UNIVERZITA PALACKÉHO V OLOMOUCI

Přírodovědecká fakulta

Katedra fyzikální chemie



Bioinformatická studie konformací páteře RNA molekul

BAKALÁŘSKÁ PRÁCE

Autor:Viktor ChromýStudijní program:B1407 ChemieStudijní obor:Aplikovaná chemieForma studia:PrezenčníVedoucí práce:Mgr. Petra Kührová, Ph.D.Rok:2019

Prohlášení autora práce

Prohlašuji, že jsem bakalářskou práci vypracoval/a samostatně s vyznačením všech použitých pramenů a spoluautorství. Souhlasím se zveřejněním bakalářské práce podle zákona č. 111/1998 Sb., o vysokých školách, ve znění pozdějších předpisů. Byl/a jsem seznámen/a s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorský zákon, ve znění pozdějších předpisů.

V Olomouci dne

Poděkování:

Rád bych poděkoval vedoucí mé bakalářské práce Mgr. Petře Kührové, Ph.D. za veškerou pomoc, rady, ochotu, zkušenosti a také za její trpělivost. Dále děkuji doc. Mgr. Pavlu Banáši Ph.D. za vytvoření skriptu a další pomoc. V neposlední řadě patří dík mé rodině a přátelům za podporu během celého studia.

Bibliografická identifikace

Jméno a příjmení autora	Viktor Chromý				
Název práce	Bioinformatická studie konformací páteře RNA molekul				
Typ práce	Bakalářská				
Pracoviště	Katedra fyzikální chemie				
Vedoucí práce	Mgr. Petra Kührová, Ph.D.				
Rok obhajoby práce	2019				
Abstrakt	 Z019 Tato bakalářská práce se zabývá analýzou páteřních konformací RNA. V teoretické části práce jsou charakterizovány základní vlastnosti RNA, páteř RNA a její konformace a také třízení do rodin podle Richardsona. Pak také v teoretické části najdeme informace o PDB databázi a o experimentálních metodách měření zpracovávaných dat. V praktické části jsou popsány zpracovávaná data a metodologie jejich zpracování. Následně jsou prezentovány výsledky deseti nejčetnějších rodin. Závěrečná diskuze mluví o využití a možnosti budoucího zkoumání. 				
Klíčová slova	RNA, páteř, rodina				
Počet stran	55				
Počet příloh	0				
Jazyk	Český				

Bibliographical identification

Autor's	first	name	and	Viktor Chromý				
surname								
Title				Bioinformatics study of backbone conformations of RNA molecules				
Type of t	hesis			Bachelor				
Departme	ent			Department of psysical chemistry				
Superviso	or			Mgr. Petra Kührová, Ph.D.				
The year of presentation				2019				
Abstract				This bachelor thesis deals with analysis of backbone conformations of RNA. The theoretical part of thesis describes basic properties of RNA, RNA backbone, and its conformations as well as separation into suits by Richardson. Moreover, the theoretical part includes information about PDB database and about experimental methods of measurement of processed data. In practical part, processed data are described and methodology of their treatment. The results of most frequent suits are presented. Final discussion talks about usage and possibility of future research.				
Keyword	S			RNA, backbone, suite				
Number of	of pages	5		55				
Number of	of apper	ndices		0				
Language Czech								

Obsah

ÚV(OD	7
Ι	TEORETICKÁ ČÁST	8
1	NUKLEOVÉ KYSELINY	9
1.1	STAVBA NUKLEOVÝCH KYSELIN	9
1.2	TYPY RNA	10
	1.2.1 Transferová tRNA	11
	1.2.2 MEDIÁTOROVÁ MRNA	11
	1.2.3 RIBOZOMÁLNÍ RRNA	12
2	KLASIFIKACE RODIN	13
2.1	ROZDĚLENÍ PODLE RICHARSONA	16
	2.1.1 POJMENOVÁVÁNÍ KONFORMACÍ RODIN PÁTEŘE RNA	16
3	PROTEINOVÁ DATOVÁ BANKA	18
3.1	HISTORIE PDB	18
4	METODY EXPERIMENTÁLNÍHO MĚŘENÍ DAT	19
4.1	RENTGENOVÁ ANALÝZA	19
4.2	NUKLEÁRNÍ MAGNETICKÁ REZONANCE	19
II	PRAKTICKÁ ČÁST	20
5	POUŽITÁ DATA A METODOLOGIE ZPRACOVÁNÍ DAT	21
6	VÝSLEDKY	22
7	ZÁVĚR	49
8	SUMMARY	50
SEZ	NAM POUŽITÉ LITERATURY	51

ÚVOD

RNA neboli ribonukleová kyselina je makromolekula nacházející se v každé buňce. Je to nezbytná součást veškeré živé hmoty. Existuje velké množství druhů RNA plnících řadu funkcí. Nejvýznamnějším úkolem RNA je kopírování a částečně i nesení genetické informace. RNA je jednovláknitá molekula nabývá však mnoha komplexních sekundárních struktur. [1,2,3]

Fosfodiesterová páteř RNA je částí, která určuje mnohé vlastnosti této makromolekuly. Páteř je popsána pomocí torzních úhlů. Pro snadnější práci s páteří a těmito úhly byla vytvořena nomenklatura popisující páteř s úhly. Způsobů, jak dělit páteř RNA a jak tyto části popisovat bylo dříve více, avšak pro zjednodušení chápání studovaných struktur bylo třeba vytvořit jednotnou nomenklaturu. To se nakonec stalo v roce 2008 kdy Richardson et. al. vytvořili systém takzvaných *rodin*. Tyto rodiny byly stanoveny jako konformace od cukru po cukr. Z analýzy všech úhlů nakonec vyplynulo 46 různých druhů konformací. Některé z těchto rodin se vyskytují často, jiné jen ve velmi ojedinělých případech. [9,20]

Cílem této práce bylo zpracovat experimentálně získaná data, a pozorovat rozdíly mezi rodinami. Zpracovávaná data byla získána z Non-redundantní PDB databáze obsahující komplexní informace. Obzvlášť tedy se zaměřením na rodiny **1a** a **1c** jelikož tyto rodiny jsou nejčastější a také hojně využívané při molekulárním modelovaní.

I TEORETICKÁ ČÁST

1 NUKLEOVÉ KYSELINY

Nukleové kyseliny jsou v hierarchii látek potřebných pro existenci živé hmoty na nejvyšším místě. Určují genetické vlastnosti všeho živého a ovlivňují její organizaci a reprodukci. Představují látku, která nese informace pro průběh všech životních procesů. [1]

1.1 Stavba nukleových kyselin

Nukleové kyseliny jsou lineární polymery, tvořené několika stavebními jednotkami s daným pořadím. Základní stavební částicí každé jednotky je pětiuhlíkatý monosacharid. V případě RNA je to β -D-ribóza. U DNA je tímto sacharidem β -D-2-deoxyribóza. Další složku mají obě tyto kyseliny společnou a tou je zbytek kyseliny fosforečné tzv. fosfátová skupina. Tato fosfátová skupina propojuje pomocí fosfodiesterové vazby pentózy přesněji jejich 3' a 5' uhlíky. Monosacharidové jednotky a fosfátové skupiny se navzájem střídají a tvoří tak pentóza-fosfátovou páteř, jež mají obě nukleové kyseliny shodnou. Fosfátové skupiny nacházející se v neutrálním prostředí jsou plně disociovány a nesou tedy záporný náboj. [2]

Třetí a zároveň poslední složkou nukleových kyselin jsou nukleové báze. Ty jsou z informačního hlediska rozhodující složkou. Z chemického pohledu jde o deriváty dvou heterocyklických dusíkatých sloučenin, a to purinu a pyrimidinu. Tyto látky mají slabě zásaditý charakter, z čehož je také odvozeno označení báze. Báze, jež mají jako dusíkatý heterocyklus purin jsou adenin (označován A) a guanin (označován G). Pyrimidinové báze jsou cytosin (označován C), thymin (označován T) a uracil (označován U). Zde můžeme vidět další rozdíl mezi DNA a RNA a to v tom, že v RNA se vyskytuje uracil a v DNA je místo něj zastoupen thymin, jenž je methylovaným derivátem uracilu. [2]

Kovalentní struktura RNA se liší od té v DNA ve dvou aspektech. Jak bylo zmíněno dříve, a jak také ukazuje název, monosacharidová jednotka v RNA je ribóza namísto deoxyribózy. Ribóza obsahuje 2'-hydroxylovou skupinu, který není přítomná v deoxyribóze. Následkem toho mohou v RNA navíc od standardních $3' \rightarrow 5'$ vazeb vznikat také $2' \rightarrow 5'$ vazby. Tato vazba má vliv na odstraňování

intronů a přidávání exonů při formování mRNA. Každý fosfodiesterový můstek má negativní náboj. Tento negativní náboj odpuzuje nukleofilní skupiny jako například hydroxidový ion. Následkem tohoto je, že fosfodiesterové vazby jsou mnohem méně náchylné k hydrolitickým reakcím než jiné estery jako například estery karboxylových kyselin. Tato odolnost je velmi důležitá po udržování celistvosti informace uchovávané v nukleové kyselině. Absence 2'-hydroxilové skupiny v DNA následně ještě zvyšuje tuto odolnost k hydrolýze. To je pravděpodobně důvodem, proč je DNA nositelem genetické informace na rozdíl od RNA. [3]

Název nukleosidy nesou ty glykosidy, na nichž jsou pomocí β -glykosidové vazby na ribóze nebo deoxyribóze navázány nukleové báze. Tato vazba se vždy vyskytuje mezi C1 atomem (ζ deoxy)-ribózy a mezi N9 dusíkem (viz obr. 1) purinových bází či N1 dusíkem pyrimidinových bází (viz obr. 1). [2]



Obrázek 1 Strukturní vzorec RNA ("R" značí purinovou nebo pirimidinovou bázi)

1.2 Typy RNA

Přestože je RNA jednovláknová molekula, některé z jejich modifikací se skládají do sekundárních struktur. Tyto sekundární struktury vznikají párováním bází a vzniku vnitrořetězcových helikálních segmentů. Molekula RNA se v buňkách běžně vyskytuje v třech základních typech a to ribosomální (rRNA), transferová (tRNA) a mediátorová (mRNA). [4]

1.2.1 Transferová tRNA

Transferových RNA je řada druhů a jsou obsaženy v každé buňce a její podíl je 10-15% celkového obsahu RNA. Její hlavní funkcí je vazba aminokyselin což zahrnuje aktivaci a označení antikodonem. Tyto molekuly RNA hrají významnou roli v syntéze polypeptidových řetězců bílkovin. Sekundární struktura všech tRNA má tvar připomínající jetelový list. Tato struktura se skládá ze čtyř helikálních částí, v nichž jsou komplementární báze spárovány pomocí vodíkových můstků. Tři z těchto čtyř úseků jsou zpětné smyčky. V molekule tRNA můžeme nalézt minoritní báze jako například methylguanin, dimethylguanin, methylinosin či dihydrouracil. Terciální struktura tRNA je ve tvaru písmene L. Pro správnou funkci musí tRNA splňovat čtyři podmiňující klíčové vlastnosti:

- Schopnost navázat v molekule mRNA se na specifický kodon. K tomuto navázáni dochází prostřednictvím trojice nukleotidů (antikodon) jež se nacházejí v antikodonové smyčce.
- Vázat kovalentně aminokyselinu. Toto je zajištěno pomocí akceptorové sekvence na 3'-konci řetězce skládajícího se z nukleotidů CCA. Na volný 3'-hydroxyl se váže aminokyselina esterovou vazbou.
- Rozpoznatelnost enzymem, který katalyzuje vznik kovalentní vazby mezi tRNA a aminokyselinou.
- Návaznost na ribozomy. [2]

1.2.2 Mediátorová mRNA

Podíl mRNA z celkové RNA v buňce je 5-10%, její strukturou je rozvinutý řetězec. Molekulová hmotnost i počet nukleotidů je různorodý z toho důvodu, že vniká pomocí přepisů genů transkripcí. [4]

Funkcí mRNA je přenos informace o deoxynukleotidu uložené v genu, která slouží v procesu syntézy proteinů. Toto je důvodem jejího pojmenování jako messanger. Životnost mRNA je relativně krátká. V buňkách bakterií přežívá pouze několik sekund až hodinu a v eukaryotických buňkách hodiny až dny. Toto přináší výhodu v možnosti pozměňování syntézy bílkovin. [2]

1.2.3 Ribozomální rRNA

Ribozomální RNA je nejvíce zastoupenou RNA v buňce a to s obsahem okolo 80% z celkové RNA v buňce. Má tři různé molekulové hmotnosti a počty nukleotidů na základě toho o jakou ze tří rRNA se jedná. 28S rRNA má molekulovou hmotnost 1200 kDa a 3700 nukleotidů, 18S rRNA má molekulovou hmotnost 550 kDa a 1700 nukleotidů a 5S rRNA má molekulovou hmotnost 36 kDa a 120 nukleotidů. Je stavebním materiálem ribozomů a přímo se účastní translace. [4]

Ribozomální RNA vytváří nadmolekulovou strukturu ribozomů a to společně s proteiny. V prokaryontních organismech nacházíme tři odlišné rRNA a to 23S, 16S a 5S. Zato u eukaryotních buněk nalézáme čtyři rozdílné druhy rRNA a těmi jsou 28S, 18S, 5,8S a 5S. V nukleoproteinovém souboru ribozomu vytváří všechny druhy rRNA speciální sekundární struktury. Některé části řetězců těchto rRNA jsou komplementárně propojeny pomocí vodíkových můstků a vytvářejí helikální segmenty. Dále se ve strukturách vyskytují zpětné smyčky a výdutě. Je zajímavé, že struktury rRNA z různých organismů jsou velmi obdobné, a to i když se jejich nukleotidové sekvence relativně liší. [2]

2 KLASIFIKACE RODIN

Z dnešních znalostí RNA víme, že její buněčná role je daleko rozsáhlejší, než jenom přenos genetické informace, ale zahrnuje také katalýzu, rozpoznávání molekul a také genetickou kontrolu. Tím pádem se RNA může chovat jako složená makromolekula s překvapivými paralelami k proteinům. Znalost třídimenzionální struktury RNA může být tedy kritická pro porozumění strukturních mechanismů, jenž se podílejí na konformačních změnách RNA, vázání ligandů a proteinů a katalýze. Navzdory tomu, že zájem o studium terciálních struktur RNA stoupá, vývoj a úspěch výpočetních nástrojů pro strukturní modelování RNA zaostávají za jejími protějšky v případě proteinů. Toto je částečně způsobeno náročností určování experimentálních struktur RNA s dostatečným rozlišením a také spletitostí volných torzních úhlů RNA páteře v každém nukleotidu. V posledních dvou dekádách množství, rozmanitost, a kvalita řešených struktur RNA ohromně rozrostla a toto umožnilo, aby byli struktury RNA páteře a bází studovány ve větším detailu. [5,6,7,8,9]

Fosfodiesterovou páteř RNA můžeme popsat pomocí šesti torzních úhlů pojmenovaných α , β , γ , δ , ε , ζ (obrázek 2). Torzní úhel je úhel mezi dvěma rovinami kdy každá rovina je tvořena třemi atomy. Tyto dvě roviny mají vždy dva atomy společné. Torzní úhly jsou tedy tvořeny čtyřmi atomy. Páteřní úhly mezi sebou korelují a jejich hodnoty jsou vynuceny samotnou stavbou RNA. Hodnoty těchto úhlů se udávají buď v rozsahu 0° až 360°, nebo -180° až 180°. Ne všechny hodnoty jsou pro daný úhel dovolené a jejich povolené hodnoty ukazuje Obrázek 3. [10,11,12]



Obrázek 2 Páteř RNA s torzními úhly a rozdělením na rodinu, nukleotid, dinukleotid a dva heminukletoidi. 1. heminukleotid je popsán písmenem, 2. heminukleotid je popsán číslem



Obrázek 3 Konformační kotouč ukazující možné rozsahy úhlů v nukleotidech. (Převzato z [11])

I když jsou párování a skládání bází dominantnějšími determinanty 3D struktur RNA, tak specifické konformace a interakce RNA páteře jsou nezbytné pro RNA katalýzu, vázání léků a aptamerů a nebo pro interakce mezi RNA a proteiny. Získat úplné údaje o všech atomech v případě páteře RNA je mnohem náročnější než je tomu u bází avšak tyto data jsou velmi žádaná, jelikož interagující molekuly vnímají RNA v celém detailu a se všemi jejími parametry.

Při určování struktur RNA molekul pomocí rentgenové krystalografie či nukleární magnetické rezonance je relativně snadné lokalizovat báze a to jako velké rigidní jednotky. Obě tyto strukturní techniky mají však daleko větší problém s páteří RNA a to převážně kvůli velké proměnlivosti mnoha torzních úhlů v každé podjednotce nukleové kyseliny. Toto vede k velkému množství šumu v datech a také k systematickým chybám v databázi pozorovaných konformací.

Analýza páteřních struktur RNA je poměrně velký problém pro uspokojivé vyřešení pomocí automatické analýzy. To je způsobeno velkou proměnlivostí dat, multidimenzionální podstatou páteře RNA a také velkým množstvím šumu v těchto datech. Často je tedy třeba využít zjednodušení těchto dat nebo využívané metody. [13,14,15]

Tento problém se rozhodlo vyřešit několik skupin vědců vytvořením klasifikace páteře RNA v úplné nebo téměř úplně atomární přesnosti. *Hershkovitz et al.* (2003) analyzovali ribosomální RNA pomocí spojení hodnot jednotlivých α , γ , δ a ζ torzních úhlů v nukleotidu fosfát k fosfátu. Tento tým vědců také jako první přišel s ideou pojmenováním konformace pomocí jednoznakového názvu, který by mohl být využit pro vyhledávání a pojmenovávání jednotlivých struktur. Definovali abecedu 37 nukleotidových konformací. Tato skupina také vyvinula multiresoluční přístup, kde se řešení mění pomocí zjednodušení skupin atomů RNA do *pseudo* objektů majících umístění a orientaci. *Murray et al.* (2003) navrhli jednotku od cukru po cukr neboli tak zvanou rodinu jak může být vidět na obrázku číslo 4. Byla jimi vytvořena také rozsáhlá databáze RNA na úrovni těchto rodin, kde nakonec definovali 42 páteřních rotamerů. *Schnider et al.* (2004) studovali dinukleotidové jednotky s třemi fosfáty a dvěma bázemi, obsahující také χ torzní úhel báze,

s celkovým počtem čtrnácti torzních úhlu. Definovali 32 dinukleotidových konformací. [12,16,17,18]

Nejběžnější páteřní konformace identifikované v těchto třech studiích byly shodné navzdory odlišným technikám, které byly použity. Tyto tři skupiny vědců se následně spojily, aby přezkoumali a přehodnotily jejich předchozí definice a výsledky a vytvořili jeden společný a konsensuální systém. [19]

2.1 Rozdělení podle Richarsona

Richardson et al. (2008) na základě všech předchozích výzkumů vytvořili jednotný systém, který je navržen tak, aby poskytoval stručnou a čistě definovanou nomenklaturu potřebnou jak pro databázi, tak pro výpočetní použití a také umožňoval překlad a porovnání s rezidui, rodinami a dinukleotidovýmy systémy z předchozích výzkumů. Tato nomenklatura je definována pro struktury RNA páteře za použití všech torzních úhlů páteře. Použitím několika nezávislých analytických metod bylo identifikováno 46 diskrétních často se vyskytujících konformací. Většina těchto konformací odpovídá určitým vlastnostem RNA. [20]

2.1.1 Pojmenovávání konformací rodin páteře RNA

Těmto konformacím bylo přiděleno jméno obsahující dva znaky, a to číslo a malé písmeno (např. "**1a**"). Toto jméno odráží kombinaci všech sedmi dihedrálních úhlů $\delta \epsilon \zeta \alpha \beta \gamma \delta$, které se nacházejí v tak zvané "rodině", což je jednotka od cukru po cukr (viz obr. 4). Každý ze znaků popisuje jeden z heminukleotidů. Číslo anebo číslu podobný znak popisuje úhly $\delta \epsilon \zeta$ (např. "**1**", "**3**", "**&**") v první části konformace a písmeno nebo písmenu podobný znak popisuje úhly $\alpha \beta \gamma \delta$ (např. "**a**", "**d**", "[") v druhé části konformace. Díky tomu může být tento stavebnicový systém použit také pro popis tradičních nukleotidových jednotek (např. "**a**1") nebo k popisu dinukleotidů (např. **a1a1**). První nukleotidový zbytek, který formuje část rodiny je uváděn jako *i-1* a druhý nukleotidový zbytek je označován jako *i* (viz obr. 6).

Tato dvouznaková označení by měla být jednoduše považováno jako stručná, napůl libovolná jména pro oblasti v sedmidimenzionální prostoru dihedrálních úhlů. Specifické číslo a písmeno pro každou konformaci rodiny znázorňuje průměrné hodnoty všech sedmi dihedrálních úhlů.

Názvy konformací byly vytvořeny tak, aby reflektovali pozice cukrů v rodině. Liché číslo v názvu konformace rodiny udává střední hodnotu pro δ *i-1* dihedrální úhel u ribózy, s C3'-endo konformací. Sudé číslo tedy na druhou stranu udává střední hodnotu pro δ *i-1* dihedrální úhel u ribózy, s C2'-edno konformací. Pro účely tohoto systému je individuální ribóza s δ v rozsahu od 55° do 110° považována jako C3'-endo konformace a ribóza s δ od v rozsahu od 120° do 175° jako C2'-endo konformace. V těchto hodnotách δ se však mohou také vyskytovat i jiné konformace cukrů, avšak je velmi málo dobře ověřených výsledků při vysokém rozlišení, které byly analyzovány jednotlivě. V tomto systému bylo použito dvanáct numerických k pojmenovávání znaků rodin. Těmito znaky jsou 0;1;2;3;4;5;6;7;8;9;#;&. [20]

Druhý znak názvu konformace rodiny, malé písmeno, popisuje konformaci cukru *i*. První půlka abecedy (**a**, **c**-**n**) popisuje rodiny C3'-endo konformací cukru. Druhá půlka abecedy (**b**,**o**-**z** a [) popisuje rodiny s C2'-endo konformací cukru. Velká písmena byla ponechána pro specifikování sekvence bází. Dvanáct znaků bylo připsáno pro heminukleotidy končící se středními hodnotami δ , které indikují C3'-endo konformaci. Devět znaků pro C2'- endo konformace. Každé písmeno specifikuje určitou kombinaci hodnot úhlů $\alpha\beta\gamma\delta$, avšak písmena byla vybrána tak aby poskytovala mnemotechnické asociace, pokud to bylo proveditelné (například **a** popisuje A-formu RNA). Těmito znaky jsou **a; b; c; d; e; f; g; h; i; j; L; m; n;**

Pro zvláštní případy byly pro oba heminukleotidy přiřazeny dva další znaky. Rodiny, pro které nejsou definovány všechny dihedrální úhly jsou označeny pomocí "_" (na koncích řetězce nebo mezery s neuspořádanou 3D strukturou). Neobvyklé konformace, které nebyly klasifikovány, jsou označovány pomocí "!", ty jsou nesprávné nebo zvláště zajímavé.

Tím pádem každý shluk rodin má dvouznakové jméno, jako například **1a** pro A-formu RNA. Poněvadž heminukleotidy αβγδ a δεζ sdílejí hodnotu δ, tím pádem musí mít navazující rodina kompatibilní konformaci ribózy. [20]

3 PROTEINOVÁ DATOVÁ BANKA

Proteinová datová banka (PDB) je celosvětová databáze obsahující experimentálně stanovené třídimenzionální struktury proteinů, nukleových kyselin a dalších biologických makromolekul. Tyto archívy obsahují souřadnice atomů, informace o primárních a sekundárních struktur, strukturní experimentální krystalografická data a také odkazy do řady dalších vědeckých databází. Tyto data jsou do PDB vkládány z měření rentgenovými strukturními analýzami, nukleární magnetickou rezonancí a také z elektronové mikroskopie. [21, 22, 23]

3.1 Historie PDB

Díky rozvoji rentgenové strukturní analýzy, se na přelomu šedesátých a sedmdesátých let skupina vědců zabývajících se krystalografií, v čele s Walterem Hamiltonem a Edgarem Mayerem v Brookhavenské národní laboratoři rozhodla vytvořit software, určen k uchování dat o strukturách proteinů a jejich atomových souřadnic. Tento byl vytvořen tak aby data v něm uložená byla jednoduše pro geometrické a grafické analýzy. V roce 1971 byla PDB představena v magazínu Nature jako spojené úsilí Cambridgeského krystalografického centra a Brookhavenské národní laboratoře.

Již v roce 1976 byla PDB využívána více než třiceti laboratoři po celém světě. Její popularita a využití se obzvláště rozrostlo po rozšíření internetu a dnes je nezbytnou součástí výzkumu vědců zabývající se strukturní biologií, bioinformatikou, biochemií, biofyzikou a mnoha dalšími obory zkoumajícími proteiny, nukleové kyseliny a další biologické makromolekuly. [24]

4 METODY EXPERIMENTÁLNÍHO MĚŘENÍ DAT

4.1 Rentgenová analýza

Jedna z metod strukturní analýzy využívá energeticky velmi bohatého (od 100 eV do 10 MeV) rentgenové záření. Toto záření je součástí elektromagnetického spektra a to v rozsahu vlnových délek od 10⁻³ nm do 10¹ nm. Analýza pomocí tohoto záření se zakládá na absorpci a emisi rentgenového záření, ale může také zahrnovat i jiné vzájemné působení mezi pozorovaným vzorkem a zářením. Základním principem této metody je přechod valenčních elektronů. Vnitřní elektron je vyražen pomocí zdrojového záření nebo pomocí jiného, rychle letícího elektronu a takto vzniklou díru zaplňuje elektron z vyšší vrstvy, přičemž se uvolní energie v podobě rentgenového záření. V závislosti na protonovém čísle můžeme pozorovat energetické rozdíly mezi stejnými hladinami rozdílných atomů. Energie rentgenového záření je tedy funkcí protonového čísla. [25, 26]

4.2 Nukleární magnetická rezonance

Metoda nukleární magnetické rezonance (NMR) je založena na principu měření změny magnetického momentu (spinu) jádra atomu. Při této metodě se využívá záření o vlnové délce od 0,5 až 100 m neboli radiové frekvence od 4 až 900 MHz. Tato metoda je použitelná pouze na atomy s nenulovým magnetickým momentem. To jsou takové atomy, které mají alespoň nukleonové či protonové číslo liché. Nejčastěji se tedy k měření využívají vodík ¹H, uhlík ¹³C fluor ¹⁹F nebo fosfor ³¹P. Působením vnějšího magnetického pole dochází k rozštěpení energetické hladiny jádra atomu a přechodům mezi vzniklými energetickými stavy. Na magnetické moment jádra má také vliv okolí atomu, díky tomu můžeme studovat okolí sledovaného atomu a tím pádem je tedy tato metoda velice vhodná pro pozorování strukturní stavby molekul. [26, 27, 28]

II PRAKTICKÁ ČÁST

5 POUŽITÁ DATA A METODOLOGIE ZPRACOVÁNÍ DAT

Jako zdroj dat byla použita Non-Redundantní RNA databáze tvořené RNA strukturami získanými rentgenovou krystalografií s atomovým rozlišením. Databáze je volně dostupná na http://rna.bgsu.edu/rna3dhub/nrlist/. Struktury v této databázi jsou nukleotidy, oligonukleotidy a také biologicky funkční molekuly RNA. Databáze obsahuje kompletní informace o délkách vazeb, úhlech a stereochemii RNA nukleotidů, jakož i o jejich preferovaných interakcích. Zahrnuje všechny typy párování a skládání bází a interakce mezi bázemi a páteří. Klíčovým parametrem pro tyto data je jejich neredundantnost (unikátnost) a to proto, aby se zabránilo neopakování použitých dat. [29]

Pro analýzu bylo použito celkem 2999 struktur s celkovým počtem 2057323 konformací. Struktury z databáze byly zpracovány pomocí skriptu vytvořeným Doc. Banášem. Tento skript u každé struktury anotoval jednotlivé rodiny a vypsal hodnoty všech dihedrálních úhlů. Následně byly data roztřízeny po jednotlivých rodinách a dále zpracovávány pomocí programovacího jazyka AWK. Grafy histogramů byly vytvořeny pomocí programu Gnuplot. Navíc modely jednotlivých rodin získány z PDB databáze byly vizualizovány pomocí programu PyMOL. Reprezentativní modely byly vybrány z rodin anotovaných výše zmíněným programem.

6 VÝSLEDKY

Zkoumaná data byla rozdělena podle příslušných rodin. V pozorovaném data setu nebyly nalezeny rodiny **&a** a **#a**. Tyto rodiny jsou velmi neobvyklé, a to z důvodu toho, že **&** a **#** reprezentují heminukleotidy se specificky neobvyklými hodnotami úhlu ε. Asi 22,47 % dat nemohlo být zařazeno do rodin, neboť při rozřazování jejich dihedrální úhly nepadly do žádné anotované rodiny. V této práci jsem se zabýval pouze rodinami, které byli zastoupeny alespoň deseti tisíci konformacemi. Toto kritérium splnilo deset rodin a těmi jsou **1a**, **1c**, **1b**, **1L**, **1g**, **2a**, **1m**, **7a**, **1**[a **1f**. V tabulce 1 je zobrazen počet studovaných konformací u každé rodiny a jejich procentuální zastoupení z celku 2057323 konformací.

Rodina **1a** se nejčastěji vyskytuje jako A-forma RNA. Rodiny **1c** a **1g** jsou nejběžněji nalézány jako součást GNRA. GNRA smyčku můžeme popsat pomocí rodin následovně: N**1aG1gN1aR1aA1cN1a**. **1c** pak také funguje jako opačná rodina k **1a** (tzv. "crankshaft" verze 1a rodiny). Rodina **1b** vede do 2' rodin, naopak **2a** z 2' rodin vychází. Role rodiny **1L** je jako tzv. "+ β shoulder" k rodině **1a** a na druhou stranu tzv. "- β shoulder" k **1a** tvoří **1m**. Rodina **7a** otáčí skládání řetězce. Nejlepší interkalační konformací je rodina **1**[. Rodina **1f** plní funkci "+ β shoulder" k rodině **1c** anebo otáčení skládání řetězce. Rodiny 5z, 4s a #a pomáhají formovat S tvar v S-motivu. Tyto rodiny nebyly pro malý výskyt dále do hodnocení zahrnuty.[20]

U vybraných rodin byly vypočítány průměrné hodnoty a směrodatné odchylky jednotlivých úhlů. Ty jsou uvedeny v tabulce 2. V tabulce jsou jednotlivé buňky barevně odlišeny v závislosti na hodnotě úhlu dle přiložené škály.

Tabulka 1 Počet studovaných konformací každé rodiny a jejich procentuální zastoupení.

Rodina	Počet	%
1a	1152691	56.03
1c	110898	5.39
1b	36377	1.77
1L	34608	1.68
1g	25841	1.26
2a	22383	1.09
1m	17635	0.86
7a	11971	0.58
1[10715	0.52
1f	10233	0.50

Tabulka 2 Hodnoty jednotlivých úhlů u daných rodin se směrodatnou odchylkou. Buňky jsou obarveny podle rostoucí hodnoty úhlu od červené, přes žlutou po zelenou (viz obr. 4)

	1a	1c	1b	1L	1g	2a	1m	7a	1[1f
δ-1 (°)	82.2±4.7	82.1±4.8	83.8±4.9	83.1±5.4	81.2±4.8	144.7±7.4	84.2±5.2	83.8±4.9	83.1±4.5	82.4±4.7
ε (°)	213.5±12.9	198.2±13.2	217.5±16.1	249.3±11.0	218.4±16.4	259.8±17.6	212.5±21.7	223.9±22.2	219.1±17.5	204.8±17.6
ζ (°)	288.9±10.8	285.8±14.9	285.7±13.9	261.7±14.2	289.9±15.6	288.6±19.9	299.2±14.5	222.2±14.6	289.7±15.4	306.5±18.3
α (°)	294.1±11.7	152.8±14.1	301.3±13.1	307.5±14.2	168.8±16.2	286.5±15.6	276.1±17.2	302.2±16.0	292.2±14.9	184.7±18.3
β (°)	173.1±11.4	202.5±17.9	176.9±14.7	135.8±21.1	158.1±21.9	189.4±18.4	216.5±11.0	152.1±18.6	222.4±14.9	143.9±19.4
γ (°)	55.0±9.2	175.4 <mark>±12.1</mark>	59.9±10.9	60.9±13.3	53.3±8.7	54.0±9.7	54.2±11.6	52.4±9.5	51.7±10.6	167.5 ± 12.5
δ (°)	81.7±4.4	84.3±4.6	143.5±7.1	81.3±5.3	84.7±4.7	84.1±5.0	85.3±5.7	83.1±5.0	143.7±6.9	84.6±5.1



Pro představu, jak tyto rodiny vypadají byly vybrány vzorové příklady. Jednotlivé atomy jsou označeny pomocí barev, a to zeleně uhlík, červeně kyslík, oranžovou fosfor a modrou dusík. V tabulce 3 jsou zapsány PDB kódy struktur ze kterých byly tito reprezentativní zástupci vybráni a také je zde číslo centrálního fosforu neboli fosforu před druhým cukrem.



Obrázek 4 Vzorové příklady rodin 1a, 1c, 1b, 1L, 1g a 2a



Obrázek 5 Vzorové příklady rodin 1m, 7a, 1[a 1f

Tabulka 3 Označení struktur v PDB databázi, ve které se nacházejí vzorové příklady a pozice centrálního fosforu jednotlivých rodin

Rodina	PDB	Fosfor
1a	1L2X	11
1c	1L2X	a28
1b	1SDS	d208
1L	1 S 72	1460
1g	1 S 72	1864
2a	1 S 72	1711
1m	1 S 72	1940
7a	1LNT	a6
1[1C0A	b658
1f	1EHZ	22

Z hodnot jednotlivých úhlů byly vytvořeny histogramy reprezentující rozložení dat (Obrázky 6-19). Následující grafy histogramů pomáhají lépe rozpoznat rozdíly mezi jednotlivými rodinami. Jen bych poznamenal, že pro rozdílné množství dat u jednotlivých rodin, nebyly histogramy škálovány na stejný

počet dat, aby histogramy byly dobře čitelné i u rodin s malou populací. U rodiny **1a** je škála na ose y neboli na ose udávají počet dat, od 0 do 140000. Rodina **1c** je škálována od 0 do 14000, následně **1b** od 0 do 4000, **1L** od 0 do 3500, **1g** od 0 do 3000, **2a** od 0 do 2500, **1m** od 0 do 1800, **7a** od 0 do 1400, **1**[od 0 do 1400 a finálně **1f** od 0 do 1200.





Obrázek 6 Histogramy úhlu δ -1 u rodin 1a, 1c, 1b, 1L a 1g



Obrázek 7 Histogramy úhlu δ-1 u rodin 2a, 1m, 7a, 1[a 1f





Obrázek 8 Histogramy úhlu ε u rodin 1a, 1c, 1b, 1L a 1g



Obrázek 9 Histogramy úhlu ε u rodin 2a, 1m, 7a, 1[a 1f





Obrázek 10 Histogramy úhlu ζ u rodin 1a, 1c, 1b, 1L a 1g



Obrázek 11 Histogramy úhlu ζ u rodin 2a, 1m, 7a, 1[a 1f





Obrázek 12 Histogramy úhlu α u rodin 1a, 1c, 1b, 1L a 1g



Obrázek 13 Histogramy úhlu α u rodin 2a, 1m, 7a, 1[a 1f





Obrázek 14 Histogramy úhlu β u rodin 1a, 1c, 1b, 1L a 1g



Obrázek 15 Histogramy úhlu β u rodin 2a, 1m, 7a, 1[a 1f



Obrázek 16 Histogramy úhlu y u rodin 1a, 1c, 1b, 1L a 1g



Obrázek 17 Histogramy úhlu y u rodin 2a, 1m, 7a, 1[a 1f



Obrázek 18 Histogramy úhlu δ u rodin 1a, 1c, 1b, 1L a 1g



Obrázek 19 Histogramy úhlu δ u rodin 2a, 1m, 7a, 1[a 1f

Jak je zřetelné z histogramů pro úhel δ -1 (obrázky 6 a 7) je hodnoty tohoto úhlu mají malý rozptyl. Je to úhel cukru a tak je pochopitelné, že se jeho hodnoty nebudou nikterak moc lišit. Jedinná rodina jejiž hodnoty úhlu δ -1 se odlišují od zbytku pozorovaných rodin je **2a**. Ta se posouvá z hodnot okolo 80°až 85°do hodnoty 144°. Jak je již zmíněno výše jedná se o rodinu, která vychází z 2' rodin. To má zjevně za následek posun hodnot tohoto úhlu.

Zde u úhlu ε (obrázky 8 a 9) je dobře pozorovatelné, že histogramy jsou více rozprostřené a pohybující se v hodnotách od 200° do 220°. Lehce se vychylují pouze rodiny **1L** a **2a**. U **2a** dochází k posunu na hodnoty kolem 260° a je tomu tak pravděpodobně z důvodu hodnoty předchozího úhlu δ -1. U úhlu **1L** dochází k posunu k 250° a to může být způsobeno jeho funkcí jako "+ β shoulder" k rodině **1a** a tím pádem tato předchozí rodina.

Data úhlu ζ se svým rozprostřením příliš neliší od úhlu ε . Tento úhel je však orientován v okolo 290°. Odchylují se pouze dvě rodiny, a to **1L** 260° a **7a** do 220°. Jelikož **7a** otáčí skládání řetězce není divu, že se razantně liší od ostatních rodin. U rodiny **1L** to může být spojeno s její funkcí, ale jak je možno vidět z těchto pozorování lze usoudit, že i přes označení **1** se první heminukleotid této rodiny značně liší od ostatních. Toto je zajímavý poznatek a ukazuje nám, že tato rodina by byla vhodným tématem k dalšímu pozorování.

Druhý heminukleotid začíná úhlem α . U tohoto úhlu můžeme vidět, že data jsou souměrně rozděleny. U sedmu rodin se střed histogramů pohybuje v okolí 300°, liší se pouze rodiny **1c** kde je střed v 152°, **1g** se středem v 169° a **1f** mající střed v 185°. **1c** a **1g** jsou součástí GNRA a jejich tvar je tedy pravděpodobně takový, aby vyhovoval stabilitě řetězce. **1f** může sloužit i k otáčení řetězce, což by rozdílná hodnota úhlu α podporovala.

Úhel β je, jak je možno pozorovat z grafů histogramů, nejvíce diverzifikovaným úhlem. Pro každou rodinu se tento úhel liší nejen tvarem píku histogramu, ale také jeho polohou.

Úhel γ je má zcela opačné chování oproti úhlu β . Středy píků jsou u všech však dvou rodin orientovány na téměř stejném místě a to v 50° až 55° a všechny píky mají identické tvary (v případě **1f** je tvar jiný ale může být způsobeno malým

počtem pozorovaných dat). Polohou středu se liší pouze **1c**, která má pík v 175° a **1f** s píkem v 167°.

Posledním úhlem druhého nukleotidu je úhel δ . Opět jako úhel δ -1 je velmi stabilní, s daty koncentrovanými ke středu, který se u většiny rodin nachází v oblasti 80° až 85°. Odlišují se pouze rodiny **1b** a **1**[. **1b** vede do 2' rodin a tím pádem se tato odlišující hodnota v úhlu 143° vysvětluje. U rodiny **1**[je tato odchylka do 143° možná spjata se schopností interkalace.



Rozložení úhlu ε proti úhlu ζ

Obrázek 20 2D histogramy rozložení úhlu ε proti úhlu ζ u rodin 1a, 1c, 1b, 1L, 1g a 2a



Obrázek 21 2D histogramy rozložení úhlu ε proti úhlu ζ u rodin 1m, 7a, 1[a 1f



Rozložení úhlu α proti úhlu γ

Obrázek 22 2D histogramy rozložení úhlu α proti úhlu γ u rodin 1a, 1c, 1b, 1L, 1g a *2a*



Obrázek 23 2D histogramy rozložení úhlu α proti úhlu γ u rodin 1m, 7a, 1[a 1f

V další části jsem se zaměřil více na **1a** a **1c** rodinu. Tyto rodiny jsou potenciálně zajímavé v rámci molekulárně dynamických simulací. Jak již bylo zmíněno výše, **1a** rodina se vyskytuje všude tam kde je A-forma RNA a **1c** předchází **1a** při přechodu struktury se zalomením do A-formy. To, že **1a** je charakteristická pro A-formu plyne z toho, že tato rodina je nejčastěji se vyskytující, a to s počtem 1152691 konformací z celkového počtu 2057323, což reprezentuje 56,03 % všech námi zkoumaných konformací. Tato rodina je velmi hojně zastoupená, což je poznat i na menší rozdrobenosti dat a nízkém počtu odlehlých hodnot. Rodina **1c** je druhou nejpočetnější rodinou s počtem 110898 konformací což odpovídá 5,39 % z celkového počtu dat.

Obě rodiny mají shodný první heminukleotid **1** (δ -1, ε , ζ) ale i tak se můžou v těchto úhlech nepatrně lišit. Není to pravděpodobné v úhlu δ -1 ale můžeme to čekat u úhlů ε a ζ . Je to způsobeno tím, že na tyto úhly mají vliv i atomy z druhého

heminukleotidu rodiny. Jejich shodu potvrzují i zkoumaná data. Tato skutečnost je velmi dobře pozorovatelná na grafech 1D histogramů (Obrázek 6, 8, 10) jednotlivých úhlů tak i na grafech 2D histogramů (obrázek 20), kde je viditelné, že průniky úhlů ε a ζ jsou ve stejných hodnotách.

U druhého nukleotidu (úhly α , β , γ a δ) popsán písmen, a to zde buď to **a** nebo c, je viditelný značný rozdíl. Grafy histogramů jednotlivých úhlů mají jednak jiný tvar ale také jinde orientovaný střed (Obrázek 12, 14, 16, 18). Toto lze také pozorovat u grafů 2D histogramů popisující úhly α a γ (Obrázek 22). Jediná shoda může být nalezena u úhlů δ avšak tento úhel je velmi stabilní dáno nejen našim pozorováním ale také tím, že je to úhel cukru. Obecně však lze říci, že hodnoty úhlů těchto dvou rodin jsou více méně shodné, až na úhly alfa a gama, které vykazují tzv. α/γ flip. Tento flip může být potenciálně problematický v molekulárně dynamických simulacích [30]. U některých struktur bylo pozorováno problematické chování. Jedním z velkých problémů je popis skládání UNCG smyček [31]. U UUCG smyčky je poslední guanin v NMR struktuře anotován jako 1a rodina. V dynamice se velmi často stává, že přechází do 1c rodiny (tj. dojde k α/γ flipu). Zdá se, že tento flip ochrání smyčku před rozpadem, neboť pokud by se tomuto flipu zabránilo (držením páteře v hodnotách typických pro 1a rodinu) dojde velmi rychle k spontánnímu rozpadu smyčky. Ze simulací UUCG vyplývá, že je ve struktuře smyčky existuje nějaké vnitřní pnutí, které je v rámci silového pole ff99bsc0 χ_{OL3} [32, 33, 34] opraveno α/γ flipem.[30] Toto má dalekosáhlé důsledky nejenom v rámci klasické molekulárně dynamické simulace (MD), ale vede k problematickému skládání UUCG smyčky v pokročilých MD metodách. Neboť při skládání UUCG smyčky se vychází z rozbaleného stavu a silové pole bude preferovat 1a rodinu a tím pádem, bude mít guanin problém se správně natočit pro sbalení UUCG smyčky. Bohužel toto není jediný případ kde silové pole pro RNA selhává. Např. u simulace pseudoknotu (PDB kód 1L2X) dochází k přechodu rodiny 1c na 1a a už nedojde k navrácení zpět. [30] Z výše uvedeného vyplývá, že sledování a popis rodin je velmi důležité a čím dál více se jeví, že současné silové pole pro RNA má problém správně popsat struktury, kde se vyskytují 1c rodiny.

7 ZÁVĚR

Cílem této práce bylo analyzovat konformace RNA páteře z PDB databáze. Tyto data byla roztřízena do rodin a rodiny následně analyzovány. Pro detailnější analýzu byly vybrány jen ty rodiny, jež byly dostatečně reprezentovány tak aby nevznikala příliš velká statistická chyba což bylo deset rodin. Nejprve byly rodiny pozorovány jako celek a následně pak jednotlivé úhly páteře zvlášť.

Prvním krokem byl výpočet průměrných hodnot jednotlivých úhlů u příslušných rodin a také jejich směrodatných odchylek. Následně byly vytvořeny histogramy všech úhlů a z nich zpracovány grafy tak aby data byla lépe reprezentována. Nebyly vytvořeny pouze jednoduché histogramy ale také 2D histogramy, které hlouběji ukazují na rozdíly mezi rodinami.

Více pozornosti bylo věnováno rodinám **1a** a **1c**. Tyto rodiny byly nejen nejvíce populované ale jsou zároveň velmi zajímavé v kontextu molekulárně dynamických simulací. **1a** je rodina tvořící A-formu RNA a **1c** je rodinou předcházející přechodu na A-formu. U těchto rodin dochází k tzv. α/γ flipu. Tento jev je jasně potvrzen pozorovanými daty.

V chování všech rodin vzniká velké množství výjimek a různých nepravidelností, které je stále obtížné objasnit. Tato oblast poznání je stále velmi málo probádána a bude tedy potřeba dalších výzkumů pro její úplné pochopení.

8 SUMMARY

Aim of this thesis was to analyze conformations of RNA backbone from PDB database. These data were divided into suits and these suits were then analyzed. For more detail analysis, I have chosen ten the most frequent suits. At first, I divided all suits into families and this was following calculation of all backbone angles for each suit.

First step was calculation of average values of individual angles in respective suits as well as their standard deviation. Subsequently histograms of all angles were created and from them graphs were compiled for better representation of data. Not only simple histograms were created but also 2D histograms, which deeper shows differences between suits.

More attention was given to suits **1a** and **1c**. These suits were not only mostly populated, but they are also very interesting in context of molecular dynamic simulations. **1a** is a suit creating A-form of RNA and **1c** is a suite previous to transition to A-form. These suits differ only in α/γ flip. This phenomenon is obviously conformed by observed data.

In all of suits there is a huge amount of exceptions and various irregularities, which are still difficult to clarify. This field of knowledge is still very little explored and for total understanding of and therefore further researches will be needed.

SEZNAM POUŽITÉ LITERATURY

1. Ferenčík M.; Škárka B.; Novák M.; Turecký L., Biochémia, Slovak academic press s.r.o., Bratislava, 2000.

2. Kodíček M.; Valentová O.; Hynek R., Biochemie: chemický pohled na biologický svět, Vysoká škola chemicko-technologická v Praze, Praha, 2015.

3. Berg J. M.; Tymoczko J. L.; Styer L., Biochemistry 5th edition, W. H. Freeman and Company, New York, 2002.

4. Vodrážka Z., Biochemie, Akademia, Praha, 2002. ISBN 80-200-0438-1

5. Gesteland, R.F.; Cech T.; Atkins J.F., The RNA world: the nature of modern RNA suggests a prebiotic RNA 3rd edition, Cold Spring Harbor Laboratory Press, New York, 2006.

6. Tinoco J.I.; Bustamante C., How RNA folds, J.Mol. Biol. 1999, 293, 271-281.

7. Kim S. H.; Berman H. M.; Seeman N. C.; Newton M. D., Seven basic conformations of nucleic acid structural unit. Acta cryst. **1973**, 29, 703-710.

8. Humphris-Narayanan E.; Pyle A. M., Discrete RNA Libraries from Pseudo-Tosrion Space, Journal of Molecular Biology **2012**, 421, 6-26.

9. Schneider B.; Morávek Z.; Beman H.M., RNA conformational classes, Nucleic Acids Res. **2004**, 32, 1666-1677.

 Dock-Bregeon, A. C.; Chevrier, B.; Podjarny, A.; Johnson, J.; de Bear, J. S.;
 Gough, G. R.; Gilham, P. T.; Moras, D., Crystallographic Structure of an RNA Helix: [U(UA)6A]2. J. Mol. Biol. **1989**, 209, 459-474.

11. Stephen Neidl, Principles of Nucleic Acid Structure, Oxford University Press, Oxford, 2008.

12. Blackburn G. M.; Gait M. J.; Loakes D.; Wiliams D.M.; Nucleic acids in chemistry and biology, Oxford University Press, New York, 1996.

Ban N.; Nissen P.; Hansen J.; Moore P.B.; Steitz T.A., The complete atomic structure of the large ribosomal subuint at 2.4 Å resolution, Science 2000, 289, 905–920.

14. Schuenzen F.; Tocilj A.; Zarivach R.; Harms J.; Gluehmann M.; Janell D.; Bashan A.; Bartels H.; Agmon I.; Franceschi F.; et al., Structuce of functionally activated small ribosomal subunit at 3.3 Å resolution, Cell **2000**, 102, 615-623.

15. Wimberley B.T.; Brodenson D.E.; Clemons W.M.; Morgan-Warren R.J.; Carter A.P.; Vonrhein C.; Hartsch T.; Richardson D.C., Structure of 30S ribosomal subuint, Nature **2000**, 407, 327-339.

16. Hershkovitz E.; Tannenbaum E.; Howerton S.B.; Sheth A.; Tannenbaum A.; and Williams L.D., Automated identification of RNA conformational motifs: Theory and aplication to Hm LSU 23S rRNA, Nucleic Acids **2003**, 31, 6249-6257.

17. Hisoao C.; Mohan S.; Hershkovitz E.; Tannenbaum A.; and Williams L.D., Single nucleotide choreography, Nucleic Acids **2006**, 34, 1481-1491.

18. Murray L.W.; Arendall III W.B.; Richardson D.C.; and Richardson J.S., RNA backbone is rotameric, Proc. Natl. Acad. Sci. **2003**, 100, 13904-13909.

19. Leontis N.B.; Altman R.B.; Berman H.M.; Bernner S.E.; Brown J.W.; Engelke D.R.; Harvey S.C.; Holbrook S.R.; Jossinet F.; Lewis S.E.; et al., The RNA Ontology Consortium: An open invitation to the RNA comunity, RNA **2006**, 12, 553-541.

20. Richardson J.S.; Schneidr B.; Murray L.W.; Kapral G.J.; Immormino R.M.; Headd J.J.; Richardson D.C.; Ham D.; Hershkovits E.; Williams L.D.; Keating K.S.; Pyle A.M.; Micallef D.; Westbrook J.; and Berman H.M., RNA backbone: Concensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution), RNA **2008**, 14, 465-481.

 Sussman J. L.; Lin D.; Jiang J.; Manning N. O.; Prilusky J.; Ritter O.; Abola E.
 E., Protein Data Bank (PDB): Database of Three-Dimensional Structural Information of Biological Macromolecules, Acta Cryst. **1998**, D54, 1078-1084. 22. Berman H. M.; Westbrook J.; Feng Z.; Gilliland G.; Bhat T. N.; Weissig H.; Shindyalov I. N.; Bourne P. E., The Protein Data Bank, Nucleic Acids Research **2000**, 28, 235-242.

23. Meyer E. F., The first years of the Protein Data Bank, Protein Sci. **1997**, 6(7), 1591-1597.

24. Berman H. M., The Protein Data Bank: a historical perpective, Acta Cryst. **2008**, A64, 88-95.

25. Němcová I.; Čermáková L.; Rychlovský P.; Spektrometrické analytické metody, Karolinum, Praha, 2004.

26. Klouda P., Moderní analytické metody, Pavel Klouda, Ostrava, 2003.

27. McMurry J., Organická chemie, VUTIUM, Brno, 2007.

28. Teng Q., Structual Biology: Practical NMR Aplication, Springer Science, New York, 2013.

29. Leontis N.; Westhof E., RNA 3D Structure Analysis and Prediction, Springer-Verlag Berlin Heidelberg, Berlin, 2012.

30. Zgarbova M.; Jurečka P.; Banáš P.; Havrila M.; Šponer J.; Otyepka M., Noncanonical α/γ Backbone Conformations in RNA and the Accuracy of Their Description by the AMBER Force Field, J. Phys. Chem. B **2017**, 121(11), 2420-2433.

31. Kührová P.; Mlýnský V.; Zgarbová M.; Krepl M.; Bussi G.; Best R.B.; Otyepka M.; Šponer J.; Banáš P., Improving the Performance of the Amber RNA Force Field by Tuning the Hydrogen-Bonding Interactions, J Chem Theory Comput **2019**, 10.1021/acs.jctc.8b00955.

32. Perez A.; Marchan I.; Svozil D.; Sponer J.; Cheatham T. E.; Laughton C. A.; Orozco M., Refinenement of the AMBER force field for nucleic acids: Improving the description of alpha/gamma conformers. Biophys *J* **2007**, *92*(11), 3817-3829.

33. Zgarbova M.; Otyepka M.; Sponer J.; Mladek A.; Banas P.; Cheatham T. E., 3rd; Jurecka P., Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. J Chem Theory Comput **2011**, *7*(9), 2886-2902.

34. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A second generation force field for the simulation of proteins, nucleic acids, and organic molecules (vol 117, pg 5179, 1995). J Amer Chem Soc **1996**, *118* (9), 2309-2309.