Univerzita Karlova

Přírodovědecká Fakulta

Katedra genetiky a mikrobiologie

Struktura DNA a nástroje její analýzy

Structure of DNA and tools of its analysis

Daniel Berdár

Bakalářská práce

Vedoucí práce: prof. Ing. Bohdan Schneider, CSc., DSc.

Praha 2021

Čestné prohlášení

Prohlašuji, že jsem bakalářskou práci na téma „Struktura DNA a nástroje její analýzy“ vypracoval samostatně a s použitím uvedené literatury.

V Praze dne: Podpis:

Poděkování

V první řadě bych rád bych poděkoval svému školiteli prof. Bohdanu Schneiderovi, za přiležitost vypracovávat u něj bakalářskou práci a možnost vzdělávat se v jeho laboratoři, na Biotechnologickém Ústavu AV ČR v.v.i.

Velké díky pak patří Mgr. Jakubu Svobodovi, za konzultace a četné rady týkající se této práce.

Abstrakt

Model struktury DNA dvoušroubovice je již známý od roku 1953. Plasticita této molekuly však umožňuje vznik většímu množství konformací, které mohou vznikat výsledkem sekvence nukleotidů či prostředí ve kterém se molekula nachází. Strukturní biologie DNA je poměrně rozsáhlá disciplína, která nachází uplatnění od biotechnologického a medicínského výzkumu až po nanotechnologie.

Tato práce se zabývá charakterizací nejdůležitějších elementů DNA z hlediska jejich struktury a stability. Čtenáři dále představuje stěžejní spektrální a difrakční metody pro řešení struktur DNA. Na závěr, práce popisuje bioinformatický nástroj pro submolekulární klasifikaci a validaci těchto struktur.

Klíčová slova:

DNA, nukleové kyseliny, strukturní biologie, strukturní bioinformatika

Abstract

Model of DNA double-helix structure is know since 1953. Plasticity of this molecule, however, gives way to many more conformations, that are a result of nucleotide sequence or chemical conditions. Structural biology of DNA is a broad discipline, which finds use everywhere from biotechnological and medical research to nanotechnologies.

This thesis concerns itself with characterisation of the most important DNA elements, especially their structure and stability. Introduces the reader to crucial spectral and diffraction methods for solving structures, and a bioinformatical tool for submolecular classification and validation od these structures.

Keywords:

DNA, nucleic acids, structural biology, structural bioinformatics

Obsah

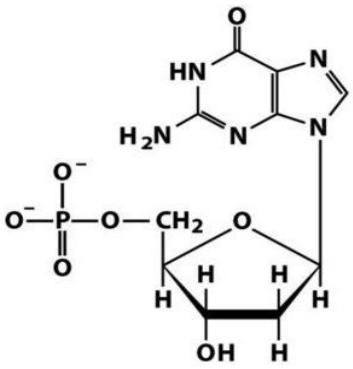
1. Úvod 1
2. Základy stavby DNA 1
2.1. Nukleotid 1
2.2. Pentózo-fosfátová páteř 2
2.3. Báze, párování a stabilita 2
3. Torzní prostor 5
3.1. Glykosidický úhel χ 5
3.2. Páteřní torze α, β, γ, δ, ε, ζ 6
3.3. Vnitřní torzní úhly τ 6
4. Duplex 9
4.1. Konformace duplexu 9
5. Nekanonické struktury 12
5.1. Triplex 12
5.2. Guaninový tetraplex 13
5.3. I-motiv 14
6. Nástroje strukturní analýzy 15
6.1. Difrakce paprsků X na monokrystalech 15
6.2. Spektroskopie nukleární magnetické rezonance 17
6.3. Práce s daty a databáze 18
6.4. PDB a mmCIF formáty 19
7. Dinukleotidové konformační třídy 21
7.1. Konformační abeceda pro nukleové kyseliny 22
7.2. DNATCO 23
8. Závěr 26
9. Bibliografie 27

1. Úvod

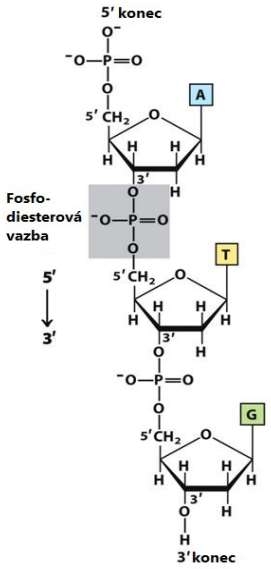
Nukleové kyseliny jsou biologické polymery, které uchovávají genetickou informaci v organismech a účastní se dalších klíčových biologických dějů, které ve svém souboru tvoří přenos genetické informace označovaný jako centrální dogma molekulární biologie. Jde o replikaci, transkripci a translaci. První struktura duplexu deoxyribonukleové kyseliny (DNA) pak byla díky práci J. Watsona, F. Cricka, R. Franklinové, M.Wilnkinse a dalších vyřešena pomocí rentgenové difrakce v roce 1953 (Franklin & Gosling, 1953; Watson & Crick, 1953; Wilkins et al., 1953). Dnes již víme, že DNA vykazuje výrazně větší komplexitu, než pouhý B-DNA duplex. Struktura DNA je plastická a silně závislá na podmínkách jako je například sekvence, hydratace, iontová síla rozpouštědla a pH.

2. Základy stavby DNA

2.1. Nukleotid



Obrázek 1.1 – Deoxyguanosinmonofosfát (dGMP). Převzato a upraveno z Boyle (2005).

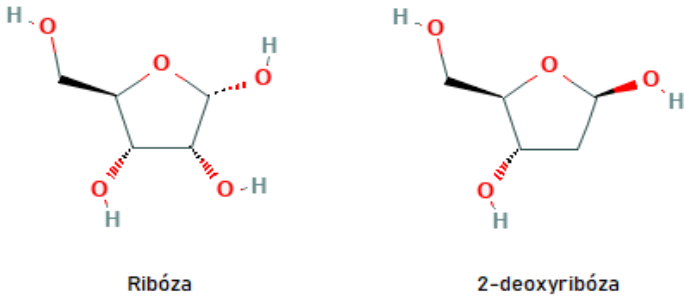


Obrázek 1.2 - Polynukleotid. Převzato a upraveno z Boyle (2005).

Nukleotidy jsou monomerní jednotkou DNA. Samy mají poměrně komplexní chemickou strukturu, která se dá rozdělit na tři hlavní části: pentóza, dusíkatá báze a fosfát (Obrázek 1.1). Sacharid figurující v DNA je 2-deoxyribóza (oproti ribóze v RNA). Na uhlíku C1' je deoxyribóza spojena glykosidickou vazbou s jednou ze čtyř dusíkatých bazí typických pro DNA – adenin (A), thymin (T), cytosin (C) a guanin (G). Na uhlíku C5' je pentóza navázána

esterovou vazbou na fosfát. Řetězením nukleotidů pak vzniká polynukleotidové vlákno (Obrázek 1.2), které je hlavní funkční jednotkou DNA.

Výše zmíněná deoxyribóza je aldopentóza odvozena od ribózy a liší se absencí hydroxylové skupiny na C2', což má mimo jiné za následek nižší tendenci DNA hydrolyzovat, oproti RNA (Obrázek 2).



Obrázek 2 - Ribóza a deoxyribóza. Převzato a upraveno z databáze PubChem (Kim et al., 2021)

2.2. Pentózo-fosfátová páteř

Polynukleotidový řetězec je vázán pomocí fosfodiesterové vazby, tak se označuje série dvou esterových vazeb a fosfátu C5'-O5'-P-O3'-C3', které spojují nukleotidy. Toto navázání je mimo jiné důležité i pro určení směru polynukleotidového vlánka, tzv. 5' konec s volnou fosfátovou skupinou a 3' konec s volnou OH skupinou na C3'. Směr, nebo také direkcionalita, DNA je klíčová vzhledem k biologické aktivitě této molekuly a její schopnosti uchovat informace.

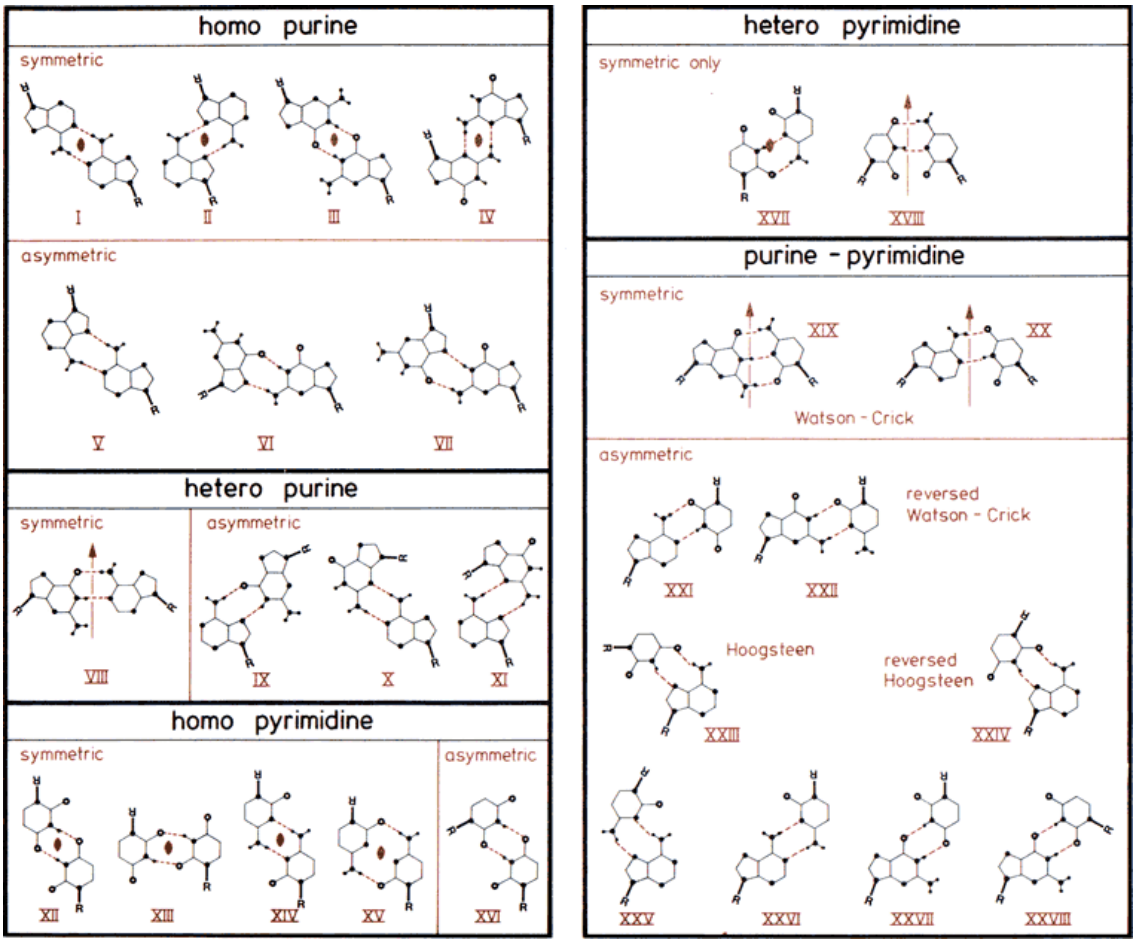
2.3. Báze, párování a stabilita

Dusíkaté báze, adenin, thymin, cytosin a guanin, také označované jako nukleobáze, jsou fundamentální část genetického kódu. Nejsou důležité jen pro ukládání informací do polynukleotidového řetězce, ale jejich interakce jsou klíčové pro strukturu a stabilitu DNA. Báze se rozdělují na deriváty pyrimidinu (C, T) a purinu (A, G).

K interakcím mezi bázemi patří například vodíkové můstky, které zprostředkovávají rozpoznávání komplementárních vláken. Páry bazí totiž asociují pomocí vodíkových můstků, kde v typickém Watson-Crick (W-C) uspořádání jsou mezi A-T páry dva vodíkové můstky a mezi G-C páry tři vodíkové můstky (Watson & Crick, 1953).

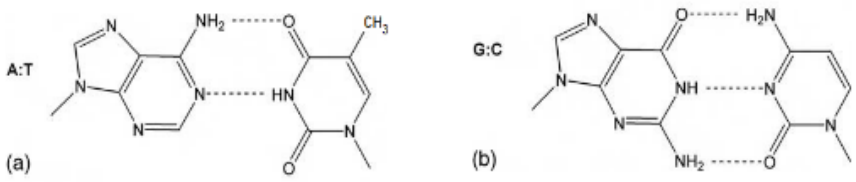
Pro stabilitu DNA jsou pak kritické π-π patrové interakce mezi bázemi, což jsou nekovalentní interakce typické pro aromatické molekuly. Patrové interakce jsou do velké míry zodpovědné za tzv. helikální ‚twist' DNA a za rozdílnou termostabilitu vlákna podle zastoupení nukleobazí. Ačkoliv stabilita sekvencí obsahující větší množství G-C párů se často přisuzuje tomu, že tvoří tři vodíkové můstky, oproti dvěma v A-T párech, tak realitou spíše je, že ke stabilitě G-C bohatých sekvencí budou výrazně přispívat právě i patrové interakce (Yakovchuk et al., 2006). Termostabilita G-C bohatých sekvencí je tedy i způsobena tím, že sekvence párů C-G, G-C a C-G, C-G disponují nejstabilnějšími patrovými interakcemi (Svozil et al., 2010).

Páry bazí také zaujímají širokou škálu vzájemných geometrií, rozlišených na základě toho, zda se jedná o purin-purin, purin-pyrimidin nebo pyrimidin-pyrimidin pár a jejich inklinaci. Tyto geometrie jsou popsány Saengerem v tabulce reprodukované na Obrázku 3 (Saenger, 1984).

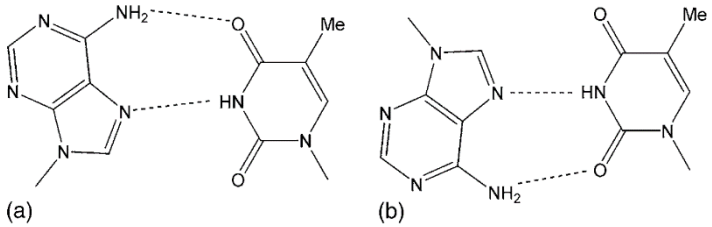


Obrázek 3 - Saengerova tabulka konformací nukleobazí podle typu bazí a vzájemné geometrie. Pro DNA je stěžejní XIX a XX nebo-li W-C páry. A XXIII, XXIV zvané Hoogsteenovo a obrácené Hoogsteenovo párování (Saenger, 1984).

Z těchto geometrií jsou pro DNA nejvýznamnější Watson-Crickovo (Obrázek 4) a Hoogsteenovo párování (Obrázek 5). Vznik Hoogsteenova páru je podmíněn rotací jedné báze kolem glykosidické vazby. Jedna báze musí být v -*anti* konformaci a druhá v -*syn* konformaci podél vazby (glykosidický torzní úhel dále diskutován v 3.1.). Pro A-T pár pak platí, že N7 adeninu se páruje s vodíkem na N3 thyminu, místo N1 jak je tomu u W-C párování (Hoogsteen, 1963).



Obrázek 4 – Kanonické párování bazí A-T (a) a G-C (b). Převzato a upraveno z (Neidle, 2008).



Obrázek 5 - A-T Hoogsteenovo párování (a) a reverzní A-T Hoogsteenovo párování (b) (Neidle 2008).

Párování se však nemusí lišit pouze geometrií, ale může vznikat i tzv. mismatch. Tedy nestandardní páry bazí; G-A, G-T, A-A, G-G, T-T, C-C, C-T a A-C. Mismatch bazí má pak mírně destabilizující efekt na na duplex. Míra destabilizace však záleží ovšem i na okolních bazích, protože destabilizující efekt je podmíněn i nevýhodnými π-π interakcemi (Allawi & SantaLucia, 1998). Mismatch však nemusí nutně deformovat samotnou strukturu duplexu (Kolenko et al., 2020).

3. Torzní prostor

Komplikovaná a dynamická struktura nukleových kyselin vychází i z velké strukturní dynamiky páteře. V následující části textu bude probrán torzní prostor DNA, který se obecně dělí na tři části; úhel kolem glykosidické vazby, úhly na pentózo-fosfátové páteři a pak vnitřní torze popisující konformaci ribózového kruhu.

Torzní úhel je případ dihedrálního úhlu, tedy úhel svíraný dvěma rovinami. Definice je pro systém čtyř po sobě jdoucích atomů A-B-C-D. Kde rovina AB je definována body A, B, C a rovina CD je definována body B, C, D. Úhel, který tyto dvě roviny svírají je torzní úhel atomů A a D okolo vazby B-C (Jones et al., 2008).

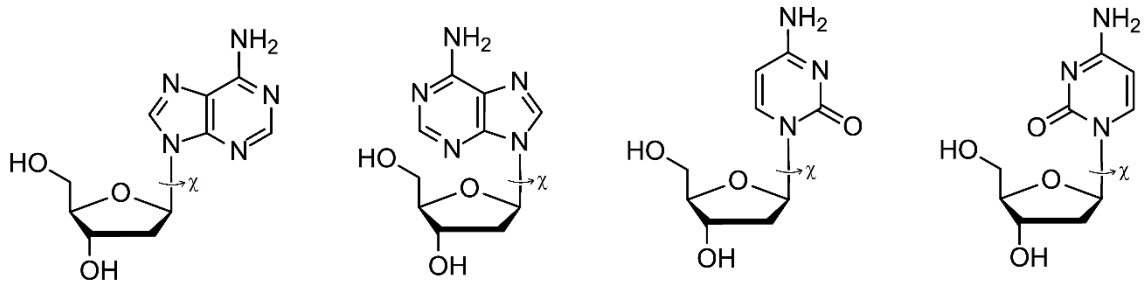
3.1. Glykosidický úhel χ

Deoxyribózu s dusíkatou bází spojuje tzv. glykosidická vazba, která je pro puriny C1'-N9 a pro pyrimidiny C1'-N1. Jedná se o σ-vazbu, báze tedy může rotovat kolem osy definované C-N. Relativní orientaci deoxyribózy a navázané báze lze popsat pomocí dihedrálního úhlu χ. Úhel χ je definován čtyřmi atomy, pro puriny jde o sekvenci O4'-C1'-N9-C4 a pro pyrimidiny O4'-C1'-N1-C2. Ačkoliv se může zdát, že báze kolem úhlu χ mohou rotovat s relativní volností, realita je, že v důsledku stereochemie DNA existují jen dvě stabilní konformace báze vůči deoxyribóze, označované jako *anti* a *syn* (Obrázek 6).

Stéricky preferované intervaly hodnot těchto konformací jsou (Neidle, 2008):

Anti: $-120^{o} > \chi > 180^{o}$

Syn: $0^{o} > \chi > 90^{o}$



Obrázek 6 - Znázornění konformací báze kolem glykosidické vazby.
Převzato a upraveno z Boyle (2005).

U bazí A, T a C platí, že konformace *anti* je energeticky výhodnější, protože dochází k menšímu počtu stérických kolizí. Konformace *syn* je nejméně destabilizující

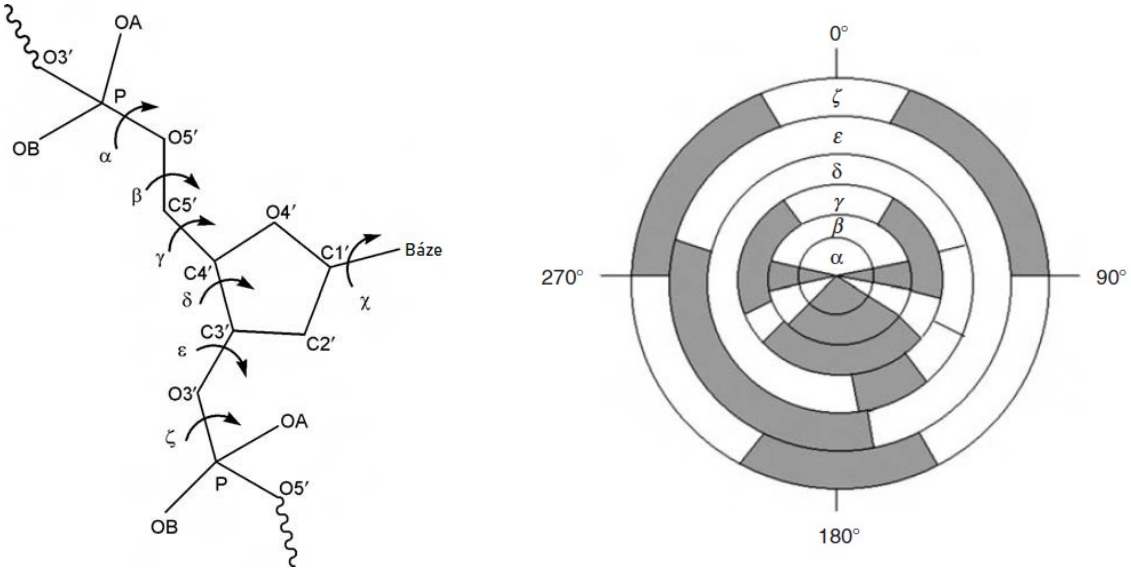
v guanosinových nukleotidech, díky elektrostatickým interakcím mezi exocyklickou N2 amino skupinou guaninu a kyslíkem na fosfátu (Neidle, 2008).

3.2. Páteřní torze α, β, γ, δ, ε, ζ

Pentózo-fosfátová páteř má šest různých torzních úhlů, značených α, β, γ, δ, ε, a ζ, rozmístěné po délce „páteře“ (Obrázek 7). Obdobně jako u úhlu χ jsou, ze stérických důvodů, hodnoty těchto torzních úhlů většinou omezeny na diskrétní intervaly ve kterých DNA nabývá stabilní konformace (Sundaralingam, 1969).

Teoreticky tak vlákno DNA může nabývat velkého množství stabilních konformací. To, že skutečné DNA oligonukleotidy nejsou pozorovány v takto rozmanitém spektru konformací je mimo jiné následek párování bazí, které svojí asociací do jisté míry „omezuje“ volnost páteřní torze (Neidle, 2008).

Změny torze v DNA korelují, tzn. tyto změny často nejsou nezávislé. Příkladem je korelace úhlu γ s glykosidickým úhlem χ, vnitřními torzními úhly deoxyribózy a páteřním úhlem α pozorovaná v nukleotidech (Schneider et al., 1997).



Obrázek 7 - Vlevo zobrazeny torzní úhly na páteři. Vpravo vyznačeny intervaly, kterých jednotlivé úhly mohou běžně nabývat. Převzato a upraveno z Neidle (2008).

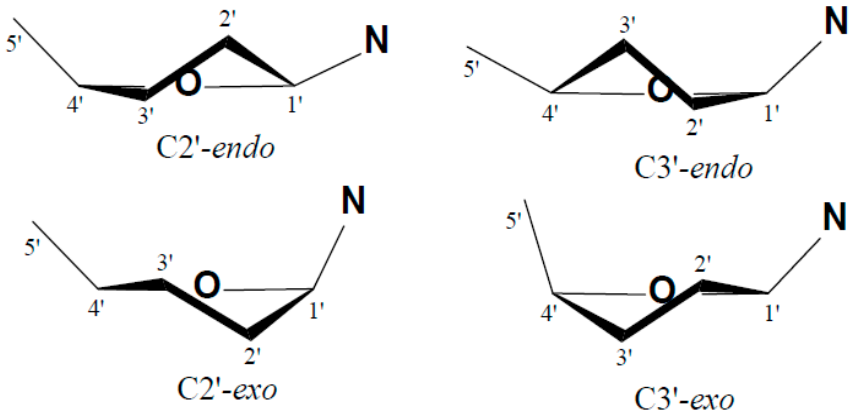
3.3. Vnitřní torzní úhly τ

V DNA se deoxyribóza nachází výhradně v cyklické podobě jako derivát pentózového cukru, furanosy, kdy její pětičlenný kruh je tvořen atomem kyslíku a čtyřmi atomy uhlíku.

Deoxyribóza však nezaujímá planární strukturu, naopak dochází k vychylování, nebo také „krabatění“ roviny cyklu (angl. *sugar puckering*). Samotné vychylování vychází hlavně z tetrahedrální geometrie sp^3 hybridizovaného uhlíku, které nepodporuje vznik striktně planárních struktur. Nemalou roli však hrají i substituenty na deoxyribóze – energeticky nejvýhodnější konformace má substituenty co nejdále od sebe (Olson, 1981).

Existují dva způsoby popisu vnitřní struktury deoxyribózy. Kvalitativně lze konformaci popsat jako vychýlení jednoho, nebo častěji dvou, atomů z roviny definované zbylými atomy (Obrázek 8). Kde označení *-endo* definuje vychýlení směrem k nukleobázi, která je sama vychýlena „nad” rovinu cyklu, a označení *-exo*, označuje vychýlení směrem od nukleobáze. Konformace *-exo* či *-endo* v jednom směru pak musí být kompenzována, vzhledem k geometrii uhlíku, menším vychýlením ve směru druhém. Například pro konformaci C2’-*endo* dochází k menšímu vychýlení C3’-*exo*. V izolovaných deoxyribonukleotidech i polynukleotidových řetězcích byly nejčastěji pozorovány C2’-*endo* a C3’-*endo* konformace (Arnott & Hukins, 1972).

Pomocí NMR spektroskopie bylo zjištěno, že výskyt C2’-*endo* a C3’-*endo* závisí na bázi v nukleotidu. Puriny vykazují vyšší zastoupení C2’-*endo* konformace a pyrimidiny C3’-*endo* (Meints et al., 2001; Van Dam et al., 2002).



Obrázek 8 - endo a exo konformace ribózového cyklu. Převzato a upraveno z Boyle (2005).

Druhý, kvantitativní, způsob popisu konformace ribózového cyklu je definováním pěti vnitřních torzních úhlů τ_0-τ_4 (Obrázek 9). Každý z torzních úhlů τ je definován sekvencí čtyř atomů. Samotná konformace se pak rozlišuje pomocí pseudorotačního úhlu P a maximálního úhlu vychýlení τ_m (Altona & Sundaralingam, 1972).

Pro P platí:

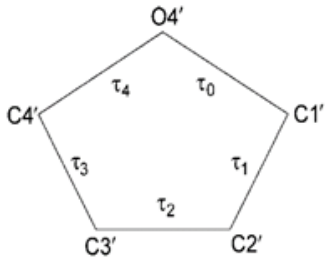
$$\mathrm{tg}\, P = \frac{(\tau_4 + \tau_1) - (\tau_3 + \tau_0)}{2\tau_2(\sin 36° + \sin 72°)}$$

a pro τ_m:

$$\tau_m = \frac{\tau_2}{cosP}$$

Pak lze přibližně vyjádřit libovolný úhel τ_j:

$$\tau_j = \tau_m cos[P + 0.8\pi(j - 2)]$$



τ_0 (C4'-O4'-C1'-C2')
τ_1 (O4'-C1'-C2'-C3')
τ_2 (C1'-C2'-C3'-C4')
τ_3 (C2'-C3'-C4'-O4')
τ_4 (C3'-C4'-O4'-C1')

Obrázek 9 - Vnitřní torzní úhly ve furanosovém kruhu. Převzato a upraveno z Neidle, 2008.

4. Duplex

V přírodě se DNA nachází většinou ve formě dvoušroubovice, nebo-li duplexu. Jedná se o dvě komplementární antiparalelně běžící vlákna DNA propojená vodíkovými vazbami mezi nukleobázemi.

V praxi to znamená, že se ve struktuře dsDNA nachází velké množství vodíkových můstků, které přispívají stabilitě a tvaru molekuly. Vodíkové můstky jsou stěžejní pro asociaci DNA do duplexu a jsou také stabilizujícím faktorem. Nejsou však, z hlediska snížení Gibbsovy volné energie dsDNA oproti ssDNA, jedinou složkou.

Kromě již výše zmíněných π-π interakcí, které jsou pro strukturu DNA kritické, hrají důležitou roli i hydrofilní interakce pentózo-fosfátové páteře a hydrofóbní efekt nukleobazí (Schweitzer & Kool, 1995). DNA ovšem není stabilizována pouze vodným rozpouštědlem, ale i kationty, které stíní záporný náboj pentózo-fosfátové páteře (Anastassopoulou, 2003).

Všechny strukturní formy DNA jsou rovněž v souladu s preferenčními geometriemi pentózo-fosfátové páteře (Schneider et al., 2018; Černý et al., 2020).

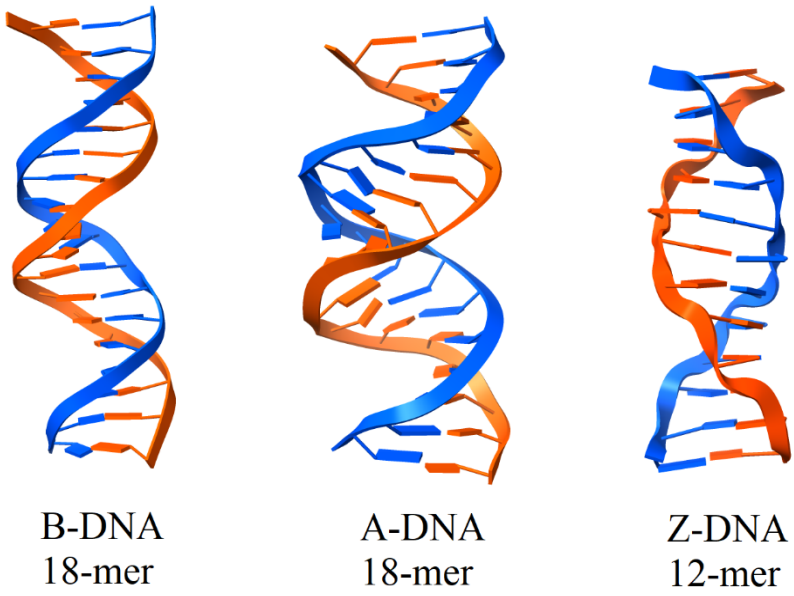
4.1. Konformace duplexu

Konformační variabilita DNA se projevuje schopností duplexu nabývat odlišných konformací, které se standardně rozdělují do tří hlavních tříd, B, A a Z (Obrázek 10). V přírodě většinou nalezneme pravotočivou B-DNA, popsanou Watsonem a Crickem, jejíž konformace je již téměř ikonická (Watson & Crick, 1953). B-DNA se pak dále dělí na BI a BII, které se od sebe mírně liší torzními parametry na pentózo-fosfátové páteři. V duplexu se rozlišují i tzv. žlábky, což označuje prostor mezi samotnými vlákny. V B-formě se jedná o velký a malý žlábek, které se liší svojí šířkou.

Další možnou formou duplexu je pak A-DNA, která je taktéž pravotočivou šroubovicí, ovšem má více bazí na otočku a je tak "kratší" a širší než-li B-DNA o stejném počtu párů bazí. Patrné jsou rozdíly ve velkém a malém žlábku, kdy původně malý žlábek je nyní výrazně širší než původní velký žlábek. A-DNA vzniká dehydratací DNA, a je dále typická pro nějaké DNA-RNA hybridní struktury (Xiong & Sundaralingam, 1998).

Existuje i levotočivá konformace dsDNA. Z těchto tří konformací má Z-DNA největší vzdálenosti mezi bázemi na stejném vlákně. Její výskyt je možný v sekvencích kde se střídá $(dC\text{-}dG)_n$ trakt (Wang & Vasquez, 2007).

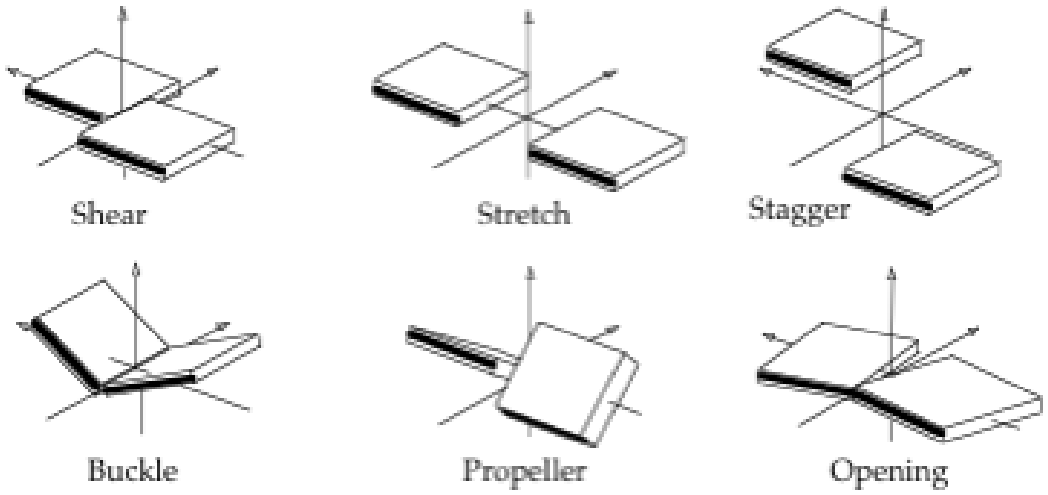
Konformace jsou popsány tzv. helikálními parametry, které popisují geometrii párování bazí, vzestup šroubovice na jeden pár, vzdálenost bazí nebo i rozměry žlábků. Hodnoty, kterých helikální parametry pro zmíněné konformace nabývají, jsou v tabulce 1. Geometrie pro páry bazí pak na obrázku 11 a geometrie pro sousední báze na jednom vlákně na obrázku 12.



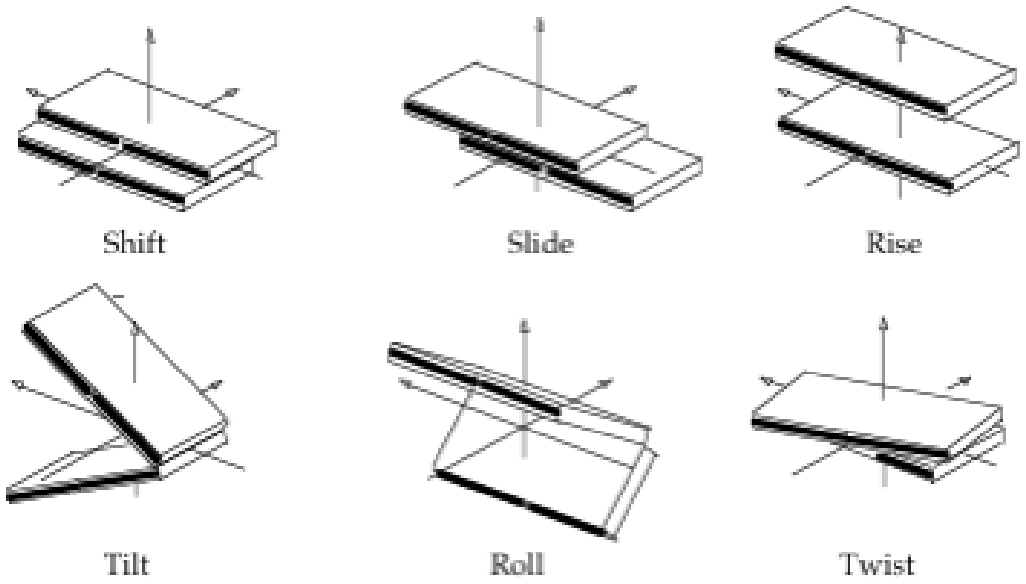
Obrázek 10 – Konformace DNA. XRD Struktury převzaty z databáze PDB. B-DNA 18mer (komplex s proteinem) 3MKY (Schumacher et al., 2010). A-DNA 18mer 6ROS (Kolenko et al., 2020). Z-DNA dodecamer 4OCB (Luo et al., 2014).

	Rise bp (Å)	Helikální twist (º)	Propeller twist/bp (º)	Inklinace /bp (º)	Počet bp na otočku
B	3.38	36.0	-15.1	2.8	10
A	2.54	32.7	-10.5	22.6	11
Z	7.25	-30.0	8.3	0.1	12

Tabulka 1 - Vybrané helikální parametry a jejich hodnoty pro konformace B, A a Z



Obrázek 11 - Popis vzájemných geometrií párů bazí (Hospital et al., 2013). Názvy ponechány v angličtině.



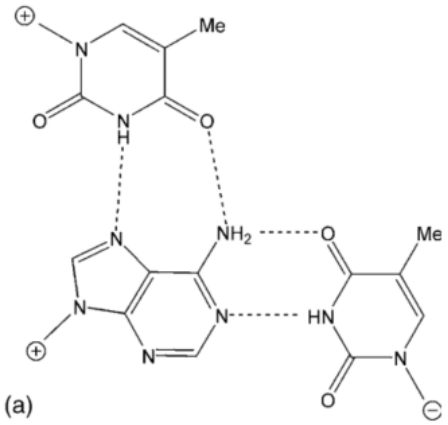
Obrázek 12 - Popis vzájemných geometrií sousedních bazí na stejném vlákně (Hospital et al., 2013). Názvy ponechány v angličtině.

5. Nekanonické struktury

Existence mnoha geometrií párování bazí, společně s komplikovaným torzním prostorem nukleových kyselin, umožňuje vznik řady odlišných struktur, než pouze antiparalelní duplex. Tato strukturní variabilita je u DNA menší než u RNA, alternativní formy jsou ale funkčně důležité.

5.1. Triplex

Triplex, nebo-li trojitý helix, někdy též H-DNA, může vzniknout, když se polypyrimidinové nebo polypurinové vlákno nachází ve velkém žlábku duplexu a asociuje s jeho polypurinovým vláknem. DNA triplexy dělíme do dvou hlavních kategorií: intermolekulární a intramolekulární triplexy (Frank-Kamenetskii & Mirkin, 1995). Intermolekulární triplex je triplex tvořen více oddělenými polynukleotidovými vlákny. Intramolekulární triplexy jsou tvořeny v rámci jednoho vlákna.



Obrázek 13 - Geometrie párování v triplexu (Neidle, 2008).

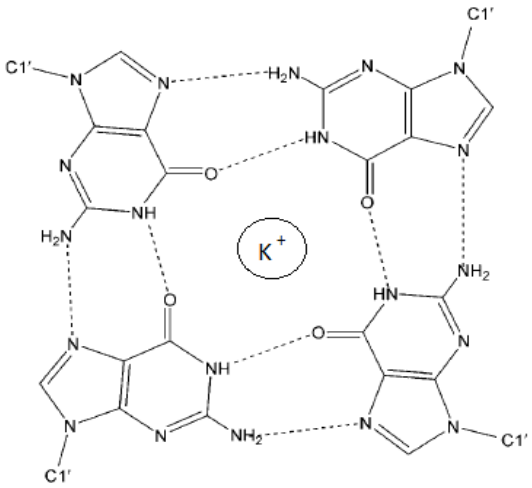
Konvence pro značení trojice nukleobazí v trojitých helixech v této práci bude X:Y-Z, kde Y-Z je pár duplexu a X:Y je pár mezi X bazí třetího vlákna a Y bazí duplexu.

Triplexy se mohou vyskytovat ve dvou hlavních konformacích. Dvě polypyrimidinová vlákna (Y) a jedno purinové vlákno (R) uspořádáné Y:R-Y nebo dvě polypurinová vlákna s jedním polypyrimidinovým uspořadaným do R:R-Y. Pro Y:R-Y se polypyrimidinové vlákno váže na polypurinové vlákno paralelně s Hoogsteenovou geometrií na páru Y:R. Zatímco pro R:R-Y dvě polypurinová vlákna asociují antiparalelně s reverzní-Hoogsteenovou geometrií na páru R:R. Pro případ antiparalelních motivů jsou typické trojice G:G-C, A:A-T a T:A-T, zatímco v paralelních motivech jsou to C^+:G-C a T:A-T .

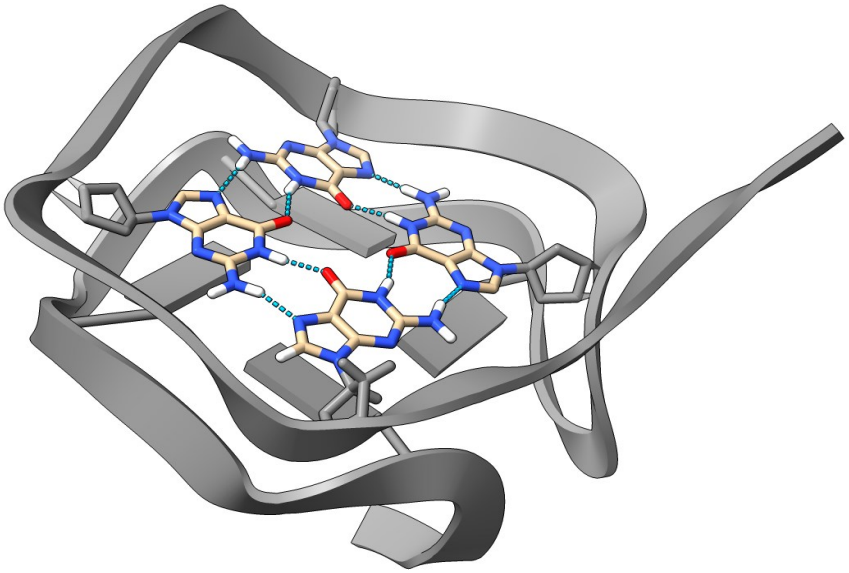
Obrázek 14 - Intermolekulární triplex DNA. Triplex-tvořící vlákno zeleně. NMR Struktura 1BWG z PDB (Asensio et al., 1999).

5.2. Guaninový tetraplex

G-tetraplexy jsou stabilní struktury tvořeny čtyřmi guaniny tvořícími tzv. tetrádu stabilizovanou jednomocným kationem, například K^+ nebo Na^+ (Obrázek 15) (Laughlan et al., 1994). Každý guanin je pak navázán na dva sousední, s jedním tvoří geometrii W-C a s druhým geometrie dle Hoogsteena. Obdobně jako u triplexu, lze G-tetraplexy rozlišit podle počtu molekul, které jej tvoří. Intramolekulární tetraplexy vznikají na jednom poly(G) vlákně a intermolekulární tetraplexy mohou vznikat ze dvou až čtyř polynukleotidových vláken (Fry, 2007). Počet molekul účastnících se v tvorbě tetraplexu ovlivňuje jak možné orientace pentózo-fosfátové páteře, tak zda se guaniny nachází v *anti* nebo *syn* konformaci.



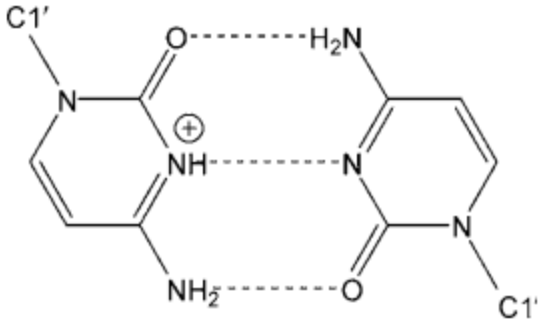
Obrázek 15 - Guaninová tetráda se znázorněným umístěním kationu. Převzato a upraveno z Neidle (2008).



Obrázek 16 - Struktura intramolekulárního tetraplexu 2MS9 (Chung et al., 2015). Pro přehlednost zobrazeny pouze dvě tetrády.

5.3. I-motiv

I-motiv je struktura DNA, která se tvoří u poly(C) vláken za mírně kyselých podmínek, pH 4.2 - 5.2, (Choi et al., 2011). Avšak může se vyskytovat i za fyziologických podmínek v poly(C) oblastech některých telomerických a promotorových sekvencí (Zeraati et al., 2018). V kyselém pH dochází k výskytu protonovaného C-C^+, páru, který tvoří základ duplexu. Následně dochází k interkalaci těchto dvou duplexů, za vzniku i-motivu. Reakce struktury na změny pH z ní činí cíl snah o vyvinutí pH-citlivých nanostruktur (Day et al., 2014)



Obrázek 17 - Párování cytosinu s protonovaným cytosinem. Převzato a upraveno z Neidle (2008).

6. Nástroje strukturní analýzy

Před samotnou klasifikací struktur nukleových kyselin, je třeba jejich strukturu nejprve změřit. Ke stanovení a validaci struktur a vlastností biomakromolekul se používá celá řada spektrálních a difrakčních metod od hmotnostní spektrometrie až po neutronovou difrakci. V následujícím textu však budou zmíněny pouze dvě metody, které jsou pro studium struktury nukleových kyselin nepostradatelné, rentgenová difrakce na krystalech a spektroskopie nukleární magnetické rezonance.

Zároveň je třeba porozumět proč struktury měříme a jak tyto data ukládáme. Krátce bude zmíněna PDB, největší strukturně-biologická databáze, a způsob jakým makromolekulární data ukládáme.

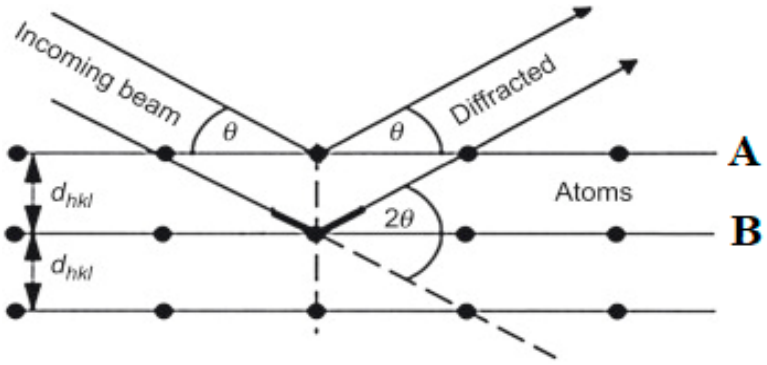
6.1. Difrakce paprsků X na monokrystalech

Rentgenová difrakce (XRD) je metoda využívána k určování chemické struktury krystalických materiálů. Materiál uspořádaný do krystalové mřížky se chová jako difrakční zesilovač pro rentgenové záření, jelikož horní hranice vlnové délky tvrdého rentgenového záření $\lambda = 10^{-10}$ *m*, což přibližně odpovídá délce chemické vazby ($l = 1{,}54x10^{-10}$ *m* pro C-C vazbu v ethanu), a umožňuje vznik interferenčního obrazce, ze kterého je možné vypočítat elektronovou hustotu.

Při srážce rentgenového paprsku s atomem dojde k elastickému rozptylu záření – změní se směr propagace záření aniž by došlo ke změně jeho kinetické energie. Rozptýlené záření spolu následně interaguje, jev který se nazývá interference. Během rozptylu vzniká mezi paprsky záření fázový rozdíl, který závisí na vzdálenosti krystalových rovin *d* (Obrázek 18). Ve většině směrů rozptylu dojde kvůli fázovému rozdílu k destruktivní interferenci, kdy se záření „vyruší“. Ovšem při specifických úhlech dopadajícího záření je rozptýlené záření ve fázi, tedy pod daným úhlem je dráhový rozdíl přesně takový, že odražené záření má celočíselný fázový rozdíl. Přesně tehdy dochází ke konstruktivní interferenci, nebo-li amplifikaci záření v určitých směrech. Takto rozptýlené záření je následně detekováno jako tzv. reflexe (Obrázek 19). Výše popsaný jev popisuje tzv. Braggův zákon:

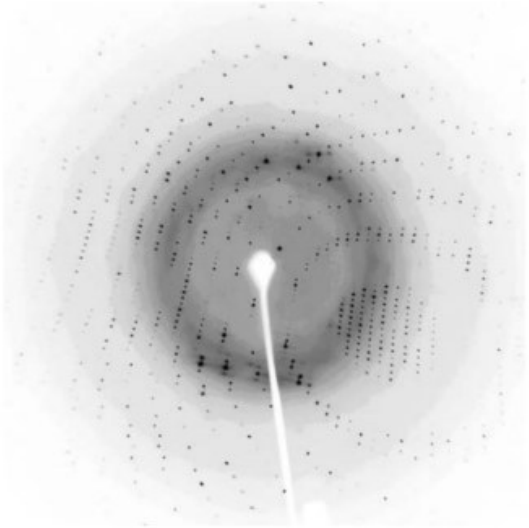
$$2d_{hkl} \sin\theta = n\lambda$$

kde λ je vlnová délka záření a θ úhel, který svírá dopadající i rozptýlené záření s krystalovou rovinou, popř. 2θ je úhel mezi trajektorií dopadajícího záření a rozptýleným zářením. A d_{hkl} je mezirovinná vzdálenost v krystalu.

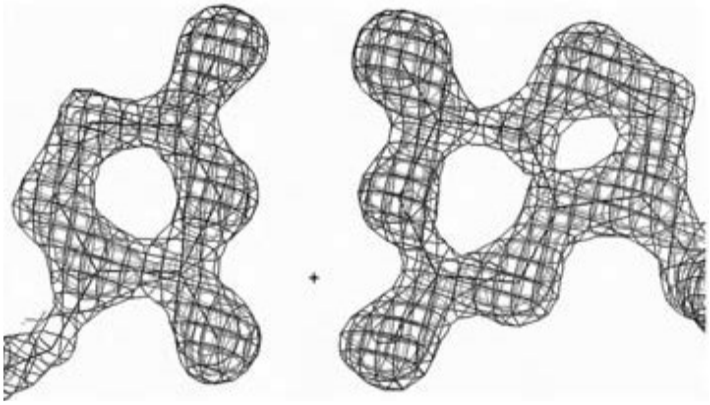


Obrázek 18 - Braggův zákon (Patrick, 2007).

Detekované reflexe jsou charakterizované amplitudou a fází. Fáze reflexí však difrakční experiment nemůže změřit. To je podstatou fázového problému. Jeho řešení spočívá buď v tzv. přímých metodách nebo v molekulárním nahrazení. Mezi přímé metody například patří zavedení těžkého atomu do krystalové mřížky (Taylor, 2003). Pokud je již dostupný homologický model, můžeme k řešení fázového problému využít molekulární nahrazení, kdy známou strukturu použijeme k získání fází (Rossmann & Blow, 1962).



Obrázek 19 - Ukázka difrakčního obrazce biopolymeru. Neidle (2008).



Obrázek 20 - Ukázka mapy elektronové hustoty G-C páru. Neidle (2008).

Z naměřených amplitud a vyřešených fází reflexí je pak z difrakčního obrazce inverzní Fourierovou transformací vypočtena mapa elektronové hustoty základní buňky v bodě (*xyz*):

$$\rho(xyz) = \frac{1}{V}\sum_{hkl} |F_{hkl}| \exp(i\Phi_{hkl}) \exp[-2\pi i(hx + ky + lz)]$$

kde V je objem základní buňky, Φ_{hkl} je fáze reflexe a $|F_{hkl}|$ amplituda strukturního faktoru vyjadřující amplitudu difraktovanou z roviny *hkl* (Taylor, 2003; Ten Eyck, 1985).

Do mapy elektronové hustoty (Obrázek 20) se následně modeluje atomární struktura řešených molekul. Ačkoliv krystalizace biomolekul je často problematická a struktura v krystalu nemusí ve všech detailech odpovídat struktuře molekuly v roztoku, relativní nenáročnost a spolehlivost této metody (oproti NMR a cryo-EM) z ní činí stále nejpouživanější metodu na poli strukturní biologie.

6.2. Spektroskopie nukleární magnetické rezonance

Nukleární magnetická rezonance (NMR) patří k jedné z nejvýznamnějších metod na poli studia molekulové struktury a dynamiky chemických procesů. Spektroskopie NMR je založena na absorpci radiofrekvenčního elektromagnetického (RF) záření, jádry atomů s nenulovým spinem, umístěných v silném magnetickém poli.

Základní podmínkou pro měření NMR spekter je nenulový jaderný spin $I \neq 0$ (^{1}H ^{13}C, ^{15}N, ^{31}P atd.), což vede k tomu, že jádro má magnetický moment. To umožňuje v silném magnetickém poli rozštěpit spinové hladiny (Zeemanův jev), které jsou v nepřítomnosti pole degenerovány. Dodáním energie ve formě RF záření pak můžeme jádra excitovat z jedné hladiny do druhé. Při vrácení do základního stavu je pak přebytečná energie vyzářena. Vyzářená energie je pak tzv. chemický posun, který je definován pro každý nuklid. NMR struktury však nerezonují v jediné frekvenci. To je důsledkem toho, že skutečné magnetické pole, jádra nacházejícího se v molekule, je superpozicí pole vnějšího a vnitřního, které je ovlivňováno diamagnetickým stíněním jader elektrony a spin-spinovými interakcemi sousedních jader. Rozdíly chemického posunu oproti definovanému posunu pro daný nuklid pak dávají informace o detailech chemického okolí daného jádra.

Tato metoda umožňuje struktury analyzovat v roztoku a pozorovat dynamiku struktury, případně i interakce (Markwick et al., 2008). Nevýhodou u biomolekul pak často bývá nutnost obohacení molekuly izotopem ^{13}C, popřípadě dalšími izotopy u vícedimenzionálních NMR, které je často velmi nákladné. NMR je dále limitována velikostí analyzované molekuly,

protože s rostoucí velikostí výrazně roste komplexita signálu, limit pro biopolymery se pohybuje v oblasti 35 kDa (Yu, 1999).

6.3. Práce s daty a databáze

Díky rozvoji spektrálních a difrakčních metod, souběžně s výpočetní technikou, jsme schopni měřit, ukládat a vizualizovat komplikované struktury jako jsou právě nukleové kyseliny či proteiny. Strukturní informace o biologicky významných molekulách pak nachází uplatnění v biotechnologickém, medicínském a materiálovém výzkumu (Burley et al., 2021). Z tohoto důvodu je výhodné mít strukturní data centralizované v přehledné a volně dostupné databázi. Největší databází biomakromolekulárních dat je PDB (Protein Data Bank), která v červenci 2021 obsahuje již více jak 180 000 struktur (Berman et al., 2000).

Každé struktuře v PDB je přiřazen unikátní čtyřmístný kód, první znak je vždy číslo od 1-9 a další tři znaky mohou být buď čísla 0-9 nebo latinská písmena A-Z. Například první B-DNA dodecamer vyřešený krystalografií má kód 1BNA (Drew et al., 1981).

XRD a NMR dohromady tvoří přibližně 95% struktur nyní dostupných na PDB. Elektronová (EM) a kryoelektronová (cryo-EM) mikroskopie také nabývají významu ve strukturní biologii, zejména ve výzkumu ribozomových komplexů. Ovšem ve výzkumu samotných nukleových kyselin figuruje cryo-EM velmi omezeně. Pouze 58 struktur nukleových kyselin, z 3779 struktur dostupných na PDB, bylo změřeno (kryo)elektronovou mikroskopií. Jedním z důvodů je, že EM a cryo-EM zatím neposkytují dostatečné rozlišení pro výzkum oligonukleotidů. Asi 62.1% struktur nukleových kyselin na PDB je řešeno pomocí XRD, a zbylých <38% primárně pomocí NMR spektroskopie (Tabulka 2).

Metoda	#	%
XRD	2347	62.1%
NMR	1366	36.1%
(cryo)EM	58	1.5%
Ostatní	11	0.3%
Celkem	3782	

Tabulka 2 – Zastoupení struktur nukleových kyselin v PDB podle experimentální metody (Berman et al., 2000).

6.4. PDB a mmCIF formáty

Jádrem strukturních dat jsou v zásadě kartézské souřadnice jednotlivých atomů v molekule. Ty jsou ve strukturních databázích doplněny popisy meziatomových interakcí, které dávají vzniknout 3D struktuře molekuly. S postupem digitální archivace makromolekulárních dat se vyvinuly formáty, které umožňují tyto data ukládat, jmenovitě PDB formát (.pdb) a formát pro makromolekulární krystalografická data, nebo-li mmCIF (.cif). Tyto formáty se od sebe v poměrně stěžejních detailech liší, v principu ale oba umožňují kromě ukládání souřadnic i informace o sekundárních strukturách, bibliografické údaje a případně i údaje o řešení, rafinaci a validaci struktury. Formát PDB byl pro archivaci makromolekulárních strukturních dat nahrazoven formátem mmCIF, protože překonává jisté limitace, které jeho předchůdce měl (Adams et al., 2019). Jmenovitě mmCIF formát je schopen lépe reprezentovat velké struktury, komplexní chemické interakce a umožňuje lépe ukládat data z nových/hybridních metod.

```
HEADER    DNA                                     13-MAY-19   6ROS
TITLE     REP RELATED 18-MER DNA
...
COMPND   4 ENGINEERED: YES
SOURCE   2 SYNTHETIC: YES;
...
EXPDTA    X-RAY DIFFRACTION
AUTHOR    P.KOLENKO,J.SVOBODA,B.SCHNEIDER
...
REMARK 200  EXPERIMENT TYPE                : X-RAY DIFFRACTION
REMARK 200  DATE OF DATA COLLECTION        : 08-MAR-18
REMARK 200  TEMPERATURE           (KELVIN) : 100
REMARK 200  PH                             : 7.0
REMARK 200  NUMBER OF CRYSTALS USED        : 1
REMARK 200
REMARK 200  SYNCHROTRON              (Y/N) : Y
REMARK 200  RADIATION SOURCE               : BESSY
REMARK 200  BEAMLINE                       : 14.2
...
SEQRES   1 A   18   DG  DG  DT  DG  DG  DG  DG  DC  DT  DT  DG  DC  DC
SEQRES   2 A   18   DC  DC  DA  DC  DC
...
ATOM      1  O5'  DG A   1     -13.887  -2.537 -12.642  1.00 92.10           O
ATOM      2  C5'  DG A   1     -13.472  -1.836 -11.488  1.00 86.11           C
ATOM      3  C4'  DG A   1     -12.213  -2.444 -10.889  1.00 84.45           C
```

Syntaxe PDB formátu

U .pdb formátu je každému řádku přiřazen klíč (HEADER, REMARK, ATOM, atd.), který je následovaný informacemi příslušící danému klíči, formátovanými dle specifických pravidel. Například v řádcích ATOM jsou právě údaje o jednotlivých atomech ve struktuře. V prvních pěti sloupcích najdeme údaje jako číslování atomů nebo náležitost k reziduu. V dalších třech

pak samotné XYZ souřadnice následované okupancí a B faktorem. Poslední sloupec udává typ atomu.

Syntaxe mmCIF se od formátu PDB výrazně liší. Místo identifikace údajů pomocí klíčů, jsou informace řazeny do takzvaných kategorií. Každá kategorie představuje tabulku, která je pomocí takzvaných klíčů či indexů vztažena k dalším kategoriím. Jednotlivé položky (informace) jsou v kategoriích popisovány tzv. tokeny. Celkově je tak mmCIF formát možné formalizovat jednoduše jako:

```
_kategorie.atribut
```

Data pak mohou být uložena jako jednoduchý pár klíč-hodnota, například:

```
_cell.entry_id  6ROS
```

Komplexnější informace jsou uloženy jako tabulka, kde každý token má více atributů a vytváří se tzv. loop tokeny. Každý atribut pak odpovídá jednomu sloupci hodnot v tabulce. Příkladem je následující tabulka:

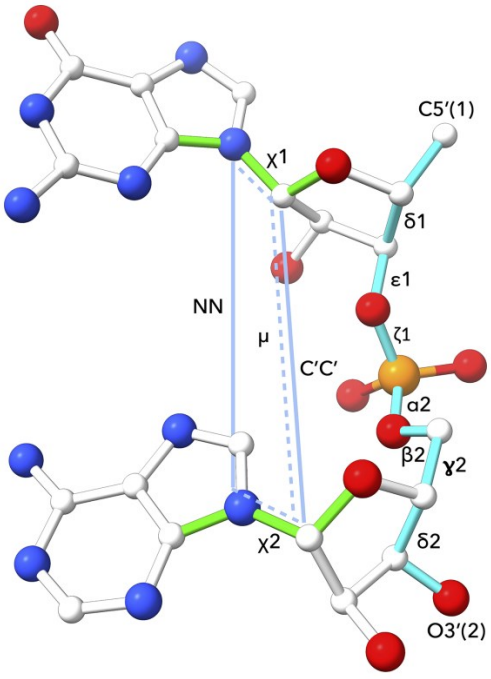
```
#
loop_
_atom_site.group_PDB
_atom_site.id
_atom_site.type_symbol
_atom_site.label_atom_id
_atom_site.label_alt_id
_atom_site.label_comp_id
_atom_site.label_asym_id
_atom_site.label_entity_id
_atom_site.label_seq_id
_atom_site.pdbx_PDB_ins_code
_atom_site.Cartn_x
_atom_site.Cartn_y
_atom_site.Cartn_z
_atom_site.occupancy
_atom_site.B_iso_or_equiv
_atom_site.pdbx_formal_charge
_atom_site.auth_seq_id
_atom_site.auth_comp_id
_atom_site.auth_asym_id
_atom_site.auth_atom_id
_atom_site.pdbx_PDB_model_num
ATOM   1   O  "O5'" . DG A 1 1  ? -13.887 -2.537 -12.642 1.00 92.10  ? 1   DG A "O5'" 1
ATOM   2   C  "C5'" . DG A 1 1  ? -13.472 -1.836 -11.488 1.00 86.11  ? 1   DG A "C5'" 1
ATOM   3   C  "C4'" . DG A 1 1  ? -12.213 -2.444 -10.889 1.00 84.45  ? 1   DG A "C4'" 1
ATOM   4   O  "O4'" . DG A 1 1  ? -11.238 -2.744 -11.929 1.00 83.67  ? 1   DG A "O4'" 1
ATOM   5   C  "C3'" . DG A 1 1  ? -11.489 -1.529 -9.921  1.00 87.48  ? 1   DG A "C3'" 1
...
```

Syntaxe mmCIF formátu

7. Dinukleotidové konformační třídy

Zatímco pro kvalitativní popis DNA mnohdy stačí klasické B, A, Z konformační třídy, pro detailnější popis strukturní plasticity molekuly DNA je třeba rigorózní řešení, které je schopné plného popisu geometrií v nukleových kyselinách (Schneider et al., 2018; Černý et al., 2020).

Za tímto účelem byl vytvořen systém dinukleotidových konformačních tříd, nebo-li NtC (angl. *diNucleotide Conformers*), který spočívá v analýze geometrií dinukleotidových segmentů v rámci jednoho vlákna (Černý et al., 2016). Tento dinukleotidový segment (angl. *dinucleotide step*) je klasifikován na základě dvanácti geometrických parametrů, z toho je sedm úhlů na pentózo-fosfátové páteři α_1, δ_1, ε_1, ζ, α_2, β_2, γ_2 (Obrázek 23), dva glykosidické torzní úhly χ_1 a χ_2 (Černý et al., 2016). Dále pak parametr *NN*, který je definován jako vzdálenost N1/9 prvního nukleotidu mezi N1/9 druhého nukleotidu. Parametr *C'C'* pak popisuje vzdálenost C1'(1) – C1'(2). Dvanáctým parametrem je tzv. pseudo-torzní úhel μ definovaný jako torze mezi glykosidickými vazbami obou nukleotidů.



Obrázek 23 - Parametry určující dinukleotidový konformer. Převzato z Černý et al., 2020.

Pomocí těchto parametrů pak bylo definováno 96+1 konformačních tříd, již výše zmíněných NtC's, tím, že 97. třída NANT je určena pro nezařazené dinukleotidy (Černý et al., 2020). Dinukleotidové konformery jsou pak zařazeny do NtC třídy se kterou mají největší

geometrickou podobnost. Samotné geometrie NtC tříd byly definovány pomocí analýzy 6870 referenčních dinukleotidů. Ty nyní tvoří standardy pro dané NtC třídy (Černý et al., 2021).

Aby byl fragment zařazen, musí každý torzní parametr ležet v rozmezí ±28º od úhlů nejbližšího NtC kandidáta. Dále pak figuruje suma těchto torzních odchylek od samotné NtC třídy. Tedy součet odchylek sedmi páteřních torzních úhlů nesmí být větší než ±60º, stejně to platí pro dva glykosidické úhly. Zároveň hodnoty pseudorotačního úhlu *P* analyzovaného fragmentu se nesmí lišit o více jak ±72º od průměrných hodnot *P* dinukleotidových standardů dané NtC třídy (Schneider et al., 2018).

	NtC	δ1	ε1	ζ1	α2	β2	γ2	δ2	χ1	χ2	μ	NN	C'C'
BI, nebo také kanonická B	BB00	138	183	258	304	180	44	138	253	258	26	4.4	4.9
BI varianta	BB01	131	181	266	301	176	49	120	248	244	26	4.3	4.9
BI-BII konformer	BB05	140	201	214	315	153	46	140	263	253	29	4.6	5.1
BII forma	BB07	144	247	169	297	141	46	141	271	260	46	5.0	5.2
Kanonická A	AA00	82	206	288	293	173	55	82	199	200	18	4.8	5.5
BI-A konformer	BA05	131	184	269	296	169	52	104	251	235	24	4.1	4.6

Tabulka 3 - Příklad vybraných NtC tříd a parametrů, které jsou pro ně definované.

NtC klasifikace má oproti klasickému rozdělení několik výhod. Zaprvé je zcela automatická a matematicky objektivní, poskytuje sjednocený protokol pro DNA i RNA a umožňuje klasifikaci na submolekulární úrovni, tedy různých ohybů a jiných lokálních porušení (Černý et al., 2021).

7.1. Konformační abeceda pro nukleové kyseliny

NtC třídy jsou seskupeny do 14+1 kódů konformační abecedy pro nukleové kyseliny (CANA), pro zjednodušení anotace hlavních strukturních elementů nukleových kyselin. CANA využívá tří-místných kódů k identifikaci konformací. Například standardní A, B, Z konformace mají v CANA kódy AAA, BBB, ZZZ a sdružují NtC konformery, které jsou pro tyto formy typické.

AAA, AAw, AAu jsou CANA kódy pro A-podobné konformery. BBB, BBw, B12, BB2, miB jsou B-konformery. B-A, A-B jsou konformery na pomezí A a B forem a ICL, OPN, SYN jsou atypické třídy popisující například absenci patrového uspořádání bazí, syn orientaci báze nebo velké vzdálenosti mezi bázemi umožňující interkalaci (Tabulka 4)

Hlavní atributy CANA kódu	CANA	počet NtC
A-konformery	AAA	6
A-podobné s netypickými α a γ torzemi	AAw	5
A-podobné se vzdálenými bázemi	AAu	3
Konformery A- do B-formy	A-B	5
Konformery B- do A- formy	B-A	8
B-konformery (kanonická B)	BBB	2
BI-konformery s netypickými α a γ torzemi	BBw	5
Konformery mezi BI- a BII-formou	B12	2
BII konformery	BB2	2
B-podobné s netypickými torzemi	miB	6
Konformery s potenciálně interkalovanými bázemi	ICL	7
Konformery s nestohovanými a často vzdálenými bázemi	OPN	33
Konformery s jednou bazí v -syn orientaci	SYN	6
Z-konformery	ZZZ	6
Nepřiřazené	NAN	1

Tabulka 4 - Tabulka CANA kódů se stručným popisem jejich hlavních strukturních atributů a počtem NtC, které jsou v nich zahrnuty.

7.2. DNATCO

Nástroj DNATCO na adrese https://dnatco.datmos.org/next umožňuje klasifikaci struktur nukleových kyselin právě pomocí univerzální strukturní abecedy, NtC a CANA (Černý et al., 2016). Struktury jsou nahrávány ve formátu mmCIF nebo PDB, buď uživatelem nebo mohou být vyhledány přímo z databáze PDB pomocí příslušného kódu struktury. Server pak přiřadí jednotlivé dinukleotidové konformace pomocí tzv. *k*-NN algoritmu (algoritmus k-nejbližších sousedů), který přiřazuje dinukleotid do NtC třídy s nejbližšími parametry (Cover & Hart, 1967; Čech et al., 2013).

Toto umožňuje relativně snadnou klasifikaci a případnou validaci struktury nukleové kyseliny. K ověření přiřazení k NtC třídě pak funguje tzv. konformační validační skóre – confal, které hodnotí zařazení na základě podobnosti geometrických parametrů (Schneider et al., 2018):

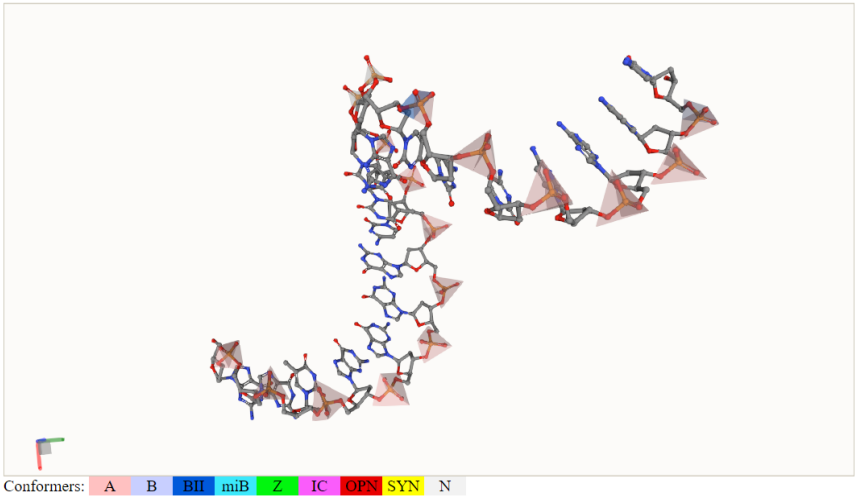
$$\text{confal} = \frac{12}{\sum_{i=1}^{12} \frac{1}{100 \exp\left[-\frac{(t_i - \langle t_i \rangle)^2}{2{c_i}^2}\right]}}$$

kde t_i jsou hodnoty pro analyzovaný dinukleotid a $\langle t_i \rangle$ jsou korespondující hodnoty přiřazeného NtC. Koeficient ${c_i}^2$ je unikátní pro každou NtC třídu tak, aby se confal rovnal 100 pro dinukleotid, jehož geometrie se perfektně překrývá s geometrií přiřazeného NtC. Confal rovný jedné je pak přiřazen geometrii dinukleotidu, který je limitní pro zařazení do NtC třídy.

Další parametr, RMSD (angl. *Root Mean Square Deviation*), posuzuje podobnost analyzovaného dinukleotidu s nejbližším referenčním dinukleotidem dané NtC třídy (Černý et al., 2020). RMSD udává průměrnou kartézskou vzdálenost mezi korespondujícími dvanácti atomy torzního prostoru konformeru a referenčního dinukleotidu:

$$RMSD = \sqrt{\frac{1}{n}\sum_{i=1}^{n} d_i^2}$$

Kde n je počet párů porovnávaných atomů a d_i je kartézská vzdálenost mezi atomy v i-tém páru (Kufareva & Abagyan, 2012). Hodnoty RMSD jsou pak na serveru DNATCO roztříděny do tří kategorií: odchylky od 0-0.5 Å, 0.5-1.0 Å a více než 1.0 Å.



Results of the assignment of 17 detected steps in 1 model(s), can be also downloaded as csv or json file. Found 15/2/0 steps in 0-0.5/0.5-1.0/1.0+Å Cartesian rmsd from reference. Average confal 61, percentile 79.

Click a row in table or a step in viewer for analysis of results. Click column headers to sort data.

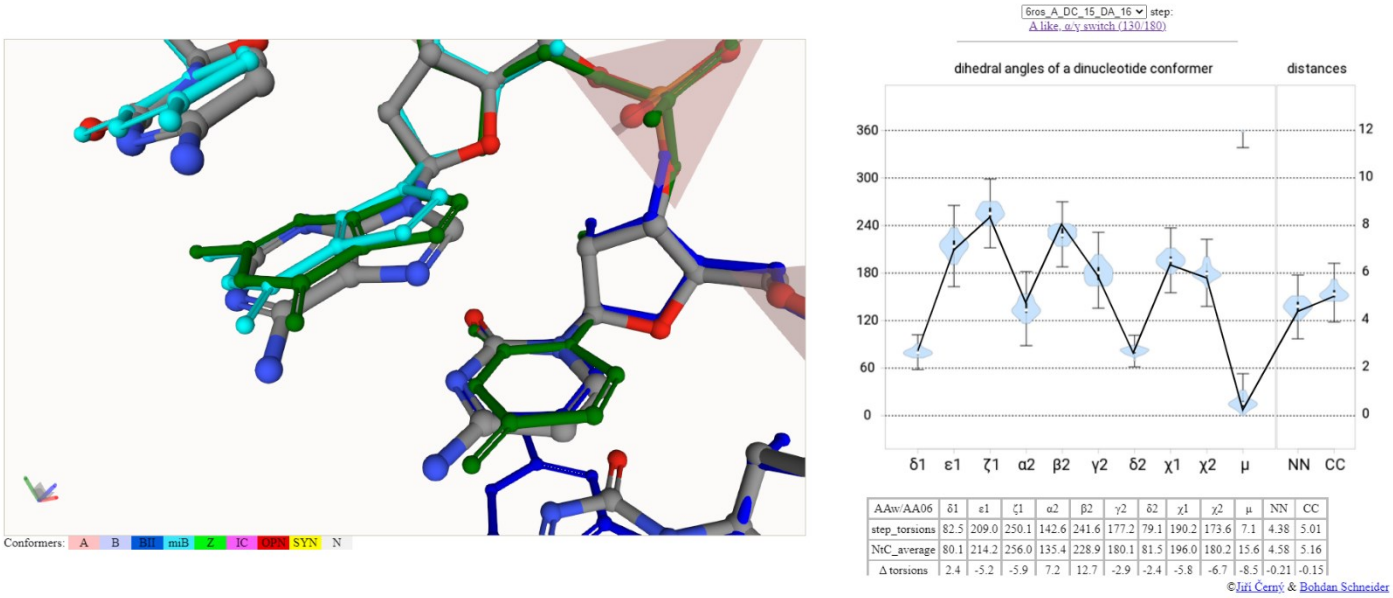
Step name	CANA	NtC	confal	rmsd
6ros_A_DG1_DG2	AAA	AA08	61	0.30
6ros_A_DG2_DT3	AAA	AA00	68	0.24
6ros_A_DT3_DG4	AAA	AA08	87	0.16
6ros_A_DG4_DG5	AAA	AA04	87	0.23
6ros_A_DG5_DG6	AAA	AA00	74	0.17
6ros_A_DG6_DG7	AAw	AA10	75	0.18
6ros_A_DG7_DC8	AAA	AA08	72	0.24
6ros_A_DC8_DT9	AAA	AA00	57	0.46
6ros_A_DT9_DT10	AAA	AA08	20	0.47
6ros_A_DT10_DG11	NAN	NANT	0	0.58
6ros_A_DG11_DC12	NAN	NANT	0	0.55
6ros_A_DC12_DC13	B-A	BA08	41	0.43
6ros_A_DC13_DC14	AAA	AA00	71	0.17
6ros_A_DC14_DC15	AAA	AA08	84	0.26
6ros_A_DC15_DA16	AAw	AA06	88	0.27
6ros_A_DA16_DC17	AAA	AA08	81	0.24
6ros_A_DC17_DC18	A-B	AB05	63	0.23

©Jiří Černý & Bohdan Schneider

Obrázek 24 - Ukázka přiřazení konformací na struktuře 6ROS.

Na obrázku 25 je pak možné vidět detaily o přiřazení segmentu 6ros_A_DC15_DA16 (CA) struktury 6ROS do NtC a CANA, včetně hodnot torzního prostoru daného dinukleotidu a odchylky od referenčních hodnot daného NtC, případně confal hodnoty pro jednotlivé torzní úhly a údaje o pseudorotaci v pentóze. CANA kód daného dinukleotidu je AAw s NtC třídou AA06.

Samotné zobrazení struktury je pak zprostředkováno pomocí webové aplikace pro 3D vizualizaci molekul, Mol* (Sehnal et al., 2021).



AAw/AA06	δ1	ε1	ζ1	α2	β2	γ2	δ2	χ1	χ2	μ	NN	CC
step_torsions	82.5	209.0	250.1	142.6	241.6	177.2	79.1	190.2	173.6	7.1	4.38	5.01
NtC_average	80.1	214.2	256.0	135.4	228.9	180.1	81.5	196.0	180.2	15.6	4.58	5.16
Δ torsions	2.4	-5.2	-5.9	7.2	12.7	-2.9	-2.4	-5.8	-6.7	-8.5	-0.21	-0.15

Obrázek 25 - Klasifikace dinukleotidového segmentu 6ros_A_DC_15_DA_16. Tmavě zelenou barvou je pak označen referenční dinukleotid pro 6ros_A_DC_15_DA_16 a modro-zeleně je vidět referenční dinkleotid pro segment 6ros_A_DA_16_DC_17.

1aoi_J_DC225_DT226	BB2	BB07	79	0.27
1aoi_J_DT226_DG227	BBw	BB11	73	0.35
1aoi_J_DG227_DA228	BBB	BB00	60	0.41
1aoi_J_DA228_DA229	BBB	BB00	94	0.20
1aoi_J_DA229_DC230	miB	BB15	64	0.21
1aoi_J_DC230_DA231	miB	BB10	71	0.32
1aoi_J_DA231_DT232	BBw	BB03	57	0.30
1aoi_J_DT232_DG233	B12	BB04	62	0.35
1aoi_J_DG233_DC234	BBB	BB00	30	0.39
1aoi_J_DC234_DC235	BBB	BB00	18	0.45
1aoi_J_DC235_DT236	BBB	BB00	37	0.34
1aoi_J_DT236_DT237	B12	BB04	81	0.23
1aoi_J_DT237_DT238	BBw	BB03	61	0.38
1aoi_J_DT238_DT239	BBw	BB02	35	0.44
1aoi_J_DT239_DG240	miB	BB13	73	0.22
1aoi_J_DG240_DA241	BBw	BB03	49	0.37
1aoi_J_DA241_DT242	BBw	BB17	26	0.34
1aoi_J_DT242_DG243	BBw	BB11	72	0.28
1aoi_J_DG243_DG244	BBB	BB00	36	0.45
1aoi_J_DG244_DA245	BB2	BB07	80	0.26
1aoi_J_DA245_DG246	BBw	BB11	60	0.37
1aoi_J_DG246_DC247	miB	BB10	21	0.42
1aoi_J_DC247_DA248	NAN	NANT	0	0.44
1aoi_J_DA248_DG249	NAN	NANT	0	0.58
1aoi_J_DG249_DT250	BBB	BB00	22	0.36
1aoi_J_DT250_DT251	BBB	BB00	80	0.28
1aoi_J_DT251_DT252	BBB	BB00	80	0.23
1aoi_J_DT252_DC253	BBB	BB00	61	0.29
1aoi_J_DC253_DC254	B12	BB04	25	0.32
1aoi_J_DC254_DA255	B12	BB04	32	0.37
1aoi_J_DA255_DA256	BBB	BB00	66	0.24
1aoi_J_DA256_DA257	B12	BB04	84	0.12

©Jiří Černý & Bohdan Schneider

Obrázek @ - Strukturní klasifikace DNA duplexu obaleného kolem histonu. Struktura 1AOI (Luger et al., 1997). Struktura má 292 nukleotidů, je tedy ukázána pouze malá výseč přiřazených konformací pro demonstraci konformační abecedy na složitějších stukturách.

8. Závěr

Mezi hlavní cíle práce patřilo seznámit čtenáře s problematikou strukturní biologie a bioinformatiky DNA. Fyziologické implikace většiny strukturních elementů, zde popisovaných, byly zanedbány pro zachování stručnosti textu.

Z tohoto krátkého přehledu je patrné, že jednoduchost ikonického B-DNA duplexu už dnes nestačí k pochopení skutečné komplexity struktury DNA. Ve skutečnosti se jedná o velmi konformačně různorodou molekulu, která je díky svým chemickým a strukturním vlastnostem, jako je párování bazí, π-π interakce či velmi složitý torzní prostor, fascinujícím cílem strukturně biologického a biofyzikálního výzkumu.

Ačkoliv tato práce se tedy primárně zabývá strukturní klasifikací, byla zmíněna i problematika stability, protože je nejen důležité umět určit a popsat strukturu, ale i vědět proč molekula může této struktury nabývat a alespoň kvalitativně chápat termodynamické a elektrostatické děje, které stojí za chováním DNA. Příkladem jsou výše zmiňované π-π interakce, jejichž důležitost pro nukleové kyseliny se stále často zanedbává.

Metody měření struktur jako XRD a NMR tvoří základ snah o to porozumět strukturám biomakromolekul. Kvůli fyzikálním bariérám nelze běžnými způsoby, jako například optickou mikroskopií, sledovat chování a strukturu molekul. Potřebujeme proto speciální spektrální a difrakční techniky, které nám umožňují nahlížet určitým způsobem na danou molekulu. A ačkoliv se může zdát, že výsledkem je vždy ‘pouze’ struktura, tak informace o použité metodě jsou klíčové, kvůli velmi odlišným způsobům jak tyto metody přistupují k řešení struktur.

Závěr práce se pak věnuje samotné strukturní bioinformatice DNA. Je zde představena nová metoda na klasifikaci struktury nukleových kyselin, která umožňuje jednotnou konformační klasifikaci pro DNA i RNA a má potenciál sloužit jako spolehlivý validační nástroj pro nové struktury. Posuzování konformace lokálně, na úrovni dinukleotidu, oproti kvalitativnímu posouzení celého oligo- či přímo polynukleotidu, umožňuje velmi přesně studovat nejen chování samotných nukleových kyselin, ale i jejich případných komplexů s proteiny.

9. Bibliografie

Adams, P. D., Afonine, P. V., Baskaran, K., Berman, H. M., Berrisford, J., Bricogne, G., Brown, D. G., Burley, S. K., Chen, M., Feng, Z., Flensburg, C., Gutmanas, A., Hoch, J. C., Ikegawa, Y., Kengaku, Y., Krissinel, E., Kurisu, G., Liang, Y., Liebschner, D., … Young, J. Y. (2019). Announcing mandatory submission of PDBx/mmCIF format files for crystallographic depositions to the protein data bank (PDB). *Acta Crystallographica Section D: Structural Biology*, *75*(4), 451–454. https://doi.org/10.1107/S2059798319004522

Allawi, H. T., & SantaLucia, J. (1998). Thermodynamics of internal C·T mismatches in DNA. *Nucleic Acids Research*, *26*(11), 2694–2701. https://doi.org/10.1093/nar/26.11.2694

Altona, C., & Sundaralingam, M. (1972). Conformational Analysis of the Sugar Ring in Nucleosides and Nucleotides. a New Description Using the Concept of Pseudorotation. *Journal of the American Chemical Society*, *94*(23), 8205–8212. https://doi.org/10.1021/ja00778a043

Anastassopoulou, J. (2003). Metal-DNA interactions. *Journal of Molecular Structure*, *651–653*, 19–26. https://doi.org/10.1016/S0022-2860(02)00625-7

Arnott, S., & Hukins, D. W. (1972). The dimensions and shapes of the furanose rings in nucleic acids. *The Biochemical Journal*, *130*(2), 453–465. https://doi.org/10.1042/bj1300453

Asensio, J. L., Brown, T., & Lane, A. N. (1999). Solution conformation of a parallel DNA triple helix with 5' and 3' triplex-duplex junctions. *Structure*, *7*(1), 1–11. https://doi.org/10.1016/S0969-2126(99)80004-5

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). The Protein Data Bank. In *Nucleic Acids Research* (Vol. 28, Issue 1, pp. 235–242). Oxford University Press. https://doi.org/10.1093/nar/28.1.235

Boyle, J. (2005). Lehninger principles of biochemistry (4th ed.): Nelson, D., and Cox, M. *Biochemistry and Molecular Biology Education*, *33*(1), 74–75. https://doi.org/10.1002/bmb.2005.494033010419

Burley, S. K., Bhikadiya, C., Bi, C., Bittrich, S., Chen, L., Crichlow, G. V., Christie, C. H., Dalenberg, K., Di Costanzo, L., Duarte, J. M., Dutta, S., Feng, Z., Ganesan, S., Goodsell, D. S., Ghosh, S., Green, R. K., Guranovic, V., Guzenko, D., Hudson, B. P., … Zhuravleva, M. (2021). RCSB Protein Data Bank: Powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Research*, *49*(1), D437–D451. https://doi.org/10.1093/nar/gkaa1038

Čech, P., Kukal, J., Černý, J., Schneider, B., & Svozil, D. (2013). Automatic workflow for the classification of local DNA conformations. *BMC Bioinformatics*, *14*(1). https://doi.org/10.1186/1471-2105-14-205

Černý, Jiři, Boziková, P., Michal Malý, Tykač, M., Biedermannova, L., & Schneider, B. (2020). Structural alphabets for conformational analysis of nucleic acids available at dnatco.datmos.org. *Acta Crystallographica Section D: Structural Biology*, *76*(9), 805–813. https://doi.org/10.1107/S2059798320009389

Černý, Jiří, Božíková, P., & Schneider, B. (2016). DNATCO: assignment of DNA conformers at dnatco.org. *Nucleic Acids Research*, *44*(W1), W284–W287. https://doi.org/10.1093/nar/gkw381

Černý, Jiří, Božíková, P., Svoboda, J., & Schneider, B. (2021). A unified dinucleotide alphabet describing both RNA and DNA structures. *Nucleic Acids Research*, *48*(11), 6367–6381. https://doi.org/10.1093/NAR/GKAA383

Choi, J., Kim, S., Tachikawa, T., Fujitsuka, M., & Majima, T. (2011). PH-induced intramolecular folding dynamics of i-motif DNA. *Journal of the American Chemical Society*, *133*(40), 16146–16153. https://doi.org/10.1021/ja2061984

Cover, T. M., & Hart, P. E. (1967). Nearest Neighbor Pattern Classification. *IEEE Transactions on Information Theory*, *13*(1), 21–27. https://doi.org/10.1109/TIT.1967.1053964

Day, H. A., Pavlou, P., & Waller, Z. A. E. (2014). I-Motif DNA: Structure, stability and targeting with ligands. *Bioorganic and Medicinal Chemistry*, *22*(16), 4407–4418. https://doi.org/10.1016/j.bmc.2014.05.047

Drew, H. R., Wing, R. M., Takano, T., Broka, C., Tanaka, S., Itakura, K., & Dickerson, R. E. (1981). Structure of a B-DNA dodecamer: conformation and dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, *78*(4), 2179–2183. https://doi.org/10.1073/pnas.78.4.2179

Frank-Kamenetskii, M. D., & Mirkin, S. M. (1995). Triplex DNA structures. In *Annual Review of Biochemistry* (Vol. 64, pp. 65–95). Annual Reviews Inc. https://doi.org/10.1146/annurev.bi.64.070195.000433

Franklin, R. E., & Gosling, R. G. (1953). Molecular configuration in sodium thymonucleate. *Nature*, *171*(4356), 740–741. https://doi.org/10.1038/171740a0

Fry, M. (2007). Tetraplex DNA and its interacting proteins. *Frontiers in Bioscience*, *12*(11), 4336–4351. https://doi.org/10.2741/2391

Hoogsteen, K. (1963). The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. *Acta Crystallographica*, *16*(9), 907–916. https://doi.org/10.1107/s0365110x63002437

Hospital, A., Faustino, I., Collepardo-Guevara, R., González, C., Gelpí, J. L., & Orozco, M. (2013). NAFlex: a web server for the study of nucleic acid flexibility. *Nucleic Acids Research*, *41*(W1), W47–W55. https://doi.org/10.1093/NAR/GKT378

Jones, R. G., Kahovec, J., Stepto, R., Wilks Michael Hess, E. S., Kitayama, T., Val Metanomski, W., Jenkins, A., & Kratochvíl, P. (2008). *Compendium of Polymer Terminology and Nomenclature IUPAC Recommendations 2008 With advice from : Compendium of Polymer Terminology and Nomenclature IUPAC Recommendations 2008*. 1–465. www.rsc.org/books

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., Yu, B., Zaslavsky, L., Zhang, J., & Bolton, E. E. (2021). PubChem in 2021: New data content and improved web interfaces. *Nucleic Acids Research*, *49*(D1), D1388–D1395. https://doi.org/10.1093/nar/gkaa971

Kolenko, P., Svoboda, J., Černý, J., Charnavets, T., & Schneider, B. (2020). Structural variability of CG-rich DNA 18-mers accommodating double T-T mismatches. *Acta Crystallographica Section D: Structural Biology*, *76*(Pt 12), 1233–1243. https://doi.org/10.1107/S2059798320014151

Kufareva, I., & Abagyan, R. (2012). Methods of protein structure comparison. *Methods in Molecular Biology*, *857*, 231–257. https://doi.org/10.1007/978-1-61779-588-6_10

Laughlan, G., Murchie, A. I. H., Norman, D. G., Moore, M. H., Moody, P. C. E., Lilley, D. M. J., & Luisi, B. (1994). The high-resolution crystal structure of a parallel-stranded guanine tetraplex. *Science*, *265*(5171), 520–524. https://doi.org/10.1126/science.8036494

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, *389*(6648), 251–260. https://doi.org/10.1038/38444

Luo, Z., Dauter, M., & Dauter, Z. (2014). Phosphates in the Z-DNA dodecamer are flexible, but their P-SAD signal is sufficient for structure solution. *Acta Crystallographica Section D: Biological Crystallography*, *70*(7), 1790–1800. https://doi.org/10.1107/S1399004714004684

Markwick, P. R. L., Malliavin, T., & Nilges, M. (2008). Structural biology by NMR: Structure, dynamics, and interactions. *PLoS Computational Biology*, *4*(9), e1000168. https://doi.org/10.1371/journal.pcbi.1000168

Meints, G. A., Karlsson, T., & Drobny, G. P. (2001). Modeling furanose ring dynamics in DNA. *Journal of the American Chemical Society*, *123*(41), 10030–10038. https://doi.org/10.1021/ja010721d

Neidle, S. (2008). Principles of Nucleic Acid Structure. In *Principles of Nucleic Acid Structure*. Elsevier Inc. https://doi.org/10.1016/B978-0-12-369507-9.X5001-8

Olson, W. K. (1981). Three-state models of furanose pseudorotation. *Nucleic Acids Research*, *9*(5), 1251–1262. https://doi.org/10.1093/nar/9.5.1251

Patrick, M. H. (2007). Crystallography made crystal clear: A guide for users of macromolecular models (3rd Ed.). *Biochemistry and Molecular Biology Education*, *35*(5), 387–388. https://doi.org/10.1002/bmb.89

Rossmann, M. G., & Blow, D. M. (1962). The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallographica*, *15*(1), 24–31. https://doi.org/10.1107/s0365110x62000067

Schneider, B., Božíková, P., Nečasová, I., Čech, P., Svozil, D., & Černý, J. (2018). A DNA structural alphabet provides new insight into DNA flexibility. *Acta Crystallographica Section D: Structural Biology*, *74*(Pt 1), 52–64. https://doi.org/10.1107/S2059798318000050

Schneider, B., Neidle, S., & Berman, H. M. (1997). *Conformations of the sugar-phosphate backbone in helical DNA crystal structures*. Biopolymers - Nucleic Acid Sciences Section. https://doi.org/10.1002/(sici)1097-0282(199707)42:1<113::aid-bip10>3.0.co;2-o

Schumacher, M. A., Piro, K. M., & Xu, W. (2010). Insight into F plasmid DNA segregation revealed by structures of SopB and SopB-DNA complexes. *Nucleic Acids Research*, *38*(13), 4514–4526. https://doi.org/10.1093/nar/gkq161

Schweitzer, B. A., & Kool, E. T. (1995). Hydrophobic, Non-Hydrogen-Bonding Bases and Base Pairs in DNA. *Journal of the American Chemical Society*, *117*(7), 1863–1872. https://doi.org/10.1021/ja00112a001

Sehnal, D., Bittrich, S., Deshpande, M., Svobodová, R., Berka, K., Bazgier, V., Velankar, S., Burley, S. K., Koča, J., & Rose, A. S. (2021). Mol* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkab314

Sundaralingam, M. (1969). Stereochemistry of nucleic acids and their constituents. IV. Allowed and preferred conformations of nucleosides, nucleoside mono-, di-, tri-, tetraphosphates, nucleic acids and polynucleotides. *Biopolymers*, *7*(6), 821–860. https://doi.org/10.1002/bip.1969.360070602

Svozil, D., Hobza, P., & Šponer, J. (2010). Comparison of intrinsic stacking energies of ten unique dinucleotide steps in A-RNA and B-DNA duplexes. Can we determine correct order of stability by quantum-chemical calculations? *Journal of Physical Chemistry B*, *114*(2), 1191–1203. https://doi.org/10.1021/jp910788e

Taylor, G. (2003). The phase problem. *Acta Crystallographica - Section D Biological Crystallography*, *59*(11), 1881–1890. https://doi.org/10.1107/S0907444903017815

Ten Eyck, L. F. (1985). Fast Fourier transform calculation of electron density maps. *Methods in Enzymology*, *115*(C), 324–337. https://doi.org/10.1016/0076-6879(85)15024-X

Van Dam, L., Ouwerkerk, N., Brinkmann, A., Raap, J., & Levitt, M. H. (2002). Solid-state NMR determination of sugar ring pucker in 13C-labeled 2′-deoxynucleosides. *Biophysical Journal*, *83*(5), 2835–2844. https://doi.org/10.1016/S0006-3495(02)75292-2

Wang, G., & Vasquez, K. M. (2007). Z-DNA, an active element in the genome. *Frontiers in Bioscience*, *12*(12), 4424–4438. https://doi.org/10.2741/2399

Watson, J. D., & Crick, F. H. C. (1953). Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, *171*(4356), 737–738. https://doi.org/10.1038/171737a0

Wilkins, M. H. F., Stokes, A. R., & Wilson, H. R. (1953). Molecular structure of nucleic acids: Molecular structure of deoxypentose nucleic acids. *Nature*, *171*(4356), 738–740. https://doi.org/10.1038/171738a0

Xiong, Y., & Sundaralingam, M. (1998). Crystal structure and conformation of a DNA–RNA hybrid duplex with a polypurine RNA strand: d(TTCTTBr5CTTC)–r(GAAGAAGAA). *Structure*, *6*(12), 1493–1501. https://doi.org/10.1016/S0969-2126(98)00148-8

Yakovchuk, P., Protozanova, E., & Frank-Kamenetskii, M. D. (2006). Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Research*, *34*(2), 564–574. https://doi.org/10.1093/nar/gkj454

Yu, H. (1999). Extending the size limit of protein nuclear magnetic resonance. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(2), 332–334. https://doi.org/10.1073/pnas.96.2.332

Zeraati, M., Langley, D. B., Schofield, P., Moye, A. L., Rouet, R., Hughes, W. E., Bryan, T. M., Dinger, M. E., & Christ, D. (2018). I-motif DNA structures are formed in the nuclei of human cells. *Nature Chemistry*, *10*(6), 631–637. https://doi.org/10.1038/s41557-018-0046-3