



**MATEMATICKO-FYZIKÁLNÍ
FAKULTA**
Univerzita Karlova

BAKALÁŘSKÁ PRÁCE

Michal Hercík

Webový plugin pro vizualizaci sady sekundárních struktur RNA

Katedra softwarového inženýrství

Vedoucí bakalářské práce: doc. RNDr. David Hoksza, Ph.D.

Studijní program: Informatika

Studijní obor: Programování a vývoj software

Praha 2023

Prohlašuji, že jsem tuto bakalářskou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů. Tato práce nebyla využita k získání jiného nebo stejného titulu.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V dne

Podpis autora

Rád bych poděkoval všem, díky kterým pro mě bylo možné tuto práci napsat. Speciálně potom svému vedoucímu práce doc. RNDr. Davidu Hokszi, Ph.D. za veškeré konzultace a odbornou pomoc.

Název práce: Webový plugin pro vizualizaci sady sekundárních struktur RNA

Autor: Michal Hercík

Katedra: Katedra softwarového inženýrství

Vedoucí bakalářské práce: doc. RNDr. David Hoksza, Ph.D., Katedra softwarového inženýrství

Abstrakt: Zkoumání RNA je důležité pro lepší pochopení evoluce nebo některých onemocnění. V této práci představujeme knihovnu napsanou v jazyce Typescript, která nabízí metody pro vizuální analýzu více sekundárních struktur RNA, nejlépe ale metody fungují při práci s dvěma nebo třemi strukturami. Pro reprezentaci používáme radial diagram, který je vygenerovaný na základě vzorové struktury. Metody použité v knihovně pro analýzu využívají způsobu generování radial diagramu a to tak, že diagramy vygenerované ze stejné vzorové struktury na sebe mapuje právě pomocí vzorové struktury, díky tomu lze sledovat rozdíly a podobnosti struktur.

Klíčová slova: bioinformatika RNA sekundární struktura web plugin

Title: Web plugin for multiple RNA secondary structure visualization

Author: Michal Hercík

Department: Department of Software Engineering

Supervisor: doc. RNDr. David Hoksza, Ph.D., Department of Software Engineering

Abstract: The study of RNA is important to better understand evolution or some diseases. In this thesis, we present a library written in Typescript that offers methods for visual analysis of multiple RNA secondary structures, but preferably work best with two or three structures. For the representation we use radial diagram, which is generated by template-based method. Methods used in the library for analysis use the template-based generation of the radial diagram by mapping diagrams generated from the same template to each other via the template, so that differences and similarities of structures can be observed.

Keywords: bioinformatics RNA secondary structure web plugin

Obsah

Úvod	3
1 Úvod do problematiky	4
1.1 RNA	4
1.1.1 Funkce	4
1.1.2 Sekundární struktura RNA	4
1.2 Vizualizace sekundárních RNA struktur	5
1.3 Podobné projekty	7
1.3.1 VARNA	7
1.3.2 RNAStructViz	8
1.3.3 Forna	9
1.3.4 R-chie	10
1.3.5 Shrnutí existujících nástrojů	11
1.4 Kreslení grafů na základě šablony	11
1.4.1 TRAVeLer	12
1.4.2 R2DT	13
2 Metody vizualizace a porovnání	15
2.1 Důsledek využívání vzorové struktury	16
2.2 Překládání struktur	16
2.2.1 Zarovnání	17
2.2.2 Průhlednost struktur	19
2.2.3 Rozmazání struktur	20
2.3 Obarvení struktur	20
2.4 Transformace na vzor	21
2.5 Mapovací čáry	22
2.6 Využití stromu	23
2.7 Zvolené metody	25
3 Dokumentace	26
3.1 Použití canvasu	26
3.1.1 Knihovna D3.js	26
3.1.2 SVG	26
3.1.3 Canvas	26
3.2 Struktura vstupních dat	27
3.3 Objektový návrh	28
3.4 Uživatelské seznámení s knihovnou	31
3.4.1 Vytvoření objektu RnaVis a vykreslení struktur	31
3.4.2 Zoom a posouvání canvasu	32
3.4.3 Zarovnání struktur	32
3.4.4 Mapovací čáry	34
3.4.5 Transformace na vzorovou strukturu	34
3.5 Demonstrace metod	35
3.5.1 Získání testovacích dat	36

Závěr	37
Seznam použité literatury	38
Seznam obrázků	40
A Přílohy	41
A.1 SQL dotazy pro získání testovacích dat	41

Úvod

RNA má více funkcí než pouze přenos genetické informace. Zkoumání struktur RNA napříč různými živočišnými druhy a sledování podobností, které mohou hrát důležitou roli v její funkci, může vést k lepšímu pochopení evoluce[16] nebo některých nemocí[9], speciálně se zkoumá například v souvislosti s rakovinou[1, 12] nebo neurologickými onemocněními[11, 13].

Funkce RNA je odvozená od její 3D struktury (terciární struktury) a ačkoliv její predikce je aktivně zkoumané téma[15, 8], je stále těžké jí získat. Mnohem více známé jsou sekundární struktury. Ty o prostorovém rozložení říkají pouze to, že spárované nukleotidy jsou blízko sebe, ale přesto nabízí dobrý popis.

Pro analýzu sekundárních struktur RNA existuje mnoho nástrojů. Nástroje se typicky soustředí na možnosti anotace, editace nebo porovnávání omezeného množství sekundárních struktur RNA a neexistuje nástroj, který by umožňoval interaktivně porovnávat mezi sebou libovolné množství sekundárních struktur RNA. Většina nástrojů navíc nenabízí možnost integrace do jiných programů, například do webových databází sekundárních struktur.

Z těchto důvodů představujeme knihovnu napsanou v jazyce Typescript, která umožňuje jednoduše vizualizovat sekundární struktury a nabízí metody pro porovnávání podobných struktur, které můžou hrát klíčovou roli k pochopení biologických mechanismů.

1. Úvod do problematiky

1.1 RNA

¹ RNA (Ribonukleová kyselina) je typ nukleové kyseliny, která je nezbytná pro různé biologické procesy. Skládá se z dlouhého řetězce nukleotidů obsahujících cukr, fosfátovou skupinu a dusíkatou bázi. Čtyři dusíkaté báze v RNA jsou adenin, guanin, cytosin a uracil. RNA je transkribována z DNA a může být nalezena všude v živých buňkách.

1.1.1 Funkce

RNA má několik důležitých funkcí v buňkách. Jednou z jejích hlavních funkcí je fungovat jako posel mezi DNA a ribozomy během syntézy proteinů. Messenger RNA (mRNA) nese genetickou informaci z DNA k ribozomům, které tuto informaci používají k vytváření proteinů. Další typy RNA, jako tRNA a rRNA, hrají také důležité role v syntéze proteinů a patří do skupiny zvané nekódující RNA (ncRNA).

ncRNA je druh RNA, která nekóduje proteiny, ale místo toho hraje regulační role v buňkách. ncRNA lze rozdělit do dvou širokých kategorií: malé ncRNA, které jsou menší než 200 nukleotidů, a dlouhé ncRNA, které jsou delší než 200 nukleotidů. Malé ncRNA, jako jsou mikroRNA a siRNA, se účastní posttranskripční regulace genů, zatímco dlouhé ncRNA se účastní různých buněčných procesů, včetně regulace exprese genů, remodelace chromatinu a inaktivace chromozomu X.

1.1.2 Sekundární struktura RNA

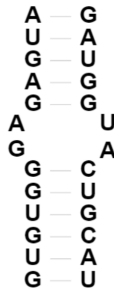
RNA molekuly se skládají do specifických tvarů díky komplementárnímu párování nukleotidů uvnitř molekuly. To vede k vytvoření sekundárních struktur, které jsou stabilizovány vodíkovými vazbami mezi bázemi. Sekundární struktury RNA hrají důležité role v různých buněčných procesech, jako je syntéza proteinů a regulace genů.

Motivy sekundární struktury RNA

Sekundární struktury RNA mohou být složeny z různých strukturálních motivů, jako jsou hairpin loop, bulge a multibranch loop.

Hairpin loop Jedním z nejčastějších typů sekundární struktury je hairpin loop¹, který se skládá z jednořetězcové oblasti RNA, která se ohýbá zpět na sebe a vytváří kmen (dvouřetězcovou oblast) a smyčku (jednořetězcovou oblast). Hairpin smyčky se často nacházejí na koncích molekul RNA a mohou hrát důležité role v stabilitě a zpracování RNA.

¹ Celá tato kapitola byla vygenerovaná pomocí ChatGPT dostupné v dubnu 2023 na adrese <https://chat.openai.com/> a upravena.



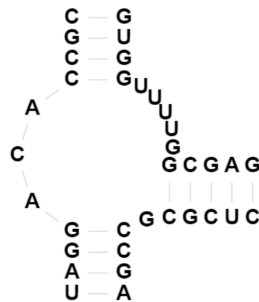
Obrázek 1.1: Ukázka motivu hairpin loop.

Bulge Dalším častým typem sekundární struktury je bulge^{1,2}, který zahrnuje jeden nukleotid, který není spárován s komplementárním nukleotidem v kmenové oblasti. Bulges mohou narušit sekundární strukturu RNA, ale také mohou být důležité v interakci mezi proteinem a RNA.



Obrázek 1.2: Ukázka motivu bulge.

Multibranch loop Multibranch loop^{1,3} je složitější sekundární struktura, která zahrnuje tvorbu více stonků a smyček. Tyto struktury mohou být důležité pro skládání a funkci RNA a často se nacházejí v molekulách RNA s katalytickou aktivitou, jako jsou ribozymy.



Obrázek 1.3: Ukázka motivu multibranch loop.

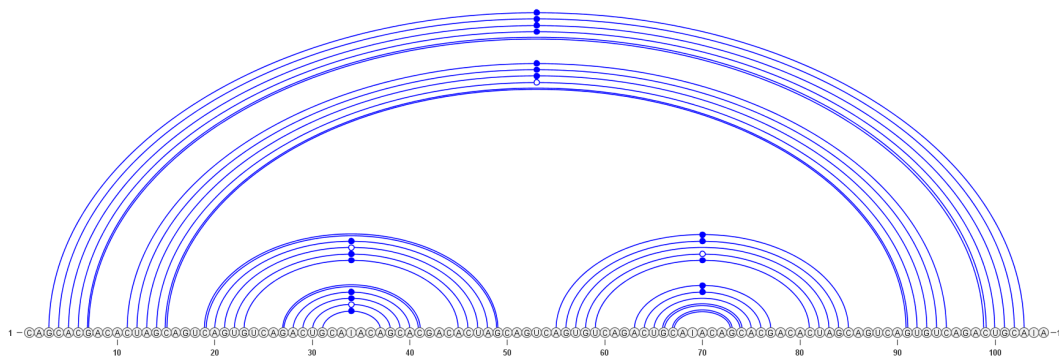
Celkově je pochopení sekundární struktury RNA zásadní pro pochopení její biologické funkce a toho, jak interaguje s jinými molekulami v buňce.

1.2 Vizualizace sekundárních RNA struktur

Pro reprezentaci sekundární RNA struktury se používají jak textové, tak grafické způsoby. Pro analýzu jsou vhodnější ty grafické, protože dokáží mnohem lépe odhalit stavbu struktury.

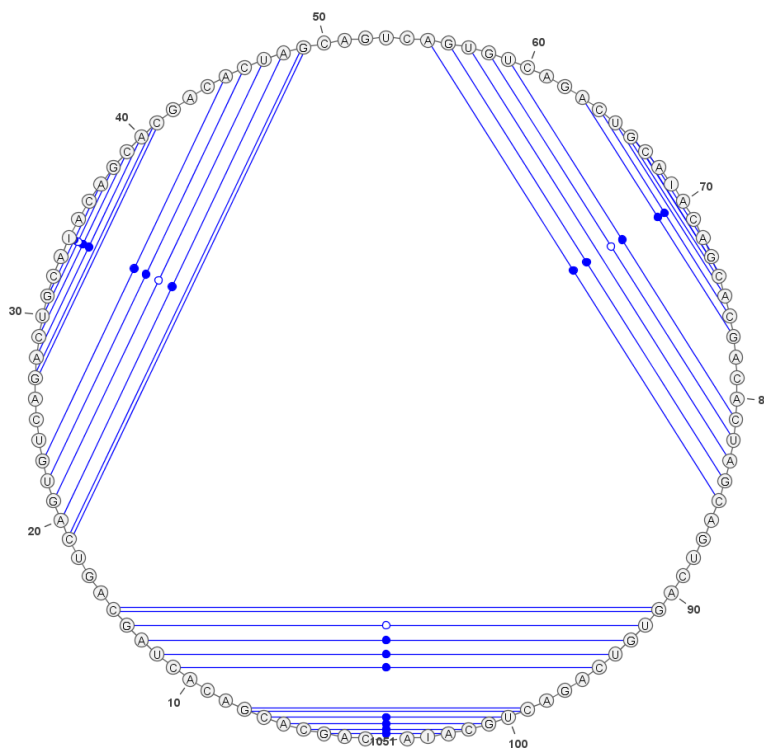
V této části představíme tři nejpoužívanější grafické reprezentace - arc diagram, circular diagram a radiate diagram. Obrázky ukázek diagramu v této části jsou získané za pomoci nástroje VARNA[3].

V arc diagramu 1.4 jsou nukleotidy zobrazeny na rovné čáře ve stejném pořadí jako v sekvenci a bázeové páry nukleotidů jsou spojeny obloukem.



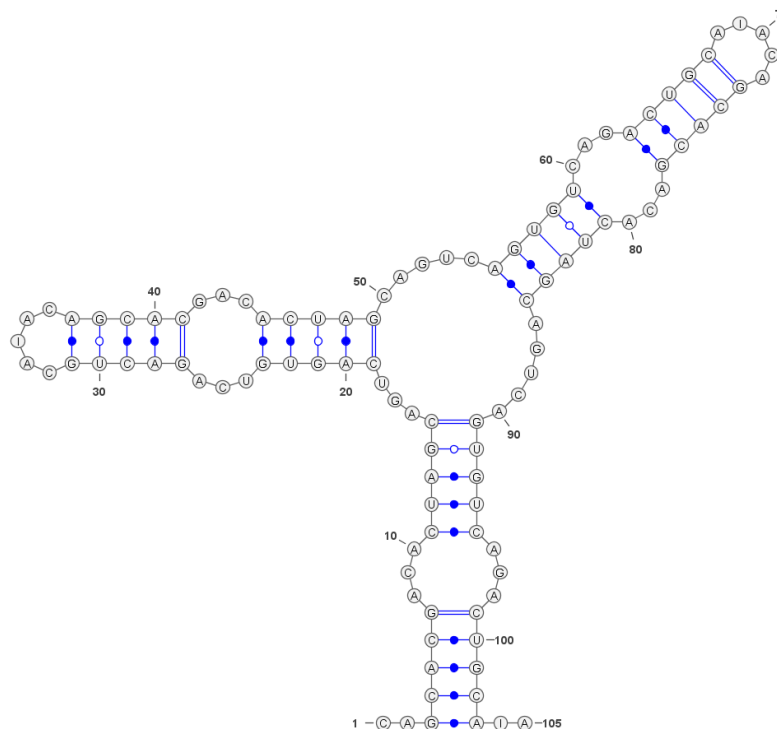
Obrázek 1.4: Ukázka arc diagramu.

Circular diagram 1.5 je velmi podobný. Nukleotidy neleží na rovné čáře, ale po obvodu kruhu. Bázeové páry jsou spojeny buď čarou nebo obloukem.



Obrázek 1.5: Ukázka circular diagramu.

V radiate diagramu 1.6 jsou pozice nukleotidů voleny tak, aby bylo možné rozeznat motivy sekundární struktury, jako jsou hairpins, bulges nebo multibranch loops.



Obrázek 1.6: Ukázka radiate diagramu.

Právě schopnost zachytit zmíněné motivy arc i circular diagramy postrádají, a proto se radiate diagram používá tam, kde je potřeba detailní vizuální analýza motivů sekundární RNA struktury a její interakce.

1.3 Podobné projekty

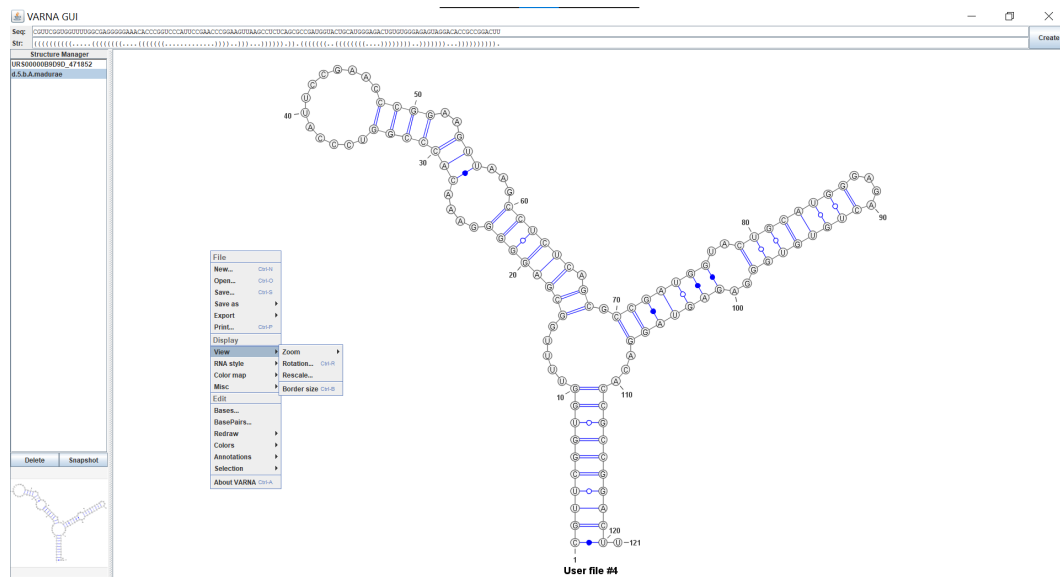
Rádi bychom čtenáře seznámili s některými nástroji, které jsou používány pro vizualizaci sekundárních RNA struktur. Většina z nich jsou programy s uživatelským rozhraním a mohlo by se proto zdát zbytečné je zmiňovat nebo porovnávat s naší knihovnou. Nicméně u níže zmíněných programů není důležité řešení samotného uživatelského rozhraní, jako především druh zvolených metod pro vizualizaci a následné porovnávání.

Z velkého množství existujících nástrojů byla snaha vybrat takové, které mají rozdílné přístupy a nabízí nejširší paletu funkcí.

1.3.1 VARNA

VARNA1.7 (Visualization Applet for RNA) je nástroj pro automatické kreslení, vizualizaci a anotaci sekundárních RNA struktur, navržený jako doprovodný software pro webové servery a databáze.

VARNA implementuje algoritmy pro vykreslení všech tří výše zmíněných diagramů, podporuje různé textové formáty pro vstup i výstup a je schopen exportovat kresbu do rastrových nebo vektorových formátů. Umožňuje ruční úpravy a strukturální anotace výsledku kresby a je považován za standard pro práci se sekundárními strukturami RNA.

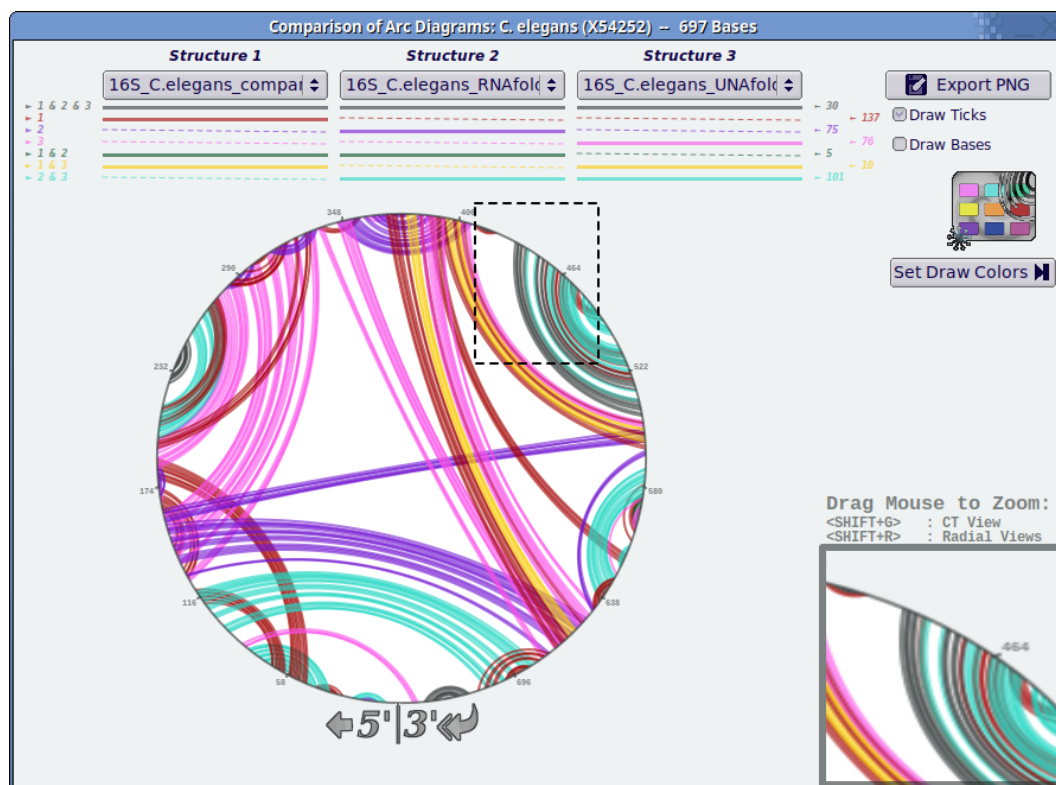


Obrázek 1.7: Snímek nástroje Varna.

1.3.2 RNAstructViz

RNAstructViz[2]1.8 je grafický nástroj pro analýzu sekundárních RNA struktur. Jeho předností je vizuální porovnání tří konfigurací v circular arc diagramu. Doplněné zabudovaným prohlížečem CT-style² souboru a prohlížečem radial diagramu podstruktury, která je přímo propojená s arc diagram oknem skrze nástroj pro výběr zoom. Mezi další funkce patří vypočítání číselných informací a možnost exportu obrázků a dat pro pozdější použití.

² CT formát souboru slouží k ukládání informace o sekvenci a bázevých pářů.



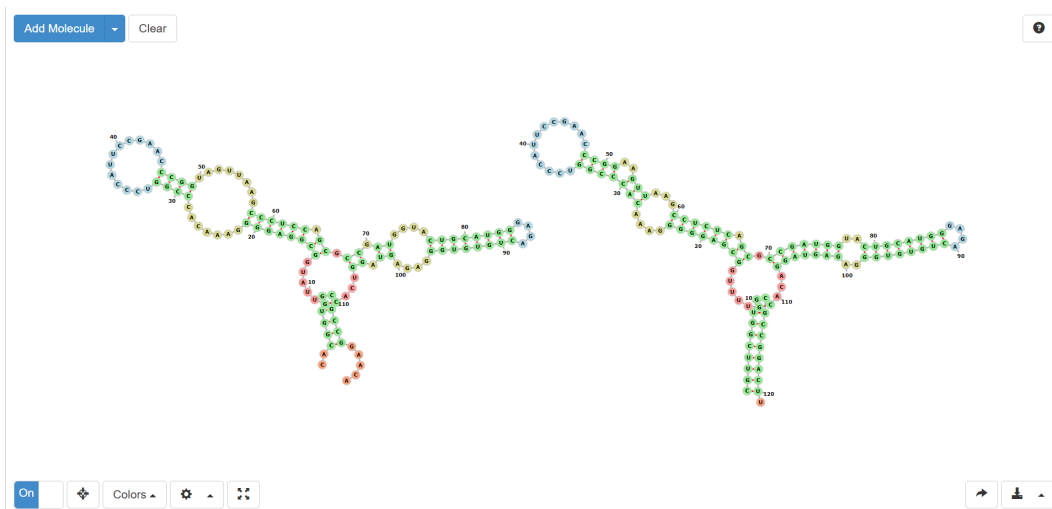
Obrázek 1.8: Snímek nástroje rnaStructViz, zobrazující tři sekundární struktury RNA.⁴

1.3.3 Forna

Forna[6]1.9 (force-directed rna) nabízí webové rozhraní a server, který umožňuje uživateli vložit sekundární RNA strukturu ve formátu dot-bracket a zobrazí ji jako force-directed graf⁵. Uživatel může následně upravit pozice přetažením myši a lze i upravovat přímo strukturu.

⁴ <https://github.com/gtDMMB/RNAStructViz/wiki/ArcDiagrams>

⁵ <https://cs.brown.edu/people/rtamassi/gdhandbook/chapters/force-directed.pdf>



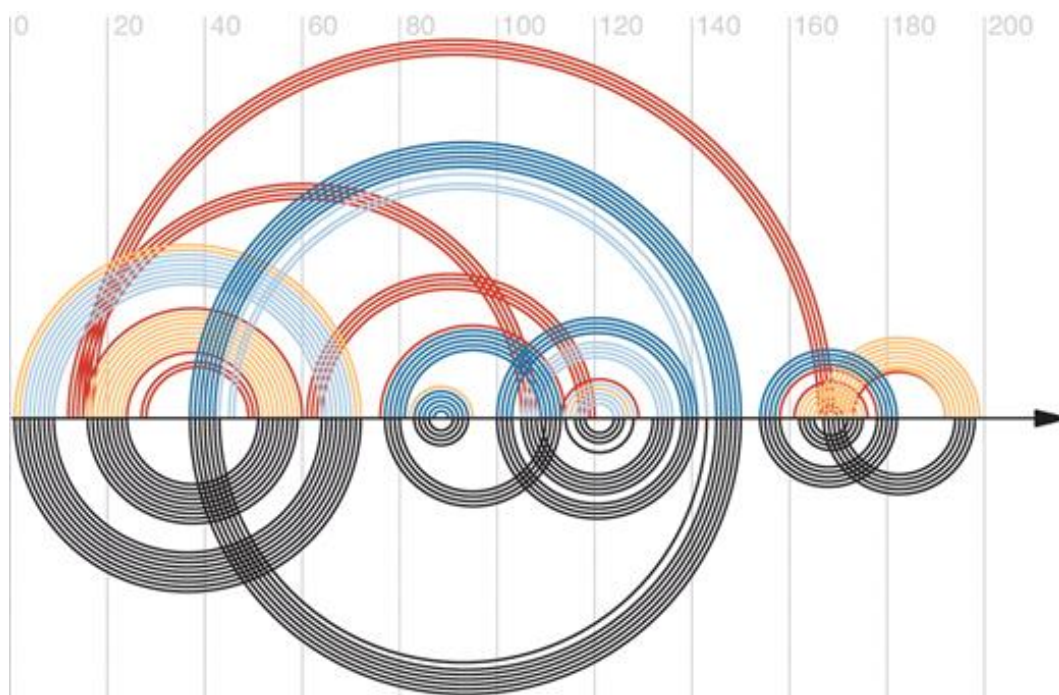
Obrázek 1.9: Snímek nástroje Forna se dvěma sekundárními strukturami.

1.3.4 R-chie

R-chie [7]1.10 je web server, který umí vygenerovat šest různých typů arc diagramu. Vývoj tohoto nástroje byl se zaměřením především na složitější struktury, které nelze hezky nakreslit v radial diagramu. R-chie umí vygenerovat diagram pro porovnávání dvou sekundárních struktur RNA. Důležitým cílem byla možnost generovat diagramy pro velké množství dat, proto také nenabízí grafické rozhraní a s ním spojenou interakci se strukturami.

Projekt také nabízí balíček napsaný v jazyce R⁶ zvaný R4RNA, který umožňuje spuštění programu lokálně a napříč operačním systémem.

⁶ <https://www.r-project.org/>



Obrázek 1.10: Výsledný arc diagram nástroje R-chie, zobrazující dvě struktury. První struktura je nad horizontální čarou a druhá pod ní.⁸

1.3.5 Shrnutí existujících nástrojů

Nástroje představené v této kapitole se soustředí především na práci s circular diagramem nebo arc diagramem, a právě pouze pro tyto diagramy nabízí nějaké metody pro porovnávání omezeného množství sekundárních struktur RNA. Forna podporuje pouze radial diagram, ale porovnávání dvou struktur, které sice jdou zobrazit vedle sebe, už nijak neusnadňuje.

Varna Podporuje všechny tři zmíněné diagramy, ale nelze ani zobrazit dvě sekundární struktury RNA vedle sebe. Velkou výhodou nástroje VARNA by byla možnost použití na webu, ale k tomu používá Java Applets⁹, které jsou od roku 2017 považované za zastaralé¹⁰.

Ze zmíněných projektů je nejpodobnější tomu našemu R-chie, který se snaží usnadnit porovnávání sekundárních RNA struktur a nabízí i knihovnu napsanou v jazyce R. Liší se pak v samotném přístupu, protože jejich rozhraní generuje pouze statické circular nebo arc diagramy.

1.4 Kreslení grafů na základě šablony

Níže jsou zmíněné dva projekty, které úzce souvisí s naší knihovnou, protože produkují data ve formátu, se kterým pracuje naše knihovna a metody použité

⁸ <https://www.e-rna.org/r-chie/>

⁹ <https://docs.oracle.com/javase/tutorial/deployment/applet/index.html>

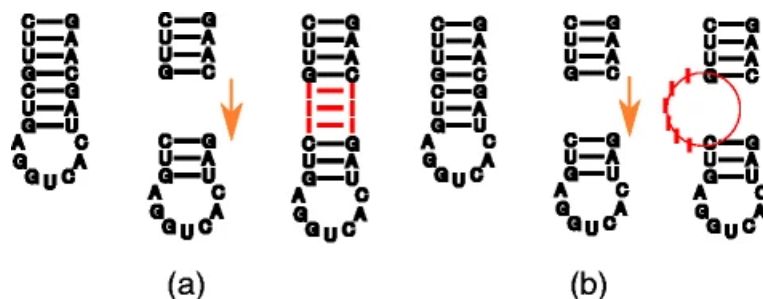
¹⁰ <https://www.oracle.com/java/technologies/javase/9-deprecated-features.html>

ke generování takových dat jsou klíčové pro naši knihovnu.

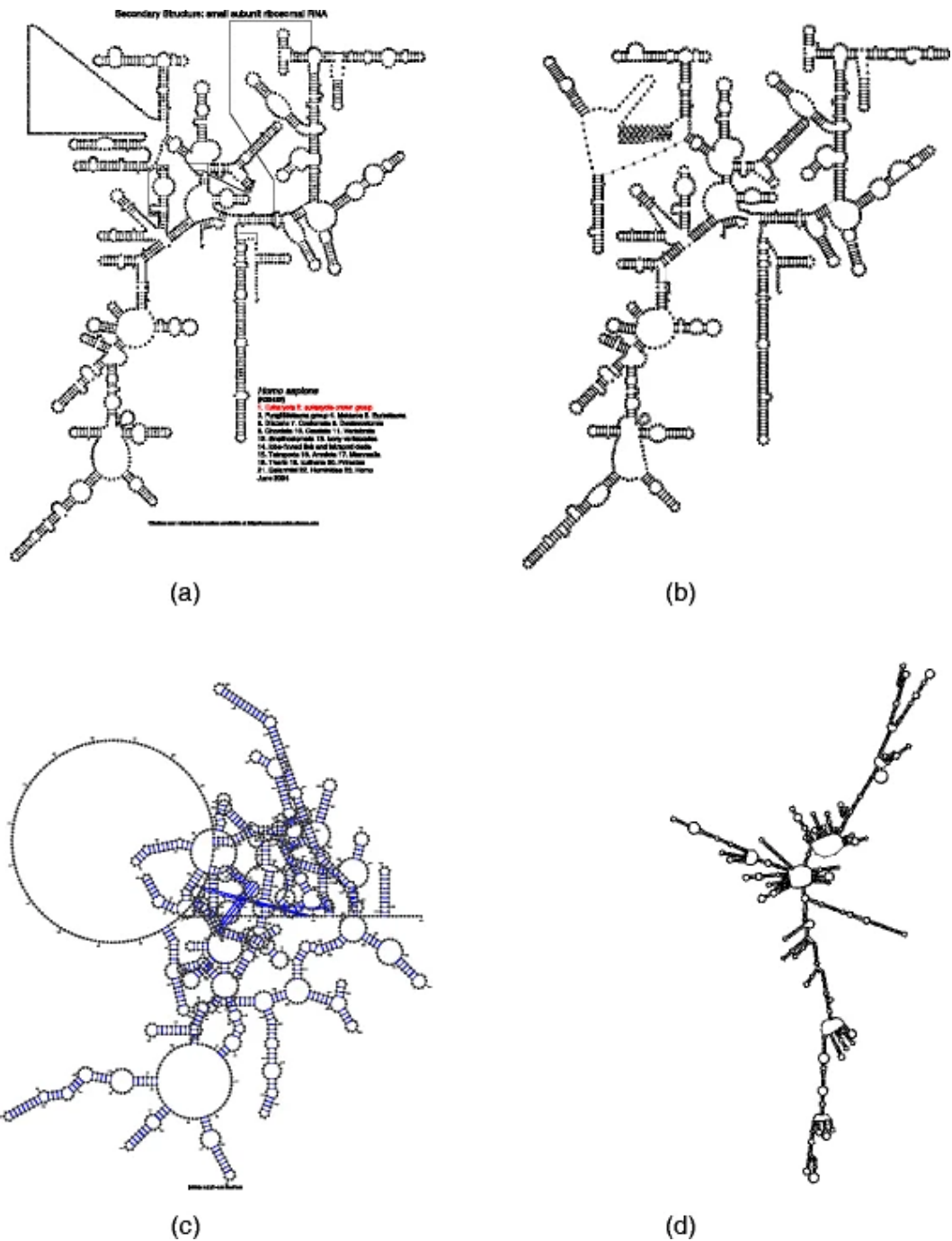
1.4.1 TRAVeLer

Traveler[4] je nástroj pro vizualizaci cílové sekundární struktury, využívající existující rozložení dostatečně podobné RNA struktury jako vzor. Traveler je založený na algoritmu, který konvertuje cílovou a vzorovou strukturu do odpovídající stromové reprezentace a využije stromovou editační vzdálenost společně s modifikací rozložení k přetvoření vzorové struktury do cílové. Traveler přijme na vstupu sekundární strukturu a vzor rozložení a na výstupu dá rozložení cílové struktury. Je to tedy command-line open source nástroj schopný rychle generovat rozložení i pro největší RNA struktury za poskytnutí dostatečně podobného rozložení.

Do vzniku Traveleru neexistoval žádný nástroj, který by dokázal velké struktury vizualizovat ve standardní notaci, se kterou jsou biologové naučení pracovat a porovnávat struktury napříč druhům.



Obrázek 1.11: Ilustrace jednoduchých editací vzorové struktury.[4]

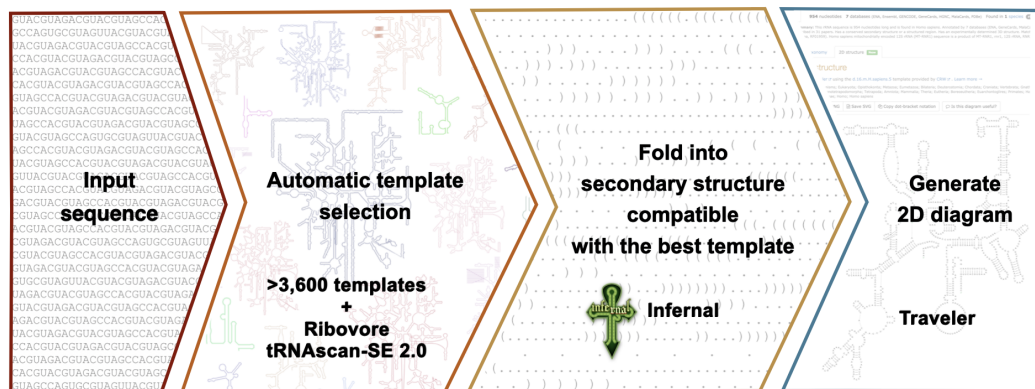


Obrázek 1.12: Rozložení malé podjednotky lidské ribozomální RNA vytvořené různými nástroji. (a) Layout ve formě, na kterou je zvyklá biologická komunita. (b) Rozložení generované Travelerem pomocí ovocné mušky jako šablony. (c) Rozložení generované nástrojem VARNA (verze 3-93). (d) Rozložení generované nástrojem RNAplot[10][4].

1.4.2 R2DT

R2DT[14] je metoda pro predikci a vizualizaci široké škály sekundárních RNA struktur ve radial diagramu. R2DT je postaveno na knihovně se 3 647 vzory reprezentujícími většinu známých RNA struktur. R2DT se používá na ncRNA

sekvencích z RNACentral¹¹ databáze a vytvořila více než 27 miliónů diagramů¹², čímž tvoří největší světovou sadu dat s 2D RNA strukturami. Pro vizualizaci neboli 2D rozložení používá R2DT právě výše zmíněný nástroj Traveler. Ostatní kroky jsou znázorněné v diagramu 1.¹³



Obrázek 1.13: Jednotlivé kroky nástroje R2DT.

¹¹ <https://rnacentral.org/>

¹² Číslo je aktuální k datu 11.4. 2023

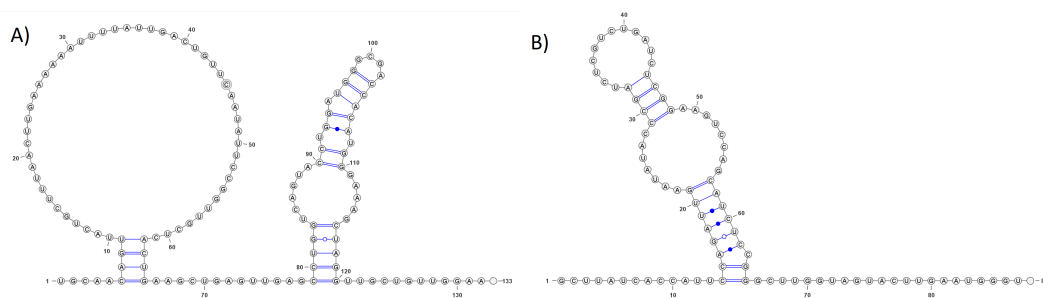
¹³ <https://github.com/rnacentral/r2dt>

2. Metody vizualizace a porovnání

Cílem této knihovny není pouze vizualizovat sekundární strukturu RNA, ale také usnadnit analýzu rozdílů a podobností mezi více strukturami RNA. Proto jsme se zaměřili na práci s radial diagramy, které nejlépe zobrazují motivy struktury a zároveň jsou nejpřirozenější reprezentací jak pro biology, kteří jsou zvyklí s takovým diagramem pracovat, tak pro širší veřejnost.

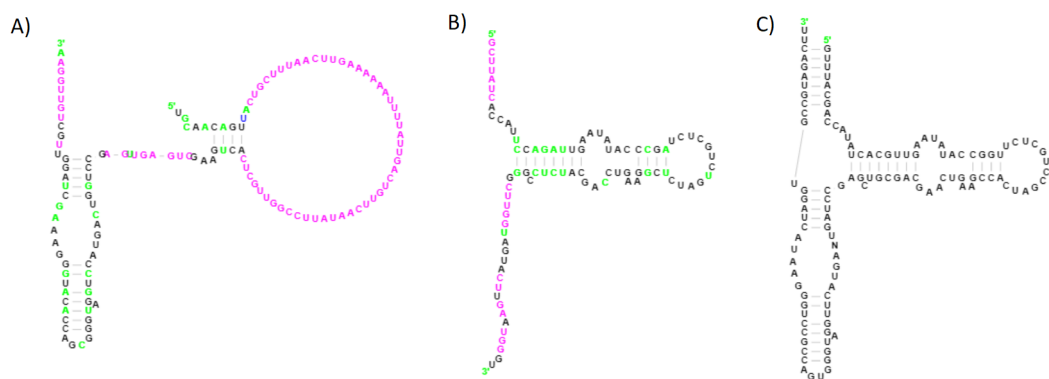
Pro generování radial diagramů používají nástroje pro vizualizaci 1.3 různé algoritmy. Nás zaujala možnost generování za použití vzorové struktury, jako to dělá nástroj Traveler. Výstupem Traveleru je soubor ve formátu JSON, který obsahuje informace o vzoru každého nukleotidu a provedených úpravách - přidání, odebrání, přejmenování a přesunutí nukleotidu.

Například pokud bychom použili nástroj Varna na zobrazení dvou struktur jako na obrázku 2.1, musíme všechny podobnosti vypořádat sami. Některé motivy sice není složité vypořádat, ale není na první pohled jasná podobnost sekvence.



Obrázek 2.1: Dvě sekundární RNA struktury s RNACentral ID A) URS00006E712C, B) URS0000AB09C9.

Na generování obou struktur používá nástroj Traveler stejnou vzorovou strukturu, která je v tomto konkrétním případě podobná obou odvozeným strukturám. Podobnost odvozených struktur ke vzorové je až na výjimky běžná, a proto jsme se rozhodli tuto podobnost předpokládat. Obrázek 2.2 je vizualizace stejných struktur, za použití nástroje Traveler spolu se vzorovou strukturou.



Obrázek 2.2: Tři sekundární RNA struktury s RNACentral ID A) URS00006E712C, B) URS0000AB09C9 vygenerované nástrojem Traveler pomocí C) vzorové sekundární RNA struktury d.5.e.S.oshimae.

Přidáním struktury, ze které jsou obě struktury odvozené, je mnohem jasnější, na která místa koukat při hledání podobností a rozdílů. Přesto nemusí být nějaká podobnost hned jasná, a proto jsme se zaměřili na způsoby, jak znázornit mapování nukleotidů na vzorovou strukturu.

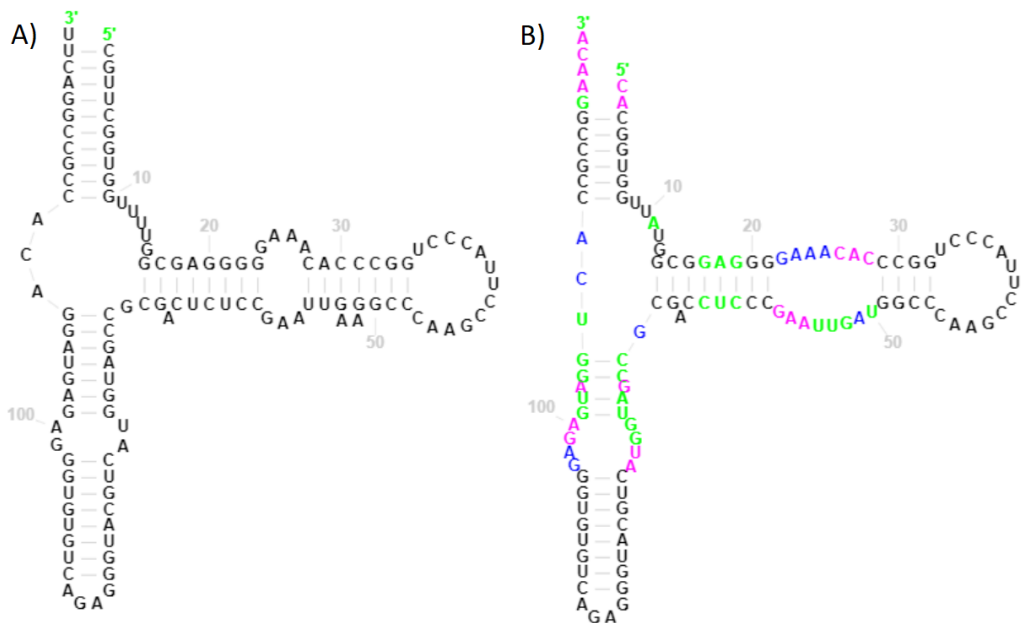
2.1 Důsledek využívání vzorové struktury

Využívání vzorové struktury pro analýzu má důležitý důsledek. Nástroj implementovaný v rámci této bakalářské práce usnadňuje analýzu N struktur, které byly vygenerované ze stejné vzorové struktury.

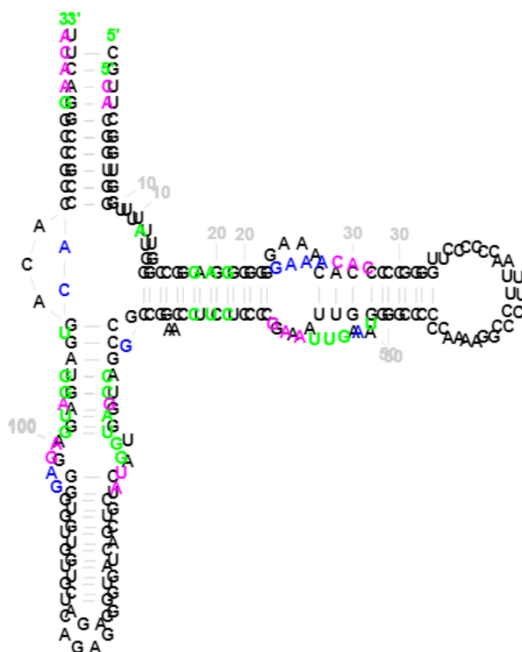
Nástroj Traveler umožňuje zvolit vzorovou strukturu, pomocí které se bude generovat rozložení. Pokud použijeme vzorovou strukturu, která není dostatečně podobná, bude i vygenerované rozložení velmi odlišné a nebude existovat mnoho nukleotidů, které půjde namapovat na vzorové nukleotidy. Z těchto důvodů lze teoreticky knihovnu, kterou představujeme v této práci použít na porovnávání jakýchkoliv struktur, protože si můžeme vynutit generování na základě stejné vzorové struktury, ale představené metody, které předpokládají podobnost nebudou užitečné.

2.2 Překládání struktur

Vygenerované sekundární struktury jsou podobné vzorové struktuře 2.3, a tím pádem bývají podobné i ostatním vygenerovaným strukturám ze stejné vzorové struktury. Dává proto smysl se pokusit struktury přes sebe přeložit 2.4, aby se spojily společné části a vynikly rozdíly. Pouhým přeložením podobných struktur přes sebe však získáme výsledek, který neposkytuje příliš zajímavé informace a je málo přehledný.



Obrázek 2.3: A) Struktura s RNACentral ID URS00000B9D9D vygenerovaná nástrojem Traveler pomocí B) vzorové struktury d.5.b.A.madurae vedle sebe.



Obrázek 2.4: A) Struktura s RNACentral ID URS00000B9D9D vygenerovaná nástrojem Traveler pomocí B) vzorové struktury d.5.b.A.madurae přeložené přes sebe.

2.2.1 Zarovnání

Je potřeba najít způsob, jak řešit problém přesunu a zarovnání struktury, protože manuální manipulace pomocí přetažení myši nebo zadávání pozice může

být zbytečně obtížná, zejména pokud se snažíme dosáhnout přesného zarovnání. Proto je velmi užitečné umožnit zarovnání sekundární RNA struktury na konkrétní nukleotid nebo skupinu nukleotidů ze vzorové struktury.

Pokud je vybrán vzorový nukleotid v pro nukleotid n z ostatních struktur, jehož vzor je v , je celá struktura přesunuta tak, aby se nukleotidy v a n překrývaly.

Skupiny nukleotidů

Pokud se snažíme najít větší skupinu nukleotidů, na které chceme strukturu zarovnat, koukáme se na posunutí pro konkrétní nukleotidy a následně vybereme buď posunutí které zarovná nejvíc nukleotidů nebo to které zarovná skupinu nukleotidů které jsou v části struktury, která nás právě zajímá.

Pro hledání konkrétního posunutí pro zarovnání jsme použili naivní způsob, který nezaručuje nejlepší možné zarovnání, ale zároveň je poměrně přímočarý a dává rozumné výsledky, díky již zmíněné podobnosti struktur.

Postupně se prochází každá struktura. V první iteraci se porovná pozice každého nukleotidu s pozicí ve vzorové struktuře. Na základě pozice se roztřídí nukleotidy. Třídění funguje tak, že do stejné skupiny indexované posunutím se přidávají všechny vzorové nukleotidy, na které se struktura zarovná daným posunutím.

V druhé iteraci se postupuje podobně, ale používají se poslední vytvořené skupiny na filtrování, tím se snažíme docílit, zarovnání na stejné nukleotidy. Filtrování pomocí skupin ignoruje všechny nukleotidy, jejich vzor není v nějaké skupině, pomocí které filtrujeme. Pokud během iterace nevzniknou žádné skupiny, řeší se iterace daná iterace zvlášť bez filtrování.

Zarovnání je popsáno pseudokódem 2.1, který vybírá z vytvořených skupin vždy to nejlepší zarovnání. Výsledek takového zarovnání je vidět na obrázku 2.5.

Listing 2.1: Pseudokód zarovnání struktur

```
def createGroup(residues, filterGroup):
    group je tabulka polí nukleotidů indexovaná souřadnicema
    group = prázdná tabulka
    for residue in residues:
        if (residue má vzorový nukleotid
            and (filterGroup is null
                or vzorový nukleotid residue je ve filterGroup)):
            t = posunutí residue, aby byl na stejné pozici,
              jako je jeho vzorový nukleotid
            přidat residue na konec pole v tabulce tempGroup
            s indexem t
    }
    return group

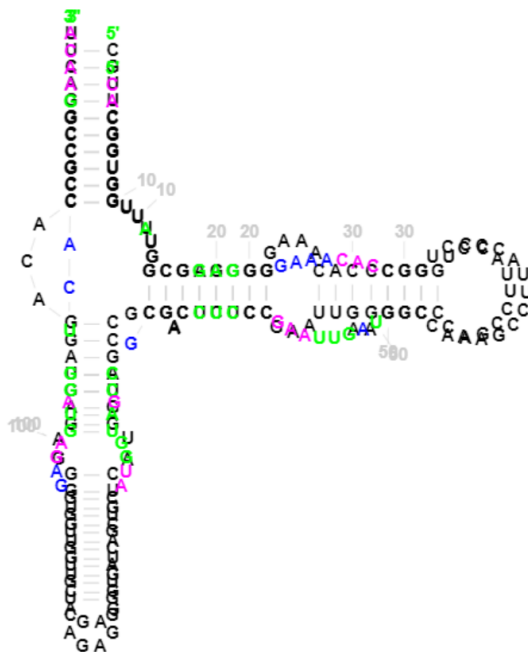
structures = pole struktur odvozených od stejné
              vzorové struktury
filterGroup je tabulka polí nukleotidů indexovaná
souřadnicema
filterGroup = null

for structure in structures:
    residues = pole nukleotidů struktury structure
    group = createGroup(residues, filterGroup)
    if (group je prázdná):
        group = createGroup(residues, null)
```

```

if (group není prázdná):
    bestShift = index (posunutí) z tabulky group, který má
                    nejdelší pole
    posuneme strukturu structure o bestShift
    filterGroup = tempGroup
    tempGroup = prázdná tabulka

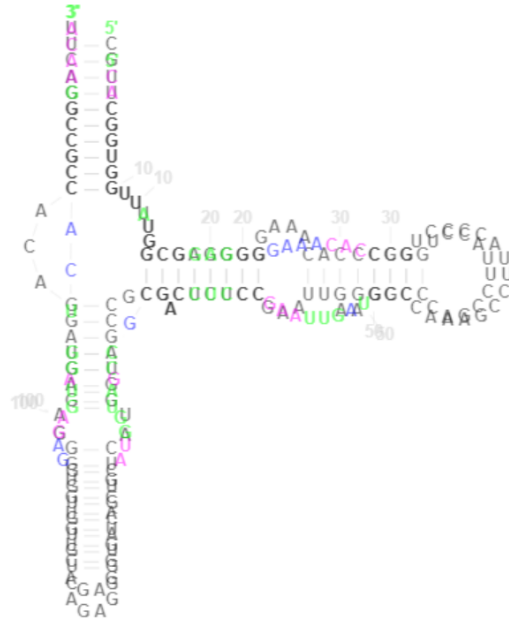
```



Obrázek 2.5: A) Struktura s RNACentral ID URS00000B9D9D vygenerovaná nástrojem Traveler pomocí B) vzorové struktury d.5.b.A.madurae přeložené přes sebe a zarovnané.

2.2.2 Průhlednost struktur

V přeložených a zarovnaných strukturách nelze snadno rozeznat, které nukleotidy jsou společné a překrývají se a které nejsou. Přidáním průhlednosti je možné tuto situaci rozlišit, protože překrývající se nukleotidy budou mít sytější barvu než ty, které se nepřekrývají^{2.6}.



Obrázek 2.6: A) Struktura s RNACentral ID URS00000B9D9D vygenerovaná nástrojem Traveler pomocí B) vzorové struktury d.5.b.A.madurae přeložené přes sebe, zarovnané a s průhledností.

2.2.3 Rozmazání struktur

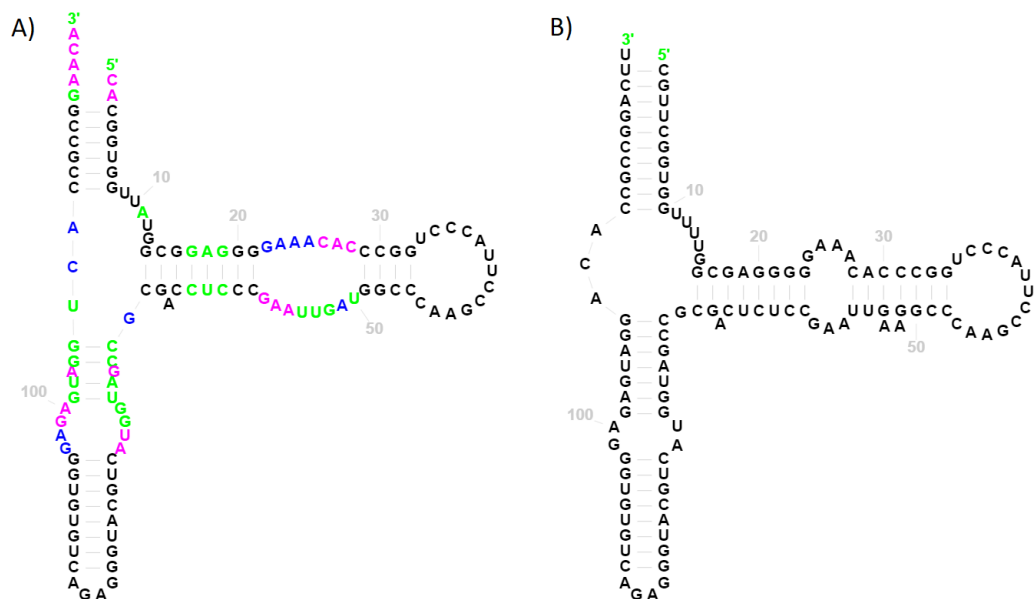
Zarovnávání struktur bohužel neřeší všechny výzvy. Výsledné obrázky se mohou zdát rozmazané. To je způsobeno tím, že ačkoli má nukleotid vzorový nukleotid, který je stejný, jeho pozice se může v rozložení mírně lišit v důsledku metody generování dat. Tento fakt může způsobit, že popisky nukleotidů vypadají rozmazaně.

Jako přímočaré řešení se může zdát posunutí jednotlivých nukleotidů, které jsou blízko sebe, aby dokonale překrývaly jejich vzory. Věříme, že by to vyřešilo zmíněný problém bez významné deformace struktury.

2.3 Obarvení struktur

Vstupní data obsahují barevné označení nukleotidů^{2.7}. Slouží k lepšímu zorientování se ve struktuře vzhledem ke vzorové struktuře. Jejich význam je následující.

Černá barva značí, že nukleotid leží na poloze vzorového nukleotidu se stejným názvem. Zelenou barvou jsou označeny ty nukleotidy jejich vzorový nukleotid bylo třeba přejmenovat. Modrou barvou jsou vyznačeny posunutý nukleotid. A poslední růžovou barvu mají nově přidáné nukleotidy.

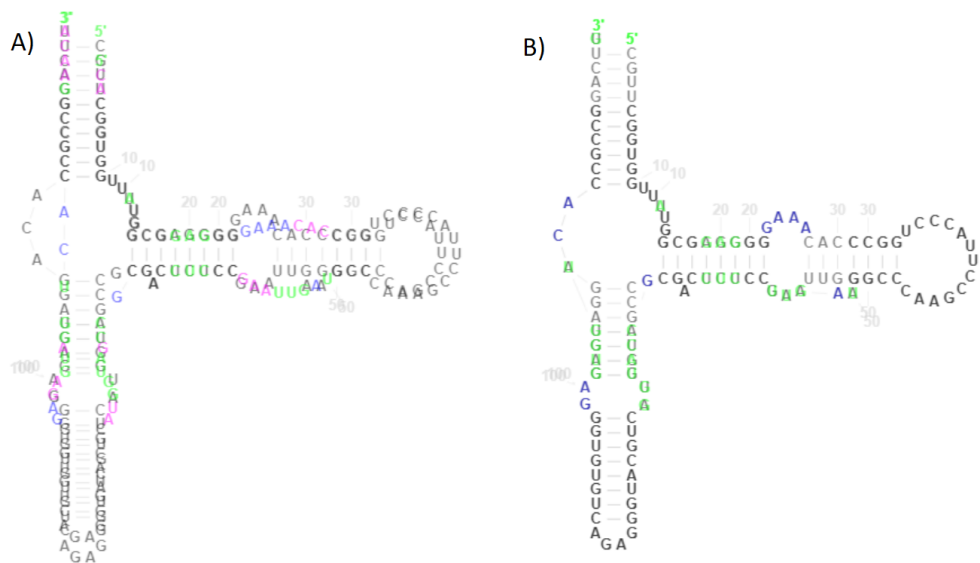


Obrázek 2.7: A) Struktura s RNACentral ID URS00000B9D9D vygenerovaná nástrojem Traveler pomocí B) vzorové struktury d.5.b.A.madurae.

2.4 Transformace na vzor

Užitečnou metodou je animovaná transformace mezi vzorovou a cílovou strukturou^{2.8}. Každý nukleotid, který má svůj vzorový nukleotid, se přemístí na pozici vzorového nukleotidu a nukleotidy, které ve vzoru nejsou, jsou skryté. Tato metoda je velmi užitečná pro práci s dvěma strukturami, které jsou si podobné, nebo pro získání počátečního přehledu o tom, co je na co namapováno.

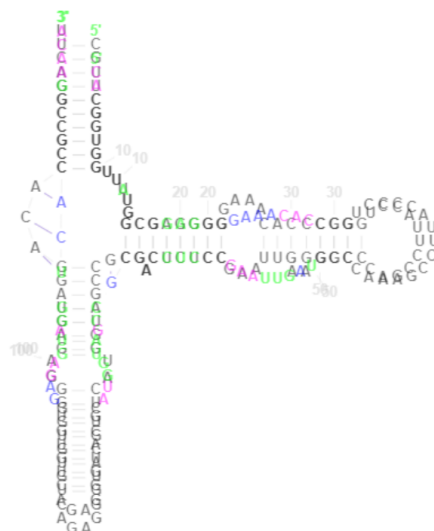
Slabou stránkou této metody je její použití při práci s více než dvěma strukturami nebo strukturami, které jsou velmi odlišné. V takových situacích se na obrazovce děje mnoho věcí a je obtížné se soustředit a vyzorovat něco užitečného.



Obrázek 2.8: Struktura s RNACentral ID URS00000B9D9D vygenerovaná nástrojem Traveler pomocí vzorové struktury *d.5.b.A.madurae* přeložené přes sebe A) před transformací a B) po transformaci.

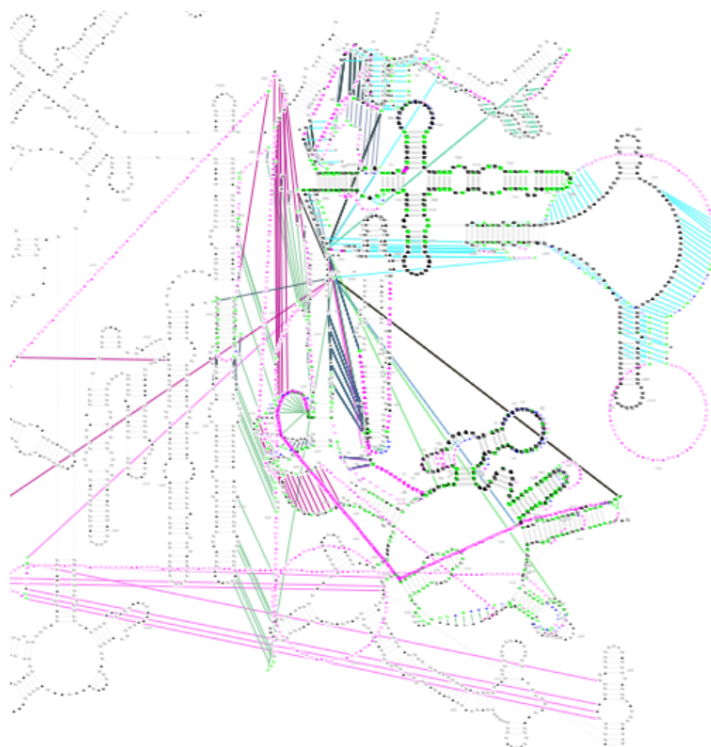
2.5 Mapovací čáry

Vědomost o tom, který nukleotid se na co mapuje, může být důležitá pro odhalení rozdílů a podobností mezi strukturami. V našem úsilí zprostředkovat tuto informaci již před transformací na vzorovou strukturu jsme přišli s metodou, která spojuje každý nukleotid s jeho vzorovým nukleotidem čarou 2.9.



Obrázek 2.9: Struktura s RNACentral ID URS00000B9D9D vygenerovaná nástrojem Traveler pomocí vzorové struktury *d.5.b.A.madurae* přeložené přes sebe s mapovacími čarami.

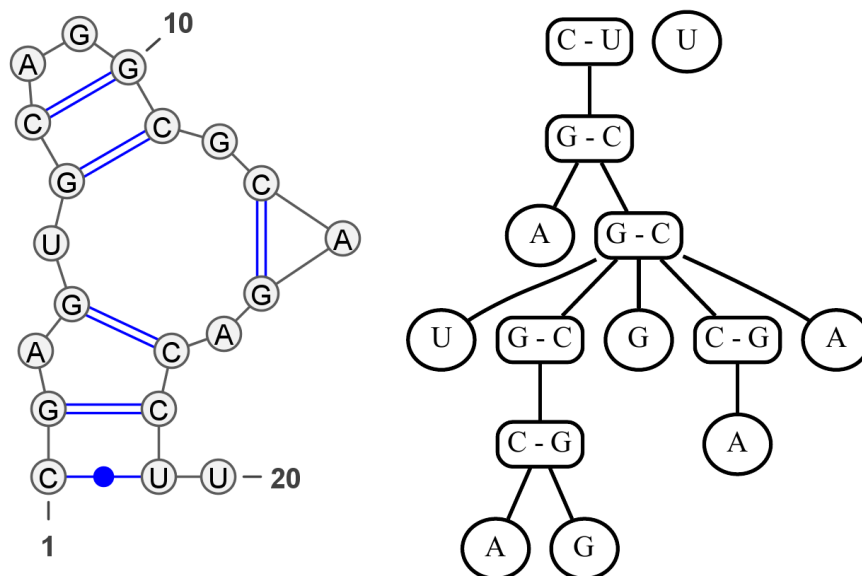
Bohužel tento způsob se zvětšující se velikostí struktury stává velmi nepřehledným^{2.10}, přesto si myslíme že můžou být užitečné.



Obrázek 2.10: Výřez z mnoha struktur vygenerovaných nástrojem Traveler pomocí vzorové struktury DD_28S_3D přeložených přes sebe s mapovacími čarami. Každá struktura má vlastní barvu mapovacích čar.

2.6 Využití stromu

V projektu Traveler byla použita stromová reprezentace pro sekundární RNA strukturu, ve které je vnitřní vrchol, tedy má více jak jednu hranu, básový pár a listem, tedy má jednu nebo žádnou hranu, je nespárovaný nukleotid. Strom pak lze vytvořit následovně. První nespárované nukleotidy ze začátku a z konce přidáme jako samostatné vrcholy. Potom jdeme postupně po básových párech. První básový pár tvoří kořen. Další básově páry připojujeme a tvoříme strom. Pokud narazíme na nespárovaný nukleotid přidáme ho jako list k poslednímu přidanému básovému páru. Větvení struktury vyústí ve větvení stromu. V obrázku 2.11 je ukázka radial diagramu sekundární struktury RNA spolu s její stromovou reprezentací.



Obrázek 2.11: Vlevo je uměle vytvořená struktura zobrazená nástrojem Varna. Vpravo je jeho stromová reprezentace.

Stromovou strukturu jsme neměli v úmyslu použít na nic konkrétního, ale chtěli jsme prozkoumat možnosti lokálních transformací struktury pro dosažení zarovnání více podobných částí, které jsou třeba jenom posunuté, jako je to například vidět na obrázku 2.12 a 2.13. Obrázky jsou výřezy dvou strukturu, které jsou zarovnané dvěma způsoby.



Obrázek 2.12: První způsob zarovnání.



Obrázek 2.13: Druhý způsob zarovnání.

Po zvážení jsme dospěli k názoru, že není možné provést úpravu struktury, která by významně nedeformovala strukturu. Jakákoli úprava části struktury znamená posunutí nějakého zbytku struktury, což může zasáhnout do jiné části. Proto by bylo nezbytné odebrat některé nukleotidy. Ty, které nejvíce překážejí, jsou ty přidané. Pokud bychom odebrali tyto přidané nukleotidy, dostali bychom se na stejný výsledek jako s transformací struktury na vzorovou strukturu.

2.7 Zvolené metody

Naše knihovna přímo podporuje překládání a zarovnání struktur na konkrétní nukleotid nebo skupinu nukleotidů bez zarovnání na úrovni jednotlivých nukleotidů pro odstranění rozmazání struktur. Lze transformovat struktury na vzorovou strukturu a upravovat průhlednost struktur.

3. Dokumentace

Vytvořili jsme Javascript knihovnu napsanou v jazyce Typescript, která nabízí jednoduchý inteface pro vytváření programů usnadňující porovnání několika sekundárních struktur RNA. V této kapitole odůvodníme volbu technologie canvas¹ pro vykreslování, seznámíme čtenáře s kódem, vstupními daty a v neposlední řadě ukážeme, jak knihovnu používat.

3.1 Použití canvasu

3.1.1 Knihovna D3.js

D3.js² je knihovna v jazyce Javascript pomáhající přivést data k životu podporující především SVG³. SVG je webový standard, který lze skvěle kombinovat s ostatními standardy jako je CSS, DOM⁴, Javascript.

D3.js knihovna klade důraz na webové standardy a tím dává uživateli možnost využívat moderní prohlížeče naplno bez dalších frameworků. S knihovnou není nejjednodušší se naučit pracovat, ale její velkou předností je rychlost a široká paleta funkcí pro vizualizace dat.

3.1.2 SVG

V počátcích jsme chtěli k zobrazování používat SVG, které V kombinaci s D3.js knihovnou je velmi silný a flexibilní nástroj.

Bohužel SVG je poměrně pomalé a je to poznat už při vykreslování tisícovek objektů. Takového počtu objektů můžeme dosáhnout pouze s jednou sekundární RNA strukturou, protože každá čára je jeden objekt a každý nukleotid se vykresluje jako dva objekty. My navíc chceme zobrazit více takových struktur a ještě s nimi dynamicky pracovat. Tím se pro nás stalo SVG nepoužitelné.

3.1.3 Canvas

Další možností bylo využití canvasu, který slibuje výrazně lepší výkon a lze ho stále jednoduše používat. Některé pro nás klíčové funkce D3.js knihovny lze využít i pro práci s canvasem. Konkrétně se jedná o zoom, posouvání a animace. U canvasu jsme nakonec i zůstali, přestože při práci s desítkami velkých struktur vykreslování není plynulé.

¹ https://developer.mozilla.org/en-US/docs/Web/API/Canvas_API

² <https://d3js.org/>

³ <https://developer.mozilla.org/en-US/docs/Web/SVG>

⁴ https://developer.mozilla.org/en-US/docs/Web/API/Document_Object_Model

3.2 Struktura vstupních dat

Výstupem nástroje Traveler jsou data ve formátu JSON, obsahující všechny potřebné informace o rozložení nukleotidů, jejich párování, velikosti popisků, barvách a tloušťkách čar. Kromě informací o rozložení lze z nich také vyčíst informace o potřebných editacích vzorové sekundární struktury. Výstupní data nástroje Traveler slouží jako vstupní data naší knihovny, proto podrobněji popíšeme jejich strukturu.

RNA je sekvence nukleotidů, to znamená, že každý nukleotid má svůj index (pořadí). V popisovaných datech je uložený index i jméno (A, C, G nebo U) pro každý nukleotid. Kromě vlastního indexu a jména je u nukleotidu n uloženo i jméno a index vzorového nukleotidu v neboli index a jméno nukleotidu v ve vzorové struktuře, ze kterého lze nukleotid n pomocí nějaké ze zmíněných editací získat. Přidaný nukleotid je nukleotid, který nevznikl modifikací nějakého nukleotidu ve vzorové struktuře a takový nukleotid má index vzorového nukleotidu -1 . Za pomoci těchto informací lze jednoduše zjistit, jestli je nukleotid přidaný nebo přejmenovaný a pokud máme k dispozici i data vzorové struktury, lze pomocí indexu porovnávat jejich souřadnice.

V rámci R2DT projektu vzniká JSON schéma⁵, které by mělo popisovat strukturu vstupních dat. Schéma je stále ve vývoji, proto aktuální výstupy Traveleru se mírně liší od schématu a je dost možné, že se formát výstupu bude v budoucnu ještě měnit a naše knihovna se jim bude přizpůsobovat.

Samotná struktura dat není složitá, ale popíšeme zde pouze tu část, kterou aktuálně využíváme, kromě toho, že ostatní data pro nás nejsou důležitá, tak jak již bylo zmíněno samotná struktura dat není pevně daná a může se měnit.

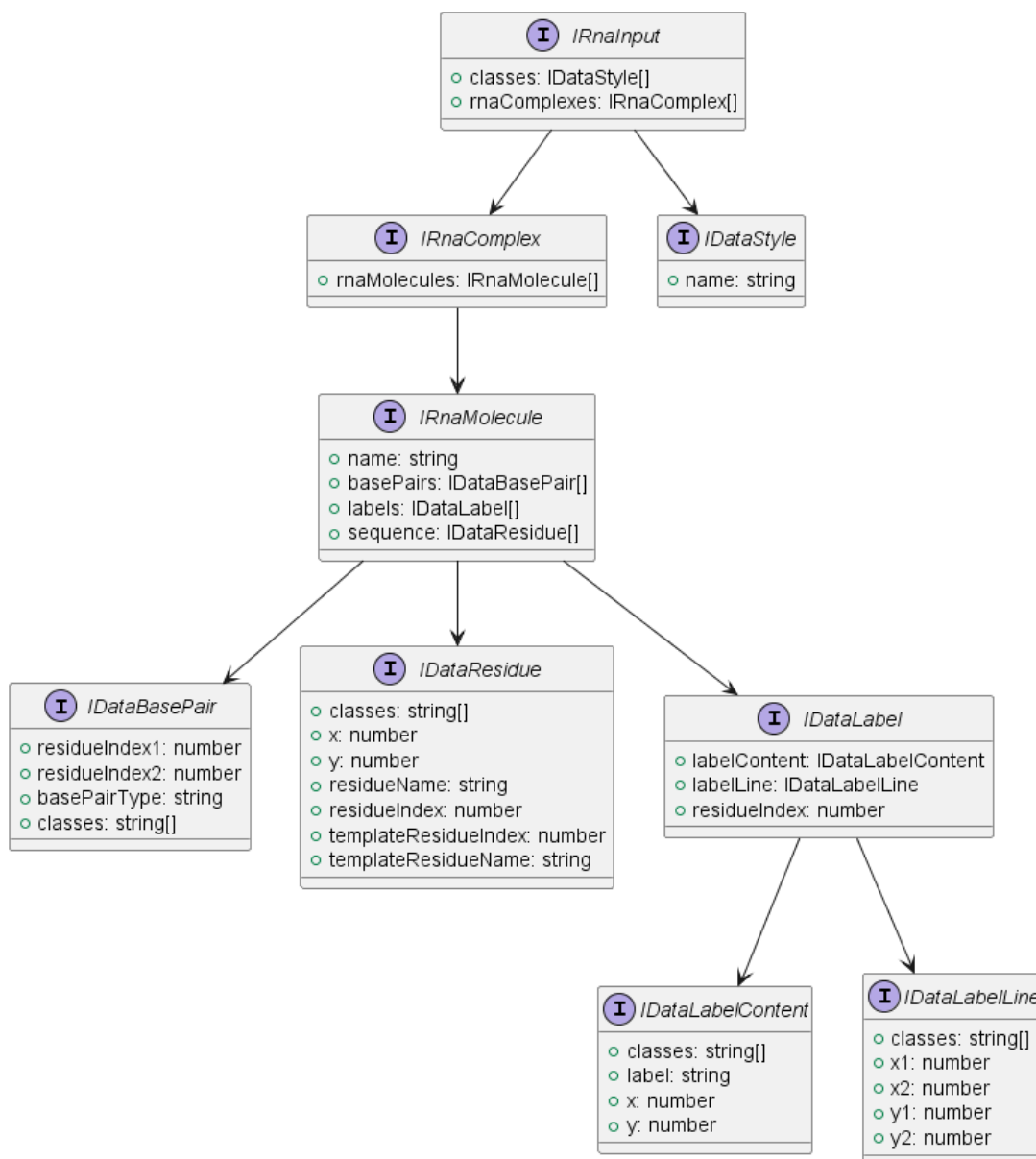
Jedná se o objekt, který má dvě položky - `rnaComplexes` a `classes`, což je pole objektů popisující třídy říkající způsob zobrazení struktury, podobně jako to kaskádové styly⁶ (CSS) diktují pro webové stránky.

`rnaComplexes` je pole polí sloužící pro popis celých skupin RNA struktur. Naše knihovna pracuje vždy pouze s nultým prvkem. Neviděli jsme důvod to dělat jinak, a pokud by se nějaký důvod našel v budoucnu, neměl by být problém naší knihovnu přizpůsobit situaci (např. rozšířením o novou metodu pro zachování zpětné kompatibility).

V rámci naší knihovny jsme vytvořili interface, který vstupní data musí splňovat. Struktura zbytku dat by měla být jasně viditelná z diagramu 3.1 těchto interfaců.

⁵ <https://github.com/LDWLab/RNA2D-data-schema>

⁶ <https://developer.mozilla.org/en-US/docs/Web/CSS>



Obrázek 3.1: Interface pro vstupní data.

3.3 Objektový návrh

V následující části se pokusíme čtenáře seznámit s objektovým návrhem naší knihovny. V této části dáme obecný pohled na strukturu tříd a později i ukážeme jak se s třídami dá pracovat.

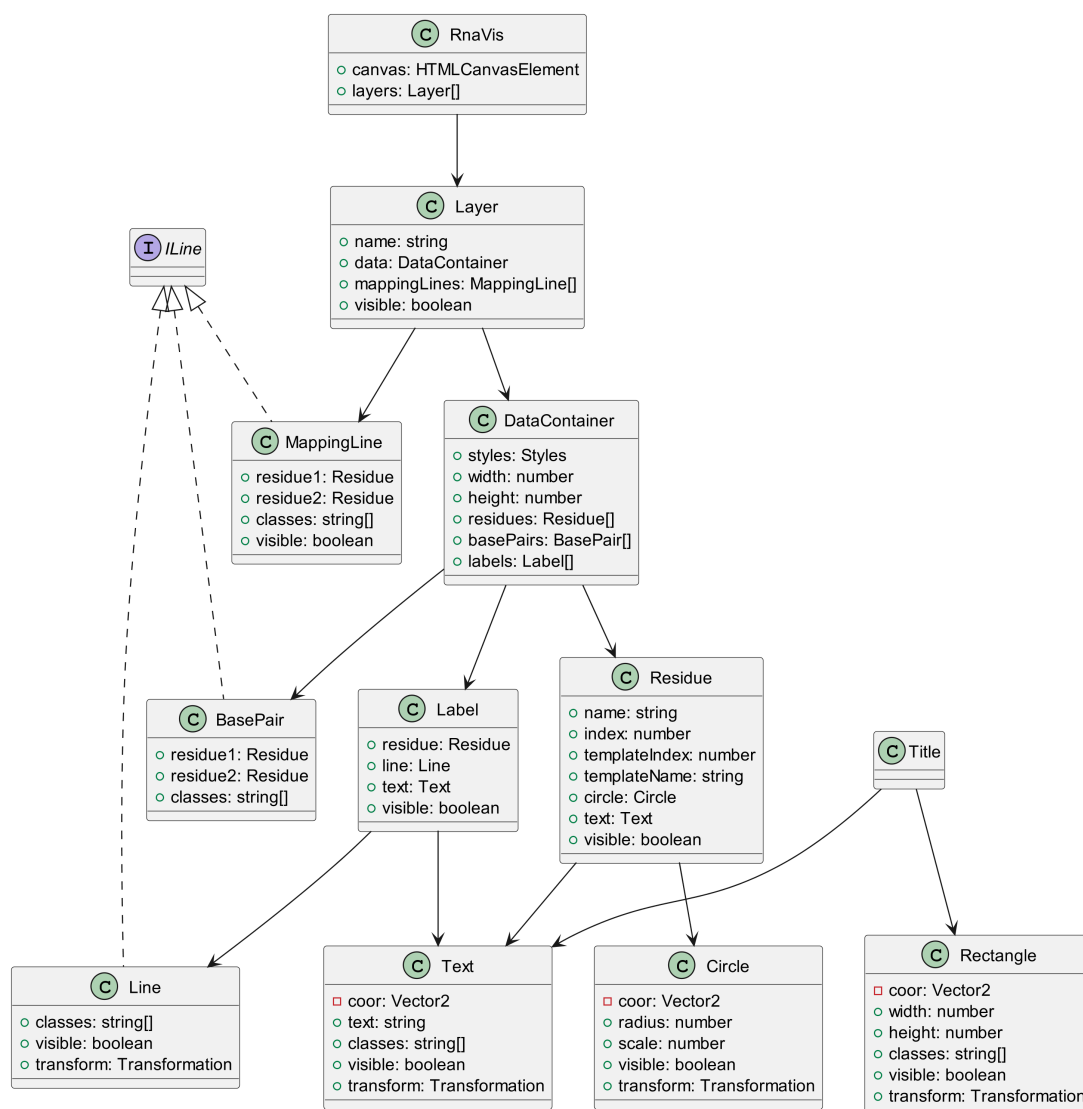
Srdcem celé knihovny je třída *RnaVis*. Má v sobě uložené vrstvy realizované třídou *Layer*, představující jednotlivé struktury. Třída *RnaVis* vykresluje struktury na canvas, nastavuje se přes ní zoom/posouvání a zpřístupňuje většinu funkcionality.

Pro určení vykreslovacích parametrů (např. font, barva, velikost) objektů máme třídu *Styles*. Před každým vykreslením se zeptáme této třídy na vykreslovací parametry pro daný objekt. Důvod, proč jsme zvolili takovéhle řešení

problému je, že jsme původně neměli jasnou představu, co a jak bude řešené, a proto nám přišlo nejjednodušší tuto část oddělit a použít stejný způsob, jaký se používá ve vstupních datech.

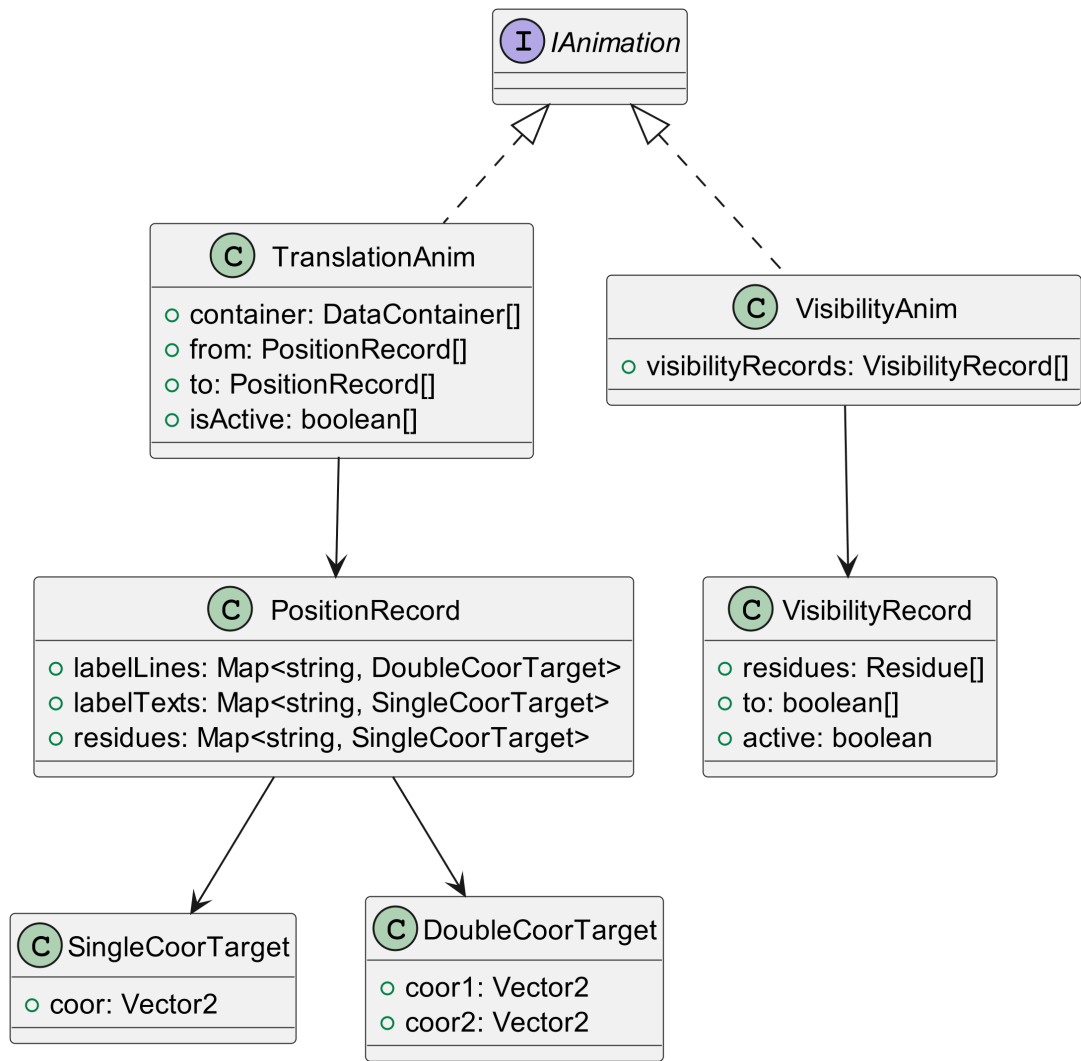
Poslední důležitou velkou částí jsou animace. Tuto funkcionalitu zpřístupňují dvě hlavní třídy - `TranslationAnim`, `VisibilityAnim`. Obě implementují interface `IAnimation`.

Diagramy tříd 3.2, 3.3, 3.4 dohromady obsahují každou třídu. Diagramama se snažíme vyjádřit obecnou strukturu, tím pádem pro přehlednost neobsahují všechny informace - všechny metody, některé privátní vlastnosti a některé závislosti. Pro podrobnější rozbor tříd a metod se doporučujeme podívat do referenční dokumentace⁷.

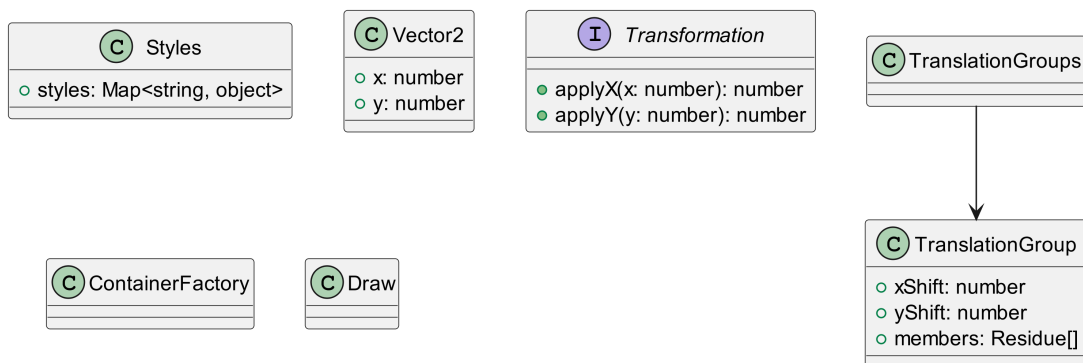


Obrázek 3.2: Diagram tříd.

⁷ <https://github.com/michalhercik/rna-visualizer/blob/main/lib/docs/README.md>



Obrázek 3.3: Diagram tříd.



Obrázek 3.4: Diagram tříd.

3.4 Uživatelské seznámení s knihovnou

S obecnou strukturou kódu jsme čtenáře seznámili a nyní ukážeme, jak lze knihovnu užívat. Bude se jednat o ukázkou implementací základních využití, jako je vykreslení struktur, zarovnání, transformace na vzorovou strukturu a některé další možnosti. Knihovna je dostupná přes npm⁸ nebo na GitHub⁹.

3.4.1 Vytvoření objektu RnaVis a vykreslení struktur

V prvním kroce importujeme veškeré potřebné zdroje^{3.1}. Použitá vstupní data `d.5.b.A.madurae.json` a `URS00000B9D9D.json` jsou dostupné na GitHub stránce projektu^{10 11}

Listing 3.1: Import

```
import {
  RnaVis,
  TranslationGroups,
  PositionRecord,
  TranslationAnim,
} from 'rna-visualizer';
import template from './d.5.b.A.madurae.json';
import structure from './URS00000B9D9D.json';
```

Následně vytvoříme canvas, na který budeme vykreslovat, nastavíme mu šířku, výšku a přidáme ho do těla dokumentu^{3.2}.

Listing 3.2: Vytvoření a nastavení canvas

```
const canvas = document.createElement('canvas');
canvas.width = 500;
canvas.height = 500;
document.body.appendChild(canvas);
```

Nyní už můžeme vytvořit samotný objekt `RnaVis`^{3.3}, který nás bude provázet v ostatních ukázkách.

Listing 3.3: Vytvoření RnaVis

```
const rnaVis = new RnaVis(canvas);
```

Přidáme do něj dvě struktury - `template` a `structure`^{3.4}. Struktury se přidávají pomocí metody `addLayer`, která má tři parametry - přidávanou strukturu, libovolný název a poslední třetí parametr je volitelný a slouží k určení, jestli chceme, aby se struktura vykreslovala a jeho defaultní hodnota je `true`, tedy defaultně se struktura vykresluje. Knihovna předpokládá, že první přidaná struktura, je vzorová struktura, ze které jsou všechny ostatní vygenerované.

⁸ <https://www.npmjs.com/package/rna-visualizer>

⁹ <https://github.com/michalhercik/rna-visualizer>

¹⁰ https://github.com/michalhercik/rna-visualizer/blob/main/demo/example_data/madurae/URS00000B9D9D_471852-d.5.b.A.madurae.json

¹¹ https://github.com/michalhercik/rna-visualizer/blob/main/demo/example_data/madurae/d.5.b.A.madurae_fromn_d.5.b.A.madurae.colored.json

Listing 3.4: Přidání struktur

```
rnaVis
  .addLayer(template, 'd.5.b.A.madurae')
  .addLayer(structure, 'URS00000B9D9D');
```

Struktury lze potom vykreslit pomocí metody `draw`3.5.

Listing 3.5: Vykreslení struktur

```
rnaVis.draw();
```

Celý kód lze zapsat následovně3.6.

Listing 3.6: Celý kód pohromadě

```
import {
  RnaVis,
  TranslationGroups,
  PositionRecord,
  TranslationAnim,
} from 'rna-visualizer';
import template from './d.5.b.A.madurae.json';
import structure from './URS00000B9D9D.json';

const canvas = document.createElement('canvas');
canvas.width = 500;
canvas.height = 500;
document.body.appendChild(canvas);

new RnaVis(canvas)
  .addLayer(template, 'd.5.b.A.madurae')
  .addLayer(structure, 'URS00000B9D9D')
  .draw();
```

V dalších ukázkách budeme předpokládat předchozí kroky a budeme na ně navazovat.

3.4.2 Zoom a posouvání canvasu

Pro přidání zoomu a posouvání stačí zavolat metodu `addZoom` na instanci třídy `RnaVis`3.7, která přidá obě funkcionality canvasu předanému jako parametr konstruktoru objektu. Interně používá metoda `D3.js` knihovnu pro zpracování zoom a posunutí canvasu.

Listing 3.7: Přidání zoom a posouvání

```
rnaVis.addZoom();
```

3.4.3 Zarovnání struktur

Zarovnání struktur2.5 docílíme tak, že zavoláme na instanci třídy `RnaVis` metodu `align`, která vrátí pole objektů typu `Vector2`, které reprezentují posunutí pro každou strukturu. Následně zavoláme na instanci třídy `RnaVis` metodu `translate`, které předáme pole obsahující posunutí a metoda posune každou strukturu o příslušné posunutí. Potom už stačí pouze vykreslit struktury metodou `draw`3.8.

Listing 3.8: Zarovnání struktur

```
const translations = rnaVis.align();
rnaVis
  .translate(translations);
  .draw();
```

Zarovnání struktur na konkrétní skupinu

Metoda `align` má volitelný parametr typu `number`, kterým můžeme vybrat, na kterou skupinu nukleotidů chceme dostat zarovnání 3.9.

Při první iteraci se porovnává první přidaná struktura (vzorová) s druhou přidanou strukturou a roztrídí se nukleotidy podle posunutí, které je zarovná. Číselný parametr potom určuje, která skupina se použije pro filtrování v dalších iteracích.

Pokud je parametr prázdný nebo menší než nula, vybere se největší skupina. Pokud je parametr větší, než počet vytvořených skupin metoda vyhodí výjimku se zprávou `groupIndex >= groups.length`.

V následující ukázce vytvoříme skupiny, které vzniknout při první iteraci a vybereme nejmenší, kterou pak použijeme pro zarovnání struktur. Skupiny vytvoříme statickou metodou `TranslationGroups.create`. První parametr je vzorová struktura, druhý parametr je struktura, kterou chceme zarovnat. Třetí parametr jsou skupiny pomocí, kterých chceme filtrovat a je volitelný a poslední parametr je minimální velikost výstupních skupin, který je taky volitelný a jeho defaultní hodnota je 5.

Listing 3.9: Zarovnání struktur na skupinu nukleotidů

```
const containers = rnaVis.getDataContainers();
const groupSizes =
  TranslationGroups.create(containers[0], containers[1], null, 20)
  .map(group => group.size());
const groupIndex = groupSizes.indexOf(Math.min(...groupSizes));
const translationsToGroup = rnaVis.align(groupIndex);
rnaVis
  .translate(translationsToGroup)
  .draw();
```

Zarovnání na nukleotid

Zarovnání na nukleotid obsahuje podobné kroky. Nejdřív si zvolíme nukleotid ze vzorové struktury, na který chceme struktury zarovnávat. Potom na instanci třídy `RnaVis` pomocí metody `getAlignmentToTempResidue` získáme posunutí pro každou strukturu a nakonec struktury zarovnáme a vykreslíme 3.10.

Listing 3.10: Zarovnání struktur na nukleotid

```
const templateResidue = rnaVis.layers[0].data.residues[42];
const translationsToResidue =
  rnaVis.getAlignmentToTempResidue(templateResidue);
rnaVis
  .translate(translationsToResidue)
  .draw();
```

Animace zarovnání

Jakýkoliv přechod je pro oko vždy příjemnější animovat a to knihovna umožňuje pomocí jednoduchého rozhraní pro animace. V následující ukázce ukážeme jak animovat zarovnání struktur.

V prvním kroce vytvoříme pro každou strukturu `PositionRecord`, který obsahuje pozice nukleotidů a popisků. V tomto případě bude obsahovat cílovou pozici pro každý nukleotid a vytvoříme ho pomocí statické metody `fromTranslation`, která bere jako parametr strukturu a posunutí struktury.

Následně vytvoříme objekt typu `TranslationAnim`, který umožňuje mimo jiné animovat transformaci struktury na daný `PositionRecord`. Potom už jenom zavoláme metodu `animate`, kterou začne celá animace. Jako parametry bere instanci třídy `RnaVis`, kterou používá pro vykreslování každého snímku animace, trvání animace v milisekundách a poslední volitelný parametr je funkce, která se zavolá na konci animace. V ukázce na konci animace jí otočíme a animujeme zpátky do původního stavu3.11.

Listing 3.11: Animace zarovnání

```
const containers = rnaVis.getDataContainers();
const animationTargets = rnaVis
  .align()
  .map((translation, i)
    => PositionRecord.fromTranslation(containers[i], translation)
  );

const anim = new TranslationAnim(containers, animationTargets);
anim.animate(rnaVis, 1500, () => {
  anim.reverse();
  anim.animate(rnaVis, 1500);
});
```

3.4.4 Mapovací čáry

Zobrazení čar, které ukazují mapování nukleotidů, je velmi jednoduché. Mapovací čáry se vytvoří hned při přidání každé vrstvy a jsou defaultně schované, proto je stačí pouze zviditelnit3.12.

Listing 3.12: Zobrazení mapovacích čar

```
rnaVis.layers.forEach(
  l => l.mappingLines.forEach(
    m => m.setVisible(true)
  )
);
rnaVis.draw();
```

3.4.5 Transformace na vzorovou strukturu

Poslední ukázkou je transformace na vzorovou strukturu. V ukázce nejdříve schováme nukleotidy, které nemají vzorový nukleotid. Potom vytvoříme opět `PositionRecord`, tentokrát pomocí statické metody `fromTemplate`, která každému nukleotidu dá pozici jeho vzorového nukleotidu. Potom už jenom animujeme přechod3.13.

Listing 3.13: Transformace na vzorovou strukturu

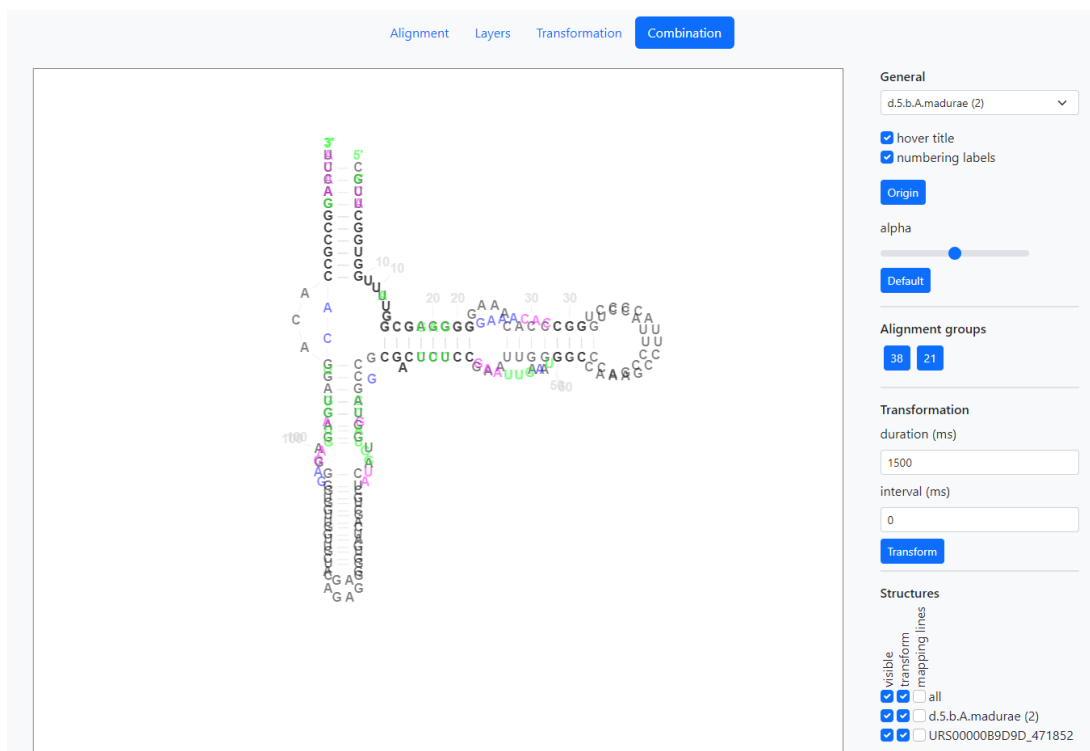
```
const containers = rnaVis.getDataContainers().slice(1);
const targetStructure = rnaVis.layers[0].data;

containers.forEach(c => {
  c.getUnmappableResidues().forEach(
    r => r.setVisible(false)
  );
});

const targets = containers.map(
  cont => PositionRecord.fromTemplate(cont, targetStructure)
);
new TranslationAnim(containers, targets)
  .animate(rnaVis, 1500);
```

3.5 Demonstrace metod

V rámci naší knihovny jsme vytvořili také webovou aplikaci¹²3.5, která demonstruje možnosti naší knihovny. Tato aplikace umožňuje uživatelům pracovat s jednotlivými metodami nebo se všemi metodami najednou pro usnadnění porovnání - překládání se zarovnáním, transformace na vzorovou strukturu, mapovací čáry a další možnosti pro zpříjemnění práce.



Obrázek 3.5: Snímek aplikace vytvořené pro demonstraci funkcionality knihovny.

¹² <https://michalhercik.github.io/rna-visualizer/>

3.5.1 Získání testovacích dat

Pro testování a demonstraci metod potřebujeme vstupní data, která budou různorodá co do velikosti struktur tak do počtu. Zároveň potřebujeme, aby vstupní data byla realistická, protože chceme řešit skutečné problémy. Otevřená databáze RNACentral používá ke generování sekundárních struktur RNA nástroj R2DT1.4.2, proto jsou data z RNACentral vhodným kandidátem.

Pro přístup k databázi lze použít jak API¹³, tak přímo posílat SQL dotazy¹⁴. Databáze bohužel nepřístupňuje mapování nukleotidů na vzorovou strukturu, které nástroj R2DT generuje, přesto je databáze užitečná a chybějící údaje lze dogenerovat.

Pro získání testovacích dat jsme využívali veřejný SQL přístup a celý postup popisujeme v následujících bodech:

1. Pomocí SQL dotazu A.1 jsme si zobrazili vzorové struktury a nějaké vybrali.
2. Následně jsme pro každou vybranou vzorovou strukturu vybrali náhodně nejvýše 20 RNA struktur, které použili danou vzorovou strukturu A.2.
3. Posledním krokem získáme mapování struktur na vzorovou strukturu. Použijeme R2DT s parametrem `-force_template <template_id>`, kde `template_id` je identifikátor vzorové struktury, která je přiřazená generované struktuře v RNACentral. Díky znalosti, kterou vzorovou strukturu použít, se přeskočí klasifikační kroky a celý proces se urychlí.

¹³ <https://rnacentral.org/api>

¹⁴ <https://rnacentral.org/help/public-database>

Závěr

Cílem této práce bylo vytvořit knihovnu pro web, pomocí které bude možné vytvářet vizualizace několika sekundárních struktur RNA a porovnávat je mezi sebou. V této práci jsme na začátek čtenáře krátce seznámili s RNA, její funkcí a vizualizací. Následně jsme představili knihovnu napsanou v jazyce Typescript, která pomocí jednoduchého rozhraní^{3.4} zpřístupňuje metody pro porovnání dostatečně podobných struktur². Knihovna nabízí plynulé vykreslování několika struktur najednou^{3.1} a je jednoduše přístupná přes npm nebo GitHub^{3.4}.

Ukázalo se, že vizualizace několika sekundárních struktur RNA je složitá. Dokázali jsme vytvořit knihovnu pro vizualizaci, ale použité metody jsou vhodné pouze pro dvě nebo tři struktury najednou, při větším množství se metody neukázaly nápomocné, především protože vizualizace přestane být přehledná. Navíc jsme se omezili na struktury, které jsou si dostatečně podobné.

V budoucnu by bylo užitečné knihovnu rozšířit o metody, které dokážou pracovat s libovolnými strukturami nebo se strukturami, které patří do stejné klasifikace^[5]. Tím bychom se přiblížili k původnímu cíli této práce. Dalším vylepšením by mohla být optimalizace kódu, aby byla knihovna schopná pracovat plynule s větším množstvím struktur.

Seznam použité literatury

- [1] BEYLERLI, O., GAREEV, I., SUFIANOV, A., ILYASOVA, T. a GUANG, Y. (2022). Long noncoding rnas as promising biomarkers in cancer. *Non-coding RNA Research*, **7**(2), 66–70. ISSN 2468-0540. doi: <https://doi.org/10.1016/j.ncrna.2022.02.004>. URL <https://www.sciencedirect.com/science/article/pii/S246805402200018X>.
- [2] CHENNEY, S., HEITSCH, C., MIZE, C., SWENSON, S., SCHMIDT, M. D., KIRKPATRICK, A. a YOON, I. (2019). Rnastructviz. URL <https://github.com/gtDMMB/RNAstructViz/wiki>. Accessed on March 29, 2023.
- [3] DARTY, K., DENISE, A. a PONTY, Y. (2009). VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, **25**(15), 1974–1975.
- [4] ELIAS, R. a HOKSZA, D. (2017). Traveler: a tool for template-based rna secondary structure visualization. *BMC Bioinformatics*, **18**(1), 487. ISSN 1471-2105. doi: [10.1186/s12859-017-1885-4](https://doi.org/10.1186/s12859-017-1885-4). URL <https://doi.org/10.1186/s12859-017-1885-4>.
- [5] GRIFFITHS-JONES, S., BATEMAN, A., MARSHALL, M., KHANNA, A. a EDDY, S. R. (2003). Rfam: an RNA family database. *Nucleic Acids Research*, **31**(1), 439–441. ISSN 0305-1048. doi: [10.1093/nar/gkg006](https://doi.org/10.1093/nar/gkg006). URL <https://doi.org/10.1093/nar/gkg006>.
- [6] KERPEDJIEV, P., HAMMER, S. a HOFACKER, I. L. (2015). Forna (force-directed RNA): Simple and effective online RNA secondary structure diagrams. *Bioinformatics*, **31**(20), 3377–3379. ISSN 1367-4803. doi: [10.1093/bioinformatics/btv372](https://doi.org/10.1093/bioinformatics/btv372). URL <https://doi.org/10.1093/bioinformatics/btv372>.
- [7] LAI, D., PROCTOR, J. R., ZHU, J. Y. A. a MEYER, I. M. (2012). R-chie : a web server and R package for visualizing RNA secondary structures . *Nucleic Acids Research*, **40**(12), e95–e95. ISSN 0305-1048. doi: [10.1093/nar/gks241](https://doi.org/10.1093/nar/gks241). URL <https://doi.org/10.1093/nar/gks241>.
- [8] LI, B., CAO, Y., WESTHOF, E. a MIAO, Z. (2020). Advances in RNA 3d structure modeling using experimental data. *Frontiers in Genetics*, **11**. doi: [10.3389/fgene.2020.574485](https://doi.org/10.3389/fgene.2020.574485). URL <https://doi.org/10.3389/fgene.2020.574485>.
- [9] LOGANATHAN, T. a DOSS C, G. P. (2023). Non-coding RNAs in human health and disease: potential function as biomarkers and therapeutic targets. *Functional & Integrative Genomics*, **23**(1), 33.
- [10] LORENZ, R., BERNHART, S. H., HÖNER ZU SIEDERDISSEN, C., TAFER, H., FLAMM, C., STADLER, P. F. a HOFACKER, I. L. (2011). ViennaRNA package 2.0. *Algorithms for Molecular Biology*, **6**(1), 26.

- [11] MODIC, M., ULE, J. a SIBLEY, C. R. (2013). Clipping the brain: Studies of protein–rna interactions important for neurodegenerative disorders. *Molecular and Cellular Neuroscience*, **56**, 429–435. ISSN 1044-7431. doi: <https://doi.org/10.1016/j.mcn.2013.04.002>. URL <https://www.sciencedirect.com/science/article/pii/S104474311300050X>. RNA and splicing regulation in neurodegeneration.
- [12] MONDAL, P. a MEERAN, S. M. (2020). Long non-coding rnas in breast cancer metastasis. *Non-coding RNA Research*, **5**(4), 208–218. ISSN 2468-0540. doi: <https://doi.org/10.1016/j.ncrna.2020.11.004>. URL <https://www.sciencedirect.com/science/article/pii/S246805402030072X>.
- [13] RANI, S., DHAR, S. B., KHAJURIA, A., GUPTA, D., JAISWAL, P. K., SINGLA, N., KAUR, M., SINGH, G. a BARNWAL, R. P. (2023). Advanced overview of biomarkers and techniques for early diagnosis of alzheimer’s disease. *Cellular and Molecular Neurobiology*. doi: 10.1007/s10571-023-01330-y. URL <https://doi.org/10.1007/s10571-023-01330-y>.
- [14] SWEENEY, B. A., HOKSZA, D., NAWROCKI, E. P., RIBAS, C. E., MADEIRA, F., CANNONE, J. J., GUTELL, R., MADDALA, A., MEADE, C. D., WILLIAMS, L. D., PETROV, A. S., CHAN, P. P., LOWE, T. M., FINN, R. D. a PETROV, A. I. (2021). R2dt is a framework for predicting and visualising rna secondary structure using templates. *Nature Communications*, **12**(1), 3494. ISSN 2041-1723. doi: 10.1038/s41467-021-23555-5. URL <https://doi.org/10.1038/s41467-021-23555-5>.
- [15] TAN, Y.-L., WANG, X., SHI, Y.-Z., ZHANG, W. a TAN, Z.-J. (2022). rsR-NASP: A residue-separation-based statistical potential for RNA 3d structure evaluation. *Biophysical Journal*, **121**(1), 142–156. doi: 10.1016/j.bpj.2021.11.016. URL <https://doi.org/10.1016/j.bpj.2021.11.016>.
- [16] YANG, H., CUI, Y., FENG, Y., HU, Y., LIU, L. a DUAN, L. (2023). Long non-coding rnas of plants in response to abiotic stresses and their regulating roles in promoting environmental adaption. *Cells*, **12**(5). ISSN 2073-4409. doi: 10.3390/cells12050729. URL <https://www.mdpi.com/2073-4409/12/5/729>.

Seznam obrázků

1.1	Ukázka motivu hairpin loop	5
1.2	Ukázka motivu bulge	5
1.3	Ukázka motivu multibranch loop	5
1.4	Ukázka arc diagramu	6
1.5	Ukázka circular diagramu	6
1.6	Ukázka radiate diagramu	7
1.7	Snímek nástroje Varna	8
1.8	Snímek nástroje rnaStructViz	9
1.9	Snímek nástroje Forna	10
1.10	Výsledný arc diagram nástroje R-chie	11
1.11	Ilustrace jednoduchých editací vzorové struktury	12
1.12	Ukázka výsledku nástroje Traveler	13
1.13	Jednotlivé kroky nástroje R2DT	14
2.1	Dvě sekundární RNA struktury	15
2.2	Dvě odvozené sekundární struktury a jedna vzorová	16
2.3	Vygenerovaná struktura vedle se vzorovou	17
2.4	Vygenerovaná a vzorová struktura vykreslené přes sebe	17
2.5	Vygenerovaná a vzorová struktura vykreslené přes sebe	19
2.6	Vygenerovaná a vzorová struktura vykreslené přes sebe	20
2.7	Vygenerovaná struktura vedle se vzorovou	21
2.8	Před a po transformaci vygenerované struktury na vzorovou	22
2.9	Ukázka mapovacích čar	22
2.10	Ukázka mapovacích čar	23
2.11	Stromová reprezentace sekundární struktury RNA	24
2.12	První způsob zarovnání	24
2.13	Druhý způsob zarovnání	25
3.1	Interface pro vstupní data	28
3.2	Diagram tříd	29
3.3	Diagram tříd	30
3.4	Diagram tříd	30
3.5	Snímek aplikace vytvořené pro demonstraci funkcionality knihovny	35

A. Přílohy

A.1 SQL dotazy pro získání testovacích dat

Listing A.1: SQL dotaz pro získání vzorových struktur

```
SELECT
  sslm.id,
  sslm.model_name,
  sslm.model_source,
  sslm.model_length,
  sslm.rna_type,
  COUNT(*)
FROM
  rnacen.rnc_secondary_structure_layout_models sslm
  JOIN rnacen.rnc_secondary_structure_layout ssl
  ON (ssl.model_id = sslm.id)
GROUP BY
  sslm.id
ORDER BY model_name
```

Listing A.2: SQL dotaz pro získání struktur vygenerovaných ze zvolených vzorových struktur

```
SELECT
  *,
  rna.seq_short
FROM
  (
    SELECT row_number()
    OVER (PARTITION BY ssl.model_id ORDER BY ssl.id ASC) As rn,
    ssl.*,
    sslm.model_name as template_name,
    sslm.model_source as template_source,
    sslm.model_length as template_length,
    sslm.rna_type
  FROM
    (SELECT
      *
    FROM
      rnacen.rnc_secondary_structure_layout_models
    WHERE
      model_name IN ('LD_SSU_3D'
        , 'RF03163'
        , 'RF03163'
        , 'RF04183'
        , 'RNAseP_e_H_sapiens_3D'
        , 'EC_SSU_3D'
        , 'TT_SSU_3D'
        , 'mt_TetT_LSU_3D'
        , 'E-Thr'
        , 'mHS_LSU_3D'
        , 'DM_LSU_3D'
        , 'DD_28S_3D'
        , 'RF02705'
        , 'RF00025'
```

```
        , 'd.5.b.Thermus.sp'  
        , 'd.16.b.S.gougerotii'  
        , 'd.5.e.S.oshimae'  
        , 'd.16.e.P.falciparum.S'  
        , 'a.16.c.C.ruber'  
        , 'HS_LSU_3D'  
        , 'TeT_LSU_3D'  
    )  
    ) sslm  
    JOIN rnacen.rnc_secondary_structure_layout ssl  
    ON ( sslm.id = ssl.model_id)  
    ORDER BY sslm.model_name  
    ) t  
    JOIN rnacen.rna rna ON (rna.upi = t.urs)  
    WHERE rn < 20
```