky (nejdále od středu je vždy kyslík O4'). Symbol v kroužku uvádí, leží-li daný atom v rovině nákresu (○), nad rovinou (+), nebo pod rovinou nákresu (-). Znaménka u vazeb udávají, zda je daný torzní úhel kladný nebo záporný.



Obrázek 3.3: Párování bází podle Watsona a Cricka.

hodnoty jednotlivých torzních úhlů (ve stupních) pro obě dvojšroubovice. Pro úplnost je uveden i úhel  $\delta$ , ačkoli je jeho hodnota jednoznačně určena pseudorotačními parametry  $P$  a  $\Psi_m$ .

Tabulka 3.1: Hodnoty torzních úhlů ideálních dvojšroubovic.

	$\alpha$	$\beta$	$\gamma$	$\delta$	$\varepsilon$	$\zeta$	$\chi$	$P$	$\Psi_m$
B-DNA	-41	136	38	139	-133	-57	78	155	40
A-DNA	-90	-149	47	83	175	-45	27	13	40
A-RNA	-80	175	49	83	-147	-78	13	13	40

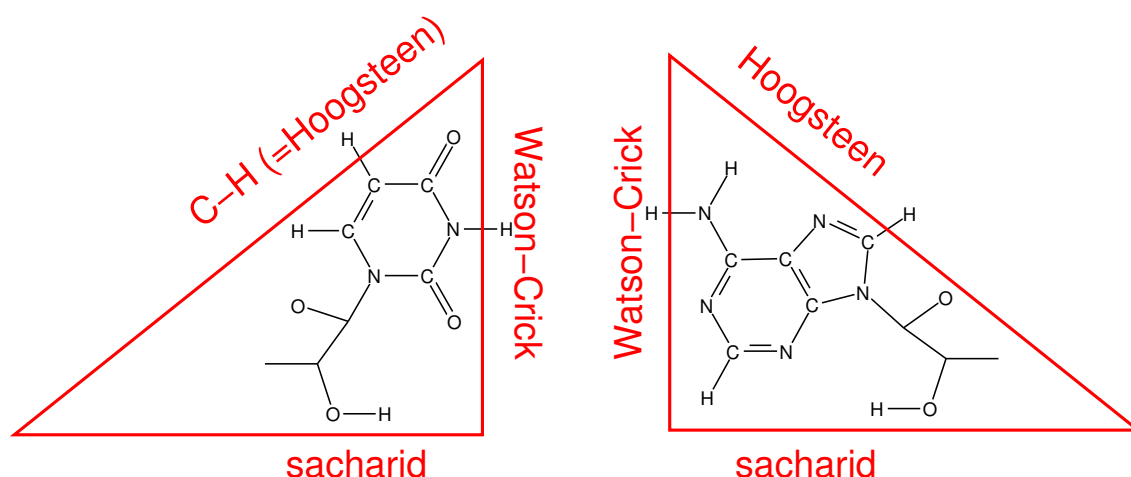
Tabulka 3.1 ukazuje, že nejvýraznější rozdíl torzních úhlů se týká konformace cukru. Zatímco (deoxy)ribosa dvojšroubovice A je v konformaci  $C3'$ -endo, deoxyribosa dvojšroubovice B je v konformaci  $C2'$ -endo (pro RNA je tato konformace nevýhodná). Vzhledem k jistým odchylkám od takto přesně definovaných konformací v reálných molekulách se často označují zmíněné konformace cukrů  $N$  a  $S$  (nalézají se v „severní“ (north) respektive „jižní“ (south) části pseudorotačního kompasu).

Jak ukazuje obrázek 3.7, cukr-fosfátová páteř ovíví dvojšroubovici zvenčí, takže zbytky kyseliny fosforečné (za běžných podmínek v disociovaném stavu, tedy záporně nabitě) jsou velmi dobře přístupné. Báze jsou naopak uvnitř dvojšroubovice, ne však zcela skryté. Mezi páteřemi jsou dostatečně široké žlábků, které umožňují jiným molekulám (nebo jiným částem téhož polynukleotidového řetězce) interagovat s bázemi. Každou bázi si můžeme zjednodušeně představit jako trojúhelník (obrázek 3.4). Tyto trojúhelníky jsou párovány jednou hranou (Watson-Crickovou), druhá hrana je přístupná z žlábků po jedné straně páteře a třetí hrana je přístupná z žlábků po druhé straně páteře.

Z vnějšího pohledu se liší vzhled žlábků mezi páteřemi jednotlivých vláken na povrchu dvojšroubovic. Dvojšroubovice B má jeden žlábek širší a hlubší (označuje se *hluboký* nebo *velký*) a druhý užší a mělký (označuje se *mělký* nebo *malý*), zatímco dvojšroubovice A má jeden žlábek velmi úzký a hluboký (označuje se *hluboký*) a druhý široký a velmi mělký (označuje se *mělký*). V obou případech do *hlubokého* žlábků směřuje hrana, která se u purinů nazývá Hoogsteenova a u pyrimidinů „C-H“ hrana<sup>5</sup>. Do *mělkého* žlábků naopak směřuje hrana v blízkosti glykosidické vazby, pro kterou byl proto navržen název „sugar-edge“ (do češtiny můžeme přeložit „sacharidová“).

Při pohledu podél osy dvojšroubovice je patrná dutina ve středu dvojšroubovice A. Dvojšroubovice A má také poněkud pozvolnější stoupání (11 reziduí na závit) ve srovnání s dvojšroubovicí B (10 reziduí

<sup>5</sup>Ve snaze o jednotnost názvosloví se často používá název „Hoogsteenova“ i pro „C-H“ hranu pyrimidinů.



Obrázek 3.4: Zjednodušený popis bází nukleových kyselin pomocí trojúhelníků (s označením jednotlivých hran).

na závit).

Pro popis vzájemné polohy bází ve dvojrůbovicích byly definovány takzvané *helikální parametry* (obrázky 3.5 a 3.6). Na první pohled je patrný rozdíl v *inklinaci*. Zatímco báze v dvojrůbovici B jsou téměř kolmé k ose, normály bází v dvojrůbovici A svírají s osou úhel asi 20°. Dvojrůbovice B a A jsou ukázány na obrázku 3.7 a na modelech A a B.

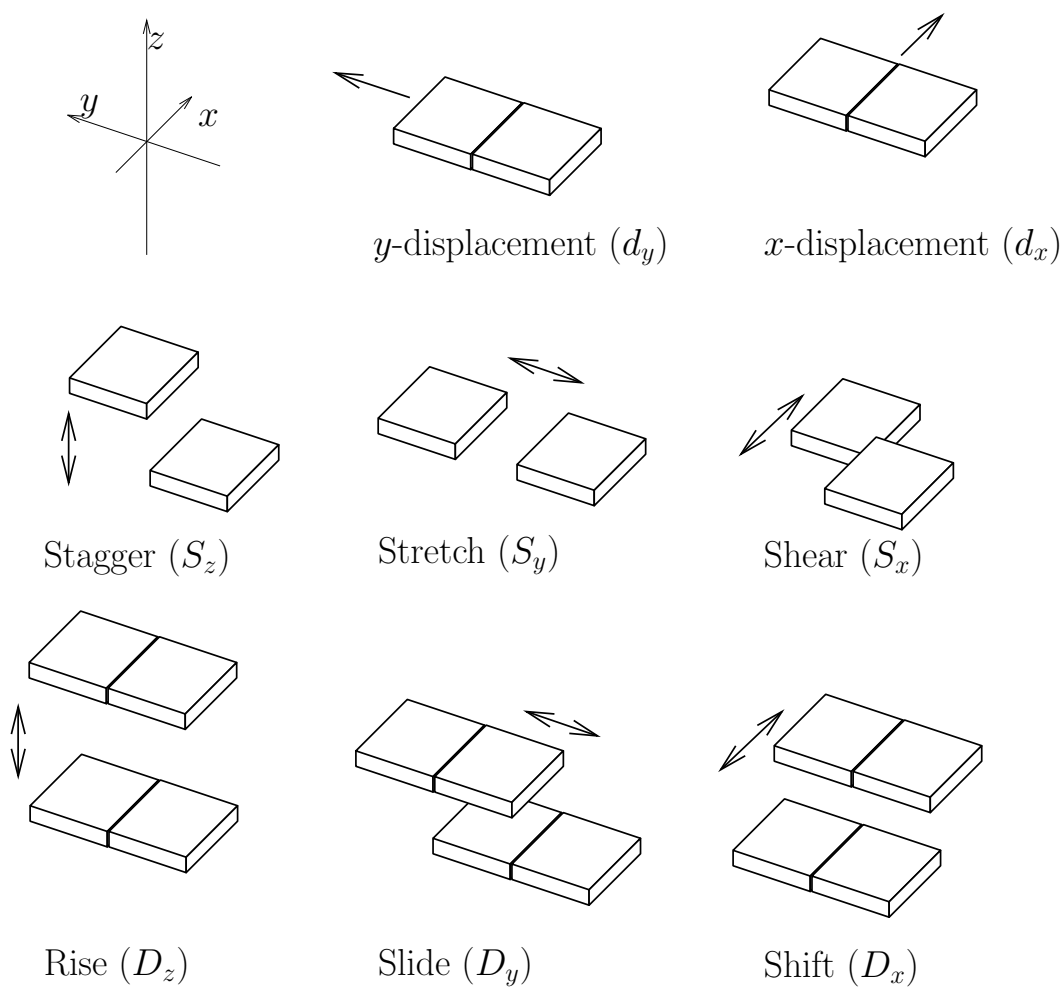
Podobně jako u proteinů se konformace DNA neomezují na popsané typické sekundární struktury. V závislosti na sekvenci a fyzikálních podmínkách se setkáváme s mnoha odlišnými strukturami (viz například zprohýbaná levotočivá dvojrůbovice Z u sekvencí s opakující se dvojicí CG).

### 3.5 Terciární struktura

DNA se jako nosič genetické informace nejčastěji vyskytuje ve formě dvojrůbovice B (i když se za určitých fyziologických podmínek a na určitých místech můžeme setkat s čtyřrůbovicovými, větvenými a jinými strukturami).

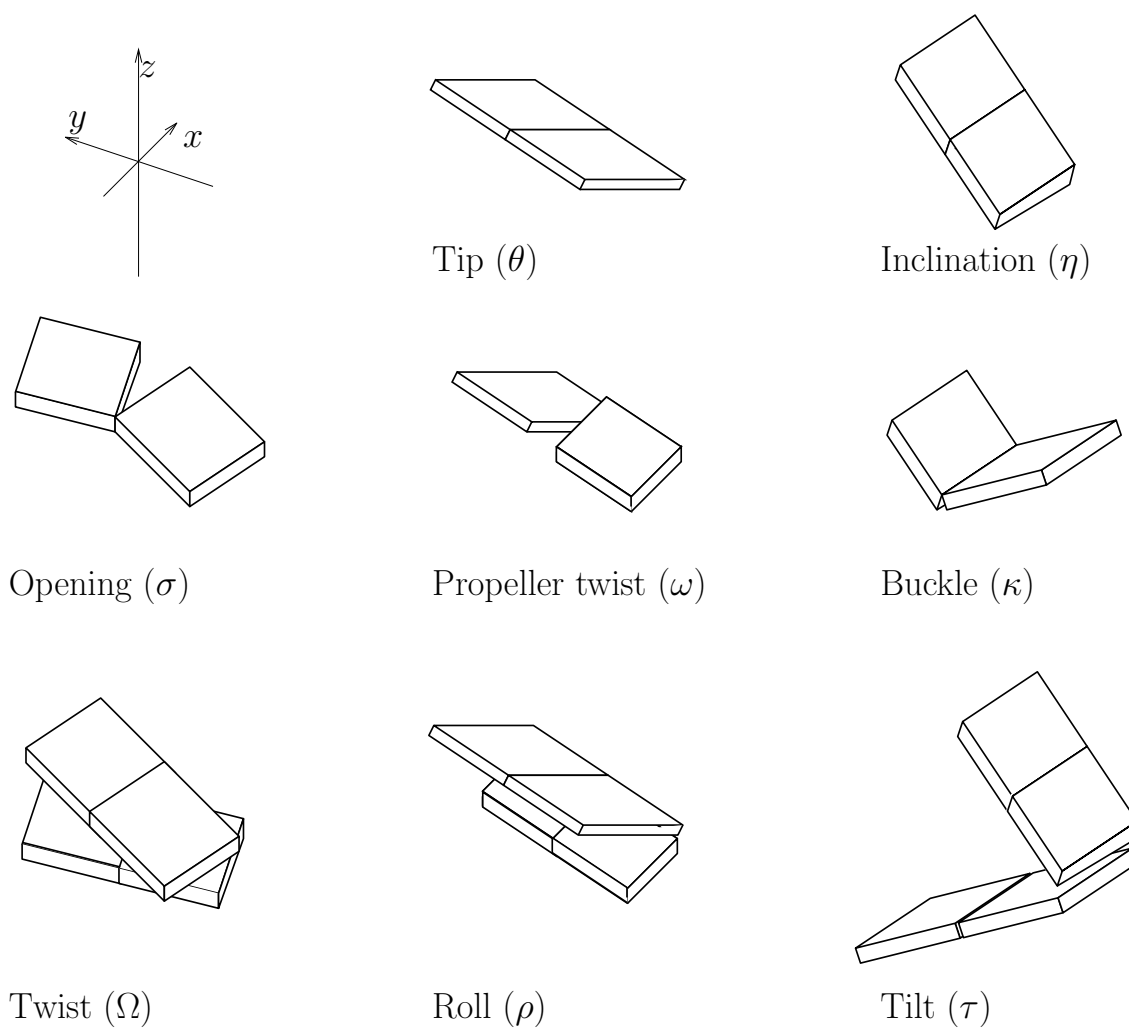
Prostorové uspořádání RNA je naproti tomu pestřejší a připomíná daleko více terciární struktury proteinů. Molekuly RNA jsou tvořeny úseky dvojrůbovic A (nazývaných *stems*, stonky) spojenými různými spojovacími prvky.

Stejně jako v proteinových ohybech se u spojovacích prvků nukleových kyselin setkáváme s odlišnými hodnotami torzních úhlů a odlišnou tvorbou vodíkových vazeb. Za zmínku stojí, že se párují i báze vyskytující se mimo pravidelné dvojrůbovice. Na rozdíl od *kanonických* párů, popsaných v sekci 3.4, je zde mnoho možností, jak síť vodíkových vazeb vytvořit. Mimo dvojrůbovice nejsou tvary párů limitovány pravidelnou strukturou, takže guanin se může dobře párovat i s uracilem, adeninem či jiným glycinem. Vodíkové vazby se tvoří nejen mezi dvěma Watson-Crickovými hranami, ale možné jsou všechny kombinace tří hran (Watson-Crickova, Hoogsteenova, sacharidová). Glykosidické vazby navíc nemusí být v poloze *cis*, ale možná je i poloha *trans*, kdy glykosidické vazby míří opačným směrem. Celkově je tak k dispozici 12 typů párování (obrázek 3.8), příklady některých nekanonických párů jsou uvedeny na obrázku 3.9. V každém páru je využita jen jedna hrana jednotlivých bází, takže báze párovaná jedním způsobem (například podle Watsona a Cricka) může stále interagovat další hranou (například Hoogsteenovou). Takovým způsobem vznikají *triplety* bází, které mohou spojovat dvojrůbovice s dalším vláknem, aniž by se výrazně měnila pravidelná dvojrůbovicová struktura.

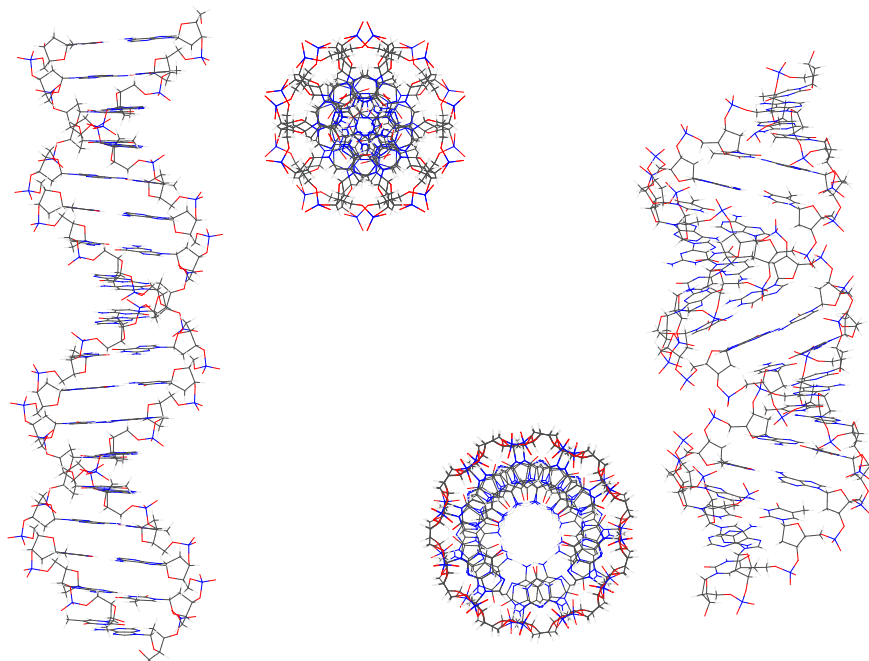


Obrázek 3.5: Translační helikální parametry. V levém horním rohu je znázorněna souřadná soustava, ve které se helikální parametry udávají. Parametry v horní řadě udávají posunutí párů bází jako celků (souběžná posunutí), parametry ve střední řadě udávají vzájemné posunutí bází v jednotlivých párech vůči sobě (protisměrná posunutí), parametry v dolní řadě udávají vzájemná posunutí sousedních párů bází. Posunutí ve směru osy  $x$ ,  $y$ ,  $z$  jsou seřazena zprava doleva.

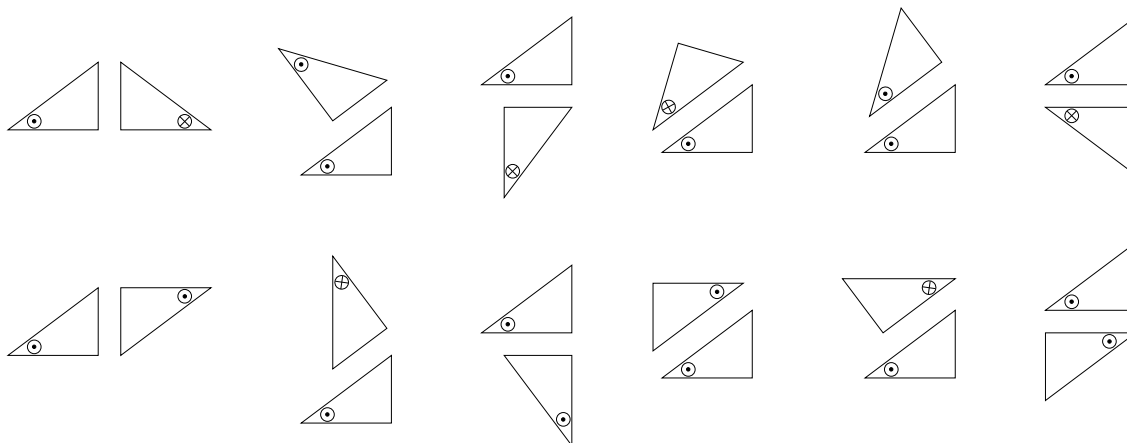




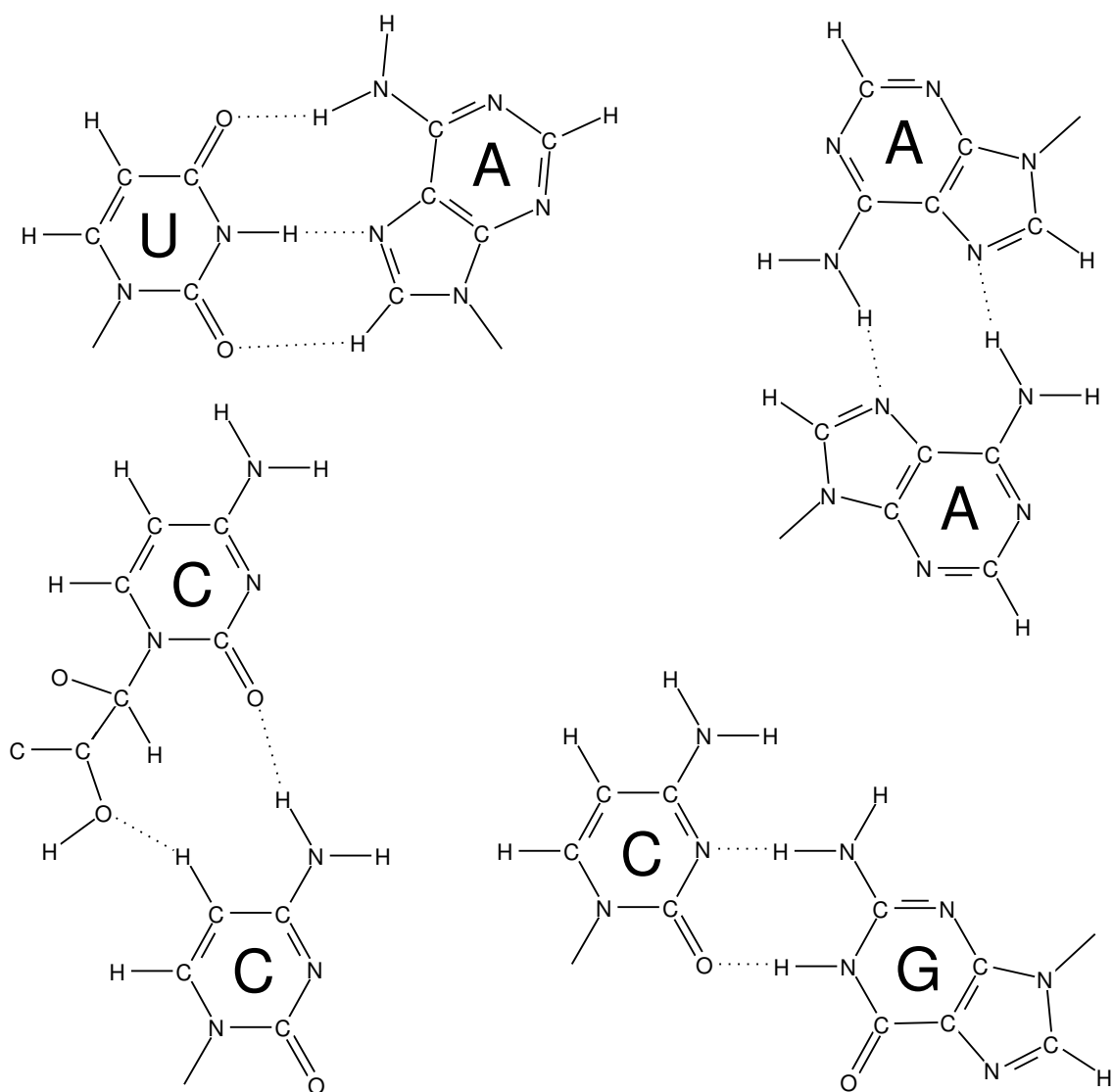
Obrázek 3.6: Rotační helikální parametry. V levém horním rohu je znázorněna souřadná soustava, ve které se helikální parametry udávají. Parametry v horní řadě udávají natočení párů bází jako celků (souběžné rotace), parametry ve střední řadě udávají vzájemná natočení bází v jednotlivých párech vůči sobě (protisměrné rotace), parametry v dolní řadě udávají vzájemná natočení sousedních párů bází. Rotace kolem osy  $x$ ,  $y$ ,  $z$  jsou seřazeny zprava doleva.



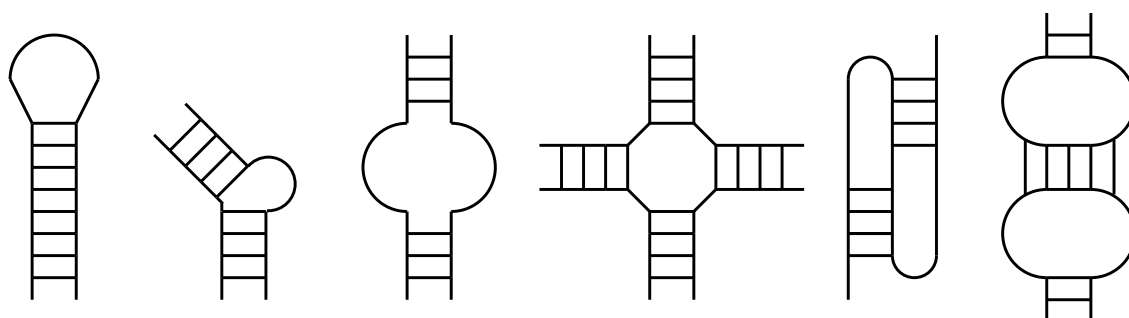
Obrázek 3.7: Modely DNA šroubovic B a A.



Obrázek 3.8: Možné způsoby párování bází. Báze jsou znázorněny jako trojúhelníky s udáním směru hlavního řetězce ( $\odot$  směřuje dopředu z roviny nákresu,  $\otimes$  směřuje dozadu za rovinu nákresu). Hrany jsou rozlišeny délkou stran trojúhelníku – nejkratší Watson-Crickova, střední sacharidová, nejdelší Hoogsteenova nebo „C-H“. Nahoře jsou znázorněna párování s *cis* uspořádáním glykosidických vazeb, dole s *trans* uspořádáním glykosidických vazeb.



Obrázek 3.9: Příklady nekanonických párů bází: vlevo nahoře *cis* párování uracilu (Watson-Crickovou hranou) s adeninem (Hoogsteenovou hranou), vpravo nahoře *trans* párování dvou adeninů Hoogsteenovými hranami, vlevo dole *cis* párování dvou cytosinů sacharidovou a „C-H“ („Hoogsteenovou“) hranou, vpravo dole *trans* párování cytosinu a guaninu Watson-Crickovými hranami.



Obrázek 3.10: Příklady strukturních motivů nukleových kyselin. Zleva vláseňková smyčka (*hairpin loops*), zlom s včleněným nepárováným nukleotidem *bulge*, *internal loop*, *four-way junction*, pseudouzle *pseudoknot*, *kissing hairpin*.

Pravidelné dvojšroubovice a struktury zmíněné v předchozím odstavci mohou tvořit supersekundární motivy podobné proteinovým (obrázek 3.10), jako vláseňkové smyčky (*hairpin loops*) a zlomy (*kinks*) v dvojšroubovici. Supersekundární struktury typické pro nukleové kyseliny jsou dále různé vnitřní smyčky (*internal loops*) a větvení (*junctions*).

Z vyšších strukturních prvků můžeme jmenovat trojšroubovicová uspořádání, kvadruplexy (útvary tvořené čtyřmi vlákny), pseudouzly (*pseudoknots*), neobvykle vzájemně párované smyčky romanticky označované *kissing hairpins* (obrázek 3.10).

Domény RNA jsou definovány stejně jako v případě proteinů. Například velká podjednotka bakteriální ribosomové RNA je tvořena šesti doménami, z nichž každá má specifickou funkci a samostatnou strukturu.

### 3.6 Kvartérní struktura a vyšší struktury

Pojem kvartérní struktury můžeme převzít tak, jak jsme si jej definovali pro proteiny. Nukleové kyseliny často tvoří funkční oligomerní komplexy s proteiny. Podobně jako proteiny se ribonukleové kyseliny podílí na stavbě nadmolekulárních útvarů (opět můžeme jmenovat ribosom).

Pravidelné šroubovice DNA jsou dále svinuty do nadšroubovicových útvarů až po úroveň chromozomů<sup>6</sup>. Zde je ovšem nutné poznamenat, že nadšroubovice jsou útvary lišící se nejen konformací, ale i konfigurací. K jejich vytvoření a zrušení je totiž nutno přerušit alespoň jedno vlákno. Nadšroubovice se obvykle popisují dvěma čísly,  $L$  (*linking number*) definuje topologii nadšroubovice (kolikrát jsou jednotlivá vlákna překřížena) a  $T$  (*twist*) definuje geometrii (zkroucení dvojšroubovice).

<sup>6</sup> Jednotlivá vlákna musí být kruhová nebo pevně uchycená například k proteinům, jinak by se nadšroubovice samovolně rozpletla.

## Kapitola 4

# Struktura oligo- a polysacharidů

### 4.1 Stavební jednotky

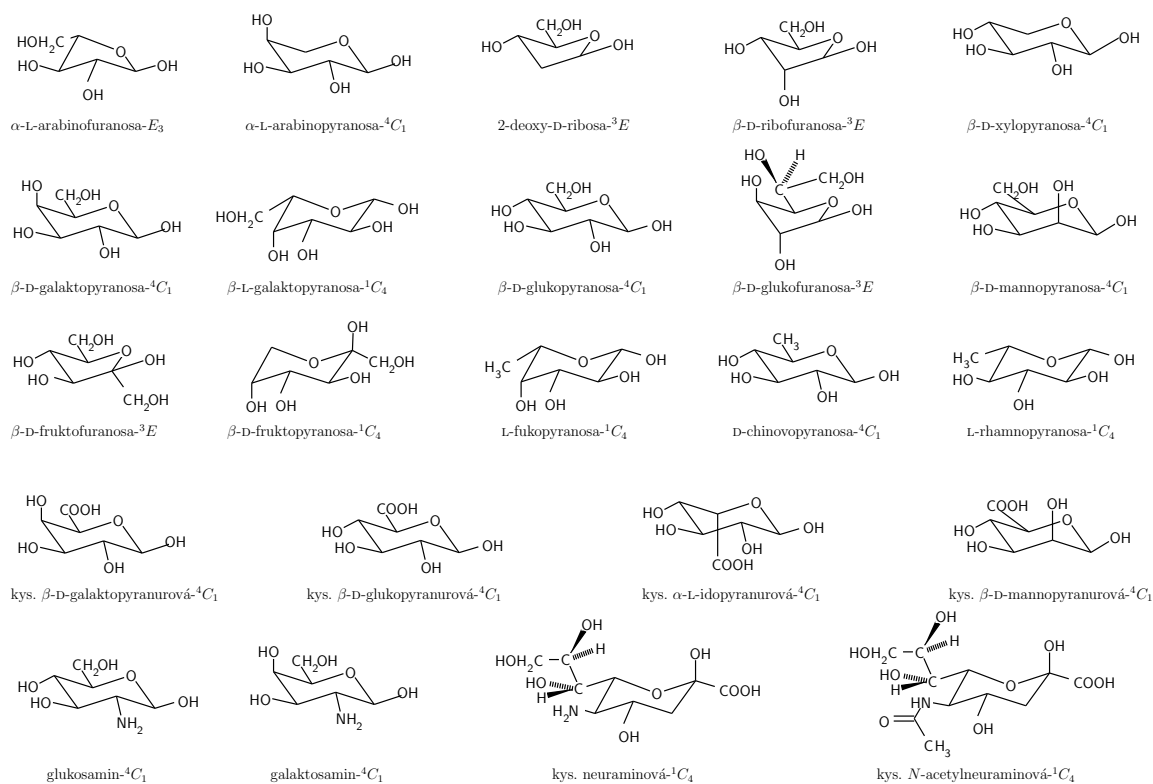
Nejnápadnějším rysem, který bude naši diskusi struktur oligosacharidů a polysacharidů provázet, je neobyčejná pestrost. Mnohem větší strukturní různorodost sacharidů ve srovnání s proteiny a nukleovými kyselinami se u oligo- a polysacharidů objevuje již na úrovni monomerních jednotek. Základní stavební jednotky se liší počtem uhlíků (nejčastější jsou hexosy a pentosy), funkční skupinou (aldosy a ketosy), geometrickou isomerií hydroxylů ( $2^n$  optických isomerů, kde  $n$  je počet chirálních uhlíků), možnou substitucí (sulfátem, karboxylem, aminem, *N*-acetylaminem), počtem uhlíků v kruhu (furanosy a pyranosy), konformací cukerného kruhu, polohou a počtem hydroxylů, které tvoří glykosidickou vazbu se sousedním reziduem. Nejběžnější monosacharidové jednotky jsou uvedeny na obrázku 4.1. Názvosloví sacharidů se řídí pokyny IUPAC-IUB.

### 4.2 Torzní úhly

Monosacharidové jednotky se vyskytují nejčastěji ve formě pětičlenných a šestičlenných kruhů. Sacharidy s pětičlenným cyklem se nazývají *furanosy*, zatímco sacharidy s šestičlenným cyklem se označují jako *pyranosy*.

Možné konformace furanos byly již probrány u nukleových kyselin. K popisu konformace šestičlenných kruhů potřebujeme tři torzní úhly. Nejběžnější konformace šestičlenných cyklů jsou takové, ve kterých vystupují dva protilehlé atomy z roviny, ve které leží zbývající čtyři atomy. Většinou je nejvýhodnější židličková konformace, ve které se pyranosy vyskytují nejčastěji. V židličkové konformaci jeden z protilehlých atomů leží nad rovinou a druhý pod rovinou. Židličkou můžeme proložit rovinu ještě jinak – tak, aby střídavě jeden z atomů ležel nad rovinou a druhý pod rovinou. Vidíme, že takto můžeme sestavit jen dvě různé židličky (v první židličce leží jedna trojice atomů nad rovinou a druhá pod rovinou, v druhé židličce se polohy trojic vymění). Proto existují jen dvě židličkové konformace, které se označují  ${}^4C_1$  nebo  ${}^1C_4$  (horní index udává číslo atomu nad rovinou, dolní index udává číslo atomů pod rovinou kruhu orientovaného tak, jak ukazuje obrázek 4.1). Která z těchto konformací je energeticky výhodnější, je dáno tím, jaké substituenty jsou vázány na uhlících kruhu. Každý uhlík může vázat kromě svých sousedů v kruhu ještě dva další atomy. Jedna z těchto vazeb míří přibližně kolmo k rovině proložené pyranosovým kruhem (tato poloha se označuje *axiální*), druhá leží zhruba v této rovině (poloha *ekvatoriální*). Ze sterických důvodů jsou nejstabilnější konformace s maximálním počtem objemných substituentů v ekvatoriální poloze ( ${}^4C_1$  pro D-glukosu).

Pokud oba protilehlé atomy leží nad rovinou kruhu, mluvíme o vaničkové konformaci. Takových vaniček můžeme sestavit šest:  ${}^1,4B$ ,  ${}^2,5B$ ,  ${}^3,0B$ ,  ${}^1,4B$ ,  ${}^2,5B$ ,  ${}^3,0B$  (horní indexy opět značí atomy nad



Obrázek 4.1: Vzorce nejčastějších monosacharidových jednotek oligo- a polysacharidů. Ze dvou možných anomerů  $\alpha$  a  $\beta$  je vždy uveden jen jeden.

rovinou a dolní indexy atomy pod rovinou). Popis jednotlivých konformací<sup>1</sup> se řídí pravidly IUPAC-IUB.

Stabilita pyranosových konformerů je tak velká, že je lze považovat za rigidní strukturní jednotky. Konformaci polysacharidového řetězce lze proto zpravidla popsat dvěma ( $\phi$ ,  $\psi$ ) nebo třemi ( $\phi$ ,  $\psi$ ,  $\omega$ ) torzními úhly. Úhel  $\phi$  definuje torzi kolem glykosidické vazby. Úhel  $\omega$  je nezbytný pokud glykosidickou vazbu tvořící hydroxyl není vázaný přímo na uhlík pyranosového nebo furanosového kruhu, ale na methylenový uhlík (například u (1 → 6) vázaných aldopyranos nebo (1 → 5) vázaných ketofuranos). Definice torzních úhlů se řídí pravidly IUPAC-IUB.

### 4.3 Primární struktura

Vzhledem k možnosti větvení a k zmíněné pestrosti monomerních jednotek je definice primární struktury oligo- a polysacharidů složitější než v případě proteinů a nukleových kyselin. Pravidla pro zápis primární struktury oligosacharidů lze nalézt v doporučeních IUPAC-IUC.

Další důležitou odlišností od proteinů a nukleových kyselin je skutečnost, že primární strukturu oligo- a polysacharidů nelze přímo vyčíst z genetického kódu. Určování primární struktury je proto v případě sacharidů mnohem obtížnější, než v případě proteinů a nukleových kyselin.

### 4.4 Sekundární struktura

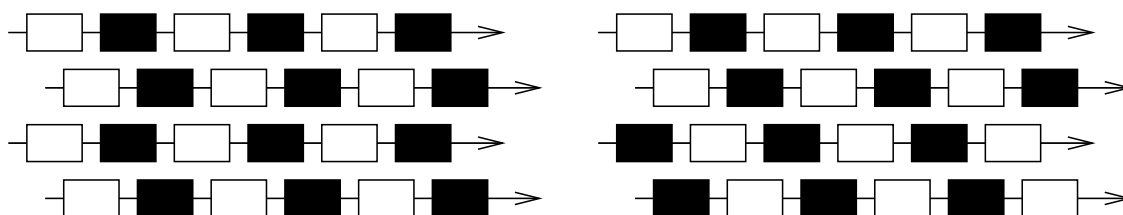
Pestrost sekundárních struktur polysacharidů ostře kontrastuje s dvojicí typických strukturních motivů proteinů a nukleových kyselin. Analogie s popisem sekundárních struktur proteinů jsou ale zřejmé. Stejně jako u proteinů, můžeme sekundární struktury polysacharidů formálně považovat za šroubovice. K jejich popisu bývají používána dvě čísla: počet monosacharidových jednotek na jeden závit  $n$  a stoupání šroubovice na jednu monosacharidovou jednotku  $h$  (nebo stoupání šroubovice na jeden závit  $P$ ). Sekundární struktury jsou opět stabilizovány vodíkovými můstky mezi závity téže šroubovice (jako u  $\alpha$ -helixu proteinů) nebo spojující dvě polysacharidová vlákna (jako v  $\beta$ -listech proteinů). Sekundární struktury polysacharidů můžeme dělit na šroubovicově vinuté (například amylosa, viz model, analogie s  $\alpha$ -šroubovicí proteinů) na natažené (formálně šroubovice, ve které na jeden závit případnou dvě rezidua, například celuloza, viz model, analogie s  $\beta$ -listy proteinů). Některé struktury polysacharidů nelze zařadit do těchto dvou kategorií, jako například velmi flexibilní dextrany vázané (1 → 6) glykosidickou vazbou nebo naopak těsně skládané bakteriální (1 → 2) glukany.

### 4.5 Terciární struktura

Sacharidy mohou tvořit velmi dlouhé řetězce, často s širokou hmotnostní distribucí. Díky větvení mohou vznikat polysacharidy přibližně kulového tvaru. Naopak přesně definované oligosacharidové struktury plní důležitou úlohu v molekulovém rozeznávání. Na rozdíl od proteinů jsou terciární struktury polysacharidů méně prozkoumány.

Příkladem architektury polysacharidů může být celuloza, základní složka buněčných stěn rostlin. Sekundárním motivem celulosy je natažený řetězec  $\beta$ -D-glukopyranos-<sup>4</sup>C<sub>1</sub> spojených (1 → 4) glykosidickou vazbou. Jde tedy formálně o šroubovici s dvěma  $\beta$ -D-glukopyranosovými jednotkami na jeden „závit“ jinými slovy, <sup>4</sup>C<sub>1</sub>-židličky  $\beta$ -D-glukopyranos se střídají tak, že kyslík pyranosového kruhu leží

<sup>1</sup>Kromě popsanych konformací, ve kterých jsou atomy ležící mimo rovinu odděleny dvěma atomy kruhu, vyskytují se vzácně i konformace, kde leží mimo rovinu atomy oddělené jen jedním atomem (šikmé židličky,  $S$ ) nebo sousední atomy (položidličky,  $H$ ). Podobně jako v případě furanos lze pyranosové konformace popsat pomocí pseudorotace. Vzhledem k tomu, že šestičlenné cykly vyžadují tři nezávislé parametry, nevystačíme s pseudorotačním kruhem, ale musíme definovat pseudorotační kouli.



Obrázek 4.2: Uspořádání řetězců v celuloze  $I_\alpha$  (vlevo) a  $I_\beta$  (vpravo). Střídající se  $\beta$ -D-glukopyranosové kruhy s kyslíkem nad a pod rovinou řetězce jsou znázorněny jako plné a prázdné obdélníky.

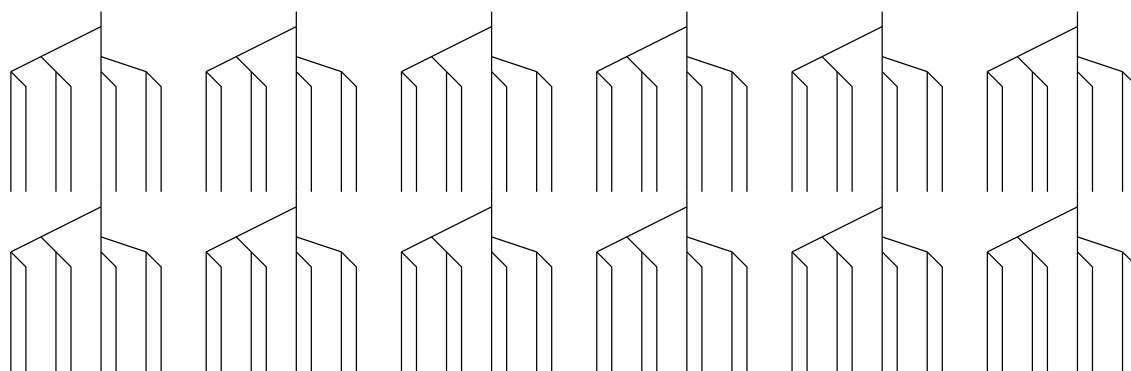
jednou nad rovinou a podruhé pod rovinou řetězce. V tomto uspořádání leží všechny hydroxyly v ekvatoriální poloze (v rovině kruhu). Vytváří se tak polární prostředí po bocích celulosového řetězce a nepolární prostředí nad a pod rovinou řetězce. Natažené řetězce se k sobě skládají nepolárními plochami a tvoří krystalovou strukturu označovanou I. Kromě hydrofobních interakcí je tato struktura stabilizována sítí vodíkových vazeb v rámci jednoho řetězce i mezi sousedními řetězci. Je zajímavé, že glykosidická vazba celulosy I je v konformaci *trans-gauche* a řetězce směřují stejným směrem (paralelní uspořádání), ačkoli energeticky výhodnější je *gauche-trans* konformace glykosidické vazby a střídavá (antiparalelní) orientace řetězců. Přirozená forma celulosy je tedy metastabilní<sup>2</sup>, na energeticky nejvýhodnější formu, označovanou II, přechází až umělým zpracováním. Celulosa I se vyskytuje ve dvou formách  $I_\alpha$  a  $I_\beta$ , které se liší vzájemným uspořádáním řetězců (obrázek 4.2). Formy  $I_\alpha$  a  $I_\beta$  se mohou střídat i v rámci jednoho celulosového vlákna, jehož délka dosahuje do 10 000  $\beta$ -D-glukopyranosových jednotek.

Jiným příkladem polysacharidu je zásobní látka rostlin, škrob. Základní stavební jednotky škrobu jsou  $\alpha$ -D-glukopyranosy-<sup>4</sup>C<sub>1</sub>, spojené (1 → 4) glykosidickou vazbou. Řetězce  $\alpha$ -D-glukopyranos mohou tvořit různé šroubovicové struktury, z nichž nejdůležitější je paralelní dvojšroubovice s výškou závitů 0,5 nm a šesti až sedmi jednotkami na jeden závit. Stejně jako v případě celulosy je důležité uspořádání hydroxylů, které směřují na povrch šroubovic, takže vnitřek šroubovic je nepolární. Škrob se skládá z nevětvených řetězců amylosy a obřích molekul amylopektinu, které jsou tvořeny přibližně milionem  $\alpha$ -D-glukopyranosových jednotek, a které obsahují i větvení prostřednictvím (1 → 6) glykosidické vazby. Větvení nijak nevadí dvojšroubovicové struktuře, naopak ji spíše stabilizuje. Dvojšroubovicové úseky mezi větveními jsou základem vrstevnaté struktury škrobových zrn (obrázek 4.3). V těchto vrstvách jsou dvojšroubovice pravidelně uspořádány jako v krystalech. Řetězce amylosy jsou nejspíše náhodně rozptýleny jako jednoduché šroubovice mezi amylopektinovými řetězci (zvláště na povrchu škrobových zrn), částečně se mohou také podílet na tvorbě dvojšroubovic.

Stejně jako jiné biomakromolekuly tvoří oligo- a polysacharidy funkční komplexy s proteiny a podílejí se na tvorbě vyšších struktur buněčné architektury. Je také důležité si uvědomit, že velkou skupinou biomakromolekul jsou glykoproteiny, jejichž oligosacharidová část je co do velikosti, důležitosti a stability konformace srovnatelná s částí proteinovou.

<sup>2</sup>Přítomnost metastabilní celulosy I je vysvětlována tím, že celulosová vlákna se tvoří krátce po biosyntéze, během které rostou celulosové řetězce jedním směrem, a na přeskládání do výhodnějšího antiparalelního uspořádání není dost času.





Obrázek 4.3: Schematické znázornění větvení amylopektinu. Rovnoběžné čáry představují paralelní dvojšroubovice.



## Kapitola 5

# Struktura biologických membrán

### 5.1 Membrány a biomakromolekuly

Biologické membrány se od biomakromolekul (proteinů, nukleových kyselin, polysacharidů) zásadně liší. Zatímco biomakromolekuly tvoří dlouhý polymerní řetězec, v němž jsou stavební jednotky pevně spojeny kovalentní vazbou, membrány se skládají z molekul, mezi kterými působí především slabé van der Waalsovy síly. Přesto můžeme nalézt důležitý rys společný membránám i biomakromolekulám. V obou případech je dobře známá konfigurace jednotlivých stavebních jednotek a popisem struktury se myslí určení geometrického uspořádání těchto stavebních jednotek.

### 5.2 Stavební jednotky

Biologické membrány jsou tvořeny látkami, které souhrnně označujeme *lipidy*. Nejde zde o název odrážející chemické složení, ale spíše fyzikální vlastnosti. Molekuly lipidů obsahují relativně malou polární, často ionizovanou část, která má výrazně hydrofilní chování. Zbytek molekuly je nepolární, výrazně hydrofobní. Molekula lipidu je tedy „vnitřně rozpolcená“ – jedna část vyhledává polární (vodné) prostředí, druhá část nepolární prostředí. Chemická struktura lipidů je velmi rozmanitá, k nejběžnějším patří následující tři typy:

- *Glycerolfosfolipidy (fosfatidy)* – k molekule glycerolu jsou vázány esterovou vazbou dvě mastné kyseliny a jeden zbytek kyseliny fosforečné. Z mastných kyselin jsou nejběžnější kyselina palmitová (šestnáctiuhlíková nasycená kyselina), osmnáctiuhlíkové kyseliny (nasycená stearová, s jednou dvojnou vazbou v poloze 9 oleová, s dvěma dvojnými vazbami v polohách 9 a 12 linoleová) a kyselina arachidonová (dvacetihlíkatá se čtyřmi dvojnými vazbami v polohách 5, 8, 11, 14). Na kyselinu fosforečnou je často další esterovou vazbou vázán kyslík ethanolaminu, serinu, cholinu, nebo inositolu (cukerný alkohol, na který bývá vázán další jeden nebo dva zbytky kyseliny fosforečné).
- *Sfingolipidy* – k molekule dvacetihlíkatého aminodiolu sfingosinu je vázána amidovou vazbou jedna mastná kyselina a jeden zbytek kyseliny fosforečné, který může být dále esterifikován (nejčastěji cholinem).
- *Steroly* – v živočišných buňkách cholesterol

### 5.3 Geometrie lipidových útvarů

Pokud přidáme lipidy ke směsi nemísitelného polárního a nepolárního rozpouštědla (voda a olej), budou ochotně tvořit rozhraní těchto dvou rozpouštědel. Hydrofobní část lipidu bude mířit do oleje a hydrofilní do vody. Pokud přidáme lipidy pouze do vodného roztoku, budou se jejich molekuly orientovat tak, aby nepolární části byly co nejméně v kontaktu s vodou (hydrofobní efekt diskutovaný v části 6.12).

Tvar útvaru, který v důsledku hydrofobního efektu vznikne, závisí na tvaru molekuly lipidu. Pokud je nepolární část lipidu malá, podobá se jeho molekula kuželu, jehož podstavu tvoří polární část. Takovéto kužely můžeme poskládat do zhruba kulového útvaru, který bude uvnitř nepolární a zvenčí polární. Takový útvar se nazývá *micela*.

Lipidy, které jsme si vyjmenovali v sekci 5.2 mají hydrofobní a hydrofilní část přibližně stejnou, takže jejich molekuly se podobají válcům. Na rozdíl od kuželů můžeme válce skládat do plochých útvarů, které se budou orientovat stejně, jako když k sobě přiložíme dva krajíce namazaného chleba. Vznikne *dvojvrstva* tvořená lipidy, které budou orientovány hydrofobní částí proti sobě. Taková dvojvrstva je základem biologických membrán<sup>1</sup>.

### 5.4 Tekutost membrán

Voda a jiné látky se mohou vyskytovat v různých skupenstvích, která se liší uspořádáním a volností molekul v trojrozměrném prostoru. V případě biomembrán se setkáváme s dvourozměrnou analogií, mluvíme o různých fázích lipidové dvojvrstvy. Při nízkých teplotách a velkém natěsnání lipidových molekul na malé ploše se tvoří dvojrozměrné krystaly. Mají-li molekuly trochu více volnosti, dochází k tvorbě dvojrozměrného gelu, tato fáze se označuje  $L_\beta$  (obrázek 5.1). Zvýšením teploty vzniká fáze kapalného krystalu  $L_\alpha$ . V ní jsou stále molekuly v těsném kontaktu, mohou se však navzájem pohybovat v rovině dvojvrstvy, jako když se snažíte razit si cestu ulicí plnou lidí. Naproti tomu je téměř nemožné, aby se molekula ze spodní strany dvojvrstvy dostala do horní a naopak. Plynné fázi by odpovídala tak nízká koncentrace lipidů, že by molekuly spolu reagovaly jen zřídka, při náhodných srážkách. To znamená, že by vůbec nedošlo k tvorbě dvojvrstvy.

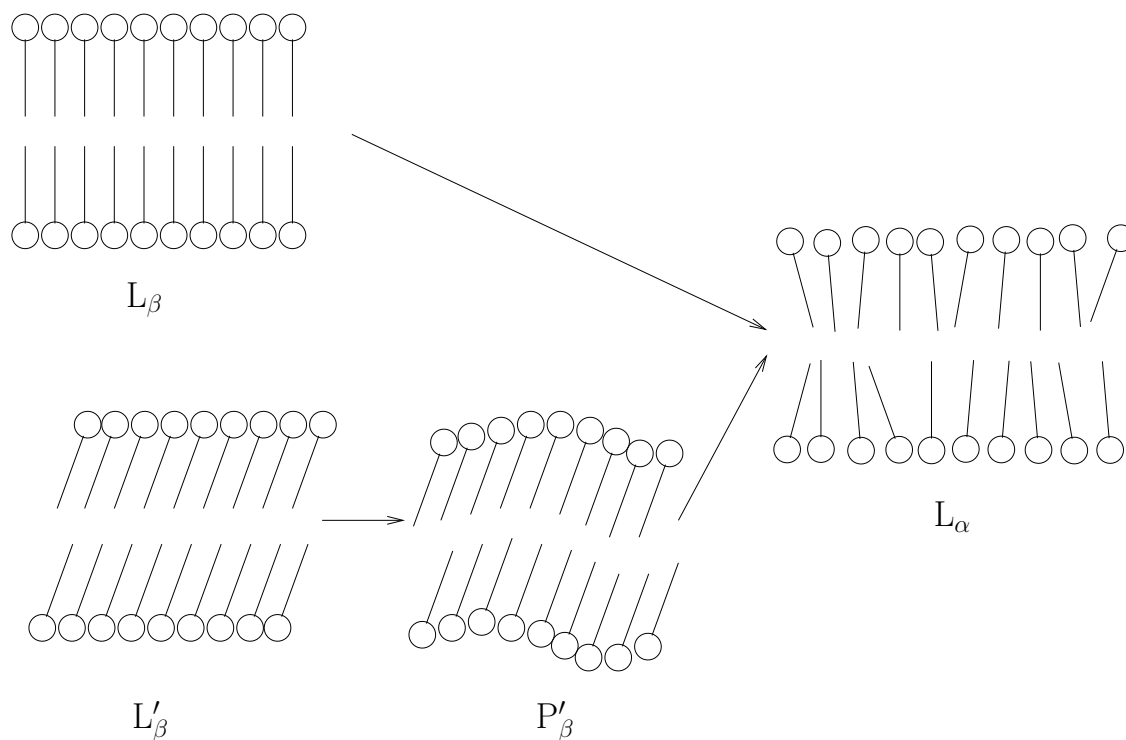
K popsáním fázovým přechodům dochází u lipidů, jejichž hydrofobní řetězce jsou orientovány kolmo k rovině dvojvrstvy. Fázové chování lipidů s šikmo orientovanými hydrofobními řetězci je o něco složitější. Při tání gelu, který se tentokrát označuje  $L'_\beta$ , dochází nejprve k vytvoření zvlněné fáze  $P'_\beta$ , ve které mají molekuly více volnosti, než v plochem gelu. Teprve dalším fázovým přechodem vzniká fáze kapalného krystalu  $L_\alpha$ .

V biomembránách se vyskytují lipidy převážně ve fázi kapalných krystalů, velmi blízko fázovému přechodu na gelovou strukturu. Mluví se často o *tekuté mozaice*, navenek tvořící jasně definovanou membránu, ve které se však mohou molekuly volně pohybovat.

### 5.5 Vliv polární skupiny

Polární část lipidu se v membránách vyskytuje v poměrně přesně definované konformaci. Tato konformace je dána požadavkem, aby molekula jako celek byla zhruba válcová a mohla tak neomezeně tvořit ploché struktury. Například v glycerolfosfolipidech obsahujících ethanolamin jsou v gelové fázi oba řetězce mastných kyselin i uhlíková páteř glycerolu orientovány kolmo k dvojvrstvě, zatímco řetězec ethanolaminu směřuje rovnoběžně s rovinou dvojvrstvy. Jde tedy o fázi  $L_\beta$ . Záměna ethanolaminu za cholin vede ke změně tvaru molekuly. Glycerol je stále orientován kolmo, ale nabitý řetězec cholinu je

<sup>1</sup>Vedle micel a dvojvrstev existuje celá řada možných uspořádání, ve kterých se budou nacházet lipidy, jejichž tvar bude něco mezi válcem a kuželem. V buněčných membránách se setkáváme s podobnou geometrií například v místě styku dvou membrán.



Obrázek 5.1: Fázové přechody lipidových dvojvrstev.

vytočen ven do rozpouštědla a řetězce mastných kyselin míří šikmo k rovině dvojvrstvy. Vzniká tedy fáze  $L'_\beta$ . Pokud obsahuje lipid volnou kyselinu fosforečnou, je polární část molekuly menší a glycerol se orientuje rovnoběžně s rovinou dvojvrstvy, aby lépe vyplnil prostor. Řetězce mastných kyselin směřují šikmo k dvojvrstvě, takže opět jde o fázi  $L'_\beta$ .

## 5.6 Vliv hydrofobního řetězce

V sekci 5.4 jsme zmínili, že tekutost membrán je dána volností pohybu jednotlivých molekul. Není proto divu, že konformace hydrofobních řetězců úzce souvisí s tekutostí biomembrány. Pokud je hydrofobní řetězec mastné kyseliny nasycený a pokud jsou všechny vazby v tomto řetězci v konformaci *trans*, zabírá tento řetězec v rovině membrány plošku o průměru 0,42 nm. Takové řetězce tedy mohou být velmi těsně poskládány a vyskytují se v tuhých gelových fázích. Rotace kolem jediné vazby do konformace *gauche(+)* vede k ohnutí řetězce, který naráz zabírá několikanásobně více místa. Takováto konformační změna tedy znamená obrovské rozvolnění struktury v daném místě. Zmíněný dramatický efekt však může být v příštím okamžiku kompenzován další konformační změnou, při které přejde vazba vzdálená o dva uhlíky od místa zlomu do konformace *gauche(-)*. Dva opačné ohyby se navzájem téměř vyruší a dvakrát zlomený řetězec bude zabírat jen asi o 10% více místa než původní.

V čem se bude lišit chování hydrofobního řetězce v případě, že jde o nenasycenou mastnou kyselinu? Pokud budou všechny vazby v konformaci (*trans*), bude zabírat mnohem více místa než nasycená. Mastné kyseliny totiž obsahují dvojně vazby v konfiguraci *cis*, které vedou k podobnému ohybu řetězce, jako konformační změny popsané v předchozím odstavci. Pokud ale bude dva uhlíky od dvojně vazby změněna konformace z *trans* na *gauche*, bude ohyb řetězce do značné míry vykompenzován.

Můžeme tedy shrnout, že změna konformace z uspořádání *trans* do polohy *gauche* může mít v různé situaci opačný účinek. V případě nasycených mastných kyselin vede ke zvýšení tekutosti, zatímco v případě nenasycených nebo již jednou konformačně deformovaných mastných kyselin vede ke zvýšení tuhosti membrány.

## 5.7 Vliv složení

Podobný účinek jako konformační změny postranního řetězce mají i změny složení membrány. Například přidavek cholesterolu do těsně uspořádané, relativně tuhé membrány tvořené nasycenými mastnými kyselinami naruší strukturu membrány a zvýší její tekutost. Naopak přidavek cholesterolu k relativně tekuté membráně obsahující značné množství nenasycených mastných kyselin vede ke zvýšení tuhosti, protože molekula cholesterolu může vyplňovat dutiny, které vznikly v důsledku přítomnosti dvojných vazeb.

## 5.8 Vyšší struktury

Tak jako se biomakromolekuly navzájem spojují a tvoří vyšší komplexy, tak i biomembrány často interagují s různými biomakromolekulami. Lipidy tak tvoří jen asi polovinu hmoty buněčných membrán. Velké množství proteinů nabývá správné konformace pouze v přítomnosti membrán. Membránové proteiny mohou lipidovou dvojvrstvou procházet, a to i několikrát, nebo k ní mohou být připojeny z jedné či druhé strany. Interakce mezi proteiny a membránou mohou být dány hydrofobním jevem, ale proteiny se mohou k lipidům vázat i kovalentními vazbami. Podobně mohou být k lipidům dvojvrstev vázány i polysacharidy. Funkce membrán není omezena jen na udržování tvaru buněk. Rozhodující část tvorby energie a regulace mnoha buněčných procesů jsou přímo spojeny s rozdělením prostoru biologickými membránami.

## Kapitola 6

# Interakce určující strukturu biomakromolekul

### 6.1 Energie a entropie

Hlavním cílem předchozích kapitol popsat strukturu (tedy konformaci) biomakromolekul. Výsledný statický geometrický model je důležitým základem pro pochopení dynamiky a funkce biologicky zajímavých molekul na atomární úrovni. V této kapitole se zamyslíme nad fyzikálními zákony, které stavby molekul určují. Tedy nad tím, proč se zkoumaná molekula vyskytuje v určité konformaci.

Hned v úvodu bychom si měli zdůraznit, že snaha o pochopení základních principů stavby molekul nás vede k velmi zjednodušenému pohledu na molekuly. A nejen nás, stejná zjednodušení používá (a často si to ani neuvědomuje) každý chemik, který kreslí na papír vzorce a montuje z kuliček a trubiček modely molekul. Prvním zjednodušením je již to, že ve vzorcích spojujeme některé atomy čárkami. Říkáme, že mezi těmito atomy existuje kovalentní vazba. Chceme zdůraznit, že mezi atomy spojenými čárkami působí síly, které zasluhují zvláštní pozornost. Jenže i mezi ostatními atomy působí síly, a to síly, které jsou popsány stejnými fyzikálními zákony. Silové působení mezi atomy, které nejsou spojeny kovalentní vazbou, může být dokonce silnější než některé kovalentní vazby.

V části 1.1 jsme na kovalentní vazbě založili rozlišování mezi konfigurací a konformací. V přírodě ale mezi konfigurací a konformací žádná jasná hranice není. Otáčení kolem některé „jednoduché“ vazby může být obtížnější, než otáčení kolem určité „dvojně“ vazby. Znamená to, že máme strukturní vzorce založené na konfiguraci a modely založené na konformaci odhodit jako zbytečné a matoucí? Určitě ne, vzorce a modely jsou velmi užitečné pomůcky, které pomáhají chápat základní pravidla a souvislosti výstavby molekul. Musíme si ale uvědomit, že pomocí nich můžeme základní principy *ilustrovat*, ne však fyzikálně *vysvětlit*. Proto v této kapitole věnujeme více pozornosti fyzikální podstatě interakcí v molekulách.

Výhodnost konformací lze popsat jedinou veličinou, *energií*. Energie má jednu zvláštnost, není možné určit její absolutní hodnotu. Nemá proto smysl ptát se, jaká je energie určité konformace. Vždy můžeme pouze porovnávat energie různých konformací mezi sebou.

Molekuly se ani v buňkách ani ve zkumavce nevyskytují v jednom exempláři, ale v obrovských počtech. Zastoupení konformací v takovém souboru molekul není dáno jen energetickou výhodností jednotlivých konformací, ale také statistikou. Můžeme si to ukázat na jednoduchém příkladu, kde budeme sledovat místo konformací různé konfigurace. Smícháme stejná látková množství dvou krátkých peptidů. Jeden bude mít sekvenci Lys-Cys-Gly (označíme si KCG podle jednopísmenných zkratk) a druhý Glu-Cys-Gly (označíme si ECG). Předpokládejme, že peptidy vytvoří dimery spojené disulfidovými můstky mezi cysteiny. Mohou tedy vzniknout homodimery KCG-KCG, ECG-ECG a heterodimer KCG-ECG.

Naším úkolem bude odhadnout, kolik kterých dimerů vznikne. Kdybychom se dívali jen na energii, tak budeme hádat, že vznikne 100 % heterodimeru, protože je energeticky výhodnější (obsahuje kladně nabitý lysin a záporně nabitý glutamát, které se budou přitahovat). Kdybychom se dívali jen na statistiku, tak budeme odhadovat 50 % heterodimeru a 25 % každého homodimeru. Skutečnost bude samozřejmě někde mezi.

Správné zastoupení konformací v souboru molekul nám poskytne *termodynamika*. Termodynamika obvykle popisuje soubor molekul ve stavu, kdy jsou jednotlivé konformace v rovnováze. V rovnováze se nevyskytuje jediná konformace s nejnižší energií. Statisticky je také výhodné mít určitou konformační pestrost. Tu popisuje termodynamika veličinou *entropie*  $S$ , úměrnou přirozenému logaritmu počtu rozlišitelných stavů  $\Omega$ . Hodnota konstanty úměrnosti závisí na jednotkách, ve kterých měříme energii. Pro jednotky  $\text{J mol}^{-1}$  je konstantou úměrnosti univerzální plynová konstanta<sup>1</sup>. Veličina, která zahrnuje oba zmíněné příspěvky, energii jednotlivých konformací i jejich pestrost, se nazývá *volná energie*. Používají se dvě definice volné energie. Obě zahrnují entropii stejným způsobem, ale liší se v zahrnutí práce spojené se změnami objemu do popisu energie. *Helmholtzova* volná energie  $A = U - TS$  zahrnuje *vnitřní energii*  $U$  vyjádřenou za předpokladu, že se objem, ve kterém jsou molekuly uzavřeny, nemůže měnit.  $T$  je termodynamická teplota. Naopak *Gibbsova* volná energie  $G = H - TS$  vychází z předpokladu, že se objem může neomezeně měnit tak, aby nedocházelo k tlakovým změnám. Místo vnitřní energie se proto používá veličina *entalpie*  $H$ , do které je zahrnuta i práce spojená s objemovými změnami. Rozdíl mezi molární vnitřní energií a molární entalpií je roven součinu tlaku a molárnímu objemu ( $H = U + pV$ ). Toto číslo je ale pro kapalnou vodu tak maličké, že v biochemii jsou hodnoty obou energií prakticky stejné. Proto je i rozdíl mezi Helmholtzovou a Gibbsovou volnou energií zanedbatelný. Protože jde o energie, můžeme spočítat pouze rozdíly  $\Delta G$  (a  $\Delta A$ ). V rovnováze dvou konformačních stavů je rozdíl jejich Gibbsových energií nulový ( $\Delta G = \Delta H - T\Delta S = \Delta H - RT \ln \Omega_1/\Omega_2 = 0$ ). Z toho vyplývá, že zastoupení jednotlivých konformerních stavů lze spočítat

$$\frac{\Omega_1}{\Omega_2} = e^{-\frac{\Delta H}{RT}}. \quad (6.1)$$

Tento vztah nám také říká, jaké hodnoty energie můžeme považovat za malé a jaké za velké. Pokud bude rozdíl  $\Delta H$  (nebo  $\Delta U$ ) výrazně menší než  $RT$ , nebude poměr konformačních stavů vůbec záviset na energii, ale bude pouze odrážet statistiku (v našem případě bude blízký jedné). Energie nižší než teplotní faktor ( $RT \approx 2,5 \text{ kJ mol}^{-1}$  při 300 K neboli 27 °C) můžeme proto zanedbat.

## 6.2 Síly působící v biomakromolekulách

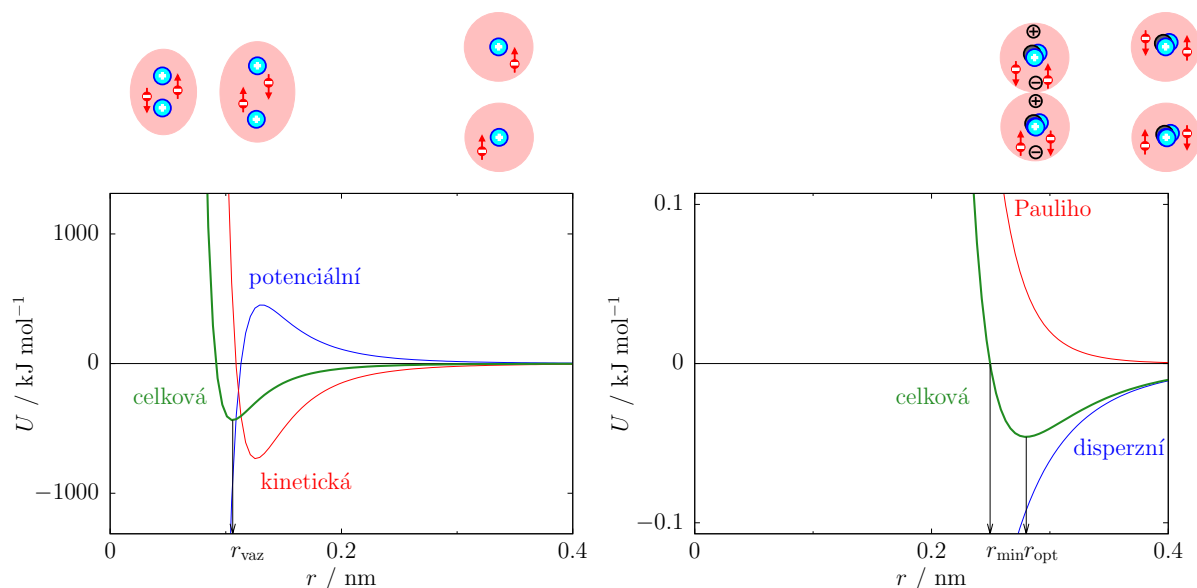
Energie související se strukturou molekul je téměř výhradně elektrického původu. Základem jejího výpočtu je Coulombův zákon, určující jakou silou na sebe působí náboje  $Q_1$  a  $Q_2$  ve vzdálenosti  $r$

$$F = \frac{1}{4\pi\epsilon_0} \frac{Q_1 Q_2}{r^2}, \quad (6.2)$$

kde  $\epsilon_0 = 8,854 \text{ C}^2 \text{ J}^{-1} \text{ m}^{-1}$  je elektrická permitivita vakua.

<sup>1</sup>Na rozdíl od energie je možné určit entropii v absolutních číslech. Například si představme, že bychom výše uvedený dimer peptidů připravili jen ze dvou molekul KCG a dvou molekul ECG. Pro přehlednost dáme molekulám KCG pořadová čísla 1 a 2 a molekulám ECG pořadová čísla 3 a 4. Existují jen dvě možnosti, jak bude vzniklá směs vypadat: buď směs homodimerů KCG-KCG + ECG-ECG nebo dvě molekuly heterodimeru KCG-ECG. Směs homodimerů může vzniknout jen jedním způsobem, když molekula číslo 1 zreaguje s molekulou číslo 2 a molekula číslo 3 zreaguje s molekulou číslo 4. Dvojice hereodimerů ale může vzniknout dvěma způsoby: když molekula číslo 1 zreaguje s molekulou číslo 3 a molekula číslo 2 zreaguje s molekulou číslo 4, nebo když molekula číslo 1 zreaguje s molekulou číslo 4 a molekula číslo 2 zreaguje s molekulou číslo 3. Homodimerní výsledek reakce tak představuje jen jeden rozlišitelný stav s nulovou entropií. Heterodimerní výsledek odpovídá dvěma rozlišitelným stavům s entropií  $S = R \ln 2 = 16,6 \text{ J mol}^{-1} \text{ K}^{-1}$ .





Obrázek 6.1: Energie atomů vodíku a kyslíku tvořících kovalentní vazbu (vlevo) a atomů helia tvořících van der Waalsovou vazbu (vpravo).

Molekuly ovšem obsahují mnoho nabitých částic (jader, elektronů). Proto je konkrétní výpočet celkové energie složitý. Z praktických důvodů se často popisují zvlášť jednotlivé typy příspěvků k celkové energii. Z těchto příspěvků si uvedeme alespoň nejdůležitější.

### 6.3 Kovalentní vazby

Kovalentní vazby určují konfiguraci molekuly. Tyto vazby popisují nejvýhodnější prostorové rozmístění elektronů z pohledu jejich vzájemných interakcí a jejich interakcí s jádry atomů. Popis kovalentních vazeb je poměrně složitý, vyžadující kvantový přístup. Zde se spokojíme s grafem popisujícím změny energie během přibližování atomů schopných tvořit kovalentní vazbu (obrázek 6.1 vlevo). Když se takové atomy přibližují, přestávají pravděpodobnost výskytu elektronů popisovat nezávislé atomové orbitály, ale elektrony začínají být ovlivněny i polem sousedního jádra. Zpočátku se tak rozšiřuje oblast jejich výskytu a klesá jejich kinetická energie<sup>2</sup>. Naopak narůstá potenciální energie kvůli interakci s elektrony sousedního atomu. S dalším přibližováním se oblast výskytu elektronů zase zmenšuje a jejich kinetická energie roste. Naopak s přibližováním k sousednímu jádru začne klesat potenciální energie elektronů. Vzdálenost, kde je celková energie nejnižší, se nazývá vazebná vzdálenost ( $r_{\text{vaz}}$  na obrázku 6.1 vlevo). Přiblížení atomů na tuto vzdálenost je spojeno s poklesem energie řádově o stovky  $\text{kJ mol}^{-1}$ .

Kovalentní vazby jsou velmi silné (viz tabulka 6.1), ale souvisí téměř výhradně s primární strukturou. Důležitou výjimkou jsou disulfidové vazby mezi cysteiny, které jsou sice kovalentní, ale určují terciární (případně kvarterní) strukturu proteinů. Jejich vliv je důležitý jen u proteinů vylučovaných ven z cytoplasmy. V cytoplasmě je totiž nadbytek peptidu glutathionu, který obsahuje cystein a tvoří disulfidové vazby s proteiny na úkor intramolekulárních disulfidových můstků v proteinech.

<sup>2</sup>Kvantová mechanika ukazuje, že není možné zároveň určit přesně polohu a hybnost elektronu (Heisenbergovy relace neurčitosti). Pokud tedy rozšíříme oblast výskytu (polohy) elektronu, můžeme zúžit rozptyl hybnosti. To vede ke snížení kinetické energie, která je úměrná druhé mocnině rychlosti a tedy i hybnosti.

Tabulka 6.1: Energie a meziatomové vzdálenosti kovalentních  $\sigma$  vazeb (červeně), disperzních interakcí (zeleně) a vodíkových vazeb (modře). Uvedené hodnoty jsou pouze přibližné a závisí na konkrétním chemickém okolí atomů.

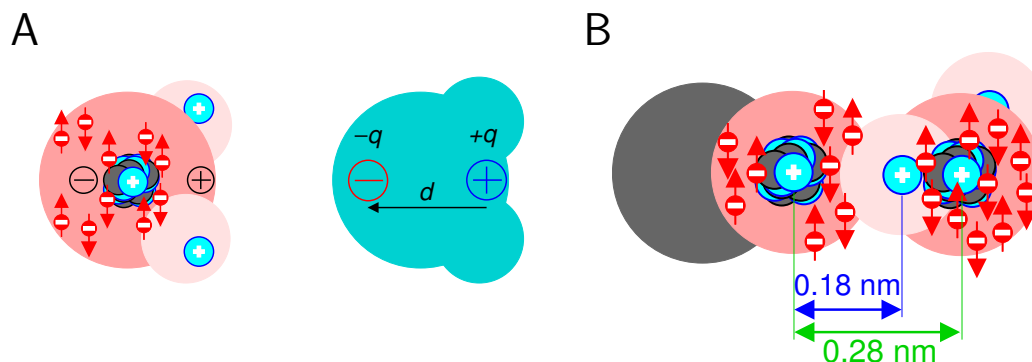
Atom...atom	$U_{\text{opt}} / \text{kJ mol}^{-1}$	$r_{\text{opt}} / \text{nm}$	$r_{\text{min}} / \text{nm}$
He...He	0,05	0,28	0,25
-H—H-	440	0,074	
-H...H-	0,50	0,24	0,20
-C...H-	0,45	0,29	0,24
-C—C-	350	0,154	
-C...C-	0,50	0,34	0,30
-N—N-	170	0,145	
-N...N-	0,85	0,31	0,27
-NH...N-	13	0,31	
-NH...O-	8	0,29	
-O—O-	150	0,148	
-O...O-	0,95	0,30	0,27
-OH...O-	21	0,28	
-OH...N-	29	0,27	

## 6.4 Van der Waalsovy interakce

Van der Waalsovy síly jsou univerzální interakce, které působí mezi polárními i nepolárními molekulami a funkčními skupinami. Pokud sledujeme změny energie během přibližování atomů *neschopných* tvořit kovalentní vazbu (obrázek 6.1 vpravo), pozorujeme pokles energie, který vzdáleně připomíná tvorbu kovalentní vazby. Optimální vzdálenosti jsou ale delší a poklesy energie výrazně menší, než pro kovalentních vazeb (v případě atomů helia na obrázku 6.1 vpravo  $0,05 \text{ kJ mol}^{-1}$ , pro atomy v biomakromolekulách  $0,5 \text{ kJ mol}^{-1}$  až  $1 \text{ kJ mol}^{-1}$ , tabulka 6.1).

Pozorovaná energie se skládá z odpudivého a přitažlivého příspěvku. Odpudivou interakcí je takzvaná *Pauliho repulze*. Při přibližování atomů *neschopných* tvořit kovalentní vazbu totiž přibližujeme elektrony jednoho atomu k již plně obsazeným orbitalům druhého atomu, což je krajně nevýhodné. Při přiblížení atomů na vzdálenost menší, než označenou  $r_{\text{min}}$ , odpudivé síly začnou převažovat a s dalším přiblížováním prudce narůstají. V chemii je nazýváme *sterické zábrany* a vymezují konformace, které se v makromolekulách vyskytnout nemohou. Například nepozorované torzní úhly a jejich kombinace v Ramachandranově diagramu.

Přitažlivou složkou van der Waalsových interakcí jsou takzvané *disperzní síly*. Jsou to interakce působící mezi elektrickými dipóly, které vznikají pouze krátkodobě v důsledku okamžitého přemístění elektronů v molekule. Pokud se elektrony ve dvou nepolárních molekulách posunou zároveň jedním směrem, sníží se energie elektrostatických interakcí mezi molekulami (ve srovnání s dvojicí stejně vzdálených molekul se symetrickým rozložením elektronů). I když pro každou dvojici interagujících atomů poklesne energie jen máličko, celkový pokles pro celou molekulu je nezanedbatelný a molekuly se proto snaží zaujímat konformace, ve kterých jsou pokud možno všechny atomy v optimální vzdálenosti k jinému atomu (včetně atomů rozpouštědla, tedy vody).



Obrázek 6.2: Schematické znázornění vody jako polární molekuly (A) a jako molekuly schopné tvořit vodíkovou vazbu (B).

## 6.5 Elektrostatické interakce

Do této kategorie budeme řadit interakce, k jejichž popisu nepotřebujeme kvantovou mechaniku (nemusíme brát do úvahy orbitaly). *Iontové interakce* představují jednoduché elektrostatické odpuzování stejně nabitých iontů a přitahování opačně nabitých iontů. Energie těchto interakcí klesá s první mocninou vzdálenosti iontů. *Dipólové interakce* jsou síly působící mezi *elektrickými dipólovými momenty*, které vznikají v molekule vytvořením nadbytku elektronů v jedné části molekuly a nedostatku elektronů (nadbytku kladného náboje) v jiné části molekuly (nebo chemické skupiny, například postranního řetězce aminokyseliny). Molekula jako celek je přitom navenek neutrální. Elektrický dipól má každá *polární vazba*, tedy kovalentní vazba mezi dvěma atomy, ve které je větší výskyt elektronů u jednoho z atomů. Na obrázku 6.2A je ukázána voda jako příklad polární molekuly. Elektrické dipóly se navzájem sčítají, takže větší části makromolekul na sebe mohou působit tímto způsobem. Energie interakcí mezi dvěma dipóly klesá s třetí mocninou jejich vzdálenosti. Protože elektrický dipólový moment má nejen velikost, ale i směr, závisí energie interakce i na vzájemné orientaci dipólů. Kromě interakcí mezi dvěma dipóly se setkáváme i s interakcemi mezi dipólem a iontem. Energie takové interakce klesá s druhou mocninou vzdálenosti iontu a dipólu a závisí na natočení dipólu vzhledem k iontu. Detailnější diskuse a odvození kvantitativních vztahů popisujících energii různých typů elektrostatických interakcí je uvedena v Dodatku A.

## 6.6 Vodíkové vazby

Vodíkové vazby neboli můstky se tvoří mezi vodíkem vázaným velmi polární vazbou na atom zvaný *donor* (nejčastěji na kyslík nebo dusík) a atomem zvaným *akceptor*, který může poskytnout elektrony pro vytvoření vazby (opět nejčastěji kyslík nebo dusík). Vliv vodíku je dobře vidět na srovnání optimální vzdálenosti dvou atomů kyslíku interagujících přímo (například dvou karbonylových skupin) s optimální vzdáleností dvou kyslíků interagujících vodíkovou vazbou. Přestože je v druhém případě mezi atomy kyslíku ještě vodík, je jejich vzájemná vzdálenost kratší (0,28 nm oproti 0,30 nm v prvním případě, tabulka 6.1 a obrázek 6.2B). Vodíkové vazby mají částečně kovalentní charakter, ale převážně jsou elektrostatické. Jejich popis je dosti složitý, protože závisí na orientaci zúčastněných dipólových momentů. Vodíkové vazby jsou slabší než vazby kovalentní (jeden vodíkový můstek snižuje energii přibližně o  $10 \text{ kJ mol}^{-1}$  až  $20 \text{ kJ mol}^{-1}$ ) a proto se nepovažují za vazby určující konfiguraci. Velmi výrazně však zvýhodňují určitou konformaci.

## 6.7 Stacking

Jako *stacking* se v anglické literatuře označují interakce aromatických aminokyselin v proteinech a zejména bázi v nukleových kyselinách. V případě benzenových kruhů jde téměř výhradně o disperzní síly. Heterocykly bázi nukleových kyselin mají také elektrické dipólové momenty, které k stackingu rovněž přispívají. Stacking je důležitý faktor stabilizující dvojšroubovice DNA, srovnatelný s párováním bázi Watson-Crickovými hranami. Interakce ploch bázi (stacking, tedy disperzní síly a dipól-dipólové interakce) snižují energii páru GC o  $70 \text{ kJ mol}^{-1}$  a energii páru AT o  $50 \text{ kJ mol}^{-1}$ , zatímco interakce hran (vodíkové vazby a dipól-dipólové interakce) snižují energii páru GC o  $120 \text{ kJ mol}^{-1}$  a energii páru AT o  $65 \text{ kJ mol}^{-1}$ .

## 6.8 Karbonylové interakce

Jako poslední příklad v našem neúplném seznamu sil v biomakromolekulách si uvedeme interakce mezi uhlíkem jedné karbonylové skupiny a kyslíkem druhé karbonylové skupiny. Energií dvou karbonylů snižuje částečně interakce jejich elektrických dipólových momentů a částečně využití antivazebného orbitalu  $\pi^*$  vazby C=O elektrony v nevazebném orbitalu  $2p_y$  kyslíku druhého karbonylu. Vhodnou orientací pro tuto interakci mají sousední karbonyly páteře proteinů v konformacích popsanych následujícími dvojicemi torzních úhlů  $(\phi, \psi)$ :  $(-60^\circ, -45^\circ)$ ,  $(-60^\circ, +135^\circ)$ ,  $(+60^\circ, +45^\circ)$ ,  $(+60^\circ, -135^\circ)$ . V prvních dvou případech jde o šroubovici  $\alpha$  a konformaci polyprolin-II (například v kolagenu), druhé dva případy jsou zrcadlové obrazy výhodné pro aminokyseliny řady D. Snížení energie není velké ( $1 \text{ kJ mol}^{-1}$  až  $2 \text{ kJ mol}^{-1}$ ), ale v součtu pro celou šroubovici není zanedbatelné. V případě konformace polyprolin-II vysvětluje, proč tuto konformaci, která není stabilizovaná vodíkovými vazbami, tvoří často flexibilní (přirozeně neuspořádané) proteiny.

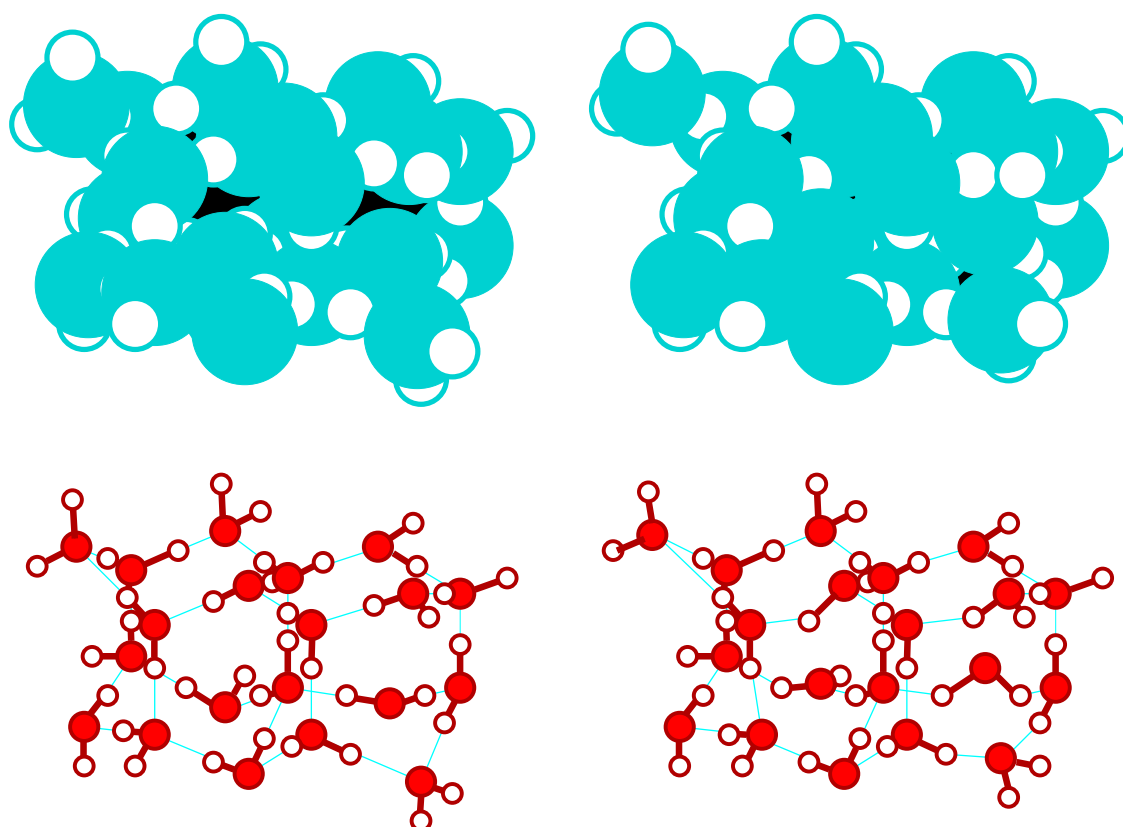
## 6.9 Voda jako rozpouštědlo

Výhodnost určité konformace je dána nejenom vzájemným energetickým působením mezi jednotlivými částmi makromolekuly, ale také interakcemi s okolními molekulami (především s molekulami rozpouštědla) a entropií celé soustavy makromolekula-rozpouštědlo. Uveďme si, jaké obecné chování můžeme očekávat u biomakromolekuly rozpuštěné ve vodném roztoku iontů solí (tedy v prostředí odpovídajícím cytoplazmě).

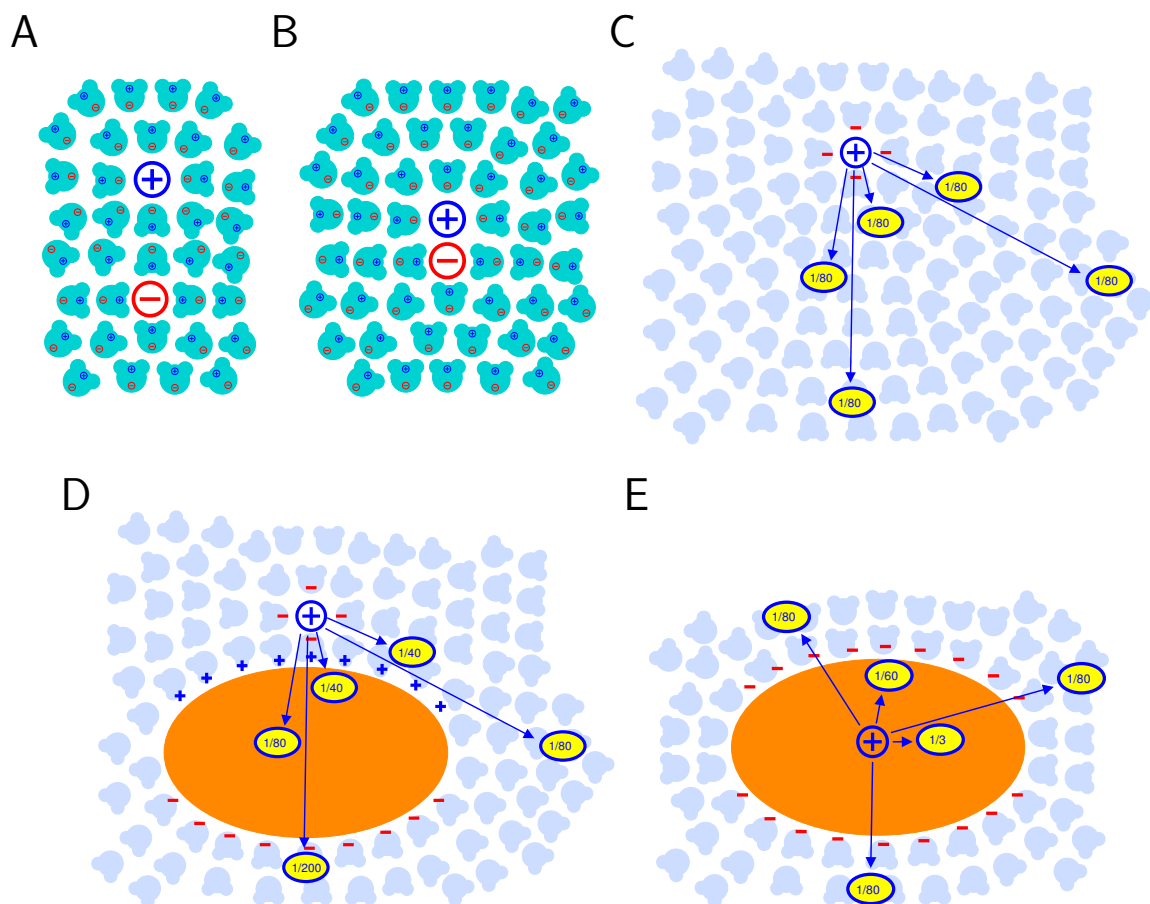
Voda je mimořádná molekula. V běžném ledu jsou molekuly ledu uspořádány v pravidelné krystalové mřížce s maximálním počtem vodíkových můstků, a to i za cenu vytvoření dutin představujících ztrátu disperzních interakcí (obrázek 6.3 vlevo). Když led roztaje, krystalová struktura se částečně zhroutí, dutiny se zaplní, ale 80 % energie vodíkových vazeb se zachová (jejich příspěvek klesne z  $-50 \text{ kJ mol}^{-1}$  na  $-40 \text{ kJ mol}^{-1}$ ). Proto si i tekutou vodu můžeme představit jako soubor molekul spojených hustou sítí vodíkových můstků. V tekuté vodě se molekuly stále otáčejí, v nové orientaci ale mohou tvořit nové vodíkové můstky s jinými molekulami vody. Vodíkové můstky se stále ruší a nové stále vznikají, ale celkový počet vodíkových vazeb je zachován bez ohledu na okamžitou orientaci jednotlivých molekul.

## 6.10 Elektrostatické interakce ve vodě

Jak ovlivní přítomnost vody elektrostatické interakce? Ionty rozpuštěné ve vodě neinteragují jenom spolu, ale také s elektrickými dipólovými momenty molekuly vody. Tyto ion-dipólové interakce představují naprostou většinu elektrostatických sil ve vzorku, interakce mezi ionty samými přispívá k celkové energii jen 1,25 %. Proto se nám zdá, že elektrostatické interakce ve vodě jsou osmdesátkrát slabší. Je



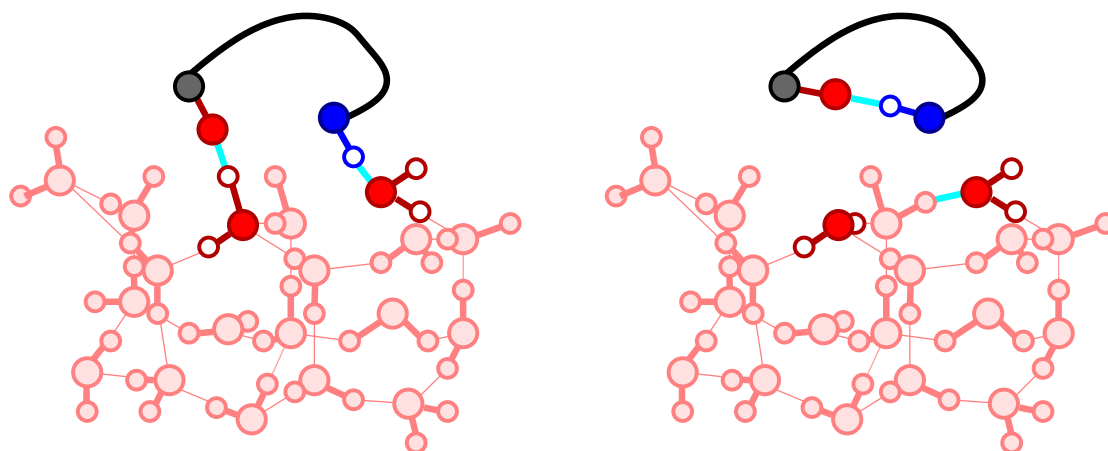
Obrázek 6.3: Struktura ledu (vlevo) a kapalné vody (vpravo). Nahoře jsou atomy kyslíku znázorněny azurově s velikostí odpovídající  $r_{\text{opt}}$  na obrázku 6.1, dole jsou pro přehlednost zmenšeny a obarveny červeně. Atomy vodíku jsou znázorněny bíle. Dutiny ve strukturách jsou v horních obrázcích vyznačeny černě. Kovalentní vazby jsou v dolních obrázcích vyznačeny hnědě, vodíkové vazby azurově.



Obrázek 6.4: A,B, orientace molekul vody kolem iontového páru. Velké symboly  $\oplus$  a  $\ominus$  představují kation a anion, malé symboly  $\oplus$  a  $\ominus$  představují parciální náboje v molekule vody. C–D, vliv orientace vody na sílu elektrostatické interakce. Symboly + a – znázorňují polarizaci v důsledku orientace vody ke kationtu. Molekula proteinu je zobrazena oranžově, čísla na žlutém pozadí představují relativní sílu interakce vzhledem k případu, kdy by kation byl ve stejné vzdálenosti ve vakuu.

třeba zdůraznit, že voda netvoří nějakou stínící bariéru mezi ionty, která by zeslabovala elektrostatické síly. K zeslabení interakcí dochází proto, že molekuly vody se v okolí nábojů mohou natočit tak, aby jejich elektrické dipólové momenty měly nejvýhodnější orientaci k iontům (obrázek 6.4A). K tomu dochází i tehdy, když jsou ionty tak blízko sebe, že se mezi ně molekula vody nevejde (obrázek 6.4B). Pokud se ionty nachází uprostřed proteinu, je elektrostatická interakce mezi nimi zeslabena mnohem méně, přibližně na čtvrtinu až polovinu (dle konkrétních aminokyselin v okolí). Menší vliv proteinu je důsledkem toho, že postranní řetězce se nemohou uvnitř proteinu natáčet tak snadno jako molekuly vody a některé z nich jsou také méně polární, než voda.

Pro zahrnutí vlivu vody se nabízí jednoduchá oprava Coulombova zákona: vynásobení konstanty  $\epsilon_0$  ve jmenovateli koeficientem 80 pro vodu a koeficientem dva až čtyři pro protein. Tento koeficient se nazývá *relativní permitivita prostředí*  $\epsilon_r$ . Tento přístup je lákavý především pro výpočetní metody, protože nijak nezpomaluje počítání. Místo jedné konstanty se prostě použije jiná. Obrázek 6.4C–D ale upozorňuje na úskalí tohoto přístupu. Zatímco ve volném roztoku lze vliv vody opravdu popsat



Obrázek 6.5: Tvorba struktur stabilizovaných vodíkovým můstkem v proteinech solvatovaných vodou. Atomy uhlíku jsou znázorněny černě, atomy dusíku modře, páteř proteinu symbolizuje černá čára. Molekuly vody, které se nezúčastňují solvatace, jsou nakresleny růžově, solvatující molekuly s použitím stejných barev jako na obrázku 6.3.

konstantou  $\epsilon_r$  (obrázek 6.4C), na povrchu proteinu jsou elektrostatické síly ovlivněny způsobem mnohem složitějším.

## 6.11 Vodíkové vazby ve vodě

Dále se podívejme, jaký vliv má prostředí vody na tvorbu vodíkových vazeb v biomakromolekulách. Ve vakuu každý vodíkový můstek mezi karbonylem a amidem páteře proteinu sníží energii o  $8 \text{ kJ mol}^{-1}$ . Pokud je ale protein solvatovaný, tvoří i v rozbaleném stavu vodíkové můstky s vodou (obrázek 6.5 vlevo). Vytvořením sekundární struktury stabilizované vodíkovým můstkem se celkový počet vodíkových vazeb nezmění, místo jedné intermolekulární vznikne jedna intramolekulární (obrázek 6.5 vpravo). Přesto má vznik vodíkového můstku v proteinu mírně stabilizující efekt.

K pochopení výhodnosti tvorby vodíkového můstku v biomakromolekule musíme prozkoumat entropické důsledky takové změny. Vytvoření vodíkové vazby mezi karbonylem v jednom místě páteře a vodíkem amidu v jiné části páteře omezí počet konformací, ve kterých se může páteř vyskytovat. Proto se entropie proteinu sníží. Tvorba vodíkového můstku v proteinu se také ale odrazí na možných orientacích molekul vody, které karbonyl a amid před vytvořením můstku solvatovaly. Místo tvorby vodíkové vazby s proteinem, která vyžadovala určitou orientaci dvou molekul vody, mohou tyto molekuly tvořit vodíkové vazby s jinými molekulami vody kolem nich. A takové vodíkové vazby se mohou vytvořit v různých orientacích. Tvorba vodíkové vazby v proteinu tak vede ke zvýšení entropie vody (zvýšení počtu stavů, ve kterých se molekula vody osvobozená ze zajetí proteinu může nacházet). Tento nárůst entropie vody převáží snížení entropie proteinu, ve kterém tvorba vodíkového můstku omezí možné konformace páteře. Z pohledu volné energie místo zisku  $-8 \text{ kJ mol}^{-1}$  vnitřní energie (nebo entalpie) proteinu získáme entropii vody. V celkové bilanci to odpovídá za pokojové teploty zisku volné energie  $\Delta G \approx -6 \text{ kJ mol}^{-1}$ . To je hodnota poměrně malá a závisující na konkrétní situaci.

## 6.12 Hydrofobní efekt

Poslední důsledek solvatace vodou, kterému věnujeme pozornost, je *hydrofobní efekt*. Dříve, než se podíváme na biomakromolekuly, prozkoumejme chování cyklohexanu, jako příkladu malé nepolární molekuly neschopné tvořit vodíkové vazby. Na obrázku 6.6A je zobrazena rovnováha mezi cyklohexanem v plynné fázi, v kapalně fázi a ve vodném roztoku. Za pokojové teploty a atmosférického tlaku preferuje cyklohexan kapalnou fázi před plynnou. Proč ale nepreferuje vodný roztok? Z pohledu entalpie (nebo vnitřní energie) je stejně výhodné pro molekulu cyklohexanu být obklopena molekulami vody nebo jinými molekulami cyklohexanu. V obou případech získáme  $-30 \text{ kJ mol}^{-1}$ . Je zřejmé, že ztrácíme nějakou entropii, protože v kapalině má cyklohexan méně volnosti, než v plynné fázi. Proč je ale ztráta entropie o tolik vyšší v případě vodného roztoku? Odpověď je naznačena na obrázku 6.6B,C.

Molekuly vody se vždy snaží orientovat tak, aby mohlo vzniknout co nejvíce vodíkových vazeb. Pokud budeme předpokládat, že si tekutá voda zachovává uspořádání ledu, můžeme nalézt šest orientací, ve kterých míří atomy vodíku vybrané molekuly vody k jedné ze sousedních molekul a mohou být donory vodíkových vazeb (obrázek 6.6B). Toto uspořádání se naruší, je-li ve vodě rozpuštěna molekula neschopná tvořit vodíkové vazby. Bude-li to molekula malá (o průměru menším než 1 nm), dokáží se molekuly vody přeskládat tak, aby celkový počet vodíkových vazeb ve vodě zůstal zachován a molekuly vody zůstaly tak blízko rozpuštěné molekuly, aby byly i výhodné disperzní interakce co nejsilnější. Cenu, kterou za to zaplatí, bude ztráta entropie. Pokud bude jedna ze sousedních molekul vody nahrazena molekulou látky, která nemůže být akceptorem vodíkové vazby (například cyklohexanem), existují jen tři orientace, ve kterých může naše být molekula vody donorem dvou vodíkových vazeb (obrázek 6.6C). Růst volné energie v důsledku reorganizace molekul vody je úměrný *objemu* rozpuštěné molekuly, nebo správněji objemu dutiny, kterou musí voda pro tuto molekulu vytvořit.

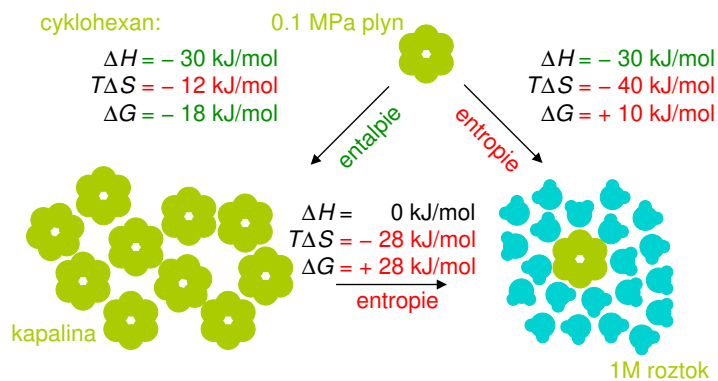
Když budeme zkoumat změny volné energie při rozpouštění nepolární molekuly větší než 1 nm, zjistíme něco jiného. Nebudeme pozorovat ztrátu entropie, ale naopak ztrátu výhodných interakcí, která se projeví zvýšením vnitřní energie. V okolí velké molekuly se již voda nedokáže přeskládat tak, aby nepřišla o některé vodíkové můstky. Ba ani nedokáže obklopit rozpuštěnou látku tak těsně, aby byly disperzní síly tak silné, jako u malé molekuly. Toto zvýšení energie nebude úměrné objemu, ale *povrchu* rozpuštěné molekuly. Změny volné energie při solvataci také závisí na tvaru solvatovaného povrchu. Voda v blízkosti plochého povrchu může vytvářet tři vodíkové vazby ze čtyř možných, zatímco molekula vody uzavřená v nepolární dutině nemůže tvořit žádnou vodíkovou vazbu.

Jakou roli hraje hydrofobní efekt ve stabilizaci nativní struktury proteinu? Pro získání zjednodušené představy můžeme vyjít z toho, co jsme si řekli o solvataci cyklohexanu. Na obrázku 6.7 jsou vlevo znázorněny čtyři hydrofobní postranní řetězce rozbaleného proteinu. Každý řetězec solvatuje osm molekul vody, které jsou tím omezeny v pohybu. Pro jednoduchost si představme, že hydrofobní postranní řetězce se chovají jako cyklohexan. Poskládáním do nativní struktury, ve které postranní řetězce interagují spolu (na obrázku 6.7 vpravo), se sníží počet molekul vody potřebných k solvataci. Uvolněné molekuly se tak stanou součástí struktury vody, ve které mohou tvořit vodíkové vazby, aniž by jejich orientace byly omezeny hydrofobní molekulou. V důsledku toho se zvýší entropie celého systému. Za nativních podmínek je tento vzrůst entropie je větší, než entropická ztráta v důsledku omezení možných konformací postranních řetězců. Celkové snížení volné energie závisí na velikosti postranního řetězce ( $-2,5 \text{ kJ mol}^{-1}$  pro alanin,  $-8 \text{ kJ mol}^{-1}$  pro leucin,  $-12 \text{ kJ mol}^{-1}$  pro fenylalanin). I když hodnoty pro jednotlivé aminokyseliny nejsou velké, vytvoření hydrofobního jádra představuje největší příspěvek ke snížení volné energie při sbalení proteinu do nativní struktury.

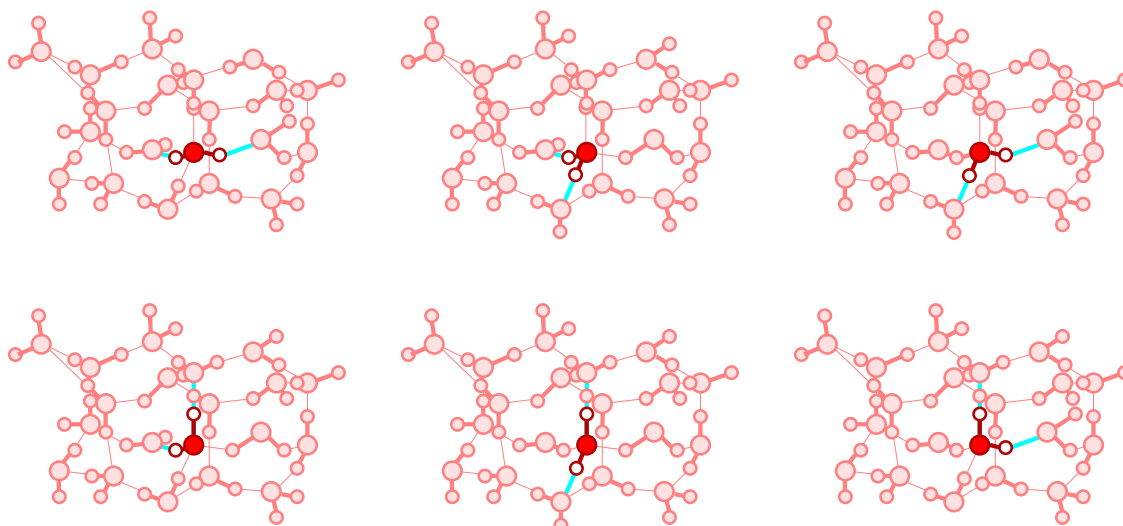
Na závěr je nutno dodat, že naše vysvětlení skládání nativní struktury proteinu je velmi zjednodušené. Jádro proteinu není tvořeno jen nepolárními postranními řetězci a ke stabilitě proteinů přispívají další vlivy. Navíc sama fyzikální podstata hydrofobního efektu není dosud plně vysvětlena. Experimenty ale přesvědčivě ukazují, že hydrofobní efekt je nejdůležitějším příspěvkem ke snížení volné energie proteinu sbaleného do pravidelné struktury a reorganizace molekul vody hraje při tom důležitou roli.



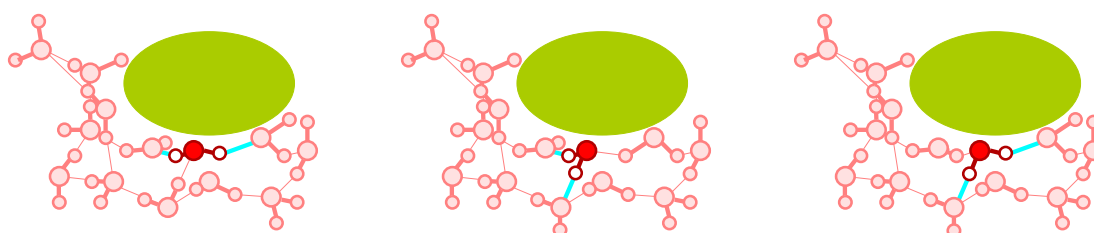
A



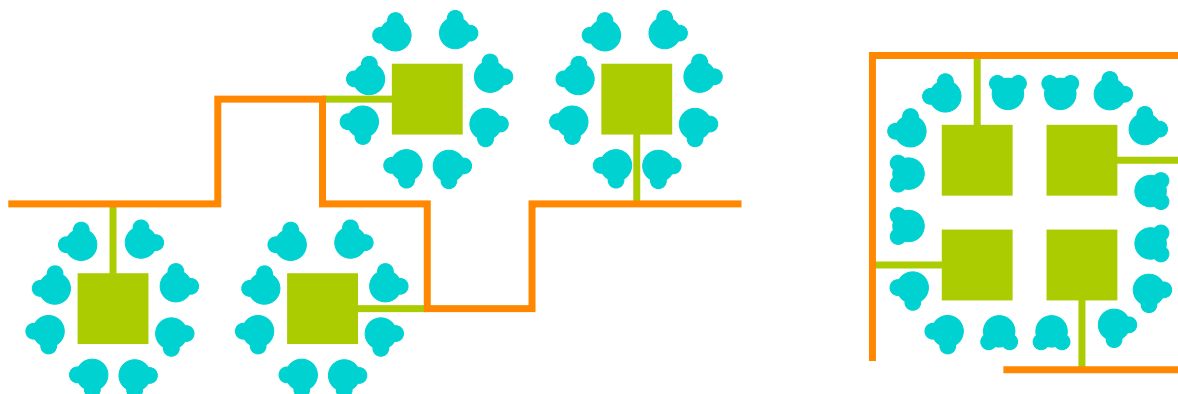
B



C



Obrázek 6.6: Hydrofobní efekt. A, bilance rovnováh mezi cyklohexanem (světle zelená barva) v plynné fázi, kapalně fázi a ve vodě (azurová barva). Výhodné příspěvky ke změně volné energie ve směru šipky jsou zelené, nevýhodné červené. B,C, příčina výrazného snížení entropie při rozpouštění cyklohexanu ve vodě. Zvýrazněná molekula vody může být donorem dvou vodíkových vazeb (B) v šesti orientacích v čisté vodě, ale (C) jen ve třech orientacích v roztoku cyklohexanu (znázorněn světle zeleně).



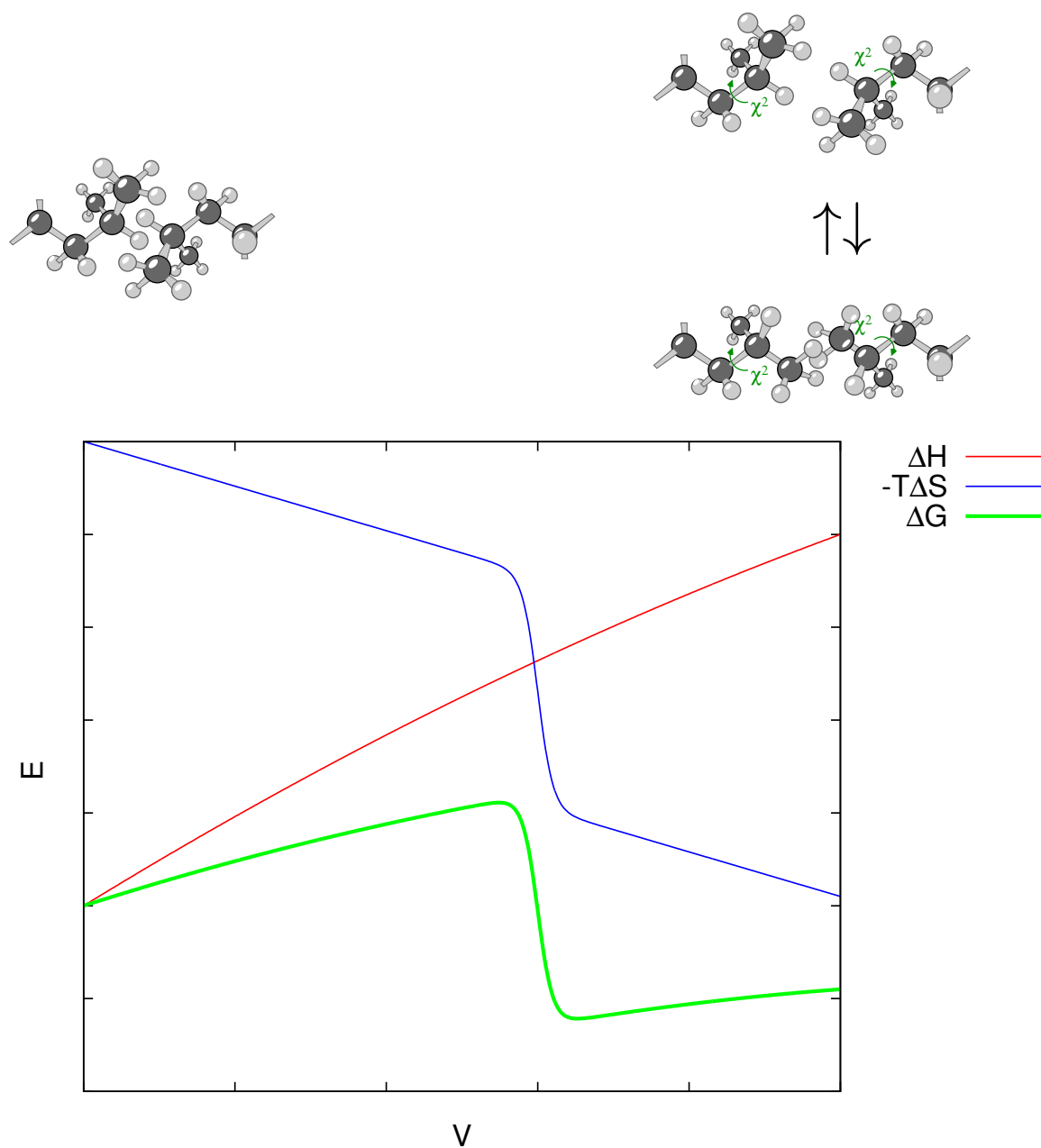
Obrázek 6.7: Stabilizace sbalené nativní konformace proteinu (vpravo) hydrofobním efektem. Hydrofobní postranní řetězce jsou nakresleny světle zeleně, páteř oranžově a molekuly vody azurovou barvou.

### 6.13 Denaturace

Jestliže jsme uvedli, že za nativních podmínek entropický zisk molekul vody převáží entropickou ztrátu postranních řetězců uvězněných v hydrofobním jádře proteinu, musíme dodat, že za odlišných podmínek tomu může být jinak. Rozdíl volné energie sbaleného o rozbaleného proteinu je velmi malý a změnou vnějších podmínek (teplota, tlak, pH, přítomnost dalších látek) můžeme rovnováhu zvrátit ve prospěch rozbaleného stavu. Konkrétním provedením tohoto procesu, zvaného *denaturace*, se tu zabývat nebudeme. Pouze si jej zjednodušeně popíšeme jako růst celkového objemu molekuly  $V$ : nativní protein je sbalený do malé kompaktní struktury, zatímco rozbalený denaturovaný protein zabírá mnohem větší prostor. Co se děje s rozdílem volné energie mezi rozbaleným a sbaleným proteinem během denaturace, docílené například zvyšováním teploty  $T$ ? Na začátku je rozdíl volné energie  $\Delta G$  kladný, rozbalených molekul proteinu je ve vzorku menšina. Jak objem postupně roste, vzdálenosti mezi atomy se zvyšují a proto slábnou disperzní síly. Důsledkem je pozvolný růst rozdílu entalpie  $\Delta H$  (nebo vnitřní energie  $\Delta U$ ). Entropie se přitom příliš nemění, proto entropický příspěvek  $-T\Delta S$  bude klesat přibližně lineárně s rostoucí teplotou. Pro jednoduchost si můžeme představit, že dva postranní řetězce uvnitř proteinu jsou do sebe zaklesnuté tak, že se nemohou volně otáčet a zůstávají v jednom rotamerním stavu. Možnost jejich rotace tedy k entropii nepřispívá, logaritmus jedničky (jeden rotamerní stav) je roven nule. V určitém okamžiku se ale protein rozvolní natolik, že naše dva postranní řetězce (a mnohé další) získají dost prostoru k rotaci. Řekněme, že se najednou každý může vyskytovat v jednom ze tří nezakrytých rotamerů. To okamžitě sníží volnou energii o  $-RT(\ln 3 - \ln 1) = -5,5 \text{ kJ mol}^{-1}$ . Podobně se v tu chvíli uvolní i další postranní řetězce a  $\Delta G$  skokem klesne na zápornou hodnotu. V tomto okamžiku se zastoupení sbalené a rozbalené formy proteinu prudce změní ve prospěch rozbaleného proteinu. Další zvyšování teploty už nic zásadního nezmění, jen se bude dále snižovat už tak nepatrné zastoupení nativního proteinu.

### 6.14 Celková bilance

Jednotlivé příspěvky k volné energii biomakromolekuly, které jsme diskutovali v této kapitole, jsou shrnuty v tabulce 6.2. Co nám tabulka říká? Kovalentní vazby, určující chemickou konfiguraci, jsou samozřejmě nejsilnější. Nekovalentní interakce jsou v buňce výrazně slabší, než ve vakuu. Výjimkou jsou iontové interakce uvnitř proteinu. Nesetkáváme se s nimi často, ale pokud se takový *iontový můstek* neboli *solný můstek* vytvoří, má na strukturu velmi výrazný vliv. Ostatní nekovalentní interakce za



Obrázek 6.8: Změny volné energie během denaturace proteinu. Růst celkového objemu molekuly  $V$  popisuje snižování kompaktnosti proteinu během zahřívání. Uvolnění postranních řetězců vedoucí k prudkému poklesu entropie je schematicky nakresleno nad grafem.

Tabulka 6.2: Srovnání energetických příspěvků určujících strukturu biomakromolekul k celkové volné energii.

Typ	$\text{kJ mol}^{-1}$	podmínky
$RT$	2,5	při 300 K (27 °C)
kovalentní vazba	350	C–C
ion-ion	460	vzdáleny 0,3 nm ve vakuu
ion-ion	150	vzdáleny 0,3 nm uvnitř proteinu
ion-ion	12	vzdáleny 0,3 nm na povrchu proteinu
dipol-dipól	30	vzdáleny 0,3 nm ve vakuu
dipol-dipól	10	vzdáleny 0,3 nm uvnitř proteinu
ion-dipól	41	vzdáleny 0,5 nm ve vakuu
ion-dipól	14	vzdáleny 0,5 nm uvnitř proteinu
vodíková vazba	10	ve vakuu ( $\Delta G \approx \Delta H$ )
vodíková vazba	6	ve vodě ( $\Delta G \approx -T\Delta S$ )
karbonylová interakce	2	v proteinu
hydrofobní efekt	8	na Leu
hydrofobní efekt	12	na Phe

běžných podmínek v cytoplasmě (v tabulce 6.2 zvýrazněné modře) přispívají k celkové volné energii méně a vzájemně srovnatelně. Příspěvek na jednu interakci není velký, většinou kolem  $10 \text{ kJ mol}^{-1}$ . V biomakromolekule je ale takových interakcí velké množství a jejich součet může být významný. To platí zejména pro hydrofobní efekt, který stabilizuje kompaktní nativní strukturu proteinu nejvíce.

Jaká je role ostatních interakcí, když hydrofobní efekt stabilizuje nativní struktury nejvíce? Hydrofobní efekt je sice nejdůležitější z pohledu volné energie, ale je strukturně nespecifický. Jde pouze o to, aby se postranní řetězce seskupily uvnitř proteinu, což, jak víme umožňují různé architektury proteinu (svazky šroubovic,  $\alpha$ -šroubovice uspořádané do mnohostěnů,  $\beta$ -soudky, TIM-soudky, atd.). To, jakou architekturu protein či jiná biomakromolekula vytvoří, určují ostatní interakce, specifické pro danou sekundární a terciární strukturu. Jejich příspěvek ke snížení volné energie je nižší, než příspěvek hydrofobního efektu, ale ne zanedbatelný.

Skutečnost, že se celková volná energie biomakromolekuly skládá z mozaiky mnoha malých a vzájemně podobných příspěvků, vysvětluje, proč je obtížné nativní konformaci proteinu spočítat a proč nelze energetický popis biomakromolekul příliš zjednodušit. Určité obecné závěry ale dělat můžeme a v této kapitole jsme tak nejednou učinili. Můžeme si je shrnout takto:

- Kovalentní vazby určují primární strukturu.
- Iontové můstky v hydrofobním prostředí uvnitř proteinu představují interakci blížící se silou kovalentní vazbě.
- Disulfidové můstky jsou důležité pro tvorbu terciární a kvartérní struktury proteinů, ale jen mimo buňku (v buňce jsou nahrazeny vazbou glutathionu).
- Torzní úhly jsou limitovány sterickou (Pauliho) repulzí.
- Voda jako rozpouštědlo má zásadní vliv.
- Kompaktní nativní struktury jsou stabilizovány především hydrofobním efektem.
- Konkrétní konformaci (sekundární a terciární strukturu) určují elektrostatické síly včetně vodíkových vazeb.

## 6.15 Termodynamika a kinetika

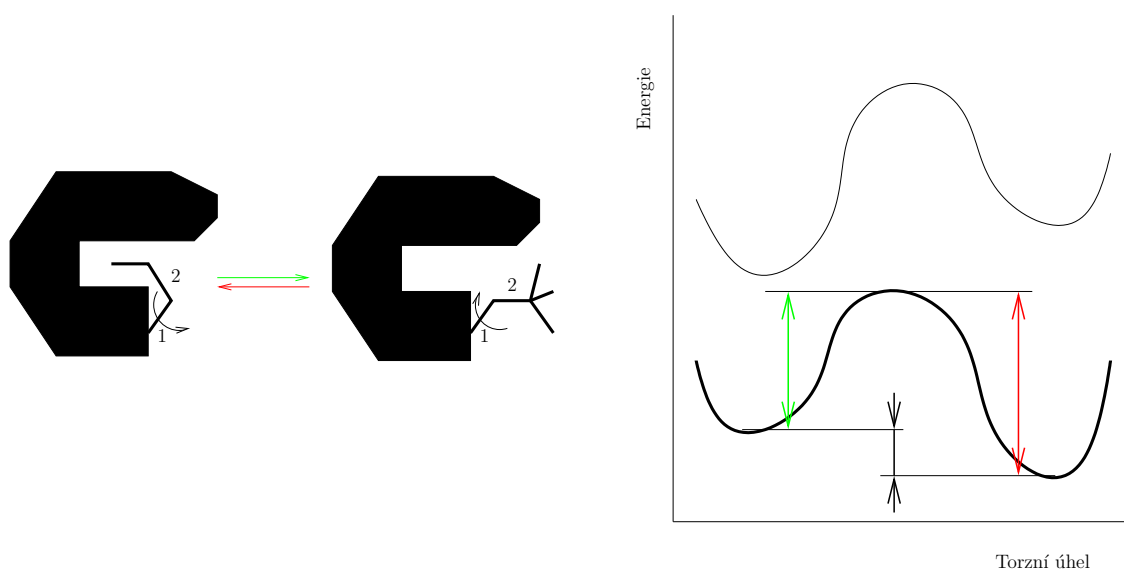
Dosud jsme se zabývali pouze energetickým popisem souborů molekul v rovnováze. Pro studium struktury není ale důležité jen energetické srovnání konformací, ale také znalost rychlosti, se kterou může jedna konformace přecházet v druhou. Popisem rychlostí takových změn konformace se zabývá *kinetika*. Rychlost přeměny konformací je opět dána energií, ne však energií nejvýhodnějších konformací, ale naopak energií nejméně výhodné konformace, kterou musí molekula projít na cestě od jedné výhodné konformace k jiné.

Srovnání kinetického a termodynamického popisu je ukázáno na obrázku 6.9. Pro zjednodušení zde popisujeme konformační změnu spojenou pouze se změnou jednoho torzního úhlu. Během rotace kolem vazby 1 se mění energie molekuly v roztoku. Tenká čára ukazuje průběh vnitřní energie jedné konformace, zatímco tlustá čára popisuje volnou energii souboru molekul v různých konformacích ostatních vazeb<sup>3</sup>. Tlustá čára tedy zahrnuje i příspěvek entropie. Rozdíl mezi údolími na křivce energie určuje pravděpodobnost, že se bude molekula vyskytovat v první konformaci, s postranním řetězcem mířícím do nitra makromolekuly, nebo v druhé konformaci, s postranním řetězcem mířícím do vodného roztoku. Naproti tomu energetický rozdíl mezi údolím a vrcholem na křivce, takzvaná *aktivační energie*, určuje rychlost, se kterou bude přecházet konformace odpovídající danému údolí na konformaci druhou. Čím je tato energetická hradba vyšší, tím bude změna pomalejší<sup>4</sup>. Zatímco u malých molekul a rotací postranních řetězců jsou aktivační energie zanedbatelné a konformační změny probíhají velmi rychle, u rotací kolem vazeb v páteři biomakromolekul se setkáváme s energetickými hradbami tak vysokými, že změny konformace jsou téměř nemožné. Tak se může stát, že molekula zůstane po celý svůj život uvězněna v jednom údolí, ačkoli jiné údolí by bylo energeticky ještě výhodnější.

---

<sup>3</sup>Vliv rozpouštědla tentokrát zanedbáváme, abychom výklad příliš nekomplikovali. Víme ale, že solvatace vodou hraje v interakcích mezi molekulami zásadní roli.

<sup>4</sup>Přesněji řečeno, tím bude menší pravděpodobnost, že ke změně dojde.



Obrázek 6.9: Srovnání termodynamického a kinetického popisu konformační změny. Vlevo je schematicky znázorněna rotace kolem vazby 1. Z možných konformací lišících se rotací kolem vazby 2 jsou znázorněny tři. Vpravo je energetický popis děje. Tenkou čarou je znázorněna závislost vnitřní energie na rotaci kolem vazby 1, tlustou čarou je znázorněna závislost volné energie na rotaci kolem vazby 1. Energetický rozdíl znázorněný uprostřed grafu energie (černě) udává, jaké bude zastoupení jednotlivých konformací v rovnovážné směsi. Energetický rozdíl vyznačený vlevo (zeleně) udává rychlost přeměny levé konformace na pravou, energetický rozdíl vyznačený vpravo (červeně) udává rychlost zpětné konformační změny.

Část II

**Lov struktur**





## Kapitola 7

# Úvod do metod strukturní biochemie

V této části skript si popíšeme metody, pomocí kterých můžeme struktury biologicky zajímavých makromolekul zkoumat. Nejdříve si ukážeme, jak můžeme získat vzorek molekuly pro experimentální metody studia struktur. Poté si představíme metody poskytující informaci o struktuře primární, o pořadí monomerních jednotek v biopolymeru. Nejvíce prostoru věnujeme zkoumání konformace, hlavnímu cíli strukturní biochemie. Zamyslíme se nad tím, co se o konformaci biomakromolekul můžeme dozvědět z výpočtu a pak se zaměříme na metody experimentální. Abychom se v různosti experimentálních přístupů zorientovali, věnujeme jim následující úvahu.

V běžném životě získáváme informace o světě kolem nás pomocí našich pěti smyslů, *čichu, chuti, zraku, hmatu a sluchu*. Každý z těchto smyslů nám podává jinou informaci o okolí. Pokud nás zajímá sklenice na stole před námi, čich a chuť nám napoví, jaký nápoj obsahuje. K tomu, abychom se dozvěděli, jaký má sklenice tvar a jak je plná, musíme ale zapojit jiné smysly. Především zrak, pokud nevidíme, tak hmat. I sluch nám ledacos napoví, sklenice různých tvarů a různě plné také různě zní, když do nich ťukneme lžičkou, nebo přejedeme vlhkým prstem po jejich okraji. Netopýři či delfini používají sluch místo zraku.

Podobně je tomu s experimentálními metodami výzkumu biologicky zajímavých molekul. Optické metody, se kterými se seznámíme v kapitole 11, se podobají čichu a chuti, popisují složení molekul, ne jejich tvar. Tak jako ochutnáním můžeme odhadnout obsah cukru či alkoholu, cirkulární dichroismus nám řekne, kolik procent skládaného listu či  $\alpha$ -šroubovice protein obsahuje, ale neposkytne nám trojrozměrný model molekuly. Pokud nás zajímá prostorové uspořádání, musíme použít jiné přístupy.

Když vás bude zajímat tvar nějakého předmětu, nejspíš se na něj podíváte. Půjde-li o něco maličkého, použijete mikroskop. Tento přímočarý přístup má ale své hranice. Mikroskopem nemůžete pozorovat objekt menší, než je vlnová délka záření, které používáte. Vlnová délka viditelného světla se pohybuje ve stovkách nanometrů. Vzdálenosti mezi atomy jsou ale o málo větší, než desetina nanometru. Když se podíváte na spektrum elektromagnetického záření, zjistíte, že správnou vlnovou délku pro pozorování atomů v molekule mají rentgenové paprsky. Rentgenový mikroskop bohužel postavit neumíme, ze dvou důvodů. Za prvé nemáme rentgenové čočky. Za druhé, i kdybychom je měli, musely by ohýbat záření s chybou menší, než vzdálenost atomů. Přesto existuje způsob, jak „vidět“ molekuly prostřednictvím rentgenových paprsků. Tuto metodu si popíšeme v kapitole 12.

Při hledání záření s vhodnou vlnovou délkou se nemusíme omezovat na elektromagnetické vlny. Kvantová mechanika ukazuje, že kromě částic světla, fotonů, mají vlnovou povahu i jiné částice. Dva druhy částic mohou být dostatečně urychleny, aby jejich vlnová délka byla v řádu desetin nanometru. Jsou to neutrony a elektrony. Neutrony můžeme použít podobným způsobem jako fotony, elektrony mají

ale jednu výhodu. Protože jsou elektrony nabité, můžeme jejich paprsek ohýbat v magnetickém poli. Díky tomu můžeme postavit elektronový mikroskop. Rozlišení nejlepších elektronových mikroskopů se blíží atomovým vzdálenostem. Musíme přitom ale používat velmi nízké dávky elektronů, abychom nepoškodili vzorek. Při elektronové mikroskopii proto zápolíme s nízkým poměrem signálu k šumu. Proto se při elektronové mikroskopii vysokého rozlišení používá průměrování většího počtu obrazů. Elektronová mikroskopie je nejužitečnější pro studium největších molekul a molekulových komplexů, pro které je použití rentgenové difrakce značně náročné. Elektronová mikroskopie a rentgenová krystalografie se proto vzájemně doplňují.

Tak, jak jsme si popsali metody založené na difrakci záření jako obdobu zraku, můžeme najít i metody představující analogii hmatu. Jde o skupinu metod označovanou jako atomová silová mikroskopie (atomic force microscopy, AFM). Při těchto metodách se používá sonda s velmi ostrým hrotem, která „osahává“ zkoumaný vzorek podobně jako jehla gramofonu „ohmatává“ povrch klasické vinylové desky.

Posledním smyslem je sluch. I pro něj můžeme mezi metodami strukturní analýzy najít přibližnou analogii. Je jí nukleární magnetická rezonance (NMR), popsaná v kapitole 14. Podobně jako při ťukání na skleničku, NMR analýza poskytuje informaci o prostorovém uspořádání zakódovanou do detegovaných frekvencí. Z tohoto pohledu jde o metodu nepřímou, podobně jako sluch: frekvence tónu odráží plnost a tvar sklenky, zpracování této informace do jasné představy sklenice je ale pro náš mozek dosti tvrdý oříšek.

## Kapitola 8

# Příprava vzorku biomakromolekul

### 8.1 Izolace z přírodního materiálu

Zpracování přírodního materiálu (rozbitých buněk, tělních tekutin a nejrůznějších produktů, které organismus vylučuje) je nejstarším způsobem získávání biomakromolekul. Výhodou tohoto postupu je snadná dostupnost větších množství takového materiálu. Nevýhodou je složitost směsi, ze které chceme naši biomakromolekulu získat.

Přírodní vzorek, z kterého chceme získat jednu čistou látku, zpravidla obsahuje tisíce jiných látek. K získání té jediné požadované je nutno navrhnout vhodný postup postupného čištění (*purifikační schéma*). Takový postup zahrnuje vhodnou kombinaci různých *separačních technik* (metod dělení látek). V tomto krátkém přehledu není možné probírat detaily purifikačních postupů, které se případ od případu liší. Zmíníme tedy pouze obecné přístupy k purifikaci různých biologicky zajímavých molekul.

Pro čištění složitých směsí biologických látek byla vyvinuta řada separačních metod. Tyto metody dělí látky na základě různé interakce molekul směsi s vnějším polem nebo s molekulami přítomnými v separačním zařízení. Volba metody závisí na množství látky, které chceme získat. K určení sekvence postačují nanogramová i menší množství, zatímco určení prostorové struktury vyžaduje několik miligramů čisté látky. Obecně lze říci, že v prvních krocích purifikačního postupu se používají metody, které sice nedělí látky příliš účinně, ale mohou zpracovat velké množství vzorku. V závěru naopak přicházejí ke slovu techniky s omezenou *kapacitou* (množstvím vzorku, které lze použít), ale vysokou účinností dělení. Navíc je třeba kombinovat metody, které dělí látky podle různých vlastností (například podle velikosti molekul, síly iontových interakcí a velikosti hydrofobního efektu), to se týká zejména proteinů.

K nejhrubším dělicím technikám patří srážení a extrakce různými rozpouštědly. Lze jimi oddělit větší skupiny látek, například nukleové kyseliny od proteinů. Poměrně hrubé je také dělení na membránách, které umožňuje třídít molekuly podle velikosti. Centrifugační metody jsou vhodné pro dělení největších molekul. Jde o techniku šetrnou, její účinnost ale není příliš vysoká. Elektroforézu a příbuzné metody je možné použít pro dělení látek tvořící ionty. Elektroforetické dělení může být velmi účinné (zejména pokud provádíme elektroforézu v tenké kapiláře). Tímto způsobem lze ale většinou zpracovat jen malá množství látek. Nejrozumnější kompromis mezi účinností a kapacitou poskytují metody kapalínové chromatografie. Velmi účinná je chromatografie na reverzní fázi, která dělí molekuly na základě hydrofobního efektu. Nevýhodou této metody je používání organických rozpouštědel, které často poruší přirozenou konformaci proteinů. Šetrnější obdobou je takzvaná hydrofobní chromatografie. Velmi oblíbená je iontoměničová chromatografie, která je poměrně účinná, šetrná a dochází při ní k zakoncentrování vzorku. Vylučovací (gelová) chromatografie nemá příliš vysokou kapacitu ani účinnost, umožňuje ale šetrné dělení molekul podle velikosti. Nejúčinnější dělení poskytují různé varianty afinitní chromatografie, které jsou založeny na vysoce specifických interakcích (například antigenu s protilátkou nebo enzymu

s inhibitorem).

Jako příklad purifikačního schématu proteinu můžeme uvést postup zahrnující srážení síranem amonným, iontoměničovou, hydrofobní a vylučovací chromatografií (v tomto pořadí). Malá množství vysoce čistého proteinu pro určení sekvence můžeme získat elektroforézou nebo chromatografií na reverzní fázi<sup>1</sup> v kapilárním provedení. Po určení částečné sekvence je pak možné vyhledat gen příslušného proteinu a větší množství vzorku připravit metodami molekulární biologie (viz sekci 8.2).

Nukleové kyseliny se obvykle získávají iontoměničovou chromatografií nebo extrakcí do směsi chloroformu a fenolu a následným alkoholovým srážením. K dělení DNA podle velikosti molekul se používá elektroforéza v gelu. Díky párování bází lze izolovat nukleovou kyselinu, která obsahuje pouze určitou sekvenci nukleotidů. Takto izolovanou DNA obvykle vnášíme metodami molekulární genetiky (sekce 8.2) do jiného organismu, se kterým se snaže pracuje. Jinou možností je určit sekvenci z malého množství nukleové kyseliny a připravit větší množství chemicky (sekce 8.9).

Srážení a extrakce slouží i k hrubému přečištění lipidů a polysacharidů. K jemnějšímu dělení slouží opět chromatografie na reverzní nebo normální fázi. Polysacharidy se často dělí pomocí iontoměničové chromatografie nebo podle velikosti vylučovací chromatografií. Protože stavba sacharidů a biomembrán není přímo geneticky kódována, jsme na studium přírodního materiálu odkázáni mnohem více, než v případě proteinů a oligonukleotidů.

Lipidové dvojvrstvy můžeme připravit poměrně jednoduše smícháním příslušných lipidů. V takovém případě ale nejde o biomembrány v pravém slova smyslu, ale spíše o modelové útvary, které postrádají membránové proteiny a jiné složky. Studium procesů, které se na membránách odehrávají, proto vyžaduje přípravu dvojvrstvy obsahující potřebné membránové proteiny. Ty jsou obvykle izolovány z přírodního materiálu s použitím vhodného detergentu, který umožní jejich rozpuštění (*solubilizaci*).

## 8.2 Metody molekulární biologie

V počátcích *molekulární biologie* se zájem badatelů soustředil na pochopení toho, jak buňka předává a zpracovává dědičnou informaci (na *molekulární genetiku*). Tento cíl vyžadoval nejprve zvládnutí experimentální práce s molekulou, která genetickou informaci nese, s DNA. Rozvoj metod práce s nukleovými kyselinami umožnil, že molekulární biologické techniky jsou dnes využívány nejen k přípravě vzorků nukleových kyselin, ale i vzorků proteinů. Metody *genového inženýrství* umožňují přípravu velkého množství proteinu v cizím organismu (se kterým se snadněji pracuje), zavedení *afinitní značky* (která dramaticky zjednodušuje purifikaci) a určitou manipulaci s proteinem (jako záměnu isotopů pro potřeby nukleární magnetické rezonance nebo záměnu síry za selen využívaný v rentgenové strukturní analýze).

Metody molekulární biologie vycházejí ze způsobu, jakým se předává genetická informace v buňce. Tento soubor složitých dějů si můžeme shrnout do několika bodů. Zaměříme se přitom na bakterie, se kterými se pracuje nejnáze.

- Nositelem genetické informace je molekula DNA. Většina dědičné informace je v buňce přítomna v jedné kopii ve formě jedné nebo více dlouhých dvojjřoubovic, zvaných *chromozomy*. Chromozom bakterie *Escherichia coli*, se kterou molekulární biologové pracují nejčastěji, je kruhová molekula obsahující čtyři miliony nukleotidů. V bakteriálních buňkách je navíc malá část genetické informace přítomna ve formě krátkých kruhových molekul DNA, zvaných *plasmidy*. Obvyklá délka plasmidů je několik tisíc nukleotidů.
- Molekula DNA je z chemického hlediska souvislý řetězec nukleotidů. Z biologického hlediska si ale tento řetězec můžeme rozdělit na kratší úseky, z nichž každý nese informaci o jedné vlastnosti

<sup>1</sup>K určení sekvence není třeba zachovat přirozenou konformaci.

buňky. Tyto úseky se nazývají *geny*. Geny jsou části informace, nikoli struktury, sekvence nukleotidů kódující jeden gen mohou být několikrát přerušeny sekvencemi, které nesouvisí s informací daného genu (takové sekvence se nazývají *introny*), nebo může určitý úsek DNA kódovat dva různé překrývající se geny.

- Během buněčného dělení se vytvoří kopie (*replika*) chromozomu, takže obě dceřinné buňky obsahují po jedné molekule chromozomové DNA. Tento proces se nazývá *replikace*.
- Buňka si svou dědičnou informaci důsledně uchovává a předává následujícím generacím. Za určitých okolností si ale mohou buňky genetickou informaci vyměňovat. Je-li například život bakterie ohrožen nějakým antibiotikem, začne bakteriální buňka ochotně přijímat plasmidy v naději, že některý z plasmidů bude obsahovat gen, který ji před antibiotikem ochrání (gen *rezistence*). Popsaná změna genetické informace se nazývá *transformace* buňky.
- Proces, kterým se určitá dědičná vlastnost projeví, se nazývá *exprese* genu. Nejběžnějším případem je syntéza proteinu kódovaného určitým genem (*exprese proteinu*).
- Má-li dojít k expresi, musí být nejdříve příslušný gen „přepsán“ do molekuly RNA. Tento proces se nazývá *transkripce*, vzniklá molekula RNA se označuje jako *mediátorová* RNA (mRNA). Vzniklá RNA je „odlitkem“, nikoli kopií určitého úseku DNA, navíc jsou z ní odstraněny nekódující sekvence nukleotidů (introny) procesem zvaným *sestřih*.
- Během exprese proteinů je genetická informace přeložena podle kódovací tabulky ze sekvence nukleotidů do sekvence aminokyselin. Tento proces interpretace genetického kódu se nazývá *translace*. K vlastnímu dekódování dochází během takzvané *aktivace* aminokyselin, kdy je každý aminokyselinový zbytek navázán na příslušnou krátkou molekulu RNA, zvanou *transferová* RNA (tRNA). Enzymy, které aktivaci řídí, musí rozpoznat, která tRNA patří ke které aminokyselině (musí „znát genetický kód“). Každá tRNA pak už rozpozná správné místo na molekule mRNA podle párování bází. Po navázání tRNA na mRNA dojde k přenosu zbytku aminokyseliny na konec rostoucího polypeptidového řetězce. Tato syntéza proteinů probíhá v útvech zvaných ribozomy.

## 8.3 Expresní vektory

Studium eukaryotních proteinů často probíhá podle naznačeného schématu „od genu k proteinu“. Jak bude zmíněno v sekci 9.1, často se i u proteinu nejprve izolovaného z přirozeného zdroje postupuje „od proteinu přes gen k proteinu“.

Eukaryotní gen se (ve formě odpovídající již sestřihané mRNA) zavede do molekuly DNA schopné nezávislé replikace a exprese (takzvaného *vektoru*, nejčastěji bývá vektorem plasmid). Tím se vytvoří takzvaný *expresní konstrukt*, kterým se *transformuje* vhodný kmen *Escherichia coli*. Dnes jsou k dispozici kmeny, se kterými se pohodlně a bezpečně pracuje a které jsou ochotny podle dodané genetické informace vyrábět velká množství proteinů pro *E. coli* zcela cizích.

K přesně definovanému vyjmutí úseku DNA se používají zvláštní enzymy zvané *restrikční endonukleasy*, které specificky rozpoznávají krátké sekvence (nejčastěji šesti nukleotidů). Vektory, do kterých úsek DNA vkládáme jsou uměle připraveny tak, aby obsahovaly:

1. větší počet míst rozpoznávaných jednotlivými restrikčními endonukleasami,
2. sekvence nezbytné pro replikaci a expresi vektoru v bakterii (sekvence určující počátek replikace, bakteriální *promotor*, což je sekvence rozpoznávaná jako začátek transkripce, místo vazby na ribozomy),

3. sekvenci kódující takzvanou *afinitní značku* (několik aminokyselinových zbytků, obvykle histidinových, které se pevně váží k určitému chromatografickému nosiči),
4. selektivní značku, které umožní rozeznat buňky, do kterých opravdu pronikl použitý vektor. Nejčastěji se používá gen rezistence k vybranému antibiotiku. Bakterie se pak pěstují v prostředí obsahujícím toto antibiotikum, takže přežijí pouze ty buňky, které přijaly expresní vektor.

## 8.4 Molekulové klonování

Pro úspěšnou expresi proteinu je nezbytné získat velké množství potomků jediné buňky, jejíž genetickou informací jsme upravili zavedením genu pro náš protein. V takovém společenství, které se nazývá *klon*, nesou všechny buňky požadovanou genetickou informaci (gen našeho proteinu). Základní postup molekulového klonování zahrnuje tři kroky:

### Příprava genu, který chceme do bakterie vložit

Předpokládejme, že máme k dispozici (například v jiném plasmidu) gen proteinu, odpovídající již sestřižené<sup>2</sup> mRNA. Tento gen můžeme rozmnožit způsobem, který napodobuje replikaci v živé buňce. Dnes se nejčastěji používá takzvaná *polymerasová řetězová reakce* (PCR z anglického *polymerase chain reaction*), kde celý proces provádíme ve zkumavce a řídíme jej pouze změnami teploty reakční směsi. Využíváme přitom zvláštní vlastnosti enzymu, který replikaci provádí. Tento enzym neumí totiž syntézu nového vlákna zahájit, potřebuje již kousek dvojšroubovice, kterou dále umí prodlužovat. Proto musí být ve směsi kromě dlouhého vlákna DNA (obsahující gen) i takzvané *primery* – krátké úseky DNA, jejichž báze se párují s konci dlouhého vlákna a vytvářejí tak potřebné zárodky dvojšroubovice.

V praxi tedy postupujeme tak, že chemicky nasyntetizujeme (nebo si necháme na objednávku vyrobit) dva krátké oligonukleotidy, které nám budou sloužit jako primery. Nesmí být příliš krátké, to by se nevytvořily stabilní zárodky dvojšroubovice, naopak syntéza dlouhých primerů by byla drahá. Vhodná délka je asi 18 nukleotidů. Primery nám umožňují nejen vybrat si, který úsek DNA chceme replikovat, ale také zavést několik koncových nukleotidů podle naší potřeby. Toho využíváme k zavedení sekvencí, které rozpoznají námi zvolené restriční endonukleasy.

Připravíme reakční směs obsahující DNA s genem, primery, volnými nukleotidy (ve formě trifosfátů, které vyžaduje replikační enzym), enzymem a dalšími látkami potřebnými pro dobrý průběh reakce.

Postupně měníme teplotu reakční směsi. Při ochlazení se spojí primery s dlouhým vláknem DNA, při optimální teplotě reakce dochází k syntéze nového vlákna DNA podle starého (obě vlákna vytvoří dvojšroubovici), při zahřátí se dvojšroubovice rozpadne na dvě vlákna. V dalším kroku tedy už začínáme se dvěma vlákny a po stejném teplotním cyklu dostaneme vlákna čtyři, v příštím kroku osm, pak šestnáct, a tak dále geometrickou řadou.

Nakonec provedeme štěpení DNA, kterou chceme do vektoru vložit, restričními endonukleasami. Po štěpení restričními endonukleasami vznikají krátké přesahující jednovláknové sekvence, kterými se později dvě molekuly DNA spojí na základě párování bází.

### Vložení genu do vektoru

Vektor (plasmid, do kterého chceme gen vložit) rozštěpíme stejnými restričními nukleasami, jakými jsme štěpili DNA obsahující náš gen. Vektory, které lze koupit, obsahují velké množství sekvencí rozpoznávaných různými restričními endonukleasami. Není proto obtížné vybrat si vhodné štěpící místo.

Po smíchání vektoru s genem dojde k párování konců obou molekul DNA. Molekuly ovšem dosud nejsou kovalentně spojeny. Dalším krokem je proto vytvoření fosfodiesterových vazeb mezi genem a vektorem pomocí specifických enzymů. Tím vznikne expresní konstrukt, který můžeme vložit do bakterie.

<sup>2</sup>Sekvence DNA odpovídající již sestřižené mRNA se označuje cDNA.

## Transformace

K situaci, kdy je bakteriální buňka ochotna přijmout cizí DNA dochází jen výjimečně. Chceme-li vložit expresní konstrukt do bakterie, musíme buňky nejdříve připravit a pak vystavit krizovým podmínkám (například zvýšené teplotě). V té chvíli se budou buňky cítit ohroženy a budou ochotny přijmout expresní konstrukt – dojde k transformaci.

Transformované buňky necháme množit na živné půdě (*médiu*) s antibiotikem, proti kterému jsou rezistentní jen buňky obsahující použitý vektor. Bakteriální kulturu potom rozetřeme na agarovou plotnu tak, aby vyrostly oddělené kolonie bakterií. Každá taková kolonie je tvořena potomky jedné buňky, jde tedy o jeden klon. Analýzou DNA jednotlivých klonů zjistíme, které opravdu obsahují vektor se začleněným genem.

Kolonii pak pipetou nebo párátkem přeneseme do živné půdy a necháme rozmnožit. Následnou selekcí se vyberou klony buněk, které opravdu obsahují vektor se začleněným genem.

## 8.5 Místně řízená mutageneze

V některých případech nechceme zkoumat pouze protein přítomný ve studovaném organismu, ale chceme jej nějakým způsobem pozměnit. Takových změn genetické informace, neboli *mutací*, lze dosáhnout pomocí metod *řízené mutageneze*.

Genetickou informaci naklonovanou do vhodného vektoru lze pozměnit zavedením mutací na přesně daném místě. Běžné strategie místně řízené mutageneze lze rozdělit na dva druhy.

1. První postup je založen na vyštípnutí dvojřetězcového fragmentu DNA v místě, které má být mutováno (přesněji řečeno mezi nejbližšími místy rozpoznávanými restrikčními nukleasami) a na nahrazení tohoto fragmentu uměle připraveným fragmentem obsahujícím požadovanou mutaci.
2. Druhý postup využívá toho, že dostatečně dlouhý (asi 18 párů bází) úsek dvojšroubovice může být účinně párován i v případě, že jeden pár bází není *komplementární* (netvoří pár dle Watsona a Cricka). Vektor s naklonovaným genem se rozdělí na jednotlivá vlákna a jedno z nich se nechá párovat s uměle připraveným oligonukleotidem obsahujícím jeden vyměněný (nepárující) nukleotid, takzvanou *bodovou mutaci*. Tento oligonukleotid pak slouží jako primer pro dosyntetizování druhého vlákna vektoru. Mutace je tak zavedena do jednoho vlákna výsledného vektoru, které je třeba izolovat. Tento druhý postup je sice složitější, ale je levnější, neboť nemusí být syntetizován celý dvojřetězcový úsek molekuly DNA.

## 8.6 Bakteriální exprese

Před expresí se transformované buňky namnoží v potřebném objemu kultivačního media a vlastní exprese se zahájí přidáním látky, která je nezbytná pro transkripci genu v plasmidu<sup>3</sup> – takzvanou *indukcí*. Nejčastěji používaným organismem je *E. coli*, která představuje dobře definovaný organismus s geneticky upravenými kmeny umožňujícími vysokou expresi cizích proteinů. Pro *E. coli* je také komerčně dostupný bohatý výběr vektorů.

Bakteriální exprese má ale i své nevýhody. Bakterie neumožňují sestřih a posttranslační modifikace eukaryotních proteinů. Proteiny mohou být vylučovány ne jako rozpustné makromolekuly ve správné

<sup>3</sup>Používané plasmidy jsou navrženy tak, že místo sloužící jako počátek transkripce, takzvaný *promotor*, je rozpoznáváno příslušným enzymem pouze tehdy, je-li přítomna určitá látka. Nejčastěji se používá promotor, který bakterie používají k zahájení transkripce genů pro odbourání laktosy. Tento promotor je běžně nefunkční, protože se v jeho blízkosti váže protein, který jej blokuje. Když ale přidáme látku, která je schopna vázat se na tento protein (používá se isopropylthiogalaktosid), protein se váže k přidané látce místo k DNA a uvolní tak promotor. Volný promotor je rozpoznán příslušným enzymem a exprese může začít.

konformaci, ale v nerozpustných shlucích zvaných *inkluzní tělíska* a disulfidové můstky nemusí být správně vytvořeny. Posledně zmíněné překážky se dají částečně obejít vhodnou volbou podmínek exprese (například snížení teploty).

## 8.7 Purifikace exprimovaných proteinů

Pro sklizení buněk je třeba rozbít buněčnou stěnu a izolovat exprimovaný protein. Pro usnadnění poslední fáze se často připravuje protein ne ve své přirozené formě, ale spojený s jiným peptidem. Druhý peptid může být buď krátký (několik aminokyselin), nebo může jít o celý protein.

Krátké peptidové *afinitní značky* umožňují snadnou purifikaci (přečištění) za vždy stejných podmínek, bez ohledu na vlastnosti exprimovaného proteinu<sup>4</sup>. Použití celého proteinu jako afinitní značky nejen umožňuje snadnou purifikaci, ale může i příznivě ovlivnit skládání proteinu (a tak omezit tvorbu inkluzních tělísek). Mezi proteiny používané k tomuto účelu patří thioredoxin, glutathion-*S*-transferasa, maltose-binding protein. Odštěpení proteinové značky je umožněno zavedením sekvence rozeznávané specifickou endopeptidasou jako faktor Xa.

## 8.8 Chemická syntéza peptidů

Syntéza biomakromolekul s přesně daným pořadím monomerních jednotek je poměrně obtížný chemický úkol. Nejčastěji se používá syntéza na pevném nosiči, na který je navázána poslední jednotka rostoucího řetězce (obrázek 8.1). Vezměme si náš oblíbený příklad proteinu a zamysleme se, jak bychom jej mohli připravit chemicky.

Každá aminokyselina obsahuje několik funkčních skupin, které spolu mohou reagovat. Chceme-li z aminokyselin syntetizovat peptid, musíme zařídit, aby

- spolu reagovaly skupiny, které chceme spojit a
- nereagovaly skupiny, které chceme zachovat nedotčeny.

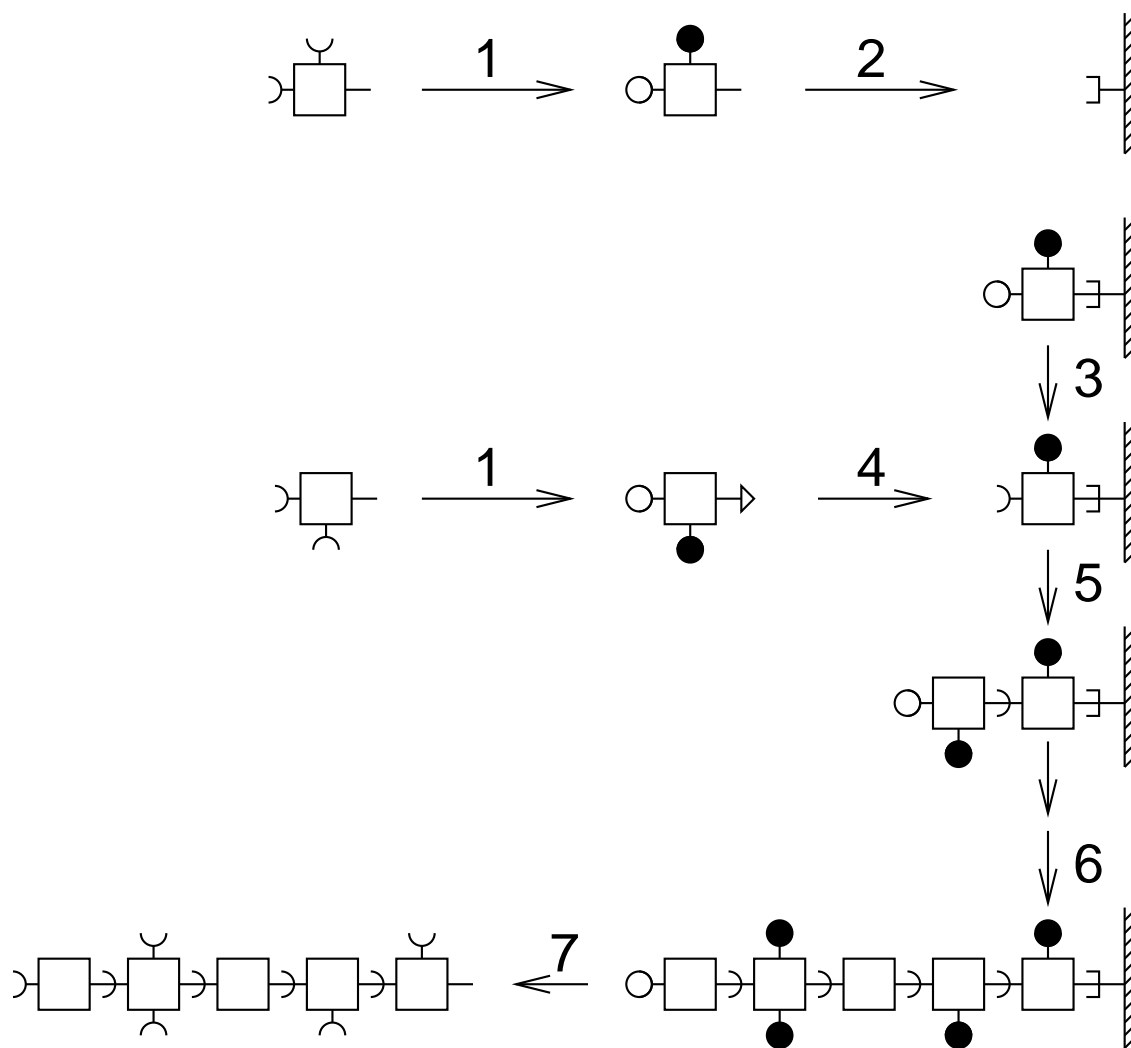
V sekci 2.2 jsme si popsali formální tvorbu peptidové vazby kondenzací amino- a karboxyskupiny dvou aminokyselin. Bez účinné katalýzy je ale tato reakce příliš pomalá, protože aminy a karboxylové kyseliny nejsou mimořádně reaktivní. K dosažení prvního cíle potřebujeme proto alespoň jednu funkční skupinu *aktivovat*. Z chemického pohledu je třeba nahradit OH-skupinu karboxylu lépe odstupující skupinou (na obrázku 8.1 je taková aktivující skupina znázorněna trojúhelníčkem).

Funkční skupiny, které spolu reagovat nemají, potřebujeme naopak ochránit, neboli *blokovat*. Po skončení syntézy musíme samozřejmě peptid *odblokovat* (šetrně odstranit chránicí skupiny), abychom získali molekulu identickou s přirozeným peptidem. Chránicí skupiny jsou na obrázku 8.1 představovány kroužkem, plný a prázdný kroužek mají ukázat, že je často dobré mít různé blokuující skupiny, které lze odštěpit nezávisle.

Obecně si postup chemické syntézy můžeme rozdělit do několika kroků, které budeme číslovat podle obrázku 8.1. Nejprve musíme připravit jednotlivé stavební jednotky s vhodně aktivovanými a blokovánými funkčními skupinami (1). Potom navážeme první (nebo poslední) monomerní jednotku na pevný nosič (2). Z navázané jednotky selektivně odštěpíme chránicí skupinu, která až dosud blokovala

<sup>4</sup>Nejčastěji se používají sekvence šesti histidinů, které komplexují dvojmocné kationty nikelnaté, kobaltnaté, zinečnaté, měďnaté. Protein se izoluje metalochelatační afinitní chromatografií. Tato metoda využívá kolon s chelátovanými kationty. Při průchodu kolonou komplexuje kation histidiny. Síla vazby je mnohonásobně vyšší pro umělé afinitní značky obsahující několik histidinů za sebou. Afinitně vázaný protein je možno vymýt snížením pH (protonace histidinů), imidazolem (kompetice s histidinem), kyselinou ethylendiamintetraoctovou (komplexuje kationty lépe než kolona), nebo odštěpením afinitní značky od proteinu (pokud je afinitní značka oddělena sekvencí rozpoznávanou specifickou endopeptidasou).





Obrázek 8.1: Obecný postup chemické syntézy biopolymeru. Jednotlivé kroky jsou popsány v textu.

konec rostoucího řetězce (3). Přidáme aktivovanou a blokovanou druhou (nebo předposlední) monomerní jednotku (4) a prodloužíme rostoucí řetězec (5). Celý postup opakujeme tolikrát, kolik jednotek má syntetizovaný polymer obsahovat (6). Nakonec hotový biopolymer uvolníme z nosiče a odstraníme všechny chránící skupiny (7).

K chemické syntéze peptidů se obvykle používá takzvaná *karbodiimidová metoda*, která je popsána v Dodatku B.1. Touto metodou je možno syntetizovat i delší peptidové řetězce. Tento postup je ale nákladný, proto se syntetizují většinou krátké peptidy, zatímco větší proteiny se připravují metodami molekulární biologie (sekce 8.2).

## 8.9 Chemická syntéza oligonukleotidů

Postup chemické syntézy nukleových kyselin vychází ze stejných principů jako peptidová syntéza. Klíčem k úspěchu je zaručit, aby se vytvořil fosfodiester mezi 5' a 3' hydroxyly a poslední OH-skupina kyseliny fosforečné zůstala volná. Nejčastěji se používá takzvaná *fosforamiditová metoda*, která je popsána v Dodatku B.2. V této metodě je málo reaktivní ester kyseliny fosforečné nahrazen reaktivním diisopropylamidem kyseliny fosforité, jejíž druhá OH skupina je chráněna  $\beta$ -kyanoethylovou skupinou.

Fosforamiditová syntéza se běžně používá k přípravě krátkých a středně dlouhých oligonukleotidů pro potřeby molekulového klonování. Pomocí metod molekulární genetiky můžeme oligonukleotidy spojovat do delších úseků DNA. Takto je možné uměle připravovat celé geny, nutná je pouze znalost sekvence.

## 8.10 Chemická syntéza oligosacharidů

Princip syntézy oligosacharidů je podobný jako v případě peptidů a oligonukleotidů. Je nutné aktivovat uhlík, který má vytvořit glykosidickou vazbu a ochránit všechny OH-skupiny kromě jediné, která se má na aktivovaný uhlík glykosidicky vázat. K aktivaci se například používají thioglykosidy. Chránící skupiny musí být různého druhu, aby bylo možné odblokovat jen vybranou OH- skupinu. Nejběžnějším chráněním je navázání benzylu nebo esterifikace kyselinou (octovou, benzoovou, pivalovou, levulovou). Chemickou syntézou je možné připravit krátké oligosacharidy včetně větvených, zdaleka však nejde o rutinně zavedené metody jako v případě peptidů a oligonukleotidů.

## 8.11 Enzymová syntéza nukleových kyselin

V sekci 8.2 jsme popsali, jakým způsobem lze získat proteiny (a nukleové kyseliny) biosyntézou v živé buňce. V částech 8.8 a 8.9 jsme se zmínili o metodách přípravy biomakromolekul chemickou syntézou. Enzymová syntéza představuje jakýsi přechod mezi biologickou a chemickou přípravou. Stejně jako v bakteriální expresi využívá složitých biologických katalyzátorů – enzymů, syntéza však neprobíhá v živé buňce, ale „ve zkumavce“ s použitím izolovaných enzymů.

Nejčastějším případem enzymové syntézy je *polymerázová řetězová reakce (PCR)*, již zmíněná v sekci 8.4. Tato metoda je poměrně jednoduchá, protože k syntéze DNA stačí jediný enzym (DNA-polymerasa). Dalšími složkami reakční směsi jsou *templát* (molekula DNA nesoucí genetickou informaci, podle které syntéza probíhá), *primery* (krátké oligonukleotidy vymezující tu část templátu, podle které syntéza opravdu proběhne), *stavební jednotky* (čtyři nukleosidtrifosfáty, které slouží jako aktivované nukleotidy) a několik málo pomocných látek. V praxi se používá enzym z bakterií žijících v horkých pramenech, takže chod reakce lze řídit pouhou změnou teploty<sup>5</sup>. Účinnost metody je mimořádná, z jediné molekuly templátu lze připravit dostatečné množství DNA, aby se s ní dalo pohodlně pracovat. Pomocí enzymové syntézy je možno připravit i molekuly RNA, je ale nutno použít jiný enzym (RNA-polymerasu).

<sup>5</sup>Enzym se neničí ani zahřátím téměř k bodu varu, takže jej není třeba během reakce přidávat.

## 8.12 Enzymová syntéza proteinů

V poslední době byla metoda podobná PCR zavedena i pro syntézu proteinů. Jde o takzvanou *bez-buněčnou expresi proteinů* (*cell-free protein expression*). Ve srovnání z PCR je bezbuněčná proteinová exprese mnohem náročnější, protože syntéza proteinů v buňkách je mnohem složitější, než replikace DNA. Expese proteinů zahrnuje transkripci (syntézu mRNA podle templátu DNA) a translaci, která probíhá na složitých nadmolekulárních útvarech – ribosomech. Navíc nejsou aminokyseliny aktivované jednoduchým navázáním zbytků kyseliny fosforečné jako nukleotidy, nýbrž vazbou na relativně složité molekuly tRNA, které zaručují správné přeložení genetické informace do sekvence aminokyselin. Proto musí reakční směs pro bezbuněčnou expresi proteinu obsahovat množství enzymů včetně celých ribozomů, izolovaných z bakteriálních buněk. Templátem je opět DNA (ve formě vhodného vektoru), takže enzymy musí zajistit i transkripci do mRNA. Jako stavební bloky slouží obvykle volné aminokyseliny, které jsou aktivovány až v reakční směsi (ta musí obsahovat potřebné enzymy a tRNA izolované z buněk). Je ale možné použít aminokyseliny již vázané na tRNA. Výhodou této možnosti je, že nejsme omezeni na 20 geneticky kódovaných aminokyselin (synteticky můžeme připravit a do molekuly proteinu zařadit i aminokyseliny, které se v přírodě nevyskytují).

V současné době nachází bezbuněčná syntéza uplatnění zejména ve dvou případech. První případ zahrnuje situace, kdy není možné protein v buňce připravit (protein je pro buňku toxický, nebo jej enzymy buňky štěpí). Druhým případem jsou situace, kdy chceme připravit protein, ve kterém jsou aminokyseliny jasně daným způsobem upraveny (začlenění nepřirozených nebo izotopově značených aminokyselin na určité místo proteinu).

## 8.13 Enzymová syntéza oligosacharidů

Pomocí enzymů izolovaných z různých organismů je možné spojovat sacharidové stavební jednotky do oligosacharidů. Princip metody se liší od enzymové syntézy proteinů a DNA v tom, že nelze použít templát, který by definoval primární strukturu oligosacharidu. Důvod je prostý – sacharidy nejsou přímo geneticky kódovány. K určení pořadí sacharidových jednotek a způsobu jejich vazby se využívá specifity jednotlivých enzymů. Každý enzym je schopen katalyzovat vznik jen určité vazby mezi určitými sacharidy. V syntéze oligosacharidu jsou proto používány v každém kroku různé enzymy, podle toho které sacharidy chceme spojit a jak. Metody enzymové syntézy oligosacharidů jsou pracné, ale vzhledem k obtížnosti chemické syntézy jsou s oblibou používány, zejména k přípravě větších množství oligosacharidů.



## Kapitola 9

# Metody určování sekvence

### 9.1 Sekvence proteinů

K určování sekvence proteinů je možno přistupovat dvěma způsoby. První přístup je založen na postupném odštěpování koncových aminokyselin a jejich analýze. Nejběžnějším příkladem je chemické odbourávání Edmanovou metodou. Druhý přístup je založen na částečném štěpení polypeptidového řetězce a analýze vzniklých štěpů (nejčastěji hmotnostní spektrometrií). Oba přístupy umožňují sekvenovat peptid pouze do určité délky (maximálně několik desítek aminokyselin). Pokud chceme znát úplnou sekvenci proteinu, můžeme použít jeden z následujících postupů.

1. Pomocí specifických enzymů<sup>1</sup> můžeme připravit kratší peptidy, které je možno sekvenovat úplně. Máme-li alespoň dvě sady peptidů s vzájemně se překrývajícími sekvencemi, můžeme kombinací částečných sekvencí určit celkové pořadí aminokyselin proteinu.
2. V posledních letech je běžnější druhý postup, kdy určíme pouze částečnou sekvenci proteinu (například přibližně 15 aminokyselin *N*-konce proteinu). Z této informace je možné odvodit nukleotidovou sekvenci kódující příslušný protein (přesněji řečeno sekvenci s několika nejednoznačnými nukleotidy v důsledku degenerace genetického kódu). Pomocí této sekvence identifikujeme gen studovaného proteinu a sekvenaci provádíme na úrovni nukleových kyselin (metody sekvenace nukleových kyselin jsou rychlejší méně pracné než metody sekvenace proteinů).
3. Nejjednodušší postup je obejít se bez sekvenace proteinů docela a změřit pouze hmotnosti peptidů vzniklých štěpením enzymy (viz postup 1) pomocí hmotnostní spektrometrie. Pokud je známa genetická informace studovaného organismu, je možné pomocí počítačové analýzy sekvence DNA vypočítat, jestli změřené hmotnosti peptidů neodpovídají nějakému proteinu kódovanému DNA. Nalezne-li se více takových peptidů odpovídajících jednomu proteinu, je velká pravděpodobnost, že skutečně studujeme tento protein. Tato metoda, zvaná *peptidové mapování*, je velmi rychlá. Lze ji samozřejmě použít pouze tehdy, je-li známá sekvence DNA daného organismu, nebo jde-li o protein již dříve zkoumaný a sekvenovaný. Je také možné, že analýza nalezne velmi podobný protein jiného organismu.

---

<sup>1</sup>Jako příklad můžeme uvést trypsin štěpící na karboxylovém konci lysinu a argininu, nebo bakteriální endopeptidasy jako peptidasa V8 (z bakterie *Staphylococcus aureus*) štěpící na karboxylovém konci kyseliny glutamové.

## 9.2 Edmanovo odbourávání

Podstatou Edmanovy metody je adice fenylisothiokyanátu na koncovou aminoskupinu proteínu. Takto derivatizovanou aminokyselinu je možno selektivně hydrolyzovat a pomocí kapalinové chromatografie určit, které aminokyselině odštěpený fenylthiohydantoin odpovídal. Celý postup se provádí v automatických sekvenátorech. Úspěšnost Edmanova odbourávání vyžaduje, aby se v každém kroku odštěpila úplně právě jedna aminokyselina. Chyby způsobené nedokonalostí se hromadí a omezují délku peptidu, který je možné sekvenovat. Kdyby štěpení probíhalo z 90 %, což je na první pohled slušná účinnost, druhý krok by proběhl pouze z 81 % ( $0,9^2$ ) a při určování sedmé aminokyseliny už bychom správného produktu získali méně než polovinu ( $0,9^7 = 0,478$ ). V praxi se většinou stanovuje sekvence zhruba patnácti aminokyselin.

## 9.3 Sekvence hmotnostní spektrometrií

Analýza derivatizovaných aminokyselin odštěpovaných v jednotlivých krocích Edmanova odbourávání vyžaduje spolehlivou chromatografickou metodu. Existuje však jiná vlastnost, kterou se aminokyseliny liší a která je mnohem jednoznačněji definována, než chromatografické chování. Touto vlastností je *hmotnost*. Přesné měření tak malé hmotnosti, jako je hmotnost jednotlivých molekul, není samozřejmě jednoduché. Metody, které nám to umožňují, se souhrnně označují jako *hmotnostní spektrometrie* (*mass spectrometry*, zkratka *MS*) a byly vyvinuty už před mnoha lety<sup>2</sup>.

Hmotnostní spektrometrie je založena na sledování pohybu iontů v elektromagnetickém poli. V principu nejjednodušší je urychlit iont v elektrickém poli a pak měřit dobu, za kterou proletí trubicí známé délky. Protože víme, jak silným polem jsme iont urychlili, známe jeho kinetickou energii. Z doby průletu umíme spočítat rychlost iontu, takže určení hmotnosti z kinetické energi  $E = \frac{1}{2}mv^2$  již nic nebrání. Tato metoda se označuje zkratkou *TOF-MS* (*Time-Of-Flight Mass Spectrometry*). Jinou možností je sledovat, jak se zakřivuje dráha iontu v magnetickém poli. Zakřivení závisí na hmotnosti a náboji iontu, takže opět můžeme jednoduše spočítat hmotnost (známe-li náboj). Zvláštním způsobem měří hmotnost takzvaná *iontová past* (*ion trap*). Tato past je tvořena elektrodami, které vytvářejí kmitající elektrické pole. V tomto poli se udrží jen ionty určitého poměru hmotnosti k náboji. Během měření měníme pole tvořící past a sledujeme, které ionty se již v pasti neudrží a vyletí do detektoru.

Největším problémem hmotnostní spektrometrie proteinů (a jiných biomakromolekul) nebylo po dlouhá léta vlastní měření hmotnosti, ale vytvoření iontu. Biomakromolekuly jsou sice v roztoku přítomny jako ionty, ale je velmi těžké tyto ionty z roztoku uvolnit a poslat do hmotnostního detektoru. Dnes známe dvě velmi účinná řešení tohoto problému. Prvním řešením je takzvaný *elektrospray*. Jde o zajímavý úkaz známý již dlouhá léta. Přivádíme-li do oblasti silného elektrického pole roztok iontů, kapičky roztoku se odpařují (ztrácejí molekuly vody), takže se zmenšuje jejich povrch a roste poměr náboje k povrchu, povrchové napětí nestačí udržet kapku pohromadě, takže se třísťí na menší a menší kapičky, které se ještě rychleji odpařují a vše pokračuje tak dlouho, až se odpaří všechny molekuly vody a prostorem letí „nahý“ iont makromolekuly.

Jiným způsobem, jak získat ionty biomakromolekul, je odpařit roztok a vzniklý prášek ostřelovat záblesky laseru. V nejčastějším provedení jsou k biomakromolekulám přidávány organické kyseliny (takzvaná *matrice*), které absorbují energii laseru, chrání makromolekulu a dodávají jí kladný náboj. Tato metoda se nazývá *Matrix-Assisted Laser Desorption Ionization* (*MALDI*).

Pomocí popsaných metod hmotnostní spektrometrie, které umožňují ionizovat peptidy ve vodných roztocích nebo vykrytalizované s vhodnou maticí, je možno studovat fragmenty peptidů. Tyto fragmenty se buď připraví před analýzou<sup>3</sup>, nebo, častěji, vznikají během vlastního měření rozpadem mole-

<sup>2</sup>Přesněji řečeno, hmotnostní spektrometrie měří poměr hmotnosti a náboje nabitě molekuly.

<sup>3</sup>K tomu lze například použít enzymy, které nespecificky odštěpují aminokyseliny z jednoho konce, jako karboxypeptidasa Y.

kul excitovaných laserovou ionizací (takzvaný *post-source decay*) nebo srážkami s molekulami vhodného plynu (*collision-induced dissociation*).

Takto získaná hmotnostní spektra mají vzhled žebříčků a z rozdílu hmotností jednotlivých fragmentů lze odvodit sekvenci původního peptidu. Rozlišení jedné hmotnostní jednotky stačí k určení většiny aminokyselin. Odlišení přibližně izobarických<sup>4</sup> aminokyselin (lysin a glutamin) vyžaduje velmi vysoké rozlišení, zatímco izomerní aminokyseliny (leucin a isoleucin) není možné odlišit bez detailního studia fragmentace postranních řetězců.

Hmotnostní spektrometry detekují pouze nabitě částice (ionty). Dojde-li ke štěpení peptidového řetězce s nábojem +1, může kladný náboj zůstat buď na *N*-koncové části peptidu (takový štěp se označuje písmenem *b*), nebo na *C*-koncové části (štěp *y*). Analýza spekter je navíc komplikována tím, že ke štěpení hlavního řetězce může docházet nejen v místě peptidové vazby, ale i mezi uhlíky (vznikají štěpy *a* a *x*) nebo mezi dusíkem a  $\alpha$ -uhlíkem (vznikají štěpy *c* a *z*). Určování sekvence proteinů je proto často založeno na počítačovém vyhodnocení spekter, které může zahrnovat všechny uvedené komplikace. Příklad sekvenace krátkého peptidu (Ala-Thr-Gly-Cys) je ukázán na obrázku 9.1. Skutečné fragmentační spektrum, získané metodou MALDI-TOF-MS, ukazuje obrázek 9.2

Sekvenace proteinů pomocí hmotnostní spektrometrie je užitečná i v případech, kdy je známá sekvence DNA (gen proteinu). Pokud totiž dochází k *posttranslačním modifikacím* proteinu (viz sekci 2.2), lze tyto modifikace z rozdílu hmotností odhalit.

## 9.4 Sekvenace nukleových kyselin

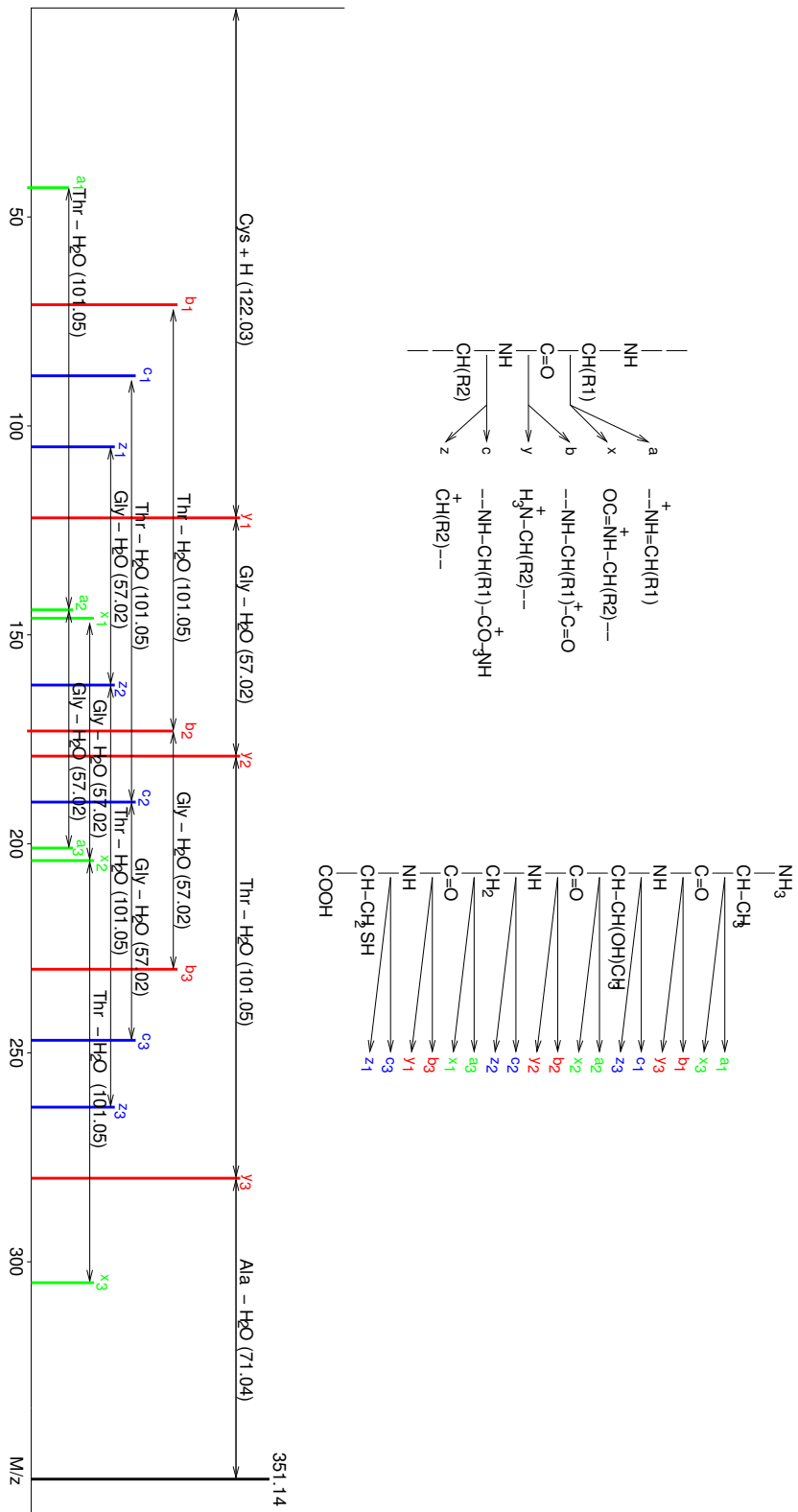
Sekvenace nukleových kyselin je založena na přípravě vzorku obsahujícího směs oligonukleotidů s jedním koncem stejným ale lišící se délkou. Taková směs je pak analyzována elektroforeticky. Podmínky tvorby směsi oligonukleotidů musí být zvoleny tak, aby v nich byly zastoupeny všechny délky. Výsledné elektroferogramy mají tvar žebříčků, ve kterých oligonukleotidy lišící se o délku jednoho rezidua mají dráhu, kterou urazí v gelu (nebo čas elektroforézy v kapiláře), odlišnou o dobře definovaný rozdíl. Vhodnou směs nukleotidů lze vytvořit buď částečným odbouráním nebo neúplnou syntézou. Obě možnosti budou popsány níže.

V principu lze pro sekvenaci nukleových kyselin použít hmotnostní spektrometrii jako u proteinů. V současné době je ale toto použití spíše výjimkou (níže popsaná enzymová sekvenace umožňuje určit sekvenci délky několika stovek nukleotidů, zatímco hmotnostní spektrometrie je limitována několika desítkami). Pokud potřebujeme zkoumat delší úseky nukleových kyselin, připravíme kratší překrývající se štěpy pomocí specifických endonukleas, podobně jako u proteinů.

## 9.5 Chemická metoda Maxama a Gilberta

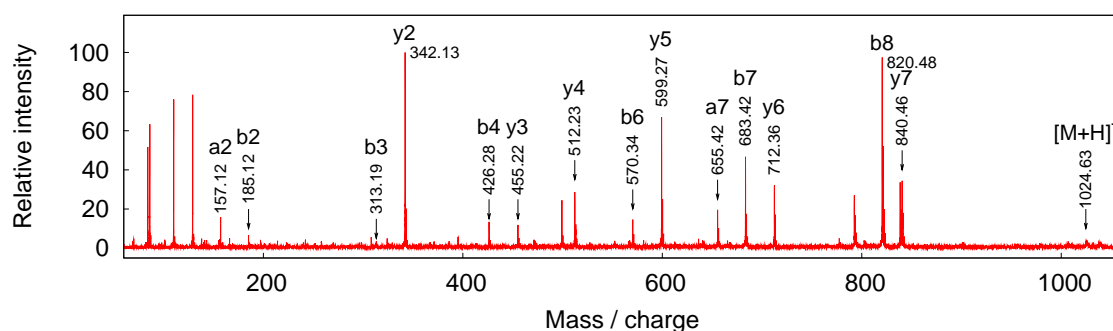
Při sekvenaci nukleových kyselin chemickou (odbourávací) metodou je nejprve třeba označit jeden konec vlákna, například 5'-konec radioaktivním fosforem <sup>32</sup>P. Poté potřebujeme čtyři chemické reakce, které přeruší vlákno v místech jednotlivých bází. Pro přerušování vlákna na pyrimidinech se používá štěpení pyrimidinového kruhu hydrazinem. Reaktivita bází klesá v pořadí uracil > cytosin > thymin, takže v případě sekvenování DNA lze zvolit podmínky, za kterých je štěpen cytosin i thymin nebo pouze cytosin (reakce thyminu je potlačena v přítomnosti 2M NaCl). V přítomnosti piperidinu pak dochází k selektivní hydrolyze fosfoesterové vazby nukleotidu s odbouraným pyrimidinovým kruhem. Při štěpení na purinech se nejprve báze methylojí dimethylsulfátem. Guanin se methyloje na dusíku

<sup>4</sup>Jako izobary se označují částice se stejnou hmotností. Pokud bychom předpokládali, že hmotnost atomu uhlíku je přesně dvanáctkrát větší než hmotnost atomu vodíku, hmotnost atomu dusíku je přesně čtrnáctkrát větší než hmotnost atomu vodíku a hmotnost atomu kyslíku je přesně šestnáctkrát větší než hmotnost atomu vodíku, byly by lysin a glutamin izobary. Protože se poměry hmotností atomů nepatrně liší od uvedených poměrů a protože lysin a glutamin obsahují různé počty atomů jednotlivých prvků, hmotnosti lysinu a glutaminu se mírně liší.



Obrázek 9.1: Příklad sekvenace tetrapeptidu Ala-Thr-Gly-Cys pomocí hmotnostní spektrometrie. Vlevo nahoře je znázorněn vznik různých fragmentů s nábojem (+1), vpravo nahoře přehled fragmentů, které mohou vzniknout štěpením jedné vazby našeho tetrapeptidu. Dole je zjednodušené spektrum s vyhodnocením rozdílů hmotností. Signál celého peptidu je znázorněn černě, fragmenty jsou obarveny podle schématu vpravo nahoře.





Obrázek 9.2: Fragmentační spektrum peptidu získané pomocí MALDI-TOF hmotnostní spektrometrie.

7, zatímco adenin se methyloje na dusíku 3. Hydrolyzou v zásaditém prostředí se přednostně štěpí *N*-glykosidická vazba s N7-methylguaninem, zatímco v kyselém prostředí se odštěpuje guanin i adenin. Fosfoesterová vazba se pak hydrolyzuje alkalicky při 90 °C. Popsané reakce se provádějí za podmínek zvolených tak, aby se každé analyzované vlákno štěpilo v průměru pouze jednou (pochopitelně na různém místě). Vzniknou tak čtyři sady radioaktivně značených oligonukleotidů, které při elektroforéze vytvoří žebříčky (obrázek 9.3), jejichž porovnáním lze snadno odečíst sekvenci studovaného vlákna DNA. Popsaná metoda je vzhledem ke své pracnosti používána málo, lze ji však přizpůsobit například výskytu nestandardních bází, kde selhává níže popsaná dideoxymetoda.

## 9.6 Enzymová neboli dideoxymetoda (Sangerova)

V současnosti je nejběžnější metoda sekvenace kratších úseků DNA (několik stovek nukleotidů) založená na enzymové syntéze oligonukleotidů, ve které sekvenovaný oligonukleotid slouží jako *templát*. Přídavkem vhodného množství dideoxyanalogů trifosfátů, ze kterých je oligonukleotid sekvenován (ddATP, ddGTP, ddTTP, ddCTP) je zajištěno náhodné předčasné ukončení syntézy oligonukleotidů<sup>5</sup> (tedy tvorba žebříčku v elektroferogramu, jak je ukázáno na obrázku 9.4). Sekvence, kterou při dideoxymetodě čteme, není sekvencí vlákna DNA použité jako templát, ale sekvencí komplementárního vlákna. Čtená sekvence tedy odpovídá již sekvenci mRNA, podle které se syntetizuje protein.

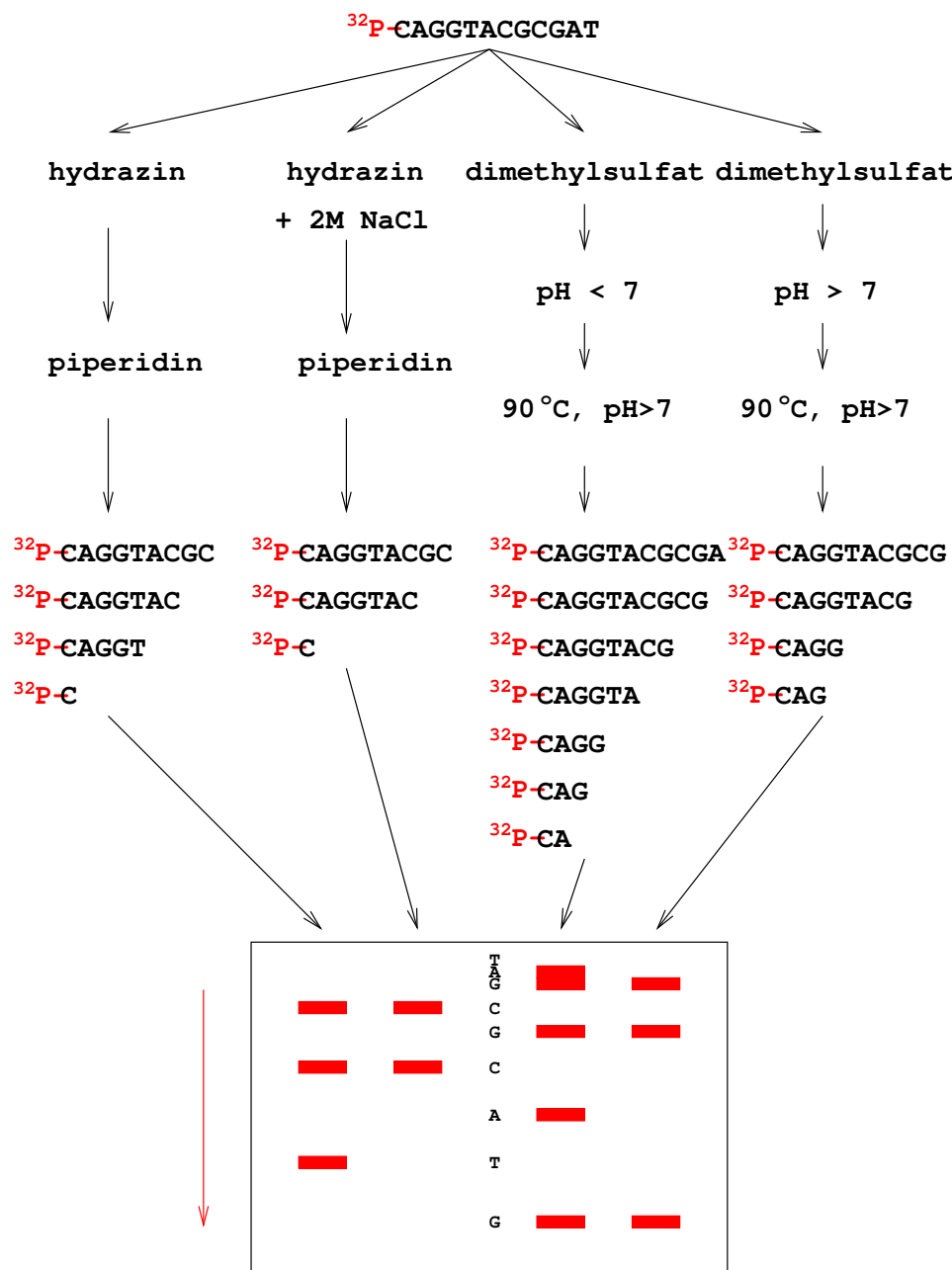
K dideoxyanalogům bývá navíc vázána nějaká fluoreskující sloučenina, kterou lze při elektroforéze snadno a citlivě detekovat. Pokud použijeme různé fluorescenční značky pro ukončení na jednotlivých bázích (G, A, T, C), můžeme sekvenaci provádět v jedné zkumavce. Obrázek 9.5 ukazuje, jak vypadá výsledek sekvenace s použitím fluorescenčních značek a dělením pomocí elektroforézy v kapiláře.

Sangerova metoda znamenala zásadní průlom v sekvenování nukleových kyselin a umožnila přečtení celého lidského genomu. Od 90. let 20. století byla vyvinuta celá řada metod, která sekvenování dlouhých úseků DNA, včetně celých genomů, výrazně urychlila. V našem základním textu se ale těmito metodami zabývat nebudeme.

## 9.7 Sekvenace oligosacharidů

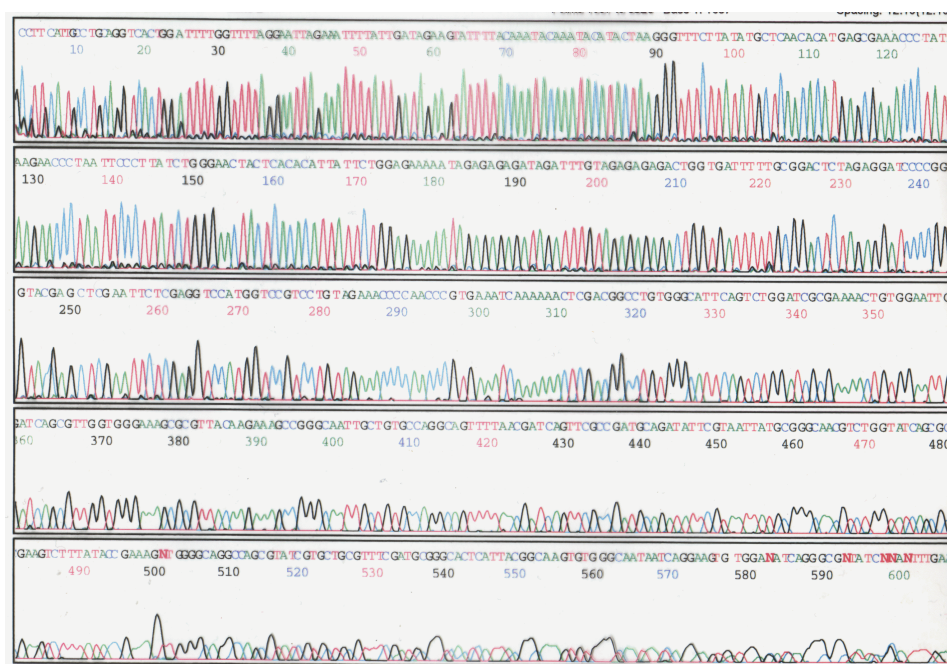
Sekvence oligosacharidů se většinou provádí pomocí hmotnostní spektrometrie. Často se kombinuje použití enzymů štěpících glykosidickou vazbu (glykosidas) specifických pro různé sacharidy a typy glykosidické vazby s fragmentací během měření. Analýza spekter je obdobná jako u proteinů, vzhledem ke

<sup>5</sup>Dideoxyanalog postrádá hydroxyl v poloze 3'. Proto po jeho zabudování nemůže dojít k dalšímu prodlužování řetězce.



Obrázek 9.3: Princip chemické metody Maxama a Gilberta. Ve spodní části obrázku je schematicky znázorněna poloha radioaktivních proužků na gelu, ve kterém byly elektroforézou rozděleny fragmenty vzniklé v jednotlivých reakcích. Šipka vlevo udává v jakém směru oligonukleotidy putovaly během elektroforézy.





Obrázek 9.5: Výstup automatického sekvenátoru využívajícího fluorescenční značky. Obrázek ukazuje záznam čtyř fluorescenčních detektorů, které snímají fluoreskující oligonukleotidy vycházející z elektroforetické kapiláry. Červeně je značen signál fluoroforu navázaného na ddTTP, modře na ddCTP, zeleně na ddATP, černě na ddGTP (obrázek poskytl Laboratoř molekulární fyziologie rostlin Přírodovědecké fakulty MU v Brně).

složitosti primárních struktur sacharidů je ale sekvenace oligosacharidů mnohem obtížnější.



# Kapitola 10

## Výpočetní metody

### 10.1 Cíl výpočetních metod ve strukturní biochemii

Jednou z nejužasnějších vlastností přírody je všeobecný řád. I ty nejsložitější děje a jevy ctí základní přírodní zákony. Dokonce můžeme říci, že pokud jde o chemii, tak ony nejzákladnější zákony již známe a umíme vyjádřit neobyčejně elegantním a praktickým jazykem matematiky. Držíme v rukou Kámen mudrců, klíč k strukturám všech molekul. Tím klíčem je energie. Jak jednoduchá je ta základní poučka: Ze všech stavů, ve kterých se může určité množství atomů vyskytovat, je nejvýhodnější ten stav, který má nejnižší energii. Zdálo by se tedy, že nic nebrání velice lákavému záměru: místo zdouhavého zkoumání struktury neznámé látky tuto strukturu prostě *vypočítat*. Vždyť stačí tak málo, pomocí rovnic, které dobře známe, vypočítat energii všech stavů našeho souboru atomů a z nich vybrat stav s nejnižší energií, matematicky řečeno, najít *globální minimum energie*. Tento stav s nejnižší energií, popsany souřadnicemi jednotlivých atomů, by měl být stabilní strukturou molekuly, kterou hledáme.

Ve strukturní biochemii nebude našim cílem počítat, která *konfigurace* atomů má nejnižší energii. Již sama existence izomerů naznačuje, že vedle sebe mohou stát molekuly s různou energií, aniž by se hned přeměnily v tu, která je z nich energeticky nejvýhodnější. Stavebnicové uspořádání biomakromolekul je také založeno na tom, že může existovat obrovské množství konfigurací lišící se pořadím základních stavebních kamenů. Toto pořadí neboli sekvenci určujeme experimentálně pomocí metod popsaných v kapitole 9.

Cílem strukturní biochemie je určit *konformaci* molekul s předem známou konfigurací. Experimentální metody určování konformace biomakromolekul jsou náročné a zdouhavé. Možnost *vypočítat* konformaci molekuly se známou konfigurací je proto lákavá. Výhody výpočtu ve srovnání s experimentálním určováním struktury jsou zřejmé na první pohled. Ušetříme práci, čas a peníze, které bychom museli vynaložit na experiment. Bez nebezpečí můžeme studovat molekuly jedovaté nebo nestabilní. Můžeme dokonce předvídat struktury molekul, které dosud nebyly vytvořeny. Navíc můžeme sledovat pohyby molekul, které jsou příliš rychlé nebo pomalé pro experimentální metody. Umíme ale takové výpočty provádět?

Pro malé molekuly je analýza konformací poměrně snadná. Výpočty energií nejsou sice tak jednoduché, jak by se mohlo na první pohled zdát, ale dobré počítače si s nimi celkem slušně poradí. Přechody mezi jednotlivými konformacemi jsou rychlé, takže můžeme předpokládat, že ve zkumavce máme takové počty jednotlivých konformerů, jaké předpovídají jejich energie. Vše se ale komplikuje, když přecházíme ke složitějším molekulám. Výpočty energií jsou složitější a složitější, takže musíme používat metody čím dál tím více zjednodušené a méně přesné. Přestože i u velkých molekul zpravidla hledáme konformaci s nejnižší energií, nemáme jistotu, že se v ní molekula opravdu nachází. Přechody mezi stavy s různou konformací jsou totiž pro velké molekuly obtížné a může se stát, že molekula po celý

svůj život setrvává v konformaci s vyšší energií. Následující stránky by měly ukázat, kam až ve výpočtech konformací biomakromolekul umíme dnes dojít a jakým způsobem se potýkáme s obtížemi, které ze složitosti biologicky zajímavých molekul plynou.

## 10.2 Modely biomakromolekul

Abychom odlišili náš popis struktury molekuly od skutečné molekuly, budeme vypočítanou strukturu nazývat *modelem molekuly*. Slovo *model* budeme používat v trochu obecnějším významu, než je konstrukce z trubiček a kuliček, kterou jsme nazvali modelem v sekci 1.1. Model bude pro nás popisem všech vlastností struktury, které nám výpočet poskytne. Samotný výpočet, kterým model získáme, můžeme označit jako *molekulové modelování*.

Modely používané ve výpočtech se značně liší v tom, jak věrně molekulu popisují. Nejbližší skutečnosti jsou modely, které se snaží postihnout chování jader i elektronů v atomech molekul. Takový popis vyžaduje použití *kvantové mechaniky*, protože na chování částice tak malé, jako je elektrony, se velmi silně projeví, že se řídí zákony kvantové mechaniky. *Kvantových metod* je celá řada. Některé jsou velmi přesné, jiné využívají různých zjednodušení, která urychlují výpočet. Ale i ty nejrychlejší kvantové metody jsou příliš pomalé pro rutinní výpočty konformací biologicky zajímavých makromolekul. Kvantové metody se používají hlavně k výpočtům chování malých molekul a ke kontrole správnosti jednodušších přístupů. Je také možné, ale zdaleka ne jednoduché, použít kvantovou mechaniku k popisu malé části biomakromolekuly, kde dochází k chemickým změnám, které kvantový popis vyžadují. Stručný přehled kvantových metod je v Dodatku C.

Náročnost kvantových výpočtů vede ke snahám o zjednodušení popisu molekul. My se budeme nejvíce věnovat modelům, ve kterých se snažíme věrně zachytit polohy jednotlivých atomů (přesněji jejich jader), ale vzdáváme se popisu elektronů. Bez elektronů nejsme samozřejmě schopni popsat fyzikálně správné chemické vazby. Musíme proto nalézt náhradní popis interakcí mezi atomy, jak si uvedeme v části 10.3. Pro takový popis se vžil název *molekulová mechanika*. Tento přístup zrychluje výpočet natolik, že můžeme zkoumat i celé proteiny a poměrně dlouhé úseky nukleových kyselin. Konformace biomakromolekul vypočítané molekulovou mechanikou pouze na základě sekvence nejsou natolik spolehlivé, abychom se obešli bez experimentálních metody určování struktury. Přesto hraje molekulová mechanika ve strukturní biochemii zásadní roli. Jednak dokáže dostatečně spolehlivě popsat chování molekul, pokud již máme představu o jejich struktuře. Dále umožňuje dobře předpovídat konformaci biomakromolekul na základě známých struktur molekul s podobnou sekvencí. Především ale molekulovou mechaniku využívají všechny takzvané experimentální metody určování struktur. Ve většině případů totiž experimentální chyby a neúplnost naměřených dat vyžadují, aby detaily výsledné struktury byly *vypočítány* spolehlivou metodou.

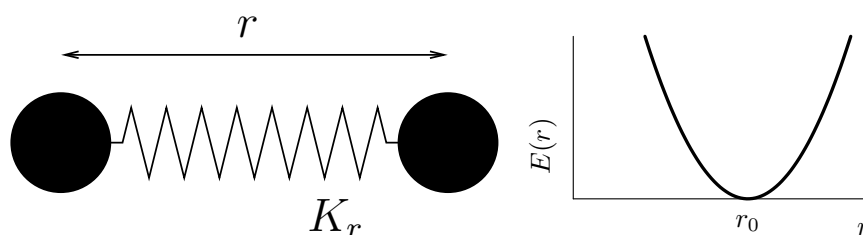
Výpočty můžeme urychlit ještě více, pokud polohy jednotlivých atomů nahradíme polohami větších částí molekuly. Takovéto *hrubozrné modely* používáme k popisu velkých struktur (například biologických membrán nebo kvartérních struktur).

## 10.3 Molekulová mechanika

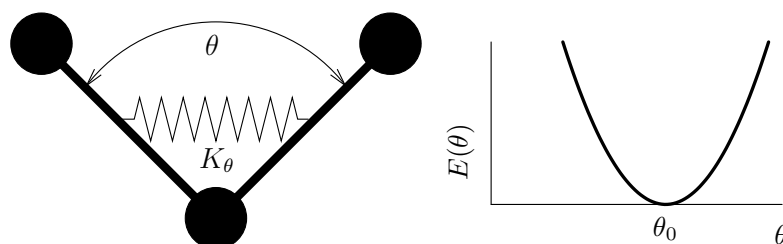
Rozdíl mezi kvantovou mechanikou a molekulovou mechanikou je ve vztahu k přírodním zákonům. Základní myšlenka kvantových výpočtů je: „Čím věrněji budou naše metody popisovat přírodní zákony, tím spíš bude jejich výsledek blízký skutečnosti.“ Naopak metody molekulové mechaniky se drží hesla: „Nestaráme se o to, jestli jsou naše metody v souladu s přírodními zákony, ale dbáme o to, aby jejich výsledek byl blízký skutečnosti“.

Pravidla molekulové mechaniky jsou sice svým způsobem umělá, nejsou ale vymyšlena bez ohledu na fyziku. Molekulová mechanika se opírá o popis klasické mechaniky Newtonovy. Místo abychom studovali





Obrázek 10.1: Zjednodušený popis energie spojené se změnami vazebné délky.



Obrázek 10.2: Zjednodušený popis energie spojené s deformací vazebného úhlu.

jak se chová molekula, vlastně studujeme, jak by se choval její model z kuliček spojených pružinkami. Všimněme si, jaký je to rozdíl od kvantové mechaniky. V obou případech oddělíme od sebe pohyby elektronů a jader. V kvantové mechanice se výpočet soustředí především na elektrony, protože právě elektrony tvoří vazby a určují chemické vlastnosti molekul. V molekulové mechanice naopak ignorujeme elektrony téměř úplně. Něčím je ovšem musíme nahradit.

Vzájemné působení atomů (ve skutečnosti zprostředkované elektrony) napodobujeme tím, že atomy různě spojujeme ideálními pružinkami (obrázek 10.1). Jak si jistě vzpomenete ze střední školy, energii dvou kuliček<sup>1</sup> spojených ideální pružinou<sup>2</sup> můžeme popsat rovnicí 10.1

$$E = K_r(r - r_0)^2, \quad (10.1)$$

kde  $r$  je okamžitá vzdálenost kuliček,  $r_0$  je rovnovážná vzdálenost kuliček a  $K_r$  je tuhost pružiny. V molekulové mechanice stačí dosadit za  $r_0$  optimální vazebnou vzdálenost a máme energetický popis dvojice atomů spojených kovalentní vazbou.

Energii spojenou s deformací vazebných úhlů můžeme popsat podobně (obrázek 10.2)

$$E = K_\theta(\theta - \theta_0)^2, \quad (10.2)$$

kde  $\theta$  je okamžitá hodnota vazebného úhlu a  $\theta_0$  je nejvýhodnější hodnota vazebného úhlu.

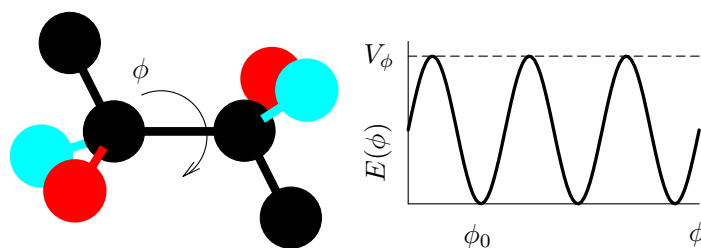
Na torzní úhel už s pružinkou nevystačíme. Musíme se ptát, jak se asi bude měnit energie, když budeme točit kolem chemické vazby. Na základě jednoduchých úvah<sup>3</sup> dospějeme k rovnici 10.3

$$E = \frac{V_\phi}{2}(1 + \cos n(\phi - \phi_0)), \quad (10.3)$$

<sup>1</sup>Přesněji dvou hmotných bodů, tedy kuliček nekonečně malých.

<sup>2</sup>Z pohledu matematiky se taková pružina chová jako *harmonický oscilátor*.

<sup>3</sup>Kdybychom měli například dvě velké skupiny spojené disulfidovou vazbou, budeme očekávat, že nejvýhodnější bude torzní úhel  $\phi = \pm 180^\circ$  (*trans*) a nejméně výhodný torzní úhel  $\phi = 0^\circ$  (*cis*). Takový průběh energie by dobře popsala funkce  $\frac{1}{2}(1 + \cos \phi)$ . V případě ethanu budou nejvýhodnější torzní úhly  $\phi = \pm 180^\circ$  (*trans*) a  $\phi = \pm 60^\circ$  (*gauche*) a nejméně výhodné torzní úhly  $\phi = \pm 120^\circ$  a  $\phi = 0^\circ$ . Tomuto průběhu by odpovídala nejlépe funkce  $\frac{1}{2}(1 + \cos 3\phi)$ .



Obrázek 10.3: Příklad zjednodušeného popisu energie spojené se změnou torzního úhlu.

kde  $V_\theta$  je energie nejméně výhodné konformace,  $n$  požadovaný počet minim energie,  $\phi$  okamžitá hodnota torzního úhlu a  $\phi_0$  optimální hodnota torzního úhlu (obrázek 10.3).

Zde je dobré si uvědomit, v čem se model molekuly popsaný rovnicemi 10.1 až 10.3 liší od fyzického modelu z kuliček a tyčinek. V případě fyzického modelu jsou torzní úhly omezeny jen požadavkem, aby do sebe kuličky přímo nenarážely. Naopak vazebné délky a úhly jsou přesně dané. Takovým pevně nastaveným vnitřním souřadnicím se říká *tvrdé vazné podmínky*. V modelu popsaném rovnicemi 10.1 až 10.3 nejsou optimální hodnoty vazebných délek a úhlů takto striktně fixovány, ale pouze zvýhodněny nižší energií. Takové omezení některých souřadnic takzvanými *měkkými vaznými podmínkami* se anglicky nazývá *restraints*<sup>4</sup>. Měkké vazné podmínky lépe popisují skutečné chování molekuly, ale vyžadují, aby se vazebné délky a úhly pro každou konformaci počítaly. Když chceme urychlit výpočet, můžeme využít tvrdé vazné podmínky. Za vazebné délky a úhly dosadíme hodnoty, které známe dostatečně přesně z experimentu a nezajímají nás možné odchylky od těchto hodnot. Tím přestanou být vazebné délky a úhly souřadnicemi (jsou to předem známé konstanty) a roli souřadnic hrají pouze torzní úhly.

Umělé vztahy představované rovnicemi 10.1 až 10.3 molekulová mechanika kombinuje s elektrickými interakcemi mezi atomy, které jsme popsali v kapitole 6. Elektrické síly můžeme popsat rovnicemi shrnutými v dodatku A. Nejčastěji se používá elektrostatická energie dvou nábojů, která klesá s jejich vzdáleností (rovnice 15.2) a energie disperzních sil, které klesají se šestou mocninou vzdálenosti atomů (rovnice 15.93). Mezi atomy navíc působí velmi silné odpudivé síly (Pauliho repulze), které brání jejich kontaktu. Tyto síly zjednodušeně popisujeme funkcí klesající s dvanáctou mocninou vzdálenosti atomů. Silová působení popsaná v tomto odstavci můžeme shrnout do rovnice 10.4

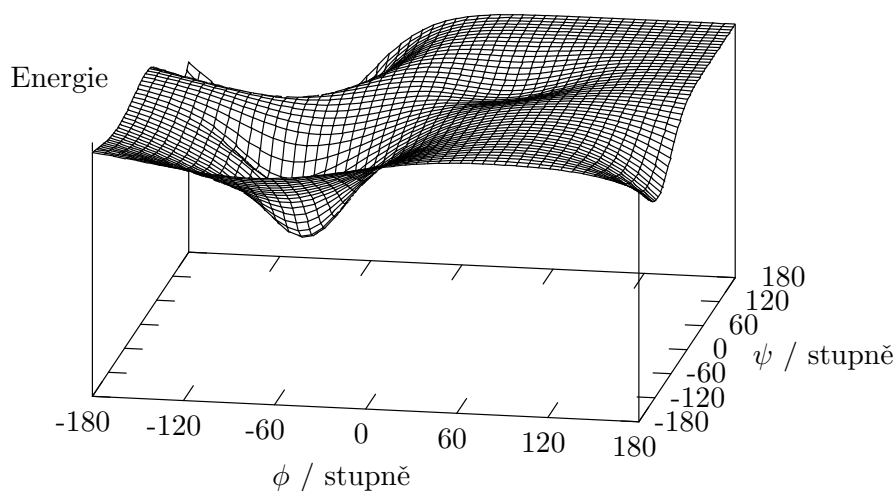
$$E = \frac{A}{r^{12}} - \frac{B}{r^6} + C \frac{Q_1 Q_2}{r}, \quad (10.4)$$

kde  $Q_1$  a  $Q_2$  jsou náboje atomů,  $r$  jejich vzdálenost a  $A$ ,  $B$ ,  $C$  konstanty úměrnosti.

## 10.4 Hyperplocha potenciální energie

Všimněme si, že energie vyjádřená v rovnicích 10.1 až 10.4 vždy závisí buď na vzdálenosti atomů, nebo na vazebném či torzním úhlu – prostě na nějakých souřadnicích atomů. Kdybychom vzali jen dvě takové souřadnice, třeba úhly  $\phi$  a  $\psi$  peptidové vazby, mohli bychom si přímo graf energie jako funkce souřadnic zobrazit. Na trojrozměrném grafu, jehož základnou by byl Ramachandranův diagram, bychom viděli zprohýbanou plochu připomínající krajinu. Údolí by odpovídala výhodným konformacím s minimální energií a pohoří by představovala maxima energie (obrázek 10.4). Tato plocha nad Ramachandranovým grafem se nazývá *hyperplocha potenciální energie* (zkratka *PES* z „potential energy surface“). V makromolekule máme souřadnic mnohem více, takže plochu bychom museli kreslit ne nad dvourozměrným diagramem, ale nad mnoharozměrným prostorem (takzvaný *konformační prostor*). To sice neumíme, ale matematickému určení takové nepředstavitelné hyperplochy a hledání jejich minim nic nebrání.

<sup>4</sup> *Restraint* je omezení, nátlak.



Obrázek 10.4: Ilustrace hyperplochy potenciální energie nad Ramachandranovým diagramem. Nejbližší hluboké údolí odpovídá konformaci  $\alpha$ -struktur, za sedlem se skrývá údolí odpovídající konformaci v  $\beta$ -lístech. Graf znázorňuje závislost potenciální energie proteinu na změnách konformace jedné z aminokyselin.

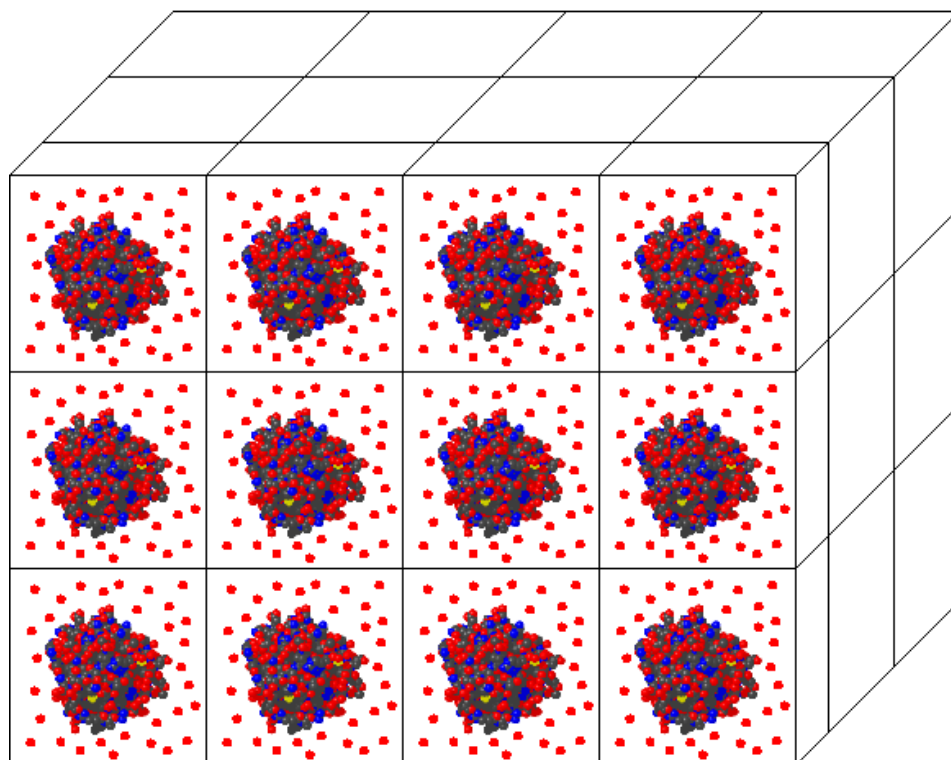
## 10.5 Silové pole

Celkovou energii atomů v molekule tedy můžeme popsat jako součet příspěvků jednotlivých atomů podle rovnic 10.1 až 10.4. K tomu, abychom takový součet uměli spočítat, musíme samozřejmě znát číselné hodnoty všech konstant  $K$ ,  $V$ ,  $A$ ,  $B$ ,  $C$ ,  $r_0$ ,  $\theta_0$ ,  $\phi_0$ . Odkud je ale vzít, když jsme molekulu popsali zcela uměle? Jak máme vědět, jakou musí mít tuhost fiktivní pružinka spojující atomy v našem modelu, když víme, že v přírodě atomy pružinkami spojeny nejsou? Máme dvě možnosti, jak tento problém vyřešit. Můžeme změřit energii odpovídající kmitání atomů a do rovnice 10.1 dosadit za  $K_r$  takové číslo, aby nám vyšla stejná energie jako změřená. Nebo můžeme energii kmitání atomů spočítat na základě kvantové mechaniky pro malou molekulu a získat  $K_r$  obdobným postupem. V obou případech musíme samozřejmě předpokládat, že v naší molekule kmitají atomy stejně jako v molekule, kde jsme kmitání měřili, nebo jako v malé molekule, pro kterou jsme prováděli kvantový výpočet. Optimální vzdálenosti atomů zase můžeme získat z rentgenové strukturní analýzy krystalů, jak o tom bude řeč v kapitole 12.

Soubor rovnic 10.1 až 10.4 a číselné hodnoty parametrů  $K$ ,  $V$ ,  $A$ ,  $B$ ,  $C$ ,  $r_0$ ,  $\theta_0$ ,  $\phi_0$  pro seskupení atomů, která nacházíme v biomakromolekulách, se nazývají *silové pole* a představují srdce každé metody molekulové mechaniky. Postup, kterým se z experimentu nebo z kvantových výpočtů určují parametry  $K$ ,  $V$ ,  $A$ ,  $B$ ,  $C$ ,  $r_0$ ,  $\theta_0$ ,  $\phi_0$ , se označuje pojmem *parametrizace silového pole*.

## 10.6 Implicitní a explicitní solvent

Konformace biomakromolekul je do značné míry určena prostředím, ve kterém se nachází. Zjednodušený popis energie umožňuje molekulové mechanice výpočty nejen pro celé biomakromolekuly, ale i pro nejbližší okolí těchto molekul. Tak je možno studovat biologicky zajímavé molekuly v jejich přirozeném



Obrázek 10.5: Periodické okrajové podmínky pro výpočet struktury proteinu s explicitním solventem (znázorněn červenými tečkami).

prostředí, ve vodném roztoku. Tento přitažlivý úkol s sebou pochopitelně nese otázku, jakým způsobem vodu do výpočtu zahrnout. Vliv rozpouštědla se popisuje jednou ze dvou metod. První z nich, zvaná metoda *implicitního solventu*, je výpočetně velmi snadná. Využívá toho, že síla elektrických interakcí je nepřímo úměrná elektrické permitivitě prostředí. Nejjednoduššího přiblížení vodnému prostředí je tak možné dosáhnout vynásobením hodnoty elektrické permitivity vakua  $\epsilon_0$  v Coulombově zákoně (rovnice 6.2) vhodnou konstantou, zvanou relativní permitivita  $\epsilon_r$ . Používá se buď hodnota  $\epsilon_r$  konstantní, všude stejná, nebo závislá na vzdálenosti atomů. Výhody i úskalí tohoto přístupu jsme už zmínili v části 6.10.

Druhý způsob zahrnutí rozpouštědla je mnohem přímočařejší. K modelu biomakromolekuly se přímo přidávají modely molekul vody. Tato metoda *explicitního solventu* je samozřejmě výpočetně náročná, počet atomů, se kterými se počítá, musí zahrnovat i atomy vody. Není možné obalit makromolekulu jen tenkou vrstvou molekul vody, vždyť v přírodě je voda v celém prostoru buňky. Jak ale takové prostředí namodelovat, aby se nám molekuly vody nerozprchly do prostoru, nebo aby se neprojevyly vlivy stěn, do kterých bychom je uzavřeli? Elegantní řešení nabízejí takzvané *periodické okrajové podmínky*. Kolem makromolekuly naprogramujeme krabici, jejíž vnitřek pravé stěny se bude zároveň chovat jako vnější strana levé stěny. Molekula vody, která z krabice uteče pravou stěnou do ní zároveň levou stěnou znovu vstoupí. Totéž bude platit pro přední a zadní stěnu a pro dno a víko (obrázek 10.5). Takové zázračné chování lze matematicky jednoduše zařídit. Vlastně odpovídá tomu, jako kdybychom měli identické krabice s makromolekulou pravidelně poskládané vedle sebe ve všech směrech.

## 10.7 Výpočet elektrostatických interakcí

Další komplikací, kterou je třeba vyřešit pro rychlý výpočet, je daleký dosah elektrostatických interakcí. V makromolekule interaguje každý atom s každým. Počet výpočtů elektrostatických interakcí by měl být tedy přibližně úměrný druhé mocnině počtu atomů, což je pro makromolekulu velké číslo. Výpočet se výrazně zrychlí, když zanedbáme interakce vzdálených atomů. To můžeme bez obav udělat pro van der Waalovy síly, které se vzdáleností rychle klesají. Jak ukazuje rovnice 15.2 v Dodatku, elektrostatické interakce klesají se vzdáleností velmi pomalu. „Uříznutí“ interakcí v určité vzdálenosti způsobí nepřirozený skok v hodnotách energie, který vede k výpočetním chybám. Pokud se budeme snažit posunout hranici zanedbání co nejdále a zároveň používáme metodu explicitního solventu, dostáváme se do dalších obtíží. Hranice zanedbání bude větší než rozměry krabice, takže budeme občas počítat interakce atomu vody se sebou samým.

Rozumný kompromis mezi přesností a rychlostí výpočtu nabízí metoda *PME (Particle Mesh Ewald)*. Tato metoda je vhodná zejména pro výpočty nukleových kyselin, v nichž každý nukleotid obsahuje záporně nabitý fosfát a elektrostatické síly se proto projevují velmi výrazně.

Metoda PME zjednodušuje výpočet tím, že elektrostatické interakce počítá pro jednotlivé molekuly vody jen do určité vzdálenosti od biomakromolekuly. Ve větších vzdálenostech využívá skutečnosti, že energie elektrostatické interakce je úměrná *převrácené hodnotě vzdálenosti*. Díky tomu, že naše molekuly jsou umístěny v periodicky se opakujících krabicích, můžeme pro výpočty veličin závisících na převrácené hodnotě vzdálenosti použít velmi rychlou výpočetní metodu založenou na *Fourierově transformaci*<sup>5</sup>.

## 10.8 Hledání lokálního minima energie

Základní myšlenkou molekulového modelování je hledání stavu molekuly s nejnižší energií. Energie molekuly závisí na množství proměnných, které popisují stav atomů v molekule. Proto s energií zacházíme jako s *funkcí mnoha proměnných*. Kdyby energie závisela jen na jedné proměnné, našli bychom minimum snadno podle následujících pouček:

- Jestliže je první derivace funkce pro určitou hodnotu proměnné rovna nule, pak má funkce v daném bodě minimum, maximum, nebo inflexní bod.
- Jestliže je druhá derivace funkce pro stejnou hodnotu proměnné větší než nula, má funkce v daném bodě minimum (záporná druhá derivace odpovídá maximu a nulová druhá derivace odpovídá inflexnímu bodu).

Pokud jde o funkci více proměnných, nestačí k hledání minima jediné číslo. Musíme počítat derivace funkce podle všech proměnných. Průběh funkce v určitém bodě popisují vektory, které obsahují první a druhé derivace v daném bodě podle všech souřadnic. Matematicky se vektor prvních derivací označuje jako *gradient* a vektor druhých derivací jako *hessián*.

Gradient je pro hledání minima užitečný, protože ukazuje směr nejprudšího růstu funkce. Minimum poznáme podle toho, že v daném bodě bude gradient nulový a ve všech okolních bodech bude gradient směřovat pryč od daného bodu<sup>6</sup>. Když se vydáme z různých bodů proti směru gradientu, měli bychom skončit ve stejném bodě, v minimu. Pokud tomu tak opravdu je, říkáme, že náš výpočet *konvergoval*.

<sup>5</sup>Fourierova transformace převádí pomocí sumace (součtu) periodickou funkci proměnné  $x$  na funkci proměnné  $1/x$ . Pro tento výpočet existuje velmi rychlý algoritmus. S Fourierovou transformací se setkáme později při popisu zpracování difrakčních dat a signálu nukleární magnetické rezonance.

<sup>6</sup>Vedle metod založených na derivacích, jako je gradientová metoda, je možné použít i nederivační metody, jako je metoda simplexová (na počátku náhodně zvolíme proměnné, ty potom měníme a sledujeme, které změny vedou ke snížení energie). Volba nejvhodnější metody závisí na počtu proměnných, na tom jak jsme daleko od minima a na tom, jak citlivě závisí energie na jednotlivých proměnných. V průběhu minimalizace bývá užitečné metody střídát.

V praktickém výpočtu samozřejmě nedosáhneme naprostou shodu a neměnnost (nulovou derivaci). Proto musíme stanovit hranici, pod kterou už jsme ochotni odchylky zanedbat a říci, že jsme dostatečně blízko minima. Podmínky, které tuto hranici definují, se nazývají *konvergenční kritéria*.

## 10.9 Prohledávání plochy potenciální energie

Hledání lokálního minima má jedno důležité omezení: „umí běhat jen z kopce“. To znamená, že metoda dospěje vždy na dno toho údolíčka hyperplochy potenciální energie, ve kterém se už od začátku nachází. Nás ovšem zajímají všechna údolí, zejména to nejhlubší. Proto se minimalizace popsaná v části 10.8 často kombinuje s metodami, které nás umí přenést i přes pohoří potenciální energie. Tyto metody připomínají postup detektiva, který hledá ve městě podezřelého. Může k tomu přistupovat různým způsobem:

1. Detektiv může postupovat systematicky, dům od domu. Tomu odpovídá metoda *systematického prohledávání (grid search)*. Jako příklad si můžeme vzít opět Ramachandranův diagram. Budeme systematicky zvyšovat úhly  $\phi$  a  $\psi$  po pěti stupních a pro každou dvojici úhlů vypočítáme energii. Abychom prohledali celý diagram, musíme takových výpočtů provést 72 krát 72, to je 5 184.
2. Detektiv může zcela náhodně brát za kliky různých domů. Podobně my můžeme volit hodnoty úhlů  $\phi$  a  $\psi$  náhodně, hodit si vždycky kostkou nebo zatočit ruletou<sup>7</sup>. Celkem trefně se tyto metody označují *Monte Carlo*.
3. Detektiv může postupovat inteligentně a prohledat nejdříve nejpodezřelejší místa. Odpovídající metody se označují přízviskem *Single-Coordinate Driving (SCD)*. Na hyperploše potenciální energie nelozíme tvrdošíjně cesta-necesta přes všechny vrcholky, ale hledáme nejpohodlnější průsmyky. Neprozkoumáme, pravda, celý Ramachandranův diagram, ale projdeme tu část, která je pro chemika nejzajímavější (stabilní konformace). Čas, který ušetříme, můžeme věnovat přesnějším výpočtům.
4. Detektiv může sledovat pohyb ve městě a čekat, až pachatel sám dorazí na místo činu<sup>8</sup>. Takovým způsobem funguje *molekulová dynamika*, kterou se budeme zabývat v příští části.

## 10.10 Molekulová dynamika

Molekulová dynamika je metoda napodobování pohybů v molekule. Pomocí této metody můžeme nejen prohledávat konformační prostor, ale studovat i samotné molekulové pohyby. Věrna heslu molekulové mechaniky, molekulová dynamika nevychází z pohybové rovnice kvantové chemie (Schrödingerova rovnice), ale používá jednoduššího popisu klasickými Newtonovými pohybovými rovnicemi. Postup výpočtu probíhá podle následujícího scénáře.

1. Metoda začíná s modelem molekuly, který je popsán nějakými počátečními souřadnicemi atomů v molekule.
2. Silové pole nám řekne, jaká je energie celé molekuly v místě určeném danými souřadnicemi atomů.
3. Vypočítáme síly, které působí na jednotlivé atomy (síla je derivace energie podle souřadnice).

<sup>7</sup>V praxi asi použijeme funkci počítače, která generuje pseudonáhodná čísla.

<sup>8</sup>Problémem tohoto přístupu je, že detektiv místo činu dopředu nezná. Může sedět za komínem a rozhlížet se, kde se co děje, část výhledu mu ale budou zakrývat nedaleké mrakodrapy. Pokud bude pachatel loupit v jejich skrytu, zůstane neodhalen.

4. Podle Newtonova zákona síly ( $F = ma$ ) vypočítáme zrychlení atomu. Zrychlení je ovšem změna rychlosti za časový úsek a rychlost zase změna souřadnice za časový úsek. Když zvolíme dostatečně krátký časový úsek (*časový krok*), můžeme předpokládat, že se *v průběhu kroku* rychlost atomů nemění, a snadno vypočítat změnu souřadnic. Rychlost přitom udává kinetickou energii atomu v daném čase.
5. Nové souřadnice použijeme místo počátečních v kroku 1 a celý postup opakujeme. Síly, zrychlení, rychlosti a souřadnice se mění po skocích (celý postup se označuje jako *leap-frog algorithm*).

Pro molekulovou dynamiku je velmi důležitá volba správné délky časového kroku. Zvolíme-li krok příliš krátký, daleko v *konformačním prostoru* souřadnic nedorazíme. Příliš dlouhý krok nás naopak vystavuje nebezpečí, že se atomy budou pohybovat konstantní rychlostí příliš dlouho a dostanou se daleko od rovnováhy. V příštím kroku se budou snažit chybu napravit, nasadí rychlost opačným směrem, díky dlouhému kroku se dostanou ještě dál od rovnováhy opačným směrem. Molekula se tak nepřírozeně rozkmitá, dojde ke srážkám atomů a posléze celý výpočet zhavaruje. Vhodný časový krok by měl být proto výrazně kratší, než doba vlastních vibrací chemických vazeb. Protože vazba C–H kmitá s periodou asi 10 fs, používá se krok dlouhý jednu femtosekundu ( $10^{-15}$  s).

V kroku 1 našeho scénáře jsme zmínili, že výpočet zahajujeme s nějakými počátečními souřadnicemi. Tyto počáteční souřadnice, které definují startovní konformaci, musíme odněkud vzít. Za počáteční konformaci můžeme zvolit buď konformaci, kterou už známe z experimentu (třeba nepřesně), nebo náhodně sbalenou molekulu.

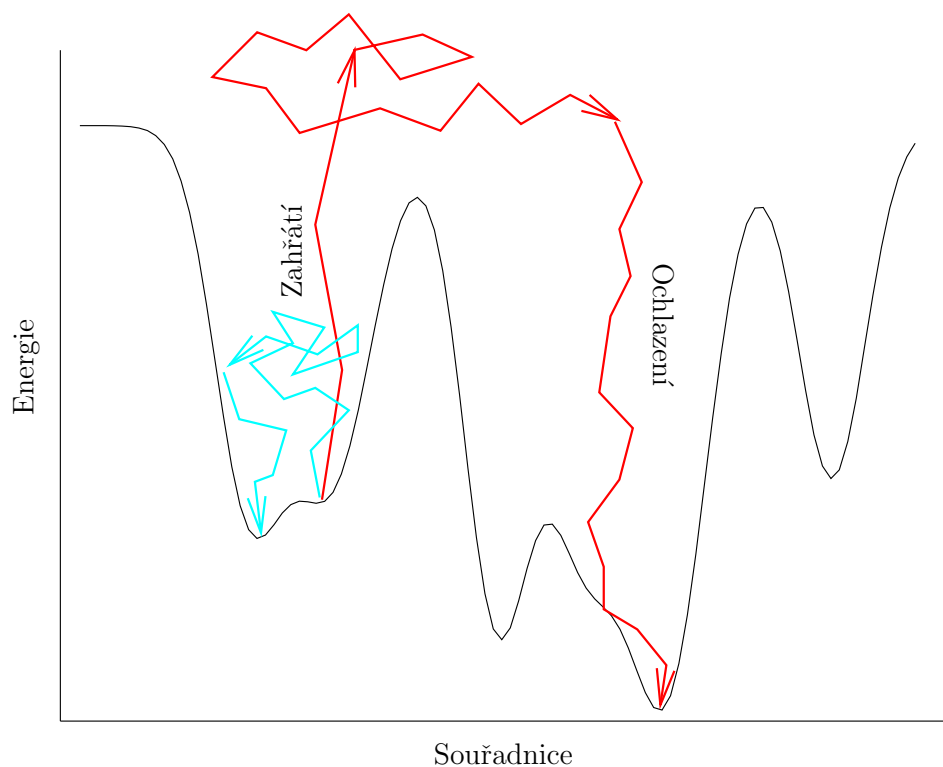
V kroku 4 jsme popsali, jak ze zrychlení určíme, o kolik se změnila rychlost atomů. Jakou rychlostí se ale pohybovaly atomy na začátku? Zde si pomůžeme tím, že pohyby atomů a molekul se navenek projeví jako *teplota*. Pokud předpokládáme, že molekula je v takzvané *tepelné rovnováze*, můžeme vypočítat statistické rozdělení rychlostí podle Maxwellovy-Boltzmannovy rovnice<sup>9</sup>. Jako hodnoty počátečních rychlostí tedy můžeme použít náhodné rychlosti, které splňují Maxwellovo-Boltzmannovo rozdělení pro zvolenou teplotu. Teplotu můžeme využít i během výpočtu. Po každém kroku nebo po několika krocích můžeme zkontrolovat vypočtené teploty a vynásobit každou z nich takovým číslem, aby výsledek odpovídal požadované teplotě. Tak vlastně donutíme molekulu udržovat stejnou teplotu, proto se této úpravě rychlostí říká *termostat*. Podobně udržujeme i stálý tlak (*barostat*).

Jak může takový výpočet molekulové dynamiky vypadat v praxi? Na začátku si zvolíme startovní konformaci biomakromolekuly, přidáme k ní molekuly vody a iontů (aby byl celý soubor navenek neutrální), vypočítáme počáteční teploty. Potom necháme makromolekulu v klidu, minimalizujeme energii vzhledem k souřadnicím atomů molekul vody a iontů. Necháme molekuly vody a ionty ustálit v krátkém výpočtu molekulové dynamiky. Takovým předběžným výpočtům říkáme *ekvilibrace systému*. Získáme tak (dosud nehybnou) makromolekulu obklopenou vhodně uspořádaným roztokem iontů. Potom zahájíme výpočet dynamiky biomakromolekuly. Nejprve postupně zvyšujeme teplotu a uvolňujeme souřadnice makromolekuly. Po dosažení cílové teploty necháme molekulu nějakou dobu ustálit a pak sledujeme změny konformace. Tuto část výpočtu označujeme jako *produkční fázi*. Průběhu změn konformace, který představuje putování v mnohazměrném prostoru, říkáme *trajektorie*.

Konformaci můžeme sledovat prostýma očima, když si co chvíli (zpravidla po několika stovkách časových kroků) souřadnice atomů zapíšeme a necháme počítač nakreslit model struktury. Každý obrázek modelu představuje snímek makromolekuly v určitém čase, spojením těchto snímků vznikne film pohybující se makromolekuly. Jiným způsobem popisu trajektorie je graf, do kterého vynášíme, jak se mění některý údaj o molekule v čase. Tímto údajem může být energie nebo střední kvadratická odchylka určité souřadnice.

Pohyblivost molekuly jsme si v úvodu této sekce zavedli proto, abychom se mohli pohybovat v celém konformačním prostoru, nejen směrem k nejbližšímu údolí. Dynamika molekul, tak jak jsme si ji

<sup>9</sup>Podle Maxwellovy-Boltzmannovy rovnice je pravděpodobnost, že se atom o hmotnosti  $m$  pohybuje při teplotě  $T$  rychlostí  $v$  úměrná  $v^2 T^{-\frac{3}{2}} e^{-\frac{mv^2}{2k_B T}}$ , kde  $k_B$  je Boltzmannova konstanta.



Obrázek 10.6: Princip metody simulovaného žhání. Při zahřátí na vysokou teplotu (červená čára) získá molekula větší kinetickou energii než nejvyšší hradby potenciální energie a může dospět do konformace odpovídající globálnímu minimu energie. Naopak při laboratorní teplotě (azurová čára) může molekula dospět pouze do konformace odpovídající lokálnímu minimu.

v předchozích odstavcích popsali, za rozumných teplot nestačí k překonání vysokých energetických pohoří. Proto se často provádí takzvané *simulované žhání* molekul, znázorněné na obrázku 10.6. To znamená, že se teplota zvýší na obrovskou hodnotu, často tisíce Kelvinů. Atomy se při této teplotě pohybují tak rychle, že jejich kinetická energie je vyšší než jakékoli pohoří na hyperploše potenciální energie. S takovou energií můžeme nad hyperplochou volně poletovat. Potom molekulu postupně ochlazuje, kinetická energie klesá, až „přistane“ v nějakém údolí na hyperploše, často daleko od startovní konformace. Za biologicky přijatelných teplot se pak molekula pohybuje v rámci nového údolí, jehož dno můžeme nalézt minimalizací energie. Když takový výpočet opakujeme s mírně odlišnými startovními podmínkami, je velká naděje, že skončíme v různých údolích. Takto můžeme prohledat velkou část konformačního prostoru a snad najít i konformaci s nejnižší energií.



# Kapitola 11

## Optické metody

### 11.1 Elektromagnetické vlnění

Cílem tohoto krátkého shrnutí principů optické spektroskopie je poskytnout základní přehled tomu, kdo zatím neměl příležitost do tajů světa molekul a elektromagnetických polí proniknout. Hned na začátku si musíme uvědomit, že vstupujeme do světa, kde se naplno projeví kvantové jevy, na které nejsme z běžného života zvyklí. Odpovídající fyzikální popis (*kvantová elektrodynamika*) by si vyžádal mnohem delší povídání. Zde se pokusíme získat alespoň hrubou představu, proč je možné dozvědět se něco o stavbě biomakromolekul z pozorování jejich interakcí s elektromagnetickým zářením.

Pokusme se nejdříve trochu popsat základní pojmy. Tedy, co je to světlo? Jistě víte, že světlo je elektromagnetické vlnění<sup>1</sup>. Tedy oscilace pole elektrických a magnetických sil, které se šíří v prostoru.

Elektrické a magnetické síly ovšem závisí nejen na tom, jak je silné pole, ale také na tom, na jak velký náboj, proud, případně elektrický nebo magnetický dipól pole působí. Proto je užitečné místo síly používat veličin vztahovaných na jednotku náboje či proudu. Elektromagnetické pole můžeme popsat čtyřmi takovými veličinami: elektrickou intenzitou  $\vec{E}$  a indukcí  $\vec{D}$  a magnetickou intenzitou  $\vec{H}$  a indukcí  $\vec{B}$ . Tyto veličiny nejsou nezávislé, ale musí splňovat takzvané Maxwellovy rovnice. Jedním z řešení Maxwellových rovnic je rovnice kulové vlny, tedy vlny, která se šíří do všech směrů v prostoru. Pokud nás zajímá šíření vln ve volném prostoru (vakuu) bez elektrických a magnetických dipólů, vystačíme si s veličinami  $\vec{E}$  a  $\vec{B}$ , protože ve vakuu  $\vec{D} = \epsilon_0 \vec{E}$  a  $\vec{B} = \mu_0 \vec{H}$  ( $\epsilon_0$  je elektrická permitivita vakua a  $\mu_0$  je magnetická permeabilita vakua).

Co je zdrojem elektromagnetické vlny? Asi nejlépe si můžeme představit zdroj vln srovnatelných s tělesnými rozměry. Tak dlouhé jsou vlny radiové<sup>2</sup>. Jedním z nejjednodušších zdrojů radiových vln je *dipólová anténa*. V nejjednodušším provedení tvoří dipólovou anténu kus rovného drátu, uprostřed přerušovaný a připojený na zdroj střídavého napětí. Taková anténa je nakreslena na obrázku 11.1.

Zdroj na obrázku ihned po zapnutí zdroje žene elektrony drátem směrem dolů. Pohybu elektronů se říká elektrický proud a ten vytváří magnetické pole. Na obrázku je nakresleno, jak vypadá takové magnetické pole v rovině kolmé na drát, přibližně v poloviční výšce. Vektory magnetické indukce v bezprostřední blízkosti drátu (modré šipky v zeleném čtverci) leží v rovině, míří podél tečen k drátu a jejich orientace se mění po směru hodinových ručiček.

Za malý okamžik (za 5 nanosekund v případě typické frekvence FM vysílače 100 MHz) se polarita zdroje napětí otočí a elektrony budou proudit nahoru. Kolem drátu tak vznikne magnetické pole s vektory indukce měnícími se proti směru hodinových ručiček (červené šipky v zeleném čtverci na

<sup>1</sup>Když budeme mluvit o světle, budeme popisovat obecné vlastnosti elektromagnetických vln.

<sup>2</sup>Radiové vlny nejsou chemii tak vzdálené, jak by se mohlo zdát. V kapitole 14 si vysvětlíme, jak se radiové vlny využívají při zkoumání molekul pomocí nukleární magnetické rezonance.

obrázku 11.1). Takové pole ale vznikne jen *těsně* kolem drátu. Změna pole se šíří prostorem rychlostí sice velkou, ale nikoli nekonečnou. Tato rychlost se nazývá *rychlost světla ve vakuu* a obvykle se značí písmenkem  $c$ . S konstantami  $\epsilon_0$  a  $\mu_0$  souvisí rychlost  $c$  vztahem  $\epsilon_0\mu_0 = 1/c^2$  a hodnota  $c$  je  $3 \cdot 10^8 \text{ m s}^{-1}$ . Touto rychlostí doputuje za 5 ns změna pole do vzdálenosti 1,5 m od drátu. Proto mají vektory magnetické indukce 1,5 m od drátu takový směr, jaký odpovídal toku elektronů před 5 ns (modré šipky). Periodické změny napětí tak způsobují magnetické pole, které má každých 5 ns opačný směr.

Kolem antény nevzniká jen magnetické pole. Maxwellovy rovnice nám říkají, že kruhová magnetická indukce měnící se v čase vytváří elektrické pole mířící tak, jak je nakresleno na obrázku 11.1 v druhém čtverci pro jednotlivé časové okamžiky. Zdrojem elektrického pole jsou také elektrony, které se periodicky hromadí na jednotlivých koncích antény a chybí na koncích opačných. Takové rozložení náboje popisujeme elektrickým dipólovým momentem. Proto se také anténě říká dipól.

Spočítat přesný tvar elektrického pole v prostoru kolem antény není vůbec jednoduché, jeden závěr ale udělat můžeme. V rovině kolmé na drát v polovině výšky míří elektrické pole (vektory elektrické indukce) rovnoběžně s drátem. Maxwellovy rovnice také říkají, že lineární elektrická intenzita měnící se v čase vytváří kruhové pole magnetické. Jak vidíme, magnetická indukce  $\vec{B}$  a elektrická intenzita  $\vec{E}$  jsou spolu těsně provázány a jejich vzájemné působení je původem tvorby vln, které se mohou šířit i vakuem. V našem případě magnetické a elektrické pole mění periodicky svůj směr v kruzích kolem antény. Magnetické pole osciluje mezi orientací proti směru a po směru hodinových ručiček, elektrické pole mezi orientací nahoru a dolů. Dostatečně daleko od drátu se zmíněné změny orientací šíří v naší rovině jako kruhová vlna, jejíž maxima a minima jsou nakreslena modře a červeně ve třetím sloupci čtverců na obrázku 11.1. Jde o vlnu *polarizovanou*, směry vektorů  $\vec{E}$  a  $\vec{B}$  jsou jasně dané orientací antény.

Jestliže jsme si řekli, že oscilace elektrického pole způsobuje rotaci pole magnetického a naopak, tak je dobré si uvědomit, že v radiové vlně se vektory  $\vec{E}$  a  $\vec{B}$  stále mění v každém bodu. Mohli bychom tedy očekávat, že z každého bodu, kde se periodicky mění  $\vec{E}$  a  $\vec{B}$ , se bude tato změna šířit jako vlna ve všech směrech. To je představa blízka přístupu známému jako Huygensův princip: Tvar šířící se vlny určíme tak, že všechny body, do kterého vlna dorazí, považujeme za zdroje kulových vln, které sečteme.

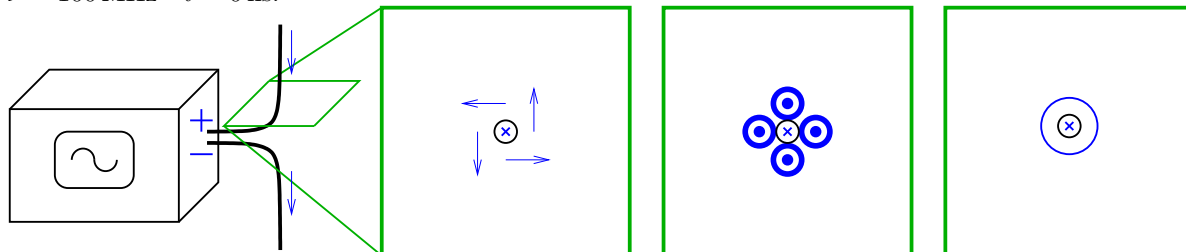
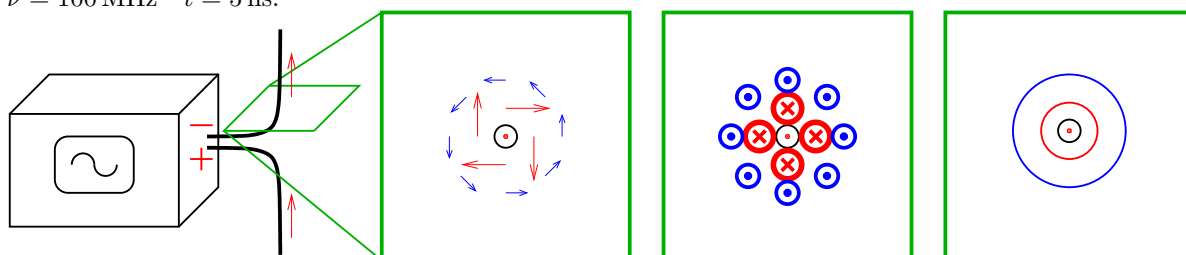
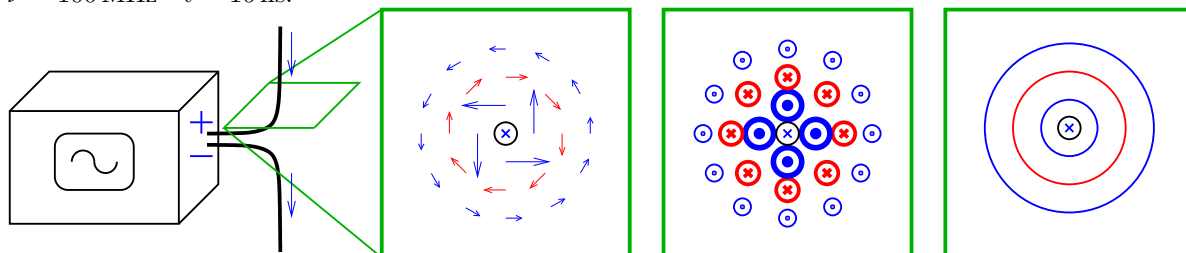
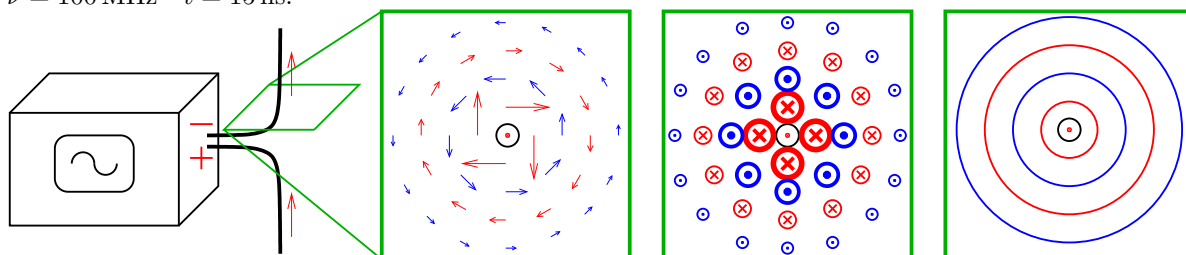
Po radiové anténě si uvedme ještě jeden zdroj elektromagnetických vln, mnohem složitější, ale bližší chemickému uvažování. Tímto zdrojem je Slunce. V obrovských teplotách, které na Slunci panují, se elektrony pohybují volně, nezávisle na jádrech atomů. Tento náhodný, tepelný pohyb představuje dívoce se měnící elektrický proud a způsobuje náhodilé změny rozložení náboje. Stejně jako u pravidelných změn rozložení elektronů v anténě, náhodné přeskupování náboje na Slunci indukuje fluktuace magnetického a elektrického pole. Matematicky můžeme takové fluktuace popsat jako součet oscilací s mnoha různými frekvencemi a v různých směrech. Tyto oscilace se šíří prostorem jako nepolarizované vlny. Dost daleko od Slunce jde o vlny kulové. A v malých rozměrech naší Země jde o vlny s téměř rovinným čelem.

## 11.2 Popis elektromagnetického vlnění

Přesný popis elektromagnetické vlny vyžaduje určení několika veličin, na které se teď podíváme blíže. Musíme si především uvědomit, které veličiny popisují změny elektromagnetického pole v čase a které v prostoru. K rozlišení časového a prostorového popisu poslouží představa vln na moři s bójkou upevněnou nad pevně daným místem mořského dna. Použijeme přitom pojmy *vlnová délka* a *perioda*, které si přesně definujeme níže.

Pokud si vlny vyfotografujeme, bude získaný snímek dobře zachycovat tvar vlny v prostoru, ale neponese žádnou informaci o změnách v čase. Na fotografii můžeme dobře změřit *vlnovou délku* jako vzdálenost sousedních hřebenů vln. Z fotografie se ale nic nedozvíme o tom, jak dlouho trvá, než bójka poklesne a zase se vznesne na hřeben vlny.

Pokud umístíme na bójku zařízení průběžně zaznamenávající, jak se mění její výška od mořského dna, získáme přesný popis oscilací v čase. Nedozvíme se ale nic o šíření vln prostorem. Záznam průběhu

$\nu = 100 \text{ MHz}$   $t = 0 \text{ ns}$ : $\nu = 100 \text{ MHz}$   $t = 5 \text{ ns}$ : $\nu = 100 \text{ MHz}$   $t = 10 \text{ ns}$ : $\nu = 100 \text{ MHz}$   $t = 15 \text{ ns}$ :

Obrázek 11.1: Ilustrace vzniku radiové vlny. Zdroj střídavého napětí mění polaritu s frekvencí  $\nu = 100 \text{ MHz}$ , jednotlivé obrázky ukazují elektromagnetické pole v intervalech 5 ns po zapnutí zdroje. V zelených rámečcích jsou (zprava doleva) nakresleny vektory  $\vec{B}$ ,  $\vec{E}$  a hřebeny vzniklé kruhové vlny. Směr vzhůru (na levém náčrtku) kolmo k zelené rovině je znázorněn tečkou, směr dolů křížkem. Směr pohybu elektronů v kolmých řezech je znázorněn tečkou nebo křížkem uprostřed zeleného rámečku.

výšky v čase bude mít podobný tvar, jako vlna na fotografii, tvar matematické funkce sinus nebo kosinus. Měření vzdálenosti dvou sousedních maxim na záznamu nám ale podá jinou informaci než změření vzdálenosti sousedních vrcholů na fotografii. Vzdálenost na záznamu odpovídá *periodě* vlny v čase.

Po plážovém úvodu se vrátíme k vlnám elektromagnetickým. Od mořské vlny se elektromagnetické vlnění liší tím, že jde o vlnění pole popsaného vektorovými veličinami  $\vec{E}$  a  $\vec{B}$ . Oba tyto vektory kmitají kolmo ke směru šíření. Jde tedy o vlnění *příčné*. Dále platí, že  $\vec{E}$  a  $\vec{B}$  jsou také vždy *vzájemně kolmé*. V dostatečné vzdálenosti od zdroje oba vektory nabývají maximálních či minimálních hodnot zároveň. Proto stačí sledovat změny jednoho z těchto vektorů (například  $\vec{E}$ ). Okamžitou velikost a směr vektoru druhého (například  $\vec{B}$ ) můžeme vždy dopočítat. Nyní si probereme si charakteristiky elektromagnetických vln (jejich souhrn je uveden v tabulce 11.1). Soustředíme se přitom na vektor  $\vec{E}$ .

- Vektor  $\vec{E}$  v každém místě prostoru kmitá v čase. Doba mezi okamžiky, ve kterých má vektor  $\vec{E}$  v určitém směru maximální hodnotu, se nazývá *perioda* a obvykle značí  $T$ . Převrácenou hodnotou periody je *frekvence*, kterou si budeme značit  $\nu$ . Jednotkou frekvence je Hertz (počet kmitů za sekundu). Matematicky se kmitání v čase dá zapsat například pomocí funkce kosinus

$$\vec{E}(t) = \vec{E}_0 \cos(2\pi t/T) = \vec{E}_0 \cos(2\pi \nu t), \quad (11.1)$$

kde  $\vec{E}$  je magnetická indukce ve směru kolmém na směr šíření vlny a  $\vec{E}_0$  je jeho maximum ( $|\vec{E}_0|$  je maximální velikost, neboli *amplituda* vektoru  $\vec{E}$ ). Faktor  $2\pi$  se někdy zahrnuje do hodnoty frekvence a vztah se pak zjednoduší na

$$\vec{E}(t) = \vec{E}_0 \cos(\omega t), \quad (11.2)$$

kde  $\omega = 2\pi\nu$  je analogií úhlové rychlosti kruhového pohybu a její jednotka je radián za sekundu.

- Změna vektoru  $\vec{E}$  se šíří prostorem rychlostí  $c$ . Jestliže je velikost vektoru  $\vec{E}$  v určitém místě a čase maximální, pak je ve stejném čase maximální také ve vzdálenosti  $c \cdot T$  od daného místa. Za dobu rovnou periodě do této vzdálenosti totiž stačí rychlostí  $c$  doputovat to, co určovalo velikost  $\vec{E}$  v původním místě. Maxima vln jsou tedy vzdálena o hodnotu  $c \cdot T$ , tato hodnota se nazývá *vlnová délka* a označuje  $\lambda$ . Převrácené hodnotě vlnové délky se říká *vlnočet*  $\tilde{\nu}$ . V případě jedné *rovinné vlny* si můžeme vždycky zvolit souřadnou soustavu tak, aby se vlna šířila ve směru osy  $x$ , vektor  $\vec{E}$  kmital ve směru osy  $z$  a vektor  $\vec{B}$  kmital ve směru osy  $y$ . Takové vlně říkáme *lineárně polarizovaná* a matematicky pak můžeme její tvar v prostoru zapsat rovnicí

$$E_z(x) = E_0 \cos(2\pi x/\lambda) = E_0 \cos(2\pi \tilde{\nu} x). \quad (11.3)$$

Pokud se rovinná vlna skládá z mnoha vln o stejné frekvenci, ale s vektorem  $\vec{E}$  kmitajícím různými směry, mluvíme o *nepolarizované* vlně.

- Šíření *kulové vlny* prostorem můžeme popsat tak, že polohu libovolného místa v prostoru zapíšeme pomocí *směrového vektoru*  $\vec{r}$  (jeho složky jsou souřadnice daného bodu). Dále nadefinujeme vektor  $\vec{s}$  tak, aby mířil ve směru šíření vlny, a aby jeho velikost byla  $|\vec{s}| = 1/\lambda$ . Obdobou rovnice 11.3 je pak

$$\vec{E}(\vec{r}) = \vec{E}_0 \cos(2\pi \vec{s} \cdot \vec{r}) = \vec{E}_0 \cos(2\pi(s_x x + s_y y + s_z z)) = \vec{E}_0 \cos(\vec{k} \cdot \vec{r}) = \vec{E}_0 \cos(k_x x + k_y y + k_z z), \quad (11.4)$$

kde jsme si zavedli vektor  $\vec{k} = 2\pi \vec{s}$  (obdoba vztahu  $\omega$  a  $\nu$ ).

- Vztahy popisující kmitání v čase a vlnění v prostoru můžeme spojit do jedné *vlňové rovnice*

$$\vec{E}(\vec{r}, t) = \vec{E}_0 \cos(\omega t - \vec{k} \cdot \vec{r}) = \vec{E}_0 \cos(\phi_t - \vec{k} \cdot \vec{r}) = \vec{E}_0 \cos(\omega t - \phi_r). \quad (11.5)$$

Proč ve všech rovnicích používáme kosinus a ne třeba sinus? A jaký je význam písmenek  $\phi$  v tomto vztahu? Pokud nás nezajímá kmitání v čase, ale jen vlnění v prostoru (což je případ difrakčních metod), můžeme časovou informaci zahrnout do nezájímavého korekčního faktoru  $\phi_t$ . Ten má význam posunutí vlny v prostoru v závislosti na tom, kolik času přesně uběhlo od zrození vlny. Pokud nás nezajímá vlnění v prostoru, ale jen kmitání v čase (což je případ nukleární magnetické rezonance), můžeme naopak prostorovou informaci zahrnout do nezájímavého korekčního faktoru  $\phi_r$ . Ten má význam časového posunu začátku sledování oscilací, který závisí na tom, jak jsme přesně daleko od zdroje vlny. Faktory  $\phi_t$  a  $\phi_r$  se nazývají *fáze* a vysvětlují, proč jsme všechny rovnice psali s funkcí kosinus. Mezi sinem a kosinem je rozdíl právě jen ve fázi. Všechny funkce kosinus bychom mohli v našich rovnicích přepsat na funkce sinus, kdybychom od argumentu kosinu odečetli fázi rovnou  $\pi/2 = 90^\circ$ . Pokud popisujeme jednu vlnu,  $\phi_t$  a  $\phi_r$  nemají žádný praktický význam a nezáleží na tom, jestli použijeme sinus, kosinus, nebo jakoukoli jejich kombinaci. Například radiové vlny s frekvencí 100 MHz se změň z kosinu na sinus, když radio zapneme o 2,5 nanosekundy později nebo s ním pohneme o 75 cm. Pokud nás ale zajímá skládání více vln, stává se mimořádně důležitým údajem *rozdíl jejich fází*. Při popisu vln se můžeme ale fázi použít také trochu jiným způsobem. V rovnici

$$\vec{E}(\vec{r}, t) = \vec{E}_0 \cos(\omega t - \vec{k} \cdot \vec{r}) = \vec{E}_0 \cos \phi \quad (11.6)$$

jsme výrazem  $\phi$  nahradili  $\omega t - \vec{k} \cdot \vec{r}$ . V tomto případě není fáze  $\phi$  nezájímavou korekční konstantou, ale popisuje změny  $\vec{E}$  v prostoru a čase. Pokud je tedy řeč o fázi, musíme dávat pozor na to, v jakém kontextu se výraz fáze používá.

Graficky lze elektromagnetické vlny znázornit různými způsoby. Na obrázku 11.2A,B jsou například vyznačena místa v prostoru, kde mají vektory  $\vec{E}$  a  $\vec{B}$  největší hodnotu v kladném (modré čáry) a v záporném (červené čáry) směru. Ještě jednodušší je zakreslit pouze směr šíření vlny, což je na obrázku 11.2A,B ukázáno světle zelenou barvou. Toto znázornění ale neudává, kde mají vlny v prostoru maximum a minimum. Můžeme také kreslit, jak se mění  $\vec{E}$  a  $\vec{B}$  v prostoru. To je na obrázku 11.2 znázorněno tmavě zelenými sinusovkami. Stejnou závislost můžeme použít pro oba vektory, musíme se jen dohodnout na tom, který směr sinusovka popisuje (na obrázku 11.2 je směr, ve kterém považujeme vektor  $\vec{E}$  nebo  $\vec{B}$  za kladný, znázorněn modrou šipkou.) Při tomto znázornění si musíme dát pozor na to, že v jednom obrázku (zde obrázek 11.2A,B) mícháme dvě veličiny (souřadnici v prostoru a elektrickou intenzitu nebo magnetickou indukci). Na obrázku 11.2C je kromě rozložení vlny v prostoru znázorněno hnědou barvou také kmitání elektromagnetického pole v čase.

Pokud srovnáváme více vln, je důležité zachytit také informaci o jejich vzájemném fázovém posunu. Na obrázku 11.3 vlevo je nakresleno několik rovinných vln s různou fází. Vlny jsou navíc podle fáze obarveny na škále mezi modrou a červenou tak, že první vlna je modrá a vlna s opačnou fází (s hodnotou  $\omega t - kx$  posunutou o  $180^\circ$ ) je červená. Na obrázku 11.3 vpravo je jiný způsob zobrazení fáze. Vychází z toho, že stejná sinusovka může popisovat dva různé pohyby vektorů  $\vec{E}$  nebo  $\vec{B}$ : kmitání  $\vec{E}$  ve směru  $z$  (nebo kmitání  $\vec{B}$  ve směru  $y$ ) a rotaci  $\vec{E}$  nebo  $\vec{B}$  kolem osy  $x$  (v rovině  $yz$ ). Kmitání vektoru  $\vec{E}$  si proto můžeme představit jako rotaci, při které nám  $\vec{E}$  uniká z našeho skutečného (reálného) směru  $z$  nikoli do směru  $y$ , ale směru kolmého k  $z$  v nějakém fiktivním (imaginárním) prostoru. Matematicky se takový pohyb zapisuje pomocí *komplexních čísel*<sup>3</sup>. Vlny s různou fází pak můžeme kreslit jako šipky znázorňující

<sup>3</sup>Komplexní číslo  $C$  se skládá ze dvou částí. Jedné z nich říkáme *reálná složka* a druhé *imaginární složka*. Tyto složky nemůžeme vzájemně míchat, chovají se podobně jako vzájemně kolmé složky vektoru: reálná část se podobá složce  $x$

Tabulka 11.1: Pojmy a veličiny popisující elektromagnetické vlnění

kmitání $\vec{E}$ , $\vec{B}$ v čase $t$ v každém bodu prostoru	perioda $T$	frekvence $\nu$	$\nu = 1/T$	$\omega = 2\pi\nu$
šíření ve směru $x$ rychlostí $c$ (rovinná vlna)	vlnová délka $\lambda$ $\lambda = cT = c/\nu$	vlnocet $\tilde{\nu}$	$\tilde{\nu} = 1/\lambda$	
šíření ve směru $\vec{r}$ rychlostí $c$ (kulová vlna)	vlnová délka $\lambda$ $\lambda = cT = c/\nu$	vlnový vektor $\vec{s}$	$ \vec{s}  = 1/\lambda$	$\vec{k} = 2\pi\vec{s}$
maximum $\vec{E}$ , $\vec{B}$	amplituda $E_0$ , $B_0$			
směr $\vec{E}$ , $\vec{B}$	polarizace			
posun počátku v čase nebo prostoru	fáze $\phi$			

odpovídající čísla. Délka šipky se rovná amplitudě vlny a úhly mezi šipkami popisují vzájemnou fázi vln. Toto zobrazení, zvané *Argandův diagram*, se často používá v krystalografii, o které bude řeč v kapitole 12. Pomocí rovnice zapíšeme vlnu v komplexním tvaru jako

$$\vec{E}(\vec{r}, t) = \vec{E}_0(\cos(\omega t - \vec{k} \cdot \vec{r}) + i \sin(\omega t - \vec{k} \cdot \vec{r})) = \vec{E}_0(\cos \phi + i \sin \phi) = \vec{E}_0 e^{i(\omega t - \vec{k} \cdot \vec{r})} = \vec{E}_0 e^{i\phi}. \quad (11.7)$$

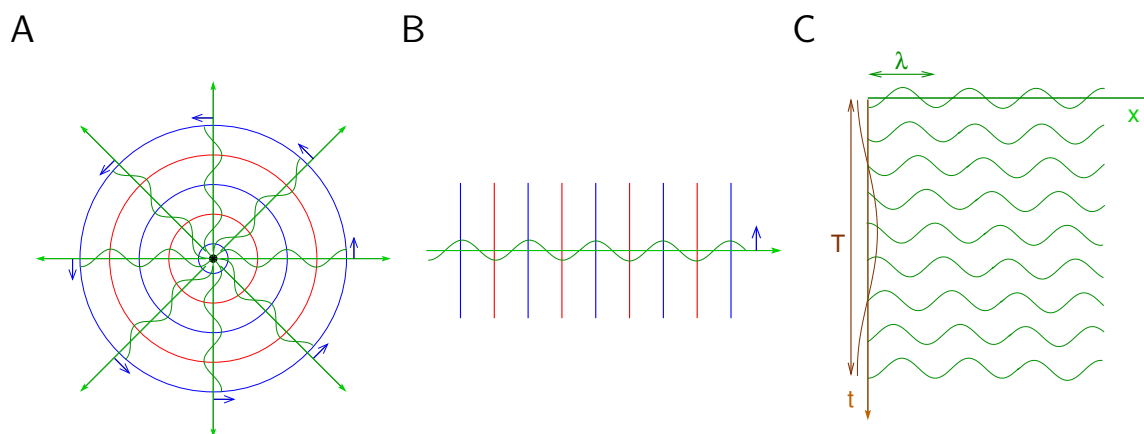
Je třeba dodat, že kromě vln s  $\vec{E}$  kmitajícím v jednom směru, také existují vlny, ve kterých se  $\vec{E}$  skutečně otáčí v rovině  $yz$ . Jak vlna postupuje, šipka znázorňující  $\vec{E}$  má konstantní délku a konec šipky se pohybuje po šroubovici kolem směru šíření vlny. Takové vlně říkáme *kruhově (cirkulárně) polarizovaná* a více si o ní řekneme v části 11.8.

### 11.3 Světlo a molekula

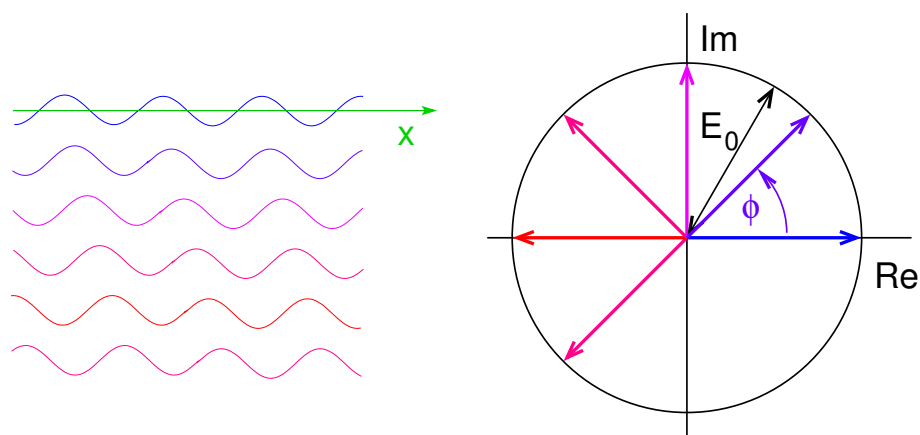
V kapitole 11.1 jsme si popsali elektromagnetickou vlnu v prostoru bez molekul, ve vakuu. Co se s touto vlnou stane, když vstoupí do prostředí tvořeného našimi molekulami? A co se stane s elektrony molekul ve vlnícím se elektromagnetickém moři? Elektrony jako elektricky nabitě částice pocítí měnící se pole a rozkmitají se s přibližně stejnou frekvencí<sup>4</sup>, jak je schematicky ukázáno na obrázku 11.4A. Tyto přesuny náboje buzené vnějším polem můžeme popsat jako elektrické dipólové momenty. Měnící se dipóly jsou ovšem samy také zdroji elektromagnetických vln, jak jsme si ukázali na příkladu radiového vysílače. Přibližný průběh elektrických siločar a magnetických indukčních čar pole tvořeného dipólovým

a imaginární složce  $y$ . Abychom reálnou a imaginární část odlišili, píšeme před imaginární část takzvanou *imaginární jednotku*  $i$ . Druhou mocninu komplexního čísla definujeme jako součin tohoto čísla s číslem s ním komplexně sdruženým (komplexně sdružená čísla se liší jen ve znaménku imaginární části):  $|C|^2 = (C_{\text{Re}} + iC_{\text{Im}})(C_{\text{Re}} - iC_{\text{Im}}) = C_0^2$ . Tento vztah vyžaduje, aby platilo  $i^2 = -1$ . Z toho, že komplexní číslo připomíná dvojrozměrný vektor, vyplývá, že je můžeme také zapsat pomocí amplitudy (velikosti neboli absolutní hodnoty)  $C_0$  a fáze (úhlu natočení)  $\alpha$ :  $C = C_0(\cos \alpha + i \sin \alpha)$ . Navíc platí velmi užitečný *Eulerův vztah*, který nám umožňuje vyjádřit komplexní číslo jako  $C = C_0 e^{i\alpha} = C_0(\cos \alpha + i \sin \alpha)$

<sup>4</sup>Pokud půjde o polární molekuly a frekvence vlnění nebude příliš vysoká, budou se celé molekuly otáčet tak, aby jejich stálé dipólové momenty měly vždy nižší energii v poli elektromagnetické vlny. Tak se chová molekula vody po dopadu mikrovlnného záření s frekvencí několika GHz. Molekula vody se přitom otočí několikrát za nanosekundu a vzorek se tím silně zahřeje, jak nám říká zkušenost s používáním mikrovlnné trouby. V případě milionkrát vyšší frekvence ultrafialových vln se voda nestací tak rychle otáčet. Proto dochází jen ke kmitání elektronů, popsanému v textu.



Obrázek 11.2: Různá grafická znázornění elektromagnetického vlnění. A, kulová vlna. B, rovinná vlna. Směr šíření vlny (paprsek) je vyznačen světle zeleně, okamžité hodnoty  $\vec{E}$  a  $\vec{B}$  popsané sinusovkou jsou nakresleny tmavě zeleně, směr vynášení kladné hodnoty  $\vec{E}$  a  $\vec{B}$  ukazuje modrá šipka. Oblasti prostoru s maximem vlny jsou vyznačeny modře, oblasti s minimem vlny červeně. C, tvar vlny v prostoru (zeleně) a kmitání vlny v čase (hnědě).



Obrázek 11.3: Fáze elektromagnetické vlny a jejich znázornění v Argandově diagramu.

momentem kmitajícím ve svislém směru je zakreslen na obrázku 11.4B. V dostatečné vzdálenosti od dipólu má elektrické a magnetické pole stejnou fázi a šíří se prostorem jako kulová vlna, jak jsme ji popsali v části 11.1. Oproti dopadající vlně má vlna šířící se z rozkmitaných elektronů fázi opožděnou<sup>5</sup> o  $90^\circ$ .

Chceme-li tedy popsat celkové pole, musíme k oscilujícímu poli dopadající vlny přičíst oscilující pole vln tvořených vybuzenými dipólovými momenty. Když bude nakonec elektromagnetická vlna prostředím našich molekul opouštět, bude jím pozměněna. Pokud budeme rozumět tomu, *jak* mění molekuly procházející vlnu, budeme umět z porovnání světla, které do vzorku s molekulami vstupuje, se světlem, které ze vzorku s molekulami vychází, získat informaci o našich molekulách.

Základem našich analýz bude sledování toho, jak se skládají vlny tvořené rozkmitanými elektrony různých atomů v různých molekulách. K tomu budeme potřebovat popsat šíření vln v různých směrech. Pokud dopadne na molekulu přibližně rovinná vlna, směr jejíhož šíření popisuje vektor  $\vec{k}$ , tak nás budou zajímat vlny rozkmitaných elektronů, jejichž šíření v různých směrech musíme popsat mnoha různými vektory  $\vec{k}'$ . Matematicky můžeme každý z těchto vektorů zapsat tak, že k původnímu vektoru  $\vec{k}$  přičteme nějaký rozdílový vektor  $\vec{q}$  (různý pro různé směry):

$$\vec{k}' = \vec{k} + \vec{q}, \quad (11.8)$$

kde velikost vektorů  $\vec{k}$  a  $\vec{k}'$  je  $1/\lambda$  a velikost vektoru  $\vec{q}$  závisí na odchylce směru  $\vec{k}'$  od směru  $\vec{k}$ . Šíření vln v různých směrech si můžeme graficky popsat jako odraz původní vlny od pomyslné roviny, znázorněné čárkovaně na obrázku 11.4C. Jak obrázek ukazuje, hodnota úhlu tohoto pomyslného odrazu  $\theta$  souvisí s velikostí vektoru  $\vec{q}$  vztahem

$$q = |\vec{q}| = 2 \frac{2\pi}{\lambda} \sin \theta. \quad (11.9)$$

V našich úvahách budeme muset rozlišovat vliv molekul malých (mnohem menších, než vlnová délka dopadajícího vlnění) od vlivu molekul velkých (rozměry se vlnové délce alespoň blízcích). V případě malých molekul budou elektrony všech atomů v molekule zdroji kulových vln, které se šíří přibližně z jednoho místa a mají tedy *stejnou fázi*. Při analýze vln šířících se z malých molekul stačí proto sledovat, jak se skládají vlny různých molekul. Proto se vlny jednotlivých elektronů budou vždycky sčítat.

V případě velké molekuly budou elektrony atomů v různých místech molekuly tak daleko od sebe, že nemůžeme rozdíly jejich poloh vůči vlnové délce zanedbat. Z molekuly se tak budou šířit kulové vlny s různými středy. Vlny z různých míst molekuly tak do určitého bodu v prostoru dorazí s *různou fází*. Proto se vlny jednotlivých elektronů téže molekuly někdy sečtou (když budou posunuty přibližně o celé násobky vlnové délky) a někdy odečtou (když budou posunuty přibližně o násobky  $\lambda/2$ ). Při analýze vln šířících se z velkých molekul proto musíme sledovat nejen, jak se skládají vlny různých molekul, ale také jak se skládají vlny šířící se z různých míst téže molekuly.

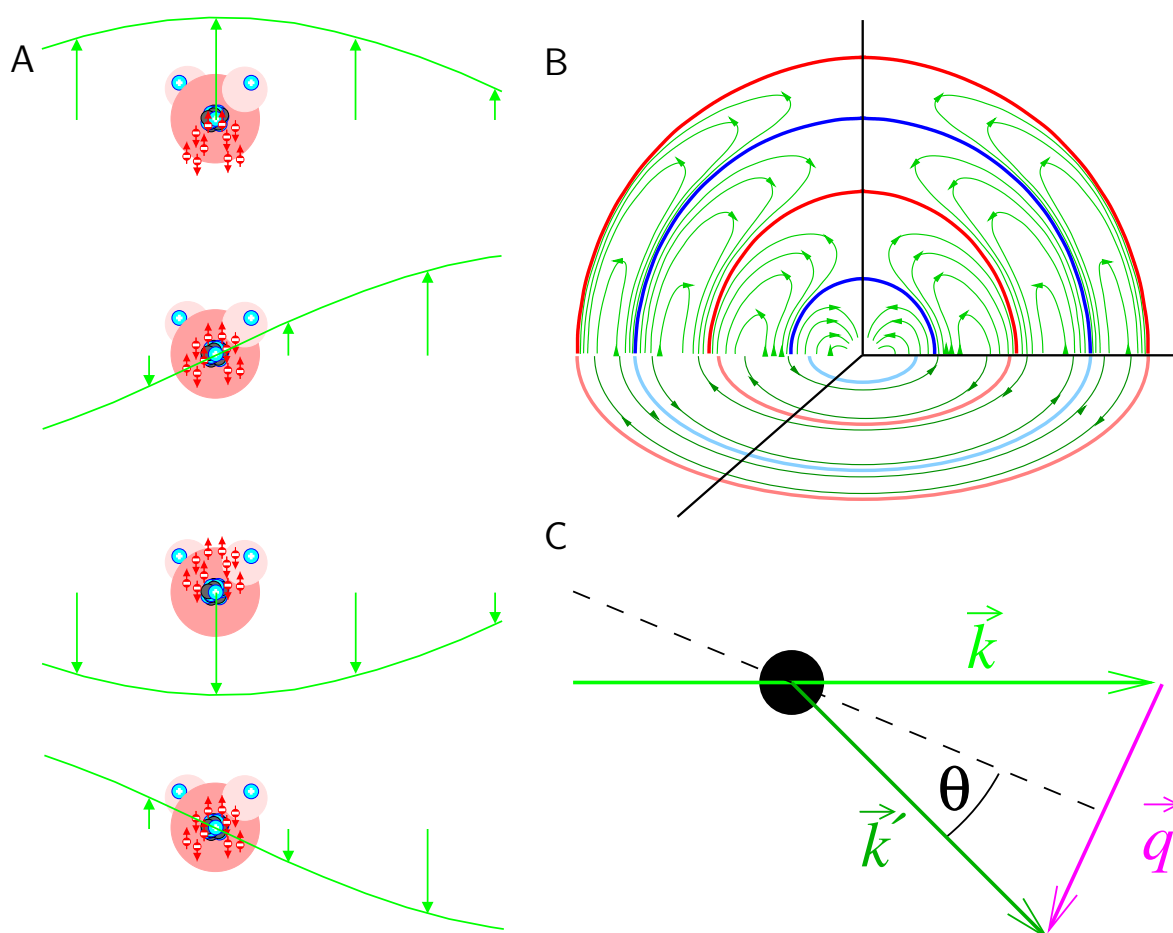
## 11.4 Světlo a krystal

Elektromagnetická vlna může dopadat na molekuly pravidelně rozmístěné a orientované (v krystalech), nebo na molekuly, jejichž polohy a orientace jsou náhodné (v plynech, kapalinách, či sklech). Nejprve se podíváme na první z uvedených případů, který si lze snáze představit.

Jak jsme si uvedli výše, v případě krystalů tvořených molekulami mnohem menšími než vlnová délka dopadajícího záření stačí vzít do úvahy skládání vln šířících se z různých molekul. Pokud jsou i vzdálenosti mezi molekulami mnohem menší než vlnová délka záření, budou elektrony v sousedních

<sup>5</sup>Z fyzikálního pohledu představuje molekula oscilátor rozkmitaný vnější silou (dopadající vlnou). Pokud, jako v našem případě, frekvence dopadající vlny rezonuje s frekvencí oscilátoru, rozkmitá se oscilátor s fázovým opožděním  $90^\circ$ .





Obrázek 11.4: A, rozkmitání elektronů v molekule vody při průchodu světelné vlny. Pro přehlednost je molekula vody nakreslena větší, než by odpovídalo vlnové délce viditelného nebo ultrafialového záření. B, elektromagnetické pole tvořené oscilujícím dipólovým momentem rozkmitaných elektronů v molekule. Elektrické pole je zakresleno ve svislém řezu jasnějšími barvami, magnetické ve vodorovném řezu matnějšími barvami. Směry vektorů  $\vec{E}$  a  $\vec{B}$  jsou vyznačeny šipkami. Modré čáry odpovídají maximům vln, červené minimům. C, směr šíření rozptýlené vlny popsán pomocí vektoru  $\vec{q}$ . Pomyslná rovina odrazu je vyznačena čárkovaně.

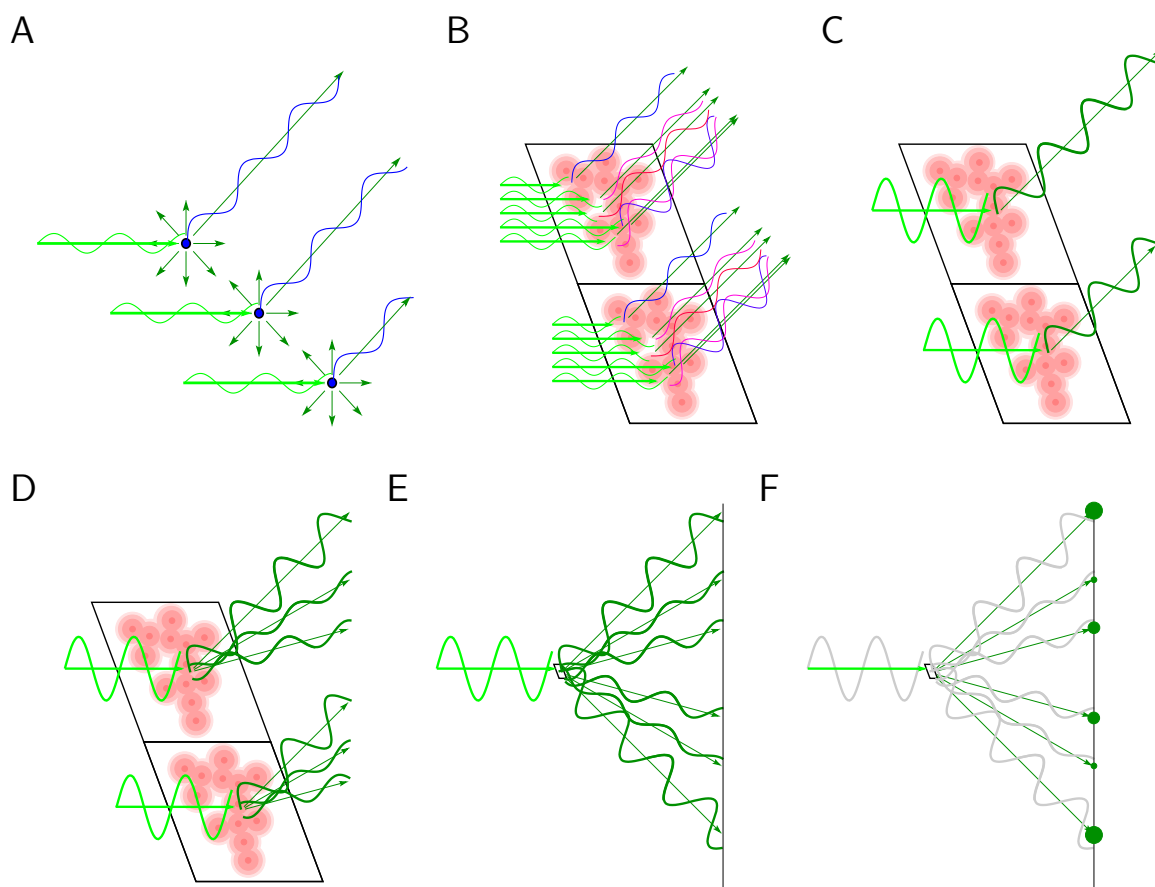
vrstvách molekul vyzařovat jen nepatrně fázově posunuté vlny. Takto mírně posunuté vlny se budou vzájemně sčítat. Krom toho se sečtou i s dopadající vlnou. Výsledkem bude vlna na každé vrstvě molekul mírně fázově opožděná, tedy celkově se pohybující trochu pomaleji, než vlna původní. V části 11.5 zjistíme, že podobně dochází k opoždování vlny při průchodu prostředím náhodně orientovaných a rozmístěných molekul, s jednou výjimkou. V případě krystalu může být opoždování vln kmitajících se stejnou frekvencí v různých směrech různé, v závislosti na symetrii krystalu. Různá rychlost postupu vln kmitajících v různých směrech nám umožňuje připravit různě polarizované světlo, jak zmíníme v části 11.8.

Pokud jsou vzdálenosti molekul v krystalu srovnatelné s vlnovou délkou dopadající vlny, mají pravidelné rozmístění a orientace molekul důležitý důsledek. Jestliže se v určitém směru sečtou vlny letící ze dvou molekul, musí se v tomto směru sečíst vlny šířící se ze všech stejně vzdálených a orientovaných dvojic molekul. Zjednodušeně je to znázorněno na obrázku 11.5A. Budeme sledovat vlny šířící se ve směru dlouhé tmavě zelené šipky. Pokud jsou molekuly (modré tečky) rozmístěny tak, že vlny letící z první a druhé molekuly mají ve zvoleném směru stejnou fázi, tak musí mít v tomto směru stejnou fázi také dvojice vln šířící se z první a třetí molekuly, stejně jako dvojice vln šířící se druhé a třetí molekuly. V krystalu je takových dvojic obrovské množství. Krystal proto mnohonásobně zesílí vlny šířící se v určitých směrech (ve kterých jsou fáze vln stejné). Při pohledu zvenčí tak máme dojem, že se v krystalu paprsek dopadajícího záření ohne do určitých směrů. Tomuto jevu se říká *ohyb světla*, neboli *difrakce*. V kapitole 12 stručně zmíníme, že ze směrů, do jakých se záření ohýbá, můžeme spočítat rozmístění molekul v krystalu.

Analýza difrakce na krystalech velkých molekul je (bio)chemicky ještě zajímavější. Difrakce totiž závisí nejen na skládání vln z různých molekul, ale také z různých míst téže molekuly. Aby byla molekula z pohledu vln velká, musíme použít záření s dostatečně krátkou vlnovou délkou (rentgenové záření). Protože různé molekuly tvoří v krystalu pravidelnou mřížku, záření se zesílí v určitých směrech, tak jako na krystalech malých molekul. V těchto směrech se ale nešíří kulové vlny ze zdrojů ve stejných bodech krystalové mřížky. Do směrů daných krystalovou mřížkou molekul míří vlnění zrozené z kmitání elektronů *v různých místech molekuly*. Protože elektrony v molekulách netvoří pravidelnou mřížku, kulové vlny pocházející z elektronů v různých částech téže molekuly mají ve směrech daných krystalem obecně různé fáze. Elektrony se navíc nenacházejí v přesně daných místech v molekule. Kvantová mechanika nám říká, že můžeme pouze hovořit o *pravděpodobnost výskytu elektronů*, neboli o *elektronové hustotě*. Takže musíme očekávat, že s určitou pravděpodobností se může šířit vlna z každého místa molekuly.

Šíření vlny v jednom náhodně vybraném směru je zjednodušeně zachyceno na obrázku 11.5B. Hustota elektronů je na obrázku znázorněna růžovou barvou. Kde je pravděpodobnost výskytu elektronu vyšší, (kde je růžová sytější), tam také s větší pravděpodobností vznikne kulová vlna. Amplituda kulových vln je proto dána elektronovou hustotou v místě jejich zrodu. Fáze těchto vln je dána polohou místa zrodu v molekule. Jednotlivé kulové vlny se skládají do výsledné vlny (znázorněné tmavě zeleně na obrázku 11.5C), která je mohutně zesílena ve směrech daných uspořádáním celých molekul v krystalu. Protože amplituda a fáze složené vlny závisí na rozložení elektronů v molekule, tedy na struktuře molekuly, říká se této vlně *strukturní faktor*. Strukturní faktor má v různých směrech zesílených krystalem různou amplitudu, podle toho jak jsou v určitém směru fázově posunuty vlny šířící se z různých míst molekuly (obrázek 11.5D). Při pohledu zvenčí vidíme ohyb (difrakci) záření do určitých směrů s různou amplitudou (danou rozložením elektronů v molekule (obrázek 11.5E)). Amplituda dopadajícího záření (přesněji její druhá mocnina) udává intenzitu<sup>6</sup> záření v různých směrech (zelené tečky na obrázku 11.5F), kterou můžeme měřit použitím vhodného detektoru. V kapitole 12 si podrobně rozebereme, jak lze rozložení elektronů a tedy tvar molekuly ze strukturního faktoru spočítat.

<sup>6</sup> Je poněkud nešťastné, že slůvkem *intenzita* se označují dvě různé veličiny, které se nám zde navíc setkávají v jednom vztahu. Na rozdíl od elektrické intenzity  $E$ , udává intenzita záření  $I$  energii záření (konkrétně plošnou hustotu zářivého výkonu, tedy změny energie za jednotku času).



Obrázek 11.5: Difrakce na krystalu. A, vlny zesílené v určitém směru (dlouhá tmavě zelená šipka) uspořádáním malých molekul (modré tečky) do pravidelné mřížky. B, vlny šířící se z různých míst molekuly, zesílené v určitém směru (dlouhá tmavě zelená šipka) uspořádáním velkých molekul (růžové tvary) do pravidelné mřížky. Sytost růžové barvy odpovídá rozložení elektronové hustoty v molekule. C, složená vlna (strukturní faktor) v daném směru. D, Strukturní faktor v různých směrech, daných uspořádáním molekul v krystalu. E, šíření vln v určitých směrech po dopadu záření na krystal. F, detekce intenzity záření (znázorněné velikostí zelené tečky) v různých směrech. Svislá čára představuje umístění detektoru.

## 11.5 Světlo a molekuly v roztoku či skle

Poněkud složitější je analýza skládání vln vzniklých rozkmitáním elektronů v *malých molekulách náhodně rozmístěných a orientovaných*, tedy molekul v roztoku či ve skle. Proto si úkol nejdříve trochu zjednodušíme a prozkoumáme, jak se skládají vlny tvořené kmitajícími elektrony v molekulách, které se nacházejí v jediné rovině kolmé na postupující rovinnou vlnu (obrázek 11.6A). Budeme přitom předpokládat, že molekuly jsou ve vzorku rozmístěny sice nepravidelně (netvoří krystalovou mřížku), ale v průměru rovnoměrně (homogenně) vyplňují prostor. Světlo postupuje jako rovinná vlna vzorkem a v každém okamžiku rozechvívá novou rovinu elektrických dipólových momentů elektronů v molekulách. Jak ukazuje obrázek 11.6B, vlny tvořené rozkmitanými elektrony různých molekul v rovině mají stejnou fázi ve směru postupu dopadající vlny, ale různou fázi v jiných směrech. Takto vzniklé vlny se šíří z dipólů kmitajících elektronů všemi směry a ve všech směrech se vzájemně téměř vyruší, kromě směrů rovnoběžných s původní (budící) vlnou. Výsledkem jsou tedy opět *rovinné* vlny. Tyto vlny se skládají s původní vlnou, oproti které jsou fázově opožděny o  $90^\circ$  (jak jsme zmínili v části 11.3) a mají mnohem menší amplitudu. Výsledná vlna bude nepatrně fázově posunuta (obrázek 11.6C).

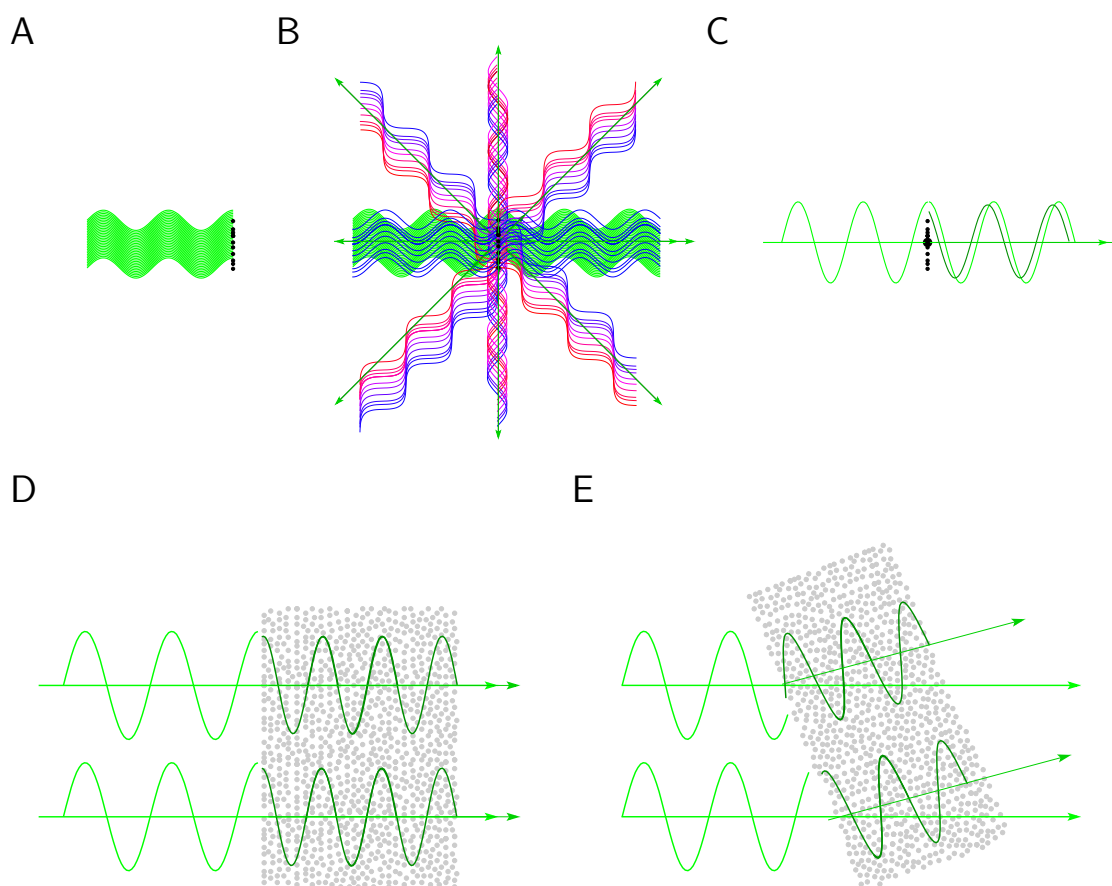
V případě chirálních molekul bude vektor  $\vec{E}$  také kmitat v rovině trochu pootočené vzhledem k rovině, ve které kmital před dopadem. Je to proto, že na elektrony působí při dopadu elektromagnetického vlnění dva druhy sil. První jsou síly dopadajícího vlnění, které se snaží rozkmitat elektrony v rovině, ve které kmitá  $\vec{E}$  vlny. Druhým typem sil jsou jeho interakce uvnitř molekuly. Tyto interakce omezují rozložení elektronů způsobem, který popisují molekulové orbitály. Například elektrony v  $\sigma$  vazbách mají větší volnost pohybu ve směru vazby, než kolmo na ni. Pokud vazba neleží v rovině dopadající vlny, směr kmitání elektronů není rovnoběžný se směrem kmitání vektoru  $\vec{E}$  vlny, ale je ovlivněn i orientací vazby v molekule. Pokud má molekula rovinu symetrie (není chirální), tak stačí molekulu otočit o  $180^\circ$  a orientace vazby se změní na svůj zrcadlový obraz. Proto se v souboru volně rotujících molekul s rovinou symetrie vlivy směru vazby vyruší. Pokud je ale molekula chirální (nemá rovinu symetrie), tak žádná rotace nemůže otočit molekulu do polohy, ve které by byla zrcadlovým obrazem molekuly v poloze původní. Z tohoto důvodu se vliv směru vazby úplně vyrušit nemůže a elektrony budou zdroji vln polarizovaných v trochu jiné rovině, než rovina kmitání  $\vec{E}$  dopadající vlny.

Popsaný zrod kulových vln, které po složení s dopadající vlnou mírně posunou její fázi, se opakuje na každé rovině. V důsledku toho se za běžných okolností vlna procházející roztokem či sklem trochu opoždí ve srovnání s vlnou šířící se vakuem<sup>7</sup> (obrázek 11.6D). Protože frekvence vlny se při průchodu prostředím molekul nemění, musí snížení rychlosti postupu vlny znamenat také zkrácení vlnové délky. V prostředí molekul je tedy postupující vlna zhuštěna. Poměr rychlosti světla ve vakuu ( $c$ ) k rychlosti šíření výsledné vlny roztokem molekul ( $c'$ ) se zpravidla označuje symbolem  $n$ . Toto číslo označujeme jako *index lomu*. Označení pochází z popisu situace, kdy světlo přechází z jednoho prostředí do druhého. Pokud dopadá rovinná vlna na rozhraní mezi prostředními šikmo, musí se šířit v jiném směru (láme se), aby zůstala spojitá (obrázek 11.6E). Úhel, pod kterým se vlna láme, můžeme spočítat právě z indexů lomu jednotlivých prostředí (Snellův zákon lomu). *Lom světla (refrakce)* je podstatou zvětšování obrazů pomocí čoček, kterému se budeme věnovat podrobněji v kapitole 13.

## 11.6 Maloúhlový rozptyl Röntgenova záření (SAXS)

Až dosud popisovaný obraz skládání vln tvořených rozkmitáním elektronů náhodně rozmístěných a orientovaných molekul se dramaticky změní, jakmile se zaměříme na molekuly *velké* (srovnatelné s vlnovou délkou dopadajícího záření). V tomto případě totiž musíme vzít do úvahy i skládání vln elektronů v téže

<sup>7</sup>Opoždování vlny znamená, že elektrony jsou zdrojem vln s fází posunutou o  $90^\circ$ . V některých případech může být fáze posunutá o  $270^\circ$  a předbíhat fázi budící vlny. Je proto nutné rozlišovat *fázovou rychlost* (která může být vyšší než rychlost světla ve vakuu  $c$ , protože fáze nenesé žádnou informaci) od *rychlosti šíření signálu* (která nemůže překročit rychlost světla ve vakuu  $c$ , protože důsledek nemůže být rychlejší než příčina).



Obrázek 11.6: Rozptyl světla na molekulách v roztoku. A, dopadající rovinná vlna (světle zelená) a molekuly nacházející se v rovině kolmé na postup vlny (černé tečky). B, šíření kulových vln způsobených rozkmitáním elektronů v molekulách v rovině kolmé na postup vlny. Vlnění v osmi vybraných směrech s barevně vyznačenou fází. Fáze ve směru šíření vlnové vlny je znázorněna modře, ostatní fáze jsou vyznačeny na barevné škále mezi modrou a červenou, kde červená odpovídá posunu o  $180^\circ$  (opačná fáze). C, výsledná vlna (tmavě zelená) složená z původní rovinné vlny (světle zelená) a vln tvořených elektrony rozkmitanými v rovině kolmé na postup rovinné vlny. D, výsledná vlna (tmavě zelená) složená z původní rovinné vlny (světle zelená) a vln tvořených elektrony rozkmitanými v celém objemu molekul roztoku (šedé tečky). E, lom šíření vlnění při šikmém dopadu rovinné vlny na roztok molekul či sklo.

molekule. Rozmístění elektronů v molekule definuje tvar molekuly a zdaleka není homogenní. Jak jsme již viděli na obrázku 11.5, složená vlna šířící se z jedné molekuly určitým směrem odráží pravděpodobnost výskytu elektronů na různých místech molekuly a vzájemný fázový posun vln v různých místech zrovných. Na rozdíl od krystalu závisí v roztoku šíření vlny i na tom, jak jsou molekuly orientovány. Skládání vln je tak v roztoku komplikováno průměrováním přes všechny možné orientace molekuly (obrázek 11.7A). Nedochází přitom ani k extrémnímu zesílení vln v určitých směrech (jako v krystalu), ani k téměř dokonalému odečtení vln ve všech směrech krom směru dopadající vlny (jako v roztoku malých molekul). Záření se s různou intenzitou šíří různými směry (obrázek 11.7B), mluvíme o *rozptylu (difúzi) záření* (anglicky *scattering*).

Jak již víme, k rozlišení jednotlivých atomů je třeba použít rentgenové záření, vlnovou délkou srovnatelné s velikostí atomů. Proto je rozptyl rentgenového záření zvláště užitečný pro studium tvaru proteinu a jiných biomakromolekul v roztoku. Metoda zvaná *SAXS* (small angle X-ray scattering, česky malouhlový rentgenový rozptyl) poskytuje poměrně detailní popis tvaru makromolekuly v roztoku, i když není schopna určit polohy jednotlivých atomů v molekule. Experiment se dosti podobá měření difrakce na krystalech (obrázek 11.7C), vyhodnocení výsledků se ale liší. Zaznamenáváme intenzitu záření rozptýleného pod různým úhlem  $2\theta$  od směru dopadajícího paprsku. Obvykle sledujeme závislost intenzity (na logaritmické škále) na velikosti vektoru  $q = |\vec{q}|$  (definovaného v rovnici 11.8), kterou snadno spočítáme z úhlu  $\theta$  (obrázek 11.7D), jiné užitečné vynesení dat zavedl rakouský fyzik Otto Kratky (obrázek 11.7E). Pro velmi malé hodnoty  $q$  (malé úhly rozptylu) platí přibližně kvadratická závislost (Guinierův vztah)

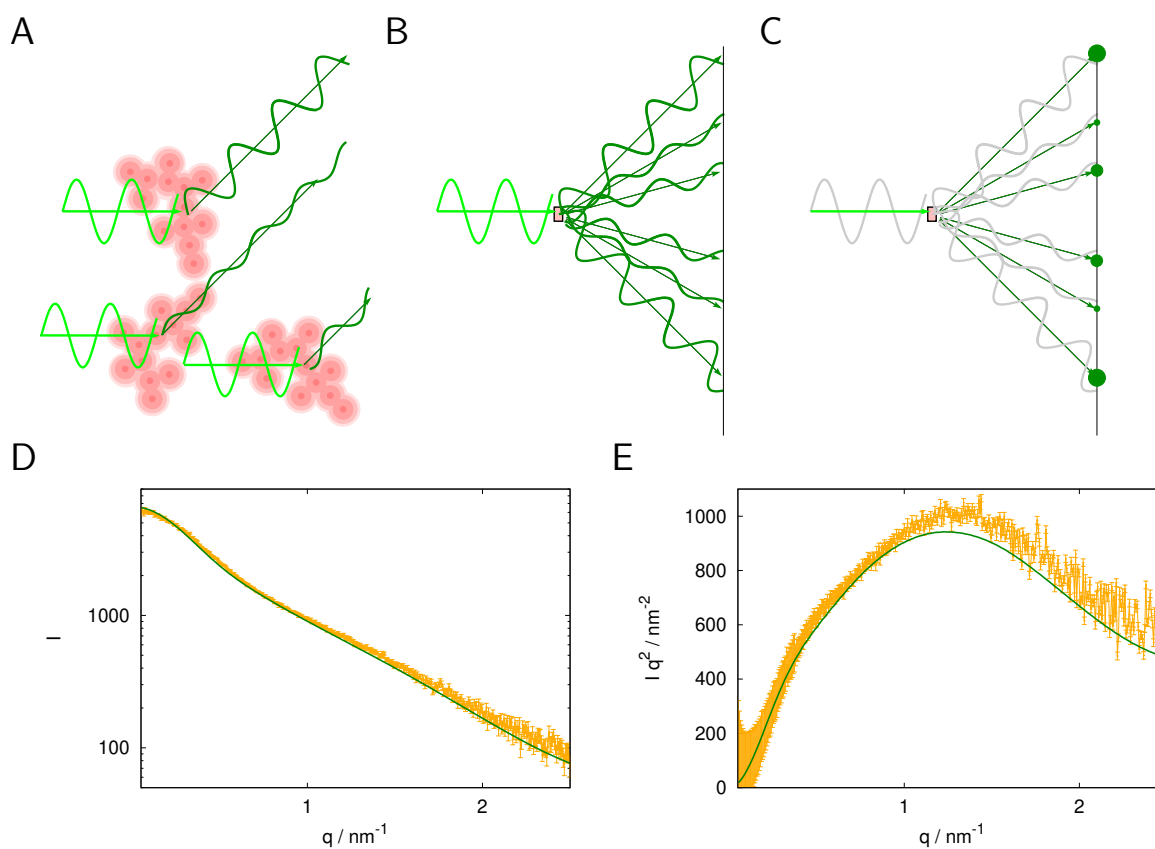
$$\ln I(q) \approx \ln I(q=0) - \frac{1}{3} R_g^2 q^2, \quad (11.10)$$

kde  $R_g$  je *setrvačný poloměr (radius of gyration)* molekuly. Z počátečního úseku závislosti  $\ln I$  na  $q$  tak můžeme snadno vypočítat velikost molekuly. Jemnější analýza je pak většinou založena na porovnání experimentu s nějakým modelem. Využíváme přitom toho, že ze známého rozložení elektronů v molekule lze intenzitu záření rozptýleného v různých směrech dobře spočítat. Pokud například chceme z měření SAXS stanovit kvartérní strukturu komplexu proteinů a terciární struktury podjednotek již známe, můžeme nechat počítač hledat takovou orientaci podjednotek v komplexu, aby nejlépe odpovídala naměřeným intenzitám. Na obrázku 11.7D,E je uveden jiný příklad, kdy byla metod SAXS použita ke kontrole výpočtu molekulové dynamiky. Pokud o tvaru molekuly žádnou představu nemáme, můžeme si zkusit skládat různé tvary z malých kuliček a porovnávat, který z takových hroznů odpovídá intenzitám rozptýleného záření nejlépe.

## 11.7 Absorpce

Až dosud jsme si vystačili s klasickým pohledem na elektrony rozkmitávané v molekulách elektromagnetickým vlněním. Abychom pochopili procesy, při kterých molekuly pohlcují energii vln, musíme použít kvantový popis. Elektron (a jiné částice) popisujeme v kvantové mechanice pomocí *vlnové funkce*. To je popis velmi podobný popisu elektromagnetické vlny, s tím rozdílem, že amplituda vlny nepopisuje elektrickou intenzitu nebo magnetickou indukci. Amplituda vlnové funkce je trochu tajemná veličina, ale její druhá mocnina popisuje něco, co už dobře známe, elektronovou hustotu v daném místě. Kvantová mechanika ukazuje, že elektron uvězněný v atomu či molekule může mít jen některé hodnoty energie a momentu hybnosti<sup>8</sup>. Kvantová mechanika navíc neumožňuje přesně určit všechny tři složky momentu

<sup>8</sup> *Moment hybnosti* je vektorová veličina popisující pohyb po zakřivených drahách. V případě pohybu částice o hmotnosti  $m$  rychlostí  $v$  po kružnici s poloměrem  $r$  je velikost momentu hybnosti rovna  $mvr$  a směr momentu hybnosti udává osu otáčení (je kolmý k rovině kružnice). Pokud je částice nabitá, má také *magnetický moment*, který je momentu hybnosti úměrný (má stejný směr). Proto můžeme různé orientace momentu hybnosti rozlišit, když umístíme molekulu do magnetického pole.



Obrázek 11.7: Maloúhlový rentgenový rozptyl (SAXS) na velkých molekulách v roztoku. A, vlny tvořené rozkmitanými elektrony v různých místech molekule s skládají do vlny s amplitudou závisující na rozmístění elektronů v molekule a fázovým posunem vln pro danou orientaci molekuly. B, podobně jako v případě difrakce na krystalu sledujeme rozptyl záření v různých směrech. C, zaznamenáváme přitom intenzitu záření rozptýleného pod různým úhlem. D, rozptylová křivka naměřená (zlatá barva) a vypočítaná z výsledků MD simulace (zelená barva) pro podjednotku delta RNA polymerasy bakterie *Bacillus subtilis*. E, stejná data v Kratkeho vynesení.

hybnosti (ve směru os  $x$ ,  $y$  a  $z$ ). Proto *stacionární stav*, ve kterém se elektron může nacházet, obvykle popisujeme jen trojicí čísel<sup>9</sup> udávajících energii, velikost momentu hybnosti a jeho složku v jednom směru. Místo hodnot těchto veličin v jednotkách J a Js většinou udáváme jen bezrozměrná *kvantová čísla*  $n, l, m_l$  (jde vlastně o násobky hodnot energie a momentu hybnosti, o které se mohou stavy lišit). Pravděpodobnost výskytu elektronu s povolenými hodnotami energie, velikosti a vybrané složky momentu hybnosti popisují vlnové funkce, kterým říkáme atomové a molekulové orbitály.

Kvantová mechanika také mění pohled na elektromagnetické vlnění. To může předávat energii jen po soustech neboli po *kvantech* o velikosti  $h\nu$ , kde  $h$  je Planckova konstanta a  $\nu$  je frekvence kmitání vlny.

Předpokládejme teď, že hodnota  $h\nu$  dopadajícího světla je shodou okolností stejná jako rozdíl mezi energiemi dvou stavů elektronů v molekule. Pokud to dovoluje tvar vlnových funkcí obou stavů, může za těchto okolností dojít k výměně energie. Molekula pohltí (*absorbuje*) dané sousto energie a použije je na změnu rozložení elektronů do energeticky náročnějšího stavu. Tuto změnu rozložení lze popsat elektrickým dipólovým momentem, kterému se říká *tranzitní* (přechodový) dipól. Velikost a směr tranzitního dipólu je možno spočítat z vlnových funkcí obou stavů. Kromě popsaného pohlcení energie molekulou může molekula energii i vyzářit. Může jít o *emisi* spontánní, danou pouze obsazením jednotlivých stavů elektrony, nebo o emisi nucenou, která je vyvolána světlem s frekvencí rezonující s rozdílem mezi energiemi jednotlivých stavů. Abyste mohli získat o elektronových stavech a přechodech mezi nimi konkrétnější představu, je v Doplňku D uveden poměrně podrobný rozbor biochemicky velmi důležitého uskupení atomů, peptidové vazby.

Na čem závisí schopnost molekul absorbovat energii světelného záření? Pravděpodobnost absorpce je tím větší, čím je frekvence záření bližší energii přechodu. Splnění této rezonanční podmínky nám popisuje *spektrum*, tedy závislost absorpce na frekvenci. Kromě toho je pravděpodobnost absorpce úměrná  $|\vec{E}_0 \cdot \vec{\mu}_T|^2 = E_0^2 \mu_T^2 \cos^2 \theta$ , kde  $\vec{\mu}_T$  je tranzitní dipólový moment a  $\theta$  je úhel mezi  $\vec{E}_0$  a tranzitním dipólem. Je-li molekula v roztoku, může se vůči světlu volně otáčet a  $\cos^2 \theta$  nahradíme střední hodnotou (což je v trojrozměrném prostoru jedna třetina).

Když světelná vlna putuje prostředím s mnoha absorbujícími molekulami (například vzorkem studovaného proteinu v roztoku), každá vrstvička molekul trochu sníží amplitudu. V důsledku toho se při průchodu vzorkem exponenciálně snižuje *intenzita světla*  $I$ , která je dána druhou mocninou velikosti elektrické intenzity  $E$ . Ve spektrometrii často měříme absorbanci  $A = \log(I_0/I)$  při průchodu vzorkem tloušťky  $l$ . Absorbance je přímo úměrná tloušťce vzorku a koncentraci absorbující látky. Konstantou úměrnosti je *absorpční koeficient*  $\varepsilon_\lambda$  (index  $\lambda$  upozorňuje, že účinnost absorpce závisí na vlnové délce)<sup>10</sup>.

Výpočty tranzitních dipólových momentů jistě nepatří mezi obvyklé činnosti biochemiků. Naopak *měření* intenzity světla prošlého vzorkem biomakromolekuly je v biochemii velmi běžné. Nejvhodnější je měření přechodů mezi orbitály aromatických systémů. V proteinech jde především o tryptofan, absorbující světlo o vlnové délce kolem 280 nm a o tyrosin, absorbující kolem 274 nm. Absorpce tyrosinu je slabší (asi čtvrtinová) ve srovnání s tryptofanem, proteiny ale obsahují většinou více tyrosinů než tryptofanů. Zmíněné vlnové délky leží v blízké ultrafialové oblasti (odpovídají frekvenci  $10^{15}$  Hz a energii  $400 \text{ kJ mol}^{-1}$ ), kde je možné měřit na běžném spektrofotometru s křemenou optikou. Máme-li štěstí,

<sup>9</sup>Ve skutečnosti musíme přidat ještě jedno číslo, které popisuje vnitřní moment hybnosti elektronu, zvaný *spin*.

<sup>10</sup>Rovnici popisující vlnu můžeme matematicky upravit tak, aby popisovala jak změnu rychlosti šíření, popsanou v části 11.5, tak exponenciální pokles intenzity, popsaný v této části. Úprava se týká vektoru  $\vec{k}$ . Zatímco ve vakuu je velikost  $k$  dána jednoduše vlnovou délkou ( $k = 2\pi/\lambda$ ), v prostředí molekul musíme tuto hodnotu vynásobit ještě takzvaným *komplexním indexem lomu*. Jak název napovídá, jde o komplexní číslo tvaru  $n - i\kappa$ , jehož reálná část je náš známý index lomu  $n$  a imaginární část  $\kappa$  je takzvaný *index absorpce*. Po dosazení do rovnice 11.7

$$\vec{E}(\vec{r}, t) = \vec{E}_0 e^{i(\omega t - (n - i\kappa)\vec{k} \cdot \vec{r})} = \vec{E}_0 e^{-\kappa\vec{k} \cdot \vec{r}} e^{i(\omega t - n\vec{k} \cdot \vec{r})} = \vec{E}_0 e^{-\kappa\vec{k} \cdot \vec{r}} (\cos(\omega t - n\vec{k} \cdot \vec{r}) + i \sin(\omega t - n\vec{k} \cdot \vec{r})) \quad (11.11)$$

vidíme, že jsme dosáhli toho, co potřebujeme. Zavedení indexu lomu  $n < 1$  nám popisuje zpomalený postup vlny a zavedení indexu absorpce  $\kappa$  nám popisuje tlumení v důsledku absorpce molekulami (imaginární číslo  $i\kappa$  dalo po vynásobení imaginární jednotkou reálné číslo  $-\kappa$ ). Z uvedených vztahů vyplývá, že běžně užívaný absorpční koeficient  $\varepsilon_\lambda$  je přímo úměrný indexu absorpce:  $\varepsilon_\lambda = 4\pi\kappa/(C\lambda \ln 10)$ , kde  $C$  je koncentrace molekul ve vzorku.



obsahuje protein navázanou skupinu, která umožňuje měření i ve viditelné oblasti (například hem). Báze nukleových kyselin představují poměrně složitý aromatický systém složený z osmi (u pyrimidinů), desíti (u adeninu), nebo jedenácti (u guaninu)  $\pi$ -orbitalů. Také báze nukleových kyselin absorbují ve stejné oblasti spektra (kolem 260 nm).

Již zmíněnou důležitou skupinou, jejíž absorpci můžeme měřit, je peptidová vazba. Nejvýraznější přechody mezi elektronovými stavy peptidové vazby jsou popsány v Dodatku D. Absorpce při přechodu  $n \rightarrow \pi_3^*$  je intenzivnější než absorpce tryptofanu, její vlnová délka (kolem 220 nm) je na hranici použitelnosti běžných spektrofotometrů. Absorpce při přechodu  $\pi_2 \rightarrow \pi_3^*$  se obvykle nevyužívá, přestože je ještě silnější. Důvodem je krátká vlnová délka (190 nm).

Oligo- a polysacharidy většinou neabsorbují ve vhodné oblasti. Pokud sacharidy neobsahují například acetylovanou aminoskupinu (tedy systém podobný peptidové vazbě), nejbližší použitelná vlnová délka je 175 nm (přechod  $n \rightarrow \sigma^*$  elektronů glykosidické vazby). Tak krátké vlnové délky vyžadují speciální zařízení (je například nutné odstranit kyslík rozpuštěný ve vodě) a využívají se málo.

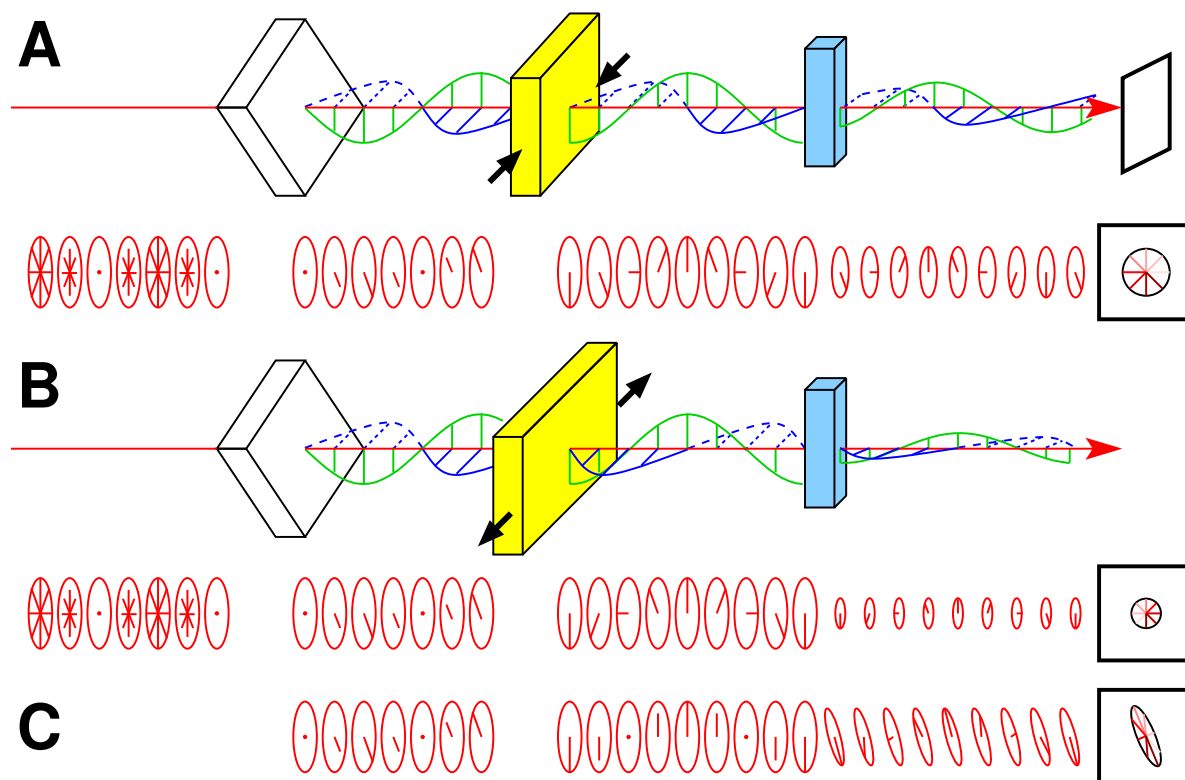
Nejčastějším využitím absorpce při elektronových přechodech (zejména aromatických skupin) je rychlé určení celkové koncentrace proteinu či nukleové kyseliny. Rozdíl mezi vlnovými délkami maximální absorpce je dostatečný k tomu, abychom odhadli i podíl proteinu a nukleové kyseliny ve směsi. Variací na toto téma je detekce biomakromolekul dělených elektroforézou nebo chromatografií. Peptidová vazba je zase výhodná proto, že je přítomna mezi každou dvojicí aminokyselin. Umožňuje nám pozorovat i krátké peptidy, které často neobsahují žádnou aromatickou aminokyselinu. Může nám ale absorpce viditelného nebo ultrafialového záření říci něco o struktuře? Na tuto otázku se pokusíme odpovědět v následujících sekcích.

## 11.8 Církulární dichroismus

V sekundárních strukturách jsou jednotlivé peptidové vazby proteinů a báze nukleových kyselin přesným způsobem vzájemně uspořádány. Toho lze využít k získání informace o sekundární struktuře. V dodatku E je zmíněno, jak se tvorba šroubovic proteinů a dvojšroubovic DNA projeví na absorpci. Ještě více se o sekundární struktuře můžeme dozvědět, když využijeme toho, že sekundární struktury jsou chirální útvary. Chirální molekuly totiž různým způsobem interagují s pravotočivým a levotočivým kruhově polarizovaným světlem. Pokud chirální molekuly světlo absorbují, je pravděpodobnost absorpce jednotlivých složek různá. Tento jev se nazývá *církulární dichroismus* (CD).

K pochopení církulárního dichroismu je užitečné si říci, jak můžeme kruhově polarizované světlo vyrobit (obrázek 11.8). Začneme se světlem lineárně polarizovaným. Vlnu kmitající v určitém směru (lineárně polarizovanou) můžeme z nepolarizovaného záření oddělit pomocí krystalu, ve kterém se vlny kmitající v různých směrech zpožďují různě (mají různý index lomu). Takovým materiálem je například křemen. Krystal (na obrázku 11.8 nakreslen bílou barvou) přitom orientujeme tak, aby vektor  $\vec{E}$  kmital v šikmé rovině, svírající úhel  $45^\circ$  se svislou i vodorovnou rovinou kolmou na směr postupu vlny. Kmitání vektoru  $\vec{E}$  v šikmé rovině znamená, že vodorovná i svislá složka vektoru kmitají se stejnou fází. Takto lineárně polarizované světlo necháme procházet takzvaným *fotoelastickým modulátorem*, nakresleným na obrázku 11.8 žlutě. Může jít o tenký hranolek oxidu křemičitého, který necháme účinkem elektrického pole vibrovat ve vodorovném směru (s frekvencí přibližně 50 Hz). V okamžiku, kdy je vibrující hranolek stlačen (obrázek 11.8A), zpomaluje<sup>11</sup> více vlnu elektrické indukce kmitající ve vodorovném směru. V okamžiku, kdy je natažen (obrázek 11.8B), je naopak vlna kmitající ve vodorovném směru

<sup>11</sup>V části 11.5 jsme si vysvětlili, že kmitání  $\vec{E}$  se na každé vrstvičce skla ve směru šíření původní vlny trochu fázově posune a tak se celá vlna postupující prostředím zpomalí. Podobně se zpomalí i vlny procházející hranolem křemenu s pravidelnou krystalovou mřížkou. Při stlačení se v krystalu křemene ocitnou atomy ve vodorovném směru blíže k sobě. Proto se šíření kmitání  $\vec{E}$  ve vodorovném směru zpomalí více, než v nestlačeném hranolu (stlačený hranol je ve vodorovném směru „hustší“). Rozmístění („hustota“) atomů ve svislém směru se při vodorovném stlačení nezmění. Proto se také nezmění rychlost šíření kmitání  $\vec{E}$  ve svislém směru.



Obrázek 11.8: Princip a měření cirkulárního dichroismu. A, průchod elektromagnetické vlny při stlačení fotoelastického modulátoru. Kmitání vektoru  $\vec{E}$  je symbolicky znázorněno červeně pod nákresem. Červené elipsy znázorňují pohled z boku na rovinu kolmou ke směru šíření vlny, černý kruh vpravo představuje pohled ve směru postupu vlny (směry vektoru  $\vec{E}$  v rostoucí vzdálenosti od pozorovatele jsou odlišeny klesající intenzitou červené barvy). Polarizátor (bílá barva) mění nepolarizované záření na lineárně polarizovanou vlnu kmitající v šikmé rovině (svírající úhel  $45^\circ$  se vvislou i vodorovnou rovinou). Vodorovná složka vektoru kmitajícího v šikmé rovině je nakreslena modře, vvislá zeleně. Stlačený fotoelastický modulátor (žlutá barva) opožďuje vodorovně kmitající složku o čtvrtinu vlnové délky. Výsledná vlna (složená z vodorovné a vvislé složky  $\vec{E}$ ) je tak pravotočivě kruhově polarizovaná. Ve vzorku (modrá barva) dochází k absorpci a opožďení pravotočivě polarizované vlny. B, průchod elektromagnetické vlny při natažení fotoelastického modulátoru. V tomto stavu vodorovná složka vektoru  $\vec{E}$  předbíhá vvislou složku a vzniká tak levotočivě kruhově polarizovaná vlna. C, součet vektorů  $\vec{E}$  z obrázků A a B. Složením vektorů  $\vec{E}$  vln po průchodu světlem nevznikne kruhově, ale elipticky polarizovaná vlna (černá elipsa vpravo). Natočení elipsy v rovině kolmé k šíření vlny závisí na rozdílu indexů lomu (způsobujícím rozdílné opožďení) levotočivě a pravotočivě polarizované složky.

zpomalena méně. Amplituda vibrací hranolku je přitom zvolena tak, aby po průchodu hranolem byla vodorovná vlna vůči svislé posunuta při stlačení o čtvrt vlnové délky zpátky a při natažení o čtvrt vlnové délky dopředu. Složením kolmých, vzájemně o  $\lambda/4$  posunutých vln vzniká kruhově polarizovaná vlna. V této vlně vektor  $\vec{E}$  nemění velikost, ale rotuje kolem směru postupu vlny. Špička šipky popisující vektor  $\vec{E}$  obíhá po šroubovici vinoucí se kolem směru postupující vlny. Při stlačení hranolku jde o vlnu pravotočivou, při natažení o vlnu levotočivou. V průmětu do roviny kolmé ke směru postupu vlny vidíme špičku šipky vektoru  $\vec{E}$  obíhat po kružnici. Při stlačení hranolku po směru hodinových ručiček (obrázek 11.8A), při natažení proti směru hodinových ručiček (obrázek 11.8B).

Takto připravené světlo oscilující mezi pravotočivou a levotočivou kruhovou polarizací necháme procházet vzorkem, nakresleným na obrázku 11.8 modře. Detektorem, který je časově synchronizován s vibracemi fotoelastického modulátoru, měříme absorpci pravotočivé a levotočivé vlny. Rozdíl mezi nimi vynásíme jako závislost na vlnové délce, tedy jako spektrum. Z historických důvodů se rozdíl absorpčních koeficientů pravotočivé a levotočivé vlny říká *elipticity*. Sečteme-li totiž různě pohlcenou pravotočivou a levotočivou vlnu, získáme místo kmitání po úsečce elipsu (obrázek 11.8C vpravo). Poměr dlouhé a krátké poloosy elipsy (*elipticity*) je přitom úměrný rozdílů zmíněných absorpčních koeficientů. Tento rozdíl je nulový vždy, když má molekula zrcadlovou rovinu symetrie (když není chirální). Elipsa je navíc otočena z důvodu rozdílného indexu lomu (rozdílné rychlosti šíření) pravotočivé a levotočivé vlny ve vzorku chirální molekuly.

Protože CD závisí přímo na chiralitě, můžeme se z CD spekter dozvědět užitečné informace o pravidelných chirálních strukturách biomakromolekul. Například u neuspořádaných proteinů pozorujeme v CD spektrech záporné maximum při 200 nm. Struktury skládaného listu vykazují pozitivní maximum při 198 nm a méně intenzivní negativní maximum při 215 nm. Konečně  $\alpha$ -šroubovice mají pozitivní maximum při 190 nm a dvě méně intenzivní negativní maxima při 208 nm a 222 nm (obrázek 11.9).

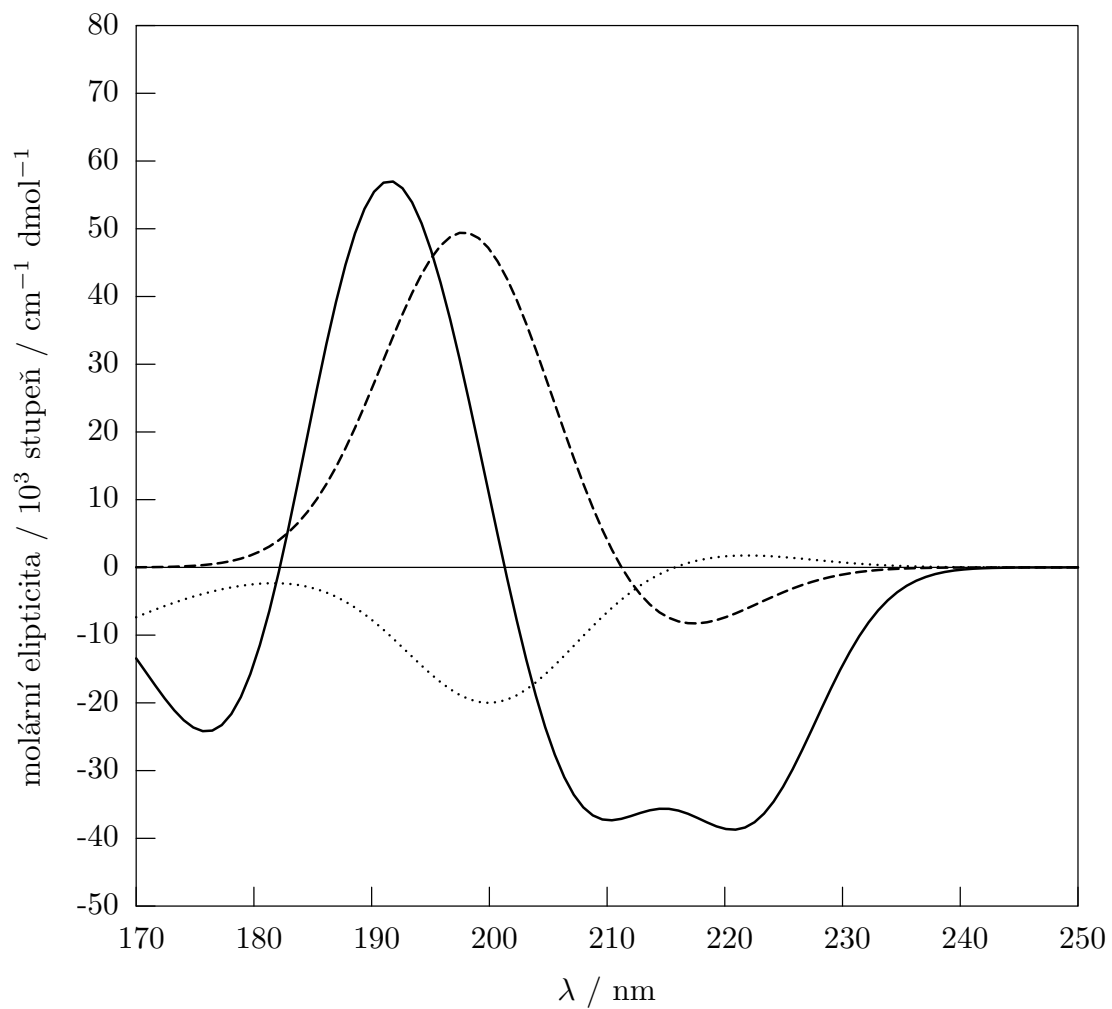
Pokud protein obsahuje různé druhy sekundární struktury, výsledné spektrum si můžeme představit jako součet tří základních spekter (náhodná konformace, skládaný list,  $\alpha$ -šroubovice) v poměru odpovídajícím zastoupení těchto tří konformací v proteinu.<sup>12</sup> Zpětně můžeme ze spektra odhadnout, jak velká část proteinu je v jednotlivých konformacích. Mohli bychom to udělat ručně, kdybychom odečítali elipticity při třech různých vlnových délkách a řešili pak soustavu tří rovnic o třech neznámých. V praxi většinou takovou analýzu svěříme počítači, který porovná naměřené spektrum se spektry známých proteinů a vyhodnotí zastoupení sekundárních struktur v našem vzorku.

Podobné rozdíly v CD spektrech můžeme pozorovat i u nukleových kyselin. Analýza spekter nukleových kyselin je ale složitější, protože každý pár bází má jiné spektrum (spektra proteinů závisí na druhu aminokyseliny velmi málo). Jako příklad si můžeme ukázat spektra dvojšroubovic DNA tvořených opakující se sekvencí CG (obrázek 11.10). O této sekvenci je známo, že kromě běžné pravotočivé dvojšroubovice B a pravotočivé dvojšroubovice A může tvořit i levotočivou dvojšroubovici Z se zlomy v průběhu cukrfosfátové páteře. Na obrázku je vidět, že pravotočivé dvojšroubovice mají podobná (ale rozlišitelná) spektra, zatímco spektrum levotočivé šroubovice Z (tečkované spektrum) má přibližně opačný průběh.

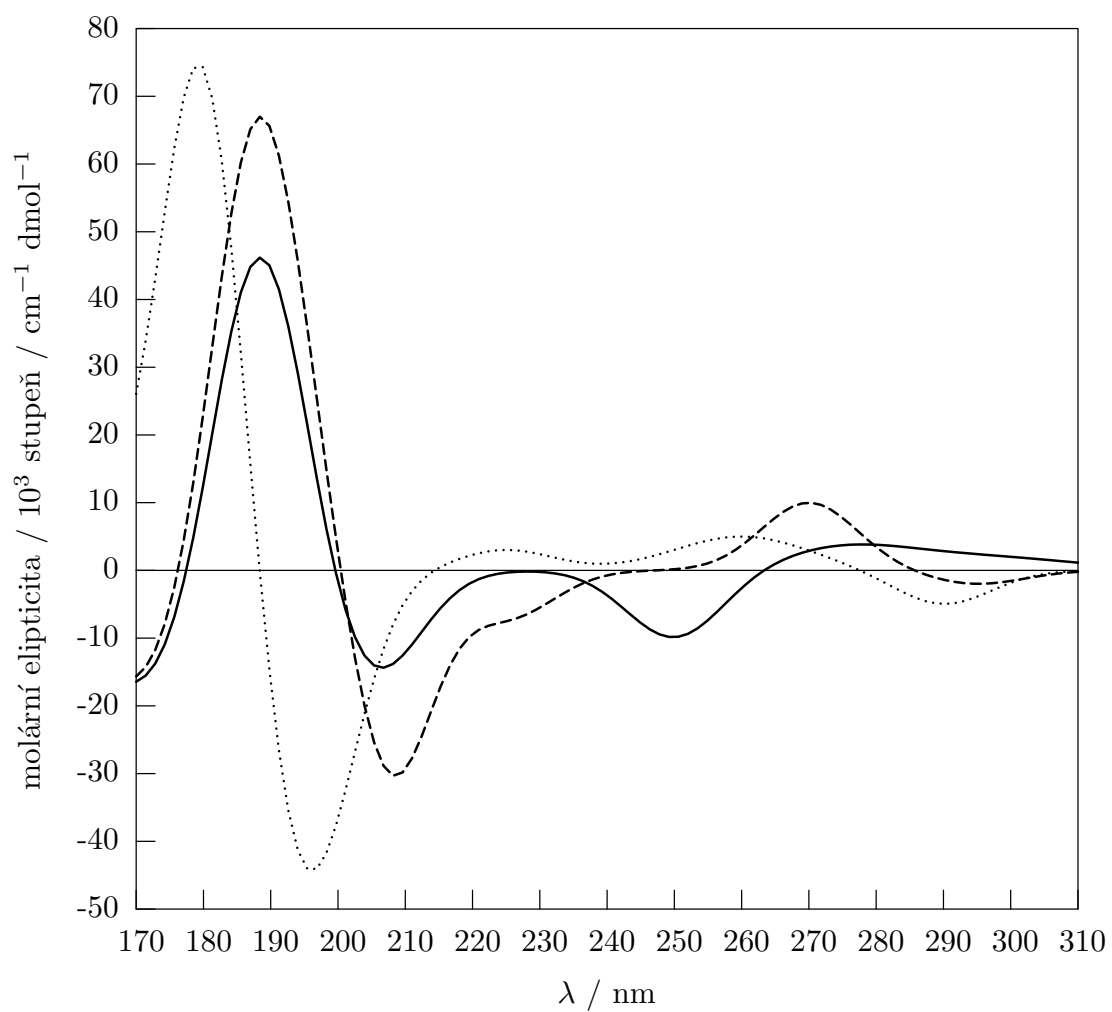
## 11.9 Vibrace jader

Zatím jsme při popisu průchodu světla prostředím molekul věnovali pozornost jen elektronům. Ani jádra atomů však neleží nehybně v určité poloze. Jádra se vůči sobě pohybují kmitavým pohybem. Polohu každého atomu můžeme popsat třemi nezávislými souřadnicemi, má tedy tři *stupně volnosti*.

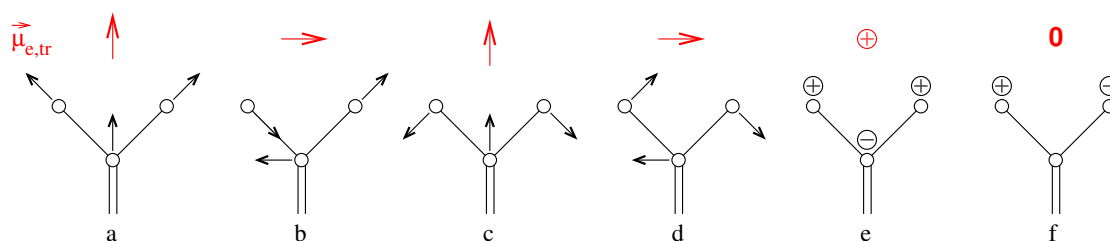
<sup>12</sup>Zatímco  $\alpha$ -šroubovice je definovaná úzkým rozmezím torzních úhlů  $\phi$  a  $\psi$ , ve skládaném listu mohou tyto torzní úhly nabývat dosti rozdílných hodnot, podle toho, jak je daný list v molekule proteinu zkrouten. Pod pojmem „náhodná struktura“ se pak skrývá průměr mnoha velmi rozdílných konformací. Rozklad spektra CD na tři základní složky proto představuje určité zjednodušení, které může v některých případech zavádět systematickou chybu do stanoveného zastoupení sekundárních struktur.



Obrázek 11.9: CD spektra  $\alpha$ -šroubovice (plná čára), skládaného listu (čárkovaně) a neuspořádaného polypeptidu (tečkovaně).



Obrázek 11.10: CD spektra dvojšroubovic DNA tvořených opakovanou sekvencí CG: dvojšroubovice B (plná čára), dvojšroubovice A (čárkovaně) a levotočivá dvojšroubovice Z (tečkovaně).



Obrázek 11.11: Vibrace tříatomové molekuly: a, symetrická valenční (in-phase stretching), b, asymetrická valenční (out-of-phase stretching), c, nůžková (in-plane in-phase bending), d, kývavá (in-plane out-of-phase bending), e, rotující (out-of-plane in-phase bending), f, twistová (out-of-plane out-of-phase bending). Na obrázku jsou vibrace v rovině papíru znázorněny šipkami, vibrace nad rovinu papíru  $\oplus$  a vibrace pod rovinu papíru  $\ominus$ . Červeně je vyznačen směr elektrického tranzitního momentu jednotlivých vibrací.

Spojíme-li dva atomy vazbou, stačí nám pro popis polohy vzniklé molekuly zase tři souřadnice. Co se stalo se třemi souřadnicemi druhého atomu? Jsou skryty v dalších stupních volnosti molekuly, které jsme dříve nazvali *vnitřními souřadnicemi*. Pro přesný popis molekuly musíme také udat její natočení v prostoru, tedy rotaci kolem dvou os kolmých na vytvořenou vazbu. Poslední stupeň volnosti nám popíše vzdálenost jader v molekule, a právě tento stupeň volnosti souvisí s vibrací (jádra nejsou stále stejně vzdálena, ale kmitají kolem rovnovážné polohy). Když přidáme k molekule další atom, zůstanou nám stále tři souřadnice pro popis polohy (těžiště) a přibude jeden další parametr pro popis natočení v prostoru (pokud nebude molekula lineární). Zbývají nám tedy ještě tři (u lineární molekuly čtyři) stupně volnosti pro popis vzájemných poloh jader, tedy vibrací.

Zkusme teď prostřední atom z naší trojice zapojit do makromolekuly a předpokládejme, že vibrace našich tří atomů jsou zcela nezávislé na vibracích makromolekuly<sup>13</sup>. V tomto zjednodušení potřebujeme tři stupně volnosti na popis polohy naší skupiny vzhledem k makromolekule. Zbývajících šest stupňů volnosti popisuje šest způsobů vibrace, které jsou znázorněny na obrázku 11.11.

Zákony kvantové mechaniky vyžadují, aby jádra kmitala jen s určitými frekvencemi. Stejně jako elektrony, i jádra se tedy v molekule vyskytují v určitých stavech. Energie procházejícího světla může být proto využita i k přeskokům mezi těmito *vibračními* stavy. Každá ze zmíněných vibrací má svou rezonanční frekvenci<sup>14</sup>. Intenzita přechodů do vyšších stavů pro jednotlivé vibrace je dána opět tranzitními dipóly (na obrázku červeně). Ty nám vlastně udávají směr polarizovaného světla rezonujícího s jednotlivými vibracemi. Energetické rozdíly jsou pro vibrace jader nižší o jeden až dva řády ve srovnání s elektrony, takže k rezonanci dochází s infračerveným světlem (frekvence  $10^{13}$ – $10^{14}$  Hz).

Kvantová mechanika také definuje podmínky, za kterých můžeme absorpci v infračervených spektrech pozorovat. Důležitá podmínka pro to, aby byl tranzitní dipólový moment nenulový a vibrační přechod povolený, se týká stálého dipólového momentu naší funkční skupiny (viz Dodatek D, rovnice 15.121). Stálý dipólový moment se musí během vibrací měnit (vibrace nesmí být symetrická). Přechody, při kterých se stálý dipólový moment nemění a jsou v absorpčních spektrech zakázané, mohou být ale povolené v takzvaných Ramanových spektrech, o kterých se zmíníme v sekci 11.10.

Frekvence a energie vibrací závisí na hmotnosti jader a síle vazeb, které je spojují. Vytvoření vodíkové vazby při vzniku sekundární struktury tedy pozmění energetické rozdíly vibračních stavů. Proto můžeme studovat sekundární strukturu i pomocí vibračních spekter. Za chvíli se seznámíme s nejčastějším příkladem takové analýzy – s pozorováním vybraných vibrací peptidové vazby. V nukleových kyselinách můžeme podobným způsobem sledovat tvorbu vodíkových vazeb během párování bází. Na konformaci nukleové kyseliny závisí vibrace fosfodiesterových vazeb.

<sup>13</sup>Uvědomme si ale, že se tak dopouštíme určité nepřesnosti. Kmitání našich tří jader *bude* spřaženo s vibracemi zbytku makromolekuly.

<sup>14</sup>Ve vibrační spektroskopii je zvykem místo frekvence  $\nu$  udávat vlnočet  $\tilde{\nu} = 1/\lambda$ , zpravidla v jednotkách  $\text{cm}^{-1}$ .

Ve vibračních spektrech můžeme pozorovat i cirkulární dichroismus, pokud použijeme polarizovaného infračerveného záření. Citlivost takového měření je ale velmi nízká.

Na závěr povídání o vibracích jader atomů se podíváme na konkrétní příklad využití vibračních spekter, na určování sekundární struktury atomů pomocí vibrací peptidové vazby. Opět použijeme zjednodušený popis, který vychází z předpokladu, že vibrace jednotlivých vazeb jsou do značné míry nezávislé na pohybech okolních atomů. Ze zkušenosti víme, že vibrace kolem  $3\,400\text{ cm}^{-1}$  je možné připsat vazbě N–H, vibrace kolem  $1\,700\text{ cm}^{-1}$  je možné připsat vazbě C=O, vibrace kolem  $1\,650\text{ cm}^{-1}$  je možné připsat vazbě C=N. Tento přístup je ale v mnoha případech příliš hrubý. Lepší výsledky poskytne zjednodušení opatrnější, které nepracuje s izolovanými dvojicemi atomů, ale rozumně zvolenými funkčními skupinami. V sekci 11.9 jsme si popsali šest možných vibrací skupiny složené ze tří atomů. Pokud se vrátíme k peptidové vazbě, situace spřažených vibrací bude ještě složitější. Můžeme popsat devět spřažených vibračních pohybů čtyř atomů ve skupině CO–NH. V literatuře se těchto devět vibračních modů označuje názvy AMID A, AMID B, AMID I, AMID II, a tak dále až po AMID VII (frekvence vibrací klesá v uvedeném pořadí). V praxi se nejčastěji využívá vibrací označovaných AMID I a AMID II (v infračervených spektrech; obrázek 11.12) nebo AMID I a AMID III (v Ramanových spektrech; obrázek 11.12), jejichž frekvence závisí na účasti vodíku ve vodíkové vazbě a tedy na druhu sekundární struktury.

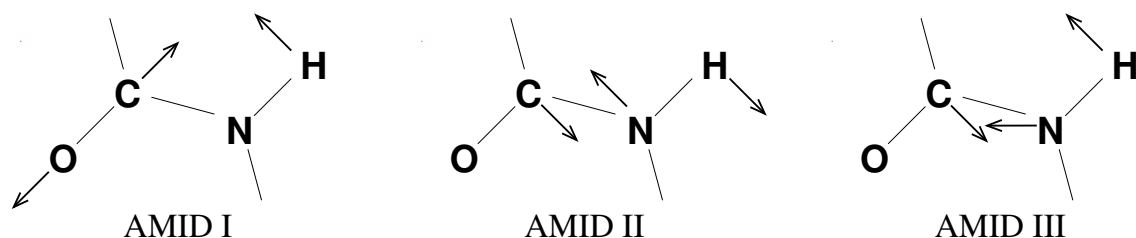
Postup při analýze sekundární struktury si můžeme ukázat na příkladu vibrace AMID I. Kdyby vodík peptidové vazby netvořil žádné vodíkové vazby, pozorovali bychom vibrace kolem vlnočtu  $1670\text{ cm}^{-1}$ . Ve vodném prostředí se tvoří ale vodíkové vazby s molekulami vody, které zpomalují vibrace AMID I. V neuspořádaných proteinech můžeme pozorovat zpomalené vibrace AMID I peptidových vazeb vystavených vodě v oblasti  $1640\text{--}1650\text{ cm}^{-1}$ . Co se stane při vytvoření sekundární struktury? Vodíkové vazby v  $\alpha$ -šroubovici jsou slabší ve srovnání s vodíkovými vazbami s vodou, proto jsou vibrace AMID I o něco rychlejší ( $1650\text{--}1660\text{ cm}^{-1}$ ). Naopak vodíkové vazby ve skládaných listech jsou silnější než vodíkové vazby s vodou. Vibrace AMID I je proto v této sekundární struktuře ještě více zpomalena ( $1625\text{--}1640\text{ cm}^{-1}$ ). Ve struktuře skládaného listu jsou také vibrace peptidových vazeb vzájemně spřaženy. Toto spřažení se projeví ve spektrech dalšími frekvencemi (okolo  $1695\text{ cm}^{-1}$ ), kterou bychom neočekávali pro izolovanou peptidovou vazbu. Další neobvyklé vibrace kolem  $1670\text{ cm}^{-1}$  jsou přisuzovány peptidovým vazbám v ohybech.

Při analýze vibrací AMID I musíme překonat některé technické překážky. První z nich je přítomnost vibrací vody okolo  $1650\text{ cm}^{-1}$ , které překrývají nejzajímavější část spektra. Moderní přístroje dokáží vibrace vody od spektra spolehlivě odečíst, dříve se zpravidla měření prováděla v těžké vodě, která vibruje pomaleji<sup>15</sup>. Další obtíž vyplývá z toho, že vibrace jednotlivých druhů sekundárních struktur se liší jen velmi málo a ve spektrech spolu splývají. Také tento problém pomáhá řešit matematické zpracování spektra. Dnes je proto možné nechat počítač rozložit spektrum na příspěvky odpovídající jednotlivým vibracím (při předem známých vlnočtech) a z poměru jejich intenzit určit zastoupení jednotlivých druhů sekundárních struktur. Spolehlivě tak lze odhadnout zastoupení skládaného listu, trochu problematické může být rozlišení  $\alpha$ -šroubovice od neuspořádané části proteinu (jak jsme si řekli, vibrace neuspořádané struktury značně závisí na tom, jestli tvoří vodíkové vazby s vodou).

## 11.10 Ramanova spektra

V sekci 11.5 jsme mluvili o rozptylu světla, při kterém nedochází k výměně energie s molekulou (*elastický rozptyl*). Dalo by se očekávat, že pokud ozáříme molekulu světlem o vlnové délce, kterou molekula neabsorbuje, bude mít rozptýlené světlo stejnou vlnovou délku. Neplatí to tak docela. Když vzorek molekul ozáříme velmi intenzivním světlem určité vlnové délky (obvykle jde o laserový paprsek), pozorujeme v rozptýleném světle nejen světlo stejné vlnové délky, ale i světlo, které vyzářily molekuly,

<sup>15</sup>Vibrace peptidové vazby jsou v těžké vodě také odlišné, protože vodík se pomalu vyměňuje s deuteriem. Aby nedošlo k nedorozumění, odlišují se vibrace deuterované peptidové vazby v literatuře čárkou (například AMID I').



Obrázek 11.12: Vybrané vibrace peptidové vazby.

u nichž nejprve došlo k přechodu do nižšího (nebo vyššího) vibračního stavu. Frekvence tohoto světla je snížena (nebo zvýšena) o příspěvek, který odpovídá rozdílu energií vibračních stavů, mezi kterými došlo k přechodu. Závislost intenzity rozptýleného světla na popsané změně frekvence se nazývá *Ramanovo spektrum*.

Na čem závisí intenzita světla pozorovaného v Ramanových spektrech? Pro odpověď se musíme vrátit do sekce 11.3. Zdrojem rozptýlené vlny jsou vybudené elektrické dipólové momenty, které odpovídají rozkmitání elektronů v molekule po dopadu světla<sup>16</sup>. Stejně jako v případě obyčejného vibračního spektra (sekce 11.9) musí během přechodu dojít ke změně tohoto dipólového momentu. Ale pozor, zde nejde o stálý dipólový moment, ale o *vybuděný* dipól. Proto není podmínkou, aby se měnil stálý dipólový moment molekuly. Podmínkou je změna schopnosti molekuly *nechat vybudit elektrické dipóly*. Jinými slovy, změna ochoty přemístit své elektrony vlivem vnějšího pole (změna *polarizovatelnosti*) během vibrace.

Ramanova spektra nám umožňují studovat vibrace stejně jako infračervená vibrační spektra. Podmínky pozorovatelnosti určité vibrace jsou ale různé (v případě molekul se středem symetrie jsou dokonce vibrační přechody povolené v infračervených spektrech zakázané v Ramanových spektrech a naopak). Symetrické vibrace jsou dobře viditelné v Ramanových spektrech, zatímco nesymetrické vibrace jsou viditelné v infračervených spektrech. Pomocí Ramanových spekter tedy můžeme zkoumat sekundární strukturu stejně jako pomocí infračervených spekter, pouze se musíme zaměřit na jiné vibrace.

<sup>16</sup>Vlnová délka dopadajícího světla je mnohem kratší, než by odpovídalo přechodu mezi vibračními stavy. Z praktických důvodů se používá vlnová délka z viditelné oblasti, kterou poskytují běžné lasery. Zajímavou modifikací metody je *rezonanční Ramanova spektroskopie*, kdy vlnovou délku dopadajícího záření volíme tak, aby byla molekulou absorbována při přechodech mezi elektronovými stavy (nikoli mezi vibračními stavy jader). Výhodou je značné zesílení citlivosti. Rezonanční Ramanova spektra můžeme pochopitelně měřit jen pro takové funkční skupiny, které absorbují ve vhodné oblasti.



## Kapitola 12

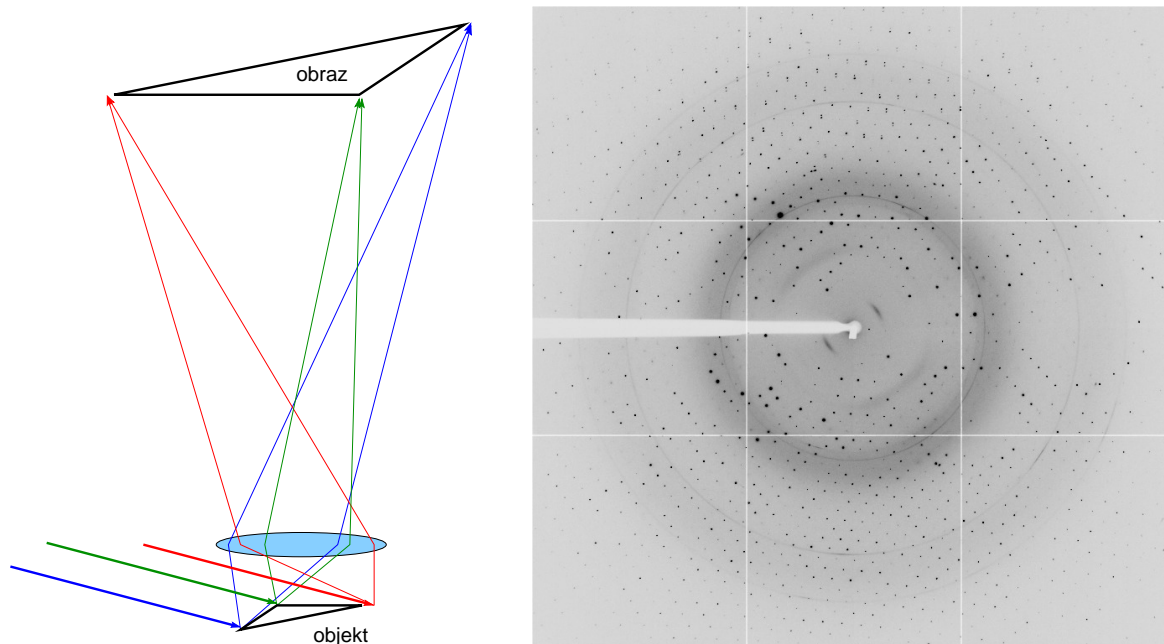
# Rentgenová krystalografie

### 12.1 Virtuální náhrada mikroskopu

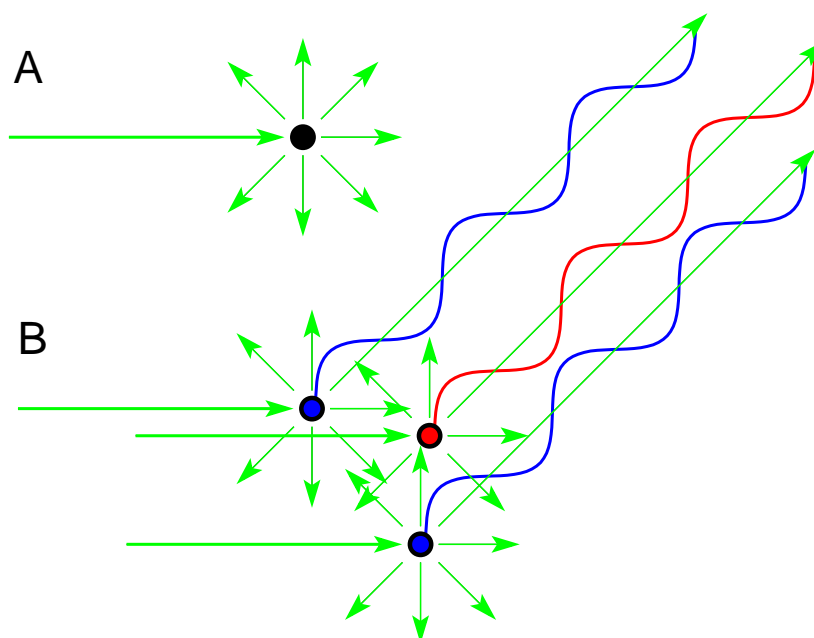
V kapitole 7 jsme narazili na nepříjemnou skutečnost. Není možné sestrojít mikroskop se zvětšením potřebným pro pozorování jednotlivých molekul, protože neexistují čočky schopné zaostřovat „světlo“ dostatečně krátké vlnové délky (rentgenové paprsky o vlnové délce kolem 0.1 nm). Čočky mikroskopu modifikují dráhu světla jednoduchým způsobem, který můžeme nahradit výpočetní operací. Podívejme se na obrázek velmi jednoduchého mikroskopu 12.1A. Co se vlastně v takovém mikroskopu děje? Na sledovaný objekt dopadá světlo, které je objektem rozptýlováno do všech směrů. Čočka mikroskopu láme rozptýlené paprsky a zaostřuje je do obrazu, který může být zvětšený a převrácený, jako v našem případě. Když nahradíme čočku (pro paprsky krátké vlnové délky nedostupnou) zařízením, které zaznamená, jak jsou jednotlivé paprsky rozptýlovány, mohli bychom z amplitud a fází rozptýleného záření vypočítat, jak by obraz vypadal. To je podstata určování tvaru molekuly pomocí rentgenové difrakce. Posvítíme na molekuly koherentním zářením vhodné vlnové délky (odpovídající vzdálenosti atomů v molekule, tedy asi 0.1 nm) a měříme, jak se na nich rozptýluje světlo. Protože citlivost měření je omezená, nepracujeme s jedinou molekulou, ale, jak je v biochemii zvykem, s velkým množstvím molekul. Připravíme krystal, ve kterém jsou molekuly uspořádány v pravidelné mřížce. Záření rozptýlené na krystalu je v určitých směrech mohutně zesíleno. Difrakční obrazec, zachycující v kterém směru je záření zesíleno, vypadá proto tak, jak ukazuje obrázek 12.1B. Každá tmavá skvrna na detektoru nám říká, že v odpovídajícím směru bylo záření zesíleno. Bez tohoto zesilovacího efektu krystalu je potřeba použít extrémně silný zdroj rentgenového záření. V současnosti jsou na světě dostupné pouze dva takové zdroje (*free electron lasers*). Většina strukturálních experimentů proto stále využívá analýzy krystalů.

Pro pochopení zesilujícího efektu krystalu a hlavně pro vyhodnocení pozorovaného difraktovaného záření je třeba porozumět tomu, co se s koherentním zářením při rozptýlu na molekulách děje. Jak jsme si již řekli v kapitole 11, elektromagnetická vlna rozkmitává elektrony molekul. V případě elastického rozptýlu je každý kmitající elektron zdrojem nového záření (o stejné vlnové délce jako budící záření), které se šíří do všech směrů (obrázek 12.2). Vlny vyzářené různými elektrony mají tedy stejnou vlnovou délku, ale v závislosti na poloze elektronů v krystalu mají různou délku optické dráhy a jsou vůči sobě různě fázově posunuté (obrázek 12.2). Právě na fázovém posunu záleží, jestli se paprsky budou skládat konstruktivně a záření se zesílí (to když mají jednotlivé vlny podobnou fázi), nebo jestli se amplitudy odečtou a vlny vzájemně vyruší (to když mají fázi opačnou).

Základní otázka tedy zní: Kdy mají rozptýlené vlny stejnou fázi? Odpověď na tuto otázku vysvětlí, proč krystal funguje jako zesilovač rozptýleného záření a proč pozorovaný difrakční obrazec krystalu obsahuje jen omezený počet skvrn.



Obrázek 12.1: Vlevo princip nejjednoduššího mikroskopu, vpravo příklad difrakčních dat.



Obrázek 12.2: Rozptyl záření na jednom atomu (A) a difrakce na více atomech (B). Záření do zvýrazněného směru rozptýlené na modře označených atomech se skládá konstruktivně (vlna šířící se daným směrem má stejnou fázi), zatímco záření rozptýlené na červeně označeném atomu je fázově posunuto.

## 12.2 Braggovy roviny

Kdy dojde k rozptylu koherentních světelných paprsků tak, aby měly všechny stejnou fázi? Existují dvě možnosti:

### 1. Paprsky mají stejnou délku optické dráhy

Rozptýlené záření se šíří ze zdroje všemi směry. Představme si, že objektem proložíme rovinu (může mít libovolnou orientaci a můžete si ji představit také jako zrcadlo, jak ukazuje obrázek 12.3A). Jeden z možných směrů šíření rozptýleného záření je takový, že úhel, pod kterým záření vychází (úhel odrazu od pomyslné roviny), se rovná úhlu dopadu zdrojového paprsku na tuto rovinu. Všechny rozptýlené paprsky, které se od pomyslné roviny „odráží“ pod stejným úhlem, pod jakým na rovinu dopadl zdrojový paprsek, mají stejnou fázi, bez ohledu na to, v kterém místě na rovinu (zrcadlo) dopadly. Obrázek 12.3A vysvětluje proč. Dva koherentní paprsky mají ve chvíli, kdy protnou úsečku HK stejnou fázi. První z nich narazí na zrcadlo pod úhlem  $\theta$  v místě označeném K. Z tohoto bodu se šíří rozptýlené záření všemi směry, nás ale teď zajímá pouze směr s úhlem odrazu  $\theta$ . Čím je tento směr vyjímecný? Tím, že dráha, kterou musí urazit druhý paprsek před dopadem ( $\Delta'$ ) navíc oproti prvnímu paprsku, je stejná jako dráha, kterou zase urazí navíc první paprsek po odrazu ( $\Delta$ ). A to bez ohledu na to, jak jsou body K a L daleko od sebe. Ve chvíli, kdy paprsky protnou úsečku LM, mají opět stejnou fázi. Pověšimněte si také, že směr paprsku se po odrazu změní o *dvojnásobek* úhlu dopadu, tedy o  $2\theta$ . Pokud budeme chtít z difrakčních obrazců určit úhel odrazu od pomyslné roviny, musíme odchylku difraktovaného paprsku od původního směru vždy vydělit dvěma. Kvůli analogii mezi odrazem světla od pomyslné roviny (zrcadla) v krystalu se difrakčním skvrnám říká *reflexe*, protože krystal si můžeme představit jako soustavu zrcadel (*reflektorů*), která odrážejí paprsky do různých směrů. Těmto pomyslným zrcadlům říkají krystalografové *Braggovy roviny*. Zákon odrazu říká, že úhel odrazu světla od zrcadla se rovná úhlu dopadu. Tato vlastnost zrcadla zajistí, že odražené paprsky si zachovají stejné relativní fáze, jako měly před dopadem na zrcadlo. Díky tomu můžeme v zrcadlech pozorovat obrazy předmětů.

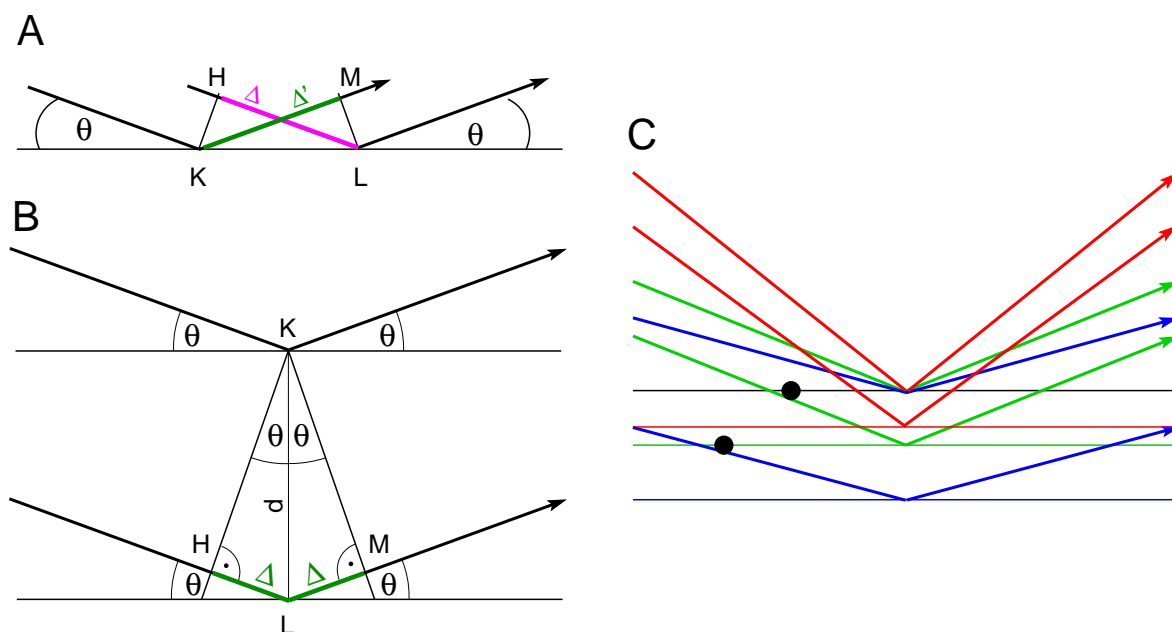
### 2. Rozdíl délky optických drah paprsků je roven celočíselnému násobku jejich vlnové délky

Zatím jsme se zabývali tím, co se stane, když se dva paprsky odrážejí od *stejně* Braggovy roviny. Co když se ale každý odrazí od *jiné* roviny<sup>1</sup> (obrázek 12.3B)? Délky optických drah těchto paprsků budou v takovém případě různé. Odražené paprsky proto nebudou mít stejnou fázi vždycky, ale jen tehdy, když bude rozdíl délek optických drah paprsků roven celočíselnému násobku vlnové délky  $\lambda$ . Jaká musí být vzdálenost Braggových rovin, aby byla tato podmínka splněna? Vysvětlení je na obrázku 12.3B. Rozdíl v délce optických drah paprsků závisí na vzdálenosti pomyslných rovin a na úhlu dopadu zdrojového paprsku (úhel odrazu musí být stejný, jako úhel dopadu). Rozdíl v délce optických drah paprsků odražených od různých rovin je roven dvojnásobku vzdálenosti  $\Delta$ . Obrázek 12.3B ukazuje, že úhel HKL je stejný jako úhel  $\theta$ . Součet úhlů v trojúhelníku je  $180^\circ$  a součet úhlů KLH a  $\theta$  je  $90^\circ$ . Pak můžeme spočítat, že  $\Delta = d \sin \theta$ . Aby měly paprsky stejnou fázi po rozptylu na různých rovinách, je třeba, aby rozdíl v délce jejich optických drah (dvojnásobek vzdálenosti  $\Delta$ ) byl roven celočíselnému násobku vlnové délky záření  $2\Delta = N\lambda$  (kde  $N$  značí přirozené číslo). Závislost mezi vlnovou délkou záření a vzdáleností mezi Braggovými rovinami je tedy  $N\lambda = 2d \sin \theta$ .

Analýza obrázku 12.3 nám poskytla důležitý vztah:

$$d = \frac{N\lambda}{2 \sin \theta}. \quad (12.1)$$

<sup>1</sup>Jak ukazuje obrázek 12.3A, fáze je stejná pro paprsky odražené od jakéhokoli místa stejné roviny, proto můžeme body K a L v rovinách libovolně posouvat.



Obrázek 12.3: Ohyb záření (difrakce) na Braggových rovinách.

Objekty, které leží na Braggových rovinách rozptylují světlo tak, že jednotlivé paprsky budou mít shodnou fázi a budou se tedy skládat konstruktivně (jejich amplitudy se budou sčítat). Naproti tomu objekty, které mají mezi sebou jinou vzdálenost než je vzdálenost Braggových rovin budou rozptylovat světlo s odlišnou fází.

Rovnice 12.1, známá jako *Braggův zákon*, nám také říká, že větší úhly ohybu pozorované v difrakčních obrazcích odpovídají menším vzdálenostem mezi Braggovými rovinami. Tato nepřímá úměra je základem konceptu takzvaného *reciprokého prostoru*, se kterým krystalografové často pracují.

Graficky lze vztah 12.1 vyjádřit obrázkem 12.3C. Zeleně, modře a červeně jsou vyznačeny Braggovy roviny různě vzdálené od roviny první (černé). Šipky odpovídajících barev ukazují směry, ve kterých má světlo rozptylované na jednotlivých rovinách stejnou fázi (celkový dráhový rozdíl  $2\Delta$  se rovná vlnové délce  $\lambda$ ).

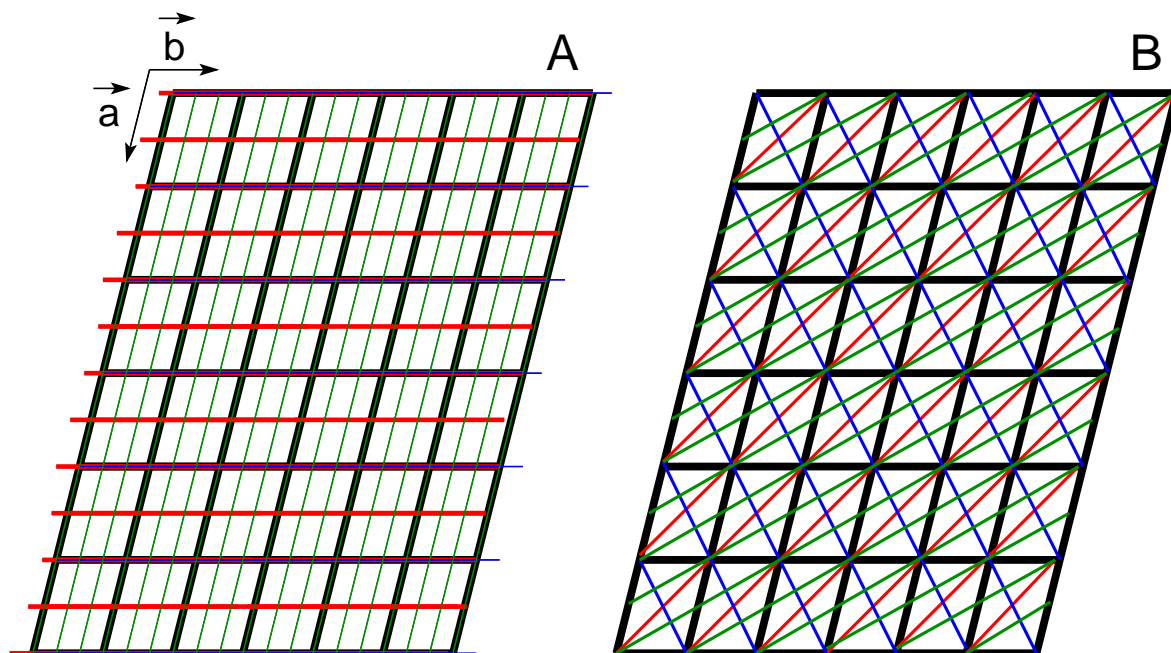
Pro pochopení difrakce je dobré zamyslet se nad tím, jaký nejmenší a největší difrakční úhel můžeme pozorovat. Rozptyl světla pod nejmenším (nulovým) úhlem nastane, když se paprsek neodchyluje od původního směru ( $\theta = 0^\circ$ ) a dráhový rozdíl mezi paprsky rozptýlenými jednotlivými elektrony je nulový. To odpovídá nekonečné vzdálenosti mezi Braggovými rovinami. Paprsek putující v původním směru proto nenese žádnou informaci o vzdálenostech, nese pouze informaci o celkovém počtu elektronů („hmotě“ objektu). Naopak největší difrakční úhel odpovídá kolmému odrazu ( $\theta = 90^\circ$ ), kdy se paprsek odráží zpět do směru, ze kterého přiletěl (směr paprsku se mění o  $180^\circ$ ). V takovém případě se dráhový rozdíl rovná polovině vlnové délky. To vysvětluje, proč nemůžeme za normálních okolností pomocí elektromagnetických (nebo jiných) vln zkoumat objekty menší, než je polovina vlnové délky záření (takzvaný difrakční limit).

## 12.3 Difrakce na krystalu

Pro vysvětlení difrakce jsme použili popis zkoumaného objektu jako soustavy rovin (zrcadel), na kterých jsou rozmístěny elektrony. Zatím jsme se ale nezabývali tím, jaká je orientace Braggových rovin v krystalu. Orientace těchto rovin určuje směry, do kterých krystal vysílá intenzivní rozptýlené záření. Nyní se blíže podíváme na to, kde se ony roviny (zrcadla) v krystalu nalézají. Ideální krystal je tvořen velkým počtem uspořádaných molekul. Nejmenší jednotka krystalu, která se pravidelně opakuje ve všech směrech, se nazývá *elementární buňka*. Jak jsme si již říkali, krystal zesiluje rozptýlené záření v určitých směrech. Intenzivní rozptýlené záření pozorujeme ve směrech, do kterých rozptylují různé elementární buňky záření se stejnou fází. Aby měly všechny paprsky v určitém směru stejnou fázi, musí Braggovy roviny procházet stejnými body ve všech elementárních buňkách. Pak jsou splněny obě podmínky, kdy dochází k rozptylu rentgenových paprsků tak, aby v určitých směrech měly všechny rozptýlené paprsky stejnou fázi. Tyto podmínky jsou splněny, když bude vzdálenost mezi rovinami (zrcadly) rovna délce hrany jedné elementární buňky v krystalu, nebo když na délku jedné hrany připadne celočíselný násobek rovin. Na obrázku 12.4A vidíme dvojrozměrný krystal s elementárními buňkami vyznačenými černě. Jednotlivé hrany buňek mají velikosti  $a$  a  $b$ . Červené roviny protínají hranu  $a$  v polovině. Skácame-li z jedné červené roviny na druhou, posuneme se proto o délku hrany  $a$  přesně po dvou skocích. Hrany  $b$  tyto roviny vůbec neprotínají. Červené roviny můžeme proto označit indexem  $(20)$ . Modré roviny jsou vzdálené o délku hrany  $a$  a hranu  $b$  neprotínají. Označují se  $(10)$ , protože do sousední buňky nám stačí jeden skok. Zelené roviny dělí hranu  $b$  na tři části, hranu ve směru  $a$  neprotínají. Označení zelených rovin je  $(03)$ .

Trochu složitější příklady Braggových rovin jsou ukázány na obrázku 12.4B. Když se podíváte na červené roviny, zjistíte, že skok na sousední rovinu nás posune o jednu hranu elementární buňky jak ve směru  $\vec{a}$ , tak ve směru  $\vec{b}$ . Jde tedy o rovinu  $(11)$ . Modré roviny jsou uspořádány podobně, ale je tu jeden rozdíl. Pokud se skokem na sousední rovinu posuneme o jednu délku hrany  $dál$  ve směru  $\vec{a}$ , zjistíme, že ve směru  $\vec{b}$  se ocitneme o jednu buňku *zpět*. Takovou rovinu označujeme  $(1-1)$ . U zelených rovin k posunutí o jednu buňku ve směru  $\vec{a}$  potřebujeme dva skoky, zatímco ve směru  $\vec{b}$  se o buňku dál dostaneme jedním skokem. Jde tedy o rovinu  $(21)$ . Krystal používané pro krystalografickou analýzu mají tři rozměry. K popisu rovin, které je protínají v různých místech tak potřebujeme tři čísla. V krystalografii se tato tři čísla označují písmenky  $h$ ,  $k$ ,  $l$  a nazývají *Millerovy indexy*.

Jak víme z kapitoly 11, k rozptylu elektromagnetického záření dochází především na elektronech. Jakou informaci o poloze elektronu můžeme získat z jediné difrakce? Vysvětlili jsme si, že všechny objekty ležící na Braggově rovině budou rozptylovat světlo se stejnou fází. Čím více objektů bude na rovinách ležet, tím silnější difrakci budeme v daném směru pozorovat. Co když bude ale polovina objektů ležet na Braggových rovinách a polovina přesně uprostřed mezi nimi? Potom bude mít difrakční skvrna velmi malou intenzitu, protože objekty uprostřed Braggových rovin budou rozptylovat světlo v daném směru přesně s opačnou fází, než objekty na rovinách. Obecně se dá říci, že intenzita difrakční skvrny nám říká, jak moc jsou objekty soustředěny na rovinách, jejichž vzdálenost a orientace odpovídá úhlu difraktovaného záření. Fázový úhel difrakční skvrny (který nemůžeme přímo změřit) udává průměrnou polohu objektu ve směru kolmém na příslušnou Braggovu rovinu, neříká ale nic o poloze objektu v této rovině. Zkusme si to ukázat na příkladu dvou elektronů, nakreslených na obrázku 12.3C jako černé body. Pokud budeme sledovat difrakci v modrém směru, budou mít paprsky rozptýlené na jednotlivých elektronech opačnou fázi a vyruší se, protože spodní elektron leží přesně uprostřed mezi černou a modrou rovinou. Pokud budeme sledovat difrakci v zeleném směru, paprsky budou mít stejnou fázi a sečtou se, protože spodní elektron leží na zelené rovině. Pokud budeme sledovat difrakci v červeném směru, budou mít paprsky různou fázi. Částečně se tedy sečtou, ale intenzita difrakce bude menší. Z intenzit difrakcí v různých směrech tak můžeme spočítat jaká je vzájemná poloha elektronů ve svislém směru (na našem obrázku), ale ne, jaká je jejich vzájemná poloha ve směru vodorovném.



Obrázek 12.4: Braggovy roviny v krystalu.

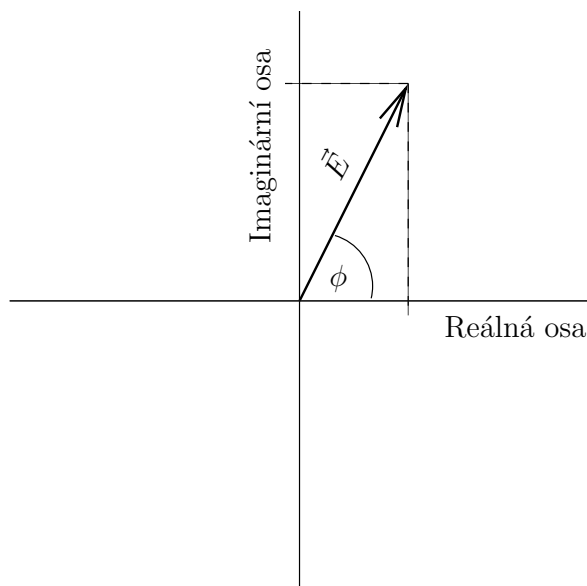
## 12.4 Strukturní faktor

Získání informace o rozmístění elektronů v krystalu představuje jádro rentgenové strukturní analýzy. Rozložení elektronů totiž vykresluje tvar molekuly, který hledáme. Pokud chceme z rozptylu rentgenových paprsků na krystalu určit rozmístění elektronů v molekulách, které tvoří krystal, musíme difrakci popsat matematickými rovnicemi.

Jak již víme z kapitoly 11, elektromagnetickou vlnu (v našem případě rentgenové záření s vlnovou délkou  $\approx 0,1$  nm) lze popsat, jako by byla složena z reálné a imaginární části (rovnice 11.7). Reálná část má tvar  $|E_0| \cos \phi$  a imaginární část má tvar  $|E_0| \sin \phi$  (je tedy posunuta o  $90^\circ$ ). Abysme zjistili, jaký bude výsledek interakce dvou vln, stačí sečíst jejich reálné a imaginární složky. Vlny tak můžeme skládat jako vektory v takzvaném *Argandově diagramu* (obrázek 12.5), ve kterém kreslíme jednotlivé složky v rovině komplexních čísel. Abychom nemuseli skládat vektory manuálně, popíšeme vektory pomocí matematického vzorce. Reálná složka  $|E_0| \cos \phi$  a imaginární složka  $|E_0| \sin \phi$  elektromagnetické vlny mohou být popsány v jedné rovnici s použitím komplexních čísel  $|E_0| \cos \phi + i|E_0| \sin \phi$ . Imaginární jednotka  $i$  značí, že složka  $|E_0| \sin \phi$  je orientována kolmo ke směru reálné složky (obrázek 12.5). Pro další zjednodušení můžeme přepsat popis vlny do exponenciální formy  $|E_0| \cos \phi + i|E_0| \sin \phi = e^{-i\phi}$  (rovnice 11.7). S použitím tohoto matematického aparátu můžeme popsat výslednou vlnu po rozptylu záření na krystalu ve směru definovaném indexy  $h, k, l$ . Komplexní číslo popisující amplitudu a fázi rozptýleného záření nazýváme *strukturní faktor*:

$$F(hkl) = V \int_{x=0}^1 \int_{y=0}^1 \int_{z=0}^1 \rho(xyz) e^{2\pi i(hx+ky+lz)} dx dy dz, \quad (12.2)$$

kde  $F(hkl)$  je komplexní číslo popisující reálnou a imaginární složku rozptýleného záření,  $h, k, l$  jsou Millerovy indexy a  $V$  je objem elementární buňky analyzovaného krystalu. Integraci provádíme



Obrázek 12.5: Argandův diagram.

přes celý objem elementární buňky ve všech třech rozměrech vyjádřených frakčními souřadnicemi  $x$ ,  $y$ ,  $z$ . Tyto souřadnice jsou definovány jako zlomky hran elementární buňky, například  $x = 0,25$  znamená, že elektron se nachází v jedné čtvrtině elementární buňky ve směru hrany  $a$ . Hodnota  $\rho(x y z)$  je tedy elektronová hustota v bodě  $x, y, z$  v rámci elementární buňky.

Při určování struktury molekul z rentgenové difrakce se ale nesnažíme spočítat hodnotu strukturního faktoru. Naopak chceme spočítat distribuci elektronové hustoty, která nám ukáže tvar molekuly. Rovnici 12.2 lze převést do formy vhodné pro výpočet elektronové hustoty pomocí Fourierovy transformace

$$\rho(x y z) = \frac{1}{V} \sum_h \sum_k \sum_l F(h k l) e^{-2\pi i(hx+ky+lz)}. \quad (12.3)$$

Jednotlivé členy rovnice jsou stejné, jako v rovnici 12.2. Integrace byla nahrazena sumací, protože k difrakci dochází pouze ve směrech definovaných parametry krystalu (podle orientace "zrcadel" určené Millerovými indexy  $(h k l)$ ).

Strukturní faktor je komplexní číslo, které se skládá z reálné a imaginární části, stejně jako světelná vlna má amplitudu a fázi. Rovnici 12.3 můžeme tedy přepsat jako

$$\rho(x y z) = \frac{1}{V} \sum_h \sum_k \sum_l |F(h k l)| e^{-2\pi i(hx+ky+lz) + i\alpha(h k l)}. \quad (12.4)$$

Oproti rovnici 12.3 jsme provedli nahrazení  $F(h k l) = |F(h k l)| e^{i\alpha(h k l)}$ , kde  $|F(h k l)|$  značí amplitudu a  $\alpha(h k l)$  fázi strukturního faktoru. Elektronová hustota je vždy reálným číslem.

## 12.5 Krystalizace biomakromolekul

Pustili jsme se do rentgenové krystalografie biologicky zajímavých makromolekul, aniž bychom si řekli, jak krystaly tak velkých molekul můžeme získat. Přitom úspěch v určování struktury biomakromolekul

závisí hlavně na tom, jestli se nám podaří připravit dostatečně veliký a kvalitní krystal, abychom jej mohli použít jako mřížku pro difrakci rentgenového záření. Pojd'me proto tuto chybu rychle napravit.

Pro pochopení procesu krystalizace je nutné vědět něco o chování roztoků pevných látek. Nazveme-li nějakou chemikálii „pevnou látkou“, znamená to, že za běžných podmínek jsou její molekuly natěsnány na sebe bez větší volnosti pohybu. Pokud jsou molekuly poskládány pravidelně, mluvíme o krystalu. Co se stane, když takovou pevnou látku vložíme do vody (nebo jiného rozpouštědla)? Bude-li takové látky málo, její molekuly se rozdělí a každá z nich se obalí poměrně pravidelnou konstrukcí z molekul rozpouštědla. Takovému stavu říkáme *nenasycený roztok*. Rozpustnost většiny látek je ale omezená. Když budeme pevnou látku do roztoku dále přidávat, dospějeme do stavu, kdy další rozpouštění nebude možné. Připravili jsme nasycený roztok. Koncentrace rozpuštěné látky se tedy nebude dále zvyšovat a přidaná látka se bude prostě hromadit v pevném stavu. Zvyšování koncentrace rozpuštěné látky můžeme dosáhnout nejen přidáváním pevné látky, ale také odebráním molekul vody – odpařováním. Také tímto způsobem dojdeme do stavu, kdy nebude možné koncentraci látky dále zvyšovat, ale část se jí začne ukládat v pevném stavu, tedy budovat krystal nebo amorfní, neperiodickou strukturu. Pro stavbu krystalu je nejobtížnější položit základy pravidelného uspořádání molekul. Proto se bude zásadně lišit situace, kdy odpařujeme roztok látky, ve kterém již krystal přítomný je, od situace, kdy odpařujeme roztok bez přítomnosti krystalu. V prvním případě se po dosažení určité koncentrace budou další molekuly ochotně zařazovat do již přítomného pravidelného uspořádání krystalu a krystal poroste. V druhém případě molekuly po určitou dobu „nebudou umět“ pravidelnou strukturu vytvořit, i když bude roztok nasycený. Koncentrace rozpuštěné látky tedy dál bude růst. V takovém případě mluvíme o *přesyceném roztoku*. Při určité koncentraci může být výhodnost pevného stavu tak výrazná, že se prakticky samovolně začnou tvořit zárodky krystalů, které ihned začnou rychle růst.

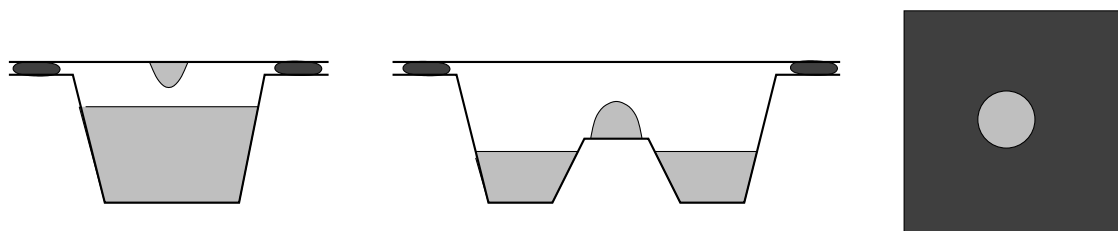
Při rychlém odpařování dochází k tomu, že koncentrace roste prudce a brzy se začne tvořit velké množství zárodků krystalů. Ty sice porostou rychle, ale brzy vyčerpají rozpuštěnou látku, takže výsledkem bude obrovské množství mikroskopických krystalků, které nejsou pro přesné určování struktury použitelné. Při pomalém odpařování naopak vzniká nejprve přesycený roztok. Pokud vložíme do přesyceného roztoku jeden malý krystal, budou se k němu molekuly rozpuštěné látky připojovat a můžeme tak vypěstovat velký pravidelný krystal. Proto je přesycený roztok, jako stav, ve kterém krystal může růst, ale samovolně nevzniká, pro pěstování velkých krystalů velmi důležitý.

Víme již, jak z malého krystalu vypěstovat velký. Kde ale vzít ten malý krystal, když určujeme strukturu zkoumané biomakromolekuly poprvé? Musíme najít takové podmínky krystalizace, kdy připravíme přesycený roztok a na okamžik zvýšíme koncentraci biomakromolekuly natolik, aby vznikl malý počet zárodků krystalů. Pak se vrátíme do přesyceného roztoku, ve kterém budou krystaly dále růst.

Oblíbeným krystalizačním postupem je *metoda visící kapky*, ilustrovaná vlevo na obrázku 12.6. Připravíme si asi mililitr roztoku látky, které krystalografové říkají srážedlo. Může to být organická nebo anorganická sůl, organické rozpouštědlo, nebo polyethylenglykol. Z roztoku srážedla odebereme několik mikrolitrů, smícháme se stejným objemem roztoku zkoumané biomakromolekuly a kápneme na vhodně upravené sklíčko. Koncentrace biomakromolekuly v kapce by měla být taková, aby za dané koncentrace srážedla byl její roztok téměř nasycený. Potom umístíme roztok srážedla do malé nádoby a sklíčkem nádobku přikryjeme tak, aby kapka visela nad hladinou srážedla. Nádobka se sklíčkem musí být dobře utěsněna, aby páry vody nemohly unikat ven.

Co se v nádobce bude dít? Koncentrace srážedla v kapce je po smíchání s roztokem biomakromolekuly poloviční ve srovnání s roztokem srážedla v nádobce. Tento rozdíl koncentrací se bude během času vyrovnávat. Pokud je srážedlo netěkavé, bude k vyrovnávání docházet tak, že molekuly vody se budou vypařovat z visící kapky a koncentrovaný roztok srážedla je bude postupně absorbovat. Protože objem kapky je mnohem menší, než objem roztoku srážedla (obvykle zhruba tisíckrát), objem visící kapky se ztrátou vody zmenší přibližně na polovinu. Během zmenšování kapky poroste koncentrace biomakromolekuly, vznikne přesycený roztok a v určitém okamžiku může dojít k vytvoření zárodku krystalu (takzvaného *nukleačního centra*). Tento krystalek začne růst a během růstu se spotřebuje část





Obrázek 12.6: Krystalizační metody visící kapky (vlevo), sedící kapky (uprostřed) a mikrodávek (vpravo).

biomakromolekul, takže jejich koncentrace v roztoku poklesne na úroveň přesyceného roztoku, v níž jsou krystaly stabilní, ale nová nukleační centra nevznikají. V ideálním případě poklesne koncentrace dříve, než se stačí vytvořit více zárodků, a v kapce nám poroste jediný krystal. Takové ideální (nebo alespoň téměř ideální podmínky) není snadné nalézt, obvykle je nutné vyzkoušet různé koncentrace řady srážedel.

Metoda visící kapky má různé alternativy. Pokud nám nestačí objem kapky, kterou kapilární síly udrží viset na víčku, můžeme použít *metodu sedící kapky* (obrázek 12.6 uprostřed). Moderní laboratoře svěřují zdoluhavé hledání vhodných podmínek automatickým robotům, pro které je výhodná *metoda mikrodávek (microbatch method)*. Při ní je do parafinového oleje umístěno velké množství kapiček s roztokem obsahujícím srážedlo a biomakromolekulu (obrázek 12.6 vpravo). V oleji nedochází k odpařování vody, takže se musíme strefit do podmínek, kdy je již na začátku koncentrace biomakromolekul taková, aby vzniklo jen malé množství zárodků krystalů. Tyto větší nároky na počáteční složení jsou vyváženy tím, že automat může vyzkoušet mnohem více počátečních podmínek.

Jak vidíme, pěstování krystalů biomakromolekul je spíše umění založené na pokusu a omylu, než vědecká disciplína. Jsme-li úspěšní, na konci našeho snažení nás čeká krásný pravidelný krystal o rozměrech desetin milimetru. Zajímavější než vnější vzhled je ale uspořádání molekul v takovém krystalu.

Možná vás překvapí, jak výrazně se krystaly biomakromolekul liší od toho, co nám při slově *krystal* většinou vytane na mysli a co zpravidla odráží naše znalosti o obsahu slánky či cukřenky. Složitě makromolekuly mají dosti nepravidelný tvar a je pro ně proto obtížné nalézt pravidelné krystalové uspořádání. Krystal proto obsahuje různé dutinky mezi jednotlivými makromolekulami, které jsou vyplněny krystalizačním roztokem. Objem roztoku je často srovnatelný s objemem zkoumané molekuly (kolísá mezi 30 % až 70 % celkového objemu krystalu makromolekul). Dutinky mezi makromolekulami jsou dost velké na to, aby se v nich mohla voda poměrně volně pohybovat. Přítomnost dutinek má pro krystalografu příjemné i méně příjemné důsledky. Značná volnost molekul vody a některých částí makromolekul způsobuje, že se makromolekuly mohou v krystalu částečně pohybovat a přesnost určování struktury je proto horší, než v případě krystalů malých molekul, které bývají těsně uspořádány. Na druhou stranu voda v dutinkách zaručuje, že ani v krystalu není biomakromolekula přítomná v „suchém stavu“. Prostředí, v kterém se molekula nachází, se spíše podobá velmi koncentrovanému roztoku a struktury biomakromolekul v krystalech se většinou velmi podobají strukturám v roztoku, o kterých bude řeč v kapitole 14.

## 12.6 Měření difrakce

Od přípravy krystalu nyní postoupíme k vlastnímu měření. Přístroj pro měření difrakce (*difraktometr*) má tři hlavní části: zdroj rentgenového záření, zařízení pro uchycení krystalu v různých polohách (*goniostat*) a detektor.

### Zdroj záření

Rentgenové záření je elektromagnetické vlnění o velmi krátké vlnové délce, tedy vysoké energii. Obecně jsou zdrojem elektromagnetických vln nabitě částice, které mění směr nebo rychlost pohybu. V klasické Roentgenově lampě dochází k vyzařování, když elektrony urychlené v silném elektrickém poli prudce brzdí při dopadu na kladnou elektrodu. Pro měření difrakce biomakromolekul se ale dnes již většinou používají jako zdroje záření *synchrotrony*. V tomto případě už lze těžko hovořit o přístroji, protože může jít o zařízení velikosti malého městečka. V synchrotronech jsou elektrony urychlené v lineárních a kruhových urychlovačích na rychlost blízkou rychlosti světla a přivedeny do takzvaného akumulárního prstence. V něm se pohybují po přibližně kruhové dráze, která se skládá z rovných úseků spojených ohybovými magnety. V každé zatáčce vyzařují elektrony úzký paprsek velmi intenzivního polarizovaného elektromagnetického záření. Protože pro většinu měření potřebujeme záření o jedné vlnové délce, je důležitou součástí zdroje *monochromátor*, zařízení, které nám umožní vybrat záření s přesně definovanou vlnovou délkou.

### Goniostat

Během měření je krystal uchycen v takzvaném *goniostatu*, který umožňuje krystalem během měření otáčet. V některých experimentech se používá goniostat, který umožňuje krystal natočit podle více os a tak jej orientovat libovolným způsobem. Většinou ale vystačíme s jednoosým goniostatem, který otáčí krystalem tam a zpátky v rozsahu několika stupňů.

### Detektor

Intenzitu paprsků difraktovaných na krystalech biomakromolekul v různých směrech měříme pomocí plošných detektorů<sup>2</sup>, zaznamenávajících intenzitu a polohu dopadajících paprsků. Příkladem může být film citlivý na rentgenové paprsky, obrazová deska<sup>3</sup> (*image plate*), nebo polovodičový detektor typu fotodiodových polí nebo CCD (*charge-coupled device*, používá se v digitálních fotoaparátech). Výsledkem měření je difrakční snímek (obrázek 12.1B)

## 12.7 Vyhodnocení dat

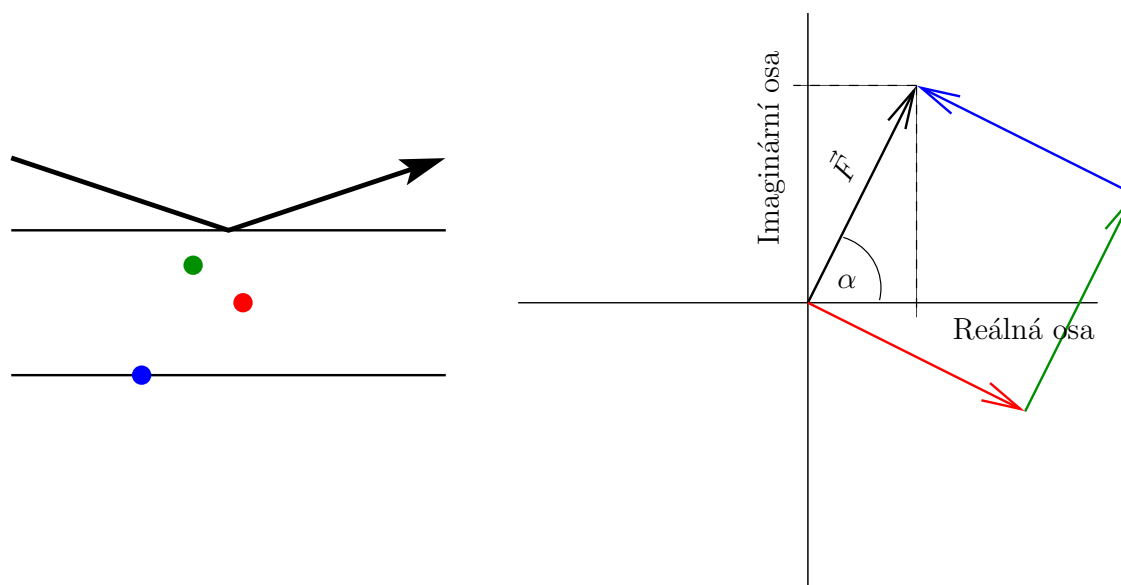
Zpracování naměřených dat můžeme rozdělit do tří kroků. Nejprve musíme určit mřížkové parametry a orientaci krystalu (vektory  $\vec{a}$ ,  $\vec{b}$ ,  $\vec{c}$ ). Známe-li mřížkové parametry a orientaci krystalu, známe „adresy“ (indexy  $h$ ,  $k$ ,  $l$ ) míst, kde můžeme reflexe pozorovat. Víme tedy, v kterém místě na detektoru máme měřit intenzitu  $I$ .

V dalším kroku musíme určit celkové intenzity jednotlivých reflexí, odečíst pozadí a stanovit chybu měření.

Ve třetím kroku redukuje se naměřené intenzity na amplitudy strukturních faktorů. Intenzity naměřené detektorem jsou úměrné druhé mocnině absolutní hodnoty strukturního faktoru. Konstanty úměrnosti ale závisí na řadě dalších faktorů: na objemu krystalu, na rychlosti, se kterou krystalem otáčíme, na absorpci záření krystalem, na polarizaci záření, na postupném poškozování krystalu dopadajícím rentgenovým zářením. Tyto vlivy se mohou lišit pro jednotlivé reflexe, proto musí být provedeny různé opravy naměřených intenzit, abychom získali porovnatelné hodnoty. Tato část zpracování výsledků difrakčního pokusu, kdy určujeme velikost amplitudy strukturních faktorů, se nazývá *redukce dat*. Pro výpočet map

<sup>2</sup>V případě malých molekul lze měřit intenzitu záření v jednom směru pomocí fotonásobiče se scintilačním krystalem, který vyzařuje viditelné světlo po dopadu rentgenového záření. V tomto případě opakujeme měření s orientacemi krystalu a detektoru, které odpovídají jednotlivým reflexím. V případě makromolekul je ale reflexí tolik, že je není možné jednotlivě proměřovat.

<sup>3</sup>Jde o moderní obdobu fotografické desky, která se nevyvolává ponořením do vývojky, ale čte pomocí laseru a fotonásobiče.



Obrázek 12.7: Strukturální faktor trojice elektronů. Vlevo je znázorněna poloha tří elektronů mezi Braggovými rovinami, vpravo Argandův diagram ukazující sčítání strukturálních faktorů jednotlivých elektronů (strukturální faktory a elektrony jsou nakresleny stejnými barvami). Černá šipka představuje výsledný strukturální faktor.

elektronové hustoty potřebujeme ještě zjistit fázi jednotlivých strukturálních faktorů. To je úkol natolik složitý, že mu věnujeme samostatnou diskusi.

## 12.8 Řešení fázového problému

Výsledkem difrakčního experimentu jsou skvrny v difrakčním obrazci. Každá z těchto skvrn odpovídá vlně šířící se daným směrem, vlně s určitou amplitudou a fází. Vlnová rovnice v komplexním tvaru (rovnice 11.7) dokumentuje, že vlnu můžeme popsat pomocí komplexního čísla. Pro přehlednost je užitečné znázornit toto komplexní číslo graficky, jako vektor v rovině, kde vodorovná souřadnice udává reálnou složku a svislá imaginární složku. Amplituda pak odpovídá délce šipky znázorňující vektor a fázi popisuje úhel, o který je šipka otočena od reálného směru. Skládání rozptýlených vln lze graficky zobrazit jako součet vektorů popisujících jednotlivé vlny. Příklad pro tři elektrony je ukázán na obrázku 12.9. Rozptyl v určitém směru odpovídá odrazu na Braggových rovinách určité vzdálenosti. Fáze vln rozptýlených na jednotlivých elektronech závisí na tom, jak kde se daný elektron nachází mezi Braggovými rovinami: elektron přesně uprostřed mezi rovinami rozptyluje s opačnou fází, než elektron ležící přímo na Braggově rovině (na obrázku mají vektory popisující rozptyl na těchto elektronech opačný směr). Výsledný černý vektor odpovídá strukturálnímu faktoru trojice elektronů, který popisuje výslednou vlnu rozptýlenou v daném směru.

Rovnice 12.2 nám říká, že strukturální faktory pro rozptyl v různých směrech (tedy pro různé body reciproké mřížky popsané Millerovými indexy  $h$ ,  $k$ ,  $l$  odpovídající jednotlivým difrakčním skvrnám) jsou dány Fourierovou transformací rozložení elektronové hustoty v elementární buňce. Pro nás je zvlášť důležité to, že Fourierova transformace funguje i opačným směrem, takže ze strukturálního faktoru bychom mohli přímo vypočítat rozložení elektronové hustoty. A struktura molekuly je dána právě výskytem elektronů. Opačnou Fourierovu transformaci můžeme matematicky popsat rovnicí 12.3.

Ukázali jsme si, že rovnici 12.3 můžeme přepsat jako

$$\rho(xyz) = \frac{1}{V} \sum_h \sum_k \sum_l |F(hkl)| e^{-2\pi i(hx+ky+lz)+i\alpha(hkl)} \quad (12.5)$$

(rovnice 12.5), kde  $|F(hkl)|$  značí amplitudu a  $\alpha(hkl)$  fázi strukturního faktoru.

Kdybychom uměli měřit amplitudu i fázi strukturního faktoru, stačilo by prostě nechat počítač zpracovat naměřená data podle rovnice 12.5. Získali bychom ihned rozložení pravděpodobnosti výskytu elektronů v elementární buňce (takzvanou *mapu elektronové hustoty*), což je vlastně tvar molekuly.

Ve skutečnosti ale během experimentu pouze naměříme, *kolik* fotonů v daném směru dopadne na detektor, ne jakou mají fázi. Počet fotonů je úměrný druhé mocnině amplitudy vlny, informaci o fázi ztrácíme. Této komplikaci říkáme *fázový problém* a velká část krystalografie je věnována jeho řešení.

## 12.9 Pattersonova mapa

Rovnice 12.3 nám říká, jak bychom získali elektronovou hustotu, kdybychom provedli Fourierovu transformaci strukturních faktorů. Patterson si položil otázku, co získáme, když místo toho spočítáme Fourierovu transformaci toho, co nám poskytne měření, tedy Fourierovu transformaci druhých mocnin amplitud strukturního faktoru. Takový výpočet je ekvivalentní rovnici 12.3, ve které použijeme pro všechny fázové úhly  $\alpha$  hodnotu  $0^\circ$ . Místo mapy elektronové hustoty získáme jinou mapu, zvanou Pattersonova mapa. Tato mapa je v krystalografii velmi užitečná.

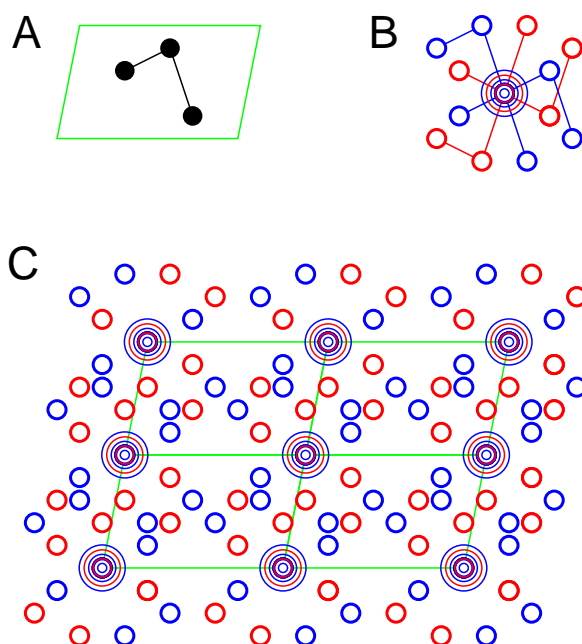
Pattersonova mapa je mapou *meziatomových vektorů*. Jinými slovy, pokud bychom v mapě elektronové hustoty našli vysokou hustotu elektronů v místě popsaném vektorem  $\vec{r}_1$  a v místě popsaném vektorem  $\vec{r}_2$ , v Pattersonově mapě bychom viděli maxima v místech pospaných vektory  $\vec{r}_1 - \vec{r}_2$  a  $\vec{r}_2 - \vec{r}_1$ . Výšky maxim by přitom odpovídaly součinu výšek maxim v místech  $\vec{r}_1$  a  $\vec{r}_2$  mapy elektronové hustoty. Na obrázku 12.8 je ukázka Pattersonovy mapy jednoduché molekuly. Pattersonova mapa je vlastně kombinace několika posunutých obrazů molekuly, s tím, že každý z obrazů je posunut tak, aby jeden atom ležel v počátku buňky. Protože mapa zobrazuje vektory typu  $\vec{r}_1 - \vec{r}_2$  i  $\vec{r}_2 - \vec{r}_1$ , najdeme v ní i převrácené obrazy molekuly.

Pokud není atomů moc, můžeme z Pattersonovy mapy určit původní polohy atomů v elementární buňce. Pattersonova mapa nám tedy umožní získat tvar molekuly i bez znalosti fáze! S rostoucím počtem atomů takový postup ale brzy selhává. Z poloh  $N$  atomů v elementární buňce můžeme sestavit  $N^2 - N$  nenulových rozdílových vektorů. Pro desetiatomovou molekulu tak získáme mapu s devadesáti maximy, kterou je možné rozumně analyzovat. Ale pro malý protein skládající se z tisíce atomů bude Pattersonova mapa obsahovat 999 000 maxim, která splynou v nerozlišitelnou směs. V kombinaci s dalšími triky je ale Pattersonův přístup základem metod řešení fázového problému v krystalografii velkých molekul.

## 12.10 Metoda molekulárního přemístění

Metoda molekulárního přemístění (*molecular replacement*) umožňuje vyřešit fázový problém v případě, že máme dobrý model molekuly, jejíž strukturu určujeme. V praxi to znamená, že již dříve někdo vyřešil strukturu dostatečně podobné molekuly, což je dnes již běžný případ. Podobnost terciárních struktur dvou proteinů většinou odráží podobnost jejich primárních struktur, takže srovnáním sekvencí získáme dobrý odhad toho, zda bude molekulární přemístění úspěšné. Pokud je známá struktura dostatečně kompletní a sekvence obsahují alespoň 40% identických aminokyselin, je velká naděje, že metoda molekulárního přemístění nám umožní vypočítat fázové úhly strukturních faktorů.

Molekulové přemístění vyžaduje, abychom modelovou strukturu vložili ve správné orientaci a do správného místa v rámci elementární buňky našeho krystalu. K tomu, abychom molekulu natočili, musíme určit tři rotační úhly, a k tomu, abychom ji umístili v elementární buňce, musíme zadat tři polohové parametry, popisující posunutí v prostoru. Obecně je molekulové přemístění šestiřozměrný



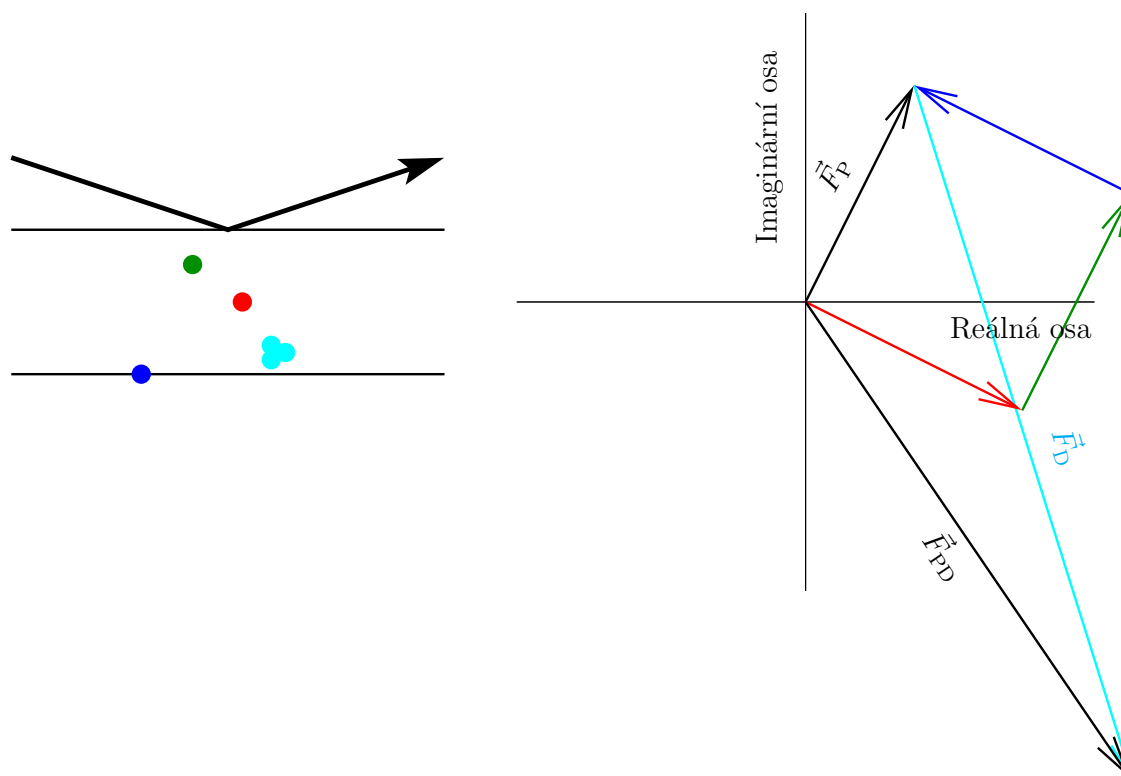
Obrázek 12.8: Pattersonova mapa. Na obrázku A je černě zobrazena jednoduchá molekula v zelené elementární buňce. Obrázek B ukazuje posunutí molekuly vedoucí k překryvu atomů. Obrázek C představuje několik buněk Pattersonovy mapy této molekuly.

problém. Většinou lze ale úkol řešit ve dvou krocích, nejdříve spočítat orientaci a pak správně natočenou molekulu posunout do správného místa.

K pochopení rotačních a translačních funkcí nám pomůže Pattersonova mapa. Ačkoli pro molekulu velikosti proteinu nezískáme rozlišené vzdálené vektory, způsob jejich akumulace je specifický pro strukturu a orientaci makromolekuly. Vektory v Pattersonově mapě můžeme rozdělit na dva druhy. *Intramolekulární* vektory (mezi atomy v téže molekule) závisí pouze na orientaci, ne na pozici molekuly v buňce. *Intermolekulární* vektory (mezi atomy různých molekul) závisí na natočení i na pozici, takže po určení správné orientace nám pomohou najít potřebné posunutí. Protože vzdálenosti v rámci molekuly jsou v průměru kratší, intramolekulární vektory jsou většinou kratší. Proto můžeme rotaci spočítat, když použijeme pouze část Pattersonovy mapy v okolí počátku.

## 12.11 Přímé metody

Pokud předpokládáme, že se molekuly skládají z atomů podobných tvarů s kladnou elektronovou hustotou, můžeme nalézt statistickou závislost mezi sadami strukturálních faktorů. Tyto vztahy nám mohou pomoci najít hodnoty fází. Zmíněné statistické závislosti jsou dostatečně výrazné pouze pro krystaly s velmi vysokým rozlišením a nepříliš velkým počtem atomů v elementární buňce (v řádu stovek). Proto jsou přímé metody použitelné k určení struktur proteinů jen vyjíměčně. V rámci experimentálních přístupů k řešení fázového problému jsou ale přímé metody užitečné.



Obrázek 12.9: Ovlivnění celkového strukturního faktoru přítomností těžkého atomu. Vlevo je znázorněna poloha tří lehkých atomů (nakreslených červeně, zeleně a modře) a jednoho těžkého atomu (nakresleného azurovou barvou) mezi Braggovými rovinami, vpravo Argandův diagram ukazující sčítání strukturních faktorů jednotlivých atomů (strukturní faktory a atomy jsou nakresleny stejnými barvami). Černá šipka představuje výsledný strukturní faktor pro molekulu bez těžkého atomu ( $\vec{F}_P$ ) a s těžkým atomem ( $\vec{F}_{PD}$ ).

## 12.12 Metody izomorfního nahrazení

Metody izomorfního nahrazení jsou založeny na předpokladu, že dokážeme krystal modifikovat tak, abychom ovlivnili strukturní faktor, ale ne strukturu a uspořádání makromolekul v krystalové mřížce. Ze způsobu, jakým změníme strukturní faktor, se pak snažíme odhadnout hodnoty fází. Důležité je, abychom zavedené změny dokázali popsat malým počtem parametrů. Proto se snažíme nahradit některé atomy makromolekuly (nebo na ni navázat) atomy s výrazně větším počtem elektronů. Takové atomy, běžně označované jako *těžké atomy*, významně ovlivní intenzity rozptýlených paprsků. Elektrony *jednoho atomu* jsou soustředěny do malého prostoru, nacházejí se tedy všechny přibližně ve stejné vzdálenosti od Braggových rovin a mají proto v podstatě stejnou fázi. Proto je příspěvek jednoho atomu ke strukturnímu faktoru úměrný počtu elektronů daného atomu. Na rozdíl od toho se elektrony *různých atomů* vyskytují v rozdílných polohách vzhledem k Braggovým rovinám a rozptylují tedy záření s rozdílnou fází. Jejich kombinovaný příspěvek ke strukturnímu faktoru je proto relativně malý. Když vezmeme do úvahy, že intenzita rozptýleného záření je úměrná druhé mocnině amplitudy strukturního faktoru (délky šipky), pochopíme, proč jeden atom uranu, který má patnáctkrát víc elektronů než uhlík, přispěje k intenzitě rozptýlených paprsků přibližně stejně, jako 225 atomů uhlíku.

Většina biomakromolekul se ovšem skládá pouze z „lehkých“ prvků. Jak jsme si řekli v sekci 12.5, krystaly biologicky zajímavých makromolekul obsahují velké množství dutinek a kanálků. Ionty těžkých

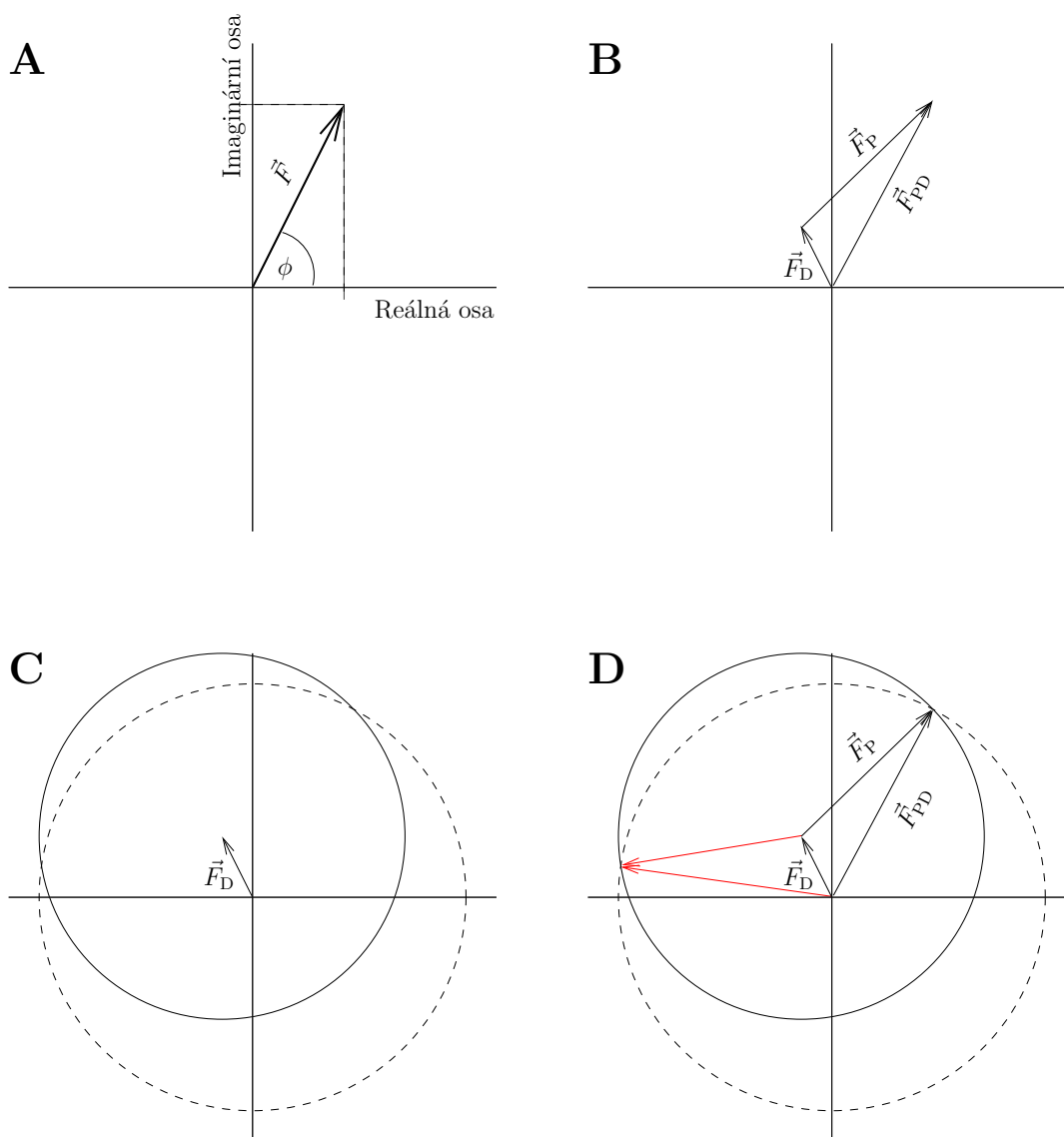
kovů mohou prostupovat těmito kanálky a dostanou se tak ke všem makromolekulám v krystalu. Většinou stačí přidat roztok soli těžkého kovu ke krystalu makromolekuly. Jinou možností je zavedení těžkého atomu přímo do biomakromolekuly ještě před krystalizací. Například v proteinech je možné nahradit v cysteinu síru selenem, v DNA použít místo thyminu bromuracil a podobně. Všechny tyto metody vycházejí z předpokladu, že těžké atomy neovlivní konformaci biomakromolekuly ani tvar krystalové mřížky – odtud i název *izomorfní nahrazení*, tedy nahrazení biomakromolekuly jejím derivátem, který má stejný tvar. Aby izomorfní nahrazení bylo úspěšné mřížkové parametry krystalů makromolekuly s těžkým kovem a bez těžkého kovu se nesmí lišit o více než 1 %.

Pokud se nám podaří připravit dva krystaly se stejnými mřížkovými parametry, jeden obsahující pouze nativní protein a druhý protein s navázanými těžkými atomy, můžeme naměřit difrakční data obou. Rozdíly v pozorovaných intenzitách rozptýleného záření odpovídajících reflexí z těchto dvou experimentů budou dány příspěvkem těžkých atomů. Z těchto rozdílů můžeme vypočítat Pattersonovu mapu, která bude zachycovat pouze rozložení těžkých atomů v elementární buňce. Pokud elementární buňka obsahuje malý počet těžkých atomů, je možné takovou Pattersonovu mapu analyzovat a určit, kde se těžké atomy v buňce nacházejí. Ze známého rozmístění těžkých atomů dokážeme Fourierovou transformací vypočítat jejich strukturní faktor včetně fáze. Jak nám to pomůže v hledání fáze strukturního faktoru proteinu samotného?

Podstatou řešení je skutečnost, že strukturní faktor krystalu s těžkým atomem můžeme vyjádřit jako součet dvou komplexních čísel. Prvním z těchto komplexních čísel je strukturní faktor těžkých atomů získaný například z Pattersonovy mapy (pro něj známe amplitudu i fázi), druhým je strukturní faktor makromolekuly, který se rovná strukturnímu faktoru krystalu tvořeného biomakromolekulou bez těžkého atomu (pro ten známe amplitudu, ale ne fázi). Sčítání komplexních čísel je výhodné zobrazit v dvojrozměrném grafu, kde na vodorovnou osu vynášíme reálnou část a na svislou imaginární část (obrázek 12.10A). Náš případ je takto graficky znázorněn na obrázku 12.10B. Ve skutečnosti ovšem známe směr jen pro šipku  $\vec{F}_D$ , u ostatních šipek známe jen délku. Řešení tohoto problému je ukázáno na obrázku 12.10C. Nejprve zakreslíme vektor  $\vec{F}_D$ , pro který známe velikost i směr. Do obrázku zakreslíme také vektor  $\vec{F}_{PD}$ , který popisuje strukturní faktor makromolekuly spolu s těžkým kovem. Pro tento vektor známe pouze amplitudu, ale ne fázi. Informaci o amplitudě nakreslíme jako kružnici se středem v počátku souřadnic a poloměrem odpovídajícím amplitudě  $|F_{PD}|$ . Nakonec zakreslíme do obrázku to, co víme o vektoru  $\vec{F}_P$ , který odpovídá strukturnímu faktoru nativní makromolekuly. Opět známe pouze amplitudu, ale ne fázi. Také  $|F_P|$  zakreslíme jako kružnici. Součet  $\vec{F}_P$  a  $\vec{F}_D$  musí odpovídat  $\vec{F}_{PD}$ . Střed kružnice znázorňující amplitudu  $|F_P|$  proto umístíme na konec vektoru  $\vec{F}_D$ . Kružnice znázorňující amplitudu  $|F_P|$  a  $\vec{F}_{PD}$  se obvykle protnou ve dvou bodech. Jednoduché izomorfní nahrazení (SIR, *Single Isomorphous Replacement*) proto neumožňuje jednoznačné určení fází strukturních faktorů makromolekuly. Je třeba připravit alespoň dva krystaly s různými těžkými atomy. Graficky lze řešení znázornit opět v Argandově diagramu, do kterého v tomto případě zakreslíme dva známé vektory (strukturní faktory těžkých atomů ve dvou derivátech) a tři kružnice s poloměry odpovídajícími amplitudám strukturních faktorů nativní makromolekuly a jejich dvou derivátů. Jednoznačné řešení fázového problému je pak dáno průsečíkem všech tří kružnic. Toto je podstata metody *vícenásobného izomorfního nahrazení* (MIR, *Multiple Isomorphous Replacement*). Metoda MIR je základní metodou řešení fázového problému krystalů makromolekul neznámé struktury.

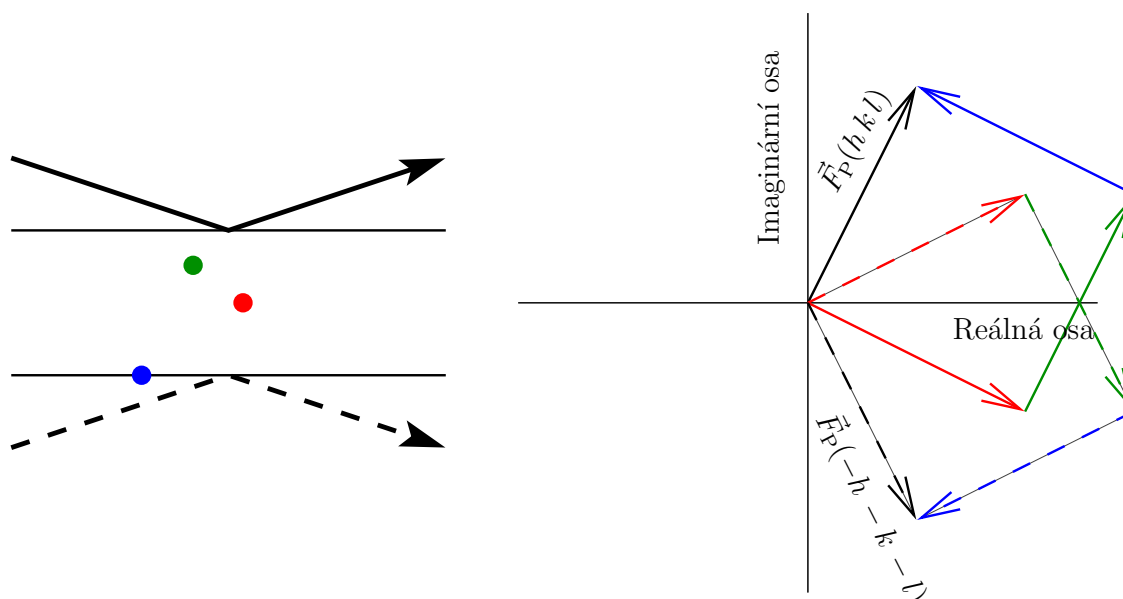
## 12.13 Metody využívající anomální rozptyl

Až dosud jsme předpokládali, že při průchodu rentgenového záření krystalem se paprsek chová k elektronům zcela netečně, že nedochází k výměně energie mezi molekulou a zářením. V takovém případě interagují všechny elektrony se zářením stejně. Pro difrakci je důležitý jen rozdíl fází jednotlivých rozptýlených paprsků určený relativní polohou elektronů mezi Braggovými rovinami, ne absolutní hodnota fáze. Proto můžeme fázi záření rozptýleného na elektronu v počátku souřadné soustavy krystalu



Obrázek 12.10: Strukturální faktor v Argandově diagramu (A) a řešení fázového problému izomorfním nahrazením (B, C a D). Na obrázku B jsou nakresleny strukturální faktory nederivatizované biomakromolekuly ( $\vec{F}_P$ ), biomakromolekuly derivatizované těžkým atomem ( $\vec{F}_{PD}$ ) a těžkých atomů ( $\vec{F}_D$ ). Fázi známe pouze pro  $\vec{F}_D$ , pro ostatní strukturální faktory známe jen amplitudy. Neznámé fáze můžeme určit způsobem ukázaným na obrázku C. Do Argandova diagramu zakreslíme známý vektor  $\vec{F}_D$ . Dále zakreslíme čárkovanou kružnici o poloměru  $|\vec{F}_{PD}|$ , která reprezentuje všechny možné vektory  $\vec{F}_{PD}$ . Nakonec zakreslíme i plnou kružnici o poloměru  $|\vec{F}_P|$ , znázorňující možné vektory  $\vec{F}_P$ . Protože víme, že součet  $\vec{F}_D$  a  $\vec{F}_P$  je rovný  $\vec{F}_{PD}$ , víme také, že počátek vektoru  $\vec{F}_{PD}$  (a tedy střed čárkované kružnice) leží v počátku vektoru  $\vec{F}_D$ , zatímco počátek vektoru  $\vec{F}_P$  musíme umístit do konce vektoru  $\vec{F}_D$  (aby na sebe šipky představující vektory navazovaly a znázorňovaly tak součet  $\vec{F}_D + \vec{F}_P = \vec{F}_{PD}$ , jak ukazuje obrázek B). Proto musíme střed plné kružnice položit do konce vektoru  $\vec{F}_D$ . Řešení ukazuje obrázek D. Místo, kam směřují vektory  $\vec{F}_P$  a  $\vec{F}_{PD}$ , musí ležet na obou kružnicích, průsečík kružnic nám tedy určuje konce obou hledaných vektorů. Obecně se ale plná a čárkovaná kružnice protnou ve dvou bodech, úloha má proto dvě možná řešení. Správné řešení je nakresleno černě, nesprávné červeně.





Obrázek 12.11: Friedelův pár v nepřítomnosti těžkého atomu. Vlevo je znázorněna poloha lehkých atomů (nakreslených červeně, zeleně a modře) mezi Braggovými rovinami, vpravo Argandův diagram ukazující sčítání strukturálních faktorů jednotlivých atomů (strukturální faktory a atomy jsou nakresleny stejnými barvami). Čárkovaně jsou nakresleny strukturální faktory pro odraz od spodní strany Braggových rovin. Černé šipky představují výsledné strukturální faktory  $\vec{F}_P(hkl)$  a  $\vec{F}_P(-h-k-l)$ . Tyto vektory mají stejnou reálnou složku a opačné imaginární složky, jde tedy o komplexně sdružená čísla.

definovat jako nulovou. Ostatní elektrony budou rozptylovat s odlišnou fází. Elektrony, které leží na Braggových rovinách, budou rozptylovat se stejnou fází ve směru definovaném Braggovou rovinou, fáze záření rozptýleného na elektronech mezi Braggovými rovinami je posunutá úměrně kolmé vzdálenosti elektronu od Braggovy roviny (plné šipky na obrázku 12.11). Stejná sada Braggových rovin popisuje ale také odraz na *opačné straně* roviny, tedy pod přesně opačným úhlem (čárkované šipky na obrázku 12.11). Jestliže odraz paprsku znázorněného plnou čarou na obrázku 12.11 odpovídá rovině s Millerovými indexy  $(hkl)$ , druhou stranu stejné roviny (od které se odráží čárkovaný paprsek) popisují Millerovy indexy  $(-h-k-l)$ . Dvojice reflexí s indexy  $(hkl)$  a  $(-h-k-l)$  se označuje jako *Friedelův pár*. Pokud nedochází k interakci záření s elektrony, mají strukturální faktory Friedelova páru fázi stejné velikosti, ale opačného znaménka. Matematicky řečeno, jde o *komplexně sdružená* čísla  $F(hkl) = F^*(-h-k-l)$ . Jinými slovy, čárkované šipky na obrázku 12.11 jsou zrcadlovým obrazem plných šipek.

Pokud ale dochází k interakci mezi zářením a elektrony, jinými slovy, pokud se frekvence záření blíží rezonanční frekvenci<sup>4</sup>, situace se změní. Pohlcený foton vybudí přechod elektronu do jiného stavu a vyzářený foton bude mít jinou amplitudu i fázi. Tento rozdíl se nazývá *anomální rozptyl*. Proto strukturální faktory Friedelova páru nejsou v přítomnosti anomálního rozptylu komplexně sdružené.

Rezonanční frekvence elektronů v atomech vodíku, uhlíku, dusíku, kyslíku a fosforu jsou příliš daleko vlnových délek rentgenového záření používaného pro analýzy makromolekul. Proto i metody využívající anomální rozptyl vyžadují přítomnost těžších atomů. Pokud určíme rozmístění atomů, pro které pozorujeme anomální rozptyl, v elementární buňce, můžeme vypočítat amplitudu i fázi strukturálních faktorů  $F_{PD}(hkl)$  a  $F_{PD}(-h-k-l)$  a vyřešit tak fázový problém. Pokud jsme schopni rozlišit intenzity re-

<sup>4</sup>Rezonanční frekvence je daná rozdílem energií dvou stavů elektronů v ozařovaných atomech, vyděleným Planckovou konstantou,  $\Delta E/h$ . V případě absorpce rentgenových paprsků mluvíme často o absorpční hraně, protože od určité frekvence je k dispozici velké množství přechodů.

flexí s opačnými hodnotami indexů  $h$ ,  $k$ ,  $l$ , můžeme tedy anomální rozptyl využít pro řešení fázového problému podobně jako v případě isomorfního nahrazení. Více informací o tomto přístupu můžete najít v Dodatku F.

K anomálnímu rozptylu dochází, když se frekvence rentgenových paprsků blíží rezonanční frekvenci přechodu mezi stavy elektronů v daném atomu. Síla anomálního rozptylu proto silně závisí na vlnové délce použitého záření. Pokud můžeme naměřit difrakční data při několika vlnových délkách blízkých absorpční hraně, můžeme získat informaci o fázi podobným způsobem, jako z měření několika různých derivátů při použití metody MIR. Tato metoda se nazývá MAD (*multi-wavelength anomalous dispersion*). V proteinové krystalografii se často měří MAD pro atom selenu, který se do molekuly zavede záměnou methioninu za selenomethionin při expresi rekombinantního proteinu. Selen, kterým v proteinu nahrazujeme síru, má silný anomální signál při vlnových délkách dostupných na synchrotronech.

## 12.14 Určení struktury z difrakčních dat

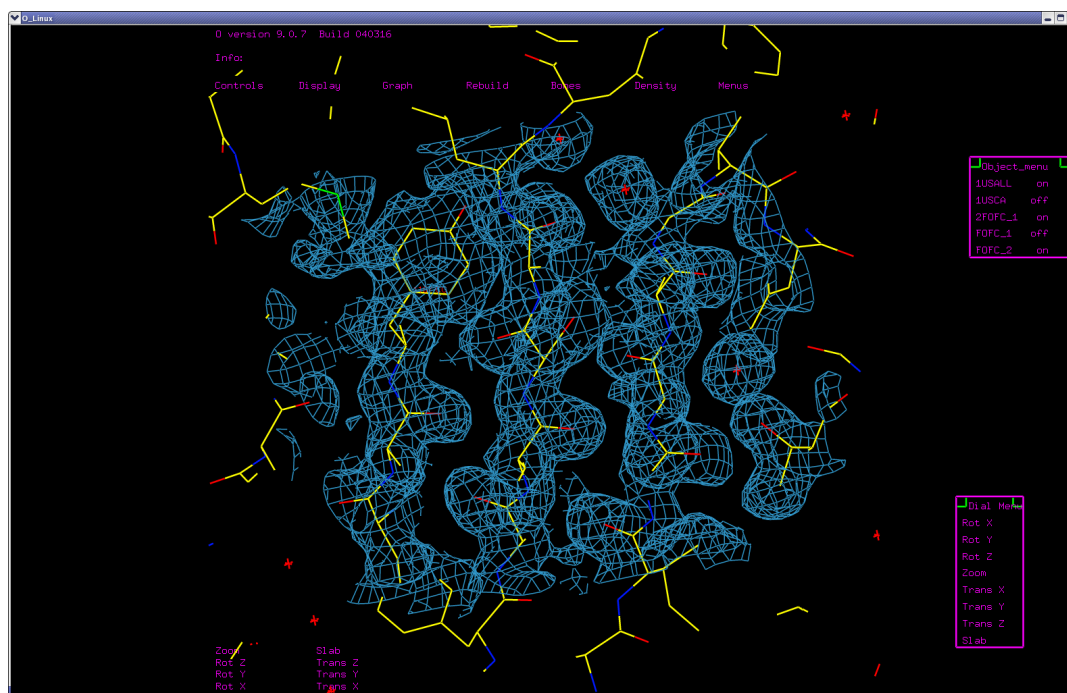
Vysvětlili jsme si, jak můžeme z intenzity reflexí určit amplitudu strukturního faktoru. Ukázali jsme si také, jak vypočítat fázi strukturního faktoru. Nic nám tedy nebrání spočítat podle vztahu 12.5 mapu elektronové hustoty. Zdálo by se tedy, že jsme u cíle našeho snažení. Ve skutečnosti nás ještě ale pořádný kus práce čeká. Je třeba postavit atomární model studované makromolekuly. Stavba modelu není vždy jednoduchá. Počáteční fáze, které použijeme k výpočtu mapy elektronové hustoty, často obsahují chyby, které způsobí, že tvar molekuly je překryt šumem, nebo se jeví jako rozmazaný. Když postavíme počáteční model molekuly, můžeme z něj vypočítat fázové úhly pro jednotlivé strukturní faktory, zkombinovat je s naměřenými amplitudami strukturních faktorů a vypočítat novou mapu. Pokud jsme model postavili správně, bude nová mapa obsahovat mnohem méně šumu a tvar molekuly v ní bude zřetelnější. Proto je nutné postupně vypočítanou strukturu zpřesňovat.

Zpřesňování struktury můžeme provádět nejen ručně, ale i s pomocí programů, které jsme si popsali v kapitole 10. Používají se přitom metody molekulové mechaniky (sekce 10.3), které jsou dostatečně rychlé i pro velké molekuly. Je zde ale důležitý rozdíl. Během výpočtu se kromě hledání konformace s nejmenší energií snažíme zároveň hledat konformaci nejlépe odpovídající experimentálně naměřeným amplitudám strukturních faktorů.

Jakým způsobem najdeme struktury, které dobře odpovídají naměřeným amplitudám? Nejjednodušší by bylo prostě provádět matematickou minimalizaci. Funkcí, kterou bychom minimalizovali, by byl rozdíl mezi experimentálně zjištěným strukturním faktorem a strukturním faktorem vypočítaným z modelu struktury. Proměnnými by byly souřadnice, které popisují polohu atomů v molekule. Aby takový postup fungoval dostatečně přesně, mělo by být naměřených intenzit aspoň desetkrát více, než proměnných. Souřadnic atomů v nepříliš velkých biomakromolekulách jsou tisíce. To znamená, že bychom měli vyhodnotit desetitisíce reflexí. Tak velký počet reflexí vyžaduje vysoké rozlišení difrakčního experimentu. Máme-li štěstí a náš krystal takové rozlišení poskytuje, můžeme získat strukturu nepříliš velkých biomakromolekul velmi rychle.

Pro větší molekuly je splnění přísných požadavků na dostatek dat obtížné. Musíme proto provádět opakovanou minimalizaci, kdy vyjdeme z počátečního modelu struktury, provedeme minimalizaci, vypočítáme nové strukturní faktory a z nich novou mapu elektronové hustoty. Na základě této vylepšené mapy upravíme náš model a celý postup opakujeme. Úpravy modelu se většinou dělají „ručně“ na obrazovce počítače, s pomocí interaktivních programů (příkladem mohou být programy Coot, XtalView nebo O). Taková práce může trvat dny až několik týdnů. Ruční úpravy struktury lze do jisté míry nahradit metodami molekulové dynamiky (sekce 10.10) a celý postup je tak možno zautomatizovat (použití molekulové dynamiky umožňují například programy Refmac, Phenix, X-PLOR a CNS). Obrázek 12.12 ukazuje část modelu molekuly proteinu zabudovaný do získané mapy elektronové hustoty.

Abychom vyhověli požadavku na dostatečný počet měřených reflexí vzhledem k počtu proměnných, je výhodné snížit počet proměnných. Můžeme předpokládat, že struktura aromatických kruhů v amino-



Obrázek 12.12: Mapa elektronové hustoty molekuly proteinu (modře) s vestavěným strukturálním modelem.

kyselinách nebo bázích nukleových kyselin není ovlivněna celkovou konformací makromolekuly. Je proto zbytečné počítat se souřadnicemi všech atomů aromatických skupin, jako by byly nezávislé. Počet stupňů volnosti je proto možné snížit na jediný torzní úhel, který popisuje orientaci aromatického kruhu jako celku. Takové dosazení pevně daných souřadnic některých atomů je příkladem použití *tvrdých vazebných podmínek (constraints)*. Můžeme použít i *měkké vazebné podmínky (restraints)* tím, že souřadnice atomů v aromatických kruzích pouze omezíme pomocí rovnic, které stabilizují jednotlivé atomy v polohách odpovídajících ideální konformaci molekuly, jak byla zjištěna při studiu struktury malých molekul, pro které jsou dostupná data s vysokým rozlišením).

Výsledkem našeho snažení je být model struktury, ve kterém známe souřadnice všech atomů. Tak tomu bývá ale málokdy. Některé části molekul mohou zaujímat v krystalu různé konformace. Mapa elektronové hustoty pak ukazuje pouze rozmazaný tvar, podle kterého není možné postavit model. Proto je běžné, že z difrakčních měření nemůžeme určit polohu několika pohyblivých aminokyselin na koncích polypeptidového řetězce nebo v pohyblivých smyčkách, či polohu některých atomů v postranních řetězcích na povrchu biomakromolekuly.

I poměrně dobře definované atomy v krystalu kmitají kolem svých středních poloh. Toto kmitání souvisí s teplotou. Kromě souřadnic atomů proto bývá ve strukturách atomů uveden také tento tepelný rozkmit. Zpravidla se uvádí jako takzvaný *B-faktor*, což je střední kvadratická odchylka od průměrné polohy<sup>5</sup> atomu.

<sup>5</sup>Tato odchylka může být různá v různých směrech, tato různost se ale u makromolekul obvykle zanedbává.

## 12.15 Kontrola správnosti

Během určování struktury jsme mnohokrát použili různých odhadů, oprav, ručních zásahů a přibližných výpočtů. Je tu tedy nebezpečí, že se nám do modelu vloudily nejrůznější chyby. Proto musíme pečlivě zkontrolovat, jestli získaná struktura nevypadá podezřele.

Nejpoužívanějším měřítkem přesnosti struktury je  $R$ -faktor, definovaný jako

$$R = \frac{\sum_i ||F_i^{\text{exp}}| - |F_i^{\text{mod}}||}{\sum_i |F_i^{\text{exp}}|}, \quad (12.6)$$

kde  $|F_i^{\text{exp}}|$  jsou amplitudy strukturních faktorů určených z experimentálních dat a  $|F_i^{\text{mod}}|$  jsou amplitudy strukturních faktorů vypočítaných z atomárního modelu. Čím je struktura přesnější, tím je  $R$  nižší. Během zpřesňování strukturního modelu by se tedy měl  $R$  snižovat. Nízkého  $R$  lze ovšem dosáhnout i uměle, když například doplníme molekuly vody tam, kde by měla být část proteinu. Abychom odhalili takováto zdánlivá „vylepšení“, rozdělíme si intenzity na dvě části. Z větší části (90 %) vypočítáme strukturu, z menší části (10 %) počítáme  $R$ . Od okamžiku, kdy se takto vypočtený  $R$  již nesnižuje (i když  $R$  určený ze všech intenzit stále klesá), víme, že se nám nedaří model zpřesňovat.

Kromě testování přesnosti je dobré se podívat, zda hodnoty vazebných délek, vazebných úhlů a torzních úhlů odpovídají hodnotám běžným pro organické molekuly. Jakékoli neobvyklé hodnoty jsou podezřelé – je větší pravděpodobnost, že jsme udělali chyby, než že jsme objevili něco dosud nevídaného.

## Kapitola 13

# Elektronová mikroskopie

### 13.1 Elektrony jako vlnění

Biomakromolekuly jsou mikroskopické objekty a jejich zkoumání nelze uspokojivě popsat zákony klasické fyziky platícími v makrosvětě. Již několikrát jsme narazili na to, že pochopení různých stránek strukturní biochemie vyžaduje kvantovou mechaniku. To platí zejména tehdy, když se zabýváme elektrony, jejichž kvantová povaha je velmi zřetelná.

V části 11.4 a v celé kapitole 12 jsme vzali kvantové chování elektronů do úvahy tak, že jsme představu elektronu nacházejícího se v určitém místě nahradili popisem rozložení *elektronové hustoty* v prostoru. Rozložení elektronové hustoty v molekule přitom představovalo objekt, který zkoumáme pomocí elektromagnetických vln, k jejichž popisu nám stačila klasická fyzika. V této kapitole si ukážeme, že elektromagnetickou vlnu můžeme nahradit elektrony. To lze pochopit jen tehdy, když pomocí kvantové mechaniky popíšeme nejen elektrony ve studované biomakromolekule, ale i elektrony, které na ni dopadají.

Kvantová mechanika nám říká, že světlo i elektrony se chovají zároveň jako částice i jako vlny. Proto můžeme to, co jsme se o vlnách naučili v kapitolách 11 a 12, použít i pro popis elektronů bombardujících zkoumané molekuly. Vlny popisující elektrony jsou ovšem podstatně tajemnější, než vlny světelné. Světelnou vlnu jsme si mohli představit jako vlnění elektrických a magnetických sil popsaných vektory  $\vec{E}$  a  $\vec{B}$ . Kvantová mechanika nám představu vlnící se klasické veličiny jako  $\vec{E}$  a  $\vec{B}$  neposkytuje. I když různé verze kvantové teorie popisují vlnění pomocí jasně definovaných matematických objektů, není snadné si tyto objekty spojit s něčím, co známe z každodenní zkušenosti. Protože korektní kvantová analýza elektronových vln je daleko za rámcem našich přednášek, budeme se často ve výkladu utíkat k analogiím se světelným vlněním.

Jakkoli se elektronové vlny světelným podobají, liší se v jedné podstatné vlastnosti. Elektrony mají na rozdíl od světla (od fotonů) elektrický náboj. Elektronovou vlnu lze tedy ohýbat v magnetickém poli, což u světla není možné. Pomocí cívek, kterými protéká elektrický proud, tak můžeme sestavit *elektromagnetické čočky*, které fungují podobně, jako čočky skleněné ve světelné mikroskopii. Elektromagnetické čočky mají přitom jednu ohromnou výhodu. Ohýbají i elektronové vlny velmi krátkých vlnových délek, zatímco pro světlo (přesněji pro rentgenové paprsky) krátkých vlnových délek vhodné čočky nemáme. Proto je možné z elektromagnetických čoček sestavit *elektronový mikroskop* schopný rozlišit jednotlivé atomy v molekulách.

Elektronová mikroskopie se v mnohém podobá mikroskopii využívající viditelné světlo. Proto při vysvětlování principů elektronové mikroskopie můžeme často využít zkušenosti s mikroskopy, se kterými jsme se začali seznamovat možná již na základní škole. Z tohoto pohledu je elektronová mikroskopie technikou přístupnější a pochopitelnější, než rentgenová krystalografie.

Z mnoha variant elektronové mikroskopie se zaměříme pouze na *transmisní elektronovou mikroskopii* (TEM). TEM je totiž pro popis struktury biomakromolekul nejužitečnější. TEM nám poskytuje, stejně jako rentgenové krystalografie, *mapu elektronové hustoty*, která popisuje tvar zkoumané molekuly.

## 13.2 Konstrukce elektronového mikroskopu

Konstrukce transmisního elektronového mikroskopu (obrázek 13.1) je velmi podobná konstrukci světelného mikroskopu. Zdrojem vlnění je takzvané *elektronové dělo* (anglicky *electron gun*, v nejjednodušším provedení ohnuté kovové vlákno zahříváné průchodem elektrického proudu, ze kterého vyletují elektrony a jsou urychlovány silným elektrickým polem).

Optickou soustavu tvoří *elektromagnetické čočky*, které kontrolovaným způsobem ovlivňují šíření elektronových vln. Na rozdíl od čoček světelného mikroskopu jde o cívky, kterými protéká elektrický proud. Magnetické pole indukované elektrickým proudem působí silou na letící nabitě částice, v našem případě elektrony, a mění jejich dráhu.

Pod zdrojem jsou umístěny dvě nebo tři *kondenzorové čočky*, které zajišťují optimální „osvit“ vzorku elektronovým vlněním. Objekt, který zkoumáme (vzorek našich biomakromolekul), umístíme na sítku (anglicky *speciment grid*) o průměru 3 mm. V elektronové mikroskopii pracujeme s velmi rozdílnými zvětšeními. Na začátku, když hledáme pozorovaný objekt, pozorujeme celou sítku, jejíž obraz zvětšujeme jen asi padesátkrát. Nakonec, když chceme rozlišit jednotlivé atomy, musíme použít zvětšení v řádu desítek tisíc ( $70\,000 \times$  až  $100\,000 \times$ ) a sledujeme jen velmi malou oblast sítky. Kondenzorové čočky nám umožňují soustředit elektronové vlny do oblasti, kterou při daném zvětšení právě sledujeme.

Pod sítkou se vzorkem je *objektivová čočka*, která zásadním způsobem určuje rozlišení a kvalitu obrazu. Následuje série *projektorových čoček*, které vytvářejí zvětšený obraz na detektoru.

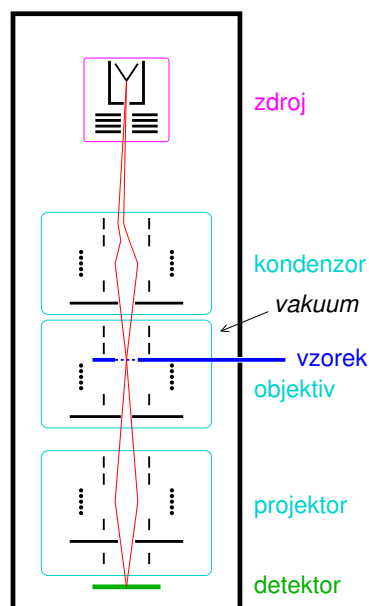
Detektorem může být film, nebo stínítko pokryté vrstvou scintilátoru, ve které dopad elektronů způsobí emisi fotonů viditelného světla, které pozorujeme. V současnosti je detektorem často kamera přímo detekující dopadající elektrony.

Celá optická soustava je umístěna v neprodyšném tubusu, ve kterém je udržováno velmi vysoké vakuum. Jinak by elektrony interagovaly s molekulami vzduchu, což by ovlivnilo elektronové vlny natolik, že by žádný obraz zkoumaného objektu nevytvořily. Nutnost práce ve vakuu nám neumožňuje pozorovat nefixované biologické vzorky. Při jejich umístění do vakua by se okamžitě odpařila voda a zničily všechny biologické struktury, včetně biomakromolekul. Proto musí být vzorky fixovány buď chemicky (zalitím do tenké vrstvy pryskyřice) nebo zamrazením v tenké vrstvičce amorfního ledu.

V elektronové mikroskopii je také nutné zaručit, aby se mikroskopem v danou chvíli pohyboval jen jeden elektron. Protože mají elektrony elektrický náboj (na rozdíl od světla neboli fotonů), vzájemně se odpuzují. Přítomnost druhého elektronu by zcela změnila vlnu tvořenou prvním elektronem a k vytvoření obrazu by nedošlo. Všechny následující úvahy o skládání (interferenci) vln se budou týkat vlnění *jediného elektronu*. To je velmi pádným důkazem platnosti zákonů kvantové mechaniky, které říkají, že obraz, který jeden elektron na detektoru vytvoří, závisí na všech cestách, kterými tento jediný elektron mohl (s různou pravděpodobností) k detektoru doputovat. Požadavek, aby se mikroskopu nacházel vždy jen jeden elektron, lze splnit poměrně snadno. Elektrony se totiž v elektronovém mikroskopu pohybují velmi rychle (desítky procent rychlosti světla), takže jeden elektron prolétne optickou soustavou mnohem dříve, než ze zdroje vystřelí další.

## 13.3 Výhody a nevýhody elektronové mikroskopie

Elektronová mikroskopie má ve srovnání s jinými metodami (mikroskopie ve viditelném světle, rentgenová krystalografie) některé důležité výhody:



Obrázek 13.1: Zjednodušené schéma transmisního elektronového mikroskopu. Elektrony jsou znázorněny červeně. Čočky mikroskopu tvoří elektromagnetická cívka (znázorněné tečkami) s clonou (vodorovná čára s mezerou uprostřed) a korekčními cívkami (svislé čárky). Ve skutečném mikroskopu tvoří kondenzor a projektor soustavy více čoček.

- Hlavní výhodou elektronové mikroskopie je, že používá elektrony s vysokou energií, odpovídající velmi krátké vlnové délce. Jak víme z kapitoly 12, krátká vlnová délka nám umožňuje rozlišit malé detaily, v případě biomakromolekul jednotlivé atomy.
- Výhodou i nevýhodou elektronové mikroskopie je, že elektrony letící z elektronového děla velmi silně interagují s elektrony v atomech vzorku. To nám umožňuje dosáhnout vysokého kontrastu, zejména pro vzorky fixované chemicky a *barvené* atomy s velkým počtem elektronů (*těžké atomy*). Většina atomů biologicky zajímavých molekul (H, C, N, O, P, S) obsahuje málo elektronů. Přídavek atomů s mnoha elektrony (Os, Au, Pb, U) vytváří silný kontrast, protože tyto atomy interagují s elektronovou vlnou mnohem silněji, než atomy biogenních prvků.
- Díky elektrickému náboji elektronů můžeme v elektronové mikroskopii použít elektromagnetické čočky, pro které platí standardní zákony optiky. Proto se elektronová mikroskopie podobá světelné mikroskopii mnohem více, než rentgenová krystalografie.
- Existují velmi intenzivní zdroje elektronů, takže můžeme použít velmi krátké expozice.
- TEM poskytuje informaci o vnitřní struktuře objektů, včetně biomakromolekul. Snímané obrázky jsou dvojrozměrné projekce trojrozměrných objektů a nesou informaci nejenom o povrchu objektů, ale i o rozložení elektronové hustoty uvnitř objektů.

Elektronová mikroskopie ale má i své nevýhody:

- Nevýhodou je nezbytnost použití vysokého vakua, což vyžaduje chemickou fixaci nebo zmrazení vzorku. Zmrazení musí být přitom provedeno tak, aby nedošlo k poškození vzorku a molekuly zůstaly ve stavu co nejbližším nativní konformaci.

- Vzorek musí být velmi tenký, abychom se vyhnuli 100% absorpci. Velmi silná interakce elektronových vln s elektrony vzorku, zmíněná výše jako výhoda, je z tohoto pohledu nevýhodou. Pokud není vzorek velmi tenký, elektronová vlna s ním interaguje natolik silně, že k detektoru vůbec nedorazí a my nepozorujeme žádný signál. Ideální tloušťka vzorků je do 200 nm, limit je asi 500 nm. Ve zkoumání struktur purifikovaných biomakromolekul nás limit tloušťky vzorků příliš neomezuje, protože velikost průměrných proteinů je několik nanometrů. Pro sledování buněk je ale omezení tloušťky vážným problémem. I malá bakterie *Escherichia coli* má délku 3  $\mu\text{m}$  a tloušťku 1  $\mu\text{m}$  a eukaryotické buňky (například fibroblasty pěstované na povrchu podložky) mají tloušťku větší, než 10  $\mu\text{m}$ .
- Dalším důsledkem silné interakce elektronů je poškození biologického vzorku. Délka měření proto musí být krátká a množství elektronů, kterým můžeme vzorek vystavit, dříve než dojde k jeho zničení, je malé. Proto při studiu biomakromolekul získáváme data s velmi špatným poměrem mezi signálem a šumem (S/N). Zatímco pravidlem analytické chemie je nepoužívat výsledky měření se  $S/N < 2$ , obrazy nebarvených biomakromolekul mívají často  $S/N < 0,1$ . Získaný obraz tedy obsahuje mnohem více šumu než signálu a pouhým okem je obtížné rozpoznat, jestli na obrázku naše molekula je nebo není.

## 13.4 Optika a rozlišení mikroskopů

Jak jsme zmínili v části 13.1, elektronové vlny se v klasické (nekvantové) fyzice nevyskytují a je proto obtížné je popsat jednoduchým a intuitivním způsobem. Proto využijeme obecné podobnosti světelných a elektronových vln a vysvětlíme si principy a omezení optiky mikroskopu na příkladu viditelného světla procházejícího skleněnou čočkou. Navážeme přitom na část 11.5, kde jsme si ukázali, že rozkmitání elektronů skla procházející světelnou vlnou vede k opoždění této vlny (zkrácení vlnové délky při zachování frekvence), případně ke změně směru jejího šíření, zvané *lom světla*.

Princip jednoduché čočky je ukázán na obrázku 13.2. Podívejme se nejdříve na to, jak se chová světlo letící z jediného bodu pozorovaného objektu. Za takový bod můžeme považovat třeba velmi malou bublinku vzduchu ve skle přesně ve středu hlavičky zeleného špendlíku na obrázku 13.2. Kulová vlna šířící se ze středu hlavičky narazí na různá místa čočky s různou fází. Z míst na povrchu čočky, kam vlna dopadla, se šíří zpomalená (zhuštěná) kulová vlna všemi směry, dokud nedospěje k protějšímu povrchu čočky. Z míst dopadů na tento povrch se pak šíří původní rychlostí všemi směry kulové vlny, z nichž čtyři je zakresleny na obrázku 13.2A. Zaostrěný obraz středu hlavičky špendlíku vznikne tam, kde budou mít tyto vlny stejnou fázi (detail na obrázku 13.2B). Stejnou fázi budou mít tehdy, když na své cestě ze středu hlavičky do místa obrazu budou mít stejný počet maxim. To platí pro bod uprostřed čárkované kružnice napravo od čočky. Všechny čtyři zakreslené vlny se tam mají čtyřicáté první maximum od středu hlavičky (ve skutečném případě viditelného světla by ovšem šlo o statisíce vlnových délek). Například vlna, která prošla spodní částí čočky, měla delší cestu, ale pomalý úsek vedoucí čočkou byl kratší, než v případě vlny, která narazila na čočku ve vodorovném směru. Cestám, po kterých dorazí světlo do stejného místa se stejným počtem maxim, říkáme stejně dlouhé *optické dráhy*.

Skutečný světelný mikroskop se skládá z většího počtu čoček. Podobně jako v elektronovém mikroskopu popsaném v části , kondenzorové čočky směřují světlo na vzorek a projektorové čočky obraz, nakreslený na obrázku 13.2A čárkovaně, dále zvětšují a promítají na detektor (je-li detektorem oko, říkáme těmto čočkám okulár).

Analýza vln ukázaná na obrázku 13.2A je poněkud nepřehledná. Proto se obvykle zakreslují jen směry šíření vln, jak je ukázáno na obrázku 13.2C. Těmto směrům říkáme *paprsky*. Na obrázku 13.2C jsou znázorněny paprsky šířící se jak z hlavičky, tak ze špičky zeleného špendlíku. Jasně ukazují, že modrý obraz špendlíku je zvětšený. Kreslení paprsků místo vln sice obrázek zjednoduší, ale může vést



k falešným představám. Mohlo by se zdát, že soustavou čoček můžeme dosáhnout libovolného zvětšení jakkoli malého objektu. Ve skutečnosti ale brzy narazíme na limit daný difrakcí, kterou paprsky ignorují.

Limit rozlišení je definován pro bodový zdroj vlnění, například maličkou dírkou v neprůsvitné fólii. Paprsky předpovídají, že záření procházející sebemenším kruhovým otvorem vytvoří při dopadu na plochu za otvorem jasnou skvrnu ve tvaru otvoru a ostře ohraničený stín všude jinde (obrázek 13.2C vlevo). Poctivá analýza šíření vln nás ale vede k jinému závěru. Za otvorem se šíří kulové vlny všemi směry, i mimo směr šíření. Proto na ploše nevznikne ostrý stín, ale skvrna s rozmazanými okraji, obklopená dalšími světelnými kroužky (obrázek 13.2C vpravo), zvané Airyho disky<sup>1</sup>. Světlo se ohýbá, míří za roh. Proč jsme si toho nevšimli u světla postupujícího roztokem molekul v části 11.5? Protože jsme zkoumali skládání vln tvořených v celé rovině kolmé na směr dopadající vlny. Z takové roviny sledujeme u běžných chemických vzorků (kyveta ve spektrofotometru) obrovskou plochu, milimetry nebo centimetry čtverečné ve srovnání s vlnovou délkou několik desetin mikrometru. V takových obrovských rozměrech ve směrech odchylovajících se od původního paprsku téměř každou vlnu vyruší fázově posunutá vlna vzniklá o kus dál. Hraniční oblast, kde už žádné „o kus dál“ není (kde například končí vzorek) je u typických vzorků zanedbatelně malá. V případě maličké dírkky ale začíná být hraniční oblast srovnatelná s osvětlenou plochou těsně za otvorem. Leckterá vlna šířící se z roviny dírkky se nepotká s fázově posunutou vlnou, která by ji vyrušila.

Vznik rozmazaného obrazu omezuje naši schopnost rozlišit dva malé objekty blízko sebe (obrázek 13.2C vpravo). Když jejich obrazy zvětšíme, zvětšíme je rozmazané a naše schopnost je rozlišit se tím nezlepší. Rozlišení ale můžeme zlepšit tím, že zvětšíme světelný kužel, který na čočku dopadá. Můžeme si to ukázat opět na obrazu špendlíku zvětšeného jednoduchou čočkou. Informace o poloze zdroje (na obrázku 13.2A střed hlavičky zeleného špendlíku) je zakódovaná v celé kulové vlně, která se ze zdroje šíří. Jak jsme si vysvětlili, zaostřený obraz vzniká tak, že se v určitém místě se stejnou fází skládají, a tedy výrazně zesílí, vlny pocházející z různých oblastí původní kulové vlny. Pokud na čočku dopadá velká část původní kulové vlny, skládá se velké množství vln z velmi různých směrů šíření původní vlny. Složená vlna má velmi intenzivní a úzké hlavní maximum, které velmi přesně lokalizuje polohu zdroje. Takové maximum snadno odlišíme od podobně intenzivního maxima složené vlny pocházející z jiného zdroje (například jiné bublinky v hlavičce špendlíku). Pokud ale na čočku dopadá jen úzký kužel kulové vlny šířící se ze zdroje, v místě zaostřeného obrazu se setkává se stejnou fází jen málo vln pocházejících z velmi podobných směrů původní kulové vlny. Složením těchto vln vzniká vlna s nevýrazným hlavním maximumem a množstvím dalších maxim. Taková maxima ze dvou blízkých zdrojů bude obtížné rozlišit, budou splývat.

Zjednodušeně je vliv šíře světelného kuželu na rozlišení znázorněn na obrázku 13.2E–G. Signál, který detektor zachytí, je zde znázorněn v zeleném rámečku. Pro pochopení rozložení intenzity záření v rovině detektoru je dobré si připomenout, že se vlny sčítají (konstruktivně interferují) nejen přesně v průsečíku modrých kružnic, znázorňujících jejich maxima na obrázku 13.2A,B, ale i v blízkosti tohoto průsečíku. Tam, kde se například setká maximum jedné vlny (oblast s  $\vec{E}$  mířícím kladným směrem) s nulou jiné vlny (oblast s  $\vec{E} = 0$ ), výsledkem jejich sečtení je stále kladná vlna (s  $\vec{E}$  mířícím kladným směrem), byť ne zesílená. Proto obraz zachycený detektorem není v našem případě obrazem středu hlavičky bod uprostřed čárkované kružnice, ale trochu rozmazaná oblast v okolí. Signál na detektoru má stejný profil jako světlo z bodového zdroje na obrázku 13.2C vpravo, s minimem přibližně ve dvou třetinách poloměru obrazu hlavičky a sekundárním maximumem (odpovídajícím prvnímu Airyho kroužku) poblíž povrchu hlavičky. Pokud na našem zjednodušeném obrázku zúžíme clonou kužel světla dopadající na čočku, k tvorbě obrazu přispějí jen dvě prostřední vlny (obrázek 13.2G). V důsledku toho bude oblast, ve které se vlny sečtou, výrazně širší a minimum na detektoru bude již mimo povrch hlavičky (obrázek 13.2H).

Hodnotu rozlišení můžeme definovat i přesně matematicky. Jako *rozlišení*  $r$  se udává taková vzdálenost dvou zdrojů, pro kterou hlavní maximum prvního zdroje leží v prvním minimu druhého zdroje a naopak.

<sup>1</sup>George B. Airy byl britský astronom, který původ kroužků vysvětlil.

Pokud tedy budeme pozorovat v hlavičce špendlíku na obrázku 13.2 dvě bublinky, jednu ve středu a druhou u povrchu hlavičky, bez clony je rozlíšíme (obrázek 13.2F), ale se clonou již ne (obrázek 13.2H). Závislost rozlišení na velikosti světelného kužele se popisuje pomocí úhlu  $\alpha$  mezi osou a pláštěm kuželové plochy. Protože pozorovaný kužel porovnáváme s délkou vlny v daném prostředí, je rozlišení úměrné  $\lambda/n$ , kde  $\lambda$  je vlnová délka ve vakuu a  $n$  je index lomu. Po započítání závislosti intenzity složené vlny na vzdálenosti od hlavního maxima lze rozlišení vypočítat

$$r \approx 0,61 \frac{\lambda}{n \sin \alpha}. \quad (13.1)$$

Tento vzoreček vysvětluje, proč lze ve světelné mikroskopii zvýšit rozlišení (tedy snížit hodnotu  $r$ ) tak, že co nejširší čočku objektivu přiblížíme těsně ke vzorku (zvýšíme  $\alpha$ ) a mezi vzorek a objektiv kápneme imerzní olej s vysokým indexem lomu. Rozlišení také můžeme zvýšit použitím modrého světla, které má kratší vlnovou délku. Jak ale víme z kapitoly 12, k rozlišení atomů v molekulách bychom potřebovali světlo s vlnovou délkou blízkou velikostem atomů (řádově 0,1 nm), což je daleko za hranicí viditelné oblasti. V případě elektromagnetického záření jde o rentgenové záření, v případě elektronového záření o vlny tvořené elektrony.

Je také dobré si uvědomit, že číslo  $r$  popisuje, *jak blízké zdroje lze rozlišit*, ne *jak přesně lze lokalizovat polohu jednoho zdroje*. Polohu (těžiště) je možné určit s přesností mnohem lepší než  $r$  i z rozmazaného signálu s Airyho disky, pokud je tento signál dost daleko od signálů jiných zdrojů. Ve světelné mikroskopii se intenzivně rozvíjejí metody *superrozlišovací mikroskopie* (*superresolution microscopy*) umožňující přesnější měření polohy zdroje, než by odpovídalo číslu  $r$ . Například je možné označit molekuly v buňce fluorescenční látkou a aktivovat jen malý počet takových fluoreskujících značek. Jednotlivé fluorescenční zdroje tak budou od sebe dostatečně daleko, abychom analyzovali zvláště jednotlivé rozmazané signály a spočítali jejich těžiště s přesností lepší, než odpovídá číslu  $r$ .

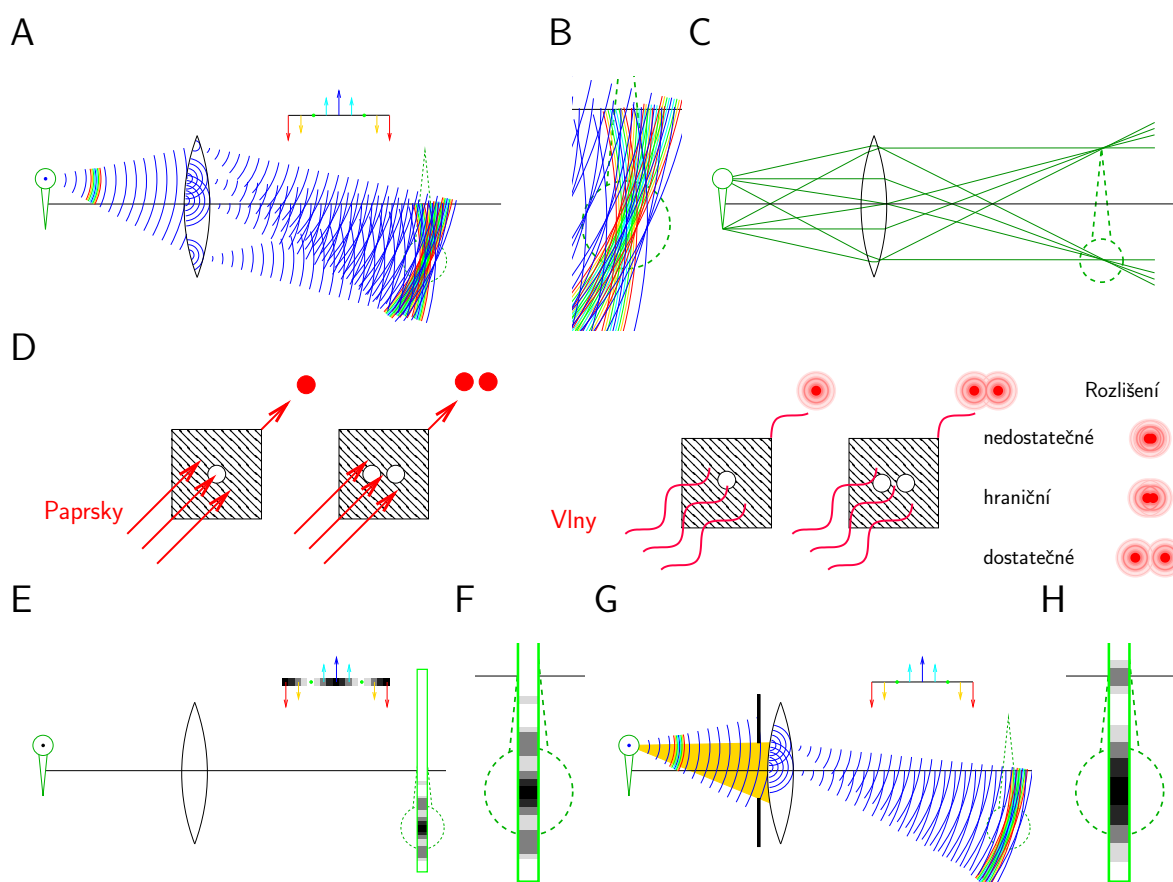
V elektronové mikroskopii jsme ale odkázáni na rozložení elektronové hustoty, které nás zajímá, a nemůžeme zpřesňovat určování poloh atomů nějakými značkami. Stejně tak nemůžeme zvyšovat index lomu prostředí, protože musíme pracovat ve vakuu. Tvar magnetického pole cívek, které používáme místo čoček, nás navíc nutí používat elektronové vlny jen ve velmi úzkém kuželu<sup>2</sup>. Jediný parametr, kterým můžeme zlepšit rozlišení, je vlnová délka.

Naštěstí je možné připravit elektrony s kinetickou energií odpovídající vlnovým délkám jednotek pikometrů (tisícin nanometrů). Takovéto vlny jsou schopny rozlišit jednotlivé atomy, i když je sledujeme jen ve velmi úzkém kuželu. Navíc elektrony s vysokou energií lépe procházejí materiálem a při jejich použití jsme méně omezeni tloušťkou vzorku.

## 13.5 Příprava vzorků pro elektronovou mikroskopii

Klasická elektronová mikroskopie je založena na metodách „barvení“ vzorků atomy, které obsahují výrazně více elektronů, než převažující biogenní prvky. K *pozitivnímu barvení* se používají acetát uranu, citrát olovnatý, oxid osmičelý a jiné soli těžkých kovů. Buněčné struktury, na které se těžké atomy vážou (například hydrofobní  $\text{OsO}_4$  se hromadí v membránách, jiné kovy interagují s proteiny) výrazně více pohlcují elektrony a pozorujeme je jako tmavá místa. Jinou možností je *negativní barvení*, kdy se například virové částice zalijí do tenké vrstvy barviva. Tam, kde se nacházejí virové částice, potom pozorujeme jasnější místa chybějícího barviva. Ke studiu velkých proteinů a zejména jejich komplexů byla používána metoda zvaná *stínování* (anglicky *shadowing*), při které byl na vzorek studovaných částic pod ostrým úhlem (například  $4^\circ$ ) naprášen těžký kov a na tvar molekul se usuzovalo ze stínů protažených ve směru naprašování. Zmíněné klasické metody poskytly zásadní informace o detailech

<sup>2</sup>Jde o obdobu fotografování s méně kvalitním objektivem. I s takovou optikou můžeme získat pěkné fotografie za dobrého osvětlení, které nám umožní použít vysoké číslo clonky. V tomto případě clona propouští do objektivu světlo jen v úzké oblasti blízko středu čočky, kde má nedokonalost optiky nejmenší vliv.



Obrázek 13.2: Princip zvětšování obrazů pomocí čoček. A, vysvětlení na základě šíření vlny z bodového zdroje ve středu hlavičky zeleného špendlíku. Barva vln udává hodnotu  $\vec{E}$  kmitající kolmo k rovině nákresu, barevná škála je ukázána nad obrázkem. B, zvětšené okolí obrazu. C, popis pomocí paprsků. D, tvorba obrazu s ostrým stínem vyplývající z popisu záření pomocí paprsků (vlevo), tvorba obrazu s rozmazaným stínem v důsledku difrakce (ohybu) záření na hraně otvoru (vpravo). E, rozložení druhé mocniny amplitudy (intenzity) složené vlny na stínítku detektoru umístěného kolmo k rovině nákresu. Řez detektorem je ukázán v zeleném rámečku, škála intenzit je uvedena nad obrázkem. F, detailní pohled na rozložení intenzity snímané detektorem v okolí obrazu. G, vliv zúžení kuželu světla clonou na zobrazení předmětu. H, rozložení intenzity snímané detektorem pro zobrazení předmětu ukázané na obrázku G.

buněčných struktur a tvarech velkých komplexů biomakromolekul, nedokázaly však rozlišit jednotlivé atomy. Molekuly látek používaných k barvení totiž vytvářejí shluky (clustery) a rozlišení je omezeno velikostí těchto clusterů. Nejlepší rozlišení makromolekul barvených těžkými atomy je 2 nm, tedy o řád horší, než vzdálenosti atomů. Navíc jsou soli těžkých kovů toxické a poškozují strukturu studovaných biomakromolekul.

Alternativou k barvení je *kryoelektronová mikroskopie*. Místo barvení vzorek zmrazíme tak rychle, že se voda nestačí uspořádat do krystalové mřížky ledu, ale vytvoří se amorfní led, který se velmi podobá struktuře tekuté vody. Krystalizaci ledu se vyhýbáme hlavně ze dvou důvodů. Při tvorbě krystalků ledu spojené s velkými objemovými změnami dochází k poškození makromolekul a difrakce elektronů na krystalové mřížce ledu značně komplikuje zobrazování molekul.

Příprava vzorku pro kryoelektronovou mikroskopii je poměrně složitá. Povrch dodávaných sítěk je hydrofobní, proto je třeba sítku před použitím opracovat v plasmovém výboji. Poté roztok biomakromolekuly nanese na sítku. Nadbytek roztoku odstraníme přitisknutím filtračního papíru (*blotting*), takže na sítku zůstanou jen tenké skvrny roztoku (s tloušťkou kolem 100 nm). K vlastnímu mražení dochází rychlým ponořením sítky do kapalného ethanu, podchlazeného v lázni kapalného dusíku. Ponořením přímo do kapalného dusíku by amorfní led nevznikl, protože teplota kapalného dusíku odpovídá jeho bodu varu. Na povrchu předmětu ponořeného do kapalného dusíku dochází k rychlému vypařování a vrstvička dusíkové páry izoluje předmět od chladné kapaliny a zpomaluje jeho ochlazování. Naopak ethan je v dusíkové lázni zchlazen hluboko pod bod varu a velmi rychle odebírá teplo z ponořené sítky. K vlastnímu manipulaci se sítkou se používá zařízení zvané vitrobot, přesto je celý postup náročný na zručnost.

Ochlazené sítky se vzorky jsou uchovány v kapalném dusíku až do chvíle měření. Zvýšení teploty by vedlo k nežádoucímu překrystalování amorfního ledu. Také tubus mikroskopu a zařízení, kterým se vzorek do evakuovaného mikroskopu vkládá, musí být při kryoelektronové mikroskopii stále udržovány při nízké teplotě.

Velkou výhodou elektronové mikroskopie je, že během rychlého zmrazení se nemění ani struktura studované molekuly, ani struktura okolní vody. Biomakromolekuly tak můžeme studovat v nativním stavu. Protože molekuly nebarvíme a používáme TEM, získáváme informaci i o vnitřní struktuře biomakromolekul: snímané obrazy jsou projekce celého objemu molekuly.

## 13.6 Typy vzorků v elektronovou mikroskopii

V kryoelektronové mikroskopii se používají různé typy vzorků. V počátcích rozvoje metody hrály důležitou roli vzorky makromolekul, které tvoří dvojrozměrné krystaly. Typickým příkladem jsou transmembránové proteiny, které tvoří vhodné krystaly pro rentgenovou difrakci, ale některé z nich je možné připravit v membránách v tak vysoké hustotě, že vytvoří dvojrozměrný krystal. Metoda je vhodná pro menší proteiny (do 200 kDa) a poskytuje rozlišení srovnatelné s rentgenovou krystalografií (0,25 nm).

Pro určování konformace biomakromolekul je nejdůležitějším přístupem *rekonstrukce jedné částice* (*single particle reconstruction*). Tato metoda využívá toho, že v mnoha případech dokážeme připravit vzorky rekombinantních proteinů, které obsahují mnoho miliard molekul ve stejné konformaci. V elektronovém mikroskopu tak můžeme nasnímat obrázky desetitisíců až statisíců molekul, se kterými můžeme pracovat, jako by to byly obrazy jedné molekuly v různých orientacích. Informaci z jednotlivých snímků můžeme zkombinovat a spočítat z ní trojrozměrnou distribuci elektronové hustoty v molekule studovaného proteinu. Metoda poskytuje nejlepší výsledky pro velké proteiny a je možné dosáhnout atomového rozlišení (i pod 0,1 nm).

Pomocí kryoelektronové mikroskopie můžeme zkoumat i *pleiomorfní* objekty, z nichž žádné dva nemají stejnou strukturu. Jde například o buněčné organely, které se vždy na molekulární úrovni poněkud liší. Obrazy takových molekul nemůžeme průměrovat, ale musíme použít přístupu *tomografie*. Obrazy

objektu v různých orientacích získáváme tak, že v mikroskopu vzorek různě otáčíme a nakláníme. Z obrazů v různých orientacích můžeme zpět spočítat trojrozměrnou strukturu studovaného objektu. Výsledná struktura bude mít nezbytně omezené rozlišení. Jednak proto, že nemůžeme získat obrazy objektu ve všech orientacích (protože objekt je fyzicky vždy umístěn v nějakém držáku), ale také proto, že každý elektron objekt poškodí, takže máme jen omezený počet obrazů, které můžeme pro daný objekt získat. Poměr S/N je proto ještě nižší, než u snímků biomakromolekul. Pokud ale studovaný objekt obsahuje více kopií téže molekuly, můžeme průměrovat obrazy takové molekuly z různých částí vzorku a lokálně zlepšit rozlišení. V nejlepších případech je možné dosáhnout rozlišení okolo 0,4 nm.

## 13.7 Tvorba a zpracování digitálního obrazu

V elektronové mikroskopii získáváme digitální obraz. Tato skutečnost může znít triviálně, ale pro práci s mikroskopem a získanými snímky je velmi důležitá. Snímky musí být digitální proto, abychom s nimi mohli pracovat v počítači a vypočítat z nich trojrozměrnou strukturu molekuly.

Digitální obrazy se vždy skládají z velkého množství políček zvaných *pixels*. Digitální černobílá fotografie s náhled plynulými přechody mezi odstíny šedi je ve skutečnosti mozaika pixelů s různou intenzitou signálu. Maximální a minimální intenzita odpovídá černé a bílé barvě (nebo naopak).

Digitální povaha obrazu vysvětluje, proč je výhodné používat kamery přímo detekující dopad elektronů, jak jsme zmínili v části 13.4. Tradičně se v elektronové mikroskopii používaly fotografické filmy, které jsou citlivé na dopad světla i elektronů. Nevýhodou filmů je, že obraz z nich musíme do digitální formy převádět skenováním. Novějším typem detektorů byly kamery využívající technologie CCD (*coupled charge device*, integrované obvody propojených kondenzátorů citlivých na světlo). Nevýhoda CCD zařízení je, že nejsou přímo citlivá na elektrony. Proto bylo nutné použít vrstvičky scintilátoru (například krystalického materiálu zvaného yttrito-hlinitý granát, YAG), který po dopadu elektronu emituje fotony, které CCD detekoval. Scintilátor ale emitoval fotony i v jiných směrech, než směr dopadu elektronu, a tyto fotony aktivovaly sousední pixely CCD kamery. Proto docházelo k lokálnímu rozmazávání signálu. Velký pokrok pro elektronovou mikroskopii znamenaly polovodičové kamery schopné detekovat elektrony přímo. Elektronová vlna tak aktivuje přímo pixel, na který dopadla. Díky tomu může být přesněji lokalizována a struktury biomakromolekul můžeme počítat s lepším rozlišením.

V počítači ukládáme digitální obraz jako seznam čísel, která popisují intenzity signálů jednotlivých pixelů. Pokud by se náš detektor skládal jen z 25 pixelů, mohl by zápis snímku na obrázku 13.3A vypadat takto

$$\text{Obraz}_1 = (0; 0; 0; 0; 0; 0; 10; 10; 0; 0; 10; 5; 5; 10; 0; 10; 5; 5; 10; 0; 0; 10; 10; 0; 0). \quad (13.2)$$

Správně můžeme číst tento záznam jenom tehdy, když víme, kterému pixelu které číslo patří.

Zápis 13.2 můžeme také vyjádřit jako *lineární kombinaci* 25 základních stavů detektoru

$$\begin{aligned} & 0 \cdot P_{11} + 0 \cdot P_{12} + 0 \cdot P_{13} + 0 \cdot P_{14} + 0 \cdot P_{15} + \\ & 0 \cdot P_{21} + 10 \cdot P_{22} + 10 \cdot P_{23} + 0 \cdot P_{24} + 0 \cdot P_{25} + \\ & 10 \cdot P_{31} + 5 \cdot P_{32} + 5 \cdot P_{33} + 10 \cdot P_{34} + 0 \cdot P_{35} + \\ & 10 \cdot P_{41} + 5 \cdot P_{42} + 5 \cdot P_{43} + 10 \cdot P_{44} + 0 \cdot P_{45} + \\ & 0 \cdot P_{51} + 10 \cdot P_{52} + 10 \cdot P_{53} + 0 \cdot P_{54} + 0 \cdot P_{55} \end{aligned} \quad (13.3)$$

kde například  $P_{12}$  označuje stav s jednotkovou intenzitou pixelu v první řadě a druhém sloupci a nulovou intenzitou všech ostatních pixelů (obrázek 13.3B).

Snímek na obrázku 13.3A můžeme ale také zapsat jako lineární kombinaci 25 hypotetických stavů detektoru uvedených na obrázku 13.3C. Tyto stavy neodpovídají aktivaci jednotlivých pixelů, ale periodickým vzorům, vlnícím se v různých směrech s různou vlnovou délkou. Jednotlivé stavy  $\Psi_{hk}$  nedefinuje poloha na detektoru, ale parametry  $h$  a  $k$  popisující kolikrát se periodický vzor zavlní ve vodorovném a svislém směru.

Periodicky se opakujícími strukturami jsme se zabývali detailně v kapitole o rentgenové krystalografii (kapitola 12). Víme, že na vzdálenosti atomů v pravidelných krystalových mřížkách závisí směr, ve kterém se dramaticky zesílí rozptýlené rentgenové záření a tedy i místo na detektoru, kde pozorujeme difrakční skvrnu. Difrakci na krystalech tedy můžeme matematicky interpretovat následujícím způsobem. Elektronovou hustotu v krystalu si můžeme rozložit na součet periodických funkcí s různou vlnovou délkou. Každá z těchto funkcí odpovídá pravidelnému vzoru, který zesiluje rozptýlené rentgenové záření v jiném směru. Intenzita difrakčních skvrn v jednotlivých směrech nám přímo říká, jak moc která periodická funkce přispívá k celkovému rozložení elektronové hustoty. Při rozptylu záření fyzicky dochází k tomu, čemu se v matematice říká Fourierova analýza, rozklad nějaké matematické funkce na součet funkcí sinus a kosinus. Proto také je matematickým základem výpočtu rozložení elektronové hustoty *Fourierova transformace* difrakčního obrazce.

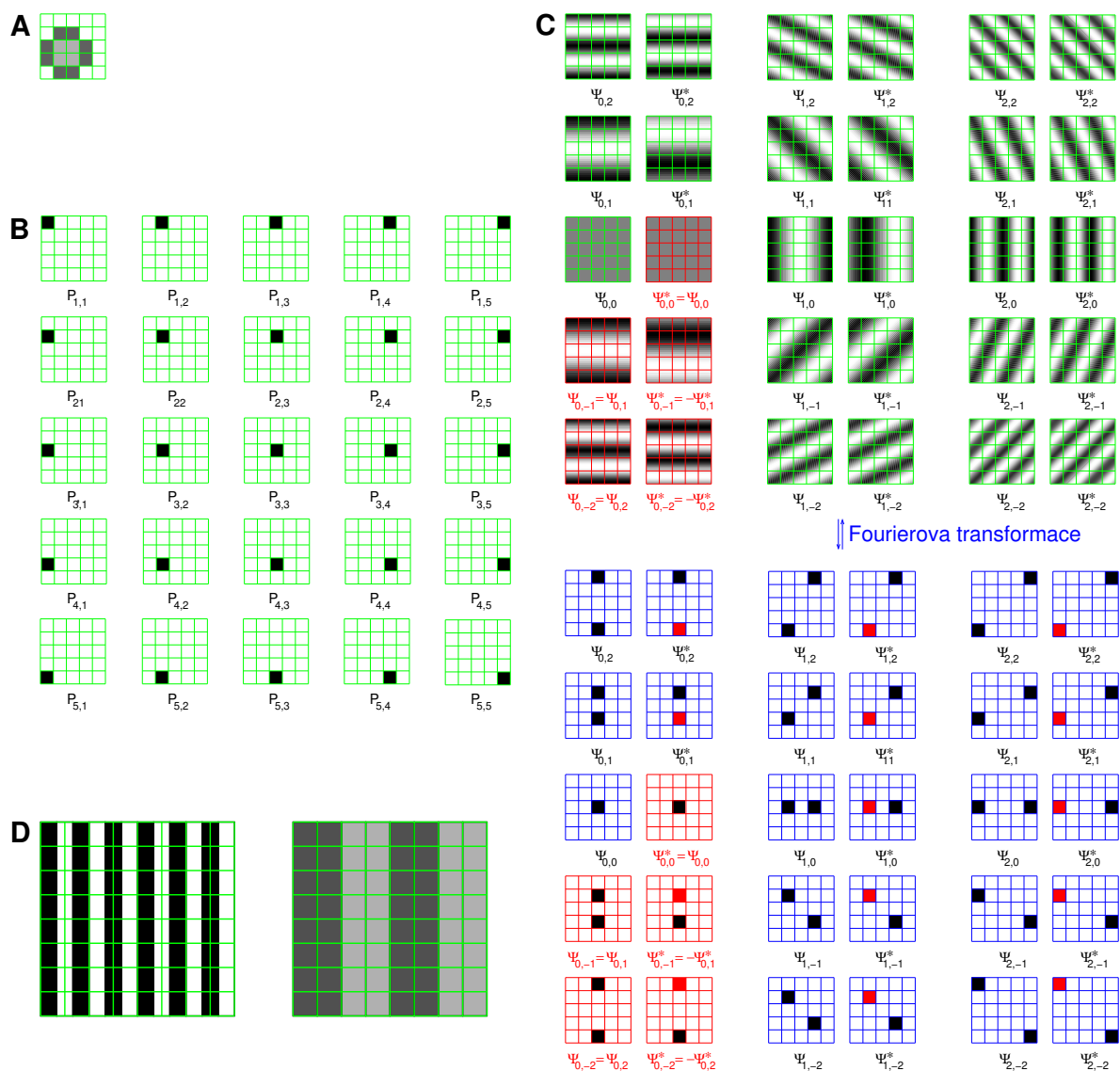
Proč ale rozkládat na periodické funkce obrazy molekul, které nejsou umístěny v pravidelných krystalových mřížkách? Každý detektor se skládá z konečného počtu  $N$  pixelů. Informace, kterou digitální obraz nese, je tedy vždy kombinací  $N$  stavů detektoru, ať už jde o jednopixelové stavy  $P_{xy}$  z obrázku 13.3B, nebo periodické stavy  $\Psi_{hk}$  z obrázku 13.3C. I když je obraz na detektoru nepravidelný, lze jej beze zbytku převést na kombinaci pravidelných vzorů  $\Psi_{hk}$ . Počítač přitom dělá to, k čemu fyzicky dochází v difrakčním experimentu, tedy Fourierovu analýzu. Parametry  $h$  a  $k$  přitom hrají stejnou roli, jako Millerovy indexy v rentgenové krystalografii. Stejně jako Millerovy indexy definují polohu skvrn v difrakčním obrazci rentgenového záření, můžeme periodické vzory  $\Psi_{hk}$  graficky znázornit jako body o souřadnicích  $h, k$ . Matematický postup, kterým ze závislosti intenzity na souřadnicích  $x, y$  spočítáme závislost na souřadnicích  $h, k$ , není opět nic jiného, než Fourierova transformace. Intenzita v místě souřadnic  $h, k$  v grafu získaném Fourierovou transformací nám říká, jak moc přispívá k původnímu obrazu periodický tvar, který má ve vodorovném směru  $h$  maxim a minim a ve svislém směru  $k$  maxim a minim. Při zpracování obrazů je velkou výhodou, že můžeme ze závislosti na  $x, y$  spočítat závislost na  $h, k$  a naopak.

Digitální povaha obrazu zaručuje, že veškerou informaci zachycenou v obraze můžeme vyjádřit pomocí konečného počtu periodických vzorů. Tento konečný počet ale zároveň omezuje rozlišení, kterého lze dosáhnout. Nejrychlejší změna periodického vzoru, kterou může detektor zachytit, je taková, kdy se maxima a minima střídají v sousedních pixelech (Nyquistův zákon). Maximální hodnota  $h$  a  $k$  je proto rovna polovině počtu pixelů ve vodorovném a svislém směru. Pokud obraz obsahuje vzor s vyšším počtem maxim, uvidíme v obraze zvlnění, které ve skutečnosti neexistuje. Zjednodušeně je to znázorněno na obrázku 13.3D, kde je znázorněn detektor s osmi pixely ve každém směru. Maximální (Nyquistova) frekvence, kterou takový detektor může rozlišit, je  $8/2 = 4$ . Na tento detektor promítáme vzor, ve kterém se ve skutečnosti vystřídá šest tmavých a šest světlých proužků na šíři detektoru. Když si spočítáme, jakou průměrnou intenzitu detekují jednotlivé pixely, zjistíme, že místo vzoru se správným počtem maxim (šest ve vodorovném směru) vidí detektor fiktivní obrazec procházející dvěma maximy. V běžném životě můžeme vidět tento efekt v televizi, pokud v ní vystupuje osoba s jemnějším vzorem oděvu, než odpovídá rozlišení obrazovky.

Popsané omezení *digitálního rozlišení* nás vede k tomu, abychom pečlivě zvážili, jaké zvětšení mikroskopu můžeme pro detektor s daným počtem pixelů použít. Pokud chceme například popsat strukturu s rozlišením 0,2 nm, musí jeden pixel dopadat vlna z oblasti maximálně 0,1 nm.

## 13.8 Amplitudový a fázový kontrast

Poměr signálu a šumu (S/N) v obrazech získaných elektronovým, ale i světelným, mikroskopem závisí na *kontrastu* mezi vlnou dopadající na detektor přímo a vlnou ovlivněnou vzorkem. Vzorek ovlivňuje vlnu dvěma způsoby. Prvním z nich je *amplitudový kontrast* způsobený částečným pohlcením vlny vzorkem. Velký amplitudový kontrast mají vzorky „barvené“ těžkými atomy. Druhým typem vlivu vzorku je *fázový kontrast*. Při průchodu vlny vzorkem dochází k jejímu opoždění, které popisujeme



Obrázek 13.3: Vlastnosti digitálního obrazu. Na obrázku A je příklad obrazu na detektoru skládajícím se z  $5 \times 5$  pixelů. Jakýkoli obraz na tomto detektoru je lineární kombinací 25 možných stavů detektoru znázorněných na obrázku B. Zároveň ale jakýkoli obraz na tomto detektoru je také lineární kombinací periodických vzorů na obrázku C. Podle Nyquistova zákona může mít periodický vzor nejvýš dvě minima a maxima v každém směru. Abychom byli schopni vyjádřit pomocí periodických vzorů každý digitální obraz, musíme vzít v potaz vlnění směřující vzhůru i dolů (směr rozlišujeme znaménkem indexu  $k$ ) a vlnění s různou fází (například kosinus a sinus, který od kosinu rozlišujeme hvězdičkou v označení vzoru). Kombinací 3 vodorovných frekvencí ( $h = 0; 1; 2$ ), 5 svislých frekvencí ( $k = -2; -1; 0; 1; 2$ ) a 2 fází bychom získali 30 vzorů, 5 se jich ale opakuje dvakrát (na obrázku C jsou vyznačeny červeně). Nezávislých vzorů je tedy 25, jak potřebujeme pro 25 pixelů. Zahrnutí záporných hodnot  $h$  (vlny postupující doleva) by vedlo jen k opakování již zahrnutých vzorů. V dolní části obrázku C je ukázáno, že Fourierovou transformací periodických vzorů získáme závislost na vlnočtech vzorů, která připomíná jednopixelové diagramy na obrázku B. V našem ilustrativním obrázku označuje červená barva zápornou hodnotu poskytnutou Fourierovou transformací. Při analýze obrazu ale většinou vynášíme absolutní hodnotu (nerozlišujeme fázi). Na obrázku D je ilustrován Nyquistův zákon (popsáno v textu).

hodnotou indexu lomu. Proto má vlna, která prošla vzorkem, posunutou fázi vůči vlně, která vzorkem neprošla. Rozdíl v intenzitě (amplitudě) vlny přitom může být nepatrný. Převážně fázový kontrast mají vzorky biomakromolekul, jejichž strukturu určujeme pomocí kryoelektronové mikroskopie. Příkladem ze světelné mikroskopie mohou být střípky skla ve vodě nebo průsvitné buňky neobsahující významné množství pigmentu. Pokud na takové vzorky zaostříme, jsou ve světelném mikroskopu velmi špatně viditelné. K zobrazování takových vzorků jsou používány speciální fázově kontrastní mikroskopy. Pokud takové zařízení nevlastníme, můžeme zlepšit kontrast tím, že zaostříme mírně pod sledovaný objekt. Okraje buněk či střípků, kde se setkávají prostředí s různým indexem lomu, jsou při podostření mnohem lépe viditelné.

Princip podostření je zjednodušeně představen na obrázku 13.4. Obraz předmětu je zvětšen optickou soustavou (znázorněnou na obrázku jako jediná čočka). V místě zaostření obrazu se sčítají se stejnou fází vlny přicházející různými cestami z pozorovaného předmětu, jak jsme si ukázali na obrázku 13.2. Pokud vzorek absorbuje záření (případ a na obrázku 13.4), je amplituda vlny prošlé vzorkem (na obrázku 13.4 znázorněna tmavě zeleně) výrazně nižší, než vlna vzorkem neovlivněná. Amplitudový kontrast, symbolicky znázorněn rozdílnou velikostí zelených elips v horní části obrázku, je významný. Pokud ale vzorek neabsorbuje (případ b na obrázku 13.4), skládá se vlna prošlá vzorkem s původní vlnou tak, že se amplituda výsledné vlny liší od amplitudy původní vlny jen zanedbatelně a kontrast mezi nimi je zanedbatelný. Jak jsme vysvětlili v kapitole 11 (část 11.3 a 11.5), je tomu tak proto, že vlna vzniklá rozkmitáním molekul vzorku je oproti dopadající vlně fázově opožděna o  $90^\circ$  a má mnohem menší amplitudu.

Situace se ale změní, pokud nebude obraz zaostřen v důsledku sférické aberace<sup>3</sup> nebo proto, že detektor úmyslně posuneme z roviny zaostřeného obrazu (*defocus*). V tom případě se v určitém pixelu detektoru setkají vlny s rozdílnou fází, které se částečně odečtou. Tak vznikne fázový kontrast, ať už vzorek záření pohlcovává (případ c na obrázku 13.4), nebo ne (případ d).

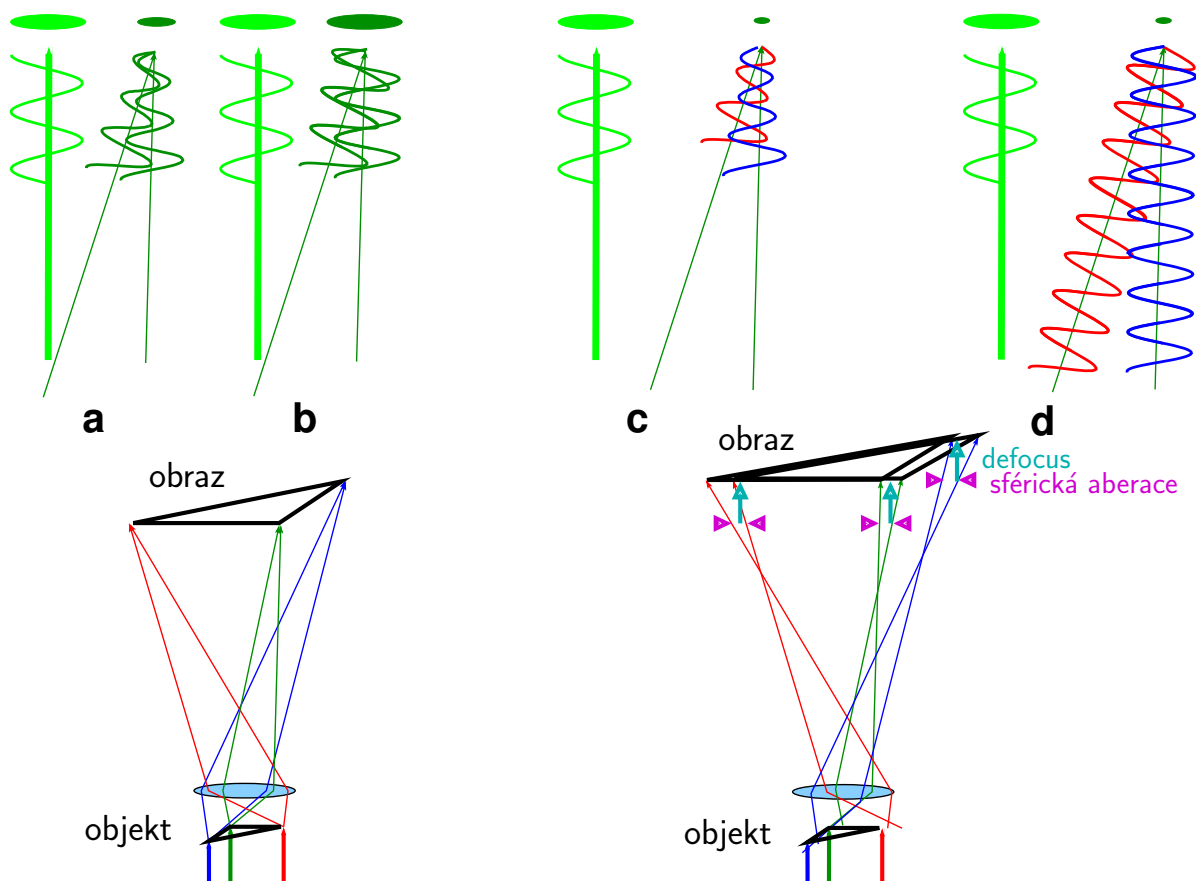
Podostření (*underfocus*) používáme i v kryoelektronové mikroskopii. Pokud je elektronový mikroskop zaostřen, vznikne v určité rovině pod objektivovou čočkou zaostřený obraz vzorku, který další soustava čoček zvětší a promítne na detektor. Jestliže není vlna vzorkem pohlcována více, než okolím (molekulami vody v amorfním ledu), nepozorujeme na detektoru žádný signál. Když ale mírně snížíme proud v objektivové čočce, zaostřený obraz vznikne v jiné rovině, dále od objektivové čočky. Na detektor bude ale stále promítána složená vlna z původní roviny, ve které není obraz zaostřený. Díky rozostření a značné sférické aberaci elektromagnetických čoček dochází k fázovému kontrastu mezi vlnou ovlivněnou a neovlivněnou vzorkem.

Chceme-li získat kvalitní data pro výpočet struktur, musíme vzniku fázového kontrastu rozumět kvantitativně. Využijeme opět analogii se světelnou vlnou. V části 11.3 jsme zmínili, že světelná vlna rozptýlená molekulami bývá fázově opožděná o  $90^\circ$  oproti dopadající vlně. Jak víme z části 13.4, v místě zaostřeného obrazu má světelná vlna stejnou fázi, jako když opouští objekt. Mimo zaostřený obraz ale není vlna ovlivněná molekulami ve vzorku posunuta vůči vlně původní o  $90^\circ$ . V závislosti na délce optické dráhy mají vlny rozptýlené vzorkem do různých směrů (a ohnuté čočkami zpět na detektor) fázové posuny v celém rozsahu  $0^\circ$ – $360^\circ$ . Vlny posunuté o  $180^\circ$  při skládání s neovlivněnou vlnou její amplitudu snižují a vytvářejí tak negativní fázový kontrast. Naopak vlny s nulovým fázovým rozdílem amplitudu původní vlny zvyšují a vedou k pozitivnímu fázovému kontrastu. Pro kvantitativní popis je výhodné využít toho, že díky Fourierově transformaci umíme obraz popsat pomocí souřadnic  $h$  a  $k$ . Na těchto souřadnicích závisí kontrast jednoduchým způsobem: v soustředných kruzích kolem počátku ( $h = k = 0$ ) osciluje mezi hodnotami  $+1$  a  $-1$  (obrázek 13.5). Tato závislost se nazývá *funkce přenosu kontrastu* (*contrast transfer function, CTF*)<sup>4</sup>. Pro vyhodnocení vlivu podostření na získaný obraz stačí převést Fourierovou transformací obraz do souřadnic  $h, k$ , vynásobit funkcí přenosu

<sup>3</sup>Sférické aberace je optická vada, kdy se světelné paprsky na okraji čočky lámou víc, než poblíž optické osy.

<sup>4</sup>Matematicky lze popsat oscilace funkcí  $\sin(\pi C_S \lambda^3 k^4 / 2 - \pi \Delta z \lambda k^2)$ , kde  $C_S$  je koeficient popisující vliv sférické aberace a  $\Delta z$  udává, o kolik mikrometrů je obraz podostřen (*defocus*). Amplituda oscilací navíc s rostoucím  $k$  klesá.





Obrázek 13.4: Rozdíl mezi amplitudovým a fázovým kontrastem. Vlevo je znázorněna tvorba zaostřeného obrazu, vpravo obrazu rozostřeného a ovlivněného sférickou aberací čočky. Na obrázcích a,c je srovnána intenzita světla letícího mimo objekt (světle zeleně vlevo) s intenzitou ovlivněnou průchodem objektem s amplitudovým kontrastem (tmavě zeleně vpravo). Na obrázcích b,d je srovnána intenzita světla letícího mimo objekt (světle zeleně vlevo) s intenzitou ovlivněnou průchodem objektem s amplitudovým kontrastem (tmavě zeleně vpravo). Intenzita záření dopadajícího na detektor je znázorněna velikostí zelené elipsy v horní části.

kontrastu, a výsledek převést Fourierovou transformací zpět do souřadnic  $x, y$ .

Skutečnost, že funkce přenosu kontrastu osciluje mezi kladnými a zápornými hodnotami, má na výsledný obraz zásadní vliv. Tvar zobrazeného objektu je sice zachován, ale získaná intenzita neodpovídá elektronové hustotě. Tmavé oblasti obrazu jsou někde zobrazeny světle a naopak (obrázek 13.5A). Částečně lze tento nedostatek odstranit tak, že se obraz převede do souřadnic  $h, k$ , v oblastech, kde je násoben zápornou funkcí přenosu kontrastu, změnímme znaménko (provedeme *inverzi kontrastu*) a po Fourierově transformaci zpět do souřadnic  $x, y$  získáme obraz podobnější skutečné elektronové hustotě (obrázek 13.5B). Neodstraníme tak ale zásadní problém, že CTF prochází nulou a pro určité hodnoty  $h, k$  tak ztratíme signál.

Tento nedostatek řešíme tak, že snímáme obrazy studovaných biomakromolekul s různou hodnotou podostření (obrázek 13.5B,C). To vede k tomu, že v různých obrazech má CTF různé polohy bodů, kde prochází nulou. Jednotlivé obrazy pak zkombinujeme a tak doplníme oblasti nulového signálu signálem z jiných obrazů (obrázek 13.5D). Tímto způsobem získáme informaci v celém rozsahu  $h, k$ , abychom mohli provést rekonstrukci struktury studované biomakromolekuly.

### 13.9 Poškození vzorku

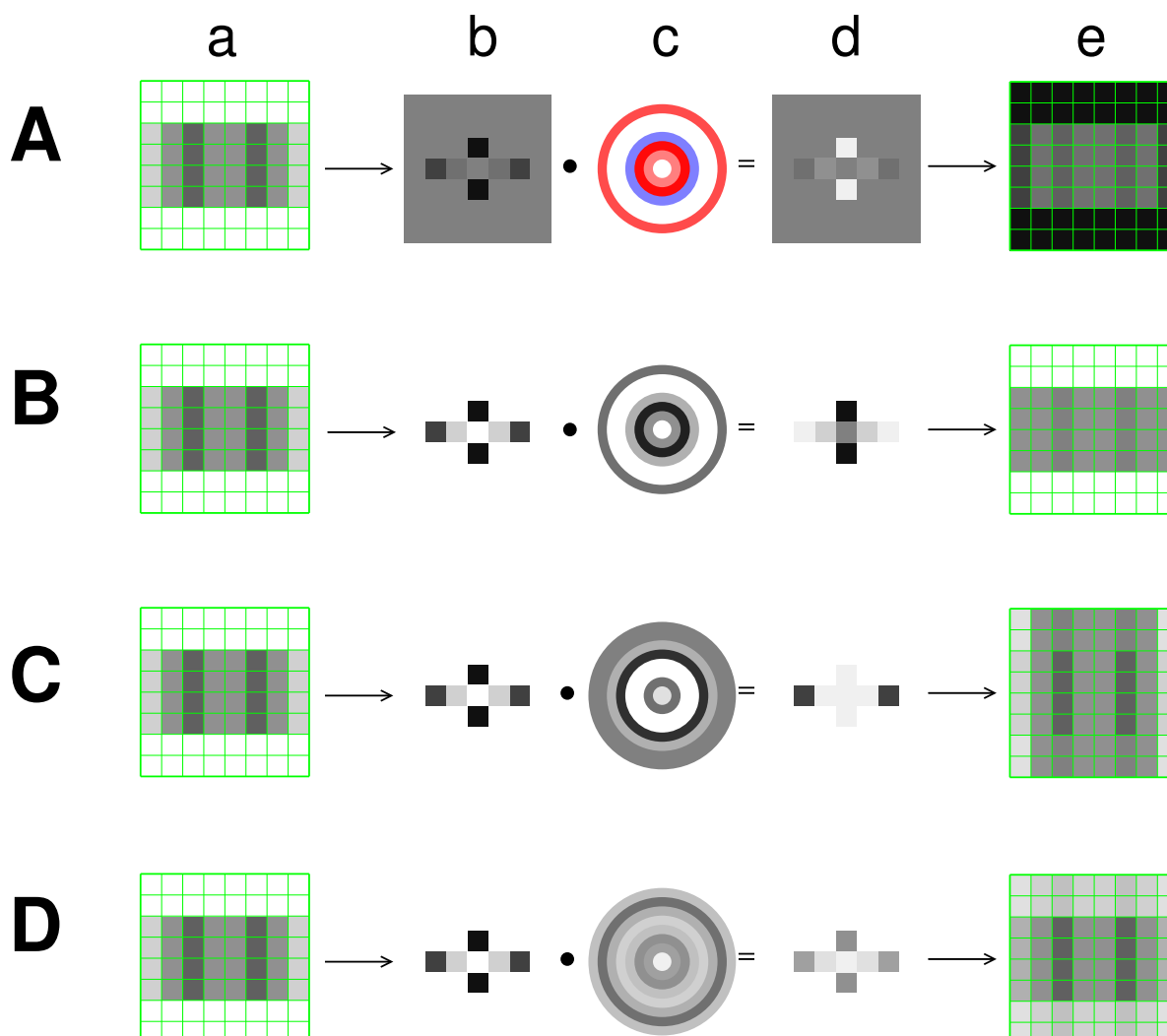
Navzdory chlazení na velmi nízkou teplotu dochází při kryoelektronové mikroskopii k významnému poškození vzorku. Na  $1 \text{ nm}^2$  dopadají stovky až tisíce elektronů a ne všechny interagují se vzorkem elasticky (bez přenosu energie). Některé elektrony předávají vzorku energii, což vede k jeho poškození. Při delším osvětlení se v ledu začínají tvořit bubliny, čemuž je třeba zabránit. Proto se snažíme dávku elektronů omezit na minimum. Například není možné zaostřit mikroskop na oblast našeho zájmu, protože již tím bychom vzorek zničili. Ostříme zpravidla na místa v okolí oblasti zájmu a předpokládáme, že průměr nastavených hodnot bude odpovídat zaostření na místo uprostřed. Naštěstí současné mikroskopy dokáží zaostřit a nasbírat data automaticky. Omezený počet elektronů, který může na vzorek dopadnout, je také hlavní příčinou nízkého poměru signálu k šumu.

### 13.10 Porovnávání dvojrozměrných obrazů

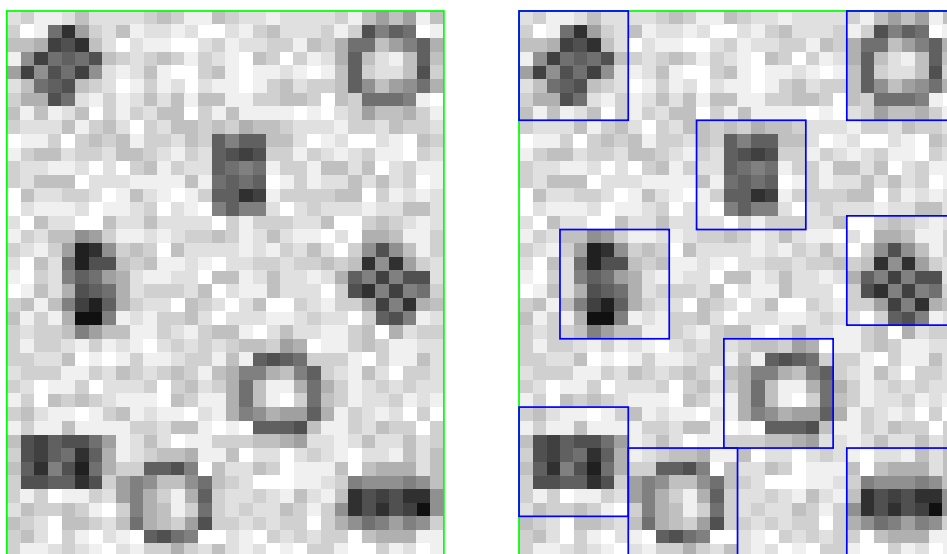
Abychom vylepšili velmi nízký poměr signálu k šumu, snažíme se nasnímat velké množství obrazů s molekulami, jejichž strukturu chceme určit. Velmi zjednodušený příklad je ukázán na obrázku 13.6, kde se hypotetický detektor skládá jen z 1 280 pixelů (ve skutečnosti jde o miliony pixelů). Pro přehlednost je na obrázku velmi vysoký poměr signálu k šumu (přibližně 4), ve skutečnosti bývá poměr o jeden až dva řády horší.

Ze snímků je třeba vyříznout obrázky jednotlivých molekul (modré rámečky v pravé polovině obrázku 13.6), jednotlivé obrázky mezi sebou porovnáváme, zjišťujeme, které jsou si nejvíce podobné, abychom je mohli zprůměrovat. Průměrováním snížíme náhodný šum a získáme obrázky, které mají lepší kontrast.

K porovnávání obrázku můžeme použít *analýzu hlavních komponent* (*principal component analysis, PCA*). Jde o přístup, kterým je možné automaticky identifikovat skupiny obrázků, které jsou si podobné. Teoreticky bychom mohli porovnávat intenzitu každého pixelu v každém modrém rámečku na obrázku 13.6 se všemi pixely všech ostatních rámečků. Z matematického pohledu by to znamenalo vynést intenzitu každého pixelu na jinou osu mnohorozměrného grafu (pro rámečky na obrázku 13.6 bychom pracovali v 64-rozměrném prostoru, protože každý rámeček má pixelů 64). Každému rámečku by pak odpovídal jeden polohový vektor v 64-rozměrném prostoru a my bychom hledali, které vektory jsou si nejpodobnější (nacházejí se nejbliž sebe). Takové srovnání by ale bylo příliš výpočetně náročné. Metoda PCA vychází z předpokladu, že určité kombinace intenzit jsou pro jednotlivé tvary typické. Tyto kombinace pak hledá a použije je jako nové osy v jiném mnohorozměrném prostoru. Zjednodušení



Obrázek 13.5: Vliv funkce přenosu kontrastu na snímané obrazy. Pokud by měl pozorovaný objekt velký amplitudový kontrast, pozorovali bychom obraz ukázaný ve sloupci *a*. Protože ale musíme pracovat s fázovým kontrastem a použít podostření, získáváme obraz ukázaný ve sloupci *e*. Fourierova transformace toho obrazu (sloupec *d*) je součinem Fourierovy transformace obrazu ze sloupce *a*, ukázané ve sloupci *b* a funkce přenosu kontrastu (CTF), ukázané ve sloupci *c*. Šipky na obrázku označují Fourierovu transformaci, tečka mezi sloupci *b* a *c* násobení. A, intenzita ovlivněna neupravenou CTF, oscilující mezi kladnými (modrá barva) a zápornými (červená barva) hodnotami, nulová hodnota Fourierovy transformace odpovídá šedé barvě. B, provedena inverze kontrastu. CTF nabývá jen kladných hodnot, znázorněných odstíny šedi. Nulová hodnota Fourierovy transformace odpovídá bílé barvě. C, jako B, ale obraz byl sníman s jiným podostřením. D, kombinace (průměr) předchozích dvou měření. CTF ve sloupci *e* je průměrem funkcí na předchozích dvou řádcích a neprochází (krom počátku) nulou.



Obrázek 13.6: Identifikace obrázků studované molekuly ve snímku elektronové mikroskopie. Vlevo zjednodušený příklad snímku, vpravo tentýž snímek s vyznačenými obrázky (modré rámečky)

je v tom, že stačí porovnat menší počet skutečně nejnepřítěžnějších kombinací a nemusíme ztrácet čas porovnáváním každého pixelu s každým. Velmi jednoduchý příklad je ukázán na obrázku 13.7 pro kombinace pouhých dvou pixelů.

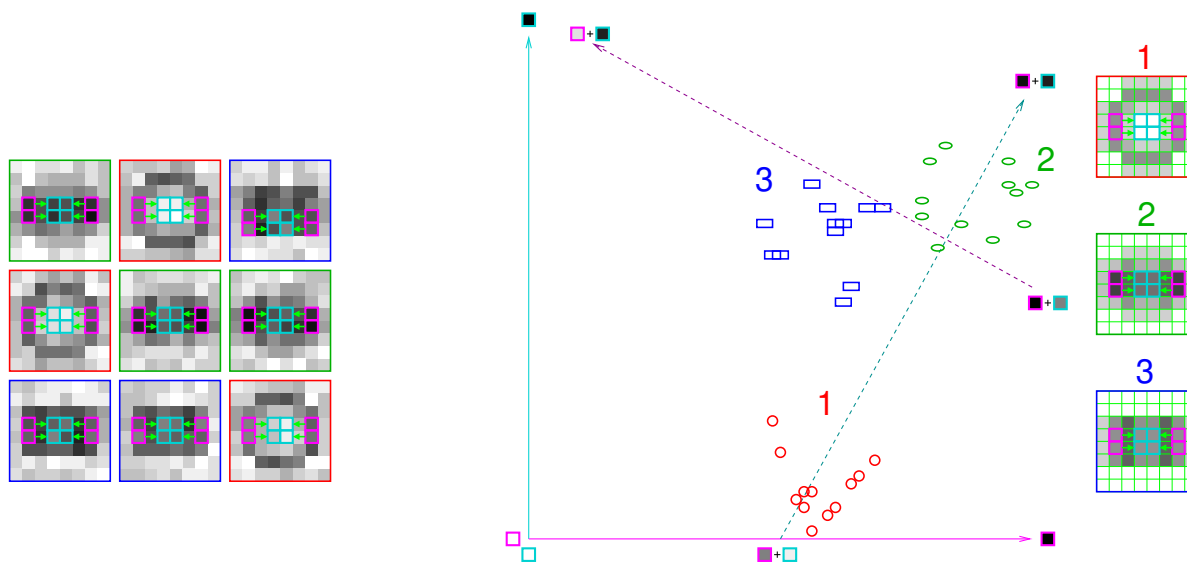
### 13.11 Rekonstrukce trojrozměrné struktury biomakromolekul

Naším konečným cílem je určit trojrozměrnou strukturu. Základním problémem našeho úkolu je, že jednotlivé obrázky představují průměty molekul v různých orientacích, ve kterých se nacházely v roztoku těsně před rychlým zmrazením. Je na nás, abychom zjistili, jak jednotlivé projekce vůči sobě zorientovat, abychom z dvojrozměrných průmětů mohli rekonstruovat trojrozměrnou strukturu. Z matematického pohledu jde vlastně o hledání tří úhlů. Dva úhly určují směr, ze kterého se na molekulu (nebo přesněji skrz molekulu) díváme (směr, kterým míří červené, modré a tmavě zelené šipky na obrázku 13.8A). Dva průměty ze stejného směru mohou být vůči sobě pootočené *v rovině snímku*. Takové vzájemné pootočení popisuje třetí úhel. Tento třetí úhel odpovídá rotaci kolem směru červené, modré, nebo tmavě zelené šipky na obrázku 13.8A<sup>5</sup>

Rekonstrukce trojrozměrné struktury představuje sérii postupných kroků, ve kterých hledáme správné orientace projekcí. V počátcích jde o proces značně zatížený náhodnými chybami, kdy doufáme, že první odhady tvaru molekuly se alespoň trochu podobají skutečné struktuře a že postupné zpřesňování nás nakonec ke správné struktuře dovede.

Obecný postup je shrnut na obrázku 13.9. Vycházíme z velkého počtu obrázků vyznačených ze snímku vzorku (šipka a), pro něž hledáme správnou orientaci vůči molekule (šipky b–d). Obrázky, které jsou projekcemi molekuly ze stejného směru, vzájemně posuneme a natočíme v rovině snímku, abychom je mohli přeložit (*alignment*, šipka b). Potřebná posunutí (translace) a natočení (rotace) ukazují světle

<sup>5</sup>Když použijeme analogie s fotografováním, tak dva úhly určují směr, ze kterého nás fotograf fotil (směr přímky v prostoru je popsán dvěma úhly), a třetí úhel říká, jestli fotograf držel fotoaparát rovně, nebo šikmo (v druhém případě budeme na fotografii vidět svou postavu nakloněnou).

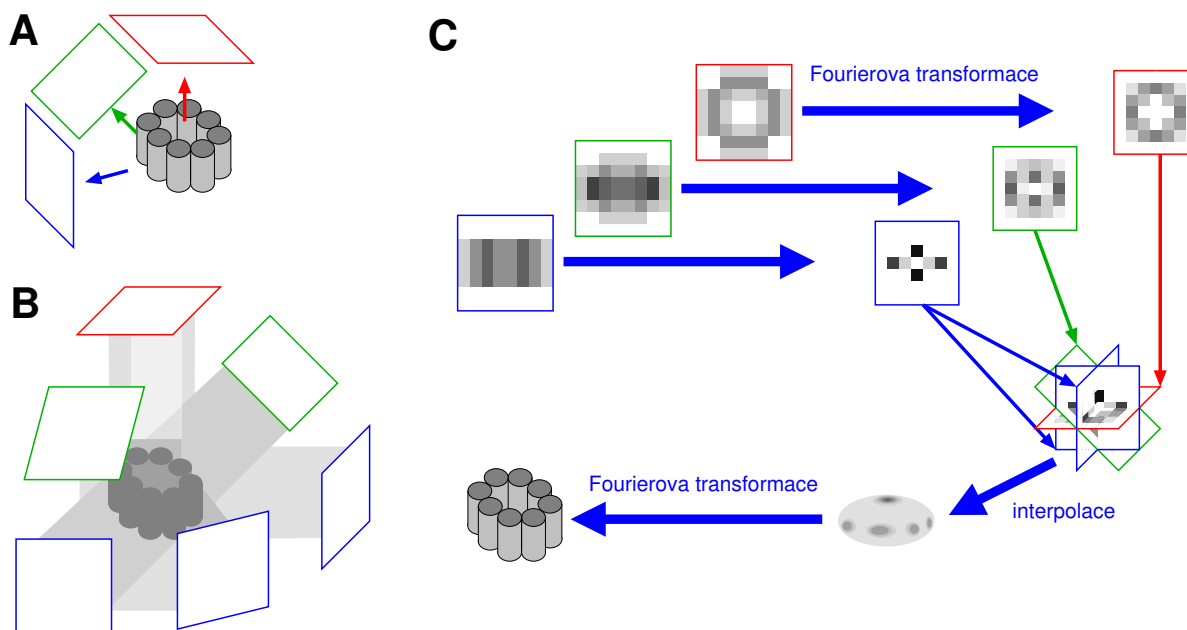


Obrázek 13.7: Princip analýzy hlavních komponent (PCA). Do grafu jsou vyneseny intenzity dvojic pixelů vyznačených na obrázku vlevo (purpurový a azurový rámeček spojený zelenou šipkou). Dvojice intenzit s 25% šumem vytvoří tři shluky bodů, které jsou rozlišeny barvou a odpovídají třem různým projekcím molekuly (projekce bez šumu jsou ukázány vpravo). Čárkované šipky představují kombinace intenzit (*hlavní komponenty*), které nejlépe rozliší shluky. Ukázaný příklad je zjednodušením, které lze ukázat na grafu. Ve skutečnosti bychom hledali hlavní komponenty v 64-rozměrném prostoru intenzit všech pixelů.

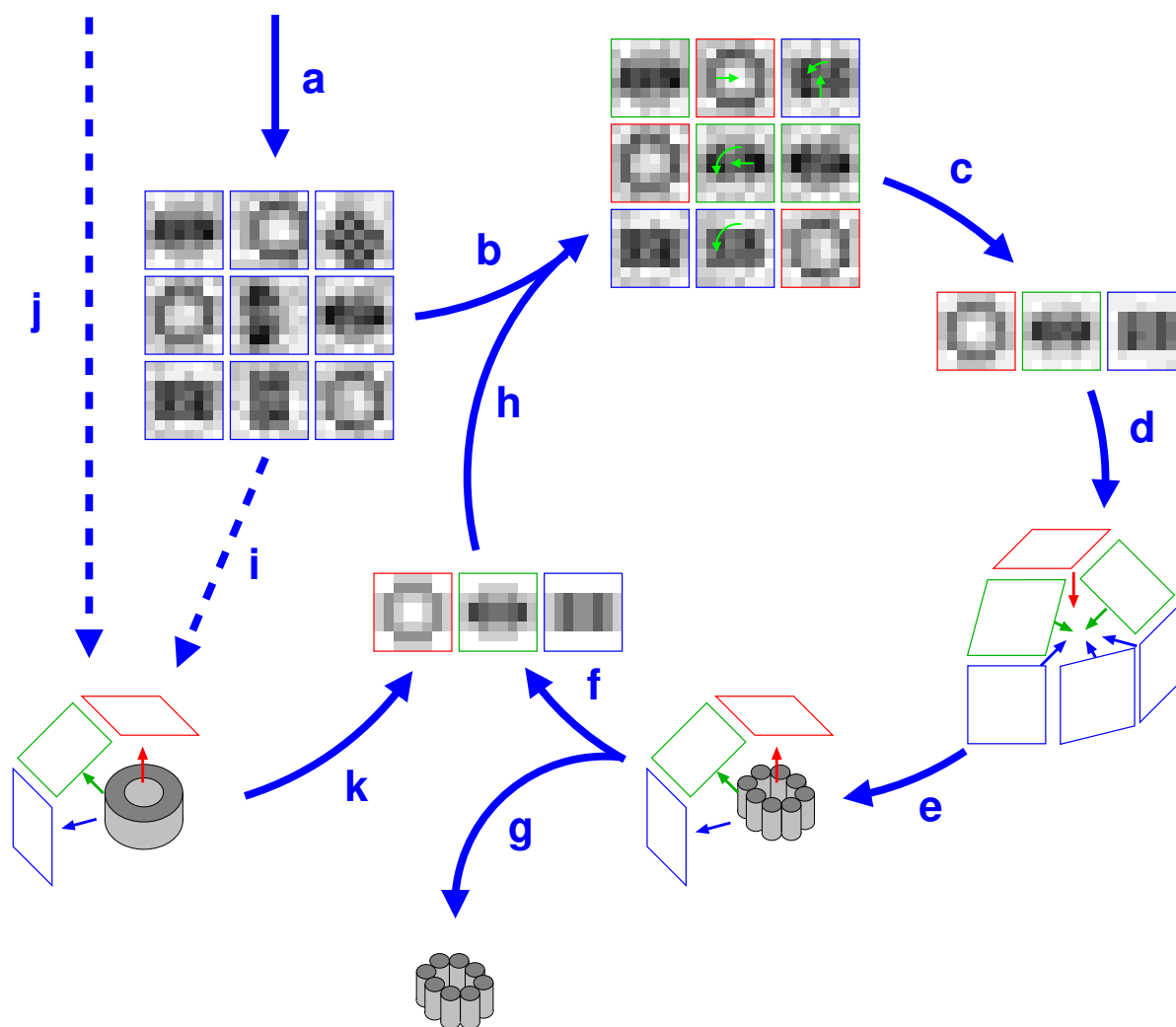
zelené šipky uvnitř obrázků. Posunutí a natočení můžeme určovat vzhledem nějakému referenčnímu obrazu, nebo bez vnější reference (například vzhledem k prvnímu vyříznutému obrázku). Obrázky porovnáváme (*2D classification*, viz část 13.10) a z těch, které odpovídají stejným úhlům projekce (stejná barva rámečku) počítáme průměrný obraz s lepším poměrem signálu k šumu (*class averaging*, šipka c). Srovnání s referenčními obrazy se známými úhly projekce zvyšuje správnost třídění obrázků podle úhlů projekce. Zprůměrovaným obrázkům přiřadíme úhly projekce (*angle assignment*, šipka d). Jako úhly projekce (směry červených, zelených a modrých šipek pro jednotlivé zprůměrované obrazy) použijeme známé hodnoty pro referenční průměty. Pokud začínáme bez vnější reference, použijeme hodnoty náhodné. Z průmětů, které jsme orientovali v prostoru, zpětně vypočítáme tvar molekuly (*3D reconstruction*, šipka e). Zprůměrované a zorientované obrazy (na obrázku čtverečky orámované červeně, zeleně a modře) můžeme považovat za stíny molekuly (obrázek 13.8A). Tvar molekuly můžeme spočítat jako oblast v prostoru, kde se protínají směry, ve kterých jsou stíny vrhány (obrázek 13.8B). Často ale provádíme rekonstrukci v *recipročném prostoru* (obrázek 13.8C). Získaný model molekuly použijeme k tomu, že vypočítáme jeho projekce do různých směrů (*re-projection*, šipka f). Takto získané projekce můžeme použít jako referenční obrazy v dalším kole výpočtů (šipka h). Do každého dalšího kola tak vstupují projekce z více a více zpřesněného modelu molekuly, které srovnáváme se stále stejnými nasnímanými obrázky. Celý postup opakujeme tak dlouho, dokud dochází k zpřesňování modelu. Výsledkem (šipka g) je mapa elektronové hustoty, velmi podobná výsledku rentgenové krystalografie. Do této mapy potom stavíme model molekuly a ověřujeme jeho správnost, tak jak jsme popsali v kapitole 12.

V popsaném iterativním postupu je kritické první kolo, kdy zatím nemáme k dispozici žádné projekce z modelu vypočítaného v předchozím kole (šipka h). Kdyby měly nasnímané obrázky molekul vysoký poměr signálu k šumu, bylo by neobjektivnější vypočítat počáteční model struktury na základě teorie Fourierovy transformace (šipka i). Základní myšlenka je následující. Fourierovou transformací jednot-

livých obrázků získáme řezy trojrozměrné Fourierovy transformace tvaru molekuly (obrázek C). Teorie Fourierovy transformace navíc říká, že jde o řezy procházející středem obrazu v recipročném prostoru. To už je dost geometrických údajů na to, abychom dokázali spočítat, jak se různé řezy se v recipročném prostoru protínají. Pak už stačí následovat postup ukázaný v dolní části obrázku 13.8C a spočítat tvar molekuly. Velký šum ve snímcích získaných kryoelektronovou mikroskopií bohužel zatěžuje tuto elegantní metodu velkou náhodnou chybou. Proto se v praxi používají a jsou dále intenzivně vyvíjeny výpočetní i experimentální postupy poskytující spolehlivější počáteční model pro rekonstrukci. Celý postup značně urychlí, když začínáme již s přibližnou znalostí tvaru molekuly (na základě dat s malým rozlišením, podobnosti s jinými molekulami a podobně, šipka j na obrázku 13.9). Rekonstrukci je možné zahájit dokonce i s obecným tvarem jako koule, válec, atd. Všechny tyto přístupy založené na subjektivní volbě startovního modelu ale nesou riziko, že nás naše volba zavede špatným směrem. Pokud budeme počítač nutit, aby ve snímcích nacházel rysy shodné s počátečním modelem, počítač je tam najde, i kdyby snímky obsahovaly čistý šum. Získaný počáteční model použijeme k tomu, že vypočítáme jeho projekce do různých směrů (šipka k), a použijeme je jako referenční obrazy v prvním kole (šipka k).



Obrázek 13.8: Rekonstrukce trojrozměrné struktury z projekcí se známou orientací. A, jednotlivé projekce (barevně orámované čtverečky) si můžeme představit jako stíny molekuly vrhané do určitých směrů (barevné šipky). B, rekonstrukci tvaru molekuly můžeme provést tak, že každý 2D obraz (stín) zpětně promítneme do směru, odkud přichází. C, rekonstrukce v recipročním prostoru. Pokud známe tvar objektu (například rozložení elektronové hustoty v molekule) ve skutečném trojrozměrném prostoru popsaném souřadnicemi  $x, y, z$ , můžeme vždy díky Fourierově transformaci popsat tento objekt v recipročním prostoru, kde hrají roli souřadnic Millerových indexy  $h, k, l$ . Stejně dobře můžeme Fourierovou transformací popis objektu vyjádřený v souřadnicích  $h, k, l$  přepsat zpátky do reálného tvaru definovaného pomocí souřadnic  $x, y, z$ . Pokud navíc máme velký počet různých dvojrozměrných projekcí objektu, můžeme je také Fourierovou transformací přepočítat do obrazů popsaných souřadnicemi  $h, k$  a zase zpět. Přitom platí, že každý obraz projekce v souřadnicích  $h, k$  představuje jednu rovinu z popisu objektu v trojrozměrném recipročním prostoru  $h, k, l$ . Navíc víme, že každá z těchto ploch prochází středem souřadné soustavy  $h, k, l$ . Proto je možné Fourierovy transformace jednotlivých průmětů použít jako řezy trojrozměrné Fourierovy transformace tvaru molekuly a chybějící body v prostoru souřadnic  $h, k, l$  odhadnout interpolací. Chyba interpolace bude tím menší, čím více různých průmětů použijeme. Ze získaného objektu v recipročním prostoru pak tvar molekuly spočítáme zpětnou 3D Fourierovou transformací.



Obrázek 13.9: Obecný postup rekonstrukce trojrozměrné struktury z dvojrozměrných snímků.



## Kapitola 14

# Nukleární magnetická rezonance

### 14.1 Magnetické chování jader

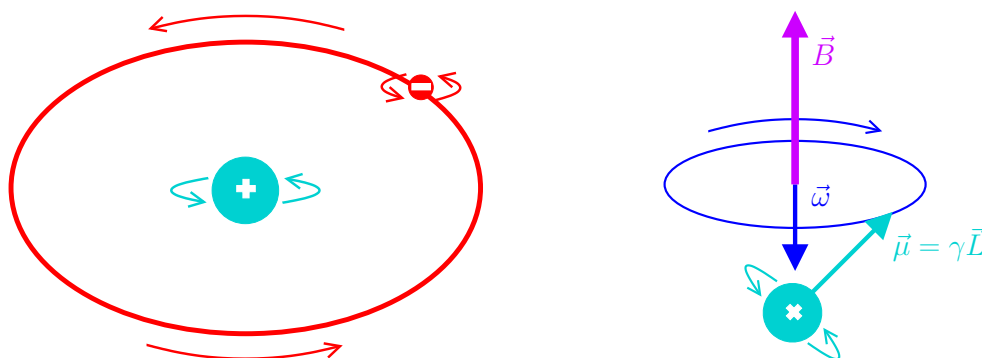
V úvodu kapitoly o využití magnetických vlastností jader atomů k určování struktury molekul je na místě otázka, odkud se magnetismus jader bere. Zdrojem magnetického pole je každý elektrický náboj, který se pohybuje. Pokud je zdrojem smyčka, kterou protéká elektrický proud, popisuje sílu takového magnetu *magnetický moment*  $\vec{\mu}$ . Magnetický moment je úměrný *momentu hybnosti*, který popisuje zakřivený pohyb náboje (elektronů) ve smyčce. Konstanta úměrnosti se nazývá *magnetogyrický poměr* a obvykle se označuje písmenem  $\gamma$ .

K pochopení momentu hybnosti dobře poslouží příklad pohybu Země kolem Slunce. Naše planeta obíhá Slunce po oběžné dráze (po *orbitě*). Tento pohyb můžeme popsat *orbitálním momentem hybnosti*, což je vektor, jehož směr určuje osu, kolem které Země obíhá, a velikost je součinem hybnosti a vzdálenosti od Slunce<sup>1</sup>. Země se ale také zároveň otáčí kolem vlastní osy. Tento pohyb, anglicky označovaný slůvkem *spin*, popisuje *vlastní moment hybnosti*.

Vyměňme teď v našem příkladě Zemi za elektron a Slunce za jádro atomu (obrázek 14.1 vlevo). Elektron se pohybuje kolem našeho jádra podobným způsobem, jako kdyby putoval kruhovou smyčkou z drátu. V elektrotechnice se velikost magnetického momentu takové smyčky definuje jako součin proudu a plochy smyčky. Směr vektoru tohoto *orbitálního magnetického momentu* je stejný jako směr orbitálního momentu hybnosti. Pokud bychom předpokládali, že elektron obíhá po kruhové dráze, nedalo by nám moc práce spočítat, že magnetogyrický poměr je roven jedné polovině poměru náboje k hmotnosti. Podobně bychom mohli popsat *vlastní magnetický moment*, spojený s otáčením elektronu kolem své vlastní osy (se spinem).

Kdybychom zkoumali pohyb elektronu kolem jádra podle zákonů kvantové mechaniky a teorie relativity, došli bychom k trochu paradoxnímu závěru, že pohyb elektronů *nemůžeme* popsat jako otáčení Země kolem vlastní osy, ale že elektrony přesto *mají* jak orbitální moment hybnosti, tak vlastní moment hybnosti (často nazývaný jednoduše spin). Proto mají elektrony také orbitální i vlastní magnetický moment. Hodnota magnetogyrického poměru vlastního momentu hybnosti se ale liší od výsledku, který jsme si uvedli v předchozím odstavci. Správnou hodnotu velmi dobře předpovídá kvantová mechanika v kombinaci s teorií relativity. Kvantová mechanika nám také říká, že ze všech stavů, ve kterých se může magnetický moment osamocenému elektronu v magnetickém poli nacházet, jsou jenom dva *stacionární* (pokud je elektron ve stacionárním stavu, setrvá v něm tak dlouho, dokud jej vnější příčina nedonutí stav změnit). V chemii se tyto stacionární stavy popisují hodnotou *spinového kvantového čísla*  $+\frac{1}{2}$  a  $-\frac{1}{2}$ , nebo symbolicky šipkami  $\uparrow$  a  $\downarrow$ . Kdyby se elektron utrhl z atomu a letěl volně prostorem, ztratil by

<sup>1</sup>Matematicky definujeme moment hybnosti  $\vec{L}$  jako vektorový součin  $\vec{L} = \vec{r} \times \vec{p} = m(\vec{r} \times \vec{v})$ , kde  $\vec{v}$  je rychlost a  $\vec{r}$  vektor popisující polohu Země vzhledem k Slunci.



Obrázek 14.1: Vlevo atom vodíku přirovnaný k Zemi obíhající kolem Slunce. Vpravo znázornění precesního pohybu magnetického momentu v magnetickém poli.

orbitální moment hybnosti, ale vlastní moment hybnosti by mu zůstal. Vlastní moment hybnosti mají i částice, ze kterých se skládá jádro. Vlastní moment hybnosti a magnetický moment jádra se skládají dosti složitým způsobem z příslušných momentů částic, které jádro tvoří a drží pohromadě. Proto se magnetogyrické poměry jader počítají velmi obtížně, dají se ale velmi přesně změřit.

Pro zkoumání struktur biomakromolekul se téměř výhradně používají jádra, která se podobají elektronu tím, že jejich magnetické momenty mají dva stacionární stavy. Z jader, která potkáváme v makromolekulách, jde o izotopy  $^{31}\text{P}$  a  $^1\text{H}$  (což je proton), která tvoří veškerá jádra fosforu a naprostou většinu jader vodíku v přírodě. Navíc k tomuto druhu jader patří  $^{13}\text{C}$  a  $^{15}\text{N}$ , ty však tvoří pouze 1,1 % jader uhlíku a 0,4 % jader dusíku, takže často musíme biomakromolekuly těmito jádry obohacovat.

Pro naši metodu je zásadní, že v přítomnosti vnějšího magnetického pole koná vektor magnetického momentu *precesní pohyb* kolem vektoru magnetické indukce tohoto pole.<sup>2</sup> Měřenou veličinou ve spektroskopii nukleární magnetické rezonance je *frekvence* tohoto precesního pohybu  $\nu$ . Frekvenci i směr precesního pohybu můžeme elegantně popsat vektorem úhlové rychlosti  $\vec{\omega}$ , jehož velikost je rovna  $2\pi\nu$  a jehož směr určuje osu precese. Směr precesního pohybu nám udává pravidlo pravé ruky (vzhledem k  $\vec{\omega}$ , obrázek 14.1 vpravo). Úhlová rychlost závisí pouze na magnetogyrickém poměru  $\gamma$  a na magnetické indukci pole  $\vec{B}$

$$\vec{\omega} = -\gamma\vec{B}. \quad (14.1)$$

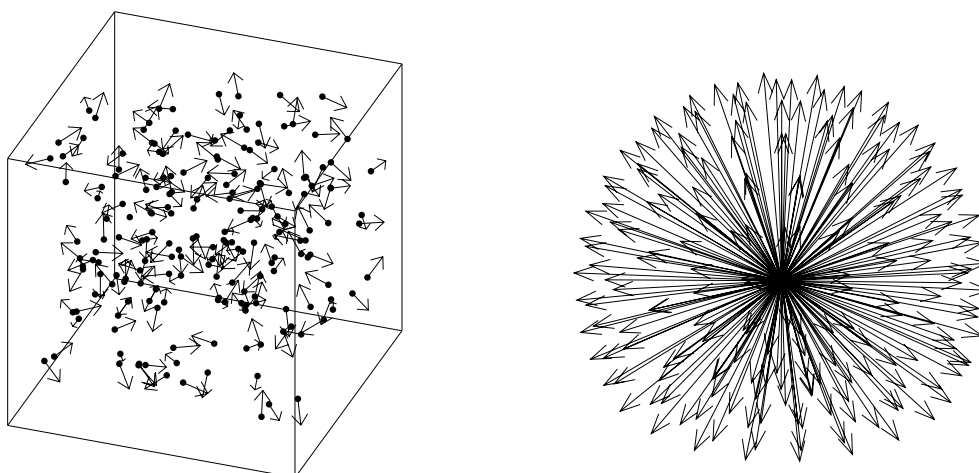
Protože  $\gamma$  je předem známá konstanta, měřením úhlové rychlosti  $\vec{\omega}$  nebo frekvence  $\nu$  vlastně měříme sílu (indukci) magnetického pole v místě jádra.

## 14.2 Ovlivnění souborů jader vnějšími poli

Kvantová mechanika je nezbytná pro správný popis magnetických momentů jednotlivých jader. Při měření spekter nukleární magnetické rezonance (NMR) ale pracujeme s makroskopickými vzorky (většinou s necelým mililitrem roztoku molekul, které zkoumáme), které obsahují miliardy miliard jader (velmi zjednodušená ilustrace je ukázána na obrázku 14.2). Proto celkové magnetické vlastnosti takových vzorků popisuje dobře klasická, nekvantová fyzika.

Popis obrovských množství jader s magnetickými momenty nutně vyžaduje statistický přístup. První otázka, kterou si položíme, bude proto ryze statistická. Jaká je pravděpodobnost, že magnetický moment

<sup>2</sup>Jinými slovy, vektor magnetického momentu se otáčí pod určitým úhlem kolem vektoru magnetické indukce, tedy kolem vektoru, který popisuje směr a sílu magnetického pole. Obdobně koná precesní pohyb roztočený setrvačnick v gravitačním poli, díky tomu se například udržíme v sedle jízdního kola.



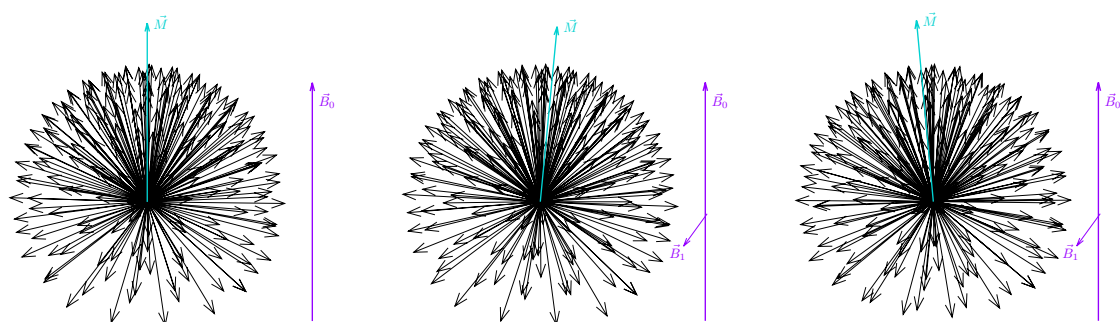
Obrázek 14.2: Vlevo zjednodušený obrázek vzorku proteinu studovaného pomocí NMR. Molekuly proteinu jsou znázorněny jako tečky. Pro každou molekulu je znázorněn jeden magnetický moment jako šipka. Vpravo znázornění orientací magnetických momentů v nepřítomnosti vnějšího magnetického pole. Počátky všech šipek z obrázku vlevo jsou umístěny do stejného bodu. Obrázek ukazuje, že magnetické momenty (šipky) míří se stejnou pravděpodobností do všech směrů.

zkoumaného jádra míří určitým směrem? V nepřítomnosti magnetického pole jsou všechny orientace magnetického momentu stejně pravděpodobné. Výsledný magnetický moment celého vzorku je součet vektorů, které popisují magnetické momenty jednotlivých jader. Protože se stejnou pravděpodobností míří do všech možných směrů, vzájemně se vyruší a výsledný magnetický moment je nulový. Abychom získali navenek pozorovatelný celkový magnetický moment, musíme zajistit, aby jádra s různou orientací magnetického momentu měla různou energii a některé orientace byly tedy více pravděpodobné, než jiné. Nejjednodušší způsob, jak to zajistit, je umístit jádra do homogenního magnetického pole, tedy do pole, ve kterém má vektor magnetické indukce, zpravidla označovaný  $\vec{B}_0$ , všude stejný směr. Směr  $\vec{B}_0$  přitom použijeme jako definici osy  $z$  souřadné soustavy, která nám pomůže magnetické chování jader popsat konkrétními geometrickými pojmy.

Po umístění vzorku do magnetického pole pomalu naroste počet magnetických momentů, které míří stejným směrem, jako  $\vec{B}_0$  (tento směr je energeticky výhodnější). Nadbytek magnetických momentů orientovaných podél  $\vec{B}_0$  se navenek projeví jako nenulový celkový magnetický moment (obrázek 14.3). Velikost celkového magnetického momentu dělená objemem vzorku se nazývá magnetizace a zpravidla označuje písmenkem  $M$ . V tepelné rovnováze je tento nadbytek malý i v nejsilnějších magnetech, protože magnetické momenty jader jsou velmi slabé a jen nepatrně přispívají k energii molekul.

### 14.3 Excitace

Přítomnost vnějšího magnetického pole sice učiní vzorek navenek magnetickým, neumožní nám ale měřit frekvenci precese magnetických momentů (alespoň ne běžným způsobem). Ačkoli náš vzorek obsahuje miliardy miliard magnetických momentů, které konají precesní pohyb s (téměř) stejnou frekvencí, detektor tuto frekvenci neregistruje, protože celkový magnetický moment je orientován podél  $\vec{B}_0$  (otáčením kolem svíslého směru se azurová šipka na obrázku 14.3 nijak nemění). Abychom mohli frekvenci měřit, musíme zajistit, aby azurová šipka (celkový magnetický moment) mířila jinam, než vektor  $\vec{B}_0$ , nejlépe



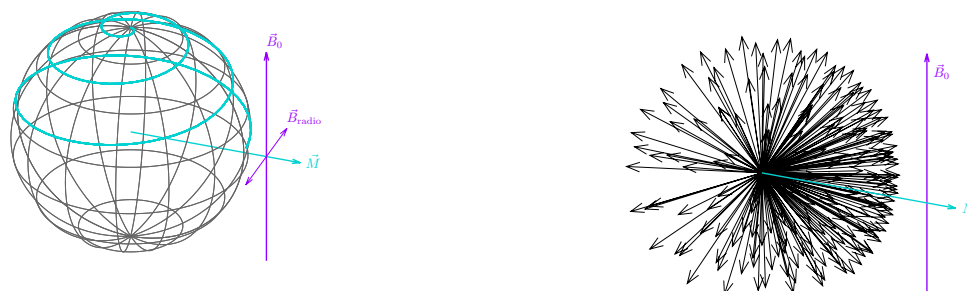
Obrázek 14.3: Vlevo je ukázána distribuce magnetických momentů jader ve vnějším magnetickém poli o indukci  $\vec{B}_0$  (fialová šipka). Celkový magnetický moment je znázorněn jako azurová šipka. Uprostřed je ukázána distribuce magnetických momentů odpovídající magnetizaci (azurová šipka) vychýlené mírně doprava dalším magnetickým polem o indukci  $\vec{B}_1$  (tenká fialová šipka), které otáčí magnetické momenty po směru hodinových ručiček. Vpravo je ukázána tatáž distribuce otočená o polovinu kruhu v důsledku precese kolem  $\vec{B}_0$ . Slabší pole o indukci  $\vec{B}_1$ , otáčející magnetické momenty po směru hodinových ručiček, v tento okamžik vrací vektor magnetizace (azurovou šipku) zpět do svislého směru.

kolmo k němu.

Již víme, že vektor magnetického momentu se otáčí kolem vektoru magnetické indukce. To nám napovídá, že k otočení azurové šipky můžeme použít další magnetické pole, s indukci  $\vec{B}_1$  kolmou ke směru  $\vec{B}_0$ .<sup>3</sup> Pokud bude vektor indukce  $\vec{B}_1$  mířit směrem k nám, bude otáčet azurovou šipku na obrázku 14.3 po směru hodinových ručiček. Ale pozor! Vektor  $\vec{B}_1$  nemůže jednoduše mířit jedním směrem kolmým na  $\vec{B}_0$ . Jakmile bude azurová šipka nepatrně vychýlena po směru hodinových ručiček ze směru  $\vec{B}_0$ , tedy na našem obrázku doprava, začne kolem  $\vec{B}_0$  konat precesní pohyb s frekvencí  $\nu_0$ . Po době odpovídající polovině periody tohoto precesního pohybu se azurová šipka otočí o půl otáčky kolem svislé osy, takže bude na obrázku mířit trochu doleva od svislého směru. To ale znamená, že v tomto okamžiku otáčení po směru hodinových ručiček (efekt  $\vec{B}_1$  stále mířícího směrem k nám) bude azurovou šipku vracet zpátky do svislého směru.

Vidíme, že vektor  $\vec{B}_1$  mířící stále jedním směrem nezpůsobí celkovou rotaci azurové šipky ze svislého směru do směru vodorovného. Potřebujeme vektor  $\vec{B}_1$ , který osciluje mezi směrem k nám a od nás se stejnou frekvencí, s jakou se vektor magnetizace (azurová šipka) otáčí kolem  $\vec{B}_0$ . V tom případě bude  $\vec{B}_1$  stále více vychylovat azurovou šipku ze svislého směru. V okamžiku, kdy je azurová šipka vychýlena doleva, míří  $\vec{B}_1$  k nám a otáčí magnetizaci po směru hodinových ručiček, tedy ještě více vlevo dolů. Ve chvíli, kdy je azurová šipka vychýlena doprava, míří  $\vec{B}_1$  od nás a otáčí magnetizaci proti směru hodinových ručiček, tedy ještě více vpravo dolů. Vektor magnetizace se tak více a více vzdaluje od svislého směru po složité spirálové dráze, znázorněné na obrázku 14.4 vlevo. Po určité době bude azurová šipka vodorovná, což je náš cíl (obrázek 14.4 vpravo). Pokud v tomto okamžiku přestane pole s indukci  $\vec{B}_1$  působit, bude celkový magnetický moment (azurová šipka) rotovat ve vodorovné rovině kolem svislé osy. Protože magnetický moment je sám zdrojem magnetického pole, vznikne tak maličké magnetické pole otáčející se s frekvencí precese magnetických momentů pozorovaných jader. Pokud bude v blízkosti cívka nebo smyčka z vodivého drátu, bude otáčející se pole v drátu indukovat střídavé elektromotorické napětí, oscilující s frekvencí precese. Toto napětí představuje signál, který v experimentu NMR měříme.

<sup>3</sup>V našich úvahách i ve skutečném experimentu bude velikost  $\vec{B}_1$  vždy mnohem menší, než velikost  $\vec{B}_0$ .



Obrázek 14.4: Vlevo znázornění pohybu vektoru magnetizace (jako azurová dráha na povrchu koule o poloměru  $|\vec{M}|$ ) v přítomnosti oscilujícího magnetického pole radiové vlny. Vpravo distribuce magnetických momentů jader po sklopení vektoru magnetizace do směru kolmého na  $\vec{B}_0$ .

## 14.4 Interakce jader s nejbližším okolím

Z popisu experimentu NMR v předchozí části nijak nevyplývá, proč by měla precese magnetického momentu některých jader zajímat chemika či biochemika. Praktické využití NMR v chemii a biochemii souvisí s tím, že vnější magnetické pole našeho magnetu není jediný vliv, kterému jsou magnetické momenty jader vystaveny.

Kromě uměle vyvolaných polí interagují jádra s magnetickými poli v jejich bezprostředním (molekulárním) okolí. Tyto interakce jsou pro nás velmi zajímavé, protože nesou informaci o rozložení elektronů a jader atomů v nejbližším okolí jádra. Jenom díky těmto interakcím můžeme použít NMR ke zkoumání vlastností molekul. Ve všech biomakromolekulách se setkáváme se třemi typy interakcí, o kterých se zmíníme v následujících oddílech trochu podrobněji.

## 14.5 Interakce s magnetickými poli párovaných elektronů

Silné magnetické pole spektrometru indukuje pohyb elektronů v molekulách. Důsledkem tohoto pohybu (pohyb nabitých částic je vlastně elektrický proud a kolem vodičů pod proudem se vytváří magnetické pole) je indukované magnetické pole, které působí proti směru vnějšího pole, které je vyvolalo. Vnější magnetické pole je proto v místě jádra částečně stíněno elektronovým obalem (obrázek 14.5A). Magnetické momenty jader jsou ale také ovlivněny poli, která vznikají v důsledku pohybu elektronů v sousedních atomech. Pokud provádíme měření v roztoku, směr a velikost těchto polí se velmi rychle mění, tak rychle, jak rychle se otáčejí molekuly v důsledku srážek s molekulami rozpouštědla.<sup>4</sup> Velmi zjednodušeně je to ukázáno na obrázku 14.5B a 14.5C. Azurové jádro na obrázku 14.5B se nachází v takové poloze vůči šedému jádru, že pohyb elektronů šedého atomů zvyšuje magnetické pole v místě azurového jádra. Pokud ale srážka s rozpouštědlem pootočí pozorovanou molulu do orientace ukázané na obrázku 14.5C, octne se azurové jádro v poloze, ve které pohyb elektronů šedého atomů magnetické pole mírně snižuje. Ve skutečných molekulách jsou pozorované magnetické momenty ovlivněny elektrony několika sousedních atomů, takže je na místě mluvit o vlivu rozložení elektronové hustoty, které není v okolí pozorovaného jádra sféricky symetrické (snad s výjimkou atomů vzácných plynů). Zmíněné malé a rychlé změny magnetického pole v důsledku srážek s rozpouštědlem jsou jedinou příčinou toho, že po vložení vzorku do magnetu se zvýší počet magnetických momentů mířících podél vnějšího magnetického

<sup>4</sup>Jde o dobře známý Brownův pohyb, z něhož nás teď zajímá pouze náhodné pootáčení molekuly (rotace), nikoli posun v určitém směru (translace). Krom pohybu celých molekul přispívají ke kolísání magnetických polí také vnitřní pohyby molekul. Vliv vnitřních pohybů je sice menší, ale z chemického a biologického pohledu jsou vnitřní pohyby zajímavější.

pole a dojde ke vzniku navenek pozorovatelné celkové magnetizace. Jak k tomu dochází? Zkusíme si to popsat opět s pomocí klasické fyziky. Není to sice korektní, ale umožní nám to získat intuitivní představu o hodně složitém, ale důležitém jevu.

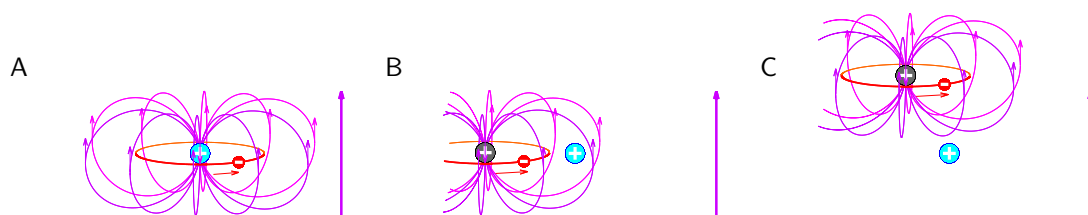
Magnetické momenty jsou absolutně netečné k mechanickému otočení molekuly. Ať molekulu natočíme jakkoli, magnetické momenty konají precesní pohyb kolem směru vnějšího magnetického pole. Magnetické momenty ale cítí změny magnetických polí. Pokud v určitém okamžiku srážka s rozpouštědlem roztočí pozorovanou molekulu náhodou právě takovou rychlostí, jako je rychlost precesního pohybu, dojde k tomu, co jsme si popsali pro radiové vlny. Rozkmitání magnetického pole (v našem případě v důsledku nárazu molekuly rozpouštědla) ve směru kolmém na vnější pole s frekvencí rezonující s frekvencí precese určitého magnetického momentu začne otáčet magnetický moment po spirále ukázané na obrázku 14.4 vlevo. Po spirále se můžeme pohybovat nahoru nebo dolů, v průměru budou častější ty, které budou magnetický moment tlačit do energeticky výhodnějšího směru (tedy do směru rovnoběžného s vnějším polem). Na rozdíl od radiových vln, které mají stále stejnou frekvenci, otáčení molekuly zasažené rozpouštědlem bude mít tu správnou rychlost a směr jen na okamžik, než přijde další náraz, který vzácnou souhru pokazí. Magnetický moment se proto posune po azurové spirále jen maličko a pak dlouho čeká, až zase přijde ta správná srážka, aby se posunul po spirále o kousek dál. Mezitím možná dojde ke srážce, která vrátí magnetický moment zpátky. Jak jsme si ale řekli, jde o náhodný jev, který nakonec v souboru všech magnetických momentů povede k takovému rozložení magnetických momentů, které bude odpovídat termodynamické rovnováze při dané teplotě.

Tak jako kolísání magnetického pole kolmo na vnější pole vede k ustavení rovnovážného rozložení magnetických momentů a vzniku celkové magnetizace po vložení vzorku do magnetu, tak samozřejmě navrací magnetizaci zpět do rovnováhy poté, co jsme ji radiovými vlnami vychýlili do kolmého směru. Díky kolmým výchylkám tedy signál získáváme (kdyby nevznikla v rovnováze celková magnetizace, nebylo by co sklápat radiovými vlnami do měřitelného směru), ale i ztrácíme. Na ztrátu signálu má ale vliv i kolísání rovnoběžné s vnějším konstantním magnetickým polem  $\vec{B}_0$ , tedy v našem uspořádání svislé výchylky. Kolem svislého směru (směru  $\vec{B}_0$ ) totiž magnetické momenty konají precesní pohyb s frekvencí, kterou měříme. Pokud se všechny magnetické momenty pohybují se stejnou frekvencí, tedy *koherentně*, celková magnetizace během měření rotuje se stejnou, přesně danou, frekvencí kolem  $\vec{B}_0$ . V důsledku pohybů molekul se ale každé jádro nachází v maličko jiném a stále se měnícím magnetickém poli. Protože frekvence precese je přímo úměrná magnetickému poli, tak se v každém okamžiku otáčí každý magnetický moment s nepatrně odlišnou frekvencí. Po delší době toto kolísání frekvencí vede k „rozostření“ celkové magnetizace a tedy úbytku měřeného signálu. Tento efekt je zvláště silný u velkých, pomalu se pohybujících molekul. To je jeden z důvodů, proč je velmi obtížné pomocí NMR zkoumat proteiny skládající se z více než 200 aminokyselin.

Diskuse kolísání frekvencí přirozeně vede k otázce, jakou frekvenci v experimentu NMR vlastně měříme. Asi nepřekvapí, že jde o frekvenci průměrnou, tedy o průměr frekvencí magnetického momentu pozorovaného jádra ve všech různě natočených molekulách v roztoku. Rozdílné rozložení elektronové hustoty způsobuje, že různá jádra téhož izotopu (například proton v methylové skupině a proton v aromatickém systému) mají různou průměrnou precesní frekvenci. Ve spektrech pozorujeme takzvaný *chemický posun*, většinou označovaný  $\delta$ . Protože precesní frekvence umíme měřit velmi přesně, dokážeme dobře rozlišit chemický posun jader v různých skupinách. Chemický posun se obvykle vyjadřuje v milióntinách precesní frekvence (jednotka ppm) vůči dohodnuté srovnávací molekule<sup>5</sup> a pro malé molekuly je nejčastěji využívanou informací o struktuře. Hodnoty chemického posunu protonů v různých funkčních skupinách jsou uvedeny na obrázku 14.6. Vidíme, že jde vždy o hodnoty nižší, než pro volný proton ( $H^+$ ).

Rozložení elektronové hustoty samozřejmě obsahuje cennou informaci o struktuře (vzpomeňte na

<sup>5</sup>Fyzikálně nejpřirozenější by bylo srovnání s volným protonem. Volné protony ale nejsou chemikálie, kterou bychom si mohli objednat z katalogů chemických firem. Proto se používají jiné, stabilní látky: v organické chemii tetramethylsilan (TMS) a v biochemii 3-(trimethylsilyl)-propan-1-sulfonová kyselina, která je dobře rozpustná ve vodě.



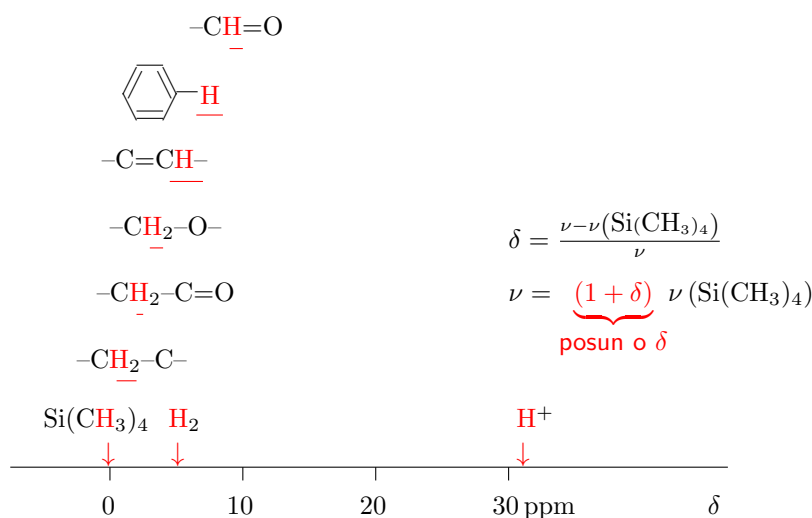
Obrázek 14.5: A, Klasický popis interakce pozorovaného magnetického momentu s orbitálním magnetickým momentem elektronu téhož atomu. Pozorované jádro je znázorněno azurovou barvou, elektron červeně. Vnější homogenní magnetické pole je znázorněno tlustou fialovou šipkou, magnetické pole elektronu je znázorněno tenkými fialovými magnetickými indukčními čarami (jejich hustota a tečna v daném místě udávají velikost a směr vektoru magnetické indukce, malé šipky vyznačují směr). Elektron se v magnetickém poli pohybuje po kruhové dráze nakreslené červeně, směr pohybu elektronu je vyznačen červenou šipkou. Magnetické pole orbitálního magnetického momentu vlastního elektronu snižuje celkové magnetické pole v místě azurového jádra (malá fialová šipka v místě azurového jádra míří dolů). B, interakce pozorovaného magnetického momentu s orbitálním magnetickým momentem elektronu sousedního atomu. Sousední jádro je nakresleno šedě. V této orientaci molekuly magnetické pole orbitálního magnetického momentu elektronu sousedního atomu zvyšuje celkové magnetické pole v místě azurového jádra (indukční čáry v místě azurového jádra míří vzhůru, jak ukazují malé fialové šipky). C, při otáčení celé molekuly se azurové jádro dostane do pozice, ve které pole orbitálního magnetického momentu elektronu sousedního atomu začíná snižovat celkové magnetické pole v místě azurového jádra (indukční čáry v místě azurového jádra začínají mířit dolů).

rentgenovou difrakci). V současné době ale nedokážeme vliv interakcí s elektrony na frekvence jader ve velkých molekulách dostatečně přesně předpovědět a hodnoty chemického posunu se využívají k určování struktur jen omezeně.

## 14.6 Přímá interakce s magnetickými dipóly sousedních jader

Kromě elektronových oblaků se v sousedství pozorovaných magnetických momentů nacházejí také magnetické momenty okolních jader. Tyto magnetické momenty jsou zdroji malých lokálních polí, jejichž velikost a směr závisí na natočení celé molekuly (podobně jako velikost a směr polí způsobených orbitálními magnetickými momenty elektronů). Magnetický dipólový moment sousedního jádra může zvyšovat nebo snižovat magnetické pole v místě sledovaného jádra, v závislosti na tom, v jaké poloze se sousední jádro nachází vzhledem ke sledovanému jádru. Pokud je spojnice jader kolmá na směr indukce vnějšího pole, jako vlevo na obrázku 14.7, magnetické pole sousedního jádra má v místě pozorovaného jádra opačný směr a vnější pole proto snižuje. Pokud ale otočíme molekulu o  $60^\circ$  nebo o  $90^\circ$  (obrázek 14.7 vpravo), míří pole sousedního jádra v místě pozorovaného jádra (šikmo) vzhůru a vnější pole tak zesiluje. Pokud je magnetická indukce sousedního jádra přesně kolmá na indukcí vnějšího pole (obrázek 14.7 uprostřed), nemá magnetický moment sousedního jádra na pole v místě pozorovaného jádra žádný vliv. Úhel mezi spojnici jader a indukcí vnějšího pole, pro který toto nastane, se nazývá magický úhel.

Protože precesní frekvence jádra závisí na magnetickém poli, nese spektrum molekuly v pevném stavu (například v krystalu) informaci o vzájemné orientaci jader. V roztoku se ovšem tato vzájemná orientace stále mění v důsledku Brownova pohybu molekul. Pole sousedních jader se tedy neustále náhodně mění (fluktuují). Pokud se molekula nachází ve zcela neuspořádaném (izotropním) prostředí,



Obrázek 14.6: Typické hodnoty chemického posunu protonů v různých funkčních skupinách.

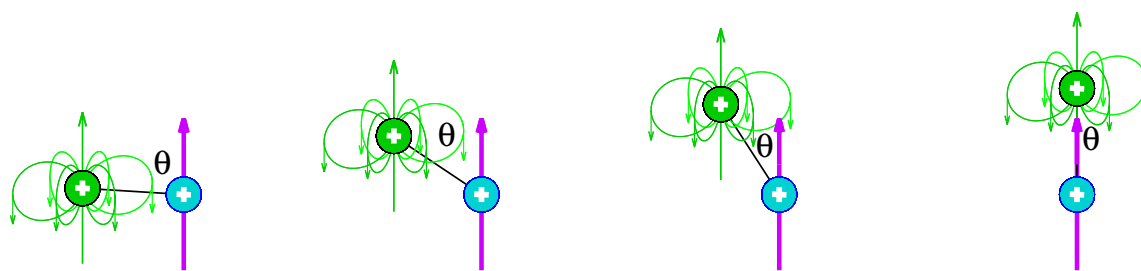
například ve vodném roztoku, průměrná hodnota polí sousedních jader je nulová. Přesto mají tato pole vliv na stav sledovaného jádra. Kolísání okamžitých polí sousedních jader vede k rovnovážnému rozložení magnetických momentů zcela stejným způsobem, jak jsme si popsali pro pole tvořená orbitálními momenty elektronů.

Dalším důsledkem interakcí s magnetickým dipólovým momentem sousedního jádra je takzvaný *nukleární Overhauserův jev* (NOE, Nuclear Overhauser Effect): pokud selektivně ozařujeme radiovými vlnami jen sledované jádro a ponecháme sousední jádro v rovnováze,<sup>6</sup> interakce mezi magnetickými momenty jader způsobí nejen to, že momenty obou jader nakonec dospějí do rovnováhy, ale také to, že rovnovážné rozložení magnetických momentů sousedního jádra (netknutých radiovými vlnami) bude dočasně narušeno. Ozářením jednoho jádra tak způsobíme nerovnováhu magnetických momentů jader v blízkém okolí, což můžeme sledovat jako změnu intenzity měřeného signálu. Protože účinnost tohoto jevu klesá s šestou mocninou vzdálenosti mezi jádry, používá se NOE velmi často k měření vzdálenosti mezi jádry.

Srovnajme si teď vlivy orbitálních magnetických momentů elektronů a vlastních magnetických momentů jader. *Dipól-dipólové interakce* mezi jádry a nerovnoměrné rozložení indukovaných polí elektronů, nazývané *anizotropie chemického posunu*, závisí na orientaci molekuly a v důsledku pohybů molekul vedou ke kolísání lokálních polí. Toto kolísání způsobuje *relaxaci*, tedy návrat rozložení magnetických momentů do termodynamické rovnováhy. Průměrná hodnota kolísajících polí v izotropním prostředí je v případě dipól-dipólové interakce nulová, v případě interakcí s orbitálními momenty elektronů nenulová, závislá na rozložení elektronové hustoty v okolí pozorovaného jádra a popsaná relativním číslem zvaným *chemický posun*. Jedním z projevů relaxace je nukleární Overhauserův jev, způsobený dipól-dipólovou interakcí, který slouží k měření vzdáleností mezi jádry.

<sup>6</sup> V praxi se, zvláště pro velké molekuly, používá trochu jiné uspořádání. Průběžně se z rovnováhy vychylují obě jádra, ale každé s jinou frekvencí, danou jeho frekvencí precese. Signál se pak zpracovává jako dvourozměrné spektrum.





Obrázek 14.7: Interakce pozorovaného magnetického momentu s magnetickým dipólovým momentem sousedního jádra. Pozorované jádro je znázorněno azurovou barvou, sousední jádro zeleně. Vnější homogenní magnetické pole je znázorněno tlustou fialovou šipkou v místě azurového jádra, magnetické pole sousedního jádra je znázorněno zelenými magnetickými indukčními čarami (jejich hustota a tečna v daném místě udávají velikost a směr vektoru magnetické indukce, malé šipky vyznačují orientaci). Čtyři obrázky zachycují čtyři orientace molekuly pohybující se Brownovým pohybem v roztoku,  $\theta$  je úhel mezi černou spojnicí jader (nemusí mezi nimi být kovalentní vazba) a fialovým směrem indukce vnějšího magnetického pole. Azurové jádro se během rotace postupně dostává z oblasti, ve které pole zeleného jádra snižuje vnější magnetické pole (zelené šipky mířící dolů), do oblasti, kde pole zeleného jádra vnější pole zvyšuje (zelená šipka mířící nahoru).

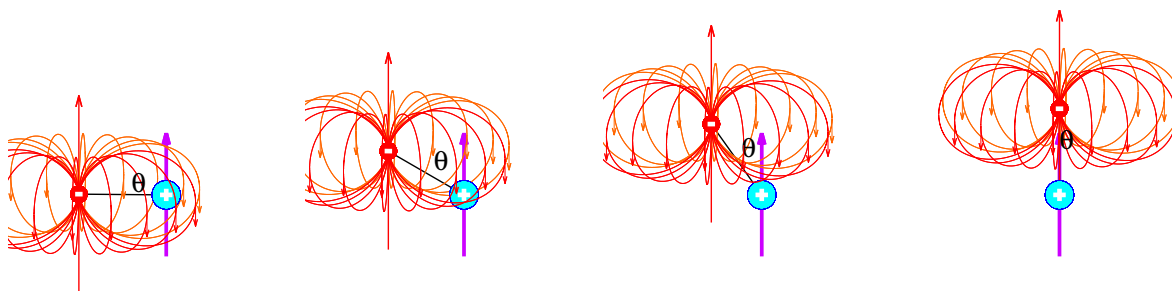
## 14.7 Vliv sousedních jader zprostředkovaný elektrony vazeb

Na první pohled se zdá, že v diskusi interakcí s orbitálním magnetickým momentem elektronů a vlastními magnetickými momenty okolních jader jsme vyčerpali všechna možná pole v okolí pozorovaných magnetických momentů. Přesto jsme se nezmínili o jednom zajímavém a důležitém druhu interakcí, nazývaném *J*-interakce (*J-coupling*). Zatímco jsme dosud dávali vždy přednost klasickému popisu velkých souborů magnetických momentů, v případě *J*-interakce vyjdeme z mikroskopického popisu dvojice interagujících jader. Vlastně začneme ještě jednodušším případem, interakcí magnetického momentu jádra s vlastním magnetickým momentem elektronu (obrázek 14.8). Ponecháme přitom stranou, že ve skutečnosti není poloha elektronu přesně lokalizována, že náš popis je klasický, ne kvantový. Jak lze čekat, tato interakce se velmi podobá dipól-dipólové interakci s magnetickým momentem jiného jádra (obrázek 14.8), která se v izotropním prostředí zprůměruje k nule. S jednou důležitou výjimkou.

V jednom případě magnetické pole v místě pozorovaného jádra nezávisí na vzájemné orientaci jádra a elektronu (a nemá tedy smysl počítat průměr přes všechny orientace). Jde o případ, kdy se elektron vyskytuje přesně v místě jádra (obrázek 14.9A). V našem klasickém obrázku to samozřejmě vypadá jako nesmysl, ale jádro a elektron nejsou kulečnickové koule. Kvantová mechanika mluví pouze o pravděpodobnosti výskytu elektronu v určitém místě a tuto pravděpodobnost popisují orbitály.<sup>7</sup> V nejjednodušším případě atomu vodíku se elektron v základním stavu nachází v orbitalu *s*, a ten má v místě jádra nenulovou hodnotu (obrázek 14.9B). Elektron se tedy očividně může s určitou (nenulovou) pravděpodobností nacházet v místě jádra. V tomto případě má magnetické pole elektronu v místě jádra stejný směr jako magnetický moment elektronu (obrázek 14.9A). Energetický příspěvek této interakce je proto přímo úměrný  $-\vec{\mu}_{\text{elektron}} \cdot \vec{\mu}_{\text{jádro}} = -|\mu_{\text{elektron}}||\mu_{\text{jádro}}| \cos \alpha$ , kde  $\alpha$  je úhel mezi směry magnetických momentů jádra (na obrázku 14.9A nevyznačený) a elektronu. Tato interakce je nazývána *Fermiho kontaktní* a tvoří většinou nejdůležitější část *J*-interakce mezi dvěma jádry.

Poloklasický popis *J*-interakce je ukázán na obrázku 14.9C,D. Nakreslený elipsoid znázorňuje vazebný orbital  $\sigma$  obsazený dvěma elektrony, které podle Pauliho principu musí mít opačný spin. Umístění elektronů do určitých bodů na obrázku 14.9C je z kvantového pohledu chybné, můžeme pouze hovořit o pravděpodobnosti, že se elektron v daném místě vyskytuje. A tvar orbitalu napovídá, že s nenulovou

<sup>7</sup>Přesněji řečeno, druhé mocniny amplitud vlnových funkcí, které jednotlivé orbitály představují.



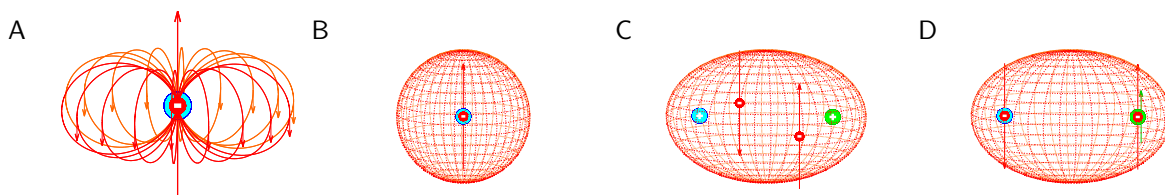
Obrázek 14.8: Interakce pozorovaného magnetického momentu s magnetickým dipólovým momentem elektronu. Pozorované jádro je znázorněno azurovou barvou, elektron červeně. Vnější homogenní magnetické pole je znázorněno tlustou fialovou šipkou v místě azurového jádra, magnetické pole vlastního magnetického momentu elektronu je znázorněno červenými magnetickými indukčními čarami (jejich hustota a tečna v daném místě udávají velikost a směr vektoru magnetické indukce, malé šipky vyznačují směr). Čtyři obrázky zachycují čtyři vzájemné polohy elektronu vůči jádru,  $\theta$  je úhel mezi černou spojnicí jádra a elektronu a fialovým směrem indukce vnějšího magnetického pole. Azurové jádro se během rotace postupně dostává z oblasti, ve které pole elektronu snižuje vnější magnetické pole (červené šipky mířící dolů), do oblasti, kde pole elektronu vnější pole zvyšuje (červená šipka mířící nahoru).

pravděpodobností se oba elektrony mohou vyskytovat v místě jader, jak je symbolicky naznačeno na obrázku 14.9D. Stejně jako u interakcí diskutovaných výše nás zajímá, jak změni  $J$ -interakce magnetické pole v místě pozorovaného jádra. Pro jednoduchost si představme, že magnetický moment sousedního jádra míří vzhůru (zelená šipka na obrázku 14.9D). Z pohledu Fermiho kontaktní interakce je energeticky nejvýhodnější stejná orientace magnetického momentu jádra a elektronu (který se právě vyskytuje v místě sousedního jádra). Protože elektrony mají opačný spin a tedy i opačný magnetický moment, tak nejvýhodnější orientace magnetického momentu elektronu v místě pozorovaného jádra bude dolů. Tento elektron tedy bude pole v místě pozorovaného jádra snižovat (v místě, kde se nachází elektron, má jeho pole stejný směr jako jeho magnetický moment).

Shrňme si, co jsme zjistili.  $J$ -interakce je vzájemné působení magnetických momentů jader zprostředkované elektrony vazeb. Protože jde zejména o Fermiho kontaktní interakce, tak příspěvek v magnetickém poli v místě jádra nezávisí na orientaci molekuly v magnetickém poli. Proto se také tato interakce někdy nazývá *skalární*, nezávislá na směru. Pokud míří magnetický moment sousedního jádra stejným směrem jako indukce vnějšího magnetického pole, tak  $J$ -interakce celkové pole v místě pozorovaného jádra snižuje. Pokud by naopak magnetický moment sousedního jádra mířil opačným směrem než vnější pole, tak by  $J$ -interakce celkové pole v místě pozorovaného jádra zvýšila. Úvahu lze rozšířit na jádra spojená několika vazbami.<sup>8</sup> Jádro s magnetickým momentem orientovaným ve směru vnějšího pole prostřednictvím elektronů vazeb bude snižovat magnetické pole v místě jader od kterých jej dělí sudý počet kovalentních vazeb a zvyšovat magnetické pole v místě jader od kterých jej dělí lichý počet kovalentních vazeb. V případě tří (a více) vazeb hraje zásadní roli také prostorové rozložení elektronové hustoty, tedy torzní úhel. Interakce je nejsilnější pro torzní úhly odpovídající uspořádání s interagujícími jádry v poloze *trans*, nejslabší interakce odpovídají torzním úhlům  $\pm 90^\circ$ . Graf závislosti interakční konstanty mezi jádry spojenými třemi vazbami ( $^3J$ ) na torzním úhlu se nazývá *Karplusova křivka*. Příklad si ukážeme v části 14.13.

Obdobně závisí  $J$ -interakce přes dvě vazby na vazebném úhlu a  $J$ -interakce přes jednu vazbu na

<sup>8</sup>  $J$ -interakce je pozorovatelná mezi jádry spojenými dvěma, třemi, výjimečně až pěti kovalentními vazbami.



Obrázek 14.9: A, magnetické pole vlastního magnetického momentu elektronu nacházejícího se v místě jádra. B, orbital s atomu vodíku má v místě jádra nenulovou hodnotu. C, vlastní magnetické momenty dvou elektronů v téměř vazebném orbitalu mají opačný spin a tedy i magnetický moment. D, energeticky nejvýhodnější orientace magnetického momentu jádra je souběžná s magnetickým momentem elektronu nacházejícího se v místě jádra. V důsledku toho je energeticky výhodná opačná orientace magnetických momentů dvou jader spojených jednou kovalentní vazbou, tvořenou dvěma elektrony s opačným magnetickým momentem. Pozorované jádro je azurové, sousední jádro a jeho magnetický moment jsou znázorněny zeleně, ostatní barvy jsou použity jako na obrázku 14.8.

délce vazby. Vazebné délky a úhly jsou ale téměř stejné ve všech konformacích. Zdálo by se tedy, že tyto  $J$ -interakce jsou nezajímavé. Jejich neměnnost je ale velmi užitečná. Můžeme jich totiž využít ke *korelaci* interagujících jader, která bude *nezávislá na konformaci*. Užitečnost takových korelací si ukážeme v sekci 14.11.

Nyní bychom měli samozřejmě postoupit od popisu jedné dvojice jader ke statistickému popisu miliard miliard takových dvojic ve skutečném vzorku. Takový popis je ale velmi složitý. Proto se spokojme s tím, že si řekneme, že při měření dvou jader citících vzájemnou  $J$ -interakci pozorujeme pro každé jádro dvě precesní frekvence lišící se o hodnotu interakční konstanty označované  $J$ . Tyto frekvence se při pohybu molekul nemění, pokud nedochází ke změnám konformace v místě jader (vazebných délek, úhlů a zejména torzních úhlů). Rozdíl precesních frekvencí způsobený  $J$ -interakcí také nezávisí na vnějším magnetickém poli, hodnota rozdílu vyjádřená v Hz je stejná ve všech spektrometrech. Pokud je jádro ovlivněno  $J$ -interakcemi s více okolními jádry, vlivy těchto interakcí se kombinují.

## 14.8 Spektrometr a experiment NMR

Jak vypadá spektrometr, tedy zařízení, které nám umožní magnetické momenty ovlivňovat tak, jak jsme si popsali, a měřit jejich frekvenci? Nezbytnou součástí spektrometru je *silný magnet*, vytvářející velmi homogenní magnetické pole o indukci  $\vec{B}_0$ . Jako zdroj slabšího oscilujícího magnetického pole  $\vec{B}_1$  využíváme magnetickou složku elektromagnetických vln. Stejně jako v případě optické spektroskopie tedy vystavujeme jádra elektromagnetickému vlnění. Frekvence tohoto vlnění musí rezonovat s frekvencí precese magnetického momentu jádra, které pozorujeme. Proto hovoříme o nukleární magnetické *rezonanci*. Energetické rozdíly mezi orientacemi magnetických momentů jader jsou velmi malé, proto je i frekvence elektromagnetického pole v NMR mnohem nižší (v oblasti velmi krátkých radiových vln, řádově  $10^8$  Hz), než frekvence záření v optické spektroskopii. Vidíme, že další důležitou částí spektrometru NMR je *radiový vysílač*.

Nejjednodušší experiment NMR vypadá tak, že vzorek ozařujeme radiovými vlnami po dobu, během které se sklopí vektor magnetizace  $\vec{M}$  z původního směru rovnoběžného s  $\vec{B}_0$  do roviny kolmé na

$\vec{B}_0$ .<sup>9</sup> Po vypnutí elektromagnetického záření bude  $\vec{M}$  v této rovině rotovat kolem  $\vec{B}_0$ . Co to znamená z praktického pohledu měření? Zatímco na začátku, v rovnovážném stavu, směřoval vektor nehybně jedním směrem, po ozáření po určitou dobu  $\vec{M}$  rotuje. Jádra sama se nám tak stala zdroji proměnného elektromagnetického pole, tedy *vysílači elektromagnetického záření*. Přesněji řečeno, jádra byla vždycky zdroji proměnných elektromagnetických polí, ale tato pole se navzájem rušila. Ozářením radiovou vlnou jsme ale soubor jader ovlivnili takovým způsobem, že se jejich proměnná pole počítají a tvoří měřitelné elektromagnetické záření.

Toto záření má frekvenci shodnou s precesní frekvencí magnetických momentů jader a můžeme je zachytit *detektorem* v principu podobným *rozhlasovému přijímači*.<sup>10</sup> Precesní frekvence stejných jader v různých místech molekuly se přitom mírně liší v důsledku interakcí s elektrony a ostatními jádry, které jsme si podrobně popsali v části 14.4. Signál zachycený radiovým přijímačem tedy nemá jednu frekvenci, ale skládá se z množství vln s mírně odlišnými frekvencemi. Podobně vypadá signál ve frekvenčně modulovaném rozhlasovém vysílání na velmi krátkých vlnách. V rozhlasovém studiu se zachytí akustické vlny hlasu zpěváka či redaktora mikrofonom, převedou na elektrické oscilace se stejnou frekvencí (v rozsahu hertzů až kilohertzů) a smísí s vysokofrekvenční radiovou vlnou (o frekvenci kolem 100 MHz). Vzniklá vysokofrekvenční radiová vlna je tak *modulovaná* nízkofrekvenčním *audio-signálem*. Tato vlna je zesílena, vysílací anténou šířena do prostoru a posléze zachycena anténou vašeho rozhlasového přijímače. V něm se z modulované radiové vlny odstraní vysokofrekvenční složka,<sup>11</sup> a zesílená nízkofrekvenční audio-složka rozhlasového signálu je poslána do reproduktoru, který ji převede na zvukové vlnění věrně kopírující akustické vlnění zachycené ve studiu.

Přijímač spektrometru funguje úplně obdobně. Pro jednoduchost si můžeme představit, že jsme jádra ozářili radiovou vlnou se stejnou frekvencí, jako je precesní frekvence srovnávací látky, pomocí které definujeme nulu na škále chemického posunu. Chemický posun pak hraje stejnou roli jako frekvence hlasu zpěváka v rozhlasovém vysílání. Po vypnutí vysílače zachytí anténa přijímače signál generovaný magnetickými momenty jader, který je modulovaný chemickým posunem. Ze zesíleného signálu jader přijímač spektrometru odstraní vysokofrekvenční složku, složku o stejné frekvenci, se kterou jsme jádra ozářovali. Ze signálu zbyde směs vln s frekvencemi odpovídající chemickým posunům jader v různých místech molekuly. Takový výsledek měření signálu protonů je ukázán na obrázku 14.10A pro velmi jednoduchou molekulu, pro acetaldehyd ( $\text{CH}_3\text{—CH=O}$ ) se čtyřmi atomy vodíku ve dvou chemických skupinách, třemi v methylové skupině a jedním v aldehydové.

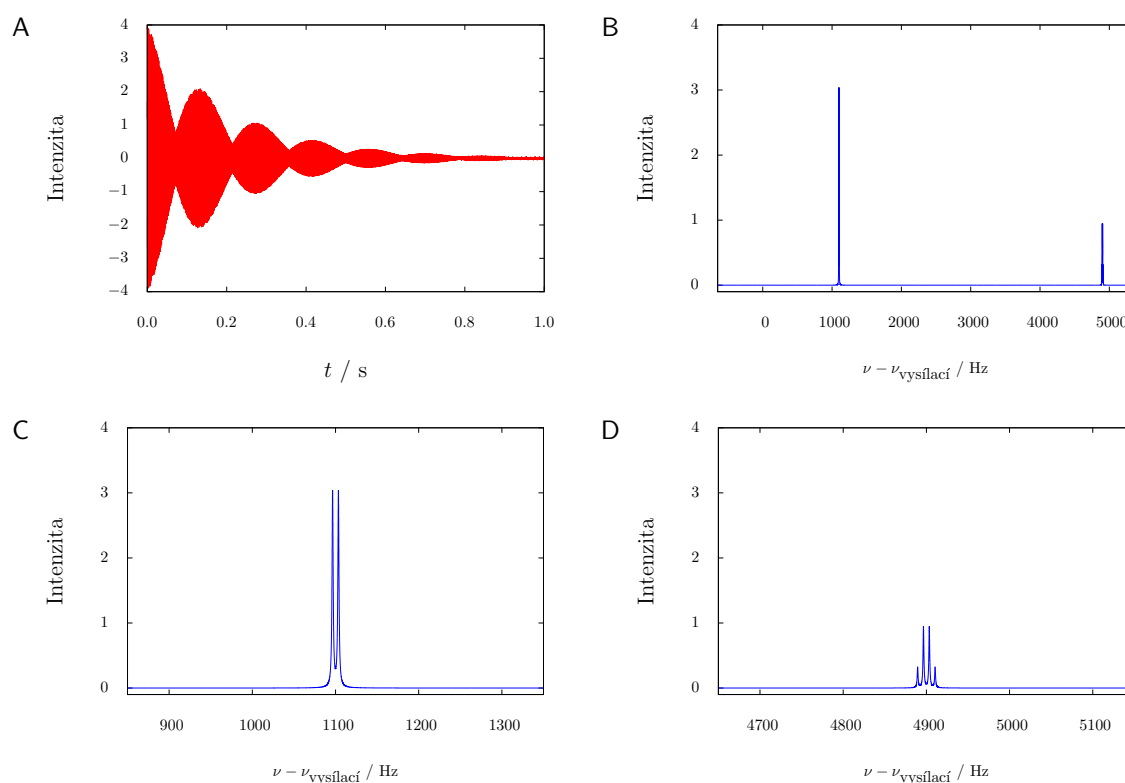
Obrázek 14.10A pěkně ukazuje, že odečtení frekvencí přímo ze záznamu měření signálu je i pro nejjednodušší molekuly prakticky nemožné. Proto záznam převádíme do digitální formy a dále zpracováváme. Nejčastěji ke zpracování používáme Fourierovu transformaci. Na rozdíl od rentgenové krystalografie pracujeme se závislostí měřeného signálu na čase. Výsledkem Fourierovy transformace je v případě NMR *spektrum* intenzity signálu pro různé frekvence. Tato intenzita je blízká nule pro frekvence, které se v signálu nevyskytovaly, ale vysoká pro hodnoty blízké frekvencím precesního pohybu magnetických momentů v měřeném vzorku. Spektrum acetaldehydu je ukázáno na obrázku 14.10B. Při pohledu na spektrum okamžitě vidíme, že trojice protonů methylové skupiny koná precesní pohyb s frekvencí asi o 1 000 Hz vyšší, než je frekvence radiových vln vysílače, zatímco frekvence aldehydového protonu je téměř o 5 000 vyšší než vysílací frekvence.

Obrázky 14.10C,D nám připomínají, že jsme při povídání o modulaci chemickým posunem zanedbali, že magnetické momenty protonů interagují také mezi sebou prostřednictvím elektronů tří kovalentních vazeb, které leží mezi methylovými vodíky a aldehydovým vodíkem. Ve skutečnosti pozorujeme pro

<sup>9</sup>V běžných spektrometrech je tato doba dlouhá přibližně 10  $\mu\text{s}$ , závisí na konkrétním spektrometru, nastavení výkonu vysílače i složení vzorku. Při studiu proteinů a nukleových kyselin se snažíme získat co nejsilnější signál, tedy sklopit magnetizaci přesně do kolmého směru. Proto vždy před vlastním experimentem měříme, jak přesně je dlouhá doba potřebná ke sklopení magnetizace o 90°, ze svislého do vodorovného směru. Tomuto pomocnému měření říkáme *kalibrace 90° pulzu* radiových vln.

<sup>10</sup>Přirovnání k radiu je vskutku na místě. Ve spektrometrech s magnetem o indukci 9 až 12 tesla vysílají jádra uhlíku na stejné frekvenci jako vaše oblíbená FM stanice.

<sup>11</sup>Na tuto vysokofrekvenční složku je přijímač naladěn.

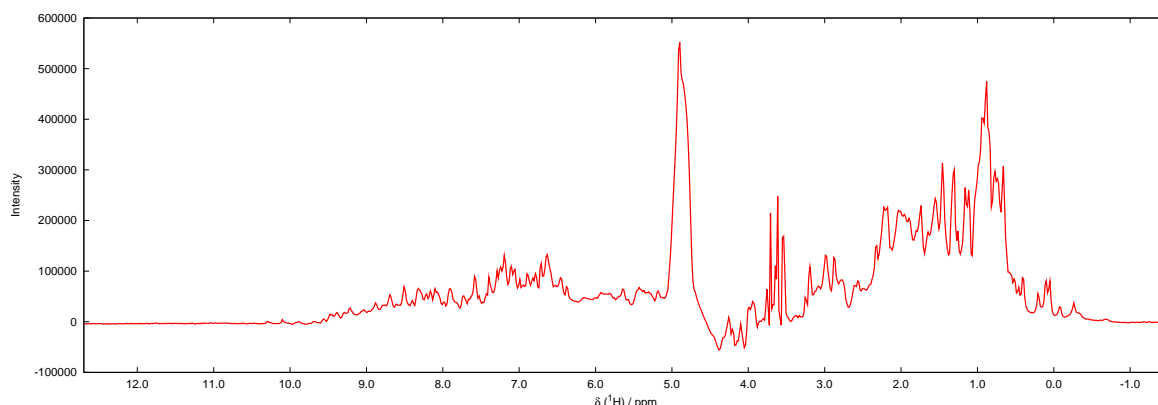


Obrázek 14.10: A, časový záznam signálu magnetických momentů protonů acetaldehydu po odstranění vysokofrekvenční složky. B, spektrum získané Fourierovou transformací tohoto signálu. C, detail části spektra okolo precesní frekvence protonů v methylové skupině. D, detail části spektra okolo precesní frekvence protonů v aldehydové skupině. Signál byl nasimulován pro magnetické pole o indukci 11,75 T a pro frekvenci radiové vlny vysílače odpovídající precesní frekvenci tetramethylsilanu.

methylové protony signály o dvou frekvencích lišících se v důsledku interakce s aldehydovým vodíkem. Podobně signál aldehydového protonu je rozštěpen na čtyři signály v důsledku interakce s protony methylové skupiny.

## 14.9 Vícerozměrná spektroskopie a korelace frekvencí

V částech 14.2 a 14.3 jsme si popsali princip moderního spektrometru, který používá krátké ozařovací pulzy elektromagnetického záření. Při studiu malých molekul nám často stačí jednoduchý experiment s jedním ozařovacím pulzem, který sklopí magnetizaci pozorovaných jader do vodorovné roviny. Po nějakou dobu (přibližně sekundu) magnetizace každého jádra v této rovině rotuje s precesní frekvencí magnetických momentů, které ji tvoří. Během této doby můžeme měřit signál indukovaný v přijímací anténě, než náhodné interakce s okolím přivedou magnetické momenty zpět do rovnováhy. V případě biomakromolekul s takovým experimentem nevystačíme, vzniklá spektra obsahují příliš mnoho frekvencí, které se často překrývají, a spektrum není možné vyhodnotit (obrázek 14.11). Naštěstí ozařovací pulzy umožňují nejen počáteční sklopení magnetizace, ale i další manipulace s magnetickými momenty jader v průběhu experimentu. Tyto manipulace mohou být poměrně složité. Jejich cílem je vytvořit v molekule takové stavy, při kterých se jaderné magnetické momenty nepohybují nezávisle jeden na druhém, ale jsou



Obrázek 14.11: Příklad jednorozměrného NMR spektra proteinu (myší protein MUP I).

určitým způsobem provázány. Toto provázání mimo jiné umožňuje, že můžeme v jednom experimentu měřit více frekvencí. Mluvíme pak o *vícerozměrných spektrech*.

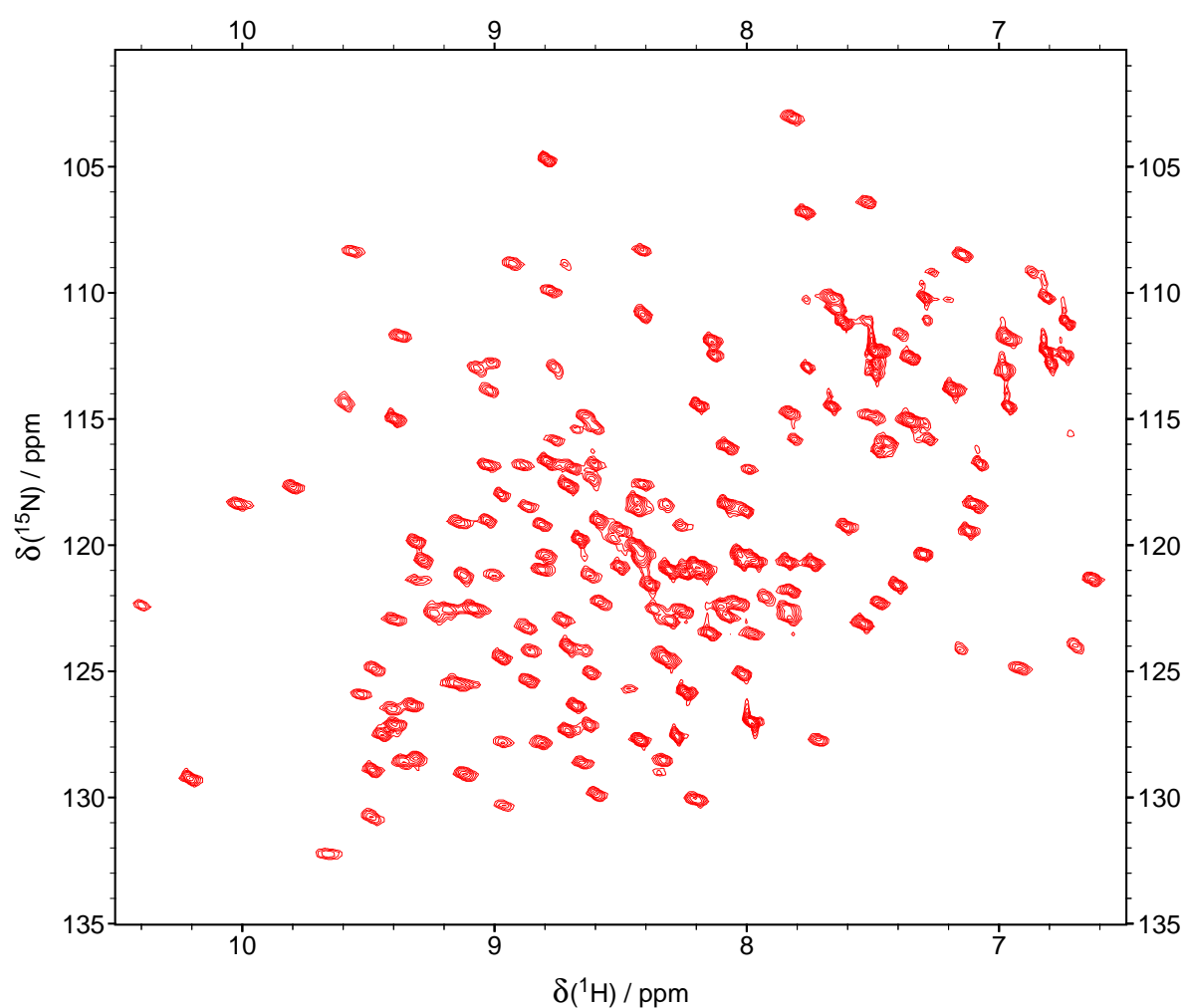
Vícerozměrná spektra neobsahují jen informaci o rezonančních frekvencích jednotlivých jader, ale nesou v sobě i údaj o tom, která jádra mají navzájem provázané pohyby magnetických momentů. Říkáme, že mezi takovými jádry existuje *korelace*. Frekvence, které ve vícerozměrných spektrech koreluje, mohou být frekvence magnetických momentů stejného izotopu (nejčastěji  $^1\text{H}$ ) v různých místech molekuly. Pak mluvíme o *homonukleární korelaci*. Může jít ale o frekvence jader různých prvků. Můžeme například korelovat frekvenci jádra  $^1\text{H}$  s frekvencí  $^{13}\text{C}$ ,  $^{15}\text{N}$ , nebo  $^{31}\text{P}$ . V takovém případě, nazývaném *heteronukleární korelace*, zpravidla potřebujeme, aby tyto izotopy byly přítomny v téměř všech jádrech daného typu (například, aby  $^{13}\text{C}$  tvořil asi jedno procento atomů uhlíku, jak v přírodě, ale aby alespoň 95 % uhlíku mělo jádro  $^{13}\text{C}$ ). Můžeme proto říci, že studium proteinů pomocí NMR bylo umožněno rozvojem dvou metod. První z nich je exprese rekombinantních proteinů umožňující značení proteinů izotopy  $^{13}\text{C}$ ,  $^{15}\text{N}$ ,  $^2\text{H}$ . Druhou je technika NMR spektroskopie ve více dimenzích a při více frekvencích, která nám umožňuje heteronukleární korelace dosáhnout.

Příklad dvourozměrného spektra s heteronukleární korelací je ukázán na obrázku 14.12. V tomto spektru jsou provázány (korelovány) frekvence jader  $^1\text{H}$  a  $^{15}\text{N}$ . Použitá korelace je založena na interakcích mezi magnetickými momenty  $^1\text{H}$  a  $^{15}\text{N}$ , a to interakcích zprostředkovaných elektrony tvořící kovalentní vazbu mezi vodíkem a dusíkem.<sup>12</sup> Ve srovnání s měřením jednorozměrných spekter precesních frekvencí  $^1\text{H}$  a  $^{15}\text{N}$  má tento *dvojrozměrný experiment* několik výhod:

1. Frekvence  $^{15}\text{N}$  měříme s mnohem větší citlivostí, než v jednorozměrném experimentu, který začíná vychýlením magnetizace  $^{15}\text{N}$ . Je tomu tak proto, že v našem dvojrozměrném experimentu začínáme sklopením magnetizace  $^1\text{H}$  a signál  $^1\text{H}$  také měříme. Precesní frekvenci  $^{15}\text{N}$  měříme pouze nepřímo, experiment je navržen tak, že signál  $^1\text{H}$  je *modulován* frekvencí magnetizace  $^{15}\text{N}$ , která se uprostřed experimentu ocitá ve vodorovné rovině. Tak je signál měřen s citlivostí odpovídající velikosti celkové magnetizace  $^1\text{H}$ , která je mnohem větší, než magnetizace  $^{15}\text{N}$ .
2. Ve spektru nevidíme frekvence všech jader vodíků (těch je v molekule zkoumaného proteinu několik tisíc), ale pouze frekvence magnetických momentů protonů vázaných přímo na dusík (těch je jenom přibližně dvě stě). Proto se frekvence jader v různých aminokyselinách a různých chemických skupinách aminokyselin mnohem méně překrývají.

<sup>12</sup>Korelace se dosáhne tím, že se polarizace magnetických momentů protonů přemění na polarizaci magnetických momentů  $^{15}\text{N}$ . Technicky toho docílíme pomocí pulzů oddělených prodlevami, během kterých interagují magnetické momenty jader  $^1\text{H}$  a  $^{15}\text{N}$  prostřednictvím elektronů ve vazbě N–H.

3. Tím, že jsou signály rozprostřeny ve dvou dimenzích, jsou mnohem lépe rozlišeny. Pokud jsou shodou okolností téměř stejné frekvence dvou protonů, většinou jsou dostatečně odlišné frekvence dusíků  $^{15}\text{N}$ , na které jsou tyto protony navázány. Při pečlivé prohlídce spektra jsme tak schopni rozlišit dvojice frekvencí téměř všech NH skupin v proteinu.
4. Asi nejdůležitější výhodou tohoto experimentu je, že nese *chemickou informaci*: říká nám, který vodík je navázán na který dusík. V samotném spektru jsou jednotlivé atomy popsány jejich frekvencemi. Abychom poznali, která frekvence odpovídá kterému atomu, musíme provést analýzu zvanou *přiřazení frekvencí atomům*, kterou si popíšeme v části 14.11.



Obrázek 14.12: Příklad dvojrozměrného NMR spektra proteinu (myší protein MUP I), poskytujícího korelace  $^1\text{H}$  a  $^{15}\text{N}$ . Spektrum je ukázáno jako vrstevnicová mapa, na svislé a vodorovné ose jsou vyneseny přesné frekvence  $^{15}\text{N}$  a  $^1\text{H}$ , vrstevnice popisují intenzitu signálu.

## 14.10 Postup strukturní analýzy proteinů pomocí NMR

Zatím jsme se snažili pochopit co to NMR je a jak je možné získat spektra NMR. Změňme teď úhel pohledu a podívejme se na to, co nám NMR spektroskopie může říci o molekulách, které studujeme.

Interakce jader se svým okolím mohou být využity k získání informací o struktuře molekuly. Obecná strategie určování struktury pomocí NMR se dá rozdělit na dva druhy experimentů.

První typ experimentů slouží k *přiřazení naměřených frekvencí odpovídajícím atomům*. Znamená to, že vyjdeme ze známé *konfigurace* biomakromolekuly a, obrazně řečeno, na co nejvíce atomů nalepíme štítek, na kterém bude napsána hodnota frekvencí. V těchto experimentech zpravidla měříme dvourozměrná a trojrozměrná spektra, která *korelují* frekvence jednotlivých jader a umožňují tak jejich přiřazení. Je-li například v trojrozměrném spektru každý signál definován trojicí frekvencí  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ , koreluje takové spektrum trojice atomů H, C, N, které k sobě nějakým způsobem „patří“. Podle druhu experimentu může „patřit“ znamenat například to, že se dané atomy vyskytují v téže skupině H—C—N. Výsledkem prvního typu experimentů je tedy tabulka udávající rezonanční frekvence jednotlivých jader v molekule. Každé jádro v tabulce je tak označeno jeho rezonanční frekvencí.

Druhý typ experimentů opět využívá korelací, ale tentokrát již vztahujících se ne k topologii, ale k prostorové struktuře. Například korelace dvou protonů v experimentu sledujícím nukleární Overhauserův jev mezi jádry znamená, že dané protony jsou si blízko v prostoru. Kvantitativní měření této korelace (NOE) umožní přibližně určit vzdálenost daných jader. Výsledkem druhého typu experimentů je opět tabulka, která tentokrát shrnuje poznatky o trojrozměrné struktuře, získané z NMR experimentů. Bez využití tabulky první bychom ale věděli pouze, že například protony o frekvenci 8,73 ppm a 9,64 ppm jsou vzdáleny 0,3 až 0,4 nm. Kombinace výsledků experimentů prvního a druhého typu nám však umožní říci, že takto blízko jsou ve struktuře například vodíky peptidových vazeb osmé a dvacáté aminokyseliny studovaného proteinu.

## 14.11 Sekvenční přiřazení

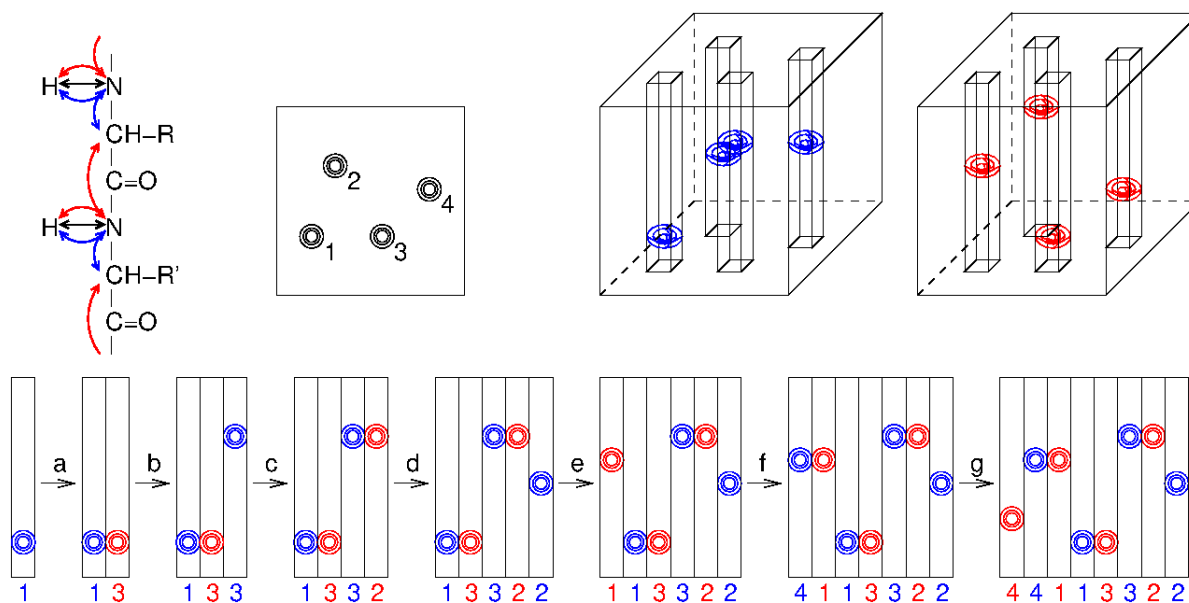
Přiřazení frekvencí atomům se obvykle provádí ve dvou krocích. Nejprve se přiřadí frekvence jader peptidové páteře (*sekvenční přiřazení*) a v druhém kroku se provede *přiřazení postranních řetězců*.

Jak jsme si již řekli, interpretace jednorozměrných protonových spekter proteinů (obrázek 14.11) je téměř nemožná ze dvou důvodů. V důsledku obrovského množství signálů je spektrum špatně rozlišené a není známé, který signál náleží kterému atomu. Připravíme-li protein značený  $^{15}\text{N}$  a měříme-li dvourozměrné spektrum s  $^1\text{H}$ - $^{15}\text{N}$  korelací, rozlišení se dramaticky zlepší rozvinutím spektra do druhé dimenze (frekvence  $^{15}\text{N}$ ) a snížením počtu signálů (spektrum obsahuje pouze  $^1\text{H}$ - $^{15}\text{N}$  korelace, tedy jen protony vázané na dusík). Jdeme-li dále a označíme-li protein  $^{15}\text{N}$  a  $^{13}\text{C}$ , můžeme měřit spektra trojrozměrná, která nejen že budou ještě lépe rozlišená, ale umožní i určit, které aminokyselině daný signál náleží. Jinými slovy, heteronukleární korelaci (korelaci mezi jádry různého typu, která jsou spojena kovalentní vazbou) můžeme použít pro sekvenční přiřazení.

Postup sekvenčního přiřazení si můžeme ukázat na příkladu krátkého peptidu, složeného z pěti aminokyselin. Obrázek 14.13 ukazuje jedno dvojrozměrné spektrum takového peptidu, které poskytuje korelaci  $^1\text{H}$ - $^{15}\text{N}$ , a dvě trojrozměrná spektra, označovaná HNCA a HN(CO)CA.<sup>13</sup> Pokud pentapeptid neobsahuje NH skupiny v postranních řetězcích, uvidíme v dvojrozměrném spektru čtyři signály (měřitelné budou pouze signály NH skupiny v peptidové vazbě, koncová aminoskupina signály neposkytne, protože její protony snadno disociují a během měření se rychle vyměňují s protony vody).

<sup>13</sup>NMR experimenty používané pro studium biomakromolekul se často označují podle jader, která jsou v nich korelována. Například zkratka HN(CO)CA nám říká, že jsou korelovány vodík, dusík a uhlík  $\alpha$  (CA). Uhlík  $\alpha$  je přítom korelován prostřednictvím karbonylového uhlíku (CO). V experimentu se polarizace magnetických momentů  $^{15}\text{N}$  převádí na polarizaci magnetických momentů jádra  $^{13}\text{C}$  v karbonylu a dále na polarizaci magnetických momentů  $^{13}\text{C}^\alpha$ . Frekvence karbonylového uhlíku ale není měřena, proto je CO v závorce.





Obrázek 14.13: Příklad sekvenčního přiřazení pomocí experimentů HNCA a HN(CO)CA. Nahoře je schema korelace v jednotlivých experimentech (černě korelace  $^1\text{H}$  a  $^{15}\text{N}$ , modře korelace v HNCA, červeně korelace v HN(CO)CA) a naměřená spektra. Dole je postup sekvenčního přiřazení.

Trojrozměrná spektra rozšiřují korelaci na uhlíky  $\alpha$ . Ve spektru HNCA (na obrázku 14.13 modře) jsou korelovány protony a dusíky NH skupiny (v peptidové vazbě) aminokyseliny číslo  $k$  s  $\alpha$ -uhlíkem téže  $k$ -té aminokyseliny. Naopak ve spektru HN(CO)CA (na obrázku 14.13 červeně) jsou korelovány protony a dusíky NH skupiny aminokyseliny číslo  $k$  opět s  $\alpha$ -uhlíkem, ale tentokrát aminokyseliny číslo  $k - 1$  (tedy prostřednictvím karbonylového uhlíku). Všimněte si, že promítnutím spektra HNCA nebo HN(CO)CA získáme stejnou korelaci, jako vidíme v dvojrozměrném spektru. Jinými slovy, signály v trojrozměrných spektrech leží vždycky v „komínku“ stojícím nad místem signálu v dvojrozměrném spektru („komínky“ jsou také znázorněny na obrázku 14.13).

Na začátku vyhodnocování spekter nevíme, který signál odpovídá které aminokyselině. Proto si signály ve spektrech očíslováme libovolně (ale stejně ve všech spektrech). Příklad číslování je na obrázku 14.13 uveden v dvojrozměrném spektru. Další postup je ukázán v dolní části obrázku 14.13. Vybereme si jeden signál v HNCA spektru, například signál označený 1 v  $^1\text{H}$ - $^{15}\text{N}$  korelaci. Pro tento signál si vyřízneme z HNCA spektra proužek, který představuje řez prostředkem komínku. V této chvíli můžeme říci pouze to, jaká frekvence jader  $\alpha$ -uhlíku odpovídá NH skupině popsané dvojicí frekvencí  $^1\text{H}$  a  $^{15}\text{N}$ , kterou jsme si označili číslem 1. Pak se snažíme nalézt v HN(CO)CA spektru signál se *stejnou* frekvencí  $^{13}\text{C}$ , odpovídající téměř  $\alpha$ -uhlíku (krok „a“ v dolní části obrázku 14.13). V HN(CO)CA spektru samozřejmě najdeme korelaci této  $^{13}\text{C}$ -frekvence s jinou NH skupinou, a to s NH skupinou *následující* aminokyseliny. V našem případě jde o NH skupinu označenou číslem 3. Zprostředkovaně přes frekvenci  $\alpha$ -uhlíku se tak dozvíme, které dvě NH skupiny následují v sekvenci po sobě. V kroku „b“ přiložíme k dvojici proužků proužek z HNCA spektra, který odpovídá stejné NH skupině jako proužek z HN(CO)CA spektra (číslo tři). Krok „c“ je opakováním kroku „a“ v HN(CO)CA spektru hledáme signál se stejnou frekvencí  $^{13}\text{C}$ . V našem případě jej nalezneme v proužku 2. Pak připojíme proužek 2 z HNCA spektra (krok „d“ a hledáme v HN(CO)CA spektru signál se stejnou frekvencí  $^{13}\text{C}$ . Žádný takovýto signál ale nenalezneme. Znamená to, že sekvenční informace je zde přerušena (v našem případě proto, že jsme dorazili na konec peptidu, při přiřazování frekvencí proteinů je často sekvenční informace přerušena prolinem, který

neobsahuje NH skupinu). Pokračujeme proto opačným směrem. V kroku „e“ přiložíme k proužku 1 spektra HNCA proužek 1 spektra HN(CO)CA. V kroku „f“ hledáme, jaký signál v HNCA spektru má stejnou frekvenci  $^{13}\text{C}$  jako signál 1 ve spektru HN(CO)CA. V našem případě jde o signál číslo 4. V kroku „g“ proto přiložíme proužek 4 spektra HN(CO)CA. Teď bychom měli hledat v HNCA spektru signál se stejnou frekvencí  $^{13}\text{C}$ , žádný takový signál ale už ve spektru nenalezneme, protože jsme dorazili na začátek peptidu. Výsledek analýzy nám ukazuje, že očíslované NH skupiny v sekvenci následují v pořadí 4-1-3-2.

Srovnáním HNCA a HN(CO)CA spektra tedy porovnááme korelace téže NH skupiny s  $\alpha$  uhlíkem dvou sousedních aminokyselin. Zprostředkovaně, prostřednictvím frekvence  $\alpha$ -uhlíku, se tak dozvíme, které dvě NH skupiny následují v sekvenci po sobě. Můžeme tedy určit pořadová čísla všech NH skupin. K sekvenčnímu přiřazení se častěji využívá dvojice spekter, ve které jsou korelovány nejen  $\alpha$ , ale i  $\beta$  uhlíky (spektra HNCACB a CBCA(CO)NH<sup>14</sup>). Zahnutí  $\beta$  uhlíků snižuje možnost náhodného výskytu shodné frekvence v uhlíkové dimenzi a proto tato varianta experimentů bývá používána nejčastěji.

V reálných případech není možné nalézt sled sousedních aminokyselin podél celé délky peptidové páteře (například výskyt prolinu, který neobsahuje NH skupinu, vede k přerušení sekvence). Je ale možné jednotlivé fragmenty sekvenčně přiřazených frekvencí přiřadit úsekům primární struktury na základě hodnot chemických posunů jednotlivých jader. Využíváme přitom zejména rozdílů frekvencí uhlíku  $\beta$ , které získáme ze spekter HNCACB a CBCA(CO)NH. Například analýzou frekvencí lze rozlišit glycin (nemá žádný  $\beta$  uhlík), alanin ( $\delta \approx 17$  až  $21$  ppm), „dlouhé“ aminokyseliny E, H, K, M, P, Q, R, V, W a redukovaný cystein ( $\delta \approx 26$  až  $36$  ppm), „krátké“ aminokyseliny D, I, L, N, F, Y a cystin ( $\delta \approx 36$  až  $46$  ppm), serin a threonin ( $\delta \approx 56$  až  $66$  ppm).

## 14.12 Přiřazení postranních řetězců

Zatím jsme hovořili o korelaci mezi jádry různého druhu (heteronukleární korelace). Velmi užitečné jsou také korelace mezi jádry stejného druhu, tedy mezi jádry, která mají přibližně stejné rezonanční frekvence (homonukleární korelace). Množinu takových jader, která jsou vzájemně propojena interakcemi zprostředkovanými elektrony vazeb, nazýváme *spinový systém*. Alifatické aminokyseliny tvoří izolované spinové systémy jader  $^1\text{H}$  a  $^{13}\text{C}$ , zatímco aromatické aminokyseliny se skládají ze dvou izolovaných spinových systémů (frekvence  $^{13}\text{C}$  alifatické a aromatické části se liší natolik, že je nemůžeme považovat za stejný typ jader).

Metody NMR umožňují měřit spektra, ve kterých jsou korelována všechna jádra spinového systému (jde o takzvaný TOCSY<sup>15</sup> experiment). Protože ze sekvenčního přiřazení známe již frekvence  $\alpha$  a  $\beta$  uhlíků a  $\alpha$  a amidových protonů, homonukleární korelace v rámci aminokyseliny umožní určit, kterým atomům postranního řetězce patří jednotlivé frekvence pozorované ve spektrech. V praxi se takováto homonukleární korelace kombinuje s výše popsanou korelací heteronukleární, která umožňuje zvýšit rozlišení. Příkladem experimentu pro určování frekvencí postranních řetězců je experiment HCCH-TOCSY, ve kterém koreluje všechny uhlíky spinového systému navzájem pomocí TOCSY a každý z uhlíků navíc heteronukleárně s protony, které jsou na něj vázány. Spojení alifatických a aromatických spinových systémů aromatických aminokyselin je založeno na heteronukleárních korelacích využívajících několika jednovazebných interakcí zprostředkovaných elektrony vazeb. Například experiment (HB)CB(CGCD)HD koreluje alifatický  $\beta$  uhlík s aromatickým  $\delta$  protonem. Speciální experimenty

<sup>14</sup>Jiná zkratka tohoto spektra odráží to, že korelace jsou získány trochu jiným způsobem. Experiment HN(CO)CACB je také možné použít (a někdy nutné použít, konkrétně když jsou vodíky aminokyselin s výjimkou NH skupin označené deuteriem), ale jeho citlivost je nižší.

<sup>15</sup>Zkratka anglického *T*Otaly *C*orrelated *S*pectroscopy. Tento experiment je založen na tom, že během měření stále „opravujeme“ mírně se lišící frekvence jader v rámci spinového systému. Proto se jádra nechovají jako jádra s *přibližně* stejnou frekvencí, ale jako jádra s *přesně* stejnou frekvencí. Magnetické momenty jader s identickou frekvencí interagují tak, že ve spektrech pozorujeme korelace *všech* atomů stejného spinového systému.

jsou používány pro některé skupiny postranních řetězců (guanidinovou v argininu, indolovou v tryptofanu, imidazolovou v histidinu). Často se ale bez experimentů spojujících izolované spinové systémy téhož postranního řetězce obejdeme a atomy v takových spinových systémech přiřadíme na základě vzdáleností atomů v prostoru (viz část 14.13).

## 14.13 NMR spektra a struktura proteinů

Seznámili jsme se již s první fází studia biomarkomolekul pomocí NMR, s fází, ve které přiřazujeme frekvence jednotlivým atomům. Pojd'me se teď podívat, jak můžeme pomocí NMR určovat struktury proteinů.

V části 14.6 jsme zmínili, že přímé interakce mezi blízkými magnetickými dipólovými momenty jsou příčinou nukleárního Overhauserova efektu (NOE). Intenzitu tohoto jevu můžeme měřit pomocí experimentu zvaného NOESY (zkratka Nuclear Overhauser Effect Spectroscopy). Tento experiment využívá homonukleární korelace, která není založena na interakci magnetických momentů zprostředkované elektrony vazeb, ale na přímé interakci mezi magnetickými momenty, které jsou si blízko v prostoru. Proto intenzita signálů v NOESY spektru (příklad ukázán na obrázku 14.14) přímo odpovídá intenzitě NOE. Strmý pokles intenzity NOE s rostoucí vzdáleností jader se často využívá v určování trojrozměrné struktury proteinů a nukleových kyselin. S využitím známých vzdáleností protonů (například v  $\text{CH}_2$  skupinách nebo benzenových kruzích) můžeme NOESY spektra kalibrovat. Víme například, že vicinální protony v aromatickém kruhu (tedy protony v poloze *ortho*) jsou vzdálené 0,25 nm. Z teorie dále víme, že intenzita NOE klesá s šestou mocninou vzdálenosti. Pokud tedy ve spektru například vidíme, že signál odpovídající korelaci dvou vicinálních protonů tyrosinu číslo 47 (tedy protonů  $\text{Y47 H}^\delta$  a  $\text{Y47 H}^\epsilon$ ) má intenzitu 64 jednotek a signál mezi protony  $\alpha$  valinu 6 a alaninu 124 má intenzitu jedné jednotky, víme, že vzdálenost mezi protony  $\text{V6 H}^\alpha$  a  $\text{A124 H}^\alpha$  je dvakrát delší,<sup>16</sup> než vzdálenost protonů  $\text{Y47 H}^\delta$  a  $\text{Y47 H}^\epsilon$ . Tímto přímočarým způsobem můžeme z intenzit signálů ve spektru určit přímo meziatomové vzdálenosti (obrázek 14.15 vlevo). Na základě meziatomových vzdáleností byly určeny první struktury proteinů počítané z experimentálních dat poskytnutých NMR. Mezi různými údaji o struktuře, které NMR spektra poskytují, hrají meziatomové vzdálenosti dodnes nejdůležitější roli.

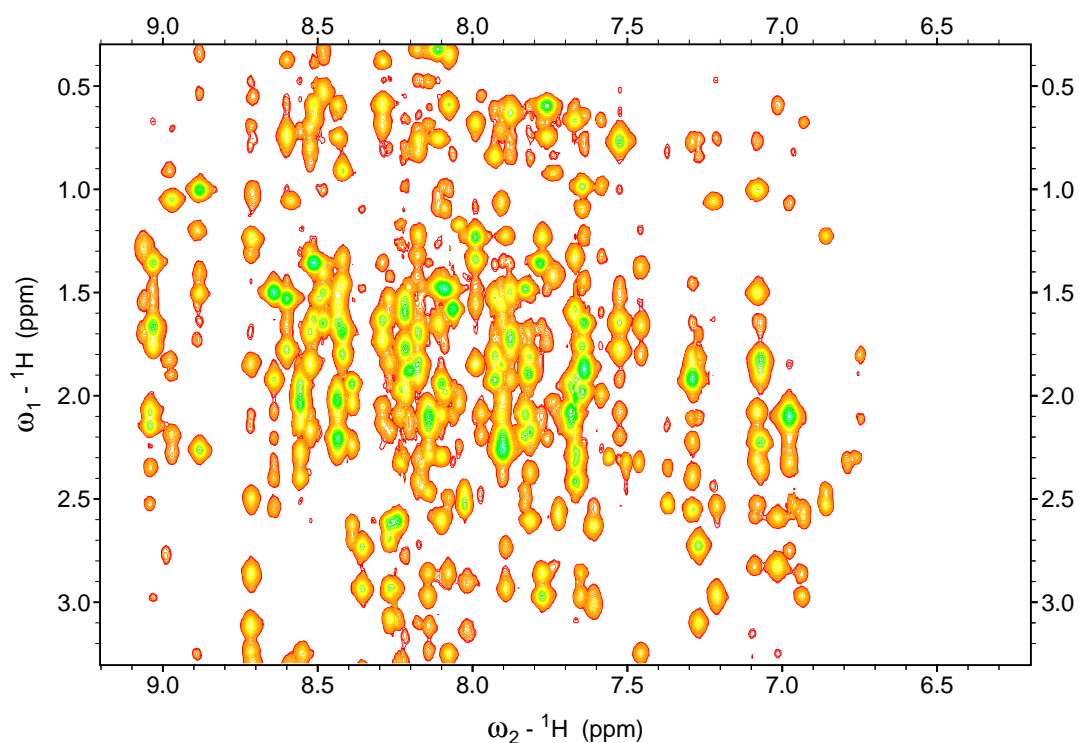
Ačkoli je NOE velmi užitečný, není jediným zdrojem informace o konformaci molekul, který NMR poskytuje. Důležitou, byť spíše doplňkovou, úlohu hrají další údaje. Jedním z nich je *J*-interakce mezi magnetickými momenty jader oddělených třemi kovalentními vazbami. Jak již zmínili v části 14.7, konstanta popisující velikost této interakce závisí na torzním úhlu prostřední vazby. Příklad závislosti několika interakčních konstant na torzním úhlu  $\phi$  je ukázán na obrázku 14.16.

Protože torzní úhly definují konformaci, je jejich určování z trívazebných *J*-interakcí velmi užitečné.<sup>17</sup> K jednoznačnému určení torzního úhlu ovšem většinou nestačí změřit jednu interakční konstantu, protože Karplusova křivka může nabývat stejné hodnoty pro čtyři různé torzní úhly. Je proto nutno měřit interakce mezi různými atomy definujícími torzi kolem téže vazby. Jak si ale ukážeme v části 14.14, i jediný typ interakční konstanty nám může hodně prozradit o sekundární struktuře.

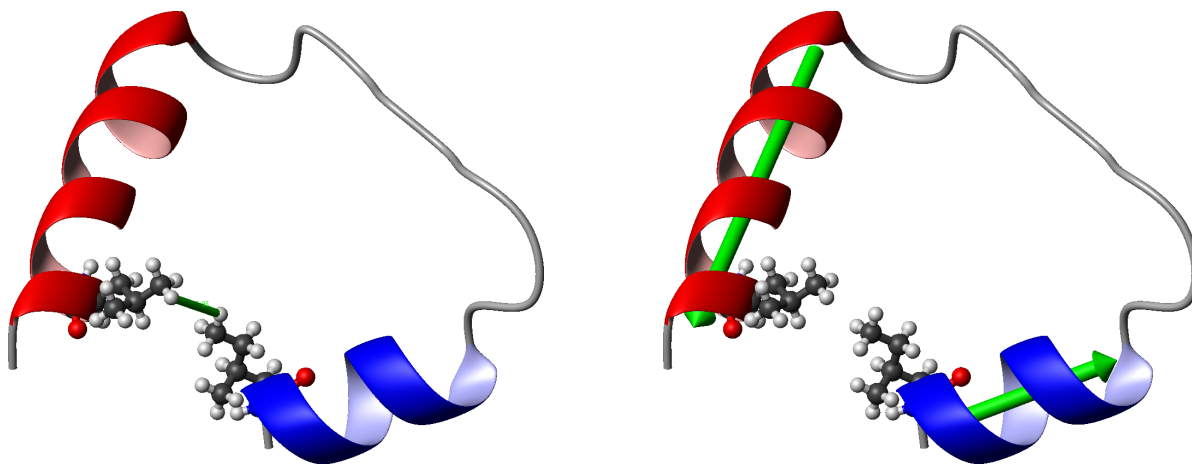
Konformaci molekul neodrážejí jen interakce mezi magnetickými momenty jader, ale i interakce s orbitálními momenty elektronů. Údaj, kterým tyto interakce v roztoku popisujeme, chemický posun, je dokonce nejdůležitější informací při strukturní analýze malých molekul založené na NMR. Chemický posun závisí na elektronové hustotě v okolí měřeného jádra. V rentgenové krystalografii či elektronové

<sup>16</sup>Dvojnásobná vzdálenost znamená  $2^6 = 64$  krát menší intenzitu.

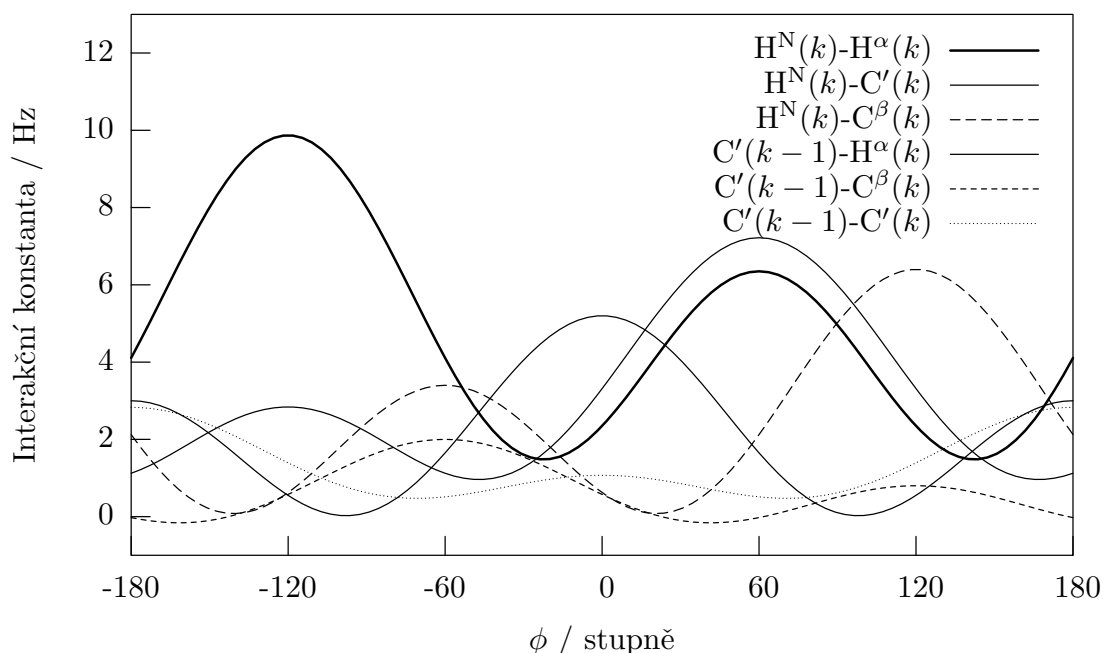
<sup>17</sup>Interakční konstanty, které intenzitu interakcí zprostředkovaných elektrony popisují, se určují buď z intenzit nebo z rozštěpení signálů ve spektrech. První postup (takzvané QJ metody) je založen na porovnání dvou spekter (nebo dvou signálů v témže spektru), z nichž v jednom se interakce uplatní a v druhém je potlačena. Druhý postup využívá nejčastěji takzvaného E.COSY principu, který umožňuje měření i velmi malých interakčních konstant. E.COSY spektra obsahují signály rozštěpené ve dvou dimenzích. Ze čtveřice signálů vzniklých štěpením se dva odečtou a zbývající dva sečtou. Pokud využijeme silné interakce v jedné dimenzi, můžeme odečíst i velmi malé štěpení (slabou interakci) v druhé dimenzi.



Obrázek 14.14: Příklad dvojrozměrného NOESY spektra proteinu, poskytujícího korelace protonů, které jsou blízko v prostoru.



Obrázek 14.15: Schematické znázornění strukturálních informací, které můžeme z NMR spekter získat pro aminokyseliny, které leží od sebe daleko v sekvenci proteinu. Vlevo je zeleně znázorněno měření vzdálenosti mezi protony methylových skupin v různých  $\alpha$ -šroubovicích. Tuto vzdálenost lze určit z hodnoty NOE mezi měřenými protony. Vpravo je zelenými šipkami ukázána vzájemná orientace dvou  $\alpha$ -šroubovic, kterou lze spočítat z hodnot zbytkových interakcí měřených v jednotlivých šroubovicích.



Obrázek 14.16: Závislosti třívazebných interakčních konstant na hodnotě torzního úhlu  $\phi$  proteinů (Karplusovy křivky).

mikroskopii je rozložení elektronové hustoty právě to, co se snažíme určit. Vypočítat rozložení elektronové hustoty biomakromolekul z chemických posunů jader je ale příliš složitý úkol. Proto při strukturální analýze proteinů a nukleových kyselin využíváme chemický posun jen jako doplňkový údaj, poskytující informaci o lokální konformaci, nejbližším okolí pozorovaného jádra. Nejčastěji využíváme empirické (odpozorované) korelace mezi hodnotami torzních úhlů a chemických posunů k tomu, abychom odhadli hodnoty torzních úhlů  $\phi$  a  $\psi$  v páteři proteinů.

Poslední údaj o konformaci, který zmíníme, vyžaduje zvláštní přípravu vzorku proteinu či nukleové kyseliny. Tímto údajem je opět přímá interakce mezi magnetickými momenty blízkých jader. Tentokrát nejde o vliv okamžitých interakcí na návrat magnetických momentů do rovnováhy, jak tomu bylo případě NOE, ale opravdu o měření průměrné hodnoty této interakce. A tato průměrná hodnota je v běžných roztocích nulová. Jak ji tedy můžeme získat?

To, co jsme si o dipól-dipólových interakcích dosud řekli, vycházelo z předpokladu, že se mohou molekuly v roztoku otáčet stejně dobře ve všech směrech. Pokud ale některý směr zvýhodníme (umístíme molekulu do *anizotropního prostředí*, například do kapalného krystalu, který se uspořádá v silném magnetickém poli spektrometru), nebude již průměr pole sousedních jader nulový. V tomto případě pozorujeme ve spektrech štěpení signálů, které závisí na částečné orientaci molekuly ve zkumavce. Měření *zbytkových dipolárních interakcí* nám umožní určit polohu vektorů mezi interagujícími jádry vůči vnější souřadné soustavě. Tato vnější souřadná soustava je určena osami, které popisují volnost molekuly rotovat v různých směrech. Tyto směry jsou zase dány směrem  $\vec{B}_0$  a druhem kapalného krystalu.

Pokud změříme zbytkovou dipolární interakci pro jednu dvojici jader (geometricky popsanou výše zmíněným vektorem směřujícím od jednoho jádra k druhému), neumožní nám to ještě jednoznačně určit orientaci této dvojice v naší vnější souřadné soustavě. Pokud ale naměříme zbytkových interakcí více v rámci částí proteinu, jejichž lokální konformace známe, například v dobře určených  $\alpha$ -šroubovicích,

můžeme opravdu spočítat, jak jsou tyto šroubovice v proteinu orientovány (obrázek 14.15 vpravo). Totéž platí například pro vzájemnou orientaci domén většího proteinu.

Zbytkové dipolární interakce jsou velmi užitečné proto, že vůbec nezávisí na vzdálenosti částí molekuly, jejichž vzájemnou orientaci určujeme. Je tomu tak proto, že všechny orientace vlastně určujeme vůči jedné společné souřadné soustavě. V tom se liší zbytkové dipolární interakce od NOE, které nám poskytnou údaje pouze o kontaktu částí molekul, jejichž protony se nacházejí v prostoru blízko sebe. Z toho vyplývá, že zbytkové dipolární interakce hrají nezastupitelnou úlohu při studiu málo kompaktních molekul, pro které potřebujeme určit vzájemnou orientaci částí, které jsou od sebe v prostoru daleko.

## 14.14 Sekundární struktura

Detailní popis struktury na základě NMR měření vyžaduje poměrně složité zpracování získaných dat, které si popíšeme v sekci 14.15. Teď se ale zkusme podívat na to, co se o struktuře můžeme dozvědět rychle a bez složitých analýz.

Na základě několika základních experimentů umíme popsat prvky sekundární struktury proteinu. Využíváme k tomu především údaje, které jsou shrnuty v následujícím výčtu a tabulce 14.1.

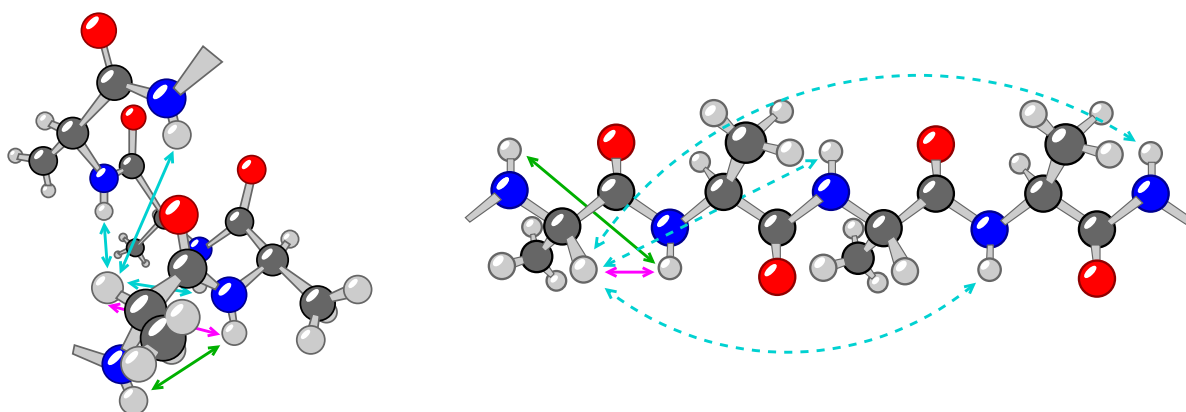
1. Hodnoty chemických posunů karbonylových uhlíků a  $\alpha$ -uhlíků jsou vyšší v  $\alpha$ -šroubovicích a nižší ve skládaných listech. Naopak hodnoty chemických posunů  $\alpha$ -protonů a  $\beta$ -uhlíků jsou nižší v  $\alpha$ -šroubovicích a vyšší ve skládaných listech. Většinu z těchto chemických posunů získáme již z experimentů, které používáme pro sekvenční přiřazení. Můžeme proto „zadarmo“ získat odhad rozložení prvků sekundární struktury ještě dříve, než vůbec začneme měřit spektra zaměřená na určení geometrie (NOE, interakční konstanty). Přestože je tato metoda velmi jednoduchá, poskytuje poměrně věrohodné výsledky.
2. Vyhodnocením NOE mezi  $\alpha$ -protony a amidovými protony peptidové vazby (obrázek 14.17) můžeme určit sekundární strukturu podle následujících pravidel. Ve skládaných listech je vzdálenost mezi  $\alpha$ -protony a amidovými protony sousední aminokyseliny (na obrázku 14.17 purpurová) menší než vzdálenost mezi dvěma amidovými protony sousedních aminokyselin (na obrázku 14.17 zelená), zatímco ve šroubovicích je tomu naopak. Navíc v  $\alpha$ -šroubovicích je malá vzdálenost mezi  $\alpha$ -protony a  $\beta$ -protony aminokyseliny vzdálené o tři rezidua. Dobře pozorovatelné jsou i NOE mezi  $\alpha$ -protony a amidovými protony aminokyselin vzdálených o tři a čtyři rezidua (azurové šipky na obrázku 14.17). Ve šroubovici  $3_{10}$  je na rozdíl od  $\alpha$ -šroubovice vzdálenost mezi  $\alpha$ -protony a amidovými protony aminokyselin vzdálených o tři rezidua kratší než vzdálenost mezi  $\alpha$ -protony a  $\beta$ -protony těchto aminokyselin.
3. V téměř rovinném uspořádání skládaného listu jsou silnější  $J$ -interakce mezi  $\alpha$ -protony a amidovými protony peptidové vazby než ve šroubovicích. Hodnoty těchto snadno měřitelných třívazebných interakčních konstant (bývají ve skládaných listech vyšší než 8 Hz a ve šroubovicích nižší než 5 Hz (tlustá čára na obrázku 14.16)). V praxi se proto obvykle sekundární struktury rozlišují pouze na základě této interakce, aniž by se měřily všechny interakční konstanty potřebné k matematicky jednoznačnému výpočtu torzního úhlu  $\phi$ . Torzní úhel  $\psi$  se zpravidla z  $J$ -interakcí nepočítá vůbec, protože příslušné interakční konstanty jsou špatně měřitelné.

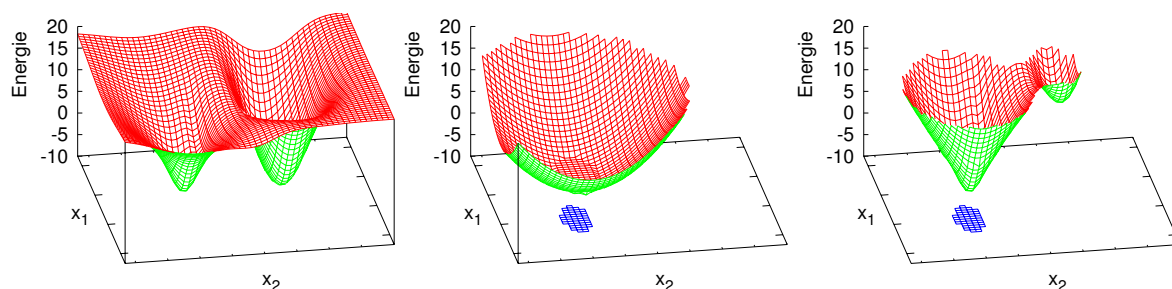
## 14.15 Výpočet konformace proteinu

Z výsledků NMR experimentů získáme velké množství geometrických údajů, zejména vzdálenosti mezi atomy vodíku, některé torzní úhly, informaci o sekundární struktuře a podobně. Na rozdíl od elektro- nové hustoty, kterou získáme z rentgenové difrakce nebo elektronové mikroskopie, z výsledků NMR

Tabulka 14.1: Údaje získané ze spekter NMR, které umožňují rozlišit sekundární struktury proteinů.

Data	$\beta$ -list	$\alpha$ -šroubovice
$\delta(C')$ , $\delta(C^\alpha)$	↓	↑
$\delta(C^\beta)$ , $\delta(H^\alpha)$	↑	↓
$ H_i^\alpha H_{i+1}^N $	0,22 nm	0,35 nm
$ H_i^N H_{i+1}^N $	0,40 nm	0,28 nm
$ H_i^\alpha H_{i+2}^N $	daleko	0,42 nm
$ H_i^\alpha H_{i+3}^N $	daleko	0,34 nm
$ H_i^\alpha H_{i+4}^N $	daleko	0,42 nm
${}^3J(H_i^N H_i^\alpha)$	> 8 Hz	< 5 Hz

Obrázek 14.17: Vzdálenosti protonů ve struktuře  $\alpha$ -šroubovice (vlevo) a skládaného listu (vpravo).



Obrázek 14.18: Příklad hyperplochy potenciální energie v MD simulaci (vlevo), umělé energie přiřazené odchylce od experimentálně určených vzdáleností (uprostřed), a součtu obou energií (vpravo). Oblast strukturních souřadnic (označených  $x_1, x_2$ ), která odpovídá výsledkům měření NMR, je vyznačena modře.

měření není na první pohled patrný tvar molekuly. Jsme v podobné situaci, jako když chceme nakreslit plán domu a k dispozici máme jen tabulku číselných hodnot různých vzdáleností. Musíme proto výsledky měření zpracovat do grafické formy. Výpočetní metody, které umožňují ze vzdáleností vytvořit trojrozměrný model, jsou označovány jako *distanční geometrie*.

Pokud bychom znali přesně hodnoty všech vzdáleností mezi jádry, poskytla by nám distanční geometrie věrný tvar molekuly<sup>18</sup>. Naměřené hodnoty jsou ale zatíženy experimentální chybou, takže určený tvar molekuly se od skutečnosti dost liší. To si můžeme sami ověřit, když se na molekulu pořádně podíváme. Víme, že konformace makromolekuly je dána hlavně torzními úhly. Vazebné délky a úhly by měly být v proteinu stejné jako ve volných aminokyselinách. Rozdíly nás varují, že navržená struktura není chemicky v pořádku.

Abychom získali chemicky rozumný model proteinu, používáme pro výpočet konformace metod molekulového modelování, které „vědí“, jak mají aminokyseliny vypadat. Nejčastěji se používá molekulová dynamika, ve které jako startovní konformaci můžeme použít třeba výsledek distanční geometrie. Mezi výpočtem struktury, jak jsme jej popisovali v kapitole 10, a využitím molekulové dynamiky v rámci určování struktury z výsledků NMR je ale důležitý rozdíl. Pokud vycházíme z výsledků NMR experimentů, používáme naměřené hodnoty (vzdálenosti atomů, torzní úhly, zbytkové dipólové interakce) jako dodatečné vazné podmínky. Molekulová dynamika se tedy nesnaží najít celkově nejvýhodnější konformaci, ale vybrat chemicky nepřijatelnější strukturu v rámci možností, kterým odpovídají experimentální data. Jinými slovy, NMR vymezí na hyperploše potenciální energie menší území a cílem molekulové dynamiky je najít v tomto území nejhlubší údolí (obrázek 14.18).

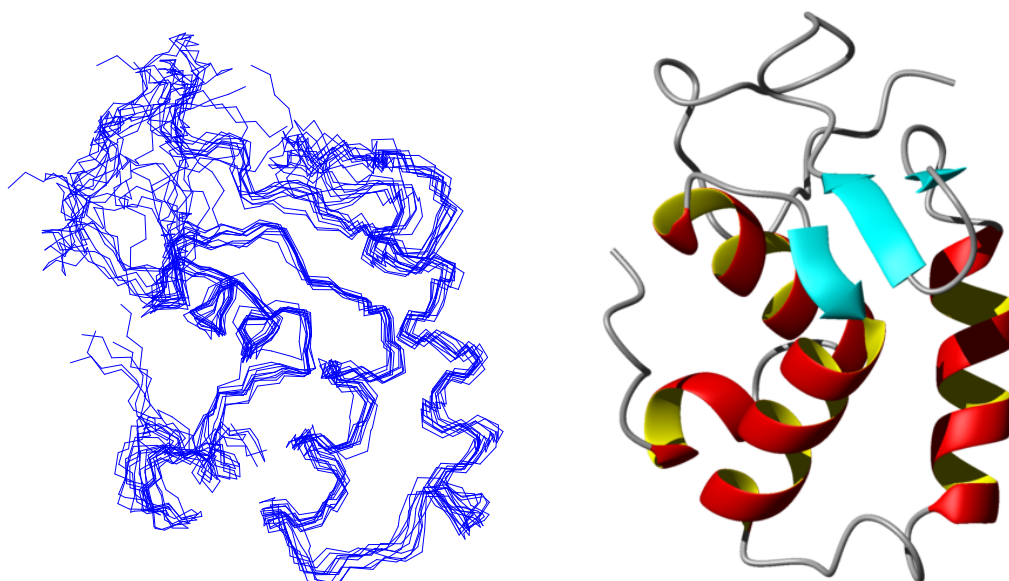
Při opakovaném výpočtu získáme trochu odlišnou strukturu. Zpravidla proto provádíme výpočtů několik a výsledné modely struktury přes sebe přeložíme (obrázek 14.19). V místech, která jsou experimentem nejlépe popsána, jsou jednotlivé modely téměř nerozlišitelné. Naopak v místech, pro která máme málo experimentálních údajů, nebo v pohyblivých částech molekuly, kde nám experiment poskytne jen průměr geometrických údajů, se modely liší.

## 14.16 Struktura nukleových kyselin pomocí NMR

Při určování struktur nukleových kyselin využívají stejné principy jako v případě proteinů. Postup práce je však v mnohém odlišný. Důvodem je nákladnost izotopového značení nukleových kyselin a odlišnost struktury, která činí NMR nukleových kyselin experimentálně obtížnější. Ve srovnání s proteiny se NOESY experimenty často používají nejen v druhé fázi řešení struktury, ale i k přiřazení frekvencí.

<sup>18</sup>Výjimkou je chiralita. Ze vzdáleností nejsme schopni určit, který z vzájemných zrcadlových obrazů je ten pravý. Naštěstí víme, že proteiny jsou stavěny z L-aminokyselin, takže nás tato nedokonalost nijak neomezuje.



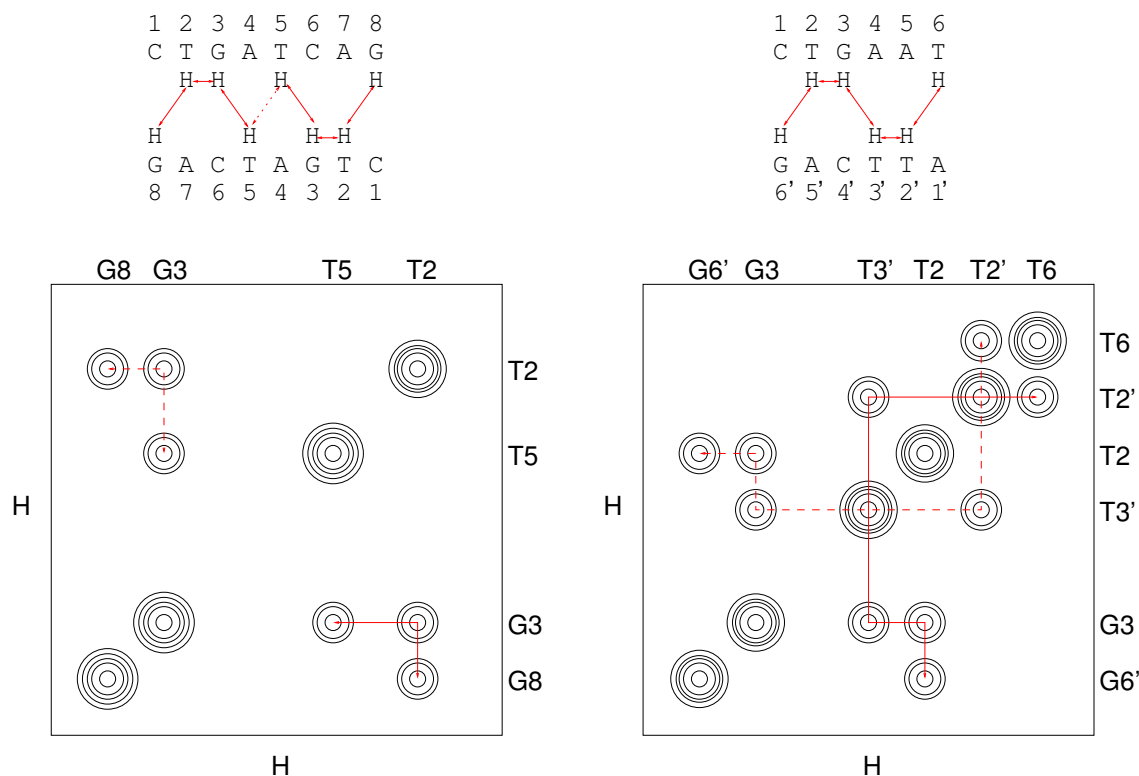


Obrázek 14.19: Peptidová páteř deseti struktur N-terminální domény podjednotky delta RNA polymerasy bakterie *Bacillus subtilis* vypočtených na základě dat (vlevo) a model struktury s nejnižší energií (vpravo).

Techniky frekvenčního přiřazení můžeme u nukleových kyselin dělit na experimenty využívající interakce zprostředkované elektrony kovalentních vazeb a na experimenty využívající NOE (dipolární interakce prostorem). Oba přístupy mají své výhody a nevýhody a v praxi se často kombinují. Zde si uvedeme jen příklady postupů založených na využití NOE.

Metody využívající NOE k frekvenčnímu přiřazení jsou založeny na korelacích protonů nacházejících se blízko sebe v prostoru. V principu tedy nevyžadují izotopové značení vzorku. V jednodušších případech lze tedy tímto způsobem studovat snadno dostupné vzorky nukleových kyselin v přirozeném zastoupení. Pokud je již k dispozici značený vzorek, kombinují se tyto metody často s heteronukleární korelací s uhlíky a protony, protože takto lze zvýšit rozlišení rozšířením spektra do další dimenze. Nevýhodou tohoto přístupu je, že vzdálenosti mezi protony a tedy i pozorované korelace závisí na konformaci dané molekuly nukleové kyseliny.

- *Korelace mezi imino (a amino) protony* V dvoušroubovicových strukturách jsou snadno pozorovatelné korelace mezi protony vodíkových můstků, které párují báze. Zpravidla jde o imino-protony, ale i protony aminoskupin jsou často pozorovatelné, zejména za nízkých teplot (vodíková vazba výrazně zpomalí výměnu protonu s protonem vody, která ve volných aminoskupinách způsobuje, že signály protonů nejsou pozorovatelné). Takto lze získat sekvenční přiřazení imino-protonů. Je nutno si uvědomit, že pouze guanin a thymin/uracil je donorem vodíku, takže korelace je patrná pouze mezi těmito bázemi (korelace „přeskakuje“ mezi jednotlivými vlákny šroubovice). Příklady spekter jsou uvedeny na obrázku 14.20. Všimněte si, že v případě symetrických dvojšroubovic jsou signály obou vláken identické.
- *Korelace báze-cukr* Pro přiřazení frekvencí jsou cenné korelace H6/H8 protonů s protony H1', H2' (silnější u dvojšroubovice A), H2'' (silnější u dvojšroubovice B) pentosového kruhu. Protože lze pozorovat korelace s vlastní (deoxy)ribosou a s cukrem předchozího nukleotidu, poskytují zmíněné korelace údaje pro sekvenční přiřazení (obrázek 14.21). Zároveň získáváme informaci o přiřazení cukru a báze v rámci jednoho rezidua.



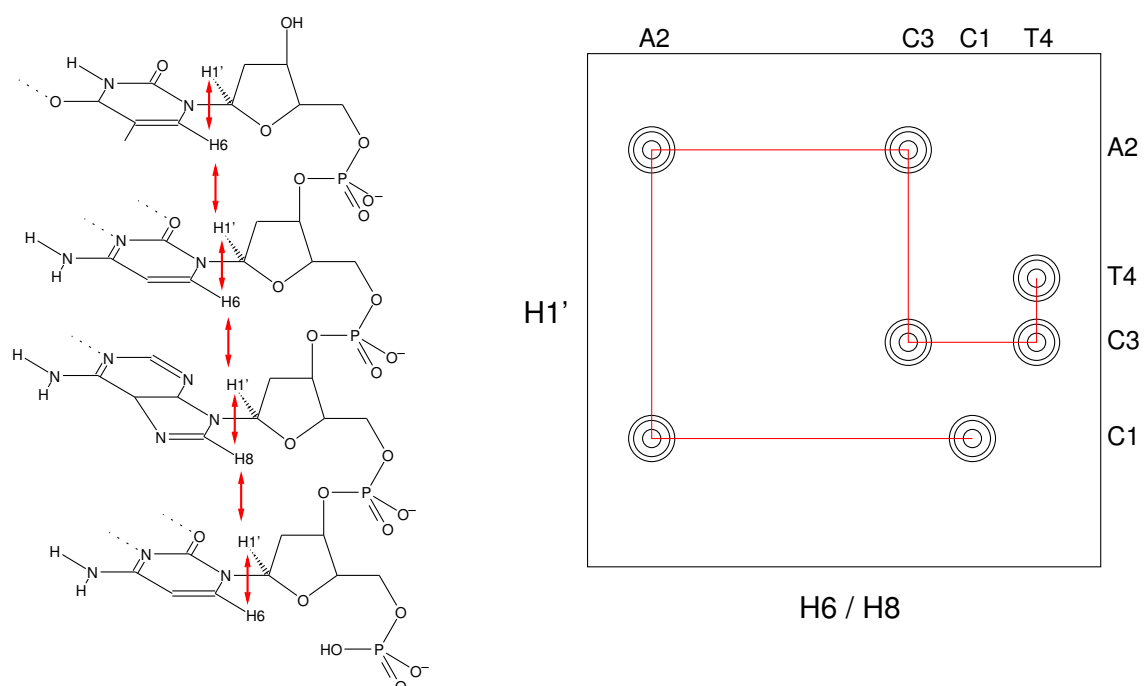
Obrázek 14.20: Příklady NOESY spekter oblasti imino-protonů bází. Červeně je znázorněna korelace mezi protony v sekvencích (tečkovaně je označena nepozorovatelná korelace mezi protony s identickou rezonanční frekvencí) a sekvenční informace ve spektrech. Vzhledem k tomu, že jde o diagonálně symetrická spektra, sekvenční informaci lze sledovat dvěma způsoby (plná a přerušovaná červená čára).

- *Korelace protonů cukru* Korelací mezi protony v rámci (deoxy)ribosového kruhu lze využít k přiřazení frekvencí protonů spinovým systémům jednotlivých bází. Navíc lze pozorovat korelace mezi protony cukrů sousedících nukleotidů. Nejsilnější bývají korelace  $H5''(i)-H2''(i-1)$  v dvojšroubovicích B a  $H5''(i)-H1'(i-1)$  v dvojšroubovicích A.

Využití NOE k určování vzdáleností mezi protony je u nukleových kyselin stejné jako u proteinů. Vzhledem k nižšímu počtu protonů v nukleových kyselinách bývá ale počet získaných meziatomových vzdáleností u nukleových kyselin nižší ve srovnání s proteiny. Při určování struktury závisí proto výsledná konformace více na závěrečném výpočtu molekulovou dynamikou, než tomu je v případě proteinů.

Torzní úhly v molekulách nukleových kyselin se také obvykle určují z hodnot třívazebných interakčních konstant. Relativně velké množství třívazebných interakcí je dostupné pro určování konformací cukrů. Při interpretaci těchto dat ale můžeme narazit na komplikace způsobené přítomností více konformací ve vzorku. Torzi kolem glykosidické vazby ( $\chi$ ) určujeme z třívazebných interakcí mezi atomy  $H1'-C2$  a  $H1'-C6$  u pyrimidinů a  $H1'-C4$  a  $H1'-C8$  u purinů. Z torzních úhlů páteře lze poměrně přesně určit  $\beta$  (z třívazebných interakcí mezi atomy  $H5''-P5'$ ,  $H5'-P5'$ ,  $C4'-P5'$ ),  $\epsilon$  (z třívazebných interakcí mezi atomy  $H3'-P3'$ ,  $C2'-P5'$ ,  $C4'-P3'$ ) a  $\gamma$  (z třívazebných interakcí mezi atomy  $H5''-H4'$ ,  $H5'-H4'$ ,  $H5''-C3'$ ,  $H5'-C3'$ ).

Výpočet struktury nukleové kyseliny na základě výsledků NMR se v principu neliší od výpočtu proteinů. Konkrétní postup výpočtu se ale odlišuje. Odlišná struktura monomerních jednotek vyžaduje



Obrázek 14.21: Vlevo schema korelací mezi protony  $H1'$  (v pentose) a  $H6$  nebo  $H8$  (v bázi), vpravo výsek oblasti NOESY spektra, která obsahuje tyto korelace. Červeně je znázorněna sekvenční informace. Protože je na obrázku ukázána pouze část spektra, sekvenční informaci můžeme sledovat jen jedním způsobem. Všimněte si, že proton první báze je korelován pouze s  $H1'$  vlastní pentosy a proton  $H1'$  poslední báze je korelován pouze s  $H6$  nebo  $H8$  protonem vlastní báze (v těchto případech chybí korelace se sousedem) – na základě toho je možné určit směr sekvence.

jiné silové pole než proteiny. Geometrických údajů z experimentu je obvykle méně, než v případě proteinů. Je tomu tak proto, že nejvíce informací poskytují atomy vodíku. Nukleové kyseliny obsahují relativně méně vodíků než proteiny. Zvláštní obtíž představuje tvar dvojšroubovice. Pokud vycházíme pouze ze vzdáleností, získáme údaje jen mezi sousedními páry nukleotidů. Tyto údaje obsahují experimentální chybu a tato chyba se postupně sčítá, takže máme velmi zkreslenou představu o tom, jestli je dvojšroubovice ohnutá. Tento nedostatek odstraňují údaje získané v anizotropním prostředí, například zbytkové dipólové interakce (viz sekci 14.6). Tyto údaje totiž odrážejí polohu vektorů mezi jádry vzhledem k vnějšímu magnetickému poli. Takto zprostředkovaně, přes vnější pole (které určuje anizotropní charakter prostředí) můžeme určit polohu skupin, které jsou v sekvenci libovolně vzdálené.

## 14.17 Struktura oligosacharidů pomocí NMR

Určování struktury (nederivatizovaných) sacharidů pomocí NMR je komplikováno nedostupností izotopově značených vzorků a špatným rozlišením protonových spekter (ze struktury vyplývá, že většina protonů má velmi podobné okolí). Sekvenční přiřazení se provádí pomocí NOE, přiřazení v rámci sacharidových jednotek je založeno na NOESY a TOCSY experimentech. Počet NOE korelací pro určování struktury mezi různými jednotkami sacharidů je většinou malý. K určování torzních úhlů v rámci jednoho cukerného kruhu lze využít třívazebné interakce mezi protony. K určování torzí kolem glykosidických vazeb je nutno měřit třívazebné interakce mezi protony a uhlíky (v přirozeném zastoupení  $^{13}\text{C}$ ). K odstranění některých ze zmíněných problémů lze použít chemickou derivatizaci (například metylaci).

## 14.18 Pohyby molekul v NMR

Během měření NMR sledujeme nejen amplitudu a frekvenci záření jednotlivých jader, ale také postupné slábnutí tohoto záření, *relaxaci*. Signál během měření slábne protože se postupně ztrácí koherence pohybů jader a protože se orientace těchto magnetických momentů v souboru molekulách vracejí do termodynamicky nejvýhodnějšího uspořádání<sup>19</sup>. Hlavní příčinou relaxace jsou změny interakcí jader s okolím (viz sekci 14.4), ke kterým dochází v důsledku pohybu molekul. Proto můžeme NMR využít nejen k popisu struktury molekuly, ale také k popisu pohybů jednotlivých atomů (*dynamiky molekuly*).

Při sledování dynamiky proteinů většinou měříme tři relaxační údaje pro jádro  $^{15}\text{N}$  v páteři proteinu (rychlost ztráty koherence, rychlost návratu do rovnováhy a nukleární Overhauserův efekt mezi  $^{15}\text{N}$  a  $^1\text{H}$  ve skupině NH). Analýza těchto údajů je poměrně jednoduchá za dvou předpokladů. Za prvé, pohyb proteinu popisujeme jako kombinaci nezávislé náhodné rotace (rotační difúze) celého proteinu a na ní nezávislého vnitřního pohybu. Za druhé, předpokládáme, že rotace celého proteinu je ve všech směrech stejná (což by platilo přesně, kdyby měla molekula proteinu tvar koule). Za těchto podmínek můžeme ze tří relaxačních údajů vypočítat tři údaje o rychlých pohybech páteře proteinu (v řádu  $10^{-12}$  až  $10^{-9}$  s). Získanými parametry jsou rychlost, s kterou se v roztoku otáčí celá molekula (ta je pochopitelně stejná pro všechna jádra), rychlost (frekvenci) a rozsah (amplitudu) vnitřních pohybů. Pokud jsou jádra ovlivněna více než jedním vnitřním pohybem, nebo pokud je tvar molekuly výrazně jiný, než kulový, potřebujeme k popisu více údajů. Proto často opakujeme měření za různých podmínek (například v různém magnetickém poli). Omezením tohoto přístupu je, že od celkové rotace rozlišíme jen pohyby rychlejší.

Pohyby, ke kterým dochází v řádu mikrosekund až milisekund, můžeme popsat také, pokud přitom dochází ke změnám chemického posunu. V tomto případě mluvíme o *konformační výměně* (chemický

<sup>19</sup>Jde o dva různé děje, které probíhají různě rychle. U velkých molekul dochází ke ztrátě koherence během desetin sekundy, ale k rovnováze se systém přiblíží až po několika sekundách. Proto signál přestáváme pozorovat již v desetinách sekundy po excitaci, ale měření můžeme opakovat nejdříve po jedné sekundě.

posun se mění v důsledku změny konformace) nebo o *chemické výměně* (chemický posun se mění v důsledku změny chemické struktury, například disociace protonu, nebo interakce s jinou molekulou). Pomocí speciálních experimentů lze určit zastoupení jednotlivých forem proteinu (například konformerů) a někdy i rychlost výměny.

Studium ještě pomalejších pohybů je snadné. Protože doba měření je přibližně jedna sekunda, můžeme pomalejší pohyby sledovat pohodlně opakovaným měřením.

Spektroskopie NMR nám tedy umožňuje získat nejen trojrozměrný obraz molekuly, ale dokonce popis ve čtyřech dimenzích (třech prostorových a jedné časové). To znamená, že určíme nejen celkovou strukturu, ale také to, které části molekuly jsou pohyblivější a jaký je rozsah a rychlost těchto pohybů. Informace o pohyblivosti může být důležitá pro pochopení funkce studované molekuly. Ačkoli se v našem povídání zaměřeném na určování struktur zmiňujeme o studiu pohybů pouze okrajově, tato aplikace je jednou z nejsilnějších stránek NMR.

## 14.19 Omezení metody a pokročilé aplikace

Použitelnost NMR pro studium biomakromolekul má v současnosti různá omezení. Některá z nich vyplývají z citlivosti metody. Uspořádání magnetických momentů jader se v dostupných magnetických polích za laboratorní teploty jen nepatrně liší od zcela náhodné orientace. Celková magnetizace jader v souboru molekul tvořícím vzorek tak odpovídá pouze tisícinám procenta magnetizace, kterou by měly zcela uspořádané magnetické momenty těchto jader. „Aktivní koncentrace“ jader ve vzorku je tak o několik řádů nižší, než celková koncentrace biomakromolekuly. To vyžaduje poměrně vysoké koncentrace vzorku (řádově milimolární) a obohacení molekul vzácnějšími izotopy  $^{13}\text{C}$  a  $^{15}\text{N}$ . S tímto omezením se snaží výrobci spektrometrů bojovat na dvou frontách: zvyšováním magnetického pole (tím se zvýhodňuje jedna orientace magnetických momentů a tak se zvyšuje celková magnetizace) a konstrukcí citlivějších detekčních zařízení.<sup>20</sup> Zvýšení citlivosti v principu umožňuje buď práci s méně koncentrovanými vzorky, nebo vynechání izotopového značení.<sup>21</sup>

Další omezení souvisí s relaxací. Obecně lze říci, že čím je větší molekula, tím pomaleji se pohybuje v roztoku, tím více času mají jádra interagovat a tím rychlejší je relaxace. Například u proteinů obsahujících více než 200 aminokyselin je relaxace tak rychlá, že začínáme mít málo času na některé důležité experimenty. Před několika lety se podařilo toto úskalí elegantně obeplout. Byly navrženy experimenty, ve kterých působí relaxace v důsledku interakcí s indukovaným polem elektronů proti relaxacím v důsledku interakcí se sousedním jádrem. Při určitém magnetickém poli tak téměř nedochází ke ztrátě koherence. Příspěvky k relaxaci v důsledku interakcí se vzdálenějšími protony lze potlačit přípravou vzorku proteinu značeného deuteriem ( $^2\text{H}$ ), v případě proteinů expresí v mediu připraveném z těžké vody ( $^2\text{H}_2\text{O}$ ). Tímto způsobem lze studovat i velké biomakromolekuly, například proteiny složené z tisíců aminokyselin.

Zvláštním případem molekul jsou *přirozeně neuspořádané proteiny*, jejichž konformace se stále mění a nedají se tedy popsat jedním strukturálním modelem. Relaxace takových molekul je většinou pomalá bez ohledu na počet aminokyselin. Není proto těžké naměřit spektra s vysokou intenzitou signálu. Je ale důležité dosáhnout dostatečného rozlišení, protože pohyby molekuly průměrují chemické posuny různých aminokyselin k téměř stejným hodnotám. V současné době jsou k dispozici experimenty, které potřebné rozlišení poskytují. Pro neuspořádané proteiny nemá cenu hledat definovanou strukturu. Pomocí analýzy chemických posunů můžeme ale spočítat zastoupení jednotlivých konformací. Pomocí metod citlivějších, než je měření NOE, *Krátkodobé kontakty nebo delší vzdálenosti můžeme studovat pomocí nepárových elektronů (volných radikálů)*. *Když takovou funkční skupinu vneseme do molekuly proteinu, urychlí ve*

<sup>20</sup>Citlivost detekce je dána poměrem signálu k šumu. Pro lepší potlačení šumu jsou v poslední době detekční zařízení chlazená plyným heliem na teplotu blízkou absolutní nule.

<sup>21</sup>Značení  $^{15}\text{N}$  je několikrát levnější než značení  $^{13}\text{C}$ . Proto je výhodné značení  $^{13}\text{C}$  vynechat a využít pouze přirozeného zastoupení izotopů uhlíku.

*svém okolí výrazně relaxaci. Můžeme buď pozorovat pokles intenzity signálu, nebo přímo měřit rychlosti relaxace. můžeme sledovat i krátkodobou tvorbu kompaktních trojrozměrných struktur.*

Dosud jsme mluvili jen o rozpustných proteinech. Velké úsilí je také vynakládáno na studium membránových proteinů. Membránové proteiny můžeme například zkoumat v roztoku, když je zabudujeme do malých lipidových micel. Strukturu proteinů nerozpustných ve vodě lze studovat v pevné fázi, nejlépe v mikrokrytalických vzorcích. Měření v pevné fázi jsou technicky náročnější, než měření v roztoku, a je obtížnější určit přesnou konformaci větších molekul. Rozdíly použitelnosti metody pro kapalné a pevné vzorky se ale snižují a struktury některých důležitých proteinů již byly určeny pomocí NMR v pevné fázi.

## Kapitola 15

# Bioinformatika

### 15.1 Co to je bioinformatika

V předchozích kapitolách jsme si popsali metody, pomocí kterých můžeme zkoumat strukturu velkých biologicky zajímavých molekul. Za poměrně krátkou dobu používání těchto metod se biochemici a biologové dozvěděli o struktuře biomakromolekul velmi mnoho. Dokonce tolik, že je již téměř nemožné se v záplavě zveřejněných struktur vyznat. Navíc předpovídání struktur ze sekvence dosáhla úrovně takové spolehlivosti, že má smysl struktury předpovězené z pořadí aminokyselin brát vážně. Protože takové množství strukturních dat nelze zvládnout lidskou silou, posedáváním v knihovně s tužkou v ruce, vznikl nový vědní obor, který se snaží vnést do záplavy informací řád s pomocí počítačů. Pro tento obor se ujal název *bioinformatika*.

Bioinformatika se narodila v osmdesátých letech dvacátého století. Matkou jí byla čile se vyvíjející počítačová technologie, otcem záměr sekvenovat co nejvíce DNA. Přčtené úseky DNA bylo nutno vzájemně srovnávat a analyzovat. Úroveň výpočetní techniky byla schopná množství dat zvládnout. O několik let později začal rychle růst počet biomakromolekul, u kterých se podařilo určit prostorovou strukturu. Také strukturní informace bylo třeba třídit. Dnes je bioinformatika nejdůležitější aplikací počítačů v biologických vědách. Chceme-li tento nový obor definovat, můžeme říci, že bioinformatika je informační technologie sloužící k uchování, zpracování a analýze biologických dat.

### 15.2 Genomové projekty

V sekci 9.4 jsme si ukázali, že určit sekvenci úseku DNA je poměrně snadné. To přivedlo molekulární biologie na myšlenku přečíst celou genetickou informaci některých organismů. Již v roce 1977 byl sekvenován genom viru  $\phi$ X174, obsahující 5 000 nukleotidů a 11 genů. V roce 1995 byla určena první úplná sekvence genomu skutečného organismu, bakterie *Haemophilus influenzae*, s 1,8 miliony nukleotidů a 1 700 genů. Cílem genomového projektu bylo získat úplné sekvence DNA modelových organismů: střevní bakterie *Escherichia coli*, kvasinky *Saccharomyces cerevisiae*, červa *Cenorhabditis elegans*, huseníčku rolního (*Arabidopsis thaliana*), mušky *Drosophila melanogaster*, myšky *Mus musculus*, psa (*Canis familiaris*) a autora celého projektu (*Homo sapiens*).

Genomový projekt se podobá postupu člověka neznalého arabštiny, který si pořídí Korán. Bude potřebovat několik dnů na to, aby se naučil arabská písmena. V tom okamžiku bude schopen celý Korán „přečíst“, tak jak my umíme „přečíst“ genom. Bude ale trvat měsíce, než náš čtenář Koránu začne rozumět tomu, co čte. Stejně tak je velký rozdíl mezi samotnou genetickou informací a funkcemi buňky, které kóduje.

Zmínili jsme se již v sekci 8.2, že vyjádření (exprese) informace uložené v DNA je složitý proces. Je proto dobré si uvědomit jednotlivé úrovně genetické informace:

1. *Genom* je informace zapsaná v DNA tvořící chromosomy, je společná všem buňkám organismu.
2. *Transkriptom* je část genomu, která je exprimována v buňce v dané fázi jejího vývoje.
3. *Proteom* je sada vzájemně interagujících molekul geneticky kódovaných proteinů, která dává buňkám její individuální charakter.

Jistě vás nepřekvapí, že ze všech modelových genomů se největšímu zájmu těší genom člověka. Sekvenace lidského genomu zahájená v polovině 80. let dvacátého století představuje dvacetileté úsilí mnoha laboratoří. Mimořádná popularita lidského genomového projektu vyprovokovala technický vývoj sekvenovacích automatů a potřebných výpočetních metod. Roku 2000 byla představena předběžná verze, lidský genom „nanečisto“ (*first draft*). K této předběžné verzi se dopracovaly nezávisle dva velké týmy: sdružení akademických laboratoří *International Consortium Human Genome Project* a komerční firma *Celera Genomics*. Dalšími dvacet let trvalo dokončení četby zvláště obtížných částí genomu. Již předběžné výsledky sekvenace ukázaly, že lidský genom obsahuje asi třicet tisíc genů, které ovšem kódují mnohem více proteinů. Stovky genů získal člověk z bakterií. Zajímavé jsou také poznatky o samovolných změnách genetické informace, o mutacích. Genetická informace se mění dvakrát rychleji u mužů než u žen. Celkem bylo odhaleno 1,4 milionu jednobodových polymorfismů (rozdílů genomů v jednom nukleotidu). Dnes je již kompletně přečtena genová informace mnoha organismů.

### 15.3 Biologické databáze

Biologické databáze jsou vlastně elektronické kartotéky biologicky zajímavých údajů. Díky své elektronické formě jsou mnohem pružnější než to, čemu se obvykle říkalo „kartotéka“ ve 20. století (plechové skříň plné kartiček z tvrdého papíru). Údaje v elektronických databázích můžeme mnohem snadněji třídít, analyzovat, porovnávat s jinými daty.

Biologických databází je mnoho druhů. Databáze můžeme třídít podle různých měřítek, například podle obsahu. Nás budou nejvíce zajímat databáze sekvencí a prostorových struktur proteinů a nukleových kyselin.

Jiným měřítkem je způsob uložení dat. Databáze může obsahovat jednotlivé položky jako jednoduché textové soubory (*flat files*, obrázek 15.1). Data mohou být ale také uspořádána jako v tabulkách, mluvíme pak o *relačních databázích*. Řádky takových tabulek tvoří jednotlivé položky, zatímco sloupce odpovídají základním charakteristikám položek (obrázek 15.2). Třetí možností je ukládání dat ve formě abstraktních objektů, takovým databázím říkáme *objektově orientované*. Jednotlivé objekty můžeme spojovat do vyšších objektů a zacházet s nimi jako s celky, což umožňuje velmi efektivní analýzu dat.

Ještě jiný způsob třídění rozlišuje tři druhy databází sekvencí podle způsobu jejich vzniku. Prvním druhem jsou databáze *primární*, které obsahují vlastní sekvence. Druhým typem jsou databáze *sekundární*, které obsahují další údaje odvozené ze sekvencí. Třetím druhem jsou databáze *složené*, které slučují několik primárních databází.

Je nutné také upozornit na to, že databáze se stále vyvíjejí. Nové vznikají a staré se slučují. Organizace, které databáze spravují se transformují nebo dokonce zanikají. A i u navenek nezměněných databází probíhá stálý a rychlý vývoj výpočetních prostředků, pomocí kterých se databáze prohledávají a analyzují. Informace o databázích proto rychle zastarává a místo konkrétních fakt je lépe pamatovat si, kde nalezneme odkazy na aktuální stav. Takový zdroj informací poskytuje *Swiss Institute of Bioinformatics* již od roku 1993 pod názvem *Expasy*. Na sklonku roku 2013 vznikla evropská „střecha“ nad databázemi a výpočetními zdroji zvaná *ELIXIR*. Internetové stránky těchto platform jsou vhodným začátkem hledání současných databází a prostředků jejich analýz.



```

DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Major urinary protein 1 precursor (MUP 1).
GN MUP1
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALE/c; TISSUE=Liver;
RX MEDLINE=87257897; PubMed=3600652;
RA Shaiban K., Gilmartin M., Derman E.;
RT "Nucleotide sequences of liver, lachrymal, and submaxillary gland
RT mouse major urinary protein mRNAs: mosaic structure and construction
RT of panels of gene-specific synthetic oligonucleotide probes.;"
RL Mol. Cell. Biol. 7:1938-1946(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=88007832; PubMed=2888770;
RA Bennett A.L., Paulson K.E., Miller R.E., Darnell J.E. Jr.;
RT "Acquisition of antigens characteristic of adult pericentral
RT hepatocytes by differentiating fetal hepatoblasts in vitro.;"
RL J. Cell Biol. 105:1073-1085(1987).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailis D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.;"
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RP SEQUENCE OF 20-41 FROM N.A.
RX MEDLINE=82265923; PubMed=7107707;
RA Krauter K., Leinwand L., D'Eustachio P., Ruddle F.,
RA Darnell J.E. Jr.;
RT "Structural genes of the mouse major urinary protein are on
RT chromosome 4.;"
RL J. Cell Biol. 94:414-417(1982).
CC -1- FUNCTION: Binds pheromones, the pheromones are released from
CC drying urine of males and affect the sexual behavior of females.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Abundant in the urine of adult male rats but
CC absent from that of females.
CC -1- SIMILARITY: Belongs to the lipocalin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M16355; AAA39767.1; -
DR EMBL: M28649; AAA39764.1; ALT_INIT.
DR EMBL: BC012221; AAH12221.1; -
DR EMBL: J00607; AAA39760.1; -
DR HSSP: P02762; 1MUP.
DR MGD: MGI:97233; Mup1.
DR InterPro: IPR002345; Lipocalin.
DR InterPro: IPR000566; Lipocaln_cytFABP.
DR Pfam: PF00061; lipocalin; 1.
DR PRINTS: PR00179; LIPOCALIN.
DR PROSITE: PS00213; LIPOCALIN; 1.
KW Glycoprotein; Signal; Pheromone-binding; Transport; Lipocalin;
KW Multigene family.
FT SIGNAL 1 18
FT CHAIN 19 180 MAJOR URINARY PROTEIN 1.
FT DISULFID 82 175 BY SIMILARITY.
FT CONFLICT 44 44 S -> F (IN REF. 2).
FT CONFLICT 127 127 K -> N (IN REF. 2).
SQ SEQUENCE 180 AA; 20648 MW; 5AA429BDA0700347 CRC64;
MKMLLLLLGL LTLVCHVAEE ASSTGRNFV EKingEWHIT IILASDKREKI EDNGNFRFLF
EQHLVLENSL VLKFTVRADE ECSELSMVA DTEKAGEYSV TYDGFNTFTI PKTDYDNFLM
AHLINERKDE TPQLMGLYGR EPDLSDDIKE RFAQLCEKRG ILRNITDLS NANRCLQARE
//

```

Obrázek 15.1: Příklad jednoduchého textového souboru.

Article ID	Journal	Volume	Pages	Year
001	Dev. Psychobiol.	19	155–161	1986
002	J. Exp. Zool.	265	558–586	1993
003	Behav. Genet.	23	51–58	1993
004	Biochem. J.	315	461–466	1995

Article ID	Author number	Author name
001	1	Drickamer, L. C.
002	1	Barkley, M.
002	2	DelgadoDeLeon, D.
002	3	Weste, R.
003	1	Williams, J. R.
003	2	Lenington, S.
004	1	Beynon, R. J.
004	2	Oliver, S.
004	3	Robertson, D. H. L.

Obrázek 15.2: Databáze odkazů na odborné články jako příklad relační databáze.

## 15.4 Databáze sekvencí proteinů

Znamé sekvence proteinů jsou uchovávány v různých primárních databázích. Nejdůležitější a nejužitečnější jsou následující dvě

- *SWISS-PROT* je nejpoužívanější databází proteinových sekvencí. Tato databáze vznikla na ženevské univerzitě, později ji převzal European Bioinformatics Institute (EBI), nyní o ni pečuje Swiss Institute of Bioinformatics (SIB). SWISS-PROT je zaměřen na vysokou kvalitu položek. Každou sekvenci provází podrobná *anotace*, tedy popis funkce, struktury, domén, posttranslačních modifikací atd. Tyto popisy jsou psány lidskou rukou, což zaručuje jejich vysokou kvalitu. SWISS-PROT obsahuje minimální počet opakujících se (a proto zbytečných, *redundantních*) položek, umožňuje účinné prohledávání a je propojen s mnoha dalšími databázemi. Příklad položky databáze SWISS-PROT je uveden na obrázku 15.1
- *TrEMBL* obsahuje všechny sekvence proteinů odvozené z databáze sekvencí DNA, která bude zmíněna v sekci 15.6. Anotace jsou tvořeny automaticky počítačem, což umožňuje velmi rychlý, ale méně kvalitní popis sekvence. Zkontrolované položky jsou potom ručně anotovány a přesouvány do databáze SWISS-PROT.

Obě zmíněné databáze byly integrovány do *UniProt KnowledgeBase* (pod názvy *UniProt/Swiss-prot* a *UniProt/TrEMBL*). UniProt KB tak můžeme považovat za příklad *složené databáze* sekvencí proteinů.

## 15.5 Sekundární databáze sekvencí proteinů

Zatím jsme hovořili pouze o databázích, které obsahují jednotlivé sekvence proteinů s příslušným popisem. Podívejme se nyní na databáze *sekundární*, které obsahují již *výsledky analýz* sekvencí. Nejdříve si ale musíme vysvětlit, co takovým výsledkem analýzy sekvence vlastně je.

Každá sekvence uložená v primární databázi je konkrétní údaj, jeden kamínek v mozaice. Když se nám podaří takových kamínků odkrýt více, začínáme vidět obraz, který mozaika znázorňuje. Několika sekvencím je společný určitý obecný rys, nějaké pravidlo, podle kterého je daný druh molekuly postaven. Například *první aminokyselina musí být methionin, druhá a třetí glycin, čtvrtá nějaká aromatická, pátá*

*cystein, šestá a sedmá mohou být libovolné, osmá musí být kyselina glutamová nebo asparagová, devátá alifatická hydrofobní, atd.* Tvorba sekundární databáze je tedy proces *abstrakce*. Zatímco v primárních databázích jsou uloženy *konkrétní* sekvence, sekundární databáze obsahují *abstraktní* principy.

Nejjednodušším příkladem abstrakce je sekvenční *motiv*. Sekvence proteinů obsahuje obvykle úseky, které jsou velmi podobné částem sekvencí jiných proteinů a úseky, které se naopak ostatním sekvencím podobají málo. Úseky sekvencí, které se s drobnými obměnami objevují v různých proteinech, se nazývají *motivy*. Podle zacházení s motivy můžeme sekundární databáze dělit na následující tři skupiny:

1. Databáze obsahující jednotlivé motivy. Příkladem může být nejstarší sekundární databáze *PROSITE*, která využívá především primární databázi SWISS-PROT a je udržována stejnou institucí (SIB). Motivy jsou v ní uloženy ve formě *regulárních výrazů*<sup>1</sup>. Záznamy v databázi PROSITE jsou uloženy ve dvou oddělených souborech, první obsahuje vlastní data a druhý podrobnou dokumentaci. Data jsou získávána automatickým vícenásobným přiložením sekvencí (viz sekci 15.9) a ručním výběrem úseků, které jsou *konzervované* (nacházejí se s malými obměnami v různých proteinech). Dokumentace je zpracována ručně. Cílem databáze PROSITE je umožnit identifikaci skupin příbuzných proteinů (*proteinových rodin*).
2. Databáze obsahující soubor všech motivů v sekvenci (takzvaný *fingerpint*, česky *otisk prstu* sekvence). *Fingerpinty* mohou být definovány na základě *frekvenčních matic* (tabulek udávajících, jak často se která aminokyselina nachází na daném místě; příkladem je databáze *PRINTS*).
3. Databáze obsahující nejen motivy, ale výsledky analýz *celých* sekvencí. Použitou abstrakcí v těchto databázích jsou *profily* (tabulky udávající, jak často se na daném místě nachází určitá aminokyselina nebo jak často na tomto místě dochází k inzerci nebo delecí<sup>2</sup>; příkladem je databáze *PROFILES*) nebo *skryté Markovovy modely* (matematické statistické algoritmy, které popisují výstavbu sekvence jako sled inzercí, delecí, nebo shod v aminokyselině na každém místě sekvence; příkladem je databáze *Pfam*).

Uvedené sekundární databáze (a některé další) jsou sdruženy pod názvem *InterPro*.

Každá sekundární databáze má své výhody a nevýhody. Je na nás, abychom se rozhodli, který přístup je nejlepší pro řešení problému, kterým se zabýváme. Také si musíme uvědomit, že sekundární databáze obsahují i chyby vyplývající z analýzy sekvencí. Každý postup vyhledávání společných rysů nese riziko, že vidíme příbuznost tam, kde ve skutečnosti není, nebo naopak příbuznost přehlédneme. Různé databáze analyzují stejné sekvence různě a mohou v nich tedy vidět různé (skutečné i jen zdánlivé) abstraktní principy. K údajům, které ze sekundárních databází získáme, je proto nutno přistupovat kriticky. Je vhodné analyzovat sekvenci s využitím různých metod a hledat, v čem se výsledky shodují (*consensus*). Zvýšíme tak pravděpodobnost správného závěru analýzy.

## 15.6 Databáze sekvencí DNA

V sekci 15.4 jsme si popsali databáze, které uchovávají sekvence proteinů. V databázích jsou proteinové sekvence uloženy jako pořadí aminokyselin pomocí nám dobře známé abecedy, která se skládá z dvaceti písmen. Ve skutečnosti je ovšem většina sekvencí proteinů určena na základě sekvence DNA a z abecedy DNA do abecedy proteinů přeložena pomocí genetické kódovací tabulky (takzvaná *koncepční*

<sup>1</sup>Zápis konkrétní sekvence je nepřetržitá řada znaků, každým znakem je jedno z dvaceti písmen symbolizující určitou aminokyselinu. Zápis motivu je obecnější. Jednotlivým znakem může být opět jedno písmeno (to znamená, že na daném místě je vždy určitá aminokyselina), nebo volba z více písmen (to znamená, že na daném místě je vždy jedna z uvedených aminokyselin), nebo speciální znak, zvaný divoká karta či zólk, za který můžeme dosadit jakoukoli aminokyselinu. Motiv, který jsme si jako příklad popsali slovně, můžeme zapsat pomocí následujícího regulárního výrazu:

M-G-G-[FYW]-C-x2-[DE]-[IVL].

<sup>2</sup>Delece je vypuštění aminokyseliny ze sekvence, inzerce je zařazení další aminokyseliny.

*translace*). Znamých sekvencí úseků DNA je ale více než těch, které tvoří geny určitého proteinu. Během desetiletí byla v laboratořích přečtena spousta úseků DNA, které nekódují žádný protein, nebo krátkých úseků, ze kterých nemůžeme určit úplnou sekvenci proteinu. Aby byl vnesen jakýsi řád i do těchto dat, bylo založeno několik databází, které ukládají sekvence úseků DNA. Nejdůležitější jsou následující tři databáze

*ENA (European Nucleotide Archive)* je evropská databáze vyvinutá v (*European Molecular Biology Laboratory*). *GenBank* je obdobná americká databáze vyvinutá v National Center for Biotechnology Information (NCBI) při National Institute of Health (NIH) v Bethesdě. *DNA Data Bank of Japan (DDBJ)* je třetí důležitou databází, vyvinutou japonským Národním ústavem genetiky. Všechny tři databáze úzce spolupracují a vyměňují si nové položky. Obsahem by se proto neměly lišit, takže rozhodnutí, kterou databázi využít, vyplývá hlavně z toho, který vyhledávací systém nám více vyhovuje.

Vedle zmíněné trojice databází, která uchovává všechny DNA sekvence s anotací, existují i databáze specializované. Jako příklad můžeme uvést databázi *GSDB*, která uchovává sekvence určené v rámci genomových projektů, nebo databázi *dbEST*, která obsahuje takzvané *exprimované sekvencní značky (EST)*<sup>3</sup>.

## 15.7 Databáze proteinových struktur

Přesná prostorová struktura byla určena jen pro malou část proteinů, jejichž sekvenci známe. I tak se známé struktury počítají na tisíce. Nepřekvapí proto, že byly vytvořeny databáze, ve kterých jsou struktury proteinů (a nejen proteinů) uloženy.

Nejdůležitější databází struktur je *Protein Data Bank (PDB)* která byla vyvinuta v Brookhaven National Laboratory a od roku 1998 je udržována v rámci Research Collaboratory for Structural Bioinformatics (RCSB). Od roku 2003 je spolu s obdobnou evropskou (*PDBe*) a japonskou (*PDBj*) databází součástí *Worldwide Protein Data Bank (wwPDB)*, která od roku 2006 zahrnuje i databázi NMR dat (*BioMagResBank*).

PDB uchovává struktury všech proteinů (ale i jiných biomakromolekul), které byly určeny rentgenovou difrakcí, nukleární magnetickou rezonancí a kryo-elektronovou mikroskopií. Struktury jsou uloženy v textových souborech, které obsahují informace o proteinu a souřadnice jeho atomů. Každá struktura je označena jedinečným identifikačním kódem, složeným ze čtyř znaků (číslic a písmen), například „1DUF“. Struktury je možné hledat také podle klíčových slov. Souřadnice atomů můžete uložit ve svém počítači, nebo si můžete trojrozměrný model molekuly prohlédnout na internetových stránkách PDB.

Kromě PDB existují i databáze obsahující analýzy struktur, podobně jako sekundární databáze obsahují analýzy sekvencí. Databáze *PDBSum* obsahuje shrnutí informací z PDB a přehledem analýzy struktury. Další databáze analyzující struktury slouží klasifikaci proteinů. Nejstaršími takovými databázemi jsou SCOP a CATH.

Databáze *SCOP (Structural Classification of Proteins)* byla vyvinuta v MRC Laboratory of Molecular Biology ve Velké Británii. Struktury jsou porovnávány kombinací automatických a ručních metod a tříděny na třech úrovních. Nejvýše stojí v hierarchii *fold*. Do stejného foldu patří proteiny se stejnými elementy sekundární struktury, ale bez známky příbuznosti. Další úroveň je *superrodina (superfamily)*, sdružující proteiny s podobnou strukturou i funkcí (což svědčí pro jejich společný původ), ale s velmi rozdílnými sekvencemi. Konečně do stejné *rodiny (family)* patří proteiny, které se podobají strukturou i sekvencí (více než 30 % aminokyselin shodných).

Odlíšnou strukturální hierarchii, obsaženou přímo v názvu, používá databáze *CATH (Class-Architecture-Topology-Homology)*, vyvinutá na University College London. Nejvýše v hierarchii stojí *třída (class)*. Každý protein je zařazen do jedné ze čtyř tříd:  $\alpha$ -proteiny,  $\beta$ -proteiny,  $\alpha$ - $\beta$ -proteiny (obsahují proteiny  $\alpha/\beta$  i  $\alpha+\beta$ ) a proteiny s nízkým obsahem sekundární struktury. Další úroveň je *architektura*

<sup>3</sup>*EST (Expressed Sequence Tag)* je krátký náhodně zvolený úsek sequence cDNA (viz sekci 8.4), pomocí kterého můžeme sledovat, jestli dochází k expresi daného genu v určité tkáni.

(*architecture*), která popisuje vzájemné prostorové uspořádání prvků sekundární struktury, aniž by se zabývala tím, jak jsou tyto prvky propojeny. Na třetí úrovni, *topology*, jsou pak proteiny dále tříděny podle topologie spojení prvků sekundární struktury. Čtvrtou úrovní je *homology*. Do skupiny se stejnou homologií patří proteiny, které mají více než 35 % sekvence shodné (předpokládá se, že mají společného předchůdce).

## 15.8 Párové přiložení (pairwise alignment) sekvencí

Chceme-li porovnat dvě sekvence, hledáme způsob, jak je přiložit vedle sebe tak, aby proti sobě bylo co nejvíce stejných aminokyselin (obrázek 15.3). Takový způsob se nazývá anglicky *pairwise alignment*, což můžeme do češtiny přeložit jako *párové přiložení*. Náš úkol je komplikován tím, že u jedné sekvence mohlo dojít k vyřazení (*deleci*) některé aminokyseliny, nebo naopak k zařazení (*inzerci*) aminokyseliny nové. Proto lze lepšího přiložení dosáhnout tehdy, když máme možnost do sekvence vkládat i mezery (anglicky *gap*). Tohoto prostředku musí být ale užíváno s mírou. Neomezeným vkládáním mezer můžeme samozřejmě dosáhnout toho, že i úplně odlišné sekvence přiložíme tak, aby proti sobě vždycky ležely stejné aminokyseliny. Již z tohoto jednoduchého příkladu je zřejmé, že párové přiložení je poměrně složitý úkol. Řešíme jej pomocí matematických modelů, jejichž vlastnosti měníme nastavením různých parametrů.

```

ulimit -t 30; /usr/molbio/bin/lalign -f -14 -g -4 -K 3 ./wwtmp/.7897.1.seq ./wwtmp/.7897.2.seq
> ./wwtmp/.7897.out LALIGN finds the best local alignments between two sequences
version 2.lu09 December 2006
Please cite: X. Huang and W. Miller (1991) Adv. Appl. Math. 12:373-381
alignments < E( 0.05):score: 42 (3 max)

Comparison of:
(A) ./wwtmp/.7897.1.seq unknown 121 bp                - 121 aa
(B) ./wwtmp/.7897.2.seq unknown 121 bp                - 121 aa
using matrix file: BL50 (15/-5), gap-open/ext: -14/-4 E(limit) 0.05

61.2% identity in 121 aa overlap (1-121:1-121); score: 579 E(10000): 4.3e-52

      10      20      30      40      50      60
unknow DHHAASCKNEWFSYNGKCYFFSTTTKTWALAQKSCSEDDATLAVIDSEKDMAFLKRYAGG
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
unknow DSHVSSCSESDVWGYQRKCYFISTVKRSWTSAQNACSEHGATLAVIDSEKDMNFKRYAGR
      10      20      30      40      50      60

      70      80      90      100     110     120
unknow LKHWIGLRNEASQTKWANGKEFNWFWNVTGSKKCVSLNHTDVASVDCEANLHWICSKPS
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
unknow EEHWVGLKKEPGHPWWSNGKEFNWFWNVTGSEKCVFLKNTFVSSMECEKNLYWICNKPY
      70      80      90      100     110     120

unknow R
      .
unknow K

61.2% identity in 121 aa overlap (1-121:1-121); score: 579 E(10000): 4.3e-52

      10      20
unknow NEWFSYNG--KCYFFSTT
      : : : : : : : : : :
unknow NNWFWNVTGSEKCVFLKNT
      90      100

```

Obrázek 15.3: Příklad párového přiložení (výstup programu *LALIGN*).

Nalézt párové přiložení celé dlouhé sekvence (*globální přiložení*) je výpočetně velmi náročné. Algoritmy párového přiložení se proto často opírají o hledání podobností v rámci krátkých úseků sekvence (*lokální přiložení*). Tak pracují nejběžnější algoritmy *FastA* a *BLAST*. Vylepšená verze algoritmu *BLAST*, která umožňuje vložení mezer, je dnes nejpoužívanější metodou prohledávání databází. Algo-

ritmus nejprve vyhledává přiložení krátkých úseků sekvence bez mezer, pak se snaží rozšířit přiložení tak, aby proti sobě leželo co nejvíce shodných aminokyselin. Pro měření shody se počítá určité *skóre*, takže hledáme páry s vysokým skóre (*high-scoring pairs*). Nakonec se přiložení vylepší vložením mezer. Abychom mohli odhadnout, nakolik přiložení odhaluje skutečnou příbuznost sekvencí, výsledky algoritmů obsahují i údaj o pravděpodobnosti, že přiložení je dílem náhody (hodnota  $p$  v algoritmu BLAST), nebo očekávaný počet úseků dané délky, které jsou náhodou shodné v prohledávané databázi (hodnota  $E$ ). Máme-li přiložení věřit, musí být hodnoty  $p$  i  $E$  co nejmenší.

Pod zkratkou BLAST se ve skutečnosti skrývá celou rodinu programů. Jedním z nich je *PSI-BLAST* (*Position-Specific Iterated BLAST*), který kombinuje metody párového a vícenásobného přiložení, o kterém bude řeč v části 15.9. V první fázi hledá nejlepší párové přiřazení. V rámci nejlepších výsledků tohoto kroku provede vícenásobné přiložení a vytvoří *profil*, o kterém jsme mluvili při diskusi sekundárních databází v sekci 15.5. Na základě profilu znovu prohledává databáze a snaží se najít vzdáleně příbuzné sekvence, které v minulém kroku přehlédl. Takovýmto opakovaným prohledáváním umí PSI-BLAST odhalit i velmi slabou podobnost. PSI-BLAST je proto jednou z nejcitlivějších metod detekce vzdálených příbuzných. Tato výhoda ale může být i nevýhodou. Do profilu se nám mohou vloučit sekvence tvořené kratičkými stále se opakujícími úseky (takzvané oblasti s nízkou komplexitou). Příkladem může být kolagen, v němž se pravidelně opakuje trojice aminokyselin Gly-Pro-Pro. V takovém případě prohledávání selže a snaží se nám vnutit zdánlivou podobnost se zcela odlišnými proteiny.

## 15.9 Vícenásobné přiložení (multiple alignment)

Často chceme vědět nejen to, jak se sobě navzájem podobají dvě sekvence. Chceme zkoumat vzájemnou podobnost v rámci celé skupiny sekvencí, která například patří do stejné proteinové rodiny. Snažíme se proto přiložit k sobě více sekvencí tak, abychom vytvořili tabulku, ve které bude na každém řádku jedna sekvence a ve sloupcích pod sebou budou co nejpodobnější aminokyseliny. Tato metoda, zvaná *multiple alignment* (*vícenásobné přiložení*), je nepostradatelná při analýze genových rodin a tvorbě sekundárních databází.

Nejlepší způsob přiložení je poskládat sekvence tak, aby každý sloupeček tabulky tvořily aminokyseliny, které jsou na stejném místě ve struktuře proteinu. Protože prostorovou strukturu známe jen pro omezený počet proteinů, musíme často skládat tabulky vícenásobného přiložení jen na základě sekvencí, podle podobnosti aminokyselin. Například kyselina asparagová by se měla nejlépe ocitnout ve sloupečku tvořeném kyselinou asparagovou. Pokud to nejde, tak by měla být ve sloupečku tvořeném zejména kyselinou glutamovou (té se nejvíce podobá nábojem) nebo asparaginem (tomu se podobá velikostí a tvarem postranního řetězce). Již méně je kyselina asparagová podobná krátkým polárním aminokyselinám a velmi se liší od aminokyselin s dlouhými nepolárními postranními řetězci.

Je možné hledat najednou nejlepší přiložení celé skupiny sekvencí. Takovým způsobem pracují *simultánní metody*, které jsou však výpočetně náročné. Častější je postupné budování tabulky přiložení pomocí *progresivních (postupných) metod*. Oblíbeným programem založeným na postupném přístupu je *CLUSTAL*. Tento program k sobě postupně přikládá dvojice sekvencí, třídí je podle vzájemné podobnosti a vytváří strom příbuznosti, jaký znáte ze šlechtických rodokmenů. Nejpodobnější sekvence odpovídají sousedním větvím stromu, zatímco málo podobné sekvence představují vzdálené větve. Rozšířenou verzí tohoto programu je *CLUSTAL W*, který se na základě mezer již vložených do sekvencí snaží vkládat mezery další a tak hledat způsoby přiložení velmi málo podobných sekvencí (obrázek 15.4). Od roku 2011 vyvíjená verze *CLUSTAL Omega* znamená výrazné zlepšení v analýze velkého množství frekvencí.



## 15.10 Předpovídání struktury ze sekvence

Když jsme si popisovali architekturu proteinů v kapitole 2, zdůrazňovali jsme, že plán struktury je obsažen již v genu, v jeho sekvenci. Celou kapitolu 10 jsme věnovali otázce, jestli umíme plán struktury bez experimentálních údajů spočítat. Odpověď byla spíše rozpačitá. Chápeme sice základní fyzikální principy, které určují, jaká konformace je nejvýhodnější. Neumíme ale příslušné rovnice přesně řešit. Výpočetní metody, které se pokoušejí co nejvěrněji popsat přírodní zákony, jsou příliš výpočetně náročné, než aby mohly být použity pro biomakromolekuly složené z tisíců atomů. Metody molekulové mechaniky, využívající místo zákonů kvantové mechaniky zjednodušující popis, jsou značně nespolehlivé, mají-li najít nejvhodnější konformaci pouze na základě sekvence. V této části si položíme otázku, jestli nám v předpovídání struktury nemůže pomoci bioinformatika. Jestli není možné pečlivou analýzou ohromného množství údajů v databázích odhalit nějakou jednoduchou souvislost mezi sekvencí a strukturou. Nemusíme rozumět fyzikálním příčinám, stačilo by jen zjistit, že jisté pořadí aminokyselin vždycky znamená určitou strukturu.

## 15.11 Předpověď sekundární struktury

Pokusme se nejprve předpovědět sekundární strukturu na základě sekvence (primární struktury). To, co o sekundárních strukturách víme z části 2.5, zchladí náš optimismus hned na začátku. Již podle definice jsou sekundární struktury *pravidelná* prostorová uspořádání, která by neměla záviset na postranních řetězcích aminokyselin (a tedy ani na sekvenci). Ve skutečnosti tato nezávislost není úplná. Statistická analýza ukazuje, že určité sekundární struktury jsou pro určité aminokyseliny o něco výhodnější a pro jiné aminokyseliny naopak méně výhodné. Byly vyvinuty algoritmy, které předpovídají, s jakou pravděpodobností bude aminokyselina v určitém místě sekvence součástí  $\alpha$ -šroubovice,  $\beta$ -listu, pravidelného ohybu, nebo nepravidelné smyčky. Tyto algoritmy mohou být založeny na statistickém vyhodnocení, analýze tvaru postranních řetězců, nebo na analýze příbuznosti (homologie). Všechny metody ale vycházejí z pravidel, která byla formulována na základě sekvencí těch proteinů, jejichž prostorová struktura byla určena experimentálně. Příklad předpovědi sekundární struktury je ukázán na obrázku 15.5.

## 15.12 Předpověď terciární struktury

Posuňme se nyní v hierarchii struktur výše a zkusme ze sekvence předpovědět, jaká bude terciární struktura proteinu. Náš úkol bude snazší, pokud je již známá prostorová struktura proteinu s podobnou sekvencí. Takové proteiny mohou posloužit jako výchozí formy (*templáty*) pro přístup zvaný *homologní modelování*. Obecně můžeme postupovat takto:

1. Ze struktury templátu vezmeme polohy uhlíků peptidové páteře a postranních řetězců. Pravidelné strukturální motivy použijeme jako základ modelu naší molekuly.
2. Snažíme se namodelovat méně pravidelné části struktury: ohyby a smyčky.
3. Navržený model dále upřesňujeme a ověřujeme jeho správnost.

Tímto postupem můžeme velmi rychle určit strukturu proteinu, aniž bychom museli provádět složitý experiment. Nezbytnou podmínkou ale je, že již známe strukturu příbuzného proteinu. Co když příbuzný protein s experimentálně určenou strukturou zatím k dispozici není? Můžeme předpovědět prostorovou strukturu přímo ze sekvence aminokyselin? Od roku 2020 to možné je. První program schopný spolehlivě předpovědět terciární strukturu ze struktury primární (ze sekvence), byl *AlphaFold2* společnosti





*DeepMind*, využívající zejména znalosti sekvencí evolučně příbuzných pořadí aminokyselin zkoumaného proteinu. My si stručně popíšeme, jak tento program pracuje.

Co nám říká srovnání sekvencí o prostorovém uspořádání molekul proteinu. Ukážeme si to na jednoduchém příkladu. Předpokládejme, že známe sekvence našeho proteinu v mnoha organismech. To je dnes běžná situace. Tyto sekvence přiložíme vedle sebe, jak jsme popsali v části 15.9. V našem proteinu je aminokyselina číslo 75 lysin a číslo 237 kyselina asparagová. To nám zatím o terciární struktuře nic neříká. Při pohledu na přílohu zjistíme, že v ostatních sekvencích je na pozici odpovídající číslu 75 v našem proteinu většinou lysin nebo arginin.<sup>4</sup> V těchto sekvencích na pozici našeho čísla 237 vždycky kyselina asparagová nebo glutamová. V některých sekvencích je ale na pozici odpovídající číslu 75 kyselina asparagová nebo glutamová. A teď pozor! V těchto sekvencích nacházíme na pozici 237 lysin nebo arginin. Co nám to říká o terciární struktuře? Na pozicích 75 a 237 jsou *vždycky* aminokyseliny s opačným nábojem. Postranní řetězce lysinu a argininu jsou nabitě kladně, postranní řetězce kyseliny asparagové a glutamové jsou nabitě záporně. Nejjednodušší vysvětlení je, že ve struktuře našeho proteinu se aminokyseliny číslo 75 a 237 elektrostaticky přitahují. Proto musí mít opačný náboj a není tak důležité, která aminokyselina je kladná a která záporná. To je ale důležitá strukturní informace. Když budeme stavět trojrozměrný model proteinu, musí být aminokyseliny číslo 75 a 237 blízko sebe a jejich postranní řetězce musí být natočené proti sobě.

Podobných příkladů bychom mohli uvést víc. Místo kladně a záporně nabitě aminokyseliny bychom mohli mít aminokyseliny s krátkým a dlouhým postranním řetězcem. Když bude řetězec aminokyseliny číslo 75 dlouhý, musí být řetězec aminokyseliny číslo 237 krátký, nebo naopak. Kdyby byly oba řetězce dlouhé, ve struktuře by si překáželi. V sekvencích se také odrazí pravidelné střídání hydrofobních a hydrofilních aminokyselin ve svazku šroubovic nebo  $\beta$ -soudku, zachování hydrofobních řetězců u aminokyselin tvořících jádro proteinu a podobně.

Vidíme, že existuje souvislost mezi zachováním typu aminokyselin na určitých pozicích v sekvenci a prostorovým uspořádáním aminokyselin na těchto pozicích v trojrozměrné struktuře. Tuto souvislost bychom mohli zkusit popsat pomocí konkrétních pravidel (například „zachování opačného náboje znamená blízkost v prostoru a otočení postranních řetězců k sobě“) a tato pravidla definovat srozumitelně pro počítač (tedy matematicky). Když bychom pak poskytli počítači údaje o zachování aminokyselin v mnoha sekvencích, měl by počítač umět spočítat trojrozměrný model proteinu. AlphaFold2 ale postupuje jinak. Využívá informatický přístup zvaný *strojové učení (machine learning)*. Úkolem počítače není jen na základě naprogramovaných pravidel vypočítat ze srovnání sekvencí prostorovou strukturu. Počítač sám hledá pravidla popisující souvislost mezi srovnáním sekvencí a prostorovou strukturou.

V první, učící fázi (*training*) poskytneme známé struktury proteinů a vícenásobná přeložení jejich sekvencí. Počítač začne pracovat se sekvencemi, známé struktury si prozatím uloží pro konečné srovnání. S vícenásobným přeložením pracuje počítač jako s tabulkou (označíme si ji  $M$ ), kde počet sloupečků je počet aminokyselin v sekvenci a počet řádků je počet přiložených sekvencí. Ze sekvence proteinu, jehož strukturu má počítač předpovědět (tuto sekvenci si označíme  $S$ ), si vytvoří další tabulku popisující vzájemné vztahy mezi jednotlivými aminokyselinami sekvence. Tato tabulka (označíme ji  $Z$ ) má stejně sloupců i řádků (podle počtu aminokyselin) a každé políčko tabulky popisuje vztah dvou aminokyselin v sekvenci. Pokud je k dispozici protein s podobnou sekvencí a známou strukturou (*templát*), je možné informace o známé struktuře (vzdálenosti, torzní úhly) zapracovat do tabulky. AlphaFold2 ale funguje i bez templátu, jen na základě vícenásobného přiložení.

První krok v předpovědi provádí modul zvaný *Evoformer*. Tento modul upravuje tabulky  $M$  a  $Z$  a vyměňuje informaci mezi nimi.  $Z$  porovnání sloupců v tabulce  $M$  upravuje informace o strukturní blízkosti aminokyselin v tabulce  $Z$ .<sup>5</sup> Naopak na základě vztahů mezi aminokyselinami v tabulce  $Z$  přidává váhu těm aminokyselinám (zapsaným jako položky na řádcích tabulky  $M$ ), které by měl výpočet

<sup>4</sup> Příslušná aminokyselina nemusí být v ostatních sekvencích vždycky sedmdesátá pátá, protože během evoluce mohlo dojít k delecí (ztrátě jedné aminokyseliny) nebo inzercí (vmezení nové aminokyseliny).

<sup>5</sup> Takto by AlphaFold2 vyhodnotil náš příklad zachování aminokyselin s opačnými náboji, protože ve sloupcích tabulky  $M$  je uložena informace o zachování aminokyselin na stejné pozici.

zohlednit nejvíc (jsou pro strukturu důležitější, protože například odpovídají aminokyselinám, které jsou podle tabulky  $Z$  blízko sebe v prostoru). Evoformer upravuje navíc tabulku  $M$  tím, že hledá vztahy mezi sekvencemi, které si zaslouhují větší pozornost. V tabulce  $Z$  zase provádí úpravy, které zajistí zachování správné geometrie.<sup>6</sup>

Druhým krokem v předpovědi je samotný výpočet trojrozměrného modelu *Strukturálním modulem*. Tento modul nejprve s využitím tabulky  $Z$  upraví hodnoty (váhy) v jednořádkové tabulce  $S$ .<sup>7</sup>, tak aby co nejlépe odpovídaly důležitosti pro strukturu. Z hodnot tabulky  $S$  vypočítá hodnoty vzájemného posunutí a natočení jednotlivých aminokyselin. Teprve potom dopočítá torzní úhly a tak určí polohy jednotlivých atomů. Nakonec se provádí krátká simulace molekulové dynamiky s optimalizací energie. Pouze v tomto kroku se používá silové pole postavené na základě chemických znalostí.

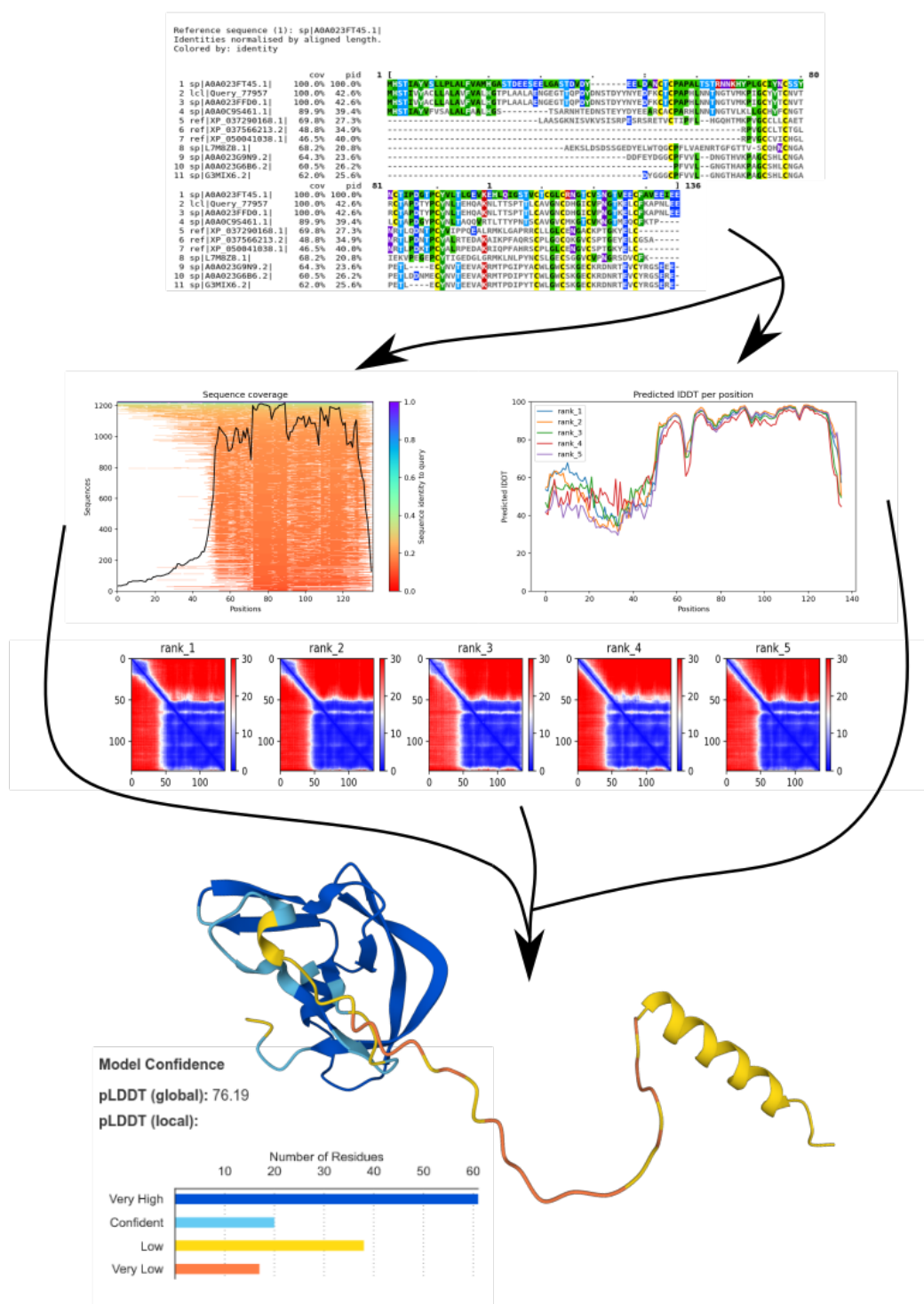
Oba kroky se několikrát opakují. V učící fázi se předpovězená struktura se porovná se strukturou získanou z experimentálních dat. Učení znamená, že se celý postup mnohokrát opakuje. Počítač přitom se mění číselné parametry, které definují jednotlivé výpočetní kroky. Změny, které předpověď struktury zlepšují, zachovává a snaží se tak nalézt parametry, které vedou k předpovědi blízké experimentálně určené struktuře. Když takové parametry nalezne pro celou skupinu struktur vybraných k učení, považuje program za „vyučení.“

Program AlphaFold2 naučený na strukturách uložených v databázi PDB předpovídá nové struktury (neznámé v době předpovědi, ale experimentálně určené později) se správností blížící se chybě experimentálních metod samotných. Z tohoto pohledu můžeme i bioinformatiku považovat za metodu určování struktury biologicky zajímavých molekul. Příklad předpovědi je ukázán na obrázku 15.6.

---

<sup>6</sup>Pokud jsou například aminokyseliny 75 a 237 vzdáleny 1 nm a aminokyseliny 237 a 541 0,5 nm, nemohou být aminokyseliny 75 a 541 od sebe dál, než 1,5 nm.

<sup>7</sup> $S$  je řádkem tabulky  $M$ , který představuje sekvenci proteinu, jehož strukturu předpovídáme.



Obrázek 15.6: Příklad předpovědi terciární struktury (výstup programu *AlphaFold2*).

Část III  
Dodatky



## A Energie různých interakcí

V jednoduchých případech lze elektrostatické působení mezi molekulami a jejich částmi popsat nepříliš složitými rovnicemi. V této části se podíváme na několik důležitých příkladů. Jde o různé varianty Coulombova zákona. Odvození uvedených rovnic často jednoduché není, v učebnicích většinou chybí a v literatuře se hledá obtížně. Zde si uvedeme i poměrně zdlouhavá odvození, aby čtenář nemusel zdlouhavě pátrat, odkud se vzaly různé mocniny vzdálenosti mezi interagujícími náboji a dipólovými momenty a závislosti na teplotě.

### A.1 Energie dvou nábojů

Energii  $U$  elektrostatické interakce mezi náboji  $Q_1$  a  $Q_2$ , které se nacházejí ve vakuu ve vzájemné vzdálenosti  $r$ , vypočítáme ze síly  $F$ , kterou se opačné náboje přitahují a stejné náboje odpuzují. Velikost této síly je

$$F = \frac{1}{4\pi\epsilon_0} \frac{Q_1 Q_2}{r^2}, \quad (15.1)$$

kde  $\epsilon_0$  je elektrická permitivita vakua ( $8,854 \cdot 10^{-12} \text{ F m}^{-1}$ ). Energie je rovná práci, která by se vykonala (nebo kterou bychom museli vynaložit) při přemístění nábojů ze vzdálenosti, kde na sebe náboje již nepůsobí (což by musela být nekonečná vzdálenost), do vzdálenosti  $r$ . Tuto práci spočítáme tak, že posčítáme všechny maličké práce  $dU$  spojené s posunutím náboje o nepatrný kousek  $dr'$ . Matematicky to lze napsat jako integrál

$$U = \int_{r_{\text{ref}}}^r F dr' = \int_{\infty}^r F dr' = \frac{Q_1 Q_2}{4\pi\epsilon_0} \int_{\infty}^r \frac{1}{r'^2} dr' = \frac{1}{4\pi\epsilon_0} \frac{Q_1 Q_2}{r}. \quad (15.2)$$

Pokud budeme chtít uvádět energii nábojů v jednom molu molekul, musíme výsledek vynásobit Avogadrovým číslem:

$$U = \frac{N_A}{4\pi\epsilon_0} \frac{Q_1 Q_2}{r}. \quad (15.3)$$

Je také dobré si uvědomit, že zatímco energie je veličina skalární, síla je vektor, má směr. Směr síly můžeme do Coulombova zákona zahrnout pomocí *jednotkového vektoru mířícího ve směru síly*. Takový jednotkový vektor můžeme zapsat jako  $\vec{r}/r$ , kde  $\vec{r}$  je vektor udávající vzájemnou polohu interagujících nábojů v dané souřadné soustavě a  $r$  je velikost tohoto vektoru, tedy vzdálenost nábojů:

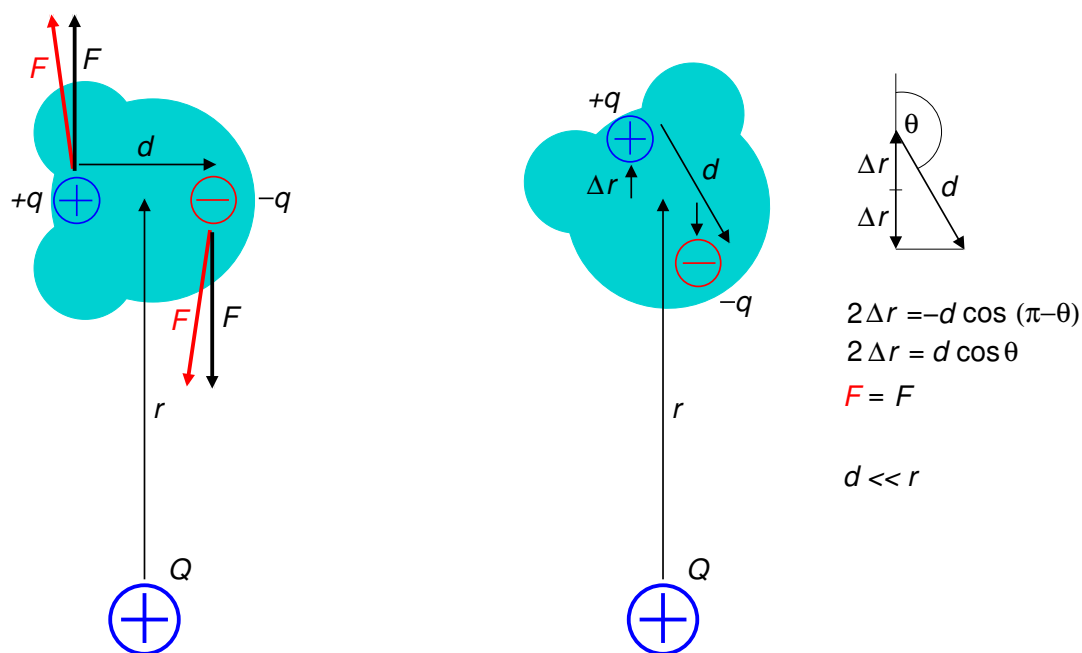
$$\vec{F} = \frac{1}{4\pi\epsilon_0} \frac{Q_1 Q_2}{r^2} \frac{\vec{r}}{r} = \frac{1}{4\pi\epsilon_0} \frac{Q_1 Q_2}{r^3} \vec{r}. \quad (15.4)$$

Další užitečnou veličinou je *elektrická intenzita*, tedy elektrická síla působící na jednotkový náboj v elektrickém poli jiného náboje o velikosti  $Q$ :

$$\vec{E} = \frac{1}{4\pi\epsilon_0} \frac{Q}{r^2} \frac{\vec{r}}{r} = \frac{1}{4\pi\epsilon_0} \frac{Q}{r^3} \vec{r}. \quad (15.5)$$

### A.2 Energie náboje a elektrického dipólu v jedné molekule

Energie  $U$  interakce mezi nábojem  $Q$  a trvalým elektrickým dipólovým momentem o velikosti  $qd$ , který odpovídá dvojici nábojů  $+q$  a  $-q$  ve vzájemné vzdálenosti  $d$ , je rovná práci spojené s otočením dipólu v poli náboje  $Q$ . Dle dohody začínáme dipól otáčet z polohy znázorněné na obrázku 15.7 vlevo. Jak je ukázáno na obrázku 15.7, na každý náboj  $Q$  působí síla směřující od nebo k náboji  $Q$  (na obrázku



Obrázek 15.7: Interakce elektrického dipólového momentu s nábojem.

červeně). Pokud je vzdálenost dipólu od náboje  $Q$  mnohem větší, než vzájemná vzdálenost nábojů  $+q$  a  $-q$ , lze síly považovat za téměř rovnoběžné (na obrázku znázorněno černě). Posunutí  $+q$  a  $-q$  vůči  $Q$  lze pak vyjádřit pomocí úhlu  $\theta$ , definovaného na obrázku 15.7 vpravo. Celkovou práci na posunutí obou nábojů o hodnotu  $\Delta r \ll r$  lze spočítat

$$\begin{aligned}
 U &= \int_r^{r+\Delta r} F_1 dr' + \int_r^{r-\Delta r} F_2 dr' = \frac{qQ}{4\pi\epsilon_0} \int_r^{r+\Delta r} \frac{1}{(r')^2} dr' - \frac{qQ}{4\pi\epsilon_0} \int_r^{r-\Delta r} \frac{1}{(r')^2} dr' \\
 &= -\frac{qQ}{4\pi\epsilon_0} \left( \frac{1}{r+\Delta r} - \frac{1}{r} - \frac{1}{r-\Delta r} + \frac{1}{r} \right) = \frac{qQ}{4\pi\epsilon_0} \frac{2\Delta r}{r^2 - \Delta r^2} \approx \frac{qQ}{4\pi\epsilon_0} \frac{2\Delta r}{r^2} = \frac{1}{4\pi\epsilon_0} \frac{qQ}{r} \frac{d}{r} \cos \theta.
 \end{aligned} \tag{15.6}$$

Po vynásobení Avogadrovou konstantou

$$U = \frac{N_A}{4\pi\epsilon_0} \frac{qQ}{r} \frac{d}{r} \cos \theta, \tag{15.7}$$

kde červená barva zdůrazňuje, čím se vztah liší od energie dvou nábojů. Porovnání s rovnicí 15.5 nám říká, že

$$U = \frac{N_A}{4\pi\epsilon_0} \frac{qQ}{r^3} dr \cos \theta = \frac{N_A}{4\pi\epsilon_0} \frac{qQ}{r^3} \vec{r} \cdot \vec{d} = N_A q \vec{E} \cdot \vec{d}, \tag{15.8}$$

čehož využijeme při zkoumání vzájemného působení dvou elektrických dipólů.



### A.3 Energie náboje a elektrického dipólu v různých molekulách

Pokud jsou náboj a elektrický dipól v různých molekulách, které se mohou vůči sobě volně otáčet, dává smysl počítat průměrnou energii pro všechny vzájemné orientace. Obecně můžeme průměr veličiny  $X$ , která závisí na jiné veličině  $u$ , spočítat jako

$$\int_u P(u)X(u)du, \quad (15.9)$$

kde  $P(u)$  je pravděpodobnost, že pro dané  $u$  bude mít  $X$  hodnotu  $X(u)$ . Tuto pravděpodobnost poskytuje Boltzmannův zákon

$$P(u) = \frac{e^{-U(u)/RT}}{Z} = \frac{e^{uw}}{Z}, \quad (15.10)$$

kde

$$u = \cos \theta, \quad (15.11)$$

$$w = -\frac{N_A}{4\pi\epsilon_0} \frac{dqQ}{r^2 RT} \quad (15.12)$$

a  $Z$  je součet všech možných hodnot  $e^{uw}$ , zvaný *partiční funkce*. Tento součet můžeme spočítat jako integrál

$$Z = \int_0^\pi P(\cos \theta) \sin \theta d\theta = -\int_1^{-1} e^{uw} du = \int_{-1}^1 e^{uw} du = \frac{e^w - e^{-w}}{w}. \quad (15.13)$$

Při výpočtu integrálu jsme využili toho, že

$$\frac{du}{d\theta} = \frac{d(\cos \theta)}{d\theta} = -\sin \theta, \quad (15.14)$$

takže  $\sin \theta d\theta = -du$ . Průměrnou energii potom můžeme spočítat jako integrál

$$\langle U \rangle = \int_{-1}^1 P(u)U(u)du = \frac{-RTw}{e^w - e^{-w}} \int_{-1}^1 ue^{uw} du = -RT \left( w \frac{e^w + e^{-w}}{e^w - e^{-w}} - 1 \right) \quad (15.15)$$

$$= -RT \frac{w(e^w + e^{-w}) - (e^w - e^{-w})}{e^w - e^{-w}}, \quad (15.16)$$

kde  $\int_{-1}^1 ue^{-uw} du$  jsme spočítali integrací *per partes*. Ačkoli jsme získali přesný výsledek, můžeme jej ještě zjednodušit. Můžeme využít toho, že exponenciální funkci lze vyjádřit jako nekonečnou řadu rostoucích mocnin

$$e^{\pm w} = \frac{1}{0!}(\pm w)^0 + \frac{1}{1!}(\pm w)^1 + \frac{1}{2!}(\pm w)^2 + \frac{1}{3!}w^3 + \dots = 1 \pm w + \frac{1}{2}w^2 \pm \frac{1}{6}w^3 + \dots \quad (15.17)$$

Protože jsme předpokládali, že  $d \ll r$ , můžeme také předpokládat, že  $w \ll 1$ . Proto můžeme zanedbat vysoké mocniny  $w$ . Jmenovatele ve výsledku rovnice 15.16 je pro velmi malá  $w$  přibližně

$$e^w - e^{-w} \approx 1 + w + \dots - 1 + w - \dots \approx 2w. \quad (15.18)$$

Při vyčíslení čitatele musíme být ale opatrnější. Kdybychom zanedbali všechny vyšší mocniny, než první, vylili bychom s vaničkou i dítě: získali bychom nulu. Když ale zachováme mocniny až do  $w^3$ , získáme

$$w(e^w + e^{-w}) - (e^w - e^{-w}) \approx w(2 + w^2) - (2w + \frac{1}{3}w^3) = (2w + w^3) - (2w + \frac{1}{3}w^3) = \frac{2}{3}w^3. \quad (15.19)$$

Celkově tedy

$$\langle U \rangle = -RT \frac{w(e^w + e^{-w}) - (e^w - e^{-w})}{e^w - e^{-w}} \approx -RT \frac{\frac{2}{3}w^3}{2w} = \frac{RTw^2}{3} = -\frac{1}{3RT} \left( \frac{N_A}{4\pi\epsilon_0} \frac{qdQ}{r^2} \right)^2, \quad (15.20)$$

kde je červeně vyznačen rozdíl od energie dipólu a náboje pevně zakotvených v jedné molekule.

#### A.4 Energie dvou elektrických dipólů v jedné molekule

Vzájemná interakce dvou elektrických dipólových momentů vlastně zahrnuje vzájemné silové působení čtyř nábojů. Jak ukazuje obrázek 15.8, k popisu vzájemné polohy dvou dipólů potřebujeme určit vzájemnou orientaci dvou vektorů v prostoru ( $\vec{d}_1, \vec{d}_2$ ). Pro zápis vektorů je výhodné použít sférické souřadnice:

$$d_{1x} = d_1 \sin \theta_1 \cos \phi_1, \quad (15.21)$$

$$d_{1y} = d_1 \sin \theta_1 \sin \phi_1, \quad (15.22)$$

$$d_{1z} = d_1 \cos \theta_1 \quad (15.23)$$

a

$$d_{2x} = d_2 \sin \theta_2 \cos \phi_2, \quad (15.24)$$

$$d_{2y} = d_2 \sin \theta_2 \sin \phi_2, \quad (15.25)$$

$$d_{2z} = d_2 \cos \theta_2 \quad (15.26)$$

Souřadnou soustavu můžeme přitom vždy zvolit tak, aby  $\phi_1 = 0$  (aby  $\vec{d}_1$  leželo v rovině  $xz$ ), jak je ukázáno na obrázku 15.8.

Analýza vzájemné interakce čtyř nábojů nebo dvou dipólů je poměrně náročná, ale značně nám ji zjednoduší vztah 15.8. Stačí nám určit, jakou intenzitu  $\vec{E}$  (tedy sílu působící na jednotkový náboj) má elektrické pole prvního dipólu v místě, kde se nachází druhý dipól. Energii pak snadno spočítáme pomocí skalárního součinu podle rovnice 15.8:

$$U = N_A q_2 \vec{E} \cdot \vec{d}_1. \quad (15.27)$$

Kde ale vzít intenzitu  $\vec{E}$ ? Obecně to je dosti složitý úkol. Intenzita  $\vec{E}$  v místě určeném polohovým vektorem  $\vec{r}$  je vektorovým součtem sil, kterými na jednotkový náboj v daném místě působí oba náboje prvního dipólu. Na obrázku 15.8 jsou intenzity odpovídající těmto silám znázorněny červenou šipkou  $\vec{E}_-$  (intenzita záporného náboje) a modrou šipkou  $\vec{E}_+$  (intenzita kladného náboje). V souřadné soustavě, použité na obrázku 15.8, mají obě intenzity nulovou složku ve směru  $y$ . Složky ve všech směrech můžeme vyjádřit



$$\sin \alpha = \frac{d_1}{2a} \sin \theta_1, \quad (15.36)$$

$$\sin \beta = \frac{d_1}{2b} \sin(\pi - \theta_1) = \frac{d_1}{2b} \sin \theta_1. \quad (15.37)$$

K vyjádření kosinů zase použijeme větu kosinovou

$$\frac{d_1^2}{4} = a^2 + r^2 - 2ar \cos \alpha \Rightarrow \cos \alpha = \frac{a^2 + r^2 - \frac{d_1^2}{4}}{2ar}, \quad (15.38)$$

$$\frac{d_1^2}{4} = b^2 + r^2 - 2br \cos \beta \Rightarrow \cos \beta = \frac{b^2 + r^2 - \frac{d_1^2}{4}}{2br}. \quad (15.39)$$

Jak ukazuje obrázek 15.8, celková intenzita ve směru  $x$  je součtem složek intenzit  $E_-$  a  $E_+$  v tomto směru:

$$E_x = E_{+,x} + E_{-,x} = E_+ \sin \beta + E_- \sin \alpha = \frac{q_1}{4\pi\epsilon_0} \frac{d_1 \sin \theta_1}{2} \left( \frac{1}{b^3} + \frac{1}{a^3} \right) = \frac{q_1}{4\pi\epsilon_0} \frac{a^3 + b^3}{a^3 b^3} \frac{d_1 \sin \theta_1}{2}. \quad (15.40)$$

Celková intenzita ve směru  $z$  je naopak rozdílem svislých složek intenzit  $E_-$  a  $E_+$ :

$$E_z = E_{+,z} + E_{-,z} = E_+ \cos \beta - E_- \cos \alpha = \frac{q_1}{4\pi\epsilon_0} \frac{1}{2r} \left( \frac{b^2 + r^2 - \frac{d_1^2}{4}}{b^3} - \frac{a^2 + r^2 - \frac{d_1^2}{4}}{a^3} \right). \quad (15.41)$$

Vzdálenosti  $a$  a  $b$  nám poskytne opět kosinová věta

$$a^2 = \frac{d_1^2}{4} + r^2 - d_1 r \cos \theta_1, \quad (15.42)$$

$$b^2 = \frac{d_1^2}{4} + r^2 - d_1 r \cos(\pi - \theta_1) = \frac{d_1^2}{4} + r^2 + d_1 r \cos \theta_1. \quad (15.43)$$

Po dosazení do čitatele rovnice 15.41

$$\begin{aligned} E_z &= \frac{q_1}{4\pi\epsilon_0} \frac{1}{2r} \left( \frac{2r^2 - d_1 r \cos \theta_1}{b^3} - \frac{2r^2 + d_1 r \cos \theta_1}{a^3} \right) = \frac{q_1}{4\pi\epsilon_0} \left( \left( \frac{1}{b^3} - \frac{1}{a^3} \right) r - \left( \frac{1}{b^3} + \frac{1}{a^3} \right) \frac{d_1 \cos \theta_1}{2} \right) \\ &= \frac{q_1}{4\pi\epsilon_0} \left( \frac{a^3 - b^3}{a^3 b^3} r - \frac{a^3 + b^3}{a^3 b^3} \frac{d_1 \cos \theta_1}{2} \right). \end{aligned} \quad (15.44)$$

Výrazy  $a^3 + b^3$  a  $a^3 - b^3$  můžeme převést na součiny

$$a^3 + b^3 = (a + b)(a^2 + b^2 - ab) \quad (15.45)$$

$$a^3 - b^3 = (a - b)(a^2 + b^2 + ab). \quad (15.46)$$

Součet  $a^2 + b^2$  můžeme snadno převést na  $2r^2 + d_1^2/2$ , ale snaha vyjádřit  $ab$ ,  $a + b$  a  $a - b$  pomocí  $r$  a  $d_1$  vede k nepřehledným odmocninám

$$\begin{aligned}
ab &= \sqrt{\left(\frac{d_1^2}{4} + r^2 - d_1 r \cos \theta_1\right) \left(\frac{d_1^2}{4} + r^2 + d_1 r \cos \theta_1\right)} = \sqrt{\left(\frac{d_1^2}{4} + r^2\right)^2 - (d_1 r \cos \theta_1)^2} = \\
&= \sqrt{\frac{d_1^4}{16} + r^4 + \frac{d_1^2 r^2}{2} - (d_1 r \cos \theta_1)^2} = r^2 \sqrt{1 + \left(\frac{d_1}{2r}\right)^4 - \left(\frac{d_1}{r} \cos \theta_1\right)^2}. \quad (15.47)
\end{aligned}$$

$$\begin{aligned}
a + b &= \sqrt{(a+b)^2} = \sqrt{a^2 + b^2 + 2ab} = \sqrt{2r^2 + \frac{d_1^2}{2} + 2r^2 \sqrt{1 + \left(\frac{d_1}{2r}\right)^4 - \left(\frac{d_1}{r} \cos \theta_1\right)^2}} \\
&= r \sqrt{2 + \frac{1}{2} \left(\frac{d_1}{r}\right)^2 + 2 \sqrt{1 + \left(\frac{d_1}{2r}\right)^4 - \left(\frac{d_1}{r} \cos \theta_1\right)^2}}, \quad (15.48)
\end{aligned}$$

$$a - b = \frac{a^2 - b^2}{a + b} = \frac{2d_1 \cos \theta_1}{\sqrt{2 + \frac{1}{2} \left(\frac{d_1}{r}\right)^2 + 2 \sqrt{1 + \left(\frac{d_1}{2r}\right)^4 - \left(\frac{d_1}{r} \cos \theta_1\right)^2}}}. \quad (15.49)$$

Pro  $d_1 \ll r$  můžeme našťestí zanedbat vyšší mocniny zlomku  $d_1/r$ , což výrazy pro  $ab$ ,  $a + b$  a  $a - b$  dramaticky zjednoduší:

$$ab = r^2 \quad (15.50)$$

$$a + b = 2r \quad (15.51)$$

$$a - b = d_1 \cos \theta_1, \quad (15.52)$$

takže

$$\frac{a^3 + b^3}{a^3 b^3} = \frac{(a+b)(a^2 + b^2 - ab)}{a^3 b^3} = \frac{(2r)(2r^2 - r^2)}{r^6} = \frac{2}{r^3} \quad (15.53)$$

$$\frac{a^3 - b^3}{a^3 b^3} = \frac{(a-b)(a^2 + b^2 + ab)}{a^3 b^3} = \frac{(d_1 \cos \theta_1)(2r^2 + r^2)}{r^6} = \frac{3d_1 \cos \theta_1}{r^4}. \quad (15.54)$$

Po dosazení do vztahů pro  $E_x$  a  $E_z$

$$E_x = \frac{q_1}{4\pi\epsilon_0} \frac{d_1 \sin \theta_1}{r^3} \quad (15.55)$$

$$E_z = -\frac{q_1}{4\pi\epsilon_0} \frac{2d_1 \cos \theta_1}{r^3}. \quad (15.56)$$

Konečně,

$$U = N_A q_2 \vec{E} \cdot \vec{d}_2 = N_A q_2 (E_x d_{2x} + E_y d_{2y} + E_z d_{2z}) = \frac{N_A}{4\pi\epsilon_0} \frac{q_1 q_2}{r} \frac{d_1}{r} \frac{d_2}{r} (\sin \theta_1 \sin \theta_2 \cos \phi_2 - 2 \cos \theta_1 \cos \theta_2), \quad (15.57)$$

kde červená barva zvýrazňuje, čím se vztah liší od energie dvou nábojů.

## A.5 Energie dvou elektrických dipólů v různých molekulách

Výpočet průměrné energie interakce elektrických dipólů ve dvou molekulách, které se mohou vůči sobě volně otáčet, jak ukazuje obrázek 15.8, je dosti zdlouhavý. Při počítání průměrů musíme integrovat přes všechny tři úhly  $\theta_1, \theta_2, \phi_2$ . Exponenciální člen Boltzmannova zákona můžeme opět psát jako

$$e^{-U/RT} = e^{uw}, \quad (15.58)$$

kde

$$w = -\frac{N_A}{4\pi\epsilon_0} \frac{q_1 q_2}{r} \frac{d_1}{r} \frac{d_2}{r} \quad (15.59)$$

podobně jako ve vztahu pro náboj a dipól, ale

$$u = \sin \theta_1 \sin \theta_2 \cos \phi_2 - 2 \cos \theta_1 \cos \theta_2. \quad (15.60)$$

Průměrnou energii pak počítáme jako

$$\langle U \rangle = \frac{\int_0^{2\pi} d\phi_2 \int_0^\pi \int_0^\pi U e^{uw} \sin \theta_1 d\theta_1 \sin \theta_2 d\theta_2}{Z} = \frac{-RTw \int_0^{2\pi} d\phi_2 \int_0^\pi \int_0^\pi u e^{uw} \sin \theta_1 d\theta_1 \sin \theta_2 d\theta_2}{\int_0^{2\pi} d\phi_2 \int_0^\pi \int_0^\pi e^{uw} \sin \theta_1 d\theta_1 \sin \theta_2 d\theta_2}. \quad (15.61)$$

Vztah se velmi zjednoduší, když si uvědomíme, že

$$u e^{uw} = \frac{d e^{uw}}{dw}. \quad (15.62)$$

Protože  $w$  nezávisí na úhlech  $\theta_1, \theta_2, \phi_2$ , můžeme podle  $w$  derivovat celý integrál. Integrál v čitateli je tak vlastně derivací partiční funkce  $Z$  podle  $w$ :

$$\langle U \rangle = \frac{-RTw \frac{dZ}{dw}}{Z}. \quad (15.63)$$

Když si uvědomíme, že

$$\frac{d(\ln Z)}{dZ} = \frac{1}{Z}, \quad (15.64)$$

můžeme psát

$$\langle U \rangle = -RTw \frac{d(\ln Z)}{dw} = -RTw \frac{d}{dw} \ln \left( \int_0^{2\pi} d\phi_2 \int_0^\pi \int_0^\pi e^{uw} \sin \theta_1 d\theta_1 \sin \theta_2 d\theta_2 \right). \quad (15.65)$$

Při výpočtu integrálu  $Z$  můžeme opět exponenciální závislost vyjádřit jako mocninnou řadu

$$e^{uw} = \frac{1}{0!}(uw)^0 + \frac{1}{1!}(uw)^1 + \frac{1}{2!}(uw)^2 + \frac{1}{3!}(uw)^3 + \dots = 1 + uw + \frac{1}{2}(uw)^2 + \frac{1}{6}(uw)^3 + \dots \quad (15.66)$$

a předpokládat, že  $w \ll 1$ . Tentokrát bude moci zanedbat všechny mocniny vyšší než druhou a integrovat výraz

$$\begin{aligned} \left(1 + uw + \frac{1}{2}(uw)^2\right) \sin \theta_1 \sin \theta_2 &= \sin \theta_1 \sin \theta_2 + w \sin^2 \theta_1 \sin^2 \theta_2 \cos \phi_2 - 2w \sin \theta_1 \cos \theta_1 \sin \theta_2 \cos \theta_2 + \\ &+ 2w^2 \cos^2 \theta_1 \cos^2 \theta_2 \sin \theta_1 \sin \theta_2 - 2w^2 \sin^2 \theta_1 \cos \theta_1 \sin^2 \theta_2 \cos \theta_2 \cos \phi_2 + \frac{w^2}{2} \sin^3 \theta_1 \sin^3 \theta_2 \cos^2 \phi_2 \end{aligned} \quad (15.67)$$

Integrály výrazů obsahujících  $\cos \phi_2$  musí být rovné nule, protože integrujeme přes celou periodu (přes všechna  $\phi_2$  od nuly do  $2\pi$ ) a integrály funkcí sinus a kosinus přes celou periodu (úměrné průměrným hodnotám těchto funkcí) jsou nulové. Ze stejného důvodu je nulový integrál výrazu

$$2 \sin \theta_1 \cos \theta_1 \sin \theta_2 \cos \theta_2 = \frac{1}{2} \sin(2\theta_1) \sin(2\theta_2), \quad (15.68)$$

protože úhly  $\theta_1$  a  $\theta_2$  integrujeme od nuly do  $\pi$ , což odpovídá celé periodě pro  $2\theta_1$  a  $2\theta_2$  (od nuly do  $2\pi$ ). Stačí nám tedy spočítat integrály třech výrazů. Výpočet usnadní substituce  $u_1 = \cos \theta_1$  a  $u_2 = \cos \theta_2$ , pro které platí  $du_1 = -\sin \theta_1 d\theta_1$  a  $du_2 = -\sin \theta_2 d\theta_2$ .

První integrál je

$$\int_0^{2\pi} d\phi_2 \int_0^\pi \int_0^\pi \sin \theta_1 d\theta_1 \sin \theta_2 d\theta_2 = \int_0^{2\pi} d\phi_2 \int_1^{-1} \int_1^{-1} du_1 du_2 = [\phi_2]_0^{2\pi} [u_1]_1^{-1} [u_2]_1^{-1} = 8\pi. \quad (15.69)$$

Druhý integrál je

$$\begin{aligned} 2w^2 \int_0^{2\pi} d\phi_2 \int_0^\pi \int_0^\pi \cos^2 \theta_1 \cos^2 \theta_2 \sin \theta_1 d\theta_1 \sin \theta_2 d\theta_2 &= 2w^2 \int_0^{2\pi} d\phi_2 \int_1^{-1} \int_1^{-1} u_1^2 u_2^2 du_1 du_2 \\ &= 2w^2 [\phi_2]_0^{2\pi} \left[\frac{u_1^3}{3}\right]_1^{-1} \left[\frac{u_2^3}{3}\right]_1^{-1} = \frac{16}{9} \pi w^2. \end{aligned} \quad (15.70)$$

Při řešení třetího integrálu navíc využijeme toho, že  $\cos^2 \phi_2 = \frac{1}{2}(1 + \cos(2\phi_2))$ . Integrál členu  $\cos(2\phi_2)$  přes dvě celé periody (pro  $\phi_2$  od nuly do  $2\pi$ , tedy pro  $2\phi_2$  od nuly do  $4\pi$ ) je nula. Stačí tedy spočítat

$$\begin{aligned} \frac{w^2}{4} \int_0^{2\pi} d\phi_2 \int_0^\pi \int_0^\pi \sin^3 \theta_1 \sin^3 \theta_2 d\theta_1 d\theta_2 &= \frac{w^2}{4} \int_0^{2\pi} d\phi_2 \int_0^\pi \int_0^\pi (1 - \cos^2 \theta_1)(1 - \cos^2 \theta_2) \sin \theta_1 d\theta_1 \sin \theta_2 d\theta_2 \\ &= \frac{w^2}{4} \int_0^{2\pi} d\phi_2 \int_1^{-1} \int_1^{-1} (1 - u_1^2)(1 - u_2^2) du_1 du_2 = \frac{w^2}{4} [\phi_2]_0^{2\pi} \left[\frac{u_1 - u_1^3}{3}\right]_1^{-1} \left[\frac{u_2 - u_2^3}{3}\right]_1^{-1} = \frac{8}{9} \pi w^2. \end{aligned} \quad (15.71)$$

Sečtením integrálů získáme  $Z = 8\pi(1 + w^2/3)$ . Podle rovnice 15.65 musíme  $\ln Z$  zderivovat podle  $w$ :

$$\frac{d(\ln Z)}{dw} = \frac{d(\ln(8\pi(1 + w^2/3)))}{dw} = \frac{d(\ln 8\pi + \ln(1 + w^2/3))}{dw} = \frac{2w/3}{1 + w^2/3} \approx \frac{2w}{3}, \quad (15.72)$$

kde jsme v posledním kroku předpokládali, že  $w \ll 1$ . Dosazením do rovnice 15.65 získáme výsledný vztah,

$$\langle U \rangle = -RTw \frac{d(\ln Z)}{dw} = -RT \frac{w^2}{3} = -\frac{2}{3RT} \left( \frac{N_A}{4\pi\epsilon_0} \frac{q_1 q_2}{r} \frac{d_1}{r} \frac{d_2}{r} \right)^2, \quad (15.73)$$

ve kterém jsou červeně zvýrazněny rozdíly od vztahu pro energii dvou iontů.

## A.6 Energie náboje a indukovaného elektrického dipólu

Vnější elektrické síly působící na elektrony vedou ke vzniku *indukovaných elektrických dipólů* i v jinak nepolárních molekulách. Velikost indukovaného dipólu je úměrná elektrické intenzitě (elektrické síle na působící na jednotkový náboj) indukujícího pole. Orientace indukovaného dipólu je dána směrem, kterým na elektrony síla působí. Protože má elektron záporný náboj, je tento směr opačný ke směru elektrické intenzity  $\vec{E}$ . Můžeme tedy psát

$$q\vec{d} = -\alpha\epsilon_0\vec{E}. \quad (15.74)$$

Konstantou úměrnosti obsahuje takzvanou *polarizovatelnost*  $\alpha$ , ochotu elektronů nechat se vnější silou posunout. Permittivita vakua  $\epsilon_0$  je zahrnuta do konstanty úměrnosti z historických důvodů.

Zkusme teď odvodit, jaká je energie molekuly s polarizovatelnými elektrony poblíž náboje  $Q$ . Jako obvykle, energii spočítáme jako práci, kterou by dipól vykonal, kdybychom jej přemístili z místa s nulovou elektrickou intenzitou (tedy z nekonečna) do dané vzdálenosti od náboje  $Q$ . Vyjdeme přitom z dříve odvozeného vztahu pro energii elektrického dipólu v poli náboje  $Q$ :

$$U = \frac{1}{4\pi\epsilon_0} \frac{qdQ}{r^2} \cos\theta = qdE \cos\theta = q\vec{d} \cdot \vec{E}. \quad (15.75)$$

Příspěvek k práci  $dU'$  při každém maličkém posunutí dipólu můžeme tedy spočítat

$$dU' = q\vec{d} \cdot d\vec{E}' = -\alpha\epsilon_0\vec{E}' \cdot d\vec{E}' = -\frac{1}{2}\alpha\epsilon_0 d(E')^2, \quad (15.76)$$

kde  $\vec{E}'$  je elektrická intenzita ve vzdálenosti  $r'$ . Celkovou energii získáme integrací všech příspěvků k práci

$$U = \int_{\infty}^r dU' = -\frac{1}{2} \int_0^E \alpha\epsilon_0 d(E')^2 = -\frac{1}{2} \alpha\epsilon_0 E^2 = -\frac{\alpha\epsilon_0}{2} N_A \left( \frac{1}{4\pi\epsilon_0} \right)^2 \frac{Q^2}{r^2}, \quad (15.77)$$

kde červená barva označuje rozdíly od vztahu pro energii dvou nábojů.

## A.7 Energie permanentního a indukovaného elektrického dipólu

Obdobně můžeme odvodit i energii dipólu indukovaného permanentním elektrickým dipólem. S využitím vztahů pro energii dvou permanentních dipólů

$$dU' = q\vec{d} \cdot d\vec{E}' = -\alpha\epsilon_0\vec{E}' \cdot d\vec{E}' = -\alpha\epsilon_0(E'_x dE'_x + E'_z dE'_z) = -\frac{1}{2}\alpha\epsilon_0(d(E'_x)^2 + d(E'_z)^2) \quad (15.78)$$

a

$$U = \int_{\infty}^r dU' = -\frac{1}{2} \int_0^E \alpha\epsilon_0(d(E'_x)^2 + d(E'_z)^2) = -\frac{1}{2} \alpha\epsilon_0(E_x^2 + E_z^2) = -\frac{\alpha\epsilon_0}{2} N_A \left( \frac{1}{4\pi\epsilon_0} \right)^2 \frac{q^2 d^2}{r^2 r^2} (1 + 3 \cos^2 \theta), \quad (15.79)$$

kde červená barva opět označuje rozdíly od vztahu pro energii dvou nábojů.



### A.8 Energie dvou indukovaných elektrických dipólů

Indukované dipóly vznikají také v nepřítomnosti vnějších elektrických sil v důsledku kmitání elektronů (elektronové hustoty) v molekulách. Fritz London, po kterém je toto silové působení pojmenováno, analyzoval energii indukovaných dipólů takto. Energie dvou interagujících indukovaných dipólů se skládá z energie prvního dipólu  $U_1$ , energie druhého dipólu  $U_2$  a energie jejich vzájemného působení  $U_{12}$ . Protože první dipól tvoří rozkmitaný elektron,  $U_1$  je energie kmitající částice, na kterou působí elektrická síla  $\vec{F}_1$  a jejíž výchylka od rovnovážné polohy popisuje vektor  $\vec{d}_1$ . Ze známých vztahů pro energii pružiny ( $U = kd^2/2$ ) a výchylku  $d$  pružiny o tuhosti  $k$  ( $\vec{F} = -k\vec{d}$ ) můžeme spočítat

$$U_1 = \frac{1}{2}k_1d_1^2 = -\frac{1}{2}F_1d_1 = -\frac{1}{2}\frac{1}{4\pi\epsilon_0}\frac{q_1^2}{d_1} = -\frac{1}{2}\frac{1}{4\pi\epsilon_0}\frac{q_1^3}{q_1d_1}. \quad (15.80)$$

Dále víme, že dipólový moment  $q_1d_1$  je úměrný elektrické intenzitě:  $q_1d_1 = -\alpha_1\epsilon_0E_1$ :

$$U_1 = -\frac{1}{2}\frac{1}{4\pi\epsilon_0}\frac{q_1^3}{q_1d_1} = \frac{1}{2}\frac{1}{4\pi\epsilon_0}\frac{q_1^3}{\alpha_1\epsilon_0E_1} = \frac{q_1^2d_1^2}{2\alpha_1\epsilon_0}. \quad (15.81)$$

Obdobně

$$U_2 = \frac{q_2^2d_2^2}{2\alpha_2\epsilon_0}. \quad (15.82)$$

Energii  $U_{12}$  známe z rovnice 15.57, kterou převedeme ze sférických do kartézských souřadnic:

$$U_{12} = \frac{1}{4\pi\epsilon_0}\frac{q_1q_2}{r}\frac{d_1}{r}\frac{d_2}{r}(\sin\theta_1\sin\theta_2\cos\phi_2 - 2\cos\theta_1\cos\theta_2) = \frac{1}{4\pi\epsilon_0}\frac{q_1q_2}{r^3}(d_{1x}d_{2x} + d_{1y}d_{2y} - 2d_{1z}d_{2z}). \quad (15.83)$$

Když sečteme všechny tři příspěvky k energii a v prvních dvou rozepíšeme  $d_1^2$  a  $d_2^2$  na jednotlivé složky, získáme

$$U = \frac{q_1^2(d_{1x}^2 + d_{1y}^2 + d_{1z}^2)}{2\alpha_1\epsilon_0} + \frac{q_2^2(d_{2x}^2 + d_{2y}^2 + d_{2z}^2)}{2\alpha_2\epsilon_0} + \frac{1}{4\pi\epsilon_0}\frac{q_1q_2}{r^3}(d_{1x}d_{2x} + d_{1y}d_{2y} - 2d_{1z}d_{2z}). \quad (15.84)$$

Pro jednoduchost budeme předpokládat, že náboje i polarizovatelnosti jsou pro oba dipóly stejné:

$$U = \frac{q^2(d_{1x}^2 + d_{1y}^2 + d_{1z}^2)}{2\alpha\epsilon_0} + \frac{q^2(d_{2x}^2 + d_{2y}^2 + d_{2z}^2)}{2\alpha\epsilon_0} + \frac{1}{4\pi\epsilon_0}\frac{q^2}{r^3}(d_{1x}d_{2x} + d_{1y}d_{2y} - 2d_{1z}d_{2z}). \quad (15.85)$$

London zde zavedl substituci

$$x_+ = \frac{d_{1x} + d_{2x}}{\sqrt{2}}, \quad x_- = \frac{d_{1x} - d_{2x}}{\sqrt{2}} \Rightarrow d_{1x} = \frac{x_+ + x_-}{\sqrt{2}}, \quad d_{2x} = \frac{x_+ - x_-}{\sqrt{2}}, \quad (15.86)$$

$$y_+ = \frac{d_{1y} + d_{2y}}{\sqrt{2}}, \quad y_- = \frac{d_{1y} - d_{2y}}{\sqrt{2}} \Rightarrow d_{1y} = \frac{y_+ + y_-}{\sqrt{2}}, \quad d_{2y} = \frac{y_+ - y_-}{\sqrt{2}}, \quad (15.87)$$

$$z_+ = \frac{d_{1z} + d_{2z}}{\sqrt{2}}, \quad z_- = \frac{d_{1z} - d_{2z}}{\sqrt{2}} \Rightarrow d_{1z} = \frac{z_+ + z_-}{\sqrt{2}}, \quad d_{2z} = \frac{z_+ - z_-}{\sqrt{2}}. \quad (15.88)$$

Po dosazení do vztahu pro energii získal vztah

$$U = \frac{q^2}{2\alpha\epsilon_0} \left( \left(1 + \frac{\alpha}{4\pi r^3}\right) (x_+^2 + y_+^2) + \left(1 - \frac{\alpha}{4\pi r^3}\right) (x_+^2 - y_+^2) + \left(1 - \frac{\alpha}{2\pi r^3}\right) z_+^2 + \left(1 + \frac{\alpha}{2\pi r^3}\right) z_-^2 \right), \quad (15.89)$$

který je formálně součtem energií šesti kmitání s výchylkami  $x_+, x_-, y_+, y_-, z_+, z_-$  a tuhostmi

$$\begin{aligned} k_1 &= \frac{q^2}{2\alpha\epsilon_0} \left(1 + \frac{\alpha}{4\pi r^3}\right), & k_2 &= \frac{q^2}{2\alpha\epsilon_0} \left(1 + \frac{\alpha}{4\pi r^3}\right), & k_3 &= \frac{q^2}{2\alpha\epsilon_0} \left(1 - \frac{\alpha}{4\pi r^3}\right), \\ k_4 &= \frac{q^2}{2\alpha\epsilon_0} \left(1 - \frac{\alpha}{4\pi r^3}\right), & k_5 &= \frac{q^2}{2\alpha\epsilon_0} \left(1 - \frac{\alpha}{2\pi r^3}\right), & k_6 &= \frac{q^2}{2\alpha\epsilon_0} \left(1 + \frac{\alpha}{2\pi r^3}\right). \end{aligned}$$

Tuhost pružiny obecně souvisí s vlastní frekvencí kmitání  $\nu$  vztahem  $k = 4\pi^2\nu^2m$ , kde  $m$  je hmotnost. Z toho vyplývá, že vlastní frekvencí kmitání pro naši šesticí můžeme spočítat

$$\nu_j = \frac{1}{2\pi} \sqrt{\frac{k_j}{2m}} = \frac{1}{2\pi} \frac{q}{2} \frac{\sqrt{1+a_j}}{\sqrt{\alpha\epsilon_0 m}} = \frac{\nu_0}{2} \sqrt{1+a_j}, \quad (15.90)$$

kde  $1 + a_j$  jsou výrazy v závorkách v seznamu tuhostí, uvedeném výše. Když vynásobíme tyto frekvence Planckovou konstantou  $h$  ( $6,626 \cdot 10^{-34}$  Js), získáme podle zákonů kvantové mechaniky energie. Výrazy  $\sqrt{1+a_j}$  můžeme zapsat jako mocninnou řadu

$$\sqrt{1+a_j} = 1 + \frac{a_j}{2} - \frac{a_j^2}{8} + \dots \quad (15.91)$$

Pro dostatečně velké vzdálenosti dipólů, kdy  $\alpha \ll r^3$ , můžeme zanedbat vyšší než druhé mocniny. Celkovou energii šesti kmitání můžeme pak spočítat

$$\begin{aligned} U &= \frac{h\nu_0}{2} \left( 2\sqrt{1 + \frac{\alpha}{4\pi r^3}} + 2\sqrt{1 - \frac{\alpha}{4\pi r^3}} + \sqrt{1 - \frac{\alpha}{2\pi r^3}} + \sqrt{1 + \frac{\alpha}{2\pi r^3}} \right) \\ &= \frac{h\nu_0}{2} \left( 2 + \frac{\alpha}{4\pi r^3} - \frac{\alpha^2}{4(4\pi)^2 r^6} + 2 - \frac{\alpha}{4\pi r^3} - \frac{\alpha^2}{4(4\pi)^2 r^6} \right. \\ &\quad \left. + 1 - \frac{\alpha}{4\pi r^3} - \frac{\alpha^2}{2(4\pi)^2 r^6} + 1 + \frac{\alpha}{4\pi r^3} - \frac{\alpha^2}{2(4\pi)^2 r^6} \right) \\ &= 3h\nu_0 - \frac{3h\nu_0}{4} \frac{\alpha^2}{(4\pi)^2 r^6}. \end{aligned} \quad (15.92)$$

První člen, nezávislý na vzdálenosti dipólů  $r$ , popisuje energii samotných indukovaných dipólů. Druhý člen je hledaná energie vzájemného působení indukovaných dipólů  $U_{12}$ . Pokud budou mít indukované dipóly různé polarizovatelnosti, bude energie jejich vzájemné interakce

$$U_{12} = -\frac{3h\nu_0}{2} \frac{\alpha_1\alpha_2}{(4\pi)^2 r^6} \frac{\nu_{0,1}\nu_{0,2}}{\nu_{0,1} + \nu_{0,2}}. \quad (15.93)$$

## B Metody chemické syntézy na pevném nosiči

### B.1 Karbodiimidová metoda syntézy peptidů

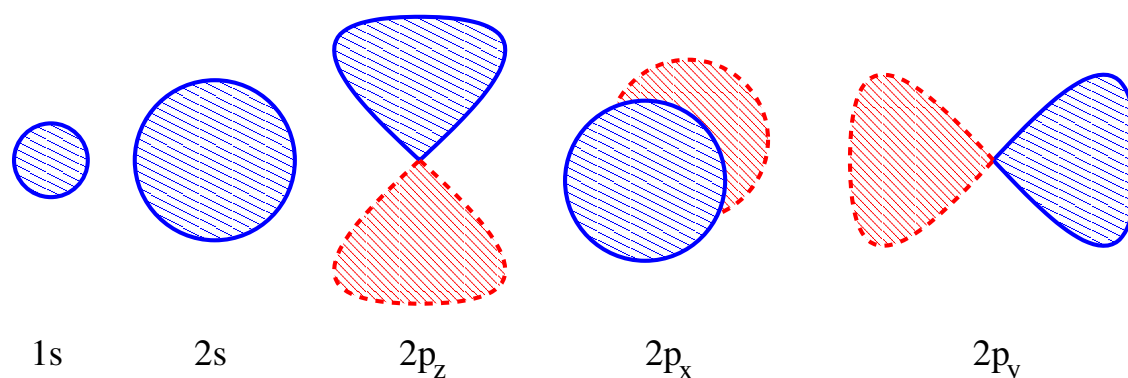
1. Zablokujeme všechny  $\alpha$ -aminoskupiny a všechny reaktivní skupiny postranních řetězců. Jako blokující skupina  $\alpha$ -aminoskupiny se používá 9-fluorenylmethyloxykarbonyl (Fmoc) nebo *tert*-butyloxykarbonyl (tBoc).
2. *C*-koncovou aminokyselinu karboxyskupinou navážeme na pevný nosič. Výhoda použití pevného nosiče je v tom, že rostoucí peptid zůstává připojený k nosiči, takže po každém kroku můžeme snadno vyměnit reakční roztok.
3. Odblokujeme  $\alpha$ -aminoskupinu *C*-koncovou aminokyseliny vázané na nosič<sup>8</sup>.
4. Přidáme předposlední aminokyselinu a aktivujeme její karboxylovou skupinu. Z praktických důvodů se používá dvoustupňová aktivace dicyklohexylkarbodiimidem a *N*-hydroxybenzotriazolem<sup>9</sup>, která výrazně omezí nežádoucí reakce. V prvním kroku dochází k adici karboxylové kyseliny na dicyklohexylkarbodiimid a meziproduct je převeden na ester benzotriazolu. Tento ester reaguje s volnou aminoskupinou poslední aminokyseliny za vzniku peptidové vazby.
5. Předchozí dva kroky opakujeme tolikrát, kolik aminokyselin má obsahovat vzniklý peptid. Pochoitelně přidáváme vždy příslušnou blokovanou aminokyselinu podle sekvence.
6. Nakonec odblokujeme všechny funkční skupiny, uvolníme peptid z nosiče a vyčistíme jej.

### B.2 Fosforamiditová metoda syntézy nukleotidů

1. Zablokujeme všechny aminoskupiny a hydroxyly. Po celou dobu syntézy je třeba chránit aminoskupiny bází a v případě RNA 2'-hydroxyl. Jako blokující skupina 5'-hydroxyly se používá dimethoxytrityl.
2. 3'-koncový nukleotid navážeme přes 3'-hydroxyl na pevný nosič.
3. Kyselinou trichloroctovou odblokujeme 5'-hydroxyl koncového nukleotidu vázaného na nosič.
4. Přidáme předposlední nukleotid s 3'-hydroxylym esterifikovaným *N*-diisopropyl- $\beta$ -kyanoethylfosforamiditem. Tato skupina reaguje v přítomnosti tetrazolu s volným 5'-hydroxylym posledního nukleotidu za vzniku triesteru kyseliny fosforité. Nezareagovaný 5'-hydroxyl nevrátně zablokujeme acetylací.
5. Jodem zoxidujeme triester kyseliny fosforité na triester kyseliny fosforečné.
6. Předchozí tři kroky opakujeme tolikrát s příslušným aktivovaným nukleotidem, kolik nukleotidů má obsahovat vzniklý produkt.
7. Nakonec odblokujeme 5'-hydroxyl prvního nukleotidu, amoniakem odštěpíme poslední nukleotid z nosiče a kyanoethyl z fosfátů, zahřátím uvolníme ostatní chráněné funkční skupiny a oligonukleotid přečistíme.

<sup>8</sup>Použití různých blokujících skupin umožňuje selektivní odblokování jedné funkční skupiny. Například v amoniaku dochází k odštěpení Fmoc, ale ne tBoc, který se naopak odštěpuje v kyselině trifluoroctové.

<sup>9</sup>nebo *N*-hydroxysukcinimidem.



Obrázek 15.9: Schematické znázornění prvních pěti orbitalů atomu vodíku. Modré šrafování odpovídá kladnému znaménku vlnové funkce, červené šrafování odpovídá zápornému znaménku vlnové funkce.

## C Kvantové metody

### C.1 Kvantová mechanika atomů

I ty největší biomakromolekuly se skládají z atomů. Atomy jsou tvořeny jádrem a elektrony. Jak postupujeme k menším a menším částicím, přestávají platit fyzikální zákony tak, jak jsme na ně zvyklí. Pro popis malých částic musíme místo naší klasické mechaniky použít obecnější metodu – *kvantovou mechaniku*. V případě elektronů je kvantové chování již tak výrazné, že nám neumožňuje říci přesně, kde se elektrony v daný okamžik nacházejí. To ale neznamená, že místo výskytu elektronů je zcela náhodné. Pro popis elektronů se používá komplexní matematická funkce (jako proměnné v ní vystupují souřadnice a čas), které fyzici říkají *vlnová funkce*. Místo toho, abychom sledovali, jak se působením vnějších sil mění poloha a rychlosti elektronů, sledujeme jak se mění vlnová funkce<sup>10</sup>. Kvantová mechanika nám pak poskytuje předpisy, jak z vlnové funkce získat informace o fyzikálních veličinách elektronů. Jednou takovou užitečnou informací je pravděpodobnost, že se v daný okamžik elektron vyskytuje v určitém místě. Tuto pravděpodobnost můžeme spočítat jako druhou mocninu (absolutní hodnoty) vlnové funkce  $\psi$ , tedy jako součin vlnové funkce a funkce s ní komplexně sdružené (tedy s opačným znaménkem imaginární části, označuje se hvězdičkou):  $\psi\psi^*$ .

V atomech se elektrony vyskytují jen v určitých *stavech*. Každý stav má určitou energii (energie se tedy nemění spojitě, ale poskakuje od stavu ke stavu) a je popsán určitou vlnovou funkcí. Vlnová funkce elektronu v atomu nebo molekule se nazývá *orbital*. Nemůžeme tedy říci, že se v určitém stavu nachází elektron v určitém místě, ale můžeme říci, že každý stav má určitou *pravděpodobnost* výskytu elektronů a že tuto pravděpodobnost můžeme spočítat pomocí uvedeného předpisu  $\psi\psi^*$ .

Nejjednodušším atomem je atom vodíku. Jeho jádro je tvořeno jediným protonem, v jehož okolí se vyskytuje jediný elektron. Mezi kvantovými chemiky se atom vodíku těší zvláštní oblibě. Vzájemné působení částic lze přesně analyzovat pouze pro dvojici částic<sup>11</sup> a atom vodíku je jedním ze vzácných případů dvojčásticové soustavy, která má nějaký chemický význam. Vlnové funkce atomu vodíku jsou natolik populární, že pronikly až do středoškolských učebnic v grafickém znázornění, uvedeném na obrázku 15.9.

Ostatní atomy mají více než jeden elektron a přesné řešení kvantových rovnic, které poskytují jejich vlnových funkce, není možné. Často se proto používá zjednodušené řešení, nazývané *Hartreeho-Fockova metoda*. Zjednodušením je zde předpoklad, že jednotlivý elektron nevidí své kolegy jako jednotlivé

<sup>10</sup>Tyto změny vlnové funkce popisuje obecný předpis známý pod jménem Schrödingerova rovnice.

<sup>11</sup>Toto platí nejen pro kvantovou mechaniku, ale i pro mechaniku klasickou. Problém tří těles je v astronomii stejně neřešitelný jako problém tří částic v kvantové fyzice.

částice, ale jako jakýsi rozmazaný průměrný záporný náboj<sup>12</sup>. Slůvko „nevidí“ zde označuje elektrostatické odpuzování mezi stejně nabitými částicemi. Z fyzikálního pohledu potom řešíme pohyb jediného elektronu v elektrickém poli tvořeném kladně nabitým jádrem a rozptýleným záporným nábojem<sup>13</sup>. Pokud předpokládáme, že obláček ostatních elektronů kolem jádra má pěkně kulový tvar, dostaneme stejné tvary atomových orbitalů jako v případě atomů vodíku. Chemici se proto neostýchají používat tyto orbitály, které jsou odvozeny z vodíkových, i pro ostatní atomy.

## C.2 Molekulové orbitály

Když sestavíme z atomů molekulu, změníme prostředí, ve kterém se elektrony pohybují. Nepřekvapí proto, že musíme hledat nové stavy elektronů. Tyto nové stavy budou popsány novými vlnovými funkcemi (novými orbitály). Je tedy nutno rozlišovat *atomové orbitály* elektronů v izolovaných atomech a *molekulové orbitály* elektronů v molekulách, jinak bychom nemohli popsat vznik chemických vazeb.

Při hledání molekulových orbitalů pochopitelně narážíme na stejnou překážku, se kterou jsme se setkali u atomů s více elektrony: kvantové rovnice nelze přesně řešit. Můžeme opět použít starý trik a hledat řešení pro jediný elektron, zatímco ostatní elektrony nahradíme jejich elektrickým polem (Hartreeho-Fockova metoda). Molekuly jsou ale složitější než atomy tím, že neobsahují jen více elektronů, ale i více jader. Naštěstí jsou jádra mnohem těžší než elektrony a pohybují se daleko pomaleji. Tyto pomalé pohyby lze bez větších chyb zanedbat, když popisujeme rychlé hemžení elektronů<sup>14</sup>. Abychom mohli tento postup použít, musíme samozřejmě vědět, kde se nacházejí jádra všech atomů v molekule. Jakmile známe polohy jader, můžeme vypočítat tvar elektrického pole, ve kterém se nacházejí elektrony a počítat jejich orbitály.

Pokud chceme spočítat neznámou strukturu, polohy jader atomů pochopitelně neznáme. V tom případě postupujeme tak, že provádíme kvantový výpočet opakovaně s různým rozmístěním jader atomů a hledáme, které rozmístění atomů je nejvýhodnější. Tento postup se nazývá *optimalizace geometrie*, jde vlastně o minimalizaci energie molekuly. O takovýchto minimalizacích budeme hovořit v sekci C.3, takže řešení problému polohy jader zatím odložíme. V následujících řádcích budeme pro jednoduchost předpokládat, že polohy jader známe.

Vraťme se teď k hledání molekulových orbitalů (jak jsme si řekli, předpokládáme že polohy jader známe). Pokud se nám nechce hledat vhodný tvar pole ostatních elektronů pro Hartreeho-Fockovu metodu, můžeme využít toho, co už známe. Kvantová mechanika nás učí, že z určitého počtu atomových orbitalů můžeme získat stejný počet molekulových orbitalů *lineární kombinací*. Srozumitelněji řečeno, jednotlivé vlnové funkce (atomové orbitály) vynásobíme vhodným koeficientem a pak je všechny sečteme. Abychom získali tolik molekulových orbitalů, kolik bylo atomových, musíme sčítání opakovat s různými koeficienty tolikrát, z kolika atomových orbitalů vycházíme. Tato metoda, srdci obyčejného chemika blízká, bývá nazývána *metodou molekulových orbitalů jako lineárních kombinací atomových orbitalů (metodou MO-LCAO)*. Důvod její obliby je jednoduchý, můžeme použít dobře známé atomové orbitály a bez velkého počítání kvalitativně popsat chemické vlastnosti molekuly.

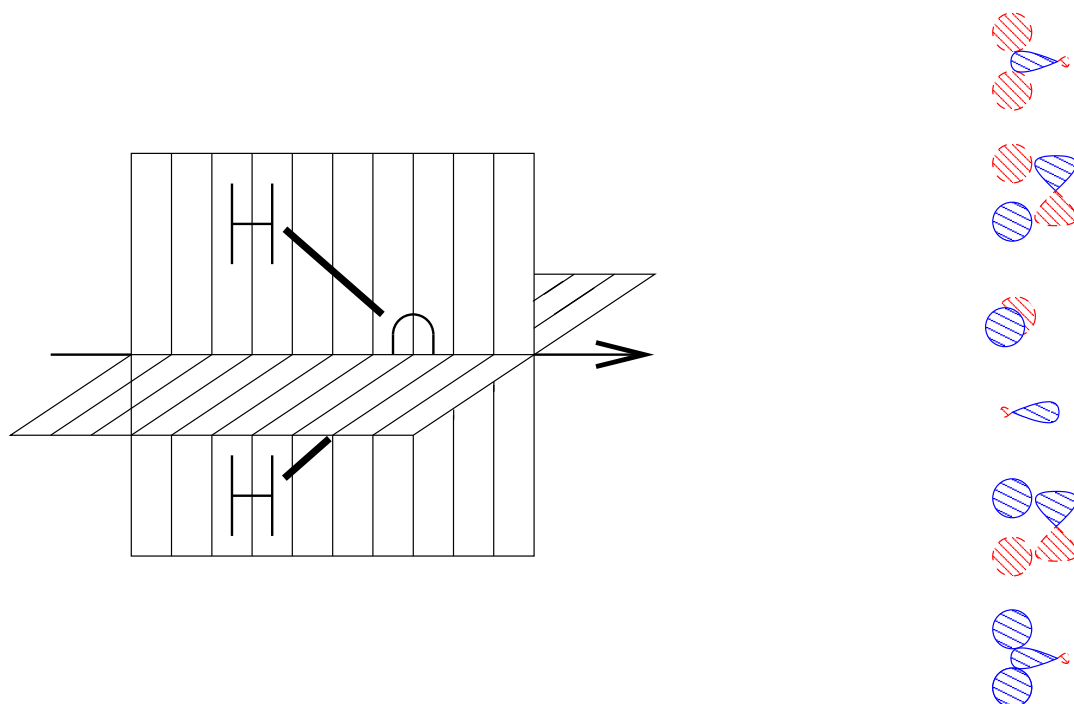
Chceme-li použít metodu MO-LCAO, musíme vědět, jakým způsobem orbitály sčítat a odčítat. Jako vodítko slouží tato obecná pravidla:

- Výsledný molekulový orbital musí odpovídat symetrii (tvaru) molekuly.
- Zpravidla stačí použít *valenční* atomové orbitály (atomové orbitály popisující elektrony, které mají v daném atomu nejvyšší energii, jsou tedy nejvolněji vázány k jádru a nejochotněji mění svůj

<sup>12</sup>Zjednodušení víceelektronového problému na jednelektronový znamená, že se elektrony budou chovat, jako by o sobě navzájem „nevěděli“. Ve skutečnosti ovšem elektrony nejsou nezávislé, jejich výskyt v atomu je *korelovan*. Chyba, které se dopouštíme tím, že tuto korelaci zanedbáme, se projeví zvýšením energie o příspěvek zvaný *korelační energie*.

<sup>13</sup>V literatuře se můžete setkat také s názvem *metoda konzistentního pole* neboli *metoda SCF* (z anglického self-consistent field). Toto označení vychází ze způsobu, jakým se hledá popis zprůměrovaného působení ostatních elektronů.

<sup>14</sup>Toto zjednodušení se nazývá Bornova-Oppenheimerova aproximace.



Obrázek 15.10: Osa symetrie a roviny symetrie molekuly vody. Vpravo je schematicky znázorněno šest molekulových orbitalů, které vyhovují symetrii molekuly vody. Orbital s nejnižší energií je nakreslen dole, orbital s nejvyšší energií nahoře. Barevné šrafování ukazuje znaménko vlnové funkce.

stav za tvorby chemické vazby).

- Atomové orbitály, které sčítáme, musí mít vhodný tvar a vzájemnou orientaci, musí se co nejvíce překrývat.
- Sčítáme anebo odečítáme orbitály, které popisují elektrony s přibližně stejnou energií.

Jako příklad si můžeme uvést hledání molekulových orbitalů vody (obrázek 15.10). Chceme získat orbitály, které věrně zobrazí tvar molekuly vody. Molekula vody je symetrická vzhledem k otočení o  $180^\circ$  kolem osy, která půlí úhel  $\angle \text{H-O-H}$ . Navíc má dvě zrcadlové roviny symetrie které se protínají ve zmíněné ose: v jedné leží všechny tři atomy a druhá je na ni kolmá. Valenční atomové orbitály, které máme k dispozici, jsou dva orbitály  $1s$  atomů vodíku, jeden orbital  $2s$  atomu kyslíku a tři orbitály  $2p$  atomu kyslíku. Když tyto orbitály přeložíme přes sebe tak, aby odpovídaly tvaru molekuly vody a aby se co nejvíce překrývaly, získáme šest molekulových orbitalů.

Atomové orbitály nemívají pro sestavování molekulových orbitalů nejvýhodnější symetrii. Proto chemici s oblibou používají jinou sadu orbitalů, než poskytuje atom vodíku. Tyto orbitály, zvané *hybridní*, lze opět získat lineární kombinací. Tentokrát ovšem sčítáme atomové orbitály téhož atomu mezi sebou. Jestliže vodíkové atomové orbitály dobře odrážely kulovou symetrii jádra vodíku, hybridní orbitály slouží jako polotovary, které jsou již konstruovány tak, aby odpovídaly okolí atomu v typických sloučeninách. Například atom uhlíku, který se ocitne v alifatické sloučenině, bývá symetricky obklopen čtyřmi sousedními atomy, které leží ve vrcholech pomyslného čtyřstěnu. Pro vytvoření molekulových orbitalů alifatických sloučenin si proto chemik přichystá vhodný hybridní orbital. Vezme jeden orbital  $2s$  a tři orbitály  $2p$  uhlíku a posčítá je tak, aby získal čtyři navlas stejné orbitály, které budou mít

stejnou symetrii jako čtyřstěn. Pro látky jako karboxylové kyseliny nebo alkeny, ve kterých je uhlík obklopen jen třemi atomy v rozích pomyslného trojúhelníka, si připraví jinou sadu tří orbitalů, které mají požadovanou symetrii. Podrobnější popis postupu tvorby molekulových orbitalů pro biologicky zajímavý příklad peptidové vazby můžete najít v dodatku D.

### C.3 Metody *ab initio*

Po kvalitativním popisu molekulových orbitalů přistupme k vlastnímu kvantitativnímu výpočtu orbitalů, které popisují biologicky zajímavé molekuly. Nejserióznější způsob, jakým lze tyto výpočty provádět, je odmítnout všechny nápovědy, které nám nabízí experiment a zkušenosti popsané v literatuře. Takový výpočet se hrdě opírá jen o obecné zákony kvantové mechaniky a základní fyzikální konstanty a označuje se jako přístup *ab initio*<sup>15</sup>.

Jak jsme si již řekli, přesný výpočet je možný jen pro nejjednodušší případy. Proto metody *ab initio* používají nejrůznějších zjednodušení a přibližná řešení hledají pomocí výkonných počítačů.

Metody *ab initio* můžeme použít k výpočtu energie molekuly v určité konformaci (takzvaný *single-point* výpočet), nebo k hledání energeticky nejvýhodnější konformace, případně k výpočtu jiných vlastností molekuly. Hledání stavu molekuly s nejnižší energií (*minimalizace energie*), které nás zajímá především, probíhá podle zásad uvedených v sekci 10.8.

1. Vypočteme energii v určité konformaci<sup>16</sup>.
2. Vypočteme síly, které působí na atomy v dané konformaci. Sílu lze spočítat jako derivaci energie v daném bodě v různých směrech, tedy *gradient*.
3. Necháme působit vypočtené síly na atomy a vypočteme, jak tyto síly změní konformaci.
4. Znovu vypočteme energii a síly, teď již pro pozměněnou konformaci.
5. Srovnáme výsledky kroku 1 s výsledky kroku 4. Pokud jsou splněna konvergenční kritéria, počítání končí. Pokud nalezneme ve výsledcích rozdíl, opakujeme kroky 2 až 4.

Během výpočtu *ab initio* je nutné počítat vlnové funkce (orbitaly), které popisují chování elektronů v molekule. V sekci C.2 jsme si ukázali, že molekulové orbitaly lze získat lineární kombinací atomových orbitalů. Nemusíme ovšem vycházet z orbitalů atomu vodíku, molekulové orbitaly můžeme vypočítat lineární kombinací nejrůznějších sad orbitalů. Taková sada funkcí, jejíž kombinací lze získat požadovaný orbital, se matematicky nazývá *báze*. Volba vhodné báze funkcí je pro výpočet velmi důležitá.

Nejpřirozenější bázi funkcí pro výpočty *ab initio* jsou staré dobré atomové orbitaly. Jde o vlnové funkce, které exponenciálně klesají se vzdáleností od jádra<sup>17</sup>. Funkce vycházející ze zjednodušeného popisu atomových orbitalů se nazývají *orbitaly Slaterova typu* (*STO*, Slater type orbitals). Jak vidíme na obrázku 15.11, exponenciální křivka je pro nulovou vzdálenost od jádra „špičatá“ (má nespojitou první derivaci). Počítání s orbitaly *STO* je ale poměrně složité. Z praktických důvodů se proto jako báze používají vlnové funkce, které mají v exponentu druhou mocninu vzdálenosti. Exponenciální křivku tak nahrazujeme Gaussovou křivkou. Samozřejmě tak získáme vlnové funkce na hony vzdálené atomovým orbitalům a nemůžeme očekávat, že tyto funkce budou dobře popisovat naši molekulu. Když ale sečteme několik Gaussových křivek, získáme křivku, která má tvar velmi podobný exponenciále. Takto dospějeme k popisu molekul, který je vhodný pro výpočty, za cenu, že báze obsahuje větší počet funkcí zvaných

<sup>15</sup>Od začátku.

<sup>16</sup>Metody *ab initio* ovšem nejsou omezeny na konformační změny, můžeme pomocí nich popsat i chemické reakce.

<sup>17</sup>Přesněji řečeno, u funkcí vypočítaných z rovnic kvantové mechaniky jde o exponenciální závislost na vzdálenosti  $r$  vynásobenou mnohočlenem, který obsahuje celočíselné mocniny vzdálenosti (například  $(a_0 + a_1r + a_2r^2)e^{-br}$ , kde  $a_0, a_1, a_2, b$  jsou konstanty). Pro vytvoření báze se mnohočleny nahrazují zjednodušeným výrazem  $r^a e^{-br}$ . Nám bude stačit, když se zaměříme na exponenciální část zjednodušených funkcí.

*orbitály Gaussova typu (GTO, Gauss type orbitals)*. Obrázek 15.11 ukazuje jednoduchý příklad upravené báze pro atom vodíku, bázi označovanou STO-3G. V ní je jeden orbital Slaterova typu (1s) nahrazen třemi orbitály Gaussova typu. Jiným příkladem je báze 6-31G. Označení ukazuje, že orbital 1s (STO) je třeba nahradit jedním GTO, který vytvoříme součtem šesti Gaussových křivek. Každý ze čtyř orbitalů STO, které odpovídají 2s a 2p orbitalům, je třeba nahradit dvěma GTO, z nichž jeden je složen ze tří Gaussových křivek a druhý je tvořen jedinou Gaussovou křivkou.

Popsané báze sice dobře odrážejí nejvýhodnější rozložení elektronů, nepopisují ale *polarizaci* (přesuny elektronů v elektromagnetickém poli). Proto se k bázi přidávají ještě takzvané *polarizační funkce*, které odpovídají vyšším atomovým orbitalům<sup>18</sup>. Mluvíme pak o *rozšířených bázích*, jejichž „nadstandardní vybavení“ se v označení báze projevuje jako hvězdička. Například báze 6-31G\* přidává sadu orbitalů d jako polarizační funkci uhlíků, zatímco báze 6-31G\*\* přidává navíc orbitály p jako polarizační funkce vodíku.

Až dosud jsme zmiňovali pouze výpočty založené na Hartreeho-Fockově metodě, založené na zjednodušujícím popisu jednoho elektronu v poli nahrazujícím ostatní elektrony. Takové zjednodušení vnáší do výpočtu chyby, protože každý elektron ve skutečnosti „cítí“ nejen průměrný náboj ostatních elektronů, ale také to, kde se ostatní elektrony nacházejí. Tyto chyby je třeba opravovat složitými výpočty. V současné době lze tento poměrně přímočarý postup použít jen pro soubory několika málo atomů. Musíme totiž opravovat zanedbání toho, že „každý ví o každém“, takže počet opravných výpočtů prudce roste pro větší molekuly (téměř s druhou mocninou počtu atomů). Existují ale metody založené na zcela odlišném principu, který je označován jako *teorie funkcionálu hustoty (DFT, density functional theory)*. Tato teorie říká, že popis elektronů, které na sebe v molekule vzájemně působí, lze nahradit výhodnějším popisem. Elektrony můžeme popsat jako částice bez náboje, které na sebe nepůsobí. Toho, aby se takové neutrální částice chovaly stejně jako elektrony v molekule, dosáhneme tím, že je umístíme do zvláštního elektrického pole. Teorie DFT dokazuje, že vždycky existuje pole, kterým lze vzájemné působení elektronů popsat. Toto pole ovšem neznáme, musíme je hledat pomocí složitých výpočtů.

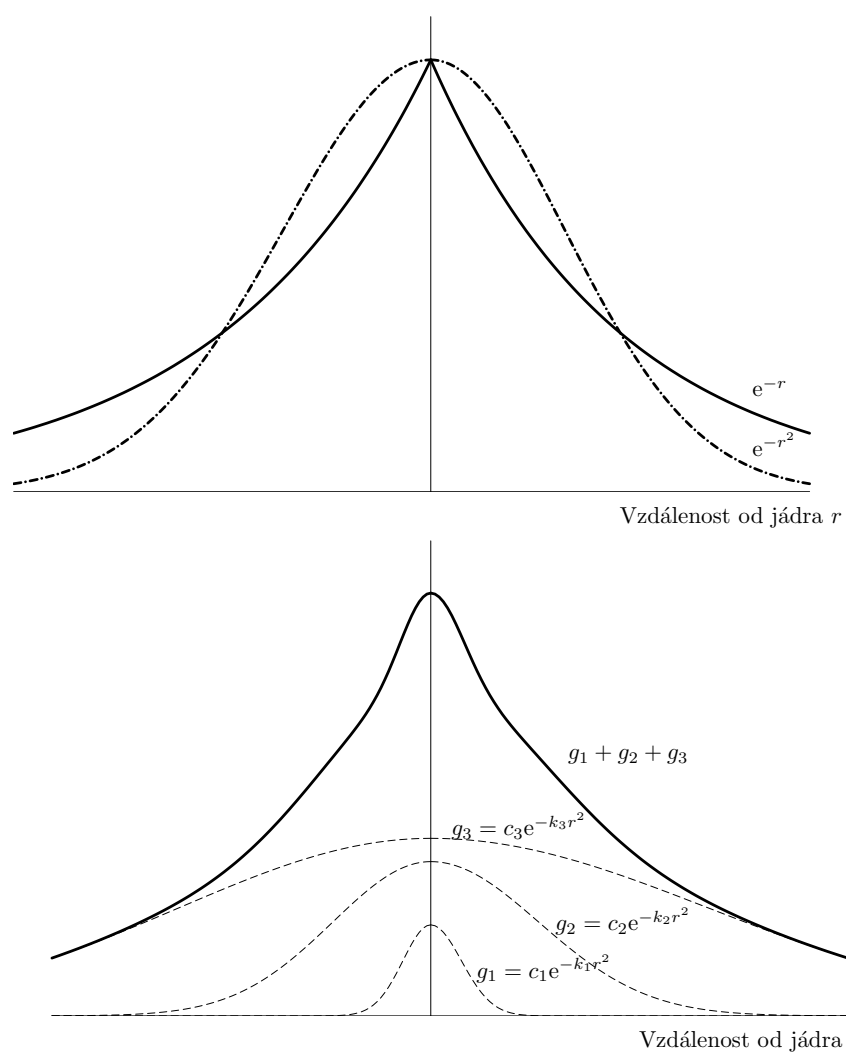
Na první pohled se zdá, že jsme si moc nepomohli. Stejně jako v Hartreeho-Fockově metodě musíme složitě dodatečně zahrnovat to, že elektrony o sobě „vědí“. Je tu ale zásadní rozdíl. Ono tajemné elektrické pole metody DFT je jen *jedno* pro celou molekulu, *stejně pro všechny elektrony*. Opravy tedy nemusíme provádět způsobem „každý s každým“, jako v Hartreeho-Fockově metodě, ale jenom jednou pro každý elektron. Pomocí metod DFT lze proto dnes provádět *ab-initio* kvantové výpočty i s několika desítkami atomů.

Dalším důležitým rozdílem mezi Hartreeho-Fockovou metodou a metodou DFT je způsob, jakým získáme kýžený výsledek – energeticky výhodné rozložení elektronů v prostoru, které udává tvar molekuly. Hartreeho-Fockova metoda je založena na tom, že z vlnových funkcí můžeme spočítat energii i rozložení elektronů v prostoru. Hledáme proto takovou vlnovou funkci, která odpovídá nejnižší energii. Z této vlnové funkce pak můžeme vypočítat rozložení elektronů (*elektronovou hustotu*) v základním stavu molekuly a jiné užitečné veličiny. Metoda DFT postupuje opačně. Hledáme přímo takové rozložení elektronů v prostoru (takovou funkci elektronové hustoty), která odpovídá nejnižší energii elektronů v tom tajemném elektrickém poli, které teorie DFT zavádí. Pokud chceme získat i vlnovou funkci, můžeme ji vypočítat zpětně z elektronové hustoty. Tím se dostáváme i k názvu metody. Mluvíme o funkcionálu hustoty, protože energii vyjadřujeme jako funkcionál funkce elektronové hustoty<sup>19</sup>.

<sup>18</sup> Orbitály, které odpovídají polarizačním funkcím, mají vyšší vedlejší kvantové číslo  $l$ , než odpovídá zaplněným orbitalům daného atomu.

<sup>19</sup> *Funkce* je matematický předpis, jak přiřadit hodnotu nějaké veličiny (v našem případě energie) určitým hodnotám proměnných (v našem případě souřadnic). *Funkcionál* je matematický předpis, jak přiřadit hodnotu nějaké veličiny (v našem případě energie) určité funkci jako celku (v našem případě funkci elektronové hustoty).





Obrázek 15.11: Graf vlevo ukazuje srovnání tvaru exponenciály (plná čára) a Gaussovy funkce (čerchovaná čára). Graf vpravo ilustruje, jak lze přibližně nahradit exponenciálu součtem tří Gaussových funkcí.

## C.4 Semiempirické metody

Náročnost metod *ab initio* vedla k mnoha snahám o zjednodušení výpočtu. Při použití těchto metod se vzdáváme snahy o fyzikálně přesný popis molekuly a raději hledáme cestičky, jak některou část výpočtu nahradit dosazením nějakého experimentálního údaje, nebo jak se určitému výpočtu zcela vyhnout.

Prvním zjednodušením semiempirických metod je, že výpočet omezují jen na elektrony s nejvyšší energií, které se podílejí na tvorbě chemických vazeb (*valenční elektrony*).

Druhým zjednodušením je, že omezíme počet funkcí, ze kterých orbitaly skládáme (zjednodušíme *bázi*). Obvykle se jako báze používají orbitaly s a p Slaterova typu.

Další zjednodušení se týká překryvů atomových orbitalů, ze kterých vyrábíme orbitaly molekulové. V sekci C.2 jsme si říkali, že pro tvorbu molekulových orbitalů (pro tvorbu chemické vazby) je nutný co nejlepší překryv atomových orbitalů. V semiempirických metodách se některé méně významné překryvy zanedbávají. Podle toho, co všechno zanedbáme, mluvíme o metodách<sup>20</sup> NDDO, INDO, CNDO. Nejobtížnější je metoda NDDO, naopak nejdřastičtější je metoda CNDO.

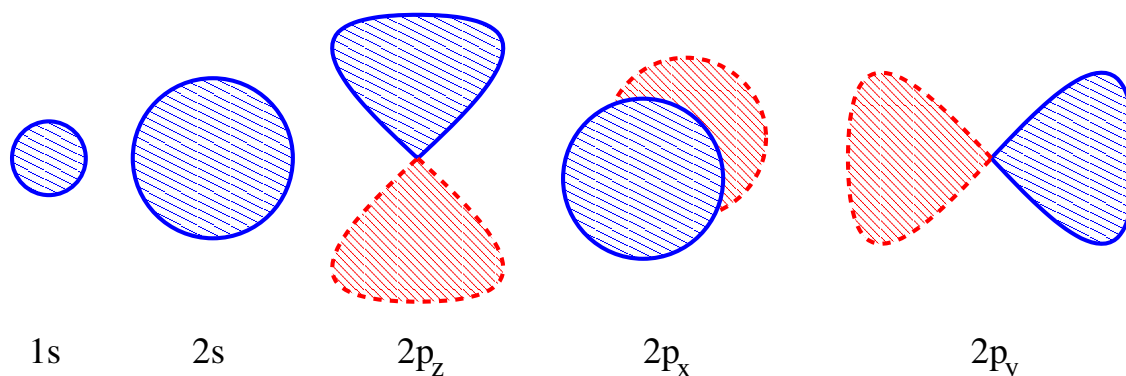
Zanedbáním překryvů se dopouštíme chyby ve výpočtu. Tuto chybu můžeme trochu napravit tím, že nahradíme číselné hodnoty překryvů nějakým opravným parametrem. Tento opravný parametr můžeme buď odhadnout z tvaru atomových orbitalů, nebo vypočítat z výsledků různých experimentů<sup>21</sup>.

Zvláštní rodinu metod tvoří takzvané *Hückelovy* metody, které zavádějí opravné parametry ne místo vypočítaných překryvů, ale přímo místo hodnot potenciální energie v rovnicích kvantové mechaniky. Parametry v Hückelových metodách vlastně odpovídají ionizačním energiím atomů (energiím, které jsou potřeba k odtržení elektronu z atomu). Proto je hledání parametrů v tomto případě velmi snadné a takzvaná *rozšířená Hückelova metoda* (*EHM, Extended Hückel Method*) poskytuje parametry pro všechny prvky periodické tabulky. EHM jsou nejlepší pro výpočty vlastností molekul obsahující atomy přechodných prvků, naopak se nepříliš hodí pro optimalizaci tvaru molekuly.

Semiempirické metody umožňují kvantové výpočty pro molekuly skládající se ze stovek atomů. Je možné vypočítat molekulové orbitaly, energie jednotlivých stavů, rozložení náboje v molekule, řady vazeb a spektrální vlastnosti molekul, které zmíníme v kapitole 11. Semiempirické metody mohou zahrnovat i vliv rozpouštědla na chování molekuly. Daň za zjednodušení výpočtů ale může být velká. Zahrnutí překryvů orbitalů do empirických parametrů a úplné zanedbání vnitřních orbitalů mohou vést k závažným chybám ve výpočtu některých veličin.

<sup>20</sup>Názvy metod jsou anglické zkratky, které popisují, jak daleko jde která metoda v zanedbání překryvů: neglect of diatomic differential overlaps, intermediate neglect of differential overlaps, complete neglect of differential overlaps.

<sup>21</sup>Zavedením opravných parametrů vznikla modifikovaná metoda NDDO (označovaná MNDO). MNDO byla dále vylepšena a nazvána Austin Model 1 (podle University of Austin). Některé hodnoty parametrů v těchto metodách bylo možné vypočítat z různých experimentů. Za zbývající byla zkusmo dosazována různá čísla. Čísla, se kterými metoda poskytovala nejlepší výsledky, byla použita jako parametry, pro které neexistovaly experimentální údaje. Později byl postup hledání zbývajících parametrů zopakován automaticky s použitím počítačové optimalizace. Metoda, která používá tyto parametry nalezené počítačem, se nazývá MNDO-PM3 (parametrická metoda číslo tři).



Obrázek 15.12: Schematické znázornění atomových orbitalů.

## D Peptidová vazba jako příklad chromoforu

Pokusme se podívat podrobněji na to, jak se interakce se světlem projeví na malé části molekuly proteinu, na peptidové vazbě. Začneme tím, že si popíšeme jednotlivé stavy elektronů této skupiny. Víme, že tyto stavy popisujeme pomocí vlnových funkcí. Abychom takovému popisu porozuměli, zkusíme si (dosti zjednodušeně) odvodit vlnové funkce peptidové vazby.

### D.1 Atomové orbitály

Ze střední školy jistě dobře znáte schematické znázornění vlnových funkcí elektronů atomu vodíku v prostoru<sup>22</sup> (Obrázek 15.12).

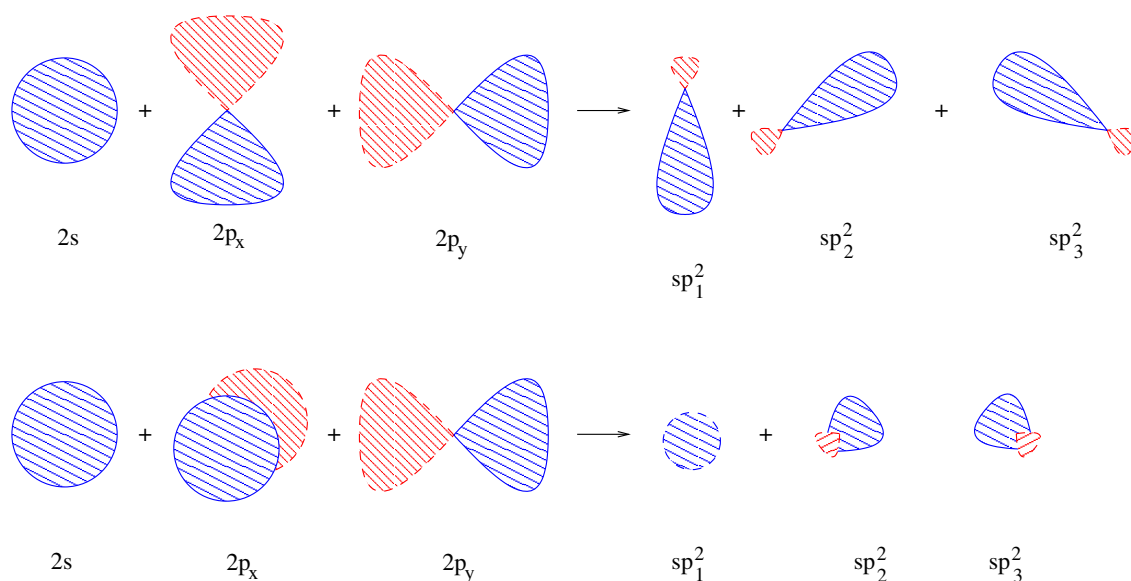
Tyto obrázky jsou vlastně prostorovou obdobou izobar na meteorologických mapách. Vyznačují místa v prostoru, kde má vlnová funkce stejnou hodnotu. Obvykle se pro každý orbital znázorňují jen dvě, „izobary“, pro jednu kladnou (zde znázorněna modře) a jednu zápornou (zde znázorněna červeně) hodnotu vlnové funkce. V peptidové vazbě máme jeden vodík, pro který nám bude stačit nejnižší orbital 1s. Dále máme v peptidové vazbě po jednom atomu uhlíku, kyslíku a dusíku. Když budeme předpokládat, že s dvojicí elektronů v nejnižším stavu 1s se nic neděje, stačí nám pro každý z těchto atomů čtyři orbitály (jeden 2s a tři 2p). To je dohromady třináct orbitalů. Pokud budeme chtít popsat vazby k sousedním  $\alpha$ -uhlíkům, budeme od každého z těchto  $\alpha$ -uhlíků potřebovat ještě jeden orbital, tedy celkem 15 atomových orbitalů.

### D.2 Hybridní orbitály

Z popsaných 15 atomových orbitalů chceme sestavit molekulové orbitály, kterých musí být také patnáct. Nové orbitály můžeme spočítat jako lineární kombinaci atomových orbitalů. To znamená, že každý atomový orbital vynásobíme určitým koeficientem a tyto příspěvky sečteme. Je výhodné tento výpočet dělat nadvakrát, nejprve sestojit takzvané *hybridní* orbitály, ze kterých se nám výsledné molekulové orbitály už budou dobře počítat.

Začneme s karbonylovým uhlíkem. Budeme předpokládat, že peptidová vazba leží v rovině  $xy$ . V této rovině tvoří uhlík tři  $\sigma$ -vazby s okolními atomy. Pro jejich popis si připravíme tři hybridní orbitály, které

<sup>22</sup> Není důvod, proč bychom si nemohli uvést definice těchto orbitalů (bez zahrnutí relativistických vlivů) jako funkcí souřadnic  $x, y, z$ :  $1s = \frac{1}{\sqrt{\pi}} \left(\frac{Z}{a}\right)^{3/2} e^{-\frac{Zr}{a}}$ ,  $2s = \frac{1}{\sqrt{2\pi}} \left(\frac{Z}{a}\right)^{3/2} \left(2 - \frac{Zr}{a}\right) e^{-\frac{Zr}{2a}}$ ,  $2p_x = \frac{1}{4\sqrt{2\pi}} \left(\frac{Z}{a}\right)^{5/2} \frac{x}{r} e^{-\frac{Zr}{2a}}$ ,  $2p_y = \frac{1}{4\sqrt{2\pi}} \left(\frac{Z}{a}\right)^{5/2} \frac{y}{r} e^{-\frac{Zr}{2a}}$ ,  $2p_z = \frac{1}{4\sqrt{2\pi}} \left(\frac{Z}{a}\right)^{5/2} \frac{z}{r} e^{-\frac{Zr}{2a}}$ , kde  $r^2 = x^2 + y^2 + z^2$ ,  $Z$  je počet protonů v jádře a  $a$  je poloměr dráhy elektronu nejbližšího jádru v klasickém Bohrově popisu.

Obrázek 15.13: Orbitaly  $sp^2$ 

budeme označovat  $sp^2$ . Vypočítáme je ze tří atomových orbitalů  $2s$ ,  $2p_x$ ,  $2p_y$ :

$$sp_1^2 = \sqrt{\frac{1}{3}}2s + \sqrt{\frac{2}{3}}2p_x \quad (15.94)$$

$$sp_2^2 = \sqrt{\frac{1}{3}}2s - \sqrt{\frac{1}{6}}2p_x + \sqrt{\frac{1}{2}}2p_y \quad (15.95)$$

$$sp_3^2 = \sqrt{\frac{1}{3}}2s - \sqrt{\frac{1}{6}}2p_x - \sqrt{\frac{1}{2}}2p_y \quad (15.96)$$

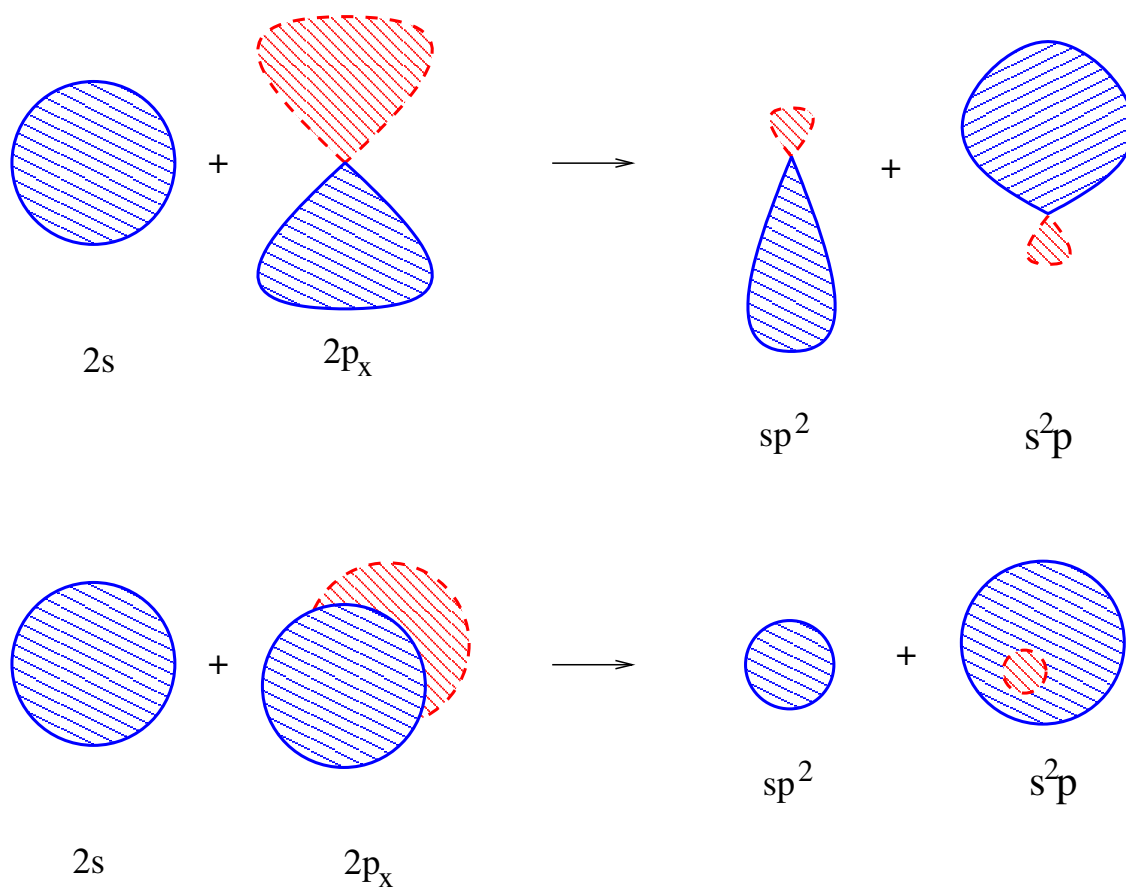
Graficky si výpočet<sup>23</sup> orbitalů  $sp^2$  můžeme znázornit tak, jak ukazuje obrázek 15.13.

V horní části obrázku jsou znázorněny orbitaly při pohledu shora na rovinu  $xy$ , v dolní části při stejném úhlu pohledu jako na předchozím obrázku. Podobným způsobem můžeme vytvořit i tři  $sp^2$  orbitaly dusíku. Trochu jiné orbitaly vypočítáme pro kyslík. Spolu s orbitalem  $2s$  vezmeme pouze jeden orbital  $2p$  ( $2p_x$ ), ze kterých sestrojíme dva orbitaly:  $sp^2$  a  $s^2p$ .

<sup>23</sup> Pokud vám není jasné, proč jsme za koeficienty zvolili právě ty podivné odmocniny, vzpomeňte, jak jsme definovali pravděpodobnost elektronu v určitém místě v prostoru:  $\psi\psi^*$  (pro naše orbitaly zvolené jako reálné funkce můžeme psát jednoduše  $\psi^2$ ). Pokud bude elektron ve stavu popsaném orbitalem  $\psi$ , musí platit, že součet pravděpodobností výskytu ve všech místech prostoru dá jedničku (někde elektron být musí). Matematicky řečeno,  $\int_V \psi^2 dV = 1$  ( $V$  značí, že integrujeme přes celý objem). Zkusme si to třeba pro orbital  $sp^2$

$$1 = \int_V (sp_1^2)^2 dV = \int_V \left( \sqrt{\frac{1}{3}}2s + \sqrt{\frac{2}{3}}2p_x \right)^2 dV = \frac{1}{3} \int_V (2s)^2 dV + \frac{2}{3} \int_V (2p_x)^2 dV + \frac{2\sqrt{2}}{3} \int_V 2s2p_x dV = \frac{1}{3} + \frac{2}{3} + 0 \quad (15.97)$$

Protože poslední integrál je nulový ( $2s2p_x$  je lichá funkce, to je dáno symetrií orbitalů), můžeme říci, že pravděpodobnost výskytu v  $sp_1^2$  je dána z jedné třetiny pravděpodobností výskytu v  $2s$  a ze dvou třetin pravděpodobností výskytu v  $2p_x$ .

Obrázek 15.14: Orbitaly s<sup>2</sup>p

$$sp^2 = \sqrt{\frac{1}{3}}2s + \sqrt{\frac{2}{3}}2p_x \quad (15.98)$$

$$s^2p = \sqrt{\frac{2}{3}}2s - \sqrt{\frac{1}{3}}2p_x \quad (15.99)$$

$$(15.100)$$

Grafické znázornění je obdobné předchozímu (obrázek 15.14).

A nakonec budeme potřebovat jeden sp<sup>3</sup> orbital od každého α-uhlíku, který vypočteme z orbitalu 2s a ze tří orbitalů 2p (takto získáme čtyři orbitaly sp<sup>3</sup>, zbývající tři bychom mohli použít na popis dalších tří σ-vazeb α-uhlíku)

$$sp_1^3 = \frac{1}{2}2s + \frac{1}{2}2p_x + \frac{1}{2}2p_y + \frac{1}{2}2p_z \quad (15.101)$$

$$sp_2^3 = \frac{1}{2}2s + \frac{1}{2}2p_x - \frac{1}{2}2p_y - \frac{1}{2}2p_z \quad (15.102)$$

$$sp_3^3 = \frac{1}{2}2s - \frac{1}{2}2p_x - \frac{1}{2}2p_y + \frac{1}{2}2p_z \quad (15.103)$$

$$sp_4^3 = \frac{1}{2}2s - \frac{1}{2}2p_x + \frac{1}{2}2p_y - \frac{1}{2}2p_z \quad (15.104)$$

### D.3 Molekulové orbitály

Konečně jsme připraveni vypočítat molekulové orbitály peptidové vazby. Podívejme se nejprve na orbitály  $\sigma$ . V peptidové skupině máme 5  $\sigma$ -vazeb, budeme tedy počítat 10  $\sigma$  orbitalů (5 vazebných a 5 antivazebných). K výpočtu použijeme jeden 1s orbital vodíku, sedm  $sp^2$  orbitalů (po třech z uhlíku a dusíku a jeden kyslíkový) a dva  $sp^3$  orbitály sousedních  $\alpha$ -uhlíků.

$$\sigma(C_1^\alpha C) = \sqrt{\frac{1}{2}}sp^3(C_1^\alpha) + \sqrt{\frac{1}{2}}sp_3^2(C) \quad (15.105)$$

$$\sigma^*(C_1^\alpha C) = \sqrt{\frac{1}{2}}sp^3(C_1^\alpha) - \sqrt{\frac{1}{2}}sp_3^2(C) \quad (15.106)$$

$$\sigma(CO) = \sqrt{\frac{1}{2}}sp_1^2(C) + \sqrt{\frac{1}{2}}sp^2(O) \quad (15.107)$$

$$\sigma^*(CO) = \sqrt{\frac{1}{2}}sp_1^2(C) - \sqrt{\frac{1}{2}}sp^2(O) \quad (15.108)$$

$$\sigma(CN) = \sqrt{\frac{1}{2}}sp_2^2(C) + \sqrt{\frac{1}{2}}sp_2^2(N) \quad (15.109)$$

$$\sigma^*(CN) = \sqrt{\frac{1}{2}}sp_2^2(C) - \sqrt{\frac{1}{2}}sp_2^2(N) \quad (15.110)$$

$$\sigma(NH) = \sqrt{\frac{1}{2}}sp_1^2(N) + \sqrt{\frac{1}{2}}1s(H) \quad (15.111)$$

$$\sigma^*(NH) = \sqrt{\frac{1}{2}}sp_1^2(N) - \sqrt{\frac{1}{2}}1s(H) \quad (15.112)$$

$$\sigma(NC_2^\alpha) = \sqrt{\frac{1}{2}}sp_3^2(N) + \sqrt{\frac{1}{2}}sp^3(C_2^\alpha) \quad (15.113)$$

$$\sigma^*(NC_2^\alpha) = \sqrt{\frac{1}{2}}sp_3^2(N) - \sqrt{\frac{1}{2}}sp^3(C_2^\alpha) \quad (15.114)$$

Graficky je tvorba orbitalů  $\sigma$  a  $\sigma^*$  ze dvou  $sp^2$  orbitalů znázorněna na obrázku 15.15.

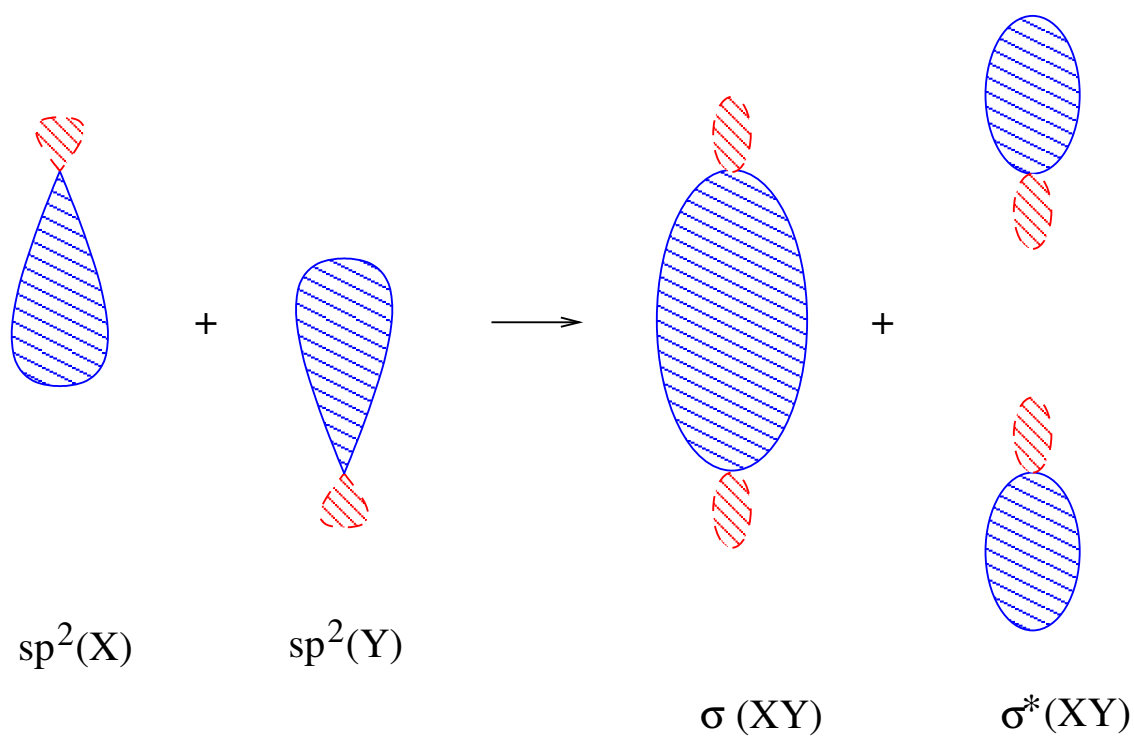
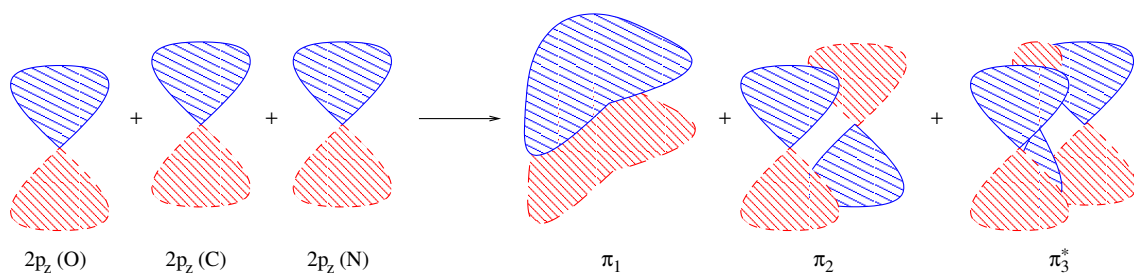
Všimněte si, že při tvorbě hybridních atomových orbitalů jsme kombinovali orbitály, které vycházely z jednoho atomu, tedy z jednoho bodu v prostoru. Teď, když tvoříme molekulové orbitály, kombinujeme atomové orbitály různých atomů (na obrázku označených X a Y), tedy vycházející ze dvou bodů v prostoru.

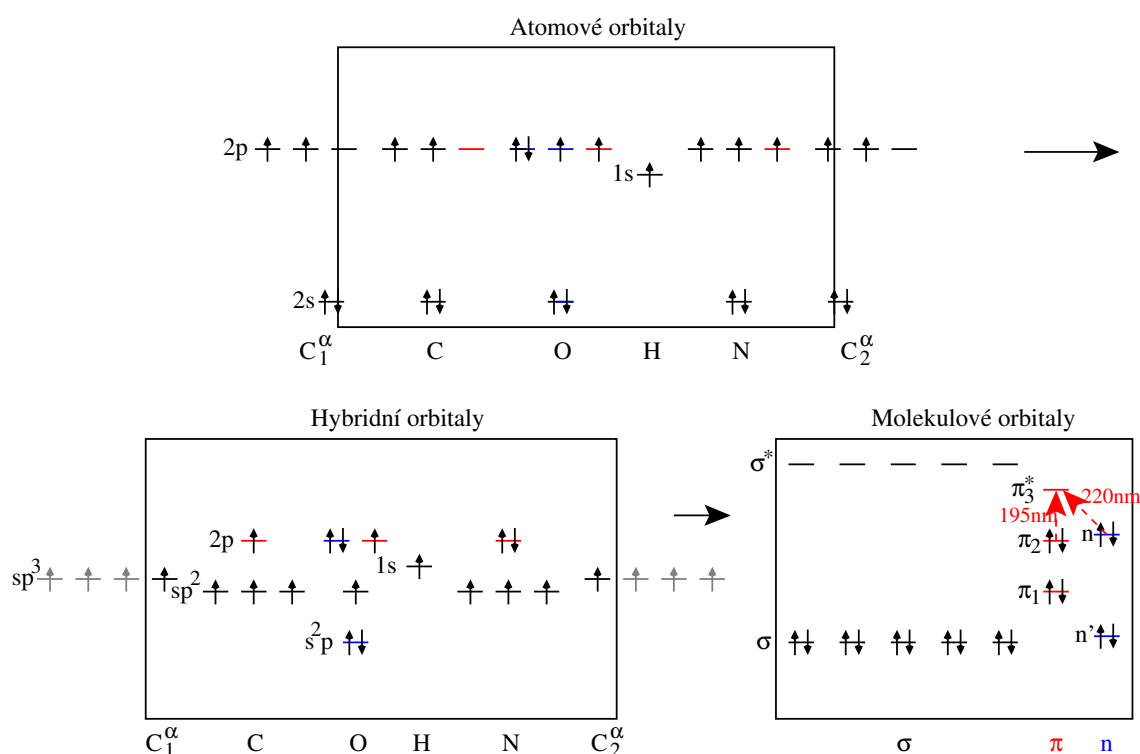
Když už máme popsané vazby  $\sigma$ , podívejme se teď na vazby  $\pi$ . Mezi původními patnácti atomovými orbitály byly tři orbitály  $2p_z$ , které jsme nepoužili k hybridizaci. Tyto orbitály leží nad a pod rovinou  $xy$  a můžeme je použít k tvorbě tří orbitalů  $\pi$  (obrázek 15.16).

$$\pi_1 = \sqrt{\frac{1}{3}}2p_z(O) + \sqrt{\frac{1}{3}}2p_z(C) + \sqrt{\frac{1}{3}}2p_z(N) \quad (15.115)$$

$$\pi_2 = \sqrt{\frac{1}{2}}2p_z(O) + \sqrt{\frac{1}{2}}2p_z(N) \quad (15.116)$$

$$\pi_3^* = \sqrt{\frac{1}{3}}2p_z(O) - \sqrt{\frac{1}{3}}2p_z(C) + \sqrt{\frac{1}{3}}2p_z(N) \quad (15.117)$$

Obrázek 15.15: Tvorba orbitalů  $\sigma$ .Obrázek 15.16: Tvorba orbitalů  $\pi$ .



Obrázek 15.17: Energetické rozložení orbitalů peptidové skupiny.

Po vytvoření  $\sigma$  a  $\pi$  orbitalů nám zbyly ještě dva nepoužité orbitály kyslíku ( $2p_y$  a  $s^2p$ ). Tyto orbitály tvoří ne vazebné molekulové orbitály  $n$  a  $n'$ .

$$n = 2p_y(O) \quad (15.118)$$

$$n' = s^2p(O) \quad (15.119)$$

#### D.4 Energie stavů a vlnové délky přechodů

Víme, že rezonanční frekvence přechodů elektronů ze základního stavu do stavu vyššího je dána rozdílem energií těchto stavů. Podívejme se tedy na energie stavů popsaných jednotlivými orbitály. Schematicky je energetické rozložení orbitalů ukázáno na obrázku 15.17.

Zleva doprava jsou nakresleny orbitály tak, jak jsme je postupně počítali. Rámeček vymezuje to, co nás zajímá (tedy skupinu CONH s vazbami na sousední uhlíky). Různé typy molekulových orbitalů jsou znázorněny barevně, pomocí barev je naznačeno, kombinací kterých atomových orbitalů jsme jednotlivé typy molekulových orbitalů získali. Dále je na obrázku zachyceno obsazení orbitalů elektrony v základním stavu. Konečně jsou na obrázku znázorněny dva přechody do vyšších stavů. Jde o přechody, které ve spektrech obvykle vidíme v ultrafialové oblasti. Světlo o vlnové délce kolem 190 nm je pohlceno při přechodu z ne vazebného  $\pi_2$  orbitalu do antivazebného  $\pi_3^*$  orbitalu. Kolem 220 nm pozorujeme absorpci při přechodu z ne vazebného  $n$  orbitalu do antivazebného  $\pi_3^*$  orbitalu.



## D.5 Tranzitní dipóly a pravděpodobnosti přechodů

Pokud chceme znát i pravděpodobnost přechodu (tedy absorpční koeficient), musíme umět spočítat také tranzitní elektrické dipólové momenty. Co to vlastně je elektrický dipólový moment? Je-li záporný náboj elektronů rozmístěn symetricky kolem kladného náboje jádra (jak by napovídaly tvary orbitalů), je dipólový moment nulový. Pokud ale dojde k přemístění elektronů takovým způsobem, že výsledný záporný náboj elektronů bude vzdálen od jádra o vzdálenost  $r$  ve směru  $\vec{r}$ , můžeme takovou polarizaci popsat elektrickým dipólovým momentem

$$\vec{\mu}_e = q\vec{r} \quad (15.120)$$

kde  $q$  je výsledný náboj. Protože kvantová teorie umí popsat pravděpodobnost výskytu elektronu (a tak i rozložení náboje) pomocí vlnové funkce, pomůže nám i nalézt dipólový moment:

$$\vec{\mu}_e = q \int_V \psi \vec{r} \psi^* dV \quad (15.121)$$

místo náboje jsme vlastně do rovnice 15.120 dosadili vyjádření pravděpodobnosti výskytu elektronu. Tak, jak jsme si v rovnici 15.121 vypočítali dipólový moment molekuly s elektrony ve stavu  $\psi$ , můžeme si spočítat i tranzitní elektronový dipólový moment přechodu ze stavu  $\psi_1$  do stavu  $\psi_2$ :

$$\vec{\mu}_{e, \text{tr}} = q \int_V \psi_2 \vec{r} \psi_1^* dV \quad (15.122)$$

Integrály v rovnicích 15.121 a 15.122 znamenají, že musíme sečíst všechny lokální příspěvky  $d\vec{\mu}_{e, \text{tr}}$  k celkovému tranzitnímu momentu  $\vec{\mu}_{e, \text{tr}}$ . Vezměme si například přechod z orbitalu 2s do orbitalu 2p<sub>z</sub>. Zkusme nejprve sčítat příspěvky k  $\vec{\mu}_{e, \text{tr}}$  podél osy  $z$ . V horní polovině (tedy pro kladné  $z$ ) mají obě vlnové funkce kladné znaménko, takže součin  $\psi_1 \psi_2$  bude kladný a příspěvek  $z\psi_1 \psi_2$  bude také kladný. Naopak v dolní polovině (pro záporné  $z$ ) mají vlnové funkce opačné znaménko, takže součin  $\psi_1 \psi_2$  bude záporný a příspěvek  $z\psi_1 \psi_2$  opět kladný ( $z$  je teď záporné). Tedy příspěvky k tranzitnímu momentu  $\vec{\mu}_{e, \text{tr}}$  budou vždy podél osy  $z$  kladné. Co když budeme podobně sčítat příspěvky podél osy  $x$ ? V horní části obrázku jsou obě vlnové funkce kladné pro kladné i záporné  $x$ . Proto se nám příspěvky v místě s kladným  $x$  vyruší s příspěvky v místě s záporným  $x$  (obě vlnové funkce jsou symetrické, sudé vzhledem k  $x$ , takže  $\psi(x) = \psi(-x)$ ). Stejně tak to dopadne i v dolní polovině obrázku, kde je zase součet  $\psi_1 \psi_2$  záporný pro kladné i záporné  $x$ . Totéž bychom zjistili i podél osy  $y$ . To znamená, že celkový příspěvek ve směru os  $x$  a  $y$  bude nulový, vektor  $\vec{\mu}_{e, \text{tr}}$  bude směřovat ve směru osy  $z$ .

Podobnou analýzou znaménka orbitalů  $\pi_2$  a  $\pi_3^*$  bychom zjistili, že tranzitní elektrický dipólový moment přechodu  $\pi_2 \rightarrow \pi_3^*$  je kolmý k ose  $z$  a směřuje od dusíku ke kyslíku. Naopak analýza orbitalů  $n$  a  $\pi_3^*$  ukazuje, že všechny složky tranzitního elektrického dipólového momentu jsou nulové. Přechod  $\pi_2 \rightarrow \pi_3^*$  je tedy *elektricky povolený*, zatímco přechod  $n \rightarrow \pi_3^*$  je *elektricky zakázaný*.

Tranzitní elektrický dipólový moment popisuje přesun pravděpodobnosti výskytu elektronu (a tedy i přesun náboje) určitým směrem. Jde tedy o popis jakési *translační* změny během přechodu. Kromě toho dochází při přechodu i k *rotačním* změnám pravděpodobnosti výskytu elektronu a tedy ke kruhovému toku náboje. Jak víte z fyziky, kruhový tok náboje vyvolá vznik magnetického pole. Kromě tranzitního elektrického dipólového momentu je tedy přechod mezi stavy popsán i tranzitním *magnetickým* dipólovým momentem. Analýza znaménka a tvaru orbitalů by nám například ukázala, že přechod  $n \rightarrow \pi_3^*$  má nenulový tranzitní *magnetický* dipólový moment, je tedy *magneticky povolený*.

Vzájemné interakce elektrických a dipólových momentů způsobují rozdílnost v absorpci levotočivě a pravotočivě polarizovaného světla (cirkulární dichroismus). Zatímco zrcadlově symetrické molekuly mají elektrické a magnetické momenty navzájem kolmé a tedy neinteragující, v chirálních molekulách k interakci dochází. Peptidová vazba, která nás především zajímá, je ovšem rovinný útvar a sama o sobě je tedy zrcadlově symetrická. V proteinech však dochází k interakcím mezi tranzitními momenty různých

vazeb. Rozdíly mezi absorpcí kolem 220 nm souvisí především s interakcemi elektrického tranzitního momentu jedné vazby s magnetickým tranzitním momentem jiné vazby. Naopak rozdíly absorpcí kolem 195 nm jsou dány vzájemným působením elektrických dipólů různých vazeb. To, že cirkulární dichroismus závisí na interakcích dipólů jednotlivých vazeb vysvětluje, proč je tato technika tak užitečná ke zkoumání sekundárních struktur, které se liší právě vzájemnou orientací peptidových vazeb.

## E Světlo a sekundární struktura

Jakým způsobem odráží schopnost absorbovat světlo sekundární strukturu (pravidelné uspořádání monomerních jednotek biopolymeru)? Představme si monomerní jednotku, jejíž elektrony mohou přecházet z jednoho stavu do druhého. Elektrony v těchto dvou stavech budou popisovat dvě vlnové funkce a přechod mezi těmito stavy bude popsán jedním tranzitním elektrickým dipólovým momentem. Utvořme z takových monomerních jednotek dimer. Každá jednotka bude moci přecházet z jednoho stavu do druhého, celkem máme čtyři stavy a dva tranzitní dipóly. Kdyby byly tyto dvě monomerní jednotky od sebe izolovány, měly by dva stavy stejnou energii (stav s první jednotkou v nižším stavu a druhou ve vyšším a stav s druhou jednotkou v nižším stavu a první ve vyšším). V dimery jsou ale tranzitní dipóly blízko sebe a navzájem interagují (první leží v elektrickém poli druhého a naopak). Proto nezískáme výše zmíněné stavy se stejnou energií, ale stavy jiné. Stav, odpovídající energeticky výhodnějšímu uspořádání dipólů (s energií o něco sníženou) a stav odpovídající energeticky nevýhodnému uspořádání dipólů (s energií o něco zvýšenou). Proto bude dimer absorbovat světlo při dvou vlnových délkách. Popsaným stavům budou samozřejmě odpovídat i nové vlnové funkce a z těchto vlnových funkcí budeme moci spočítat výsledné tranzitní elektronové dipólové momenty dimeru (budou úměrné součtu a rozdílu jednotlivých tranzitních dipólových momentů). Rozdíly energií nám budou udávat vlnové délky absorpce dimeru a výsledné tranzitní dipólové momenty budou udávat pravděpodobnost absorpce při jednotlivých vlnových délkách (absorpční koeficienty).

Pokud budeme ve výstavbě makromolekuly pokračovat, můžeme z monomerních jednotek vybudovat například pravidelnou šroubovici. Získáme tedy sekundární strukturu. Z tranzitních dipólů monomerních jednotek získáme tři výsledné tranzitní dipólové momenty. Jeden bude rovnoběžný s osou šroubovice a dva budou na ni kolmé. Energie přechodů popsanych dvěma tranzitními dipóly kolmými na osu šroubovice bude ale téměř stejná, takže ve spektru uvidíme dva absorpční pásy.

Tvorba pravidelných šroubovic ( $\alpha$ -šroubovic proteinů či dvojšroubovic DNA) se v absorpčních spektrech projeví rozštěpením do několika pásů a poklesem absorpce ve srovnání s neuspořádanou strukturou<sup>24</sup>. Ačkoli se mnoho nedovíme o strukturních detailech pravidelně uspořádaných makromolekul, můžeme pomocí poklesu absorpce sledovat tvorbu sekundárních struktur kvantitativně.

---

<sup>24</sup> Obvykle sledujeme pokles absorpce.

## F Anomální rozptyl

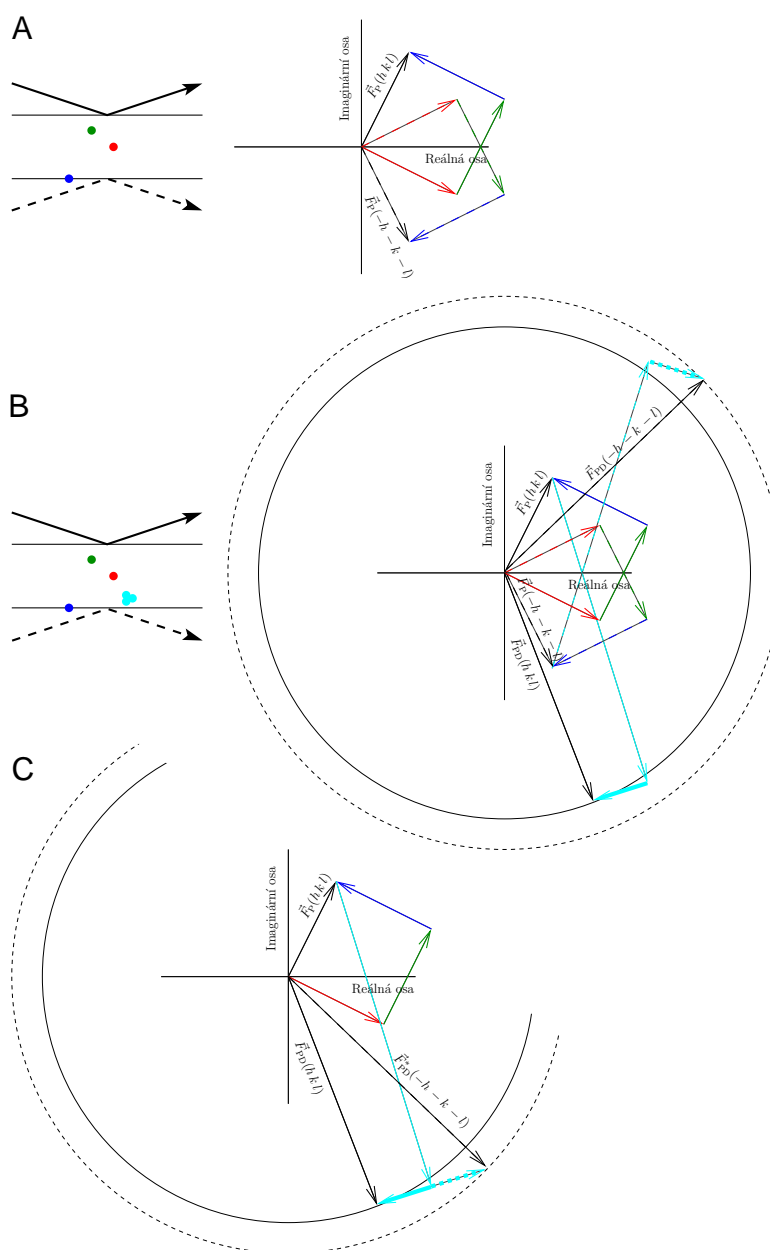
V kapitole 12.13 jsme si představili anomální rozptyl jako jev, který nám může pomoci vyřešit fázový problém. Následující text nabízí poněkud konkrétnější popis.

Zopakujme si nejdříve, co jsme si řekli o intenzitě reflexí s opačnými Millerovými indexy v nepřítomnosti těžkých atomů. Elektrony, které leží na Braggových rovinách, budou rozptylovat se stejnou fází ve směru definovaném Braggovou rovinou, fáze záření rozptýleného na elektronech mezi Braggovými rovinami je posunutá úměrně kolmé vzdálenosti elektronu od Braggovy roviny (plné šipky na obrázku 15.18A). Stejná sada Braggových rovin popisuje ale také odraz na *opačné straně* roviny, tedy pod přesně opačným úhlem (čárkované šipky na obrázku 15.18A). Jestliže odraz paprsku znázorněného plnou čarou na obrázku 15.18A odpovídá rovině s Millerovými indexy  $(hkl)$ , druhou stranu stejné roviny (od které se odráží čárkovaný paprsek) popisují Millerovy indexy  $(-h - k - l)$ . Dvojice reflexí s indexy  $(hkl)$  a  $(-h - k - l)$  se označuje jako Friedelův pár. Pokud nedochází k interakci záření s elektrony, mají strukturní faktory Friedelova páru fázi stejné velikosti, ale opačného znaménka. Matematicky řečeno, jde o *komplexně sdružená* čísla  $F(hkl) = F^*(-h - k - l)$ . Jinými slovy, čárkované šipky na obrázku 15.18A jsou zrcadlovým obrazem plných šipek.

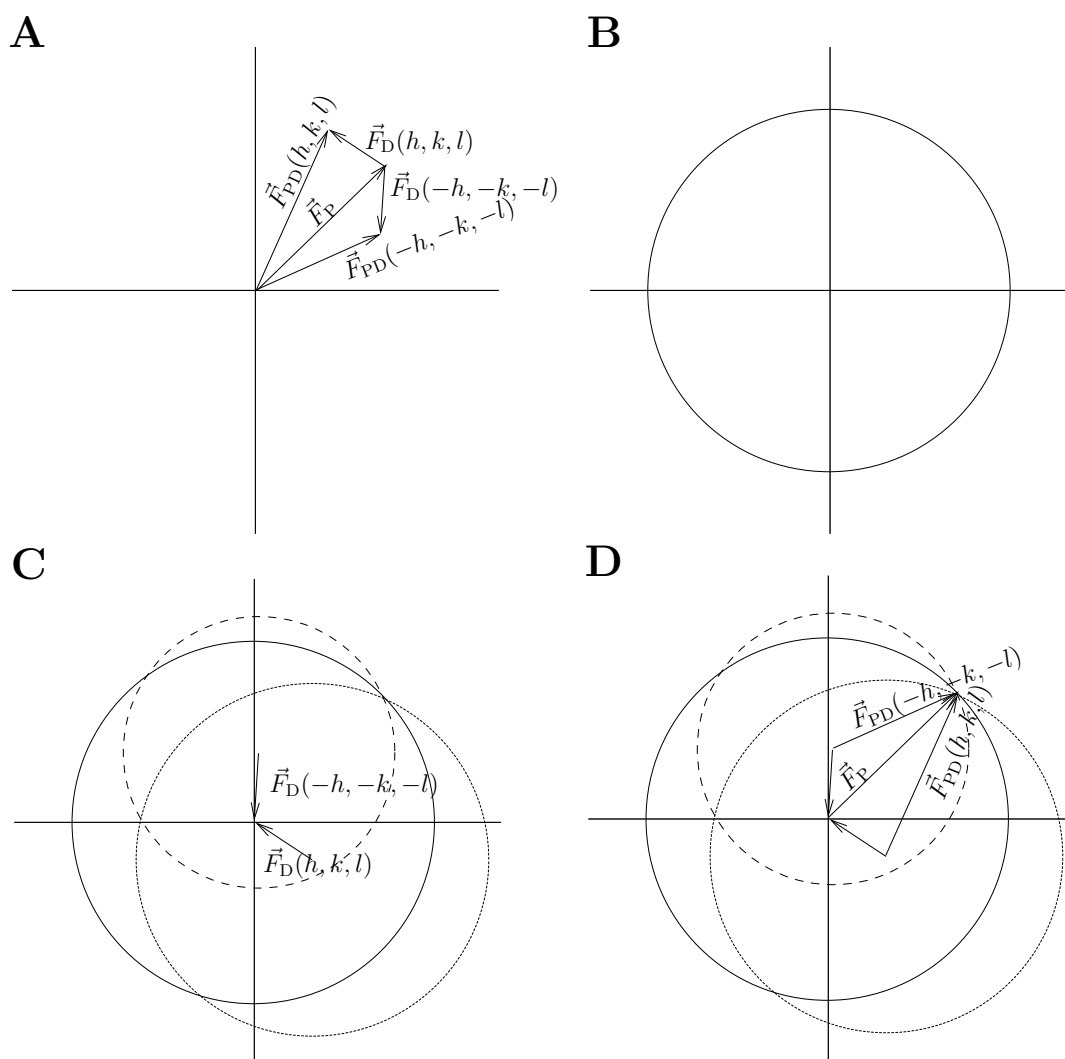
Pokud se ale frekvence záření blíží rezonanční frekvenci, pohlcený foton vybudí přechod elektronu do jiného stavu a vyzářený foton bude mít jinou amplitudu i fázi<sup>25</sup>. Tento rozdíl je znázorněn krátkou tučnou šipkou na obrázku 12.11B a nazývá se anomální rozptyl. Všimněte si, že tato šipka pootočí vektory  $\vec{F}_D(hkl)$  i  $\vec{F}_D(-h - k - l)$  ve stejném směru (změní fázi obou strukturních faktorů o stejnou hodnotu). Po přičtení příspěvků těžkého atomu (včetně anomálního rozptylu) k  $\vec{F}_P(hkl)$  a  $\vec{F}_P(-h - k - l)$  získáme výsledné strukturní faktory  $\vec{F}_{PD}(hkl)$  a  $\vec{F}_{PD}(-h - k - l)$ , které mají *různou amplitudu* – plná a čárkovaná kružnice na obrázku 15.18B mají různý poloměr. Strukturní faktory Friedelova páru nejsou v přítomnosti anomálního rozptylu komplexně sdružené  $F(hkl) \neq F^*(-h - k - l)$ . Tento rozdíl vynikne lépe, když si místo černé čárkované šipky  $\vec{F}_{PD}(-h - k - l)$  nakreslíme její zrcadlový obraz (což není nic jiného, než komplexně sdružený vektor  $\vec{F}_{PD}^*(-h - k - l)$ ). Obrázek 15.18C ukazuje, že vektor  $\vec{F}_{PD}^*(-h - k - l)$  se nepřekrývá s plnou šipkou  $\vec{F}_{PD}(hkl)$  a má i jinou amplitudu (různé poloměry kružnic). Rozdílná amplituda se projeví v difrakčním obrazci jako různá intenzita reflexí s indexy  $(hkl)$  a  $(-h - k - l)$ .

Pokud jsme schopni rozlišit intenzity reflexí s opačnými hodnotami indexů  $h, k, l$ , můžeme anomální rozptyl využít pro řešení fázového problému podobně jako v případě isomorfního nahrazení. Graficky je takový postup ukázán na obrázku 15.19.

<sup>25</sup> Pokud budou všechny rozptylovat stejně, nebude ovlivněna amplituda, ale jen fáze



Obrázek 15.18: Friedelův pár a anomální rozptyl. Obrázek A zachycuje situaci v nepřítomnosti těžkého atomu, obrázky B a C v případě, že je přítomen těžký atom vykazující anomální rozptyl. Vlevo je znázorněna poloha lehkých atomů (nakreslených červeně, zeleně a modře) a jednoho těžkého atomu (nakresleného azurovou barvou) mezi Braggovými rovinami, vpravo Argandův diagram ukazující sčítání strukturálních faktorů jednotlivých atomů (strukturální faktory a atomy jsou nakresleny stejnými barvami). Čárkovaně jsou nakresleny strukturální faktory pro odraz od spodní strany Braggových rovin. Černé šipky představují výsledné strukturální faktory pro molekulu bez těžkého atomu ( $\vec{F}_P$ ) a s těžkým atomem ( $\vec{F}_{PD}$ ). Obrázek C zachycuje srovnání strukturálního faktoru  $\vec{F}_{PD}(hkl)$  s hodnotou komplexně sdruženou se strukturálním faktorem  $\vec{F}_{PD}(-h-k-l)$  (vektor  $\vec{F}_{PD}^*(-h-k-l)$ ). Tučně azurové šipky znázorňují fázový rozdíl v důsledku anomálního rozptylu.



Obrázek 15.19: Využití anomálního rozptylu k řešení fázového problému. Na obrázku A jsou zakresleny strukturní faktory.  $\vec{F}_P$  značí strukturní faktor nativní biomakromolekuly (stejný pro  $(hkl)$  a  $(-h-k-l)$ ), který není ovlivněn anomálním rozptylem. Vektory  $\vec{F}_D$  (různé pro  $(hkl)$  a  $(-h-k-l)$ ) jsou strukturní faktory těžkých atomů, které vykazují anomální rozptyl. Vektory  $\vec{F}_{PD}$  jsou výsledné strukturní faktory derivátu s těžkými atomy. Ve skutečnosti známe pouze vektory  $\vec{F}_D(hkl)$  a  $\vec{F}_D(-h-k-l)$  a amplitudy  $|F_P|$ ,  $|F_{PD}(hkl)|$  a  $|F_{PD}(-h-k-l)|$ . Obrázky B až C ukazují princip řešení. Nejprve si nakreslíme plnou kružnici o poloměru  $|F_P|$ , která zachycuje veškerou informaci o strukturním faktoru nativní molekuly (obrázek B). Potom zaneseme do diagramu údaje o derivátech. Vektory  $\vec{F}_D(hkl)$  a  $\vec{F}_D(-h-k-l)$  můžeme zakreslit přímo. Jak jsme si vysvětlili v popisu obrázku 12.10, musí střed plné kružnice, znázorňující amplitudu  $|F_P|$ , ležet v *konci* vektoru  $\vec{F}_D$ . To platí pro  $(hkl)$  i  $(-h-k-l)$ , proto zakreslíme vektory  $\vec{F}_D(hkl)$  a  $\vec{F}_D(-h-k-l)$  tak, aby směřovaly do středu plné kružnice, jak je ukázáno na obrázku C. Nakonec zakreslíme tečkovanou a čárkovanou kružnici odpovídající strukturním faktorům  $\vec{F}_{PD}$ , které jsou ovlivněny anomálním rozptylem a liší se znaménkem indexů  $h, k, l$ . Tyto kružnice mají proto různé poloměry. Jejich posunutí vzhledem k plné kružnici, odpovídající strukturnímu faktoru nativní biomakromolekuly neovlivněné anomálním rozptylem, je dáno směry  $\vec{F}_D$  (střed kružnic leží v počátcích  $\vec{F}_D(hkl)$  a  $\vec{F}_D(-h-k-l)$ ). Podobně jako při použití isomorfního nahrazení nám řešení udávají průsečíky kružnic. Tam, kde se všechny kružnice protnou, leží konce hledaných vektorů (obrázek D).

## G Skalární interakce a přiřazení spekter NMR nukleových kyselin

Jak bylo zmíněno v kapitole 14.10, frekvenční přiřazení proteinů je založeno na jednovazebných skalárních interakcích. Výhodou tohoto přístupu je nezávislost na konkrétní konformaci. Frekvenční přiřazení nukleových kyselin pomocí skalárních interakcí naráží na to, že jednovazebné interakce často nejsou k dispozici a je nutno využívat dvou- i třívazebných interakcí. V důsledku toho je tento přístup u nukleových kyselin komplikován nízkou citlivostí a závislostí na konformaci u některých experimentů (jak dobře víme, třívazebné interakce závisí na torzních úhlech).

### Přiřazení páteře (sekvenční přiřazení)

Fosfátové skupiny páteře nukleových kyselin obsahují kyslík, který nemá vhodné izotopy pro NMR. Sekvenční přiřazení je proto založeno především na třívazebných interakčních konstantách  $^3J(\text{PC})$  a  $^3J(\text{PH})$ . První interakce využívá například HCP experiment, který poskytuje korelace  $\text{H}4'(i)\text{-C}4'(i)\text{-P}(i)$  a  $\text{H}4'(i-1)\text{-C}4'(i-1)\text{-P}(i)$  (obdoba HNCA a HN(CO)CA u proteinů). Druhou interakci využívá HPHCH experiment, který poskytuje korelace  $\text{H}5'(i)\text{-C}5'(i)\text{-P}(i)$ ,  $\text{H}5''(i)\text{-C}5'(i)\text{-P}(i)$  a  $\text{H}3'(i-1)\text{-C}3'(i-1)\text{-P}(i)$ . Třetí možností je PCCH-TOCSY experiment (varianta HCCH-TOCSY experimentu popsaného pro proteiny), který využívá  $^3J(\text{PC})$  interakce ke korelaci fosforu s pentosovým kruhem, následuje homonukleární TOCSY korelace cukerných uhlíků a heteronukleární korelace s navázaným protonem. Tak je možné korelovat fosfor s různými atomy (deoxy)ribosy.

### Přiřazení v cukrech

K přiřazení protonů a uhlíků v (deoxy)ribose se používá HCCH-TOCSY experiment, který byl popsán pro proteiny.

### Korelace báze-cukr

Ke korelaci H6 (pyrimidinů) a H8 (purinů) s H1' cukru se používá dvojice HCN experimentů, z nichž jeden excituje uhlíky alifatické a druhý aromatické. Získáme tak korelace H6-C6-N1 (pyrimidinů), H8-C8-N9 (purinů) a H1'-C1'-N1 (pyrimidinů) a H1'-C1'-N9 (purinů). Opět jde o obdoby HNCA a HN(CO)CA u proteinů. Zmíněných korelací lze dosáhnout i použitím jednoho experimentu (HCNCH), jehož citlivost je ale nižší.

### Přiřazení v bázích

Základním úkolem přiřazení v bázích nukleových kyselin je přiřazení jednotlivých protonů. Ke korelaci protonů v polohách 5 a 6 pyrimidinů lze využít jednovazebných interakcí  $^1J(\text{CH})$  a  $^1J(\text{CC})$ . Korelace vzdálenějších protonů využívá často vícevazebných skalárních interakcí, které bývají v aromatických heterocyklických spinových systémech bází poměrně intenzivní (naopak některé jednovazebné skalární interakce jsou velmi slabé). Často se používá princip TOCSY. Jako příklad můžeme uvést opět HCCH-TOCSY experiment, používaný ke korelaci H2 a H8 u adeninu, nebo HNC-TOCSY-CH experiment (na rozdíl od předchozího koreluje v prvním kroku imino-protony s uhlíky přes imino-dusík, tedy obdobně jako v HNCA experimentu), používaný ke korelaci H1 a H8 u guaninu.