

**FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY
UNIVERZITA KOMENSKÉHO
V BRATISLAVE**

VYKRESĽOVANIE RNA MOTÍVOV

Diplomová práca

**FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY
UNIVERZITA KOMENSKÉHO
V BRATISLAVE**

VYKRESĽOVANIE RNA MOTÍVOV

Diplomová práca

Študijný program: Aplikovaná informatika

Študijný odbor: 2511 Aplikovaná informatika

Školiteľ: Mgr. Tomáš Vinař, PhD.,



Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Pavol Kmeč
Študijný program: aplikovaná informatika (Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: 9.2.9. aplikovaná informatika
Typ záverečnej práce: diplomová
Jazyk záverečnej práce: slovenský

Názov: Vykresľovanie RNA motívov

Cieľ: Na vyhľadávanie RNA génov v genómoch biológovia používajú štrukturálne motívy (pripomínajúce regulárne výrazy). Úlohou je zobrazovať takéto motívy graficky pomocou techník kreslenia grafov.

Vedúci: Mgr. Tomáš Vinař, PhD.
Katedra: FMFI.KAI - Katedra aplikovanej informatiky
Dátum zadania: 08.11.2010

Dátum schválenia: 16.11.2010

doc. RNDr. Roman Ďuríkovič, PhD.
garant študijného programu

.....
študent

.....
vedúci práce

Čestné prehlásenie

Čestne vyhlasujem, že som túto diplomovú prácu vypracoval samostatne na základe pripomienok a návrhov vedúceho diplomovej práce a s použitím uvedenej literatúry.

V Bratislave 24. Februára 2013

.....

Pavol Kmeč

Pod'akovanie

Touto cestou sa chcem poďakovať vedúcemu diplomovej práce Mgr. Tomášovi Vinařovi, PhD. za všestrannú pomoc, cenné rady a pripomienky pri vypracovaní tejto práce.

Abstrakt

Autor:	Bc. Pavol Kmeč
Názov:	Vykresľovanie RNA motívov
Univerzita:	Univerzita Komenského v Bratislave
Fakulta:	Fakulta matematiky, fyziky a informatiky
Katedra:	Katedra aplikovanej informatiky
Vedúci diplomovej práce:	Mgr. Tomáš Vinař, PhD.
Počet strán:	64
Rok:	2013

Cieľom tejto práce bolo vyvinúť novú metódu na automatické vykresľovanie RNA motívov. Aplikovali sme jednu z metód silovo riadeného vykresľovania grafov a grafových štruktúr. Táto metóda postupne transformuje graf tak, že používa rôzne typy síl pôsobiace na vrcholy a podľa nich postupne posúva vrcholy na nové miesta, až kým sa graf nedostane do rovnovážneho stavu. Týmito silami je možné kontrolovať pravidlá výsledného nakreslenia grafu. Na vytvorenie grafu RNA motívu sme použili jeden zo štandardných deskriptorov používanom v nástroji RNABob. Metódu silovo riadeného vykresľovania grafov sme rozšírili a prispôbili požiadavkám, ktoré popisujú správne nakreslený RNA motív. V tejto práci sú podrobne popísane jednotlivé funkcie našej metódy a ich implementácia do nástroja, ktorý automaticky RNA motívy vykresľuje. Tento nástroj sme testovali na rôznych typoch RNA motívov a výsledky sme porovnávali s ich preferovanými ručnými nakresleniami.

Kľúčové slová: RNA motív, single strand (SSE), hélix, relácia, pružiny, vrcholy, asociácie.

Abstract

Author: Bc. Pavol Kmeč
Title: Drawing RNA motifs
University: Comenius University in Bratislava
Faculty: Faculty of Mathematics, Physics and Informatics
Department: Department of Applied Informatics
Supervisor: Mgr. Tomáš Vinař, PhD.
Number of Pages: 64
Year: 2013

The principal goal of this work was to develop a method to automatic draw RNA motifs. We have applied one method from the class of force direct graph drawing. The approach transforms graph using different types of forces. The forces move vertexes to new positions till the graph achieves a steady state. The forces manage rules of graph drawing. For the purposes of graph of RNA motif creation, we have applied one of standard descriptors used in tool RNABob. We have extended and adjusted the force direct method to rules which describe RNA motif drawing properly. Our method was implemented in a new tool to draw RNA motifs automatically. The method comprises of functions that are defined in details in the paper. We have tested the tool on a variety of types of RNA motifs. The results were compared to preferred drawing made manually.

Keywords: RNA motif, single strand (SSE), helix, relation, springs, vertexes, associations.

Predhovor

Táto práca vznikla z potreby zobrazenia sekundárnych štruktúr častí RNA, takzvaných RNA motívov. V súčasnosti existujú rôzne prístupy automatického vykresľovania RNA motívu. Metódy, ktoré sú na to používané riešia iba základné požiadavky správneho nakreslenia motívu, ako sú prekryvanie jednotlivých jeho častí alebo udržiavanie rovnakých vzdialeností medzi nukleotidmi. Cieľom tejto práce bolo vyvinúť novú metódu, ktorá bude spĺňať aj konkrétnejšie požiadavky pre správne nakreslený motív. Základom tejto metódy je metóda silovo riadeného vykresľovania grafov a grafových štruktúr. Táto metóda bola rozšírená a prispôbena požiadavkám vykresľovania RNA motívov.

Obsah

Úvod.....	1
1. Problém kreslenia RNA motívov.....	3
1.1. RNA motívy.....	3
1.2. Požiadavky pre správne vykreslený motív	6
1.3. Prehľad predchádzajúcej práce	8
2. Metódy kreslenia grafov	16
2.1. Kreslenie grafov pomocou pružín.....	16
3. Silovo riadené vykresľovanie RNA motívov.....	21
3.1. Reprézntácia RNA motívu ako graf.....	21
3.2. Metóda vykresľovania grafu.....	23
3.2.1. Funkcia FDn.....	23
3.2.2. Funkcia FDe.....	26
3.2.3. Funkcia FDa.....	29
3.2.4. Automatická rotácia SSE.....	39
3.2.5. Automatická rotácia elementov hélix a relácia.....	40
3.2.6. Zrkadlové otočenie hélixu a relácie	43
3.2.7. Udržiavanie vrcholov typu SSE v kružniciach.....	44
3.2.8. Beh algoritmu.....	52
4. Implementácia a vyhodnotenie	54
4.1. Implementačné detaily	54
4.2. Ladenie konštánt	54
4.3. Príklady vykreslených motívov	55
4.4. Rozšírenia programu	56
Záver	58
Zdroje.....	60
Prílohy.....	62
Používateľské rozhranie programu RNA viewer.....	62
Načítanie a vykreslenie RNA motívu	62
Používateľské nástroje	62
Výstupné súbory	63
CD nosič:	64

Zoznam obrázkov

Obrázok 1.1: Príklad RNA motívu.	3
Obrázok 1.2: Príklad RNA motívu obsahujúci pseudouzol.....	4
Obrázok 1.3: Príklad deskriptoru RNA motívu bulge	5
Obrázok 1.4: Zobrazenie RNA motívu bulge	5
Obrázok 1.5: Príklady možných vykreslení RNA motívu.....	6
Obrázok 1.6: Príklady vykreslenia RNA motívu.....	7
Obrázok 1.7: Príklady vykreslenia RNA motívu.....	8
Obrázok 1.8: Príklady vykreslenia RNA motívu.....	8
Obrázok 1.9: Vytvorenie kruhového grafu.	9
Obrázok 1.10: Vytváranie motívu z kruhovitého grafu.....	10
Obrázok 1.11: Vytvorenie lineárneho grafu.	10
Obrázok 1.12: Lineárny graf RNA motívu.	11
Obrázok 1.13: Príklad vytvorenia stromu RNA motívu [1].	12
Obrázok 1.14: Vytvorenie nakreslenia RNA motívu zo stromu, ktorý tvorí jeho kostru [1].	13
Obrázok 1.15: Porovnanie ručne vykresleného motívu (Obrázok A) a motívu automaticky nakresleného metódou kreslenia pomocou stromov [1] (Obrázok B).	14
Obrázok 1.16: (A) Príklad transformácie lineárneho grafu RNA motívu do planárneho nakreslenia.	15
Obrázok 2.1: Postupné znižovanie energie [15].	19
Obrázok 2.2: Príklad automatického vykreslenia grafu pomocou optimalizácie znižovania energie systému [9].	19
Obrázok 3.1: Štruktúry grafov.	21
Obrázok 3.2: Príklad štruktúry vrcholu SSE.....	22
Obrázok 3.3: Príklad štruktúry vrcholu elementu hélíx alebo relácia.	22
Obrázok 3.4: Príklad kompletnej štruktúry RNA motívu.....	22
Obrázok 3.5: Priebeh funkcie sigmoid $f(k) = \frac{1}{1 + e^{-0.01(k-1)}} - 0.5$	24
Obrázok 3.6: Príklady pôsobenia síl pružín funkcie FNn	25
Obrázok 3.7: Príklad možného vykreslenia grafu iba pomocou síl pružín.....	26
Obrázok 3.8: Príklad nájdenia hĺbky prieniku dvoch vrcholov.	27
Obrázok 3.9: Príklady pôsobenia odpudivých síl funkcie FNe	28

Obrázok 3.10: Príklad nutnosti použitia síl FDA na kontrolu uhlov medzi vrcholmi.....	29
Obrázok 3.11: Uhol alfa funkcie <i>FDA</i>	30
Obrázok 3.12: Príklad správnych uhlov, rozmiestnenia vektorov <i>u1</i> a <i>u2</i> a priradenie bodov A, B a C pre element hélíx a reláciu.	31
Obrázok 3.13: Priebeh funkcie $f(\alpha) = \sin(4 \cdot \text{rad}(\alpha)) $	32
Obrázok 3.14: Priebeh funkcie $f(\alpha) = \left \sin\left(0.5 \cdot \text{rad}(\alpha) - \frac{\pi}{8} \cdot 3\right) \right $	33
Obrázok 3.15: Priebeh funkcie $f(\alpha) = \left \sin\left(0.5 \cdot \text{rad}(\alpha) - \frac{\pi}{8}\right) \right $	33
Obrázok 3.16: Príklad elementu hélíx alebo relácie kedy sú povolené aj vnútorné uhly....	34
Obrázok 3.17: Priebeh funkcie $f(\alpha) = \left \sin\left(0.5 \cdot \text{rad}(\alpha) - \frac{\pi}{2}\right) \right $	35
Obrázok 3.18: Priebeh funkcie $f(\alpha) = \sin(0.5 \cdot \text{rad}(\alpha)) $	35
Obrázok 3.19: Príklad pôsobenia sily v asociačnom bode smerom vektoru <i>n</i> pričom po posune sa vektor <i>u2</i> postupne približuje k najbližšiemu správneuhlu.	36
Obrázok 3.20: Príklady jednotlivých možností umiestnenia vektoru <i>u2</i> medzi správnyhmi uhlami kedy $n_1 = -n/ n $	38
Obrázok 3.21: Príklad pôsobenia sily FDA.	39
Obrázok 3.22: Príklad rotácie SSE.	40
Obrázok 3.23: Príklad párovania asociačných bodov pre rotáciu elementov hélíx alebo relácie.	41
Obrázok 3.24: Rotácia jedného elementu a rotácia celej časti grafu.	42
Obrázok 3.25: Rotácia celých častí prislúchajúcich elementu <i>h2</i>	43
Obrázok 3.26: Príklad nutnosti zrkadlového otočenia elementov hélíx alebo relácie.	44
Obrázok 3.27: Príklad potreby udržiavania elementov SSE v kružniciach.	45
Obrázok 3.28: Hľadaný uhol alfa podľa ktorého sa nájde kružnica k je uhol vektorov <i>u1</i> a <i>u2</i>	46
Obrázok 3.29: Prípady kedy je polomer kružnice rovný polovičnej šírky elementu.....	47
Obrázok 3.30: Umiestenie kružnice <i>k</i>	48
Obrázok 3.31: Príklad nájdenia riadiaceho bodu <i>X</i> a všetkých pozícií znakov predpisu. ...	50
Obrázok 3.32: Príklad posunu pozícií znakov do vnútra vrcholu a prispôsobenie jeho rozmerov predpisu.	51
Obrázok 4.1: Príklad vykreslenia RNA motívov programom RNA viewer.	56

Úvod

Vykresľovanie RNA motívov je v súčasnej dobe čoraz viac potrebnéjšie. Výskumníci ktorý analyzujú RNA štruktúry často potrebujú k svojej práci vizuálnu informáciu o sekundárnych štruktúrach častí RNA takzvaných RNA motívov. V súčasnosti existuje niekoľko aplikácií, ktoré rôznymi spôsobmi napomáhajú kresliť RNA motívy. Metódy vykresľovania v nich sa líšia a sú zväčša vhodné iba na vykresľovanie úzkeho okruhu typovo rovnakých štruktúr. Zároveň je pri niektorých nástrojoch implementujúcich tieto metódy nutné si pred ich používaním naštudovať pomerne rozsiahly manuál, ktorý je pre technicky menej zdatných používateľov zložitý. Vzhľadom na kvalitu výsledku je preto niekedy potrebné kresliť RNA motív ručne.

Cieľom tejto práce je poskytnúť používateľom nástroj, pomocou ktorého je možné kresliť RNA motívy rýchlejšie a intuitívnejšie ako v podobných nástrojoch. S tým je spojený aj návrh novej metódy ich automatického vykresľovania, ktorá tvorí hlavný prínos tejto práce. Keďže RNA motív je možné reprezentovať ako neorientovaný graf s vrcholmi a hranami, rozhodli sme sa aplikovať niektorú z metód kreslenia grafov na kreslenie RNA motívov. V tejto oblasti už existuje široké spektrum rôznych metód na vykresľovanie grafov a grafových štruktúr. V tejto práci sme aplikovali jednu z nich a to metódu založenú na silovo riadenom vykresľovaní grafov. Pri tejto metóde na vrcholy pôsobia rôzne typy síl, podľa ktorých sú postupne posúvané na nové miesta, až kým sa graf nedostane do rovnovážneho stavu. Týmito silami sú kontrolované rôzne typy podmienok, ktoré majú v grafe platiť. Správnym zostavením síl pôsobiacich v grafe RNA motívu je ho potom možné vykresliť tak, aby spĺňal konkrétne požiadavky jeho správneho nakreslenia. Vytvorenie grafu RNA motívu je vykonávané z jedného zo štandardných deskriptorov popisujúceho štruktúru motívu. Ide o typ deskriptoru, ktorý sa používa v programoch na hľadanie motívov, kde je motívy potrebné vykresľovať v rámci používateľských rozhraní.

Samotná práca je rozdelená do štyroch kapitol. Obsahuje 6 grafov, 45 obrázkov a 2 prílohy. Prvá kapitola je venovaná samotnému problému kreslenia motívov, prehľadu predchádzajúcich prác a zadenovaniu konkrétnych cieľov práce. V druhej kapitole sú popísané niektoré zo silovo riadených metód kreslenia grafov, ktoré tvoria základ nami navrhutej metódy popísanej v kapitole 3. Tretia kapitola detailne popisuje novú metódu automatického vykresľovania RNA motívu, ktorá je založená na silovo riadenom vykresľovaní grafov. Sú v nej popísané jednotlivé funkcie, ktoré kontrolujú výpočet síl pôsobiacich na vrcholy a ktoré menia orientáciu jednotlivých častí grafu. V tejto kapitole je

popísaný aj algoritmus zabezpečujúci samotné vykresľovanie grafu RNA motívu použitím týchto funkcií. Záverečná kapitola je venovaná implementácií metódy, vyhodnoteniu výsledkov práce a možným rozšíreniam algoritmu.

1. Problém kreslenia RNA motívov

RNA motívy sú určité úseky sekundárnej štruktúry RNA. Sekundárna štruktúra obsahuje na rozdiel od primárnej štruktúry, ktorá nie je ničím iným ako jednoduchým reťazcom nukleotidov aj spárované úseky komplementárnych nukleotidov. Takáto štruktúra preto musí byť kreslená tak, aby boli závislosti medzi nukleotidmi viditeľné. Problém kreslenia RNA motívov nemá ustálené zadanie, pretože kreslenie ich štruktúry nie je jednoznačné, aj keď niektoré typy obrázkov sú preferované. Je možné iba vymedziť pravidlá, ktoré vychádzajú zo zvyklostí a štýlov doposiaľ používaných pri ich kreslení.

V súčasnosti existujú deskriptory, ktoré sa dajú používať na hľadanie motívov a tie potrebujeme kresliť v rámci používateľských rozhraní [6][12][5]. Preto je potrebné hľadať automatické metódy, ktoré by takéto nakreslenia RNA motívov dokázali vyrobiť. Našou úlohou je nájsť algoritmus pre kreslenie takýchto motívov.

Hlavným kritériom dobre vykresleného RNA motívu je prehľadnosť. Od toho sa časom postupne vyvinuli aj požiadavky, ktoré bližšie popisujú správne vykreslený RNA motív. Problém kreslenia RNA motívov preto priamo vychádza z týchto požiadaviek (viď. Sekcia 1.2).

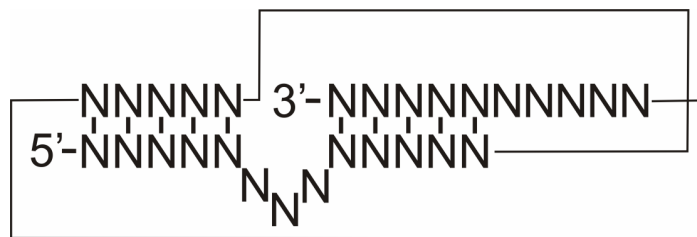
1.1. RNA motívy

Z informatického hľadiska ide o dlhý reťazec nad abecedou {A,C,G,U}. Na rozlíšenie koncov vlákna používame označenie 5' a 3', pričom sa RNA kóduje v poradí od 5' do 3'. V sekundárnej štruktúre RNA majú nukleotidy tendenciu tvoriť väzby so svojimi komplementami. A je komplementom U, C je komplementom G. RNA motív je časť takejto sekundárnej štruktúry RNA.



Obrázok 1.1: Príklad RNA motívu.

RNA motívy môžu svojím párovaním nadobúdať aj tvary takzvané pseudouzly alebo slučky, ktoré nie je možné vykresliť tak, aby boli všetky spárované nukleotidy od seba rovnako vzdialené a súčasne aby sa neprekrývali.



Obrázok 1.2: Príklad RNA motívu obsahujúci pseudouzol.

RNA motívy môžeme popísať napríklad deskriptorom. Deskriptor používajú programy na vyhľadávanie motívov v databázach sekvencií, ako napr. RNAMot [6], RNAMotif [12] alebo RNABob [5].

V tejto práci používame deskriptory programu RNABob. Tento typ deskriptora rozdeľuje sekvenciu na tri typy elementov: nespárované úseky sú označované ako *Single strand element* (SSE); spárované úseky sú označované ako *hélix* alebo *relácia*. Relácia sa od hélixu líši iba tým, že striktnejšie popisuje štruktúru spárovaných úsekov. Samotný deskriptor je zložený z dvoch hlavných častí. Prvá časť deskriptora (mapa) popisuje poradie jednotlivých elementov v motíve. Je to reťazec mien elementov oddelených medzerou, v poradí v ktorom kódujú motív, teda od 5' do 3'. Keďže hélix a relácia majú komplementárne vlákno, musia sa v mape vyskytovať dva krát. Druhý výskyt týchto elementov je označený na konci mena apostroфом. To znamená, že kódovanie pokračuje komplementárnym vláknom hélixu alebo relácie. Príkladom mapy RNA motívu je reťazec *h1 s1 h2 s2 h2' h1'*, pričom h1 a h2 reprezentujú hélixu a s1 a s2 reprezentujú dva SSE.

Druhá časť deskriptora špecifikuje jednotlivé elementy, každý na samostatnom riadku. Riadok začína menom elementu, nasleduje číslo, ktoré vyjadruje počet povolených nezhôd oproti predpisu. Za týmto číslom nasleduje reťazec (predpis) jednotlivých nukleotidov kódujúcich element. Tento reťazec obsahuje znaky A, U, C, G, ale aj niektoré špeciálne znaky, ktoré napomáhajú v situáciách, kedy nie je možné jednoznačne určiť aký nukleotid je na danom mieste. Napríklad znak *N* označuje, že sa na danom mieste môže nachádzať ľubovoľný nukleotid. Tieto dohodnuté znaky sú spísane v dokumente IUPAC kódu (International Union of Pure and Applied Chemistry). Okrem týchto znakov sa používa aj špeciálny znak ***, ktorý vyjadruje, že na danom mieste je znak *N* alebo nič. Teda reťazec **G** môže interpretovať predpis *NGN*, *NG*, *G*, *GN*. Znak *** nám teda umožňuje kódovať elementy variabilnej dĺžky. Príkladom jedného popisu elementu SSE je *s2 0 AUYA*, čo znamená, že elementu s2 je povolených nula nezhôd oproti predpisu a jeho predpis je reťazec AUYA.

Popis hélixu v druhej časti deskriptora je o niečo zložitejší, pretože obsahuje aj komplementárne vlákno. Namiesto jedného čísla, vyjadrujúceho počet povolených nezhôd oproti predpisu, sú tu dve čísla oddelené dvojbodkou: prvé číslo pre vlákno bližšie k začiatku kódovania teda k 5' a druhé číslo pre jeho komplement. Rovnako je to aj s nasledujúcim predpisom kódujúcim hélix. Tvoria ho dve samostatné reťazce oddelené dvojbodkou, vzhľadom na ktorú sú väzby v hélixe tvorené zrkadlovo. Príkladom popisu hélixu je *h1 0:0 NNNN:NNNN*, čo znamená, že hélixu h1 je povolených nula nezhôd v prvom predpise pred dvojbodkou a taktiež nula nezhôd v druhom predpise za dvojbodkou. Nasleduje samotný predpis a jeho komplementárne vlákno oddelené dvojbodkou.

Popis relácie je takmer rovnaký ako popis hélixu s rozdielom, že relácia má navyše pripojený ešte reťazec označujúci obmedzenia platiace v danej relácii. Ide o takzvanú transformačnú maticu napríklad TGYR, ktorá povoľuje G-U páry (v hélixoch sú tieto páry automaticky povolené). Príkladom popisu relácie je *r2 0:0 NNNN:NNNN TGCA*, čo znamená, že relácia r2 má taktiež ako pri hélixe povolených nula nezhôd pre prvý predpis a nula nezhôd pre jeho komplement. Taktiež nasleduje predpis relácie a jeho komplementárne vlákno oddelené dvojbodkou. Posledný reťazec transformačná matica TGCA povoľuje páry T-G a C-A v predpise.

Príklad deskriptora popisuje Obrázok 1.3 a jeho prislúchajúci RNA motív Obrázok 1.4.

```
#bulge.des
h1 s1 h2 s2 h2' s3 h1'

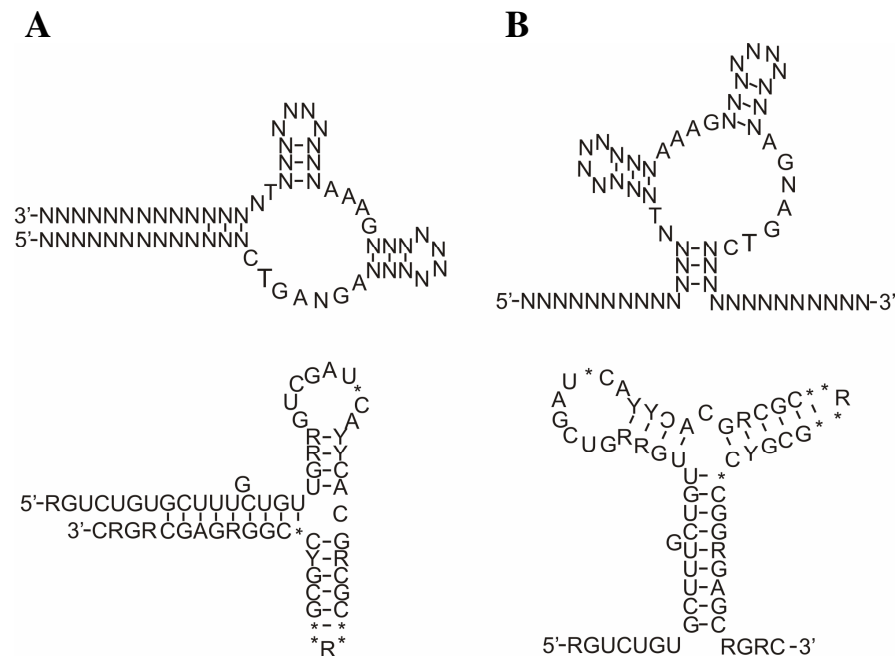
h1 0:0 NNN:NNN
s1 0 N**
h2 0:0 NNN:NNN
s2 0 GNRA
s3 0 N**
```

Obrázok 1.3: Príklad deskriptoru RNA motívu bulge



Obrázok 1.4: Zobrazenie RNA motívu bulge

Ako vidíme na Obrázok 1.3 grafická reprezentácia deskriptora je podstatne prehľadnejšia ako samotný textový zápis deskriptora. Deskriptor samotný však neobsahuje dostatok informácií na to aby príslušný motív bolo možné vykresliť jednoznačným spôsobom (viď. Obrázok 1.5).



Obrázok 1.5: Príklady možných vykreslení RNA motívu. (A) Preferovanejšie zobrazenia deskriptora. (B) Iný možný nákres motívu z rovnakého deskriptora.

Niektoré nakreslenia sú však prehľadnejšie ako iné, preferovanejšie tvorcami deskriptora. Našou úlohou je nájsť metódu vykresľovania motívu tak aby boli dodržané takéto preferencie.

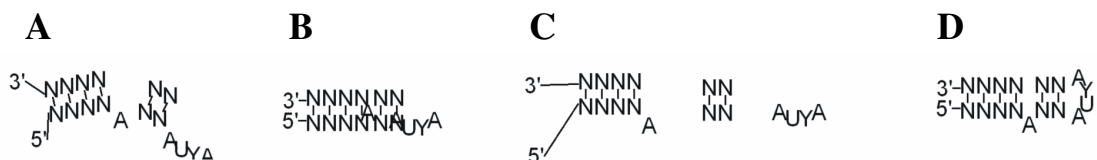
1.2. Požiadavky pre správne vykreslený motív

Správne vykreslený motív musí spĺňať jednu hlavnú podmienku, musí byť prehľadný. Taktiež musí byť z jeho nakreslenia zrejmé párovanie nukleotidov a smer kódovania. Zle nakreslený obrázok je potom taký, z ktorého nie je na prvý pohľad jasné, ktoré nukleotidy sú spárované, alebo ktoré elementy sú navzájom asociované (viď. Obrázok 1.6 A). Z toho vyplýva podmienka, že spárované nukleotidy musia mať medzi sebou rovnaké vzdialenosti. Príklady nakreslenia, ktoré porušujú túto podmienku znázorňuje Obrázok 1.6 A a Obrázok 1.6 C. Táto podmienka platí aj v prípade, keď sú vzdialenosti medzi nukleotidmi väčšie, no párovanie vyplýva zo samotnej štruktúry. (viď. Obrázok 1.7 B, Obrázok 1.8 B). Čitateľ potom nemusí na to aby pochopil obrázku sledovať postupné kódovanie pozdĺž motívu. Ďalšou chybou v nakreslenom motíve sú prekrywajúce sa

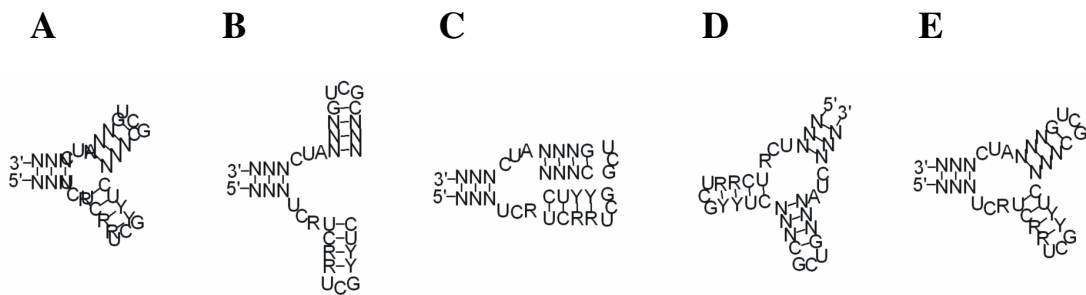
nukleotidy. Táto chyba značne narúša čitateľnosť štruktúry motívu (vid'. Obrázok 1.6 B, Obrázok 1.7 A). Preto musia byť elementy motívu orientované tak aby nedochádzalo k ich prekryvaniu.

Samostatné časti motívu, ktoré sú odlišne orientované nadobúdajú podľa možnosti buď horizontálny alebo vertikálny smer kódovania (vid'. Obrázok 1.8 C). Výnimkou je situácia, kedy by táto podmienka zapríčinila prekryvanie nukleotidov alebo nedodržanie rovnakých vzdialeností medzi nukleotidmi (vid'. Obrázok 1.7 B, Obrázok 1.8 B). Vtedy je možné orientovať časti motívu aj ináč, a to tak, aby boli uhly medzi nimi pravidelné v celom motíve pričom v horizontálnych alebo vertikálnych polohách ostávajú časti bližšie k začiatku kódovania teda k 5' (vid'. Obrázok 1.7 E). Týmto je zabezpečená celková symetrickosť nakreslenej štruktúry RNA motívu. Príklad nesprávnej orientácie častí motívu znázorňuje Obrázok 1.7 D a Obrázok 1.8 A. Príklad kedy sú síce správne orientované časti motívu, no hélix alebo relácia najbližšia k začiatku kódovania nie je v horizontálnej polohe znázorňuje Obrázok 1.7 C. Tento motív je taktiež nesprávne nakreslený.

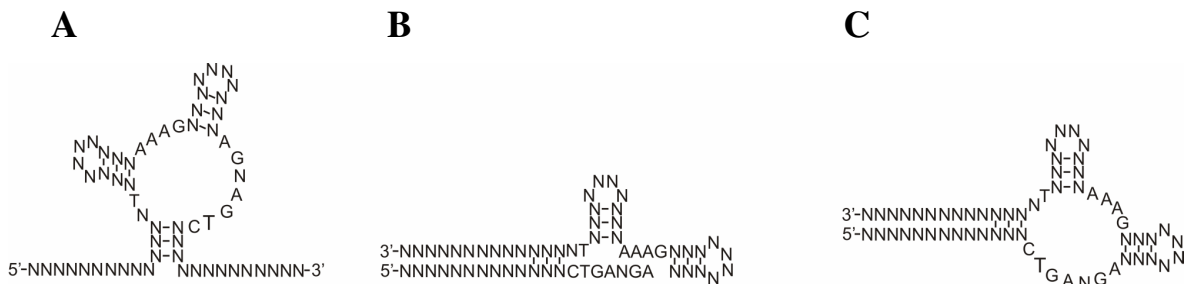
Variabilitu pri zmenách orientácie častí motívu je možné dosiahnuť ak sú nespárované úseky SSE udržiavané v kružniciach. Ide o prípady, kedy sú SSE umiestnené medzi hélixami alebo reláciami. Do týchto kružníc je potom možné zasadzovať hélixu alebo reláciu v preferovaných orientáciách, ktoré udávajú orientácie celých častí, ktorých sú začiatkami.(vid'. Obrázok 1.7 E, Obrázok 1.8 C). SSE ostávajú rovné v prípadoch, kedy nemajú z obidvoch strán susedov helixy alebo relácie (vid'. Obrázok 1.8 C).



Obrázok 1.6: Príklady vykreslenia RNA motívu.
 (A,B,C) Nesprávne vykreslené motívy. (D) Správne vykreslený motív.



Obrázok 1.7: Príklady vykreslenia RNA motívu.
 (A,B,C,D) Nesprávne vykreslené motívy. (E) Správne vykreslený motív.



Obrázok 1.8: Príklady vykreslenia RNA motívu.
 (A,B) Nesprávne vykreslené motívy. (C) Správne vykreslený motív.

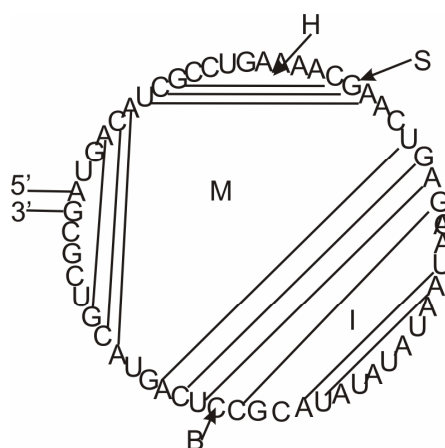
Motív nakreslení podľa týchto podmienok je estetický a zároveň je vidieť jeho kompletná štruktúra a vzťahy medzi nukleotidmi. V tejto práci je pri navrhovaní metódy pre automatické kreslenie RNA motívov kladený dôraz na dodržovanie práve týchto podmienok.

1.3. Prehľad predchádzajúcej práce

V súčasnosti existuje niekoľko nástrojov, ktoré napomáhajú vykresľovať RNA motívy. Pomocou nich je možné kresliť motívy tak, aby boli dodržané niektoré z požiadaviek správne vykresleného RNA motívu. Niektoré tieto metódy majú pomenované úseky motívu, ktoré sú špecifické svojou polohou v štruktúre. Sú to: S (Stem) označuje spárované úseky; H (hairpin) označuje úseky kde SSE na jednej zo strán uzatvára hélix alebo reláciu; B (bulge) je výstupok nespárovanej časti medzi spárovanými; I (interior loop) označuje časť motívu, kde nespárované úseky spájajú dva spárované úseky, ak je takto spájaných viacero spárovaných úsekov, je vrchol označený ako M(multi branch loop).

Jednou z metód na vykresľovanie RNA motívov je metóda vykresľovania pomocou kruhového grafu. Touto metódou je možné kresliť štruktúrou náročné a počtom elementov rozsiahle motívy pričom, sa zameriava výlučne na podmienku neprekrývania sa nukleotidov. Metóda je navrhnutá iba pre motívami, ktoré neobsahujú pseudouzly. Podstatou vykresľovania je, že sa najprv vytvorí z motívu reťazec jeho primárnej štruktúry.

Tento reťazec je vykreslený do tvaru kruhu, pričom je označený začiatok a koniec kódovania štandardnými znakmi 5' a koniec 3'. Následne sú vyznačené všetky párovania z jeho sekundárnej štruktúry a to vedením čiar medzi dvomi spárovanými nukleotidmi. Na takto vykreslenej kružnici je možné identifikovať časti RNA motívu (viď. Obrázok 1.9): znak S vyznačuje úseky, ktoré sú spojené čiarami, teda spárované nukleotidy; H vyznačuje časti kedy majú nespárované nukleotidy (tie z ktorých nevedie čiara) na svojom začiatku spárovaný nukleotid a konci jeho komplement, tie môžu neskôr vo výslednom nakreslení vytvárať kruhové tvary; B vyznačuje časti, kedy majú nespárované nukleotidy na svojom začiatku spárovaný nukleotid a konci na rozdiel od H taký spárovaný nukleotid, ktorý nie je jeho komplementom; I vyznačuje úsek medzi dvomi časťami S, kde sa nachádzajú po obidvoch stranách nespárované nukleotidy; M vyznačuje úsek, kde sa takto stretávajú viac ako dva úseky typu S. Týmto je vytvorený neorientovaný kruhovitý graf, z ktorého je možné vzhľadom na vyznačené časti vytvoriť RNA motív. Vytváranie motívu prebieha postupným približovaním pozdĺž vyznačených čiar a následným spájaním úsekov typu S. Najskôr sa k sebe priblížia pozdĺž svojich prislúchajúcich čiar všetky so svojimi komplementmi spárované úseky S. Všetky nespárované nukleotidy si musia stále udržať medzi sebou rovnaké vzdialenosti a preto sú ich konce posúvané spolu so spárovanými úsekmi. Časti motívu typu H a I môžu potom vytvárať kružnice. Následne je nutné upraviť všetky časti typu M (viď. Obrázok 1.10).

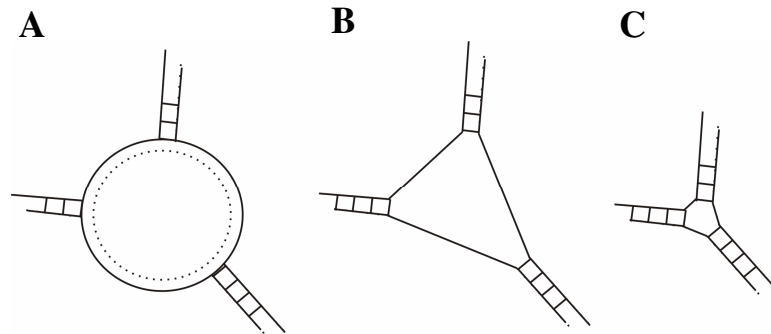


Obrázok 1.9: Vytvorenie kruhového grafu.

S (stem), H (hairpin), B (bulge), I (interior loop), M (multi branch loop). 5' a 3' označujú začiatok a koniec kódovania.

Nájde sa najkrajšia úsečka medzi koncami susedných spárovaných úsekov. Na týchto úsečkách sa nachádzajú aj zvyšné nespárované nukleotidy, ktoré sú v kruhu umiestené medzi koncami úsečky. Po skrátení vzdialeností na minimálnu hodnotu, ktorá je možná

aby sa nukleotidy neprekrývali, je RNA vo výslednom tvare. Ak sú spájané tri časti S, výsledný tvar je trojuholníkový.

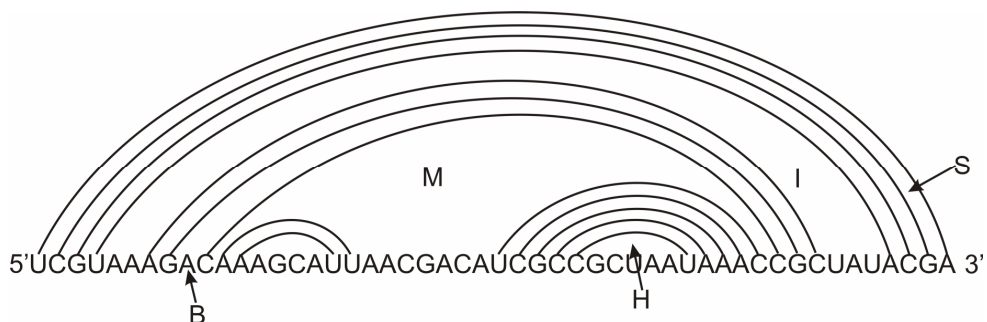


Obrázok 1.10: Vytváranie motívu z kruhového grafu.

(A) Priblíženie spárovaných úsekov S. Na obvode kruhu a na koncoch výbežkov sa nachádzajú všetky ostatné nespárované nukleotidy. (B, C) Úprava M do RNA motívu.

V prípade, že sa v jednej kružnici nachádza viacero častí M je každá riešená samostatne. Následne sa rekurzívne znovu v kruhu hľadajú a spájajú úseky M [13].

Veľmi podobná metóda je vykresľovanie RNA motívu pomocou lineárneho grafu. Na rozdiel od kreslenia motívu pomocou kruhového grafu je v tejto metóde vytvorená z primárnej štruktúry priamka. Použitie je rovnaké ako pri metóde vykresľovania pomocou kruhového grafu, pričom taktiež motívy kreslené touto metódou neobsahujú pseudouzly. Zostrojenie lineárneho grafu prebieha tak, že nukleotidy primárnej štruktúry motívu sú zoradené vedľa seba v jednej rovine, pričom je začiatok aj koniec označený štandardne 5' a 3'. Následne sú vyznačené všetky párovania zo sekundárnej štruktúry vedením čiary z nukleotidu k jeho komplementu. V takomto grafe je taktiež možné vyznačiť časti S, H, B, I a M (viď. Obrázok 1.11).



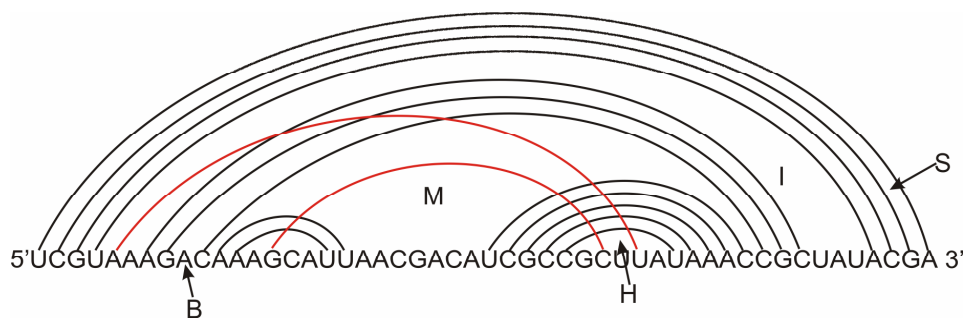
Obrázok 1.11: Vytvorenie lineárneho grafu.

S (stem), H (hairpin), B (bulge), I (interior loop), M (multi branch loop). 5' a 3' označujú začiatok a koniec kódovania.

Samotné vykreslenie RNA motívu z lineárneho grafu je v podstate rovnaké ako v kruhovom grafe. Postupne sa približujú nukleotidy v závislosti od typu úseku. Metódy

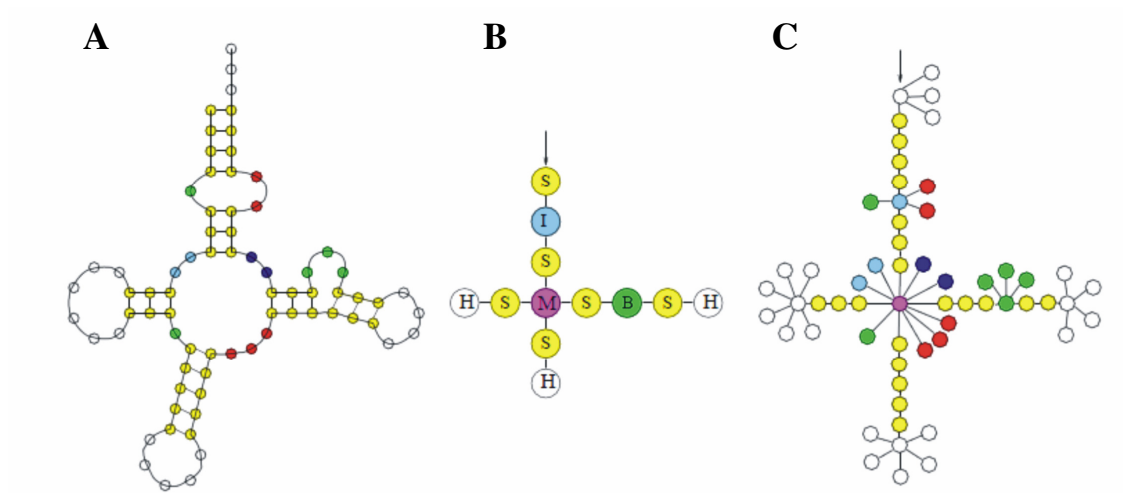
vykresľovania RNA motívu pomocou kruhového a lineárneho grafu sú implementované v nástroji jViz.RNA [7]. Rovnako obidva metódy používa nástroj na kreslenie RNA motívov VARNA [3].

Ďalšou z metód je vykresľovanie pomocou stromov. Táto metóda najskôr z motívu vytvorí strom, ktorého nakreslenie bude neskôr tvoriť kostru nakreslenia motívu. Každý z vrcholov takéhoto stromu môže byť buď typu nespárovaného nukleotidu alebo typu S, H, B, I alebo M. Každý RNA motív, ktorý neobsahuje pseudouzly možno reprezentovať takýmto stromom. Pri konštruovaní stromu je využívaný lineárny graf na to, aby sa zo štruktúry RNA odstránili všetky pseudouzly a aby boli jednotlivým úsekom motívu pripradené typy, ktoré neskôr budú reprezentovať vrcholy v strome. Najprv je vytvorený rovnako ako pri metóde kreslenia pomocou lineárneho grafu graf RNA motívu. Prekrižujúce sa hrany sú pseudouzly. Odstránené sú tie hrany, ktoré sa krížia s najväčším počtom hrán (viď. Obrázok 1.12). Grafu, ktorý už neobsahuje pseudouzly sú vyznačené miesta S, H, B, I a M rovnako ako v predchádzajúcej metóde.



Obrázok 1.12: Lineárny graf RNA motívu.
Červenou farbou sú vyznačené párovania, ktoré vytvárajú pseudouzly. Tie sú z grafu odstránené.

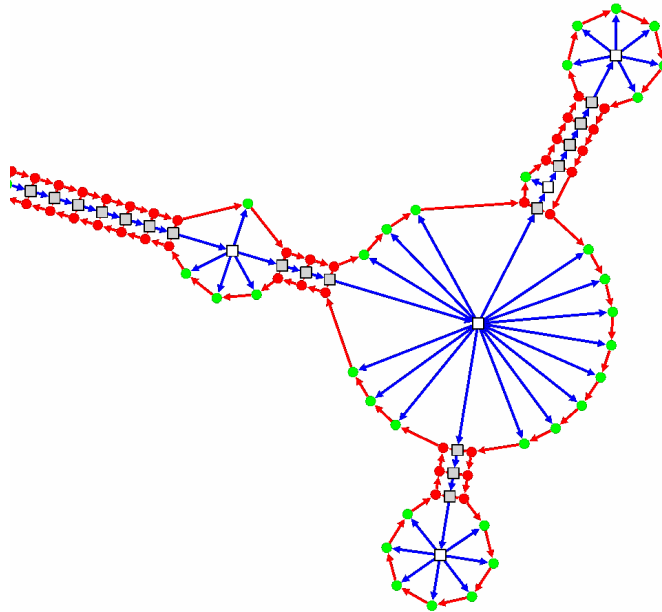
Každý nespárovaný nukleotid tvorí samostatný vrchol v strome, ktorý je spojený s prislúchajúcim úsekom H, I, B alebo M. Každé dva spárované nukleotidy tvoria jeden vrchol typu S. H má jeden centrálny vrchol, na ktorý sú pripojené všetky jeho nespárované nukleotidy a prislúchajúci vrchol S. Všetky nespárované nukleotidy, ktoré sú označené ako I sú spojené s jedným centrálnym vrcholom, ktorý ich spája s prislúchajúcimi vrcholmi S z oboch strán. Rovnako je to aj s typom B. M má taktiež centrálny vrchol a spája prislúchajúce nespárované nukleotidy a jemu prislúchajúce vrcholy S (viď. Obrázok 1.13).



Obrázok 1.13: Príklad vytvorenia stromu RNA motívu [1].

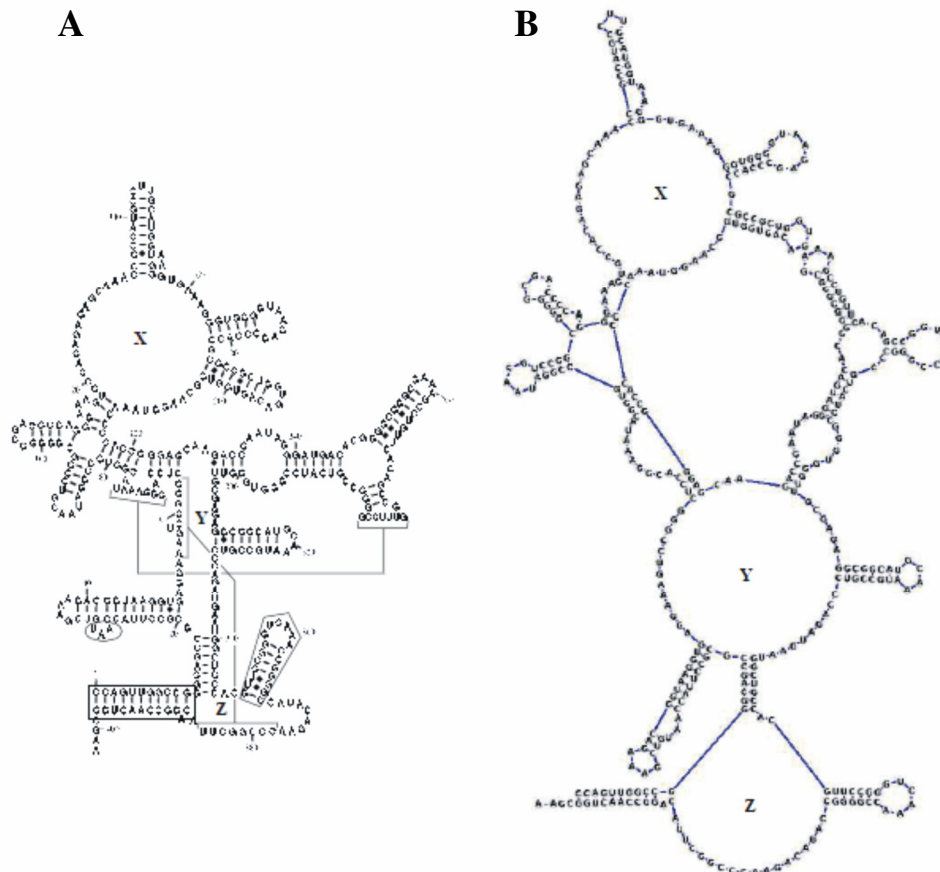
(A) RNA motív, ktorý neobsahuje pseudouzol a je možné z neho vytvoriť strom. (B) Príklad povoleného spájania jednotlivých typov vrcholov. (C) Strom motívu z obrázku A ktorý je vytvorený podľa povolených spájání typov vrcholov.

Vykreslenie takto vytvoreného stromu prebieha v dvoch etapách: najprv je nutné v centrálnych vrcholoch úsekov I, B, H a M (viď. Obrázok 1.14 biely štvorček) nájsť polomery kružníc, podľa ktorých budú následne vykresľované prislúchajúce nukleotidy a v prípade M aj prislúchajúce spárované úseky. Polomer je postupne hľadaný v najkrajnejších centrálnych vrcholoch. Kde sa hľadá taký, aby bolo možné vykresliť všetky elementy tak, aby sa neprekrývali a aby mali medzi sebou rovnaké vzdialenosti. Priestor ktorý týmto zaberú je ohraničený; následne sa rekurzívne vyberajú centrálny vrcholy, ktoré sú hlbšie v strome, pričom ich polomer závisí aj toho, či sa neprekrývajú ohraničené časti už vyrátaných polomerov; potom sa začne postupne vykresľovať motív podľa stromu zo smeru začiatku kódovania (viď. Obrázok 1.14 modré šípky) postupne po centrálnych vrcholoch a vrcholoch typu S. Ak ide o vrchol S sú jeho prislúšné nukleotidy vykreslené kolmo na smer kreslenia (viď. Obrázok 1.14 modrá šípka) ak ide o centrálny vrcholy I, B, H a M kreslia sa im prislúchajúce nukleotidy vo vzdialenosti vyrátaného polomeru od centrálného vrcholu, pričom vzdialenosti medzi nimi musia byť rovnaké ako v pri spárovaných nukleotidoch. Po prejdení celého stromu je RNA motív vo finálnom tvare (viď. Obrázok 1.14).



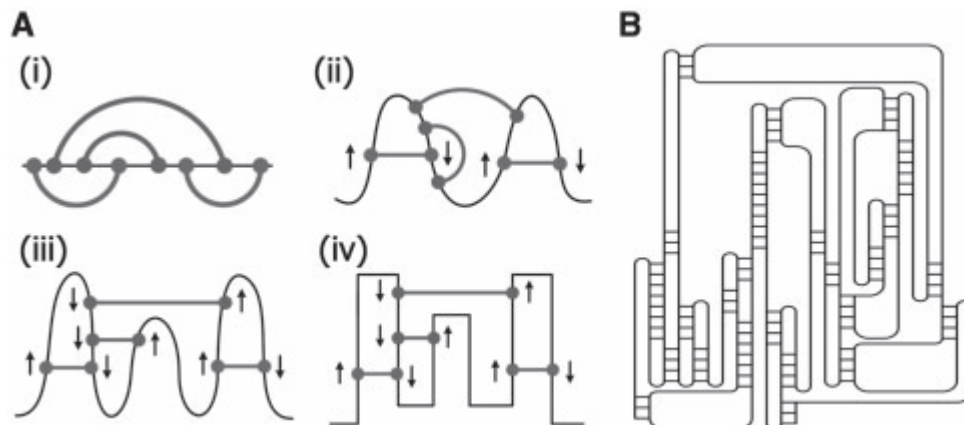
Obrázok 1.14: Vytvorenie nakreslenia RNA motívu zo stromu, ktorý tvorí jeho kostru [1]. Modré šípky vyznačujú strom motívu a smer postupného vykreslenia a červené šípky vyznačujú nakreslenie RNA motívu a smer kódovania.

Táto metóda je zameraná na dodržovanie podmienky aby nedochádzalo k prekryvaniu nukleotidov a aby boli nespárované úseky udržiavané v kružniciach. Zároveň je kladený dôraz na to, aby boli vzdialenosti medzi nukleotidmi rovnaké alebo predĺžené vzdialenosti čím kratšie [1].



Obrázok 1.15: Porovnanie ručne vykresleného motívu (Obrázok A) a motívu automaticky nakresleného metódou kreslenia pomocou stromov [1] (Obrázok B).

Všetky spomínané metódy kreslili RNA motívy, ktoré neobsahujú pseudouzly. Existuje metóda, ktorá sa zaoberá výlučne vykresľovaním motívov so pseudouzlymi. Implementuje ju nástroj PseudoViewer3 [2]. Táto metóda sa zaoberá vykresľovaním RNA motívov, ktoré obsahujú pseudouzly tak, aby vzniklo jeho planárne nakreslenie, kde sa nebudú prekrývať nukleotidy ani krížiť hrany vyznačujúce pseudouzly. Metóda ako vstupný parameter využíva lineárny graf RNA motívu, kde sú hrany spôsobujúce pseudouzly umiestnené na dolnej strane grafu (viď. Obrázok 1.16 A(i)). Postupným usporiadaním najprv hrán na hornej strane a následne usporiadaním na dolnej strane sú hrany skracované na najmenšiu možnú dĺžku tak, aby nedošlo k prekrývaniu (viď. Obrázok 1.16 A). Výsledkom je planárne zobrazenie schémy štruktúry RNA motívu.



Obrázok 1.16: (A) Príklad transformácie lineárneho grafu RNA motívu do planárneho nakreslenia. (i) Reprézantácia RNA motívu lineárnym grafom. (ii) Usporiadanie párování v hornej časti grafu. (iii) Usporiadanie párování v dolnej časti grafu. (iv) Výsledné zobrazenie planárneho nakreslenia. (B) Príklad štruktúry RNA ktorá je vykreslená pomocou transformácie A [2].

Je možné povedať, že aktuálne neexistuje metóda, ktorá by dokázala splniť všetky základné podmienky správneho nakreslenia motívu. Metódy popísané v tejto kapitole sa zameriavali iba na niektoré z nich, pričom RNA motívy, ktoré obsahovali pseudouzly neuvážovali. Metóda implementovaná v nástroji PseudoViewer3 síce riešila práve RNA motívy obsahujúce pseudouzly, no zameriavala sa výlučne iba na podmienku neprekrývania sa častí motívu. Táto práca popisuje novú metódu pomocou ktorej je možné kontrolovať aj podmienky ktoré predchádzajúce metódy zanedbávali.

2. Metódy kreslenia grafov

RNA motív sa dá reprezentovať ako graf, pričom jednotlivé elementy štruktúry (napr. SSE, helixy a pod.) sú vrcholy tohto grafu a hranami sú prepojenia týchto elementov. Z tohto hľadiska kreslenie RNA motívov vlastne spadá do bohato študovanej oblasti kreslenia grafov. Štandardnou požiadavkou je, že sa vrcholy neprekrývajú, ďalšie požiadavky je takisto možné formulovať. V tejto kapitole uvedieme prehľad niektorých metod kreslenia grafov, ktoré môžu byť užitočné pri kreslení RNA.

2.1. Kreslenie grafov pomocou pružín

Kreslenie grafov je problém ktorého, vstupom je graf $G(V, E)$, kde V je množina všetkých vrcholov, E je množina všetkých hrán v grafe G . Pričom hrany sú spojnice medzi vrcholmi a sú reprezentované čiarami alebo krivkami. Medzi vrcholmi v_i a v_j je hrana ak platí $(v_i, v_j) \in E$. Existujú rôzne štýly kreslenia pre rôzne typy grafov. RNA motív obsahujúci aj pseudouzly je možné reprezentovať ako najzákladnejší typ grafu a to neorientovaný graf. Na vykresľovanie neorientovaných grafov sa často používajú metódy, ktoré pomocou síl pôsobiacich na vrcholy kontrolujú ich vkresľovanie. V tejto práci sa zameriavame práve na silovo riadené metódy, ktoré tvoria základ aj našich metód na kreslenie RNA štruktúr popísaných v kapitole 3.

Vykresľovanie grafov pomocou silovo riadených metód sa používa hlavne na prezentáciu zložitých grafov, kde ich tvar je nositeľom informácie o ich štruktúre. Tvar grafu preto musí byť pravidelný a symetrický vzhľadom na vzdialenosti medzi vrcholmi, ktoré sú spojené hranami. Podstatou týchto metód je analógia z fyzickými zariadeniami, kde pružiny s istými vlastnosťami pôsobia svojimi príťažlivými alebo odpudivými silami na príľahlé vrcholy. Vrcholy sú potom posúvané tak, že znižujú energiu teda silu pružín, ktoré na ne pôsobia. Energia celého systému je súčet síl všetkých pružín. Výsledné nakreslenie grafu je v prípade, že sa takýto systém pružín a vrcholov dostane do stavu kedy už nie je možné ďalej znižovať jeho celkovú energiu. Tento stav nazývame rovnovážny stav grafu.

Základnou črtou všetkých silovo riadených metód pre vykresľovanie grafov je výpočet síl, ktoré pôsobia na jednotlivé vrcholy a následná postupná transformácia pozícií vrcholov tak, by sa tieto sily znižovali. Jednou zo základných metód silovo riadeného vykresľovania

grafov je metóda, kedy sa pružiny vyskytujú na miestach, kde sú vrcholy spojené hranou. Množina všetkých pružín v grafe je P kde $|P|=|E|$ a súčasne pre každé $E_i \in E$ existuje $P_i \in P$. Výpočet sily pružiny P_{ij} ktorá pôsobí na jej priľahlé vrcholy v_i a v_j je

$$P_{ij} = k_{ij} \left(|v_i - v_j| - l_{ij} \right)^2$$

k_{ij} je konštanta ktorá určuje tuhosť pružiny P_{ij} a l_{ij} je optimálna vzdialenosť medzi vrcholmi v_i a v_j . Energia celého systému je

$$E = \sum_{i=1}^{|P|} P_i$$

Táto energia je postupne znižovaná, tým že sa rieši parciálna diferenciálna rovnica pre každý z vrcholov, ktorá nájde jeho novú pozíciu. Pričom v každej iterácii, kedy je posunutý vrchol, je nutné nanovo vypočítať silu pružín, ktoré na neho pôsobia. Graf je vo výslednom nakreslení ak už nie je možné posúvaním vrcholov ďalej znižovať energiu E [10].

Iný prístup ako znižovať energiu systému je použitím optimalizačného algoritmu simulované žihanie [11]. Je to všeobecná optimalizačná technika, ktorá je ale výpočtovo „drahá“. Ide o princíp, kedy je postupne vylepšovaný aktuálny stav, kým sa nedosiahne minimálna alebo maximálna požadovaná hodnota výsledku. Problém kreslenia grafu je preto upravený na problém minimalizácie energie systému. Simulované žihanie potrebuje niekoľko parametrov, ktoré budú riadiť postupné znižovanie energie: prvý parameter je celková energia systému E ; druhým parametrom je rýchlosť znižovania energie; posledný parameter je kritériu ukončenia algoritmu. Parameter rýchlosť znižovania energie sa používa v prípadoch, kedy nie je vrcholu vypočítaná pozícia parciálnou diferenciálnou rovnicou. V takomto prípade je vrcholu vypočítaná sila a smer posunu súčtom všetkých síl pružín ktoré na neho pôsobia. Rýchlosť znižovania energie potom určuje o akú najväčšiu vzdialenosť sa má vrchol posunúť daným smerom. Ak je sila pôsobiaca na vrchol menšia ako tento parameter, vrchol sa posunie o vzdialenosť rovnajúcu sa veľkosti tejto sile. Kritériu ukončenia môže byť minimálna požadovaná hodnota energie systému, kedy po jej dosiahnutí je algoritmus zastavený. Tento parameter môže vyjadrovať aj maximálny počet iterácií, kedy už hodnota E neklesá. Po dosiahnutí tohto počtu je algoritmus zastavený. Počas behu algoritmu, kedy sú vrcholy posúvané, je nutné po každom posunutí vrcholu nanovo prepočítať sily pružín, ktoré na neho pôsobia. Použitie algoritmu simulované žihanie na znižovanie energie systému je síce z hľadiska výsledného nakreslenia vhodné no z hľadiska časovej zložitosti veľmi neefektívne. Preto sa tento algoritmus nepoužíva pri

interaktívnom vykresľovaní a pri vykresľovaní grafov, ktoré majú veľký počet vrcholov [4].

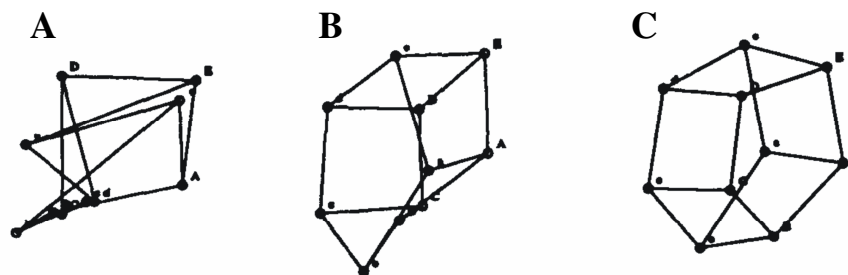
Metódy, ktoré vykresľujú grafy pomocou kontrolovania síl pôsobiacich na vrcholy prakticky iba optimalizujú jednotlivé časti spomínanej základnej metódy, podľa toho, na aký typ grafov budú aplikované a aké podmienky majú nakreslenia dodržiavať. Jednou z týchto metód je aj metóda optimalizujúca výpočet síl pružín a zároveň pripája na vrchol nový typ pružiny, ktorá bude pôsobiť výlučne odpudivými silami na všetky ostatné vrcholy v grafe. Vychádza to z požiadavky ktorá si kladie za úlohu, udržiavať rovnaké vzdialenosti aj medzi vrcholmi nespojenými hranou. Keďže zároveň ide o grafy s malým počtom vrcholov, na znižovanie energie grafu je použitý algoritmus simulované žihanie. Pre výpočet sily pružín sú definované dve funkcie, f_a ako funkcia počítajúca príťažlivé sily a f_r ako funkcia počítajúca odpudivé sily. Prvý typ pružiny vyskytujúci sa v grafe je odpudivá pružina. Každý vrchol je ňou spojený so všetkými ostatnými vrcholmi. Teda na jeden vrchol pôsobí sila $|V|-1$ odpudivých pružín. Funkcia ktorá počíta hodnotu tejto sily je f_r a je definovaná ako

$$f_r(d) = -\frac{k^2}{d}$$

Pričom d je aktuálna vzdialenosť medzi vrcholmi a k je požadovaná vzdialenosť, ktorá má byť udržiavaná medzi vrcholmi. Druhým typom je pružina pôsobiaca príťažlivými silami, ak sú od seba vrcholy príliš vzdialené a odpudivými silami, ak sú vrcholy príliš blízko pri sebe. Táto pružina pôsobí na vrcholy, ktoré sú spojené hranou. Výsledná sila tejto pružiny je súčet funkcií $f_r + f_a$ kde

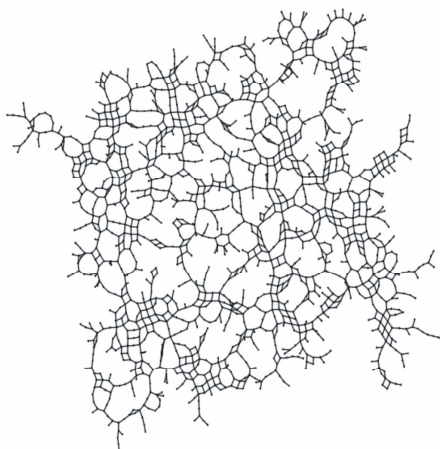
$$f_a(d) = \frac{d^2}{k}$$

Všetky pružiny, ktoré na vrchol pôsobia sú vlastne smerové vektory vychádzajúce z o stredu vrcholu smerom k príľahlým vrcholom. Ich dĺžka je veľkosť sily prislúchajúcej pružiny ktorú reprezentujú. Výsledná sila a smer pôsobenia je vektorovým súčtom všetkých takýchto vektorov [15]. Postupné znižovanie energie znázorňuje Obrázok 2.1.



Obrázok 2.1: Postupné znižovanie energie [15].
 (A) počiatočný stav. (B) Po desiatich iteráciách. (C) Finálne vykreslenie po skončení algoritmu.

Jednou z metód, ktorá sa zaoberá vykresľovaním grafov majúcich veľký počet vrcholov je metóda, ktorá optimalizuje algoritmus znižovania energie. Keďže minimalizácia energie grafu je v silovo riadených metódach časťou čo sa najviac podieľa na ich zložitosti, optimalizácia tohto výpočtu prináša možnosť kresliť touto metódou grafy z väčším počtom vrcholov. Pri tejto metóde sú v grafe identifikované časti, takzvané hrubé grafy, ktoré majú jednoduchú štruktúru a preto je možné na nich znižovať energiu podstatne rýchlejšie. Samotné znižovanie energie funguje tak, že sa z grafu vyčlenia hrubé grafy s jednoduchou štruktúrou a tvoria akoby jeden samostatný vrchol. Ostatným častiam je znižovaná energia klasickým spôsobom pomocou parciálnych diferenciálnych rovníc [10]. Potom nasleduje úprava v hrubých grafoch, kde sa energia neznižuje postupne, ale vrcholy sa umiestňujú priamo na miesta, aby boli od ich susedov v požadovanej vzdialenosti. Po ich úprave sú rekurzívne znovu v štruktúre hľadané hrubé grafy. Ak sa nejaké nájdu celý postup sa takto zopakuje, ak nie algoritmus končí [9]. Príklad vykresleného grafu pomocou tejto metódy zobrazuje Obrázok 2.2.



Obrázok 2.2: Príklad automatického vykreslenia grafu pomocou optimalizácie znižovania energie systému [9].

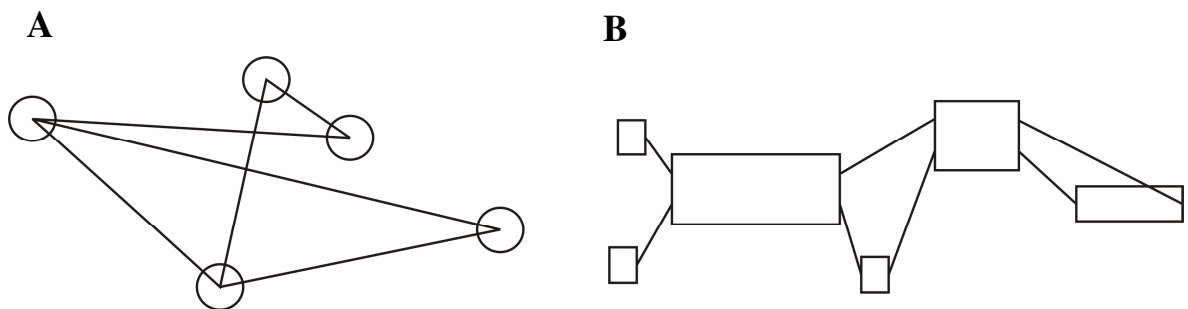
Štruktúry grafov RNA motívov sú z hľadiska kreslenia grafov jednoduché a obsahujú malý počet vrcholov. Preto je im možné znižovať energiu algoritmom simulované žihanie. Ten transformuje graf do výsledného tvaru síce pomalšie no presnejšie ako metódy, ktoré znižovanie energie optimalizujú. Pružiny, ktoré v týchto metódach rátať posúvanie môžu byť navrhnuté tak, aby kontrolovali rôzne typy podmienok. Tento fakt je využívaný pri tvorbe našej metódy na vykresľovanie RNA motívov, kde môžeme prispôbiť typy pružín požiadavkám správne nakresleného motívu.

3. Silovo riadené vykresľovanie RNA motívov

V tejto kapitole predstavíme nové riešenie problému vykresľovania RNA motívov. Na rozdiel od predchádzajúcich prístupov bolo naším cieľom zachytiť všetky podmienky správneho vykresľovania RNA motívov a to aj takých, ktoré obsahujú pseudouzly. Na kreslenie motívov sme sa rozhodli prispôbiť metódu silovo riadeného vykresľovania grafov, pretože pomocou síl je možné kontrolovať rôzne typy podmienok, ktoré majú na vykreslenom grafe platiť. Jednotlivé elementy v štruktúre deskriptora reprezentujeme ako vrcholy grafu a asociácie ako jeho hrany. Teda každý SSE, hélix a relácia budú vystupovať v grafe ako samostatné vrcholy a asociácie vytvorené pomocou mapy deskriptora ako hrany spájajúce tieto vrcholy (viď. Obrázok 3.4). Vrcholy nebudeme vykresľovať ako jednoduché body, ale komplexnejšie útvary, ktoré môžu obsahovať viacero asociačných bodov na rôznych miestach v rámci jedného útvaru a môžu byť asociované s ktorýmkoľvek asociačnými bodmi iného útvaru, alebo viacerých útvarov. Detaily metódy uvádzame vo zvyšku kapitoly.

3.1. Reprezentácia RNA motívu ako graf

V tejto podkapitole zadefinujeme problém kreslenia RNA motívov ako problém kreslenia grafov. Graf vytvorený z RNA motívu má niekoľko základných odlišností od grafov používaných v metóde kreslenia pomocou pružín. Vrcholy nie sú jednoduché body, ale štvoruholníky s viacerými asociačnými bodmi (viď. Obrázok 3.1).

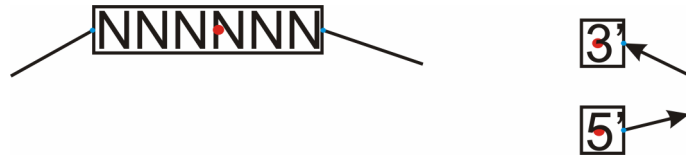


Obrázok 3.1: Štruktúry grafov.

(A) Štruktúra grafu RNA motívu. (B) Štruktúra grafu používaného pri metóde kreslenia grafov pomocou pružín.

Vrcholy môžu byť dvoch typov. Prvý typ je vrchol SSE, ktorý má dva asociačné body, ktoré môžu byť spojené s akýmkoľvek inými asociačnými bodmi v rámci grafu. Nachádzajú sa na miestach, ktoré označujú začiatok a koniec predpisu vrcholu. Jeho

Ťažisko sa nachádza v strede úsečky, ktorú tvoria asociačné body v SSE. Ťažisko zároveň udáva polohu vrcholu. Vrcholy s označením začiatku a konca kódovania motívu sú taktiež typu SSE. Majú ale iba jediný asociačný bod a ich predpis je 5' a 3' (viď. Obrázok 3.2).



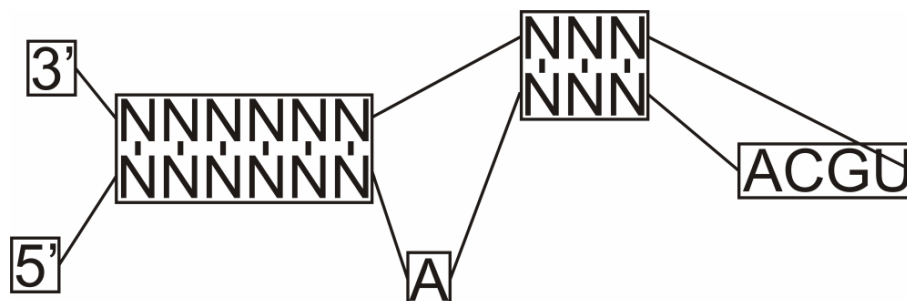
Obrázok 3.2: Príklad štruktúry vrcholu SSE.
Modré body označujú asociačné body vrcholu a červené body označujú ich ťažiská.

Druhý typ je hélix alebo relácia, ktorý má štyri asociačné body. Tak ako SSE, môžu byť spojené s akýmikoľvek asociačnými bodmi v rámci grafu. Nachádzajú sa na miestach označujúcich začiatok a koniec predpisov ich komplementárnych vlákien (viď. Obrázok 3.3).



Obrázok 3.3: Príklad štruktúry vrcholu elementu hélix alebo relácia.
Modré body označujú asociačné body vrcholu a červený bod označuje jeho ťažisko.

Ťažisko je na priesečníku úsečiek tvorených asociačnými vrcholmi a taktiež udáva polohu vrcholu. Hrany grafu sú potom spojnice medzi asociačnými bodmi susedných vrcholov vytvorené pomocou mapy v deskriptore. Každý asociačný bod má iba jednu príľahlú hranu (viď. Obrázok 3.4).



Obrázok 3.4: Príklad kompletnej štruktúry RNA motívu.

3.2. Metóda vykresľovania grafu

Naša metóda je založená na silovo riadenom vykresľovaní. Začínáme z iniciálneho nakreslenia, ktoré modifikujeme tak, že pohybujeme jednotlivými vrcholmi na základe síl, ktoré na tieto vrcholy pôsobia, až kým sa nedostaneme do rovnovážneho stavu. V našom systéme máme tri funkcie kontrolujúce rôzne druhy síl. Prvá funkcia počíta príťažlivé a odpudivé sily medzi susednými vrcholmi (FDn) spojenými hranou. Druhá funkcia (FDe) počíta iba odpudivé sily medzi vrcholmi, ktoré sú v kolízií. Tretia funkcia (FDa) počíta sily, ktoré pôsobia na jednotlivé vrcholy tak aby sa udržali správne uhly medzi susednými vrcholmi. Okrem týchto funkcií je možné v algoritme robiť na vrcholoch niekoľko základných operácií. V prvom rade môžeme pohybovať vrcholmi v smere pôsobiacich síl. Druhá operácia umožňuje rotáciu vrcholu okolo jeho ťažiska, pričom sa rotuje o taký uhol, aby výsledné pôsobenie síl na vrchol bolo minimálne. Posledná operácia umožňuje héliom a reláciám zrkadlovo vymieňať pozície vlákien s ich komplementmi. V ďalších podkapitolách popíšeme jednotlivé sily a operácie podrobnejšie.

3.2.1. Funkcia FDn

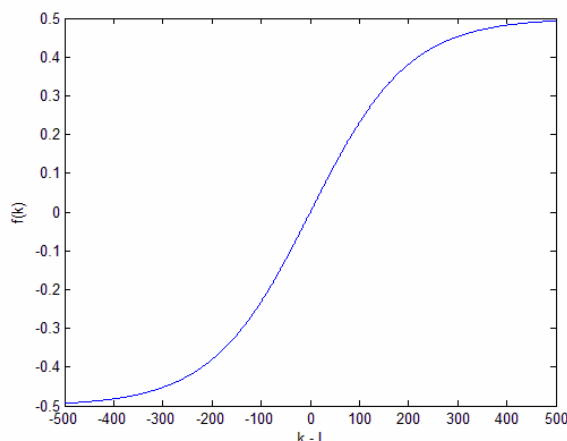
Funkcia FDn zabezpečuje, aby sa vrcholy, ktoré sú asociované hranami (pružinami), príťahovali alebo odpudzovali tak, aby bola medzi nimi dodržaná istá vzdialenosť. Teda kontroluje vzdialenosti medzi jednotlivými elementmi a tým aj vzdialenosti medzi nukleotidmi, čím dodržiava jednu z podmienok pre správne vykreslený motív. FDn je v podstate funkcia popísaná v metóde kreslenia grafov pomocou pružín v sekcii 2.1, prispôbená požiadavkám kreslenia RNA motívu.

Účinok funkcie FDn je možné kontrolovať dvoma parametrami. Prvý parameter je aktuálna dĺžka pružiny k . Druhým parametrom je funkcia $f(k)$, ktorá vypočíta veľkosť sily pružiny. Sila vypočítaná funkciou $FDn(k, f(k))$ je potom sila pružiny pôsobiacej na jej príľahlé body (asociačné body vrcholov). Každý vrchol v grafe má podľa typu dva alebo štyri asociačné body, ktoré spája hrana (pružina) s inými asociačnými bodmi ako je to popísané v štruktúre grafu RNA motívu. Keďže vzdialenosti medzi vrcholmi sú rôzne, na každý asociačný bod vrcholu pôsobí iná sila pružiny. Výsledná sila pôsobiaca na vrchol je vektorový súčet všetkých síl pôsobiacich na jeho asociačné body počítané funkciou FDn . Táto sila bude pôsobiť v ťažisku vrcholu a podľa nej bude v behu algoritmu vrchol posúvaný.

Samotný výpočet funkcie $FDn(k, f(k))$ pozostáva z niekoľkých častí. Najprv je nutné vypočítať aktuálnu dĺžku pružiny k . Pre dva asociačné body na pozíciách a a b platí $k = |a - b|$. Následne funkcia $f(k)$ je definovaná nasledovne

$$f(k) = \frac{1}{1 + e^{-0.01 \cdot (k-l)}} - 0.5$$

Premenná l je globálna konštanta, ktorá vyjadruje pokojovú dĺžku pružiny, teda dĺžku pružiny, kedy nepôsobí na svoje priľahlé body žiadnymi silami. Hodnota $k - l$ vyjadruje aktuálnu účinnú dĺžku pružiny. Ide o aktuálne predĺženie alebo skrátenie pružiny oproti pokojovej dĺžke l . Číslo $0,01$ vyjadruje ako rýchlo má sila pružiny rásť vzhľadom k maximálnej hodnote pri zmene hodnoty $k - l$, teda pri zmene aktuálnej účinnej dĺžky pružiny. Táto funkcia sa nazýva aj sigmoid je ohraničená zhora číslom $0,5$ a zdola číslom $-0,5$ (viď. Obrázok 3.5)



Obrázok 3.5: Priebeh funkcie sigmoid $f(k) = \frac{1}{1 + e^{-0.01 \cdot (k-l)}} - 0.5$.

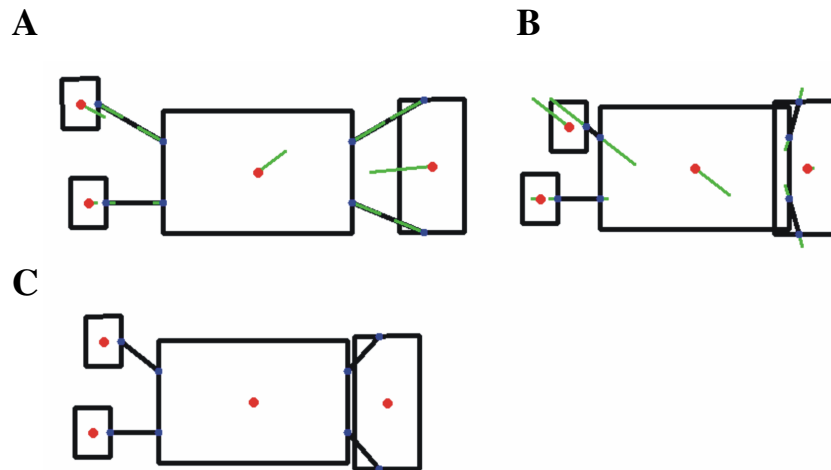
Aby bolo možné kontrolovať maximálnu respektíve minimálnu hodnotu funkcie $f(k)$ je nutné ju doplniť o ďalšiu premennú. Potom

$$f(k) = \left(\frac{1}{1 + e^{-0.01 \cdot (k-l)}} - 0.5 \right) \cdot 2d$$

Premenná d je globálna premenná, ktorá vyjadruje maximálnu veľkosť sily pružiny. Pri zväčšovaní aktuálnej účinnej dĺžky pružiny veľkosť sily pôsobiacej na jej priľahlé body stúpa po hodnotu d a pri znižovaní aktuálnej účinnej dĺžky pružiny klesá po hodnotu $-d$.

Pre každý asociačný bod je nutné vypočítať smer sily pružiny, ktorá na neho pôsobí. Veľkosť tohto smeru zodpovedá veľkosti sily ktorou pružina na bod pôsobí (viď. Obrázok

3.6). Ak bod $a_{ij} \in v_i$ je asociovaný s bodom $a_{kl} \in v_k$, a platí $v_i, v_j \in V$, potom smerový vektor u z a_{ij} do a_{kl} je $u = a_{kl} - a_{ij}$. Jednotkový vektor vektoru u je $u_1 = u/|u|$. Potom smer pôsobenia pružiny na bod a_{ij} je $u_1 \cdot FDn(k, f(k))$. Keďže FDn môže nadobúdať aj záporné hodnoty ak $k-l < 0$ (viď. Obrázok 3.5) a práve vtedy má pôsobiť na bod odpudivá sila pružiny, súčin zápornej FDn s vektorom u_1 zabezpečí otočenie vektora o 180 stupňov, teda výsledný smer sa zmení na opačný (viď. Obrázok 3.6 B)



Obrázok 3.6: Príklady pôsobenia síl pružín funkcie FNn .

Zelené čiary vyjadrujú smer pôsobenia sily FDn z konkrétneho bodu. Dĺžka čiary je potom veľkosť tejto sily. (A) Pôsobenie príťažlivých síl medzi asociačnými bodmi. (B) Pôsobenie odpudivých síl medzi asociačnými bodmi. (C) Rovnovážny stav kedy v grafe nepôsobia žiadne sily pružín.

Takto definovaná funkcia $f(k)$ síce funguje dobre, no pri použití FDn vo funkcii rotácie vrcholu, kde je nutné rotovať vrchol o taký uhol, kedy je sila pôsobenia na susedné vrcholy minimálna, nepostačuje.

Dobre definovaná funkcia $f(k)$, ktorá pri rotácií zabezpečí správne uhly medzi susednými vrcholmi je kvadratická. Definovaná je zvlášť pre príťažlivú silu a odpudivú silu. V prípade že aktuálna veľkosť pružiny k je väčšia ako pokojová l ide o príťažlivú silu, potom

$$f(k) = \frac{k^2}{l}$$

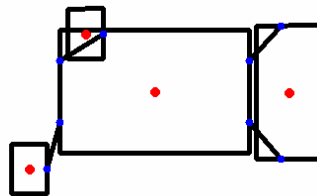
v ostatných prípadoch je $f(k)$ definovaná ako

$$f(k) = -\frac{l^2}{k}$$

Vlastnosť, že pri príťažlivých silách funkcia $f(k)$ nie je zhora ohraničená a pri odpudivých silách nie je zdola ohraničená neskôr zabezpečí správny výpočet uhla rotácie.

Smer pôsobenia sily na bod sa počíta rovnako ako v prípade funkcie sigmoid $u_1 \cdot FDn(k, f(k))$. Tu taktiež záporná hodnota funkcie $f(k)$ zabezpečí otočenie vektora o 180 stupňov. Celková energia grafu RNA motívu kontrolovaného funkciou FDn je En . Je to vlastne súčet všetkých síl pružín v grafe

Aj keď funkcia FDn kontroluje rozmiestnenie vrcholov grafu tak, aby boli dodržané predpísané vzdialenosti medzi susedmi, stále nie sú splnené všetky podmienky správneho vykreslenia RNA motívu. Aj keď energia En je nulová alebo minimálna možná nastávajú pomerne často situácie kedy sa vrcholy prekrývajú. Funkcia taktiež nemá vplyv na udržiavanie správnych uhlov medzi susednými vrcholmi (viď. Obrázok 3.7).



Obrázok 3.7: Príklad možného vykreslenia grafu iba pomocou síl pružín. Dochádza k prekrývaniu vrcholov a k nedodržaniu pravidelných uhlov medzi vrcholmi.

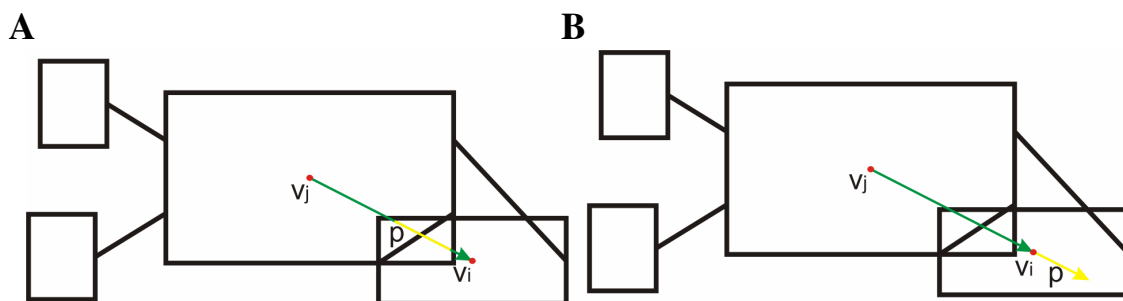
Tieto problémy postupne riešia funkcie FDe , FDa , rotácia vrcholu a zrkadlové otáčanie vlákien hélixy a relácie.

3.2.2. Funkcia FDe

Jednou z podmienok správne vykresleného RNA motívu je to, aby sa jednotlivé nukleotidy neprekrývali. V prípade RNA motívu definovanom ako graf je táto podmienka zabezpečená v prípade, že sa neprekrývajú vrcholy. Aby sa graf takto vyhýbal kolíziám je definovaná funkcia FDe , ktorá počíta silu a smer pohybu vrcholu ak dôjde ku kolízií s iným vrcholom. Hľadanie všetkých vrcholov, ktoré sú v kolízií s vrcholom pre ktorý sa výsledná odpudivá sila počíta zabezpečuje algoritmus SAT [8] [14] pre detekciu kolízií. Účinok funkcie FDe kontrolujú dva parametre. Prvý je hĺbka prieniku vrcholov v kolízií p a druhý parameter je funkcia $f(p)$, ktorá počíta odpudivú silu pôsobiacu na vrchol. Výsledkom funkcie $FDe(p, f(p))$ je smer pôsobenia sily na vrchol, ktorý je v kolízií s iným vrcholom. Veľkosť tohto smerového vektora je potom veľkosť pôsobiacej sily. V prípade, že s vrcholom v_i je v kolízií n vrcholov a sú to vrcholy v_1, v_2, \dots, v_n , pričom $v_i \neq v_1, v_2, \dots, v_n$ a $v_i, v_1, v_2, \dots, v_n \in V$ funkcia FDe počíta silu a smer pre každú dvojicu

$(v_i, v_1), (v_i, v_2), \dots, (v_i, v_n)$ a výsledná sila je potom ich vektorový súčet. Opudivá sila bude pôsobiť taktiež ako pri FDn v ťažisku vrcholu a podľa nej bude v behu algoritmu vrchol posúvaný.

Výpočet funkcie $FDe(p, f(p))$ prebieha v dvoch krokoch. Prvý krok je výpočet parametra p a druhý krok je výpočet funkcie $f(p)$. Hľadanie hĺbky prieniku dvoch vrcholov p je komplexnejší problém, pretože záleží na tom, ktorým smerom bude pôsobiť na vrchol opudivá sila. Ako dobrý a rýchly spôsob nájdenia smeru opudivej sily sa ukázal smerový vektor úsečky tvorenej ťažiskami vrcholov, ktoré sú v kolízii v smere ku vrcholu na ktorého má sila pôsobiť. Na tejto úsečke sa potom nachádza aj úsek, kde jeho dĺžka je hĺbkou prieniku p (viď. Obrázok 3.8).



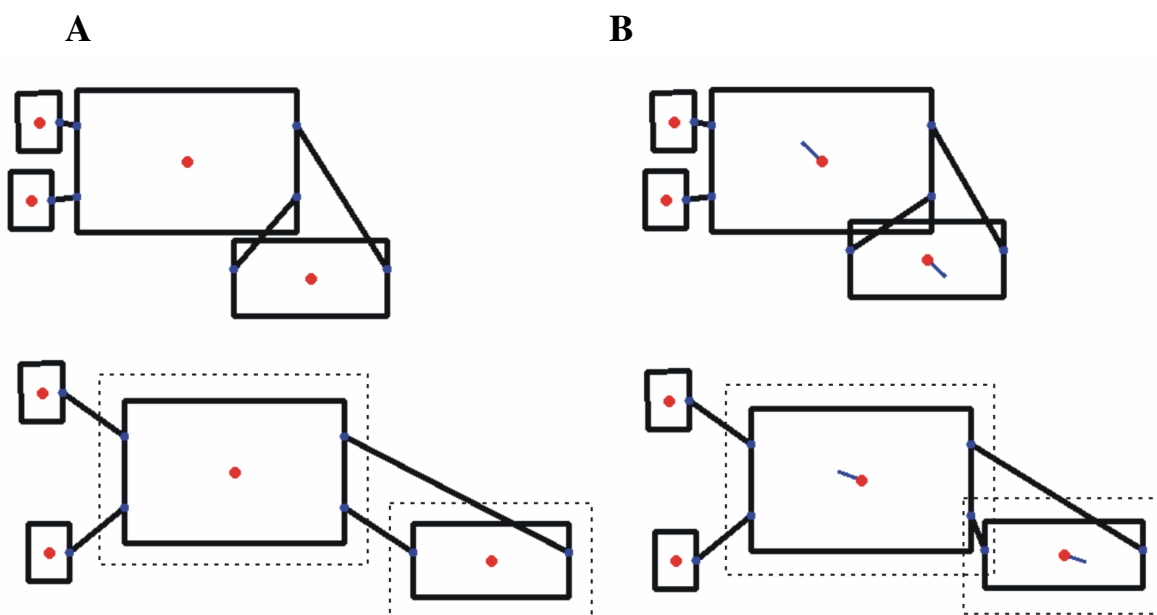
Obrázok 3.8: Príklad nájdenia hĺbky prieniku dvoch vrcholov.

Zelená šípka vyjadruje smer pôsobenia opudivej sily ktorá pôsobí na vrchol v_i . (A) Žltá čiara vyjadruje hĺbku prieniku p . (B) Žltá šípka ukazuje smer ktorým sa vrchol v_i musí pohnúť o vzdialenosť ktorá sa rovná jej dĺžke aby vrcholy v_i a v_j neboli v kolízii.

Funkcia $f(p)$ tak ako pri FDn môže mať viacero predpisov. Aby bolo možné kontrolovať maximálnu veľkosť opudivej sily pôjde o predpis funkcie sigmoid. Potom

$$f(p) = \left(\frac{1}{1 + e^{-0.01 \cdot p}} - 0.5 \right) \cdot 2g$$

Je jasné že parameter p nebude nadobúdať záporné hodnoty pretože ak sú vrcholy v kolízii, ich hĺbka je väčšia ako nula a ak nie sú v kolízii funkcia FDe a teda zároveň aj $f(p)$ nebude počítaná. Aj funkcia $f(p)$ preto nebude nadobúdať záporné hodnoty. Premenná g je globálna premenná, ktorá vyjadruje maximálnu veľkosť opudivej sily. Pri zväčšovaní hĺbky prieniku p veľkosť sily pôsobiacej na vrchol stúpa po hodnotu g (viď. Obrázok 3.9).



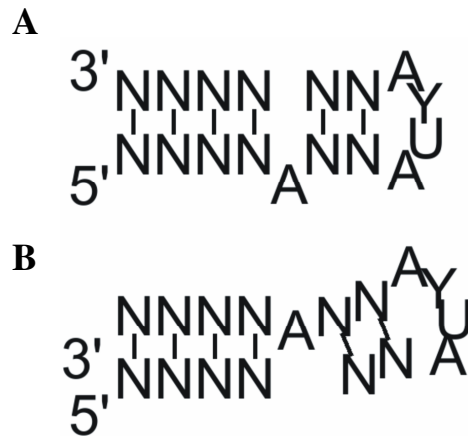
Obrázok 3.9: Príklady pôsobenia odpudivých síl funkcie FNe .
 (A) Ukážka kolízie kedy okolo vrcholu nie je ochranné pole, teda ku kolízií dochádza až vtedy keď sa vrcholy reálne prekrývajú. (B) Ukážka kolízie kedy je okolo vrcholu nastavené ochranné pole. Vtedy ku kolízií dochádza ak sa prekrývajú už ochranné polia.

Samotná funkcia $FDe(p, f(p))$ potom vypočíta smer pôsobenia odpudivej sily, kde veľkosť smerového vektoru je veľkosť tejto sily. Ak v_i je vrchol, na ktorý má sila pôsobiť a v_j je vrchol, ktorý je s ním v kolízií a $v_i, v_j \in V$, potom smerový vektor odpudivej sily u vrcholov v_i a v_j je $u = v_i - v_j$. Jednotkový vektor vektoru u je $u_1 = u/|u|$. Potom smer pôsobenia odpudivej sily je $u_1 \cdot f(p)$. Okrem premennej g je možné nastavovať pri počítaní kolízií aj to v akom okolí od vrcholu už má dochádzať ku pozitívnej detekcií. Ide o to aby sa konštantou dalo kontrolovať minimálnu vzdialenosť v akej sa majú vrcholy udržiavať. Táto konštanta sa aplikuje pri detekcií kolízií tak, že sa výška a šírka vrcholu predĺži o túto konštantu. Vznikne tak okolo každého vrcholu pole, v ktorom sa nesmie nachádzať žiaden iný vrchol (vid'. Obrázok 3.9 B).

Celková energia grafu RNA motívu kontrolovaného funkciou FDe je Ee . Je to vlastne súčet celkových odpudivých síl počítaných funkciou FDe všetkých vrcholov. Nulová hodnota En zaručuje, že žiaden vrchol nekoliduje s inými vrcholmi a ani nie je v ich nastavenom okolí.

3.2.3. Funkcia FDa

Ďalšou z podmienok správne vykresleného RNA motívu je udržiavanie častí motívu v pravidelných orientáciách. Teda je nutné udržiavať časti motívu v uhloch, ktoré nenarúšajú jeho celkový vzhľad (viď. Obrázok 3.10).

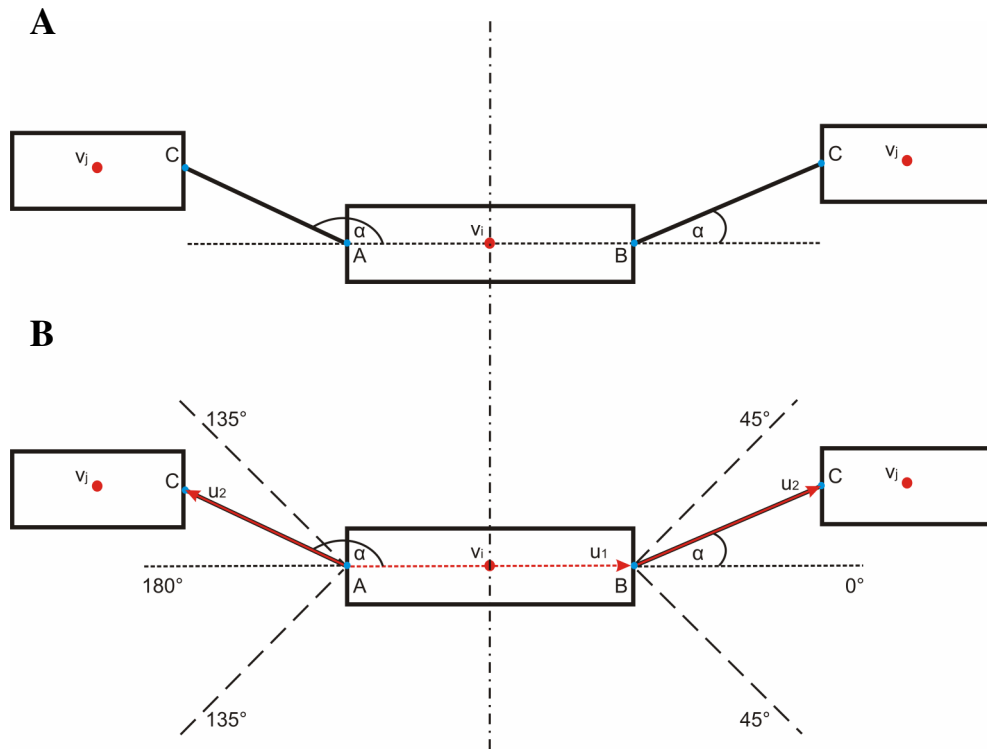


Obrázok 3.10: Príklad nutnosti použitia síl FDa na kontrolu uhlov medzi vrcholmi.

(A) Správne vykreslený RNA motív. (B) RNA motív kedy sú na vykreslenia použité iba funkcia FDn a FDe .

Aby bola dodržaná táto podmienka je nutné zaviesť ďalšiu silu, ktorá bude pôsobiť na jednotlivé asociačné body tak, aby boli dodržané správne uhly medzi susednými vrcholmi. Túto silu počíta funkcia FDa . Účinok funkcie FDa kontrolujú dva parametre. Prvý je uhol α ktorý zvierá vrchol asociačného bodu s hranou príľahlou asociácii (viď. Obrázok 3.11 A) a druhý je funkcia $f(\alpha)$, ktorá počíta silu pôsobiacu na asociačný bod. Keďže vrcholy môžu mať dva alebo štyri asociačné body, ktoré sú asociované s rôznymi vrcholmi tak ako je to popísané v štruktúre grafu RNA motívu, môže pôsobiť na každý bod rozdielna sila v rôznych smeroch. Výsledná sila pôsobiaca na vrchol je potom rovnako ako vo FDn vektorový súčet všetkých síl pôsobiacich na jeho asociačné body počítaných funkciou FDa . Táto sila bude pôsobiť v ťažisku vrcholu a podľa nej bude v behu algoritmu vrchol posúvaný. Jednotlivé asociačné body sú vo vrchole radené proti smere hodinových ručičiek, kde prvý asociačný bod je vľavo vzhľadom na ťažisko. Ak ide o hélix alebo reláciu prvý asociačný bod je vľavo dole vzhľadom na ťažisko. Výpočet funkcie $FDa(\alpha, f(\alpha))$ prebieha v dvoch krokoch. Prvý krok je výpočet parametra α a druhý krok je výpočet funkcie $f(\alpha)$.

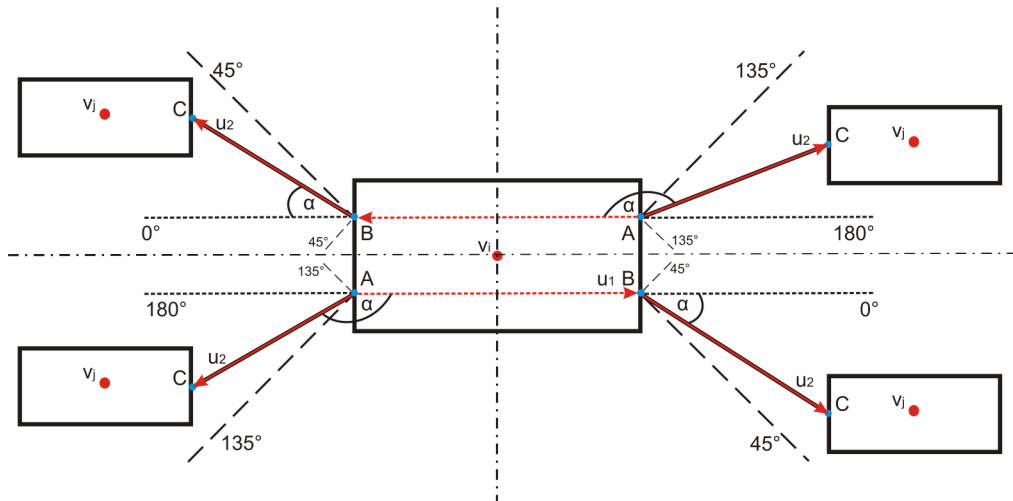
Sú dané tri body A , B , C . Bod A je prvý asociačný bod vrcholu v_i , bod B je druhý asociačný bod vrcholu v_i a bod C je asociačný bod vrcholu v_j , ktorý je asociovaný s vrcholom v_i v bodoch A alebo B (viď. Obrázok 3.11).



Obrázok 3.11: Uhol alfa funkcie F_{Da} .

Vpravo ak je počítaný prvý asociačný bod A . Vľavo ak je počítaný druhý asociačný bod B . (A) Ukážka ktoré uhly alfa je potrebné hľadať. (B) Vektory u_1 a u_2 podľa, ktorých je uhol alfa nájdený. Vyznačené uhly sú považované za správne. K nim sa bude vrchol postupne približovať.

V prípade, že ide o horné vlákno elementu hélíx alebo relácie bod A bude v poradí tretí asociačný bod a bod B bude v poradí štvrtý asociačný bod vrcholu v_i (viď. Obrázok 3.12).



Obrázok 3.12: Príklad správnych uhlov, rozmiestnenia vektorov u_1 a u_2 a priradenie bodov A, B a C pre element hélíx a reláciu.

Uhol α , ktorý zvierá vrchol a hrana príslušná asociácii je možné počítať aj ako uhol dvoch smerových vektorov u_1 a u_2 (viď. Obrázok 3.11 B). Prvým vektorom na výpočet uhla α je $u_1 = B - A$. V prípade, že funkcia FDa počíta silu pre bod A je bod C asociačný bod spojeným hranou s bodom A. Druhým vektorom je potom $u_2 = C - A$. V prípade, že funkcia FDa počíta silu pre bod B je bod C asociačný bod spojeným hranou s bodom B. Druhým vektorom je potom $u_2 = C - B$. Uhol vektorov u_1 a u_2 je možné vyjadriť vzťahom

$$\cos(u_1, u_2) = \frac{u_{11}u_{21} + u_{12}u_{22}}{\sqrt{u_{11}^2 + u_{12}^2} \cdot \sqrt{u_{21}^2 + u_{22}^2}}$$

Keďže ide o uhol dvoch vektorov, α môže nadobúdať hodnoty $\alpha = \langle 0, 180^\circ \rangle$.

Výpočet funkcie $f(\alpha)$ závisí nielen od hodnoty α , ale aj od toho ktorý z bodov A a B je ten pre ktorý funkcia FDa počíta silu. Taktiež je nutné neskôr rozlíšiť či je vrchol SSE alebo hélíx respektíve relácia. Aby bolo možné kontrolovať vzájomnú orientáciu vektorov u_1 a u_2 musia sa zdefinovať uhly, ktoré bude funkcia $f(\alpha)$ považovať za dobré teda také kedy bude sila pôsobiaca na asociačný bod nulová. Ak funkcia $f(\alpha)$ počíta silu pre bod A, tak dobré uhly budú 180° a 135° . V prípade, že funkcia $f(\alpha)$ počíta silu pre bod B, dobré uhly budú 0 a 45° . Postupným približovaním sa k týmto uhlom bude sila pôsobiaca na bod klesať po hodnotu nula a naopak pri oddiaľovaní bude sila stúpať po maximálnu hodnotu ktorou môže FDa na bod pôsobiť (viď. Obrázok 3.11 B, Obrázok 3.12). Je zřejmé, že rozmedzie medzi uhlami, kedy majú sily stúpať respektíve klesať, nie je vždy rovnaké. Preto funkcia $f(\alpha)$ musí byť prispôbovaná podľa hodnoty α a podľa pozície bodu C.

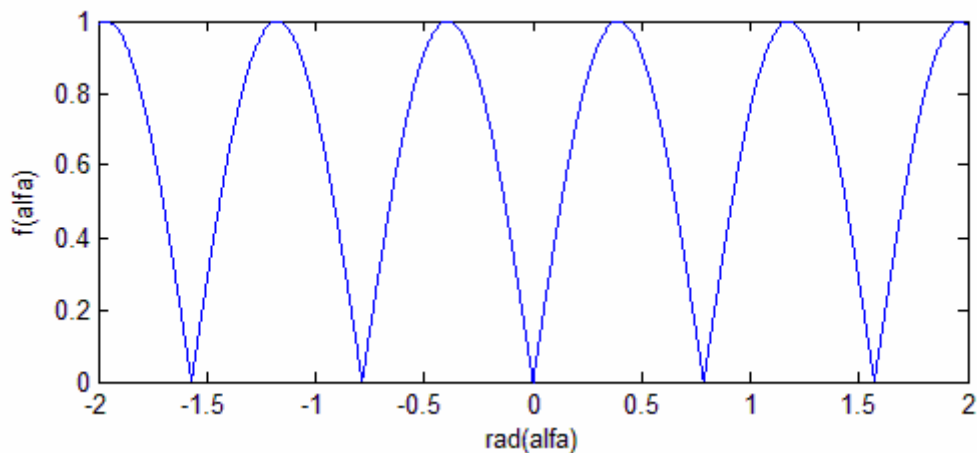
Funkcia $f(\alpha)$ počíta silu rozdielne pre dva prípady. Prvý prípad je ak funkcia $f(\alpha)$ počíta silu pre bod A , vtedy je rozdiel medzi dobrými uhlami 180° a 135° práve 45° stupňov. Maximálnu hodnotu musí preto funkcia $f(\alpha)$ nadobúdať pri hodnote uhla $157,5^\circ$. Druhý prípad je ak funkcia $f(\alpha)$ počíta silu pre bod B vtedy rozdiel medzi dobrými uhlami 0 a 45° je tiež 45° stupňov. Maximálnu hodnotu musí funkcia $f(\alpha)$ nadobúdať pri hodnote uhla $22,5^\circ$. Teda ak $\alpha \leq 180^\circ \wedge \alpha \geq 135^\circ$ v prvom prípade, $\alpha \geq 0 \wedge \alpha \leq 45^\circ$ v druhom prípade a radián uhla α je

$$rad(\alpha) = \alpha \cdot \frac{\pi}{180}$$

tak potom funkcia $f(\alpha)$ má predpis

$$f(\alpha) = |\sin(4 \cdot rad(\alpha))|$$

Takto definovaná funkcia $f(\alpha)$ je ohraničená zdola nulou a zhora jednotkou (viď. Obrázok 3.13).



Obrázok 3.13: Pribeh funkcie $f(\alpha) = |\sin(4 \cdot rad(\alpha))|$.

Aby bolo možné kontrolovať jej maximálnu hodnotu je nutné ju doplniť o ďalšiu premennú, potom

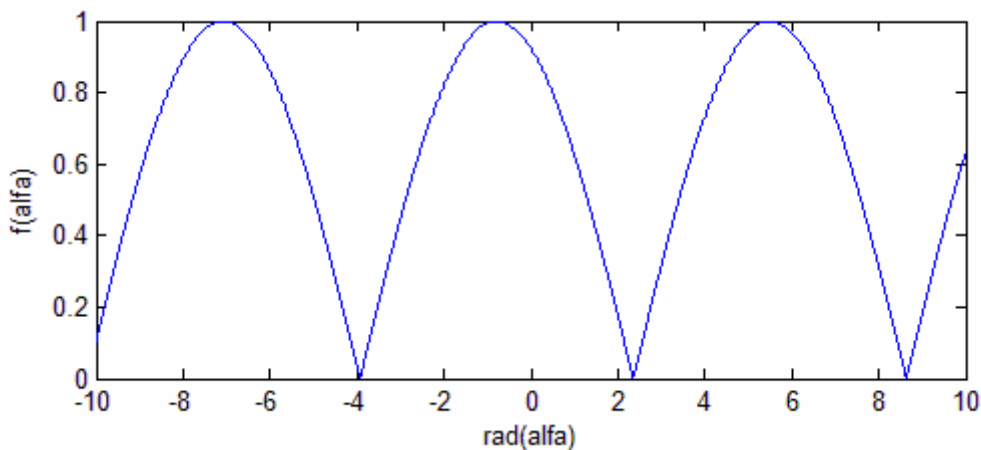
$$f(\alpha) = |\sin(4 \cdot rad(\alpha))| \cdot m$$

Premenná m je globálna premenná, ktorá vyjadruje maximálnu veľkosť sily funkcie FDA , ktorá môže pôsobiť na asociačný bod. Týmto predpisom je ale zabezpečený iba stav, kedy $\alpha \leq 180^\circ \wedge \alpha \geq 135^\circ$ v prvom prípade, $\alpha \geq 0 \wedge \alpha \leq 45^\circ$ v druhom prípade. Pre ostatné stavy je nutné definovať funkcie ktoré budú stúpať respektíve klesať v inom rozmedzí.

Ak funkcia $f(\alpha)$ počíta silu pre bod A a uhol $\alpha < 135^\circ$, tak funkcia stúpa po hodnotu $\alpha = 0$ kde $f(\alpha)$ nadobúda svoje maximum. Potom

$$f(\alpha) = \left| \sin \left(0.5 \cdot \text{rad}(\alpha) - \frac{\pi}{8} \cdot 3 \right) \right| \cdot m$$

Priebeh tejto funkcie znázorňuje Obrázok 3.14.

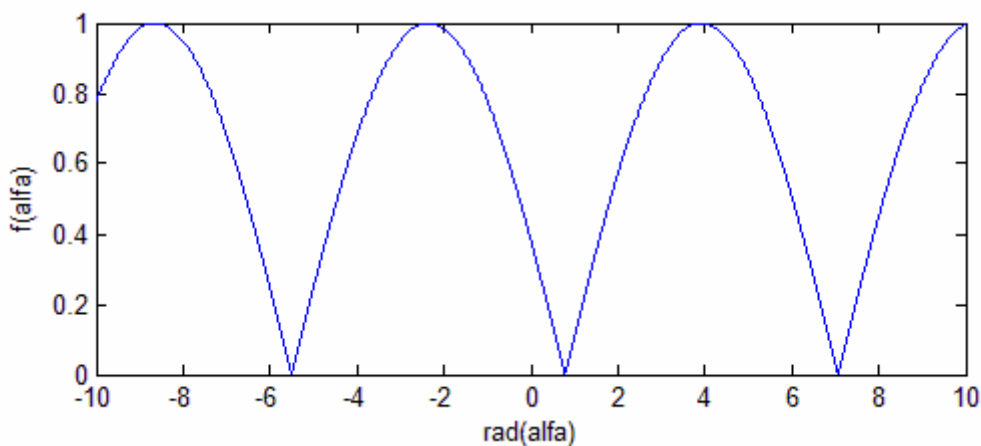


Obrázok 3.14: Priebeh funkcie $f(\alpha) = \left| \sin \left(0.5 \cdot \text{rad}(\alpha) - \frac{\pi}{8} \cdot 3 \right) \right| \cdot m$.

Ak funkcia $f(\alpha)$ počíta silu pre bod B a uhol $\alpha > 45^\circ$, tak funkcia stúpa po hodnotu $\alpha = 180^\circ$ kde $f(\alpha)$ nadobúda svoje maximum. Potom

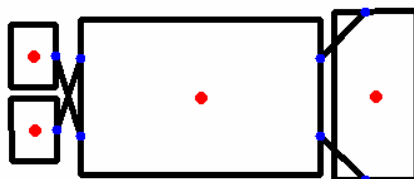
$$f(\alpha) = \left| \sin \left(0.5 \cdot \text{rad}(\alpha) - \frac{\pi}{8} \right) \right| \cdot m$$

Priebeh tejto funkcie znázorňuje Obrázok 3.15.



Obrázok 3.15: Priebeh funkcie $f(\alpha) = \left| \sin \left(0.5 \cdot \text{rad}(\alpha) - \frac{\pi}{8} \right) \right| \cdot m$.

Týmito podmienkami je zabezpečený výpočet síl tak aby žiadne sily nepôsobili na asociačné body, ak je uhol medzi ich vrcholom a príľahlou hranou 0 alebo 45° stupňov. Je nutné ale tento výpočet síl ešte ohraničiť, pretože štruktúra elementov hélíx a relácie je komplexnejšia. Každý takýto element obsahuje štyri asociačné body. Predchádzajúci výpočet povoľoval aj uhly pre asociačné body v horizontálnej línii ktoré smerujú k sebe (viď. Obrázok 3.16).



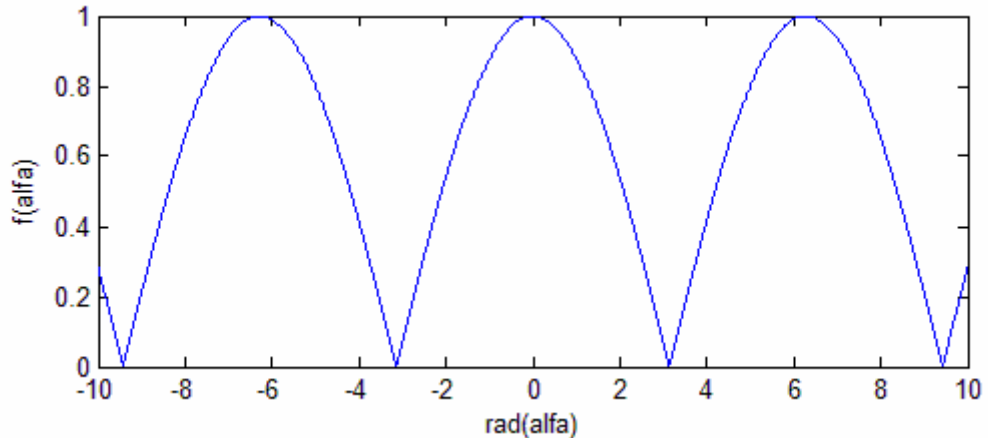
Obrázok 3.16: Príklad elementu hélíx alebo relácie kedy sú povolené aj vnútorné uhly. Môže dôjsť k stavu kedy sú uhly podľa funkcie FDA v poriadku no reálne vykreslený graf je chybný (Prekrížené asociačné hrany).

Tieto uhly je potrebné z výpočtu pre hélíx a reláciu vylúčiť. Ide o prípad kedy sa bod C nachádza pod priamkou na ktorej ležia body A a B .

Orientáciu bodu C vzhľadom na priamku na ktorej ležia body A a B je možné vypočítať nasledovne. Smerový vektor bodu A a B je $u = B - A$. Normálový vektor bodov A a B je $n = (-u_2, u_1)$, potom rovnica priamky na ktorej ležia body A a B je $n_1x + n_2y + c = 0$. Ak výsledná hodnota rovnice po osadením súradníc x a y bodu C je menšia ako nula, tak je bod C orientovaný pod priamkou a ak je hodnota väčšia ako nula tak je orientovaný nad priamkou a ak je rovná nule tak bod C leží na priamke. Ak je teda bod C orientovaný pod priamkou na ktorej ležia body A a B , ide o hélíx alebo reláciu a funkcia $f(\alpha)$ počíta silu pre bod A tak pre všetky $\alpha < 180^\circ$ funkcia stúpa po hodnotu $\alpha = 0$ kde dosiahne svoje maximum. Potom

$$f(\alpha) = \left| \sin\left(0.5 \cdot \text{rad}(\alpha) - \frac{\pi}{2}\right) \right| \cdot m$$

Priebeh tejto funkcie znázorňuje Obrázok 3.17.

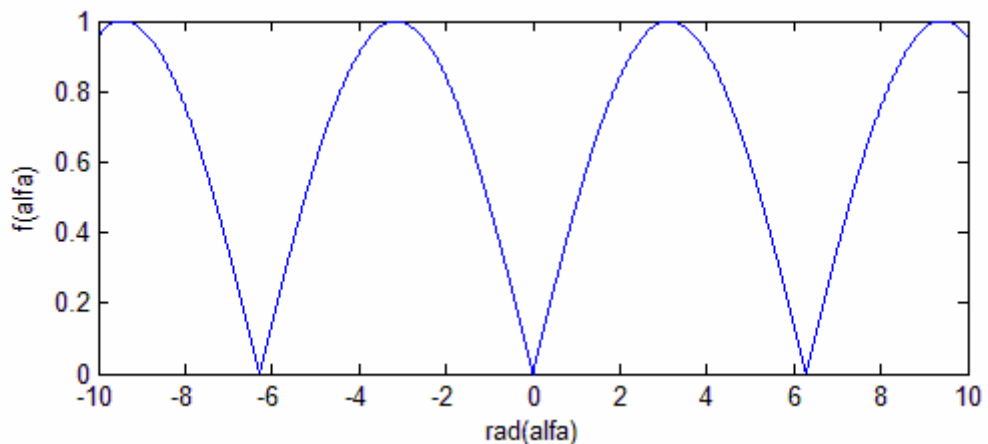


Obrázok 3.17: Priebek funkcie $f(\alpha) = \left| \sin\left(0.5 \cdot \text{rad}(\alpha) - \frac{\pi}{2}\right) \right|$.

Ak funkcia $f(\alpha)$ počíta silu pre bod B tak pre všetky $\alpha > 0$ funkcia stúpa po hodnotu $\alpha = 180^\circ$ kde dosiahne svoje maximum. Potom

$$f(\alpha) = |\sin(0.5 \cdot \text{rad}(\alpha))| \cdot m$$

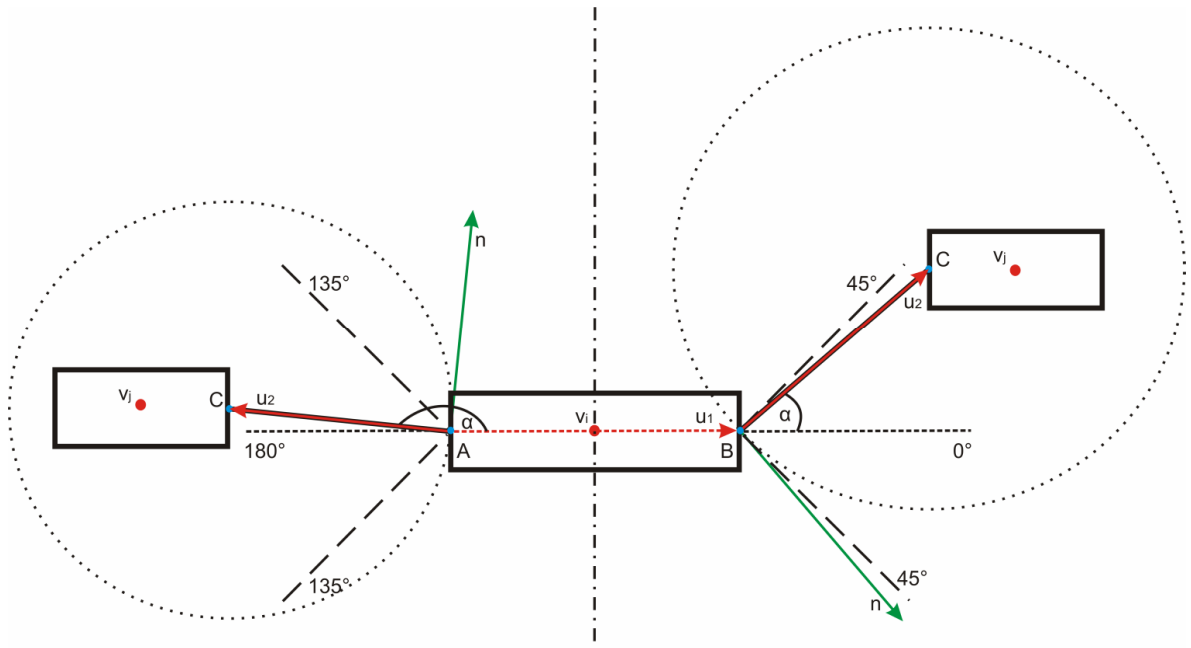
Priebek tejto funkcie znázorňuje Obrázok 3.18.



Obrázok 3.18: Priebek funkcie $f(\alpha) = |\sin(0.5 \cdot \text{rad}(\alpha))|$.

Funkcia FDa týmto vypočíta veľkosť sily, ktorá pôsobí na asociačný bod. Aby sa táto sila mohla priradiť bodu musí sa vypočítať aj smer jej pôsobenia. Pôjde o smerový vektor, ktorý bude pôsobiť v mieste výskytu asociačného bodu. Jeho veľkosť zodpovedá veľkosti sily, ktorá má na bod pôsobiť. Smer pôsobenia sily vypočítanej funkciou FDa musí mať vlastnosť, že ak sa vrchol posunie práve tým smerom tak sa uhol zvierajúci vrchol s hranou príľahlou asociačnému bodu priblíži k zadanému dobrému uhlu. Kontrolu tohto uhla

zabezpečuje pohyb vrcholu v_i po kružnici ktorej stredom je bod C a polomerom je veľkosť vektora u_2 (viď. Obrázok 3.19).



Obrázok 3.19: Príklad pôsobenia sily v asociačnom bode smerom vektoru n pričom po posune sa vektor u_2 postupne približuje k najbližšiemu správne mu uhlu. Vpravo ak je počítaný prvý asociačný bod A. Vľavo ak je počítaný druhý asociačný bod B.

Smerový vektor ktorý tento pohyb popisuje je kolmý na vektor u_2 . Práve tento vektor ktorému je veľkosť zmenená na hodnotu veľkosti sily pôsobenia je priradený asociačnému bodu.

Samotný výpočet smerového vektora pohybu je zložený z niekoľkých krokov. Základný smerový vektor pohybu je $n = (-u_{22}, u_{21})$ jednotkový vektor n je $n_1 = n/|n|$. Vektor n_1 teda zabezpečí pohyb po kružnici proti smeru hodinových ručičiek. Je zrejmé, že pohyb po kružnici musí byť v niektorých prípadoch vykonávaný aj v opačnom smere aby bolo zabezpečené, že sa aktuálny uhol približuje k dobrému uhlu najkratšou cestou, preto musí byť vektor n_1 otáčaný teda prenášobný hodnotou mínus jeden. Vektor n_1 má hodnotu $n_1 = -n/|n|$ v týchto situáciách len v prípade, že je vrchol SSE:

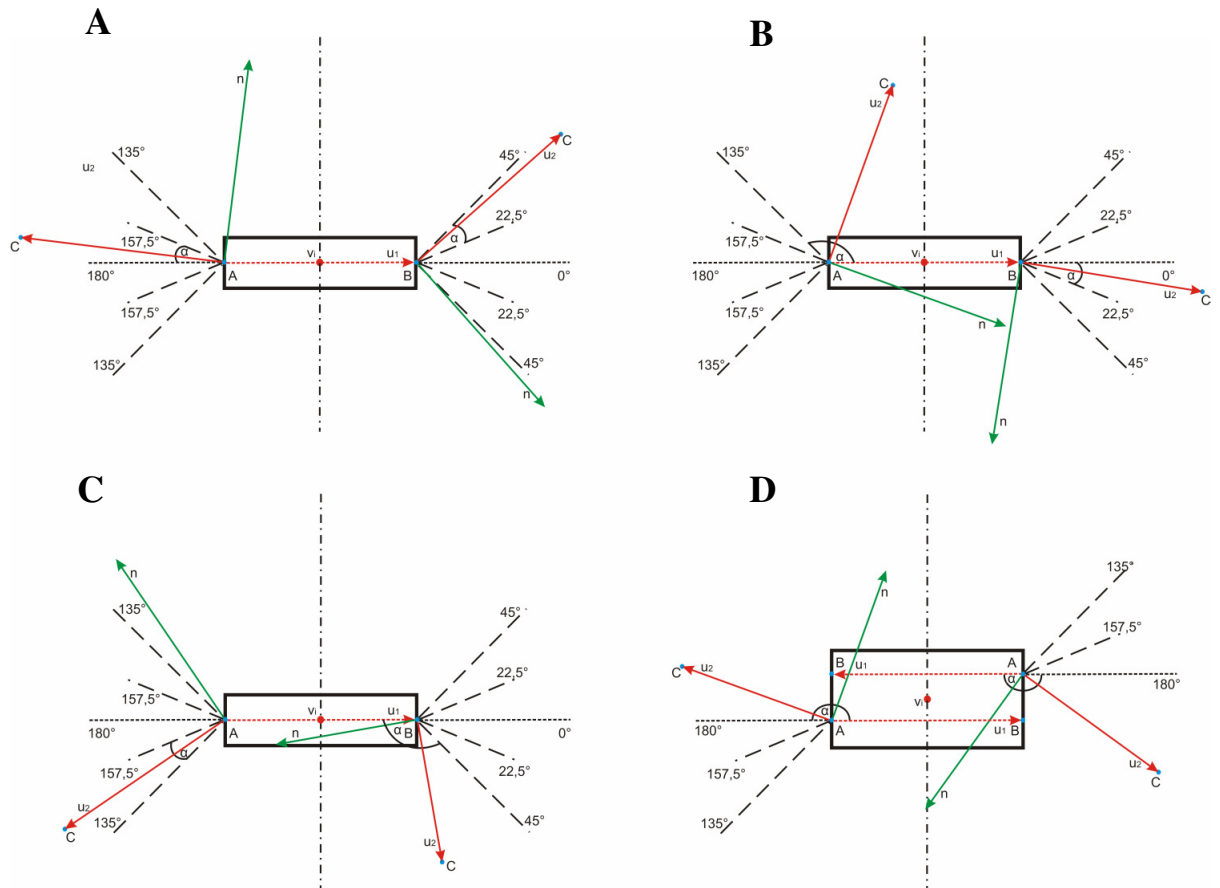
1. Funkcia $f(\alpha)$ počíta silu pre bod A, uhol $\alpha \geq 157.5^\circ \wedge \alpha \leq 180^\circ$ a orientácia bodu C je nad priamkou na ktorej ležia body A a B (viď. Obrázok 3.20 A vľavo).
2. Funkcia $f(\alpha)$ počíta silu pre bod B, uhol $\alpha \leq 135^\circ \wedge \alpha \geq 0$ a orientácia bodu C je nad priamkou na ktorej ležia body A a B (viď. Obrázok 3.20 B vľavo).

3. Funkcia $f(\alpha)$ počíta silu pre bod A , uhol $\alpha \leq 157.5^\circ \wedge \alpha \geq 135$ a orientácia bodu C je pod priamkou na ktorej ležia body A a B (vid'. Obrázok 3.20 C vľavo).
4. Funkcia $f(\alpha)$ počíta silu pre bod B , uhol $\alpha \leq 45^\circ \wedge \alpha \geq 22.5^\circ$ a orientácia bodu C je nad priamkou na ktorej ležia body A a B (vid'. Obrázok 3.20 vpravo).
5. Funkcia $f(\alpha)$ počíta silu pre bod B , uhol $\alpha \leq 22.5^\circ \wedge \alpha \geq 0$ a orientácia bodu C je pod priamkou na ktorej ležia body A a B (vid'. Obrázok 3.20 B vpravo).
6. Funkcia $f(\alpha)$ počíta silu pre bod B , uhol $\alpha \leq 180^\circ \wedge \alpha \geq 45^\circ$ a orientácia bodu C je pod priamkou na ktorej ležia body A a B (vid'. Obrázok 3.20 C vpravo).

Ak je vrchol hélix alebo relácia podmienky 1, 2 a 4 sú totožné ako pri elemente SSE.

Podmienky 3, 5 a 6 nahradia podmienky:

1. Funkcia $f(\alpha)$ počíta silu pre bod A a orientácia bodu C je pod priamkou na ktorej ležia body A a B v prípade, že ide o horné vlákno (vid'. Obrázok 3.20 D vpravo).
2. Funkcia $f(\alpha)$ počíta silu pre bod A a orientácia bodu C je nad priamkou na ktorej ležia body A a B v prípade, že ide o dolné vlákno (vid'. Obrázok 3.20 D vľavo).

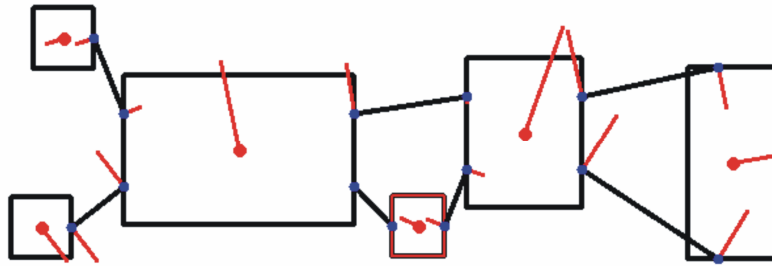


Obrázok 3.20: Príklady jednotlivých možností umiestnenia vektoru u_2 medzi správnymi uhlami kedy $n_1 = -n/|n|$.

(A, B, C) Prípady vľavo a vpravo, kedy vektor n pôsobí na asociačný bod tak aby sa vektor u_2 priblížil k najbližšiemu správne uhl. (D) Pre hélix alebo reláciu je možné vidieť vylúčené správnych uhlov ktoré sa nachádzali smerom dovnútra elementu.

Vektor n_1 má hodnotu $n_1 = n/|n|$ vo všetkých ostatných situáciách.

Výsledný smerový vektor ktorý je priradený asociačnému bodu má tvar $n_1 \cdot FDe(\alpha, f(\alpha))$. Súčet vektorov síl funkcie FDa všetkých asociačných bodov vrcholu pôsobí v jeho ťažisku a popisuje, ktorým smerom sa má vrchol pohnúť aby sa jeho uhly, ktoré tvorí so svojimi hranami približovali hodnotám 0 alebo 45°. Celková energia grafu RNA motívu kontrolovaného funkciou FDa je Ea . Je to vlastne súčet celkových síl počítaných funkciou FDa všetkých vrcholov. Nulová hodnota Ea zaručuje, že všetky vrcholy zvierajú medzi sebou uhly 0 alebo 45°. Je zrejmé, že nulová hodnota Ea nemôže nastať v prípadoch, kedy graf motívu obsahuje viacero vrcholov, ktoré musia byť pre prehľadnosť rôzne orientované. No minimálna hodnota Ea udrží takéto grafy v pravidelných tvaroch (viď. Obrázok 3.21).



Obrázok 3.21: Príklad pôsobenia sily FDa .

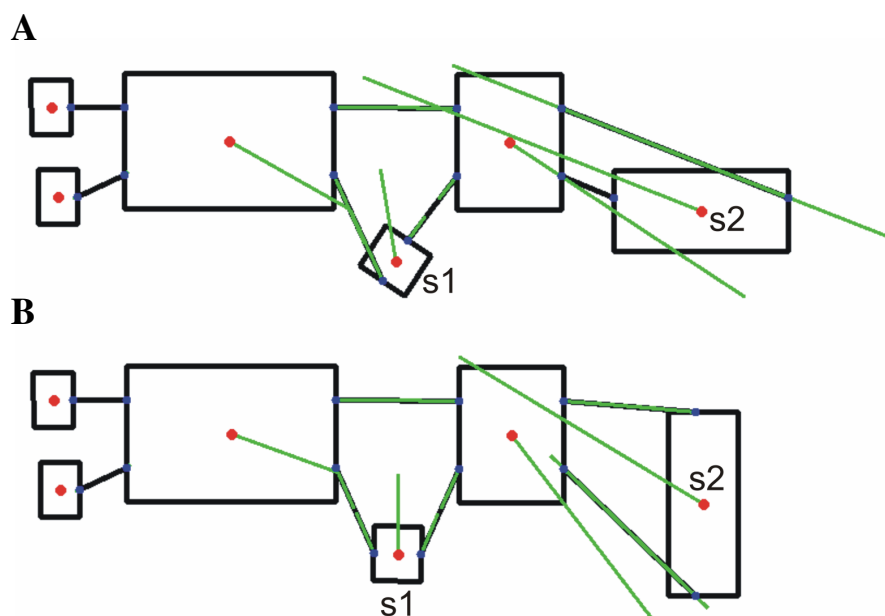
Červené čiary vyjadrujú smer pôsobenia sily FDa z konkrétneho bodu. Dĺžka čiary je potom veľkosť tejto sily. V asociačných bodoch pôsobia sily tak aby vrcholy zvierali správne uhly. V ťažisku vrcholov je výsledná sila ktorá na vrchol pôsobí.

3.2.4. Automatická rotácia SSE

Keďže každý vrchol obsahuje niekoľko asociačných bodov, ktoré majú jeho štruktúre pevné miesta a súčasne obsahuje aj svoj predpis, ktorý má taktiež svoju pevnú orientáciu v rámci vrcholu, záleží na tom ako bude celý vrchol orientovaný respektíve rotovaný. Aby bolo možné dosiahnuť rozloženie grafu do 2D priestoru kresliacej plochy a aby boli jednotlivé vrcholy natočené smerom k svojim susedom, je nutné zaviesť funkciu, ktorá bude vrchol automaticky správne otáčať. Počas behu algoritmu potom bude táto funkcia z času na čas jednotlivé vrcholy natáčať správnym smerom. Tým bude zabezpečené, že vrcholy budú počas behu algoritmu orientované tak, aby ich asociačné body smerovali k ich susedom. Jednotlivé funkcie FDn , FDe a FDa budú za týchto podmienok počítať sily, ktoré posunú vrchol správnym smerom, pretože sa otáčaním minimalizuje aj možnosť zauzlenia vrcholov. Graf RNA motívu obsahuje dva typy vrcholov, ktoré majú rozdielny tvar. Aby sa dosiahlo čo najlepšie natočenie je potrebné otáčať SSE odlišne ako hélix alebo reláciu. Je to zapríčinené aj tým, že sa na celkovej orientácii častí grafu RNA motívu podieľajú hlavne elementy hélix a relácia, ktoré majú štyri asociačné body. SSE sa potom ako jednoduchší element s dvoma asociačnými bodmi ľahšie prispôsobuje susedom.

Ako bolo naznačené v sekcii 3.2.1 výpočtu funkcie FDn sily susedov budú priamo súvisieť s natáčaním vrcholu. Aby bol SSE orientovaný smerom k svojim susedom musia ho otáčať jeho vlastné pružiny. Je ale nutné aby príťažlivá sila pružiny nebola zhora ohraničená konštantou. V hraničných prípadoch, kedy by obidve pružiny vrcholu pôsobili rovnakou teda maximálnou silou na asociačné body v rôznych orientáciách, rotácia by bola nesprávna. Z toho dôvodu bola funkcií FDn priradená kvadratická neohraničená funkcia $f(k)$ namiesto zhora aj zdola ohraničenej funkcie pre $f(k)$ sigmoid.

Samotný výpočet správneho uhlu rotácie vrcholu prebieha v kontrolovaní celkovej sily počítanej funkciou FDn , ktorá na vrchol pôsobí v jeho ťažisku. To znamená, že pre každý uhol v rozmedzí 0 až 359° sa aplikuje výpočet síl jeho pružín. Výsledný uhol rotácie je potom taký, pri ktorom bola celková sila pružín pôsobiaca v ťažisku vrcholu minimálna (viď. Obrázok 3.22).



Obrázok 3.22: Príklad rotácie SSE.
 (A) Elementy s1 a s2 pred rotáciou. (B) Elementy s1 a s2 po aplikovaní automatickej rotácie.

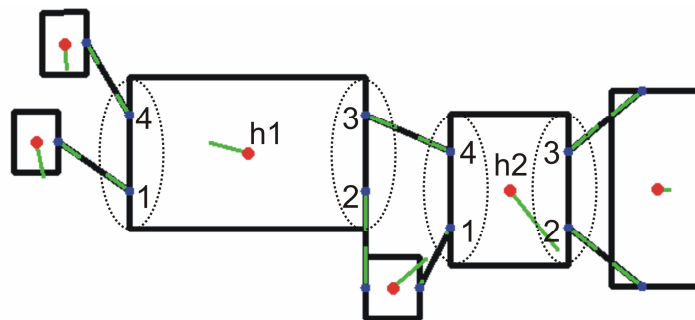
Týmto bol nájdený uhol pri ktorom na vrchol pôsobí minimálna sila pružín. Otočenie ale nebude realizované o tento uhol, pretože je nutné ponechať čas aby sa ostatné vrcholy pri svojich rotáciách jemu prispôbili. Postupne po iteráciách algoritmu sa bude vrchol otáčať približujúc k uhlu, kde pôsobí minimálna sila pružín. To zabezpečí hladšiu transformáciu grafu do jeho konečnej podoby. Zároveň tým funkcia predchádza situáciám, kedy vrchol neustále osciluje, pričom každá malá zmena pozícií susedných vrcholov vyvoláva veľkú zmenu rotácie vrcholu.

3.2.5. Automatická rotácia elementov hélíx a relácia

Rotácia elementov hélíx a relácie nemôže prebiehať rovnako ako pri SSE pretože obsahujú nie dva ale štyri asociačné body a to znamená, že na ich ťažisko pôsobia štyri nezávislé pružiny. Teda nastávajú situácie, kedy sa účinky jednotlivých síl pružín navzájom rušia a otočenie vrcholu o uhol nájdený pri minimálnej sile pôsobiacej na ťažisko nezodpovedá požadovanému stavu. Zároveň sa takýto element veľmi ťažko prispôbuje susedným vrcholom. Keďže elementy hélíx a relácia v grafe RNA motívu udávajú orientáciu celých

častí zložených z viacerých menších elementov ich správna rotácia je veľmi dôležitá. Sú dva prístupy ako zabezpečiť správnu rotáciu. Prvý prístup zohľadňuje orientáciu v okolí susedov tým, že sa podľa istých pravidiel súvisiacich práve so susedmi nájde správny uhol a samotný vrchol sa rovnako ako SSE po jednotlivých iteráciách rotuje. Druhý prístup taktiež zohľadňuje orientáciu v okolí susedov tým, že sa podľa istých pravidiel súvisiacich práve so susedmi nájde správny uhol no rotácia neprebieha iba vo vrchole, ale rotujú sa celé úseky ktoré zviaza kontrolovaný hélix alebo relácia.

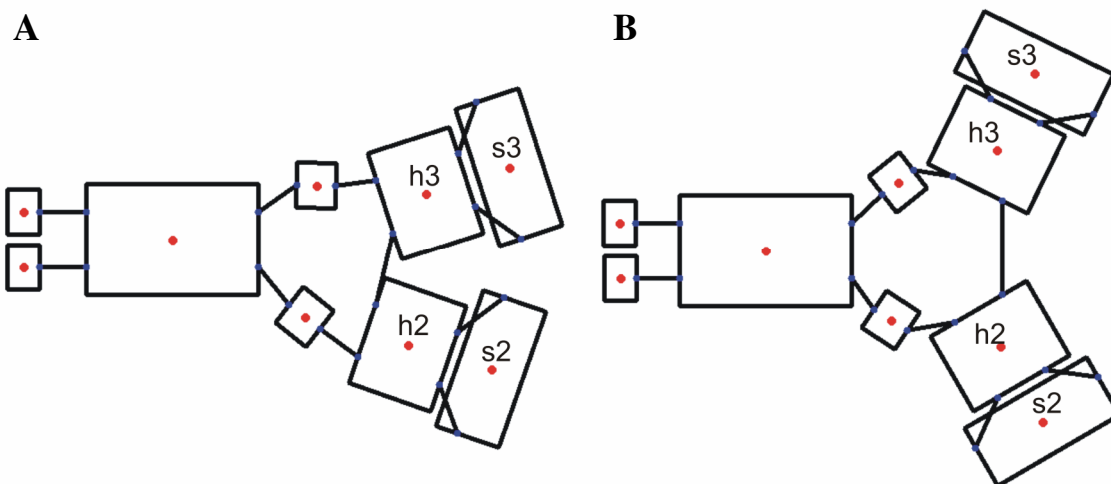
Pravidlá nájdenia správneho uhla sú v oboch prístupoch rovnaké. Líšia sa teda iba konečnou rotáciou. Prístup v rotácií SSE kedy sa správny uhol hľadal podľa síl dvoch pružín, ktoré naň pôsobia sa nemení. V elemente hélix alebo relácie sa nezávisle na sebe kontrolujú dvojice pružín vertikálnych asociačných bodov. To znamená, že prvú dvojicu tvorí v poradí prvý a štvrtý a druhú dvojicu tvorí v poradí druhý a tretí asociačný bod. Na každú dvojicu pôsobí výsledná sila spojením ich dvoch príslušných pružín (viď. Obrázok 3.23).



Obrázok 3.23: Príklad párovania asociačných bodov pre rotáciu elementov hélix alebo relácie. Rotuje sa podľa výslednej sily pružín príslušajúcich dvojici.

Vrcholu bude rovnako ako v prípade rotácie SSE postupne skúšaný uhol od 0 po 359° a pre každú dvojicu bude zaznamenaný ten uhol, kedy dosiahla minimálnu výslednú silu svojich dvoch pružín. Uhol tej dvojice, ktorý je najďalej od pôvodného uhlu v ktorom vrchol bol pôvodne je potom výsledný uhol rotácie hélixu alebo relácie.

Rotovať potom vrchol je možné rovnakým spôsobom ako v SSE, tým že sa postupne po iteráciách algoritmu bude vrchol otáčať približujúc k správnejmu uhlu. Lokálne tento prístup zabezpečuje vhodnú orientáciu vrcholu. Nastávajú ale situácie, kedy pružiny na pravej strane hélixu alebo relácie nedovolia dostatočne rotovať vrchol aj keď by to z vizuálneho hľadiska jeho ľavej strany bolo potrebné (viď. Obrázok 3.24).

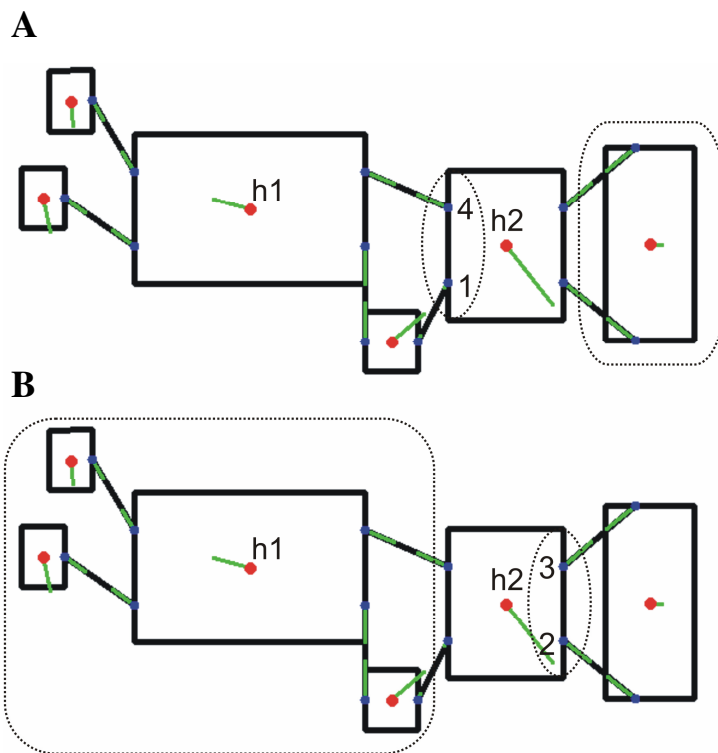


Obrázok 3.24: Rotácia jedného elementu a rotácia celej časti grafu.

(A) Príklad situácie kedy automatická rotácia elementu rotuje iba element samotný. Časti grafu ktoré prislúchajú vrcholom h2 a h3 sú nedostatočne rozostúpené. (B) Príklad situácie, kedy automatická rotácia elementu rotuje aj celú časť grafu, ktorú zvierá. Rozstupy častí prislúchajúcich elementom h2 a h3 sú dostatočné. A tým je aj celý graf pravidelnejšej štruktúry ako v obrázku (A).

Globálne preto rotovať iba samotný vrchol nestačí pretože elementy hélix a relácia udávajú orientáciu celej časti, ktorú zvierajú. Rotovať celú časť ktorú zvierá hélix alebo relácia je potrebné z hľadiska celého grafu RNA motívu.

Tento prístup v jednotlivých iteráciách zabezpečí to, že sa rotovaním hélixu alebo relácie neporušia už správne umiestnenia vrcholov nachádzajúcich sa na strane elementu kde k rotácii nemá dôjsť. Rovnako ako v prvom prístupe rotovania hélixu a relácie je prvej dvojici zloženej s prvého a štvrtého a druhej dvojici zloženej z druhého a tretieho asociačného bodu (viď. Obrázok 3.23) počítaný uhol, kedy sú ich výsledné sily priľahlých pružín najmenšie. Rozdiel je, že sa uhol nepočíta súčasne obidvom dvojiciam. Na začiatku sa náhodne vyberie dvojica, podľa ktorej sa bude element aj s prislúchajúcou časťou ktorú zvierá rotovať. Tej dvojici sa vypočíta správny uhol, kedy majú priľahlé pružiny výslednú silu minimálnu. Uhol, o ktorý sa má element s jeho prislúchajúcimi vrcholmi rotovať je rozdiel pôvodného uhla s uhlom vyrátaným. Ak ide o dvojicu zloženú z prvého a štvrtého asociačného bodu rotuje sa počítaný hélix alebo relácia a všetky vrcholy nachádzajúce sa napravo od neho (viď. Obrázok 3.25 A). Ak ide o dvojicu zloženú z druhého a tretieho asociačného bodu rotuje sa počítaný hélix alebo relácia a všetky vrcholy nachádzajúce sa naľavo od neho (viď. Obrázok 3.25 B).



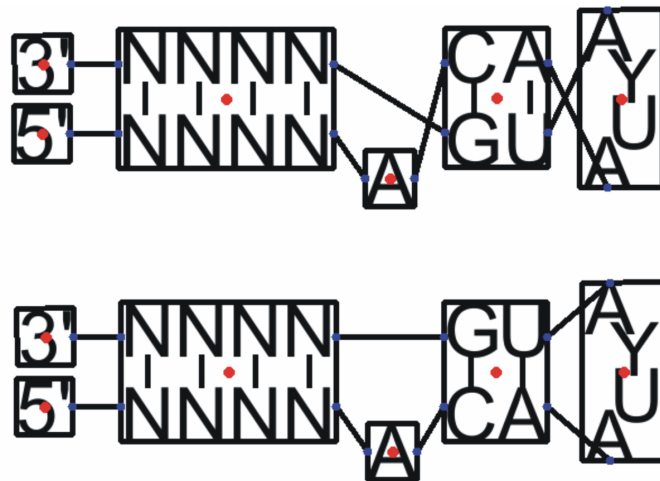
Obrázok 3.25: Rotácia celých častí prislúchajúcich elementu h2.
 (A) Prislúchajúca časť grafu dvojici asociačných bodov 1 a 4. (B) Prislúchajúca časť grafu dvojici asociačných bodov 2 a 3.

Všetky rotácie sa vykonávajú vzhľadom na ťažisko počítaného hélíxu alebo relácie. Výsledný efekt tohto prístupu je, že sa v iteráciách nemenia orientácie celých častí motívu, ktoré sú zvierané hélíxami alebo reláciami. Z tohto dôvodu nie je nutné rotovať postupne po iteráciách pretože, ak rotácia na ľavej strane vylepší stav, na pravej strane sa stav nezhorší a naopak. Po skončení behu algoritmu je potom výsledný motív podstatne pravidelnejší (viď. Obrázok 3.24 B).

3.2.6. Zrkadlové otočenie hélíxu a relácie

V štruktúre vstupného deskriptoru podľa ktorého sa graf RNA motívu zostrojil (viď. Sekcia 1.1), nie je žiadna geometrická informácia o polohe alebo orientácií jednotlivých elementov. Teda štruktúry s dvoma komplementárnymi vláknami ako hélíx a relácia nemajú informáciu o tom, ktoré z vlákien bude v rámci vrcholu v jeho hornej alebo dolnej časti. Z hľadiska nukleotidových väzieb medzi jednotlivými elementmi sa v podstate štruktúra RNA motívu nemení, ak si vlákna v rámci vrcholu aj so svojimi prislúchajúcimi asociáciami vymenia pozície. No z hľadiska vykresľovania grafu RNA motívu je nutné tieto situácie rozlíšiť aby nedochádzalo k prekriženiu hrán asociácií a tým aj k neestetickému vykresleniu celej štruktúry (viď. Obrázok 3.26 A).

Pri základnom načítaní grafu RNA motívu sú všetkým elementom typu hélix alebo relácia priradené pozície vlákien tak, že prvý predpis vlákna ktorý sa nachádza pred dvojbodkou bude kreslený v dolnej časti vrcholu a predpis druhého vlákna nachádzajúci sa za dvojbodkou bude kreslený v hornej časti. Funkciou zrkadlového otočenia je potom možné ich pozície vymeniť pričom si zachovajú svoje asociačné body aj orientáciu v rámci vrcholu (vid'. Obrázok 3.26).



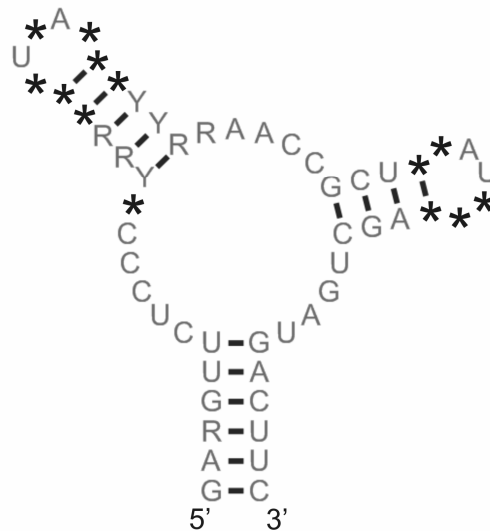
Obrázok 3.26: Príklad nutnosti zrkadlového otočenia elementov hélix alebo relácie.

Táto funkcia ale nemôže byť aplikovaná postupne na vrcholy počas behu programu ako je to pri rotácií, pretože dochádza k situáciám, kedy sa sily pružín nemenia ak si vlákna vo vrchole vymenia pozície. Tento jav zapríčiniť, že sa po skončení behu algoritmu vo výslednom grafe RNA motívu nachádzajú elementy hélix a relácia, ktorým sa ich vlastné asociačné hrany krížia. Automaticky zrkadlovo otáčať vrcholy podľa síl pružín preto nie je možné. Funkciu bude môcť pomocou užívateľského rozhrania aplikovať na vrcholy samotný používateľ po zastavení behu algoritmu, ktorý navrhne výzor RNA motívu.

3.2.7. Udržiavanie vrcholov typu SSE v kružniciach

Jednou z podmienok správne vykresleného RNA motívu je udržiavanie vrcholov typu SSE v kružniciach. Ide o také vrcholy, ktoré sú umiestnené medzi hélixami alebo reláciami. Táto podmienka spĺňa estetické kritérium kedy vrcholy typu SSE spájajú konce hélixov a relácií tak aby bol konečný graf RNA motívu čitateľný a aby bolo jasné ktoré vrcholy sú asociované a tým aj ako sa celý motív kóduje. Potreba vytvárať kružnice vychádza aj z podmienky kreslenia motívu na 2D plochu, kde sú jednotlivé časti rozdielne orientované .

Spojenie týchto častí pomocou kružníc je estetické a zároveň prehľadné (viď. Obrázok 3.27).



Obrázok 3.27: Príklad potreby udržiavania elementov SSE v kružniciach.
 Vďaka kružnici ktorú vytvárajú nespárované úseky je vykreslený RNA motív estetický a prehľadný.

Vrchol je teda potrebné v behu programu modifikovať tak, aby vzhľadom na jeho susedov menil svoj vonkajší tvar, ktorý sa prispôsobí vnútornému predpisu už kreslenému do kružnice. Z toho dôvodu sa bude vrcholu meniť výška a šírka podľa toho koľko priestoru potrebuje na svoje vykreslenie vnútorný predpis. Asociačné boby musia byť umiestnené tak ako pred transformáciou, čiže na začiatku a na konci predpisu SSE. Pri takto definovaných podmienkach všetky funkcie ako FDn, FDe a FDa budú fungovať bez zmeny.

Samotná transformácia vrcholu typu SSE má niekoľko samostatných výpočtov. Nájdenie kružnice $k = (r, s(x_s, y_s))$, podľa ktorej sa bude predpis kresliť, výpočet presných pozícií na kružnici pre jednotlivé znaky a prispôbenie tvaru vrcholu podľa veľkosti vnútorného predpisu. V prvom rade je nutné vypočítať polomer kružnice a až následne pozíciu stredu. Ako dobrý prístup sa ukázalo sledovanie pozície susedov tým, že polomer je možné vypočítať vzhľadom na uhol, ktorý zvierajú prilahlé asociačné hrany vrcholu. Tento uhol bude počítaný ako uhol dvoch vektorov u_1 a u_2 . Ak a_{i1} je prvý asociačný bod vrcholu v_i a jeho prislúchajúci asociačný bod je a_{kl} s ktorým má spoločnú asociačnú hranu, a_{i2} je druhý asociačný bod vrcholu v_i a jeho prislúchajúci asociačný bod s ktorým má spoločnú asociačnú hranu je bod a_{mn} a v_i je vrchol, ktorý sa bude transformovať, tak $u_1 = a_{kl} - a_{i1}$ a $u_2 = a_{mn} - a_{i2}$ (viď. Obrázok 3.28).



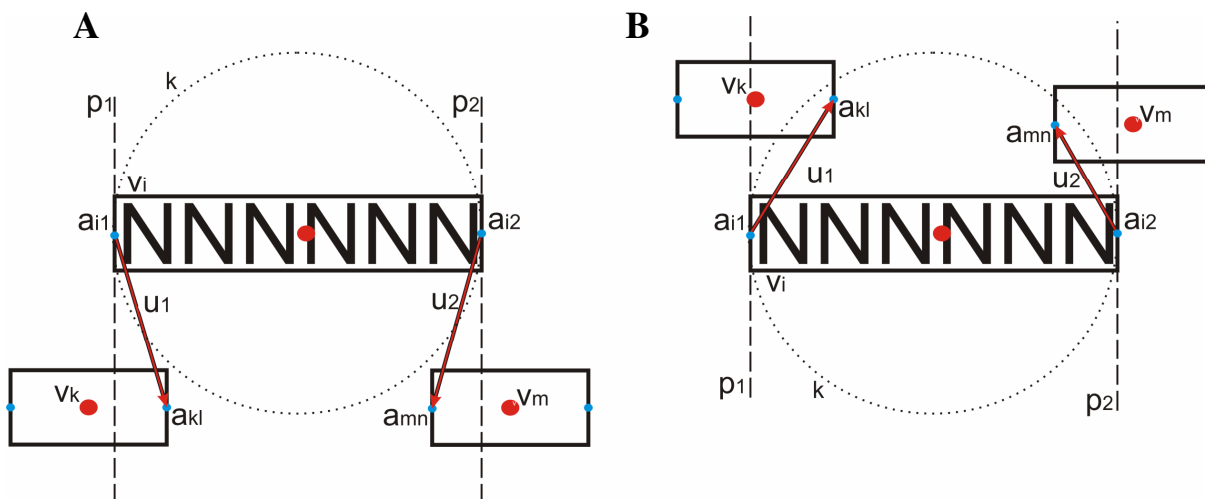
Obrázok 3.28: Hľadaný uhol alfa podľa ktorého sa nájde kružnica k je uhol vektorov u_1 a u_2 .

Uhol α ktorý zvierajú vektory u_1 a u_2 je potom počítaný pomocou kosínusovej vety rovnako ako to bolo pri výpočte funkcie FDA (viď. Sekcia 3.2.3). Funkcia $f(\alpha)$ následne vypočíta veľkosť polomeru hľadanej kružnice $r = f(\alpha)$. Ak vektory u_1 a u_2 smerujú priamo od seba, to znamená, že vrchol je v jednej línii ako jeho susedia, tak sa mu vráti pôvodný tvar aký mal pri inicializácii. Teda ide o uhol $\alpha = 180^\circ$. Takýto uhol je ale v reálnom výpočte zriedkavý preto je vrcholu navrátený pôvodný tvar ak uhol $\alpha > 170^\circ$.

Aby bolo možné umiestniť kružnicu, k tak aby na nej ležali asociačné body vrcholu polomer nesmie presahovať hodnotu polovice jeho šírky. Potom funkcia $f(\alpha)$ je zhora ohraničená touto hodnotou a je definovaná

$$f(\alpha) = \sin\left(0.5 \cdot \text{rad}(\alpha) \cdot v_i \text{šírka} + \frac{v_i \text{šírka}}{2}\right)$$

Výpočet polomeru r v závislosti od uhla vektorov u_1 a u_2 je nutné rozšíriť o prípad, kedy je iste, že polomer kružnice neklesne pod hodnotu polovice šírky vrcholu. Ide o prípad, kedy vektory smerujú k sebe tak, že sa ich koncové body nachádzajú priamo nad alebo pod vrcholom (viď. Obrázok 3.29). Kontrolovať je to možné pomocou orientácie susedných asociačných bodov a_{kl} a a_{mn} vzhľadom na priamky kolmé na úsečku a_{i1} a a_{i2} , ktoré pretínajú práve body a_{i1} a a_{i2} (viď. Obrázok 3.29).



Obrázok 3.29: Prípady kedy je polomer kružnice rovný polovičnej šírky elementu.

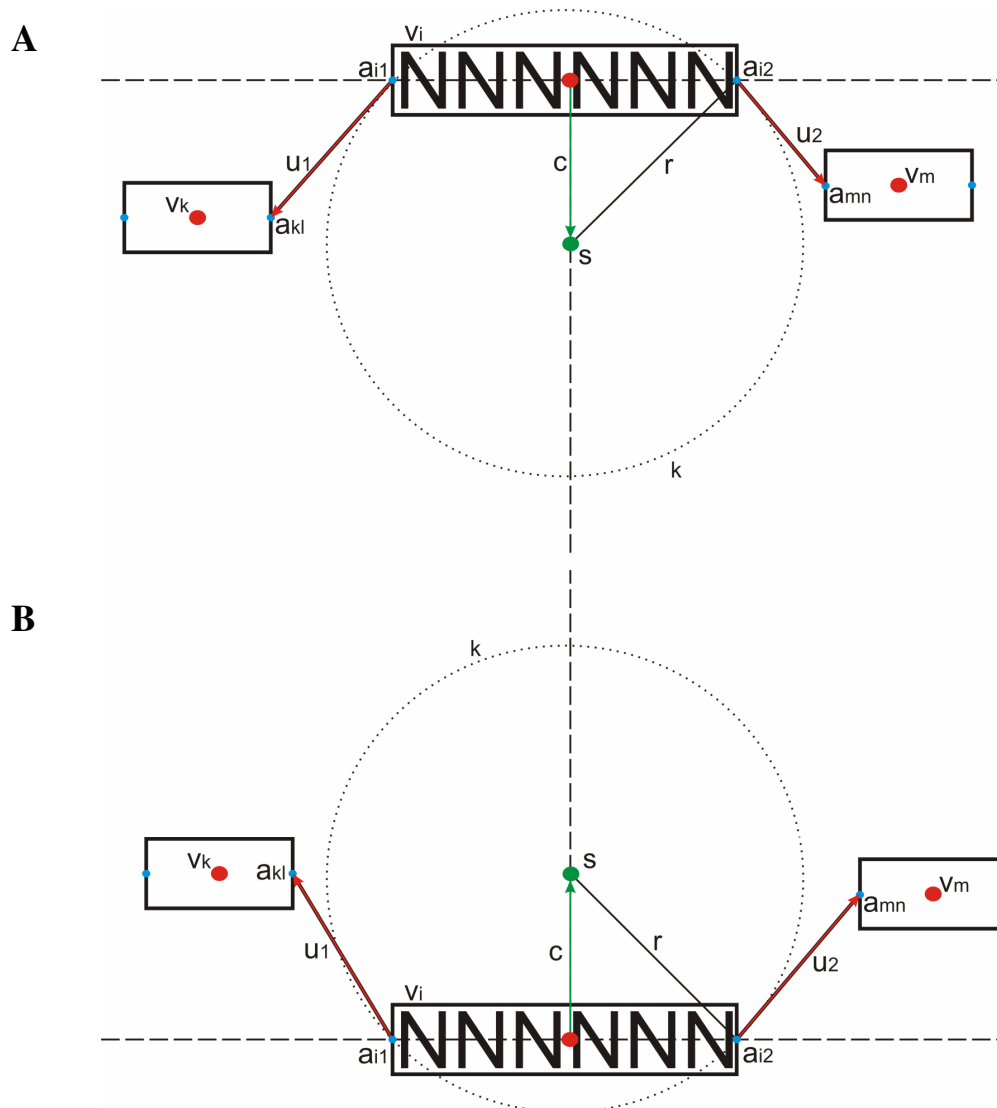
Ide o vzájomnú polohu priamky $p1$ a bodu a_{kl} a priamky $p2$ a bodu a_{mn} . (A) Prípady, kedy je orientácia susedných vrcholov nad priamkou na ktorej ležia body a_{i1} a a_{i2} . (B) Prípady, kedy je orientácia susedných vrcholov pod priamkou na ktorej ležia body a_{i1} a a_{i2} .

Orientácia je počítaná rovnako pomocou rovnice priamky ako vo FDa (vid'. Sekcia 3.2.3).

Ak sa výsledné hodnoty po dosadení bodov a_{kl} a a_{mn} do rovníc im prislúchajúcich priamok líšia tým, že sú kladné alebo záporné, tak je možné určiť, že došlo k tomuto prípadu. Funkcia $f(\alpha)$ sa vtedy neaplikuje a r je potom práve polovica šírky vrcholu.

Aby kružnica pretínala vrchol v rovnakých miesta na pravej a ľavej strane je zrejmé, že stred $s(x_s, y_s)$ sa bude nachádzať na polpriamke smerujúcej z ťažiska vrcholu, ktorá je kolmá na úsečku $a_{i1}a_{i2}$. Dôležitá je ale orientácia bodov a_{kl} a a_{mn} vzhľadom na $a_{i1}a_{i2}$. Podľa nej sa bude stred nachádzať na polpriamke buď nad alebo pod touto úsečkou. Ak po dosadení bodov a_{kl} a a_{mn} do rovnice priamky, na ktorej ležia body a_{i1} a a_{i2} budú ich výsledné hodnoty kladné, pôjde o polpriamku nad úsečkou $a_{i1}a_{i2}$ a ak záporné hodnoty, tak pôjde o polpriamku pod úsečkou $a_{i1}a_{i2}$ (vid'. Obrázok 3.30). Je nutné preto vypočítať ako ďaleko je bod s vzdialený od ťažiska vrcholu. Keďže veľkosť polomeru r ako aj šírka w vrcholu sú známe veličiny a je zrejmé, že polpriamka je kolmá na úsečku $a_{i1}a_{i2}$ vzdialenosť stredu kružnice s od ťažiska vrcholu, je potom možné vypočítať pomocou Pytagorovej vety (vid'. Obrázok 3.30).

$$c = r^2 - \left(\frac{w}{2}\right)^2$$



Obrázok 3.30: Umiestenie kružnice k .

Súradnice bodu s sú počítané ako vzdialenosť od ťažiska vrcholu o veľkosti odvesny c pravouhlého trojuholníka zloženého z bodov s , ťažiska vrcholu a asociačného bodu $ai2$. (A) Orientácia stredu s ak sú susedné vrcholy vk a vm orientované pod priamkou na ktorej ležia asociačné body $ai1$ a $ai2$. (B) Orientácia stredu s ak sú susedné vrcholy vk a vm orientované nad priamkou na ktorej ležia asociačné body $ai1$ a $ai2$.

Ďalej sa počítajú smerový vektor úsečky $a_{i1}a_{i2}$ je $u = a_{i2} - a_{i1}$, kolmý vektor na u je $n = (-u_2, u_1)$ a jednotkový vektor n je $n_1 = n/|n|$. Ak sa stred nachádza pod priamkou na ktorej ležia body a_{i1} a a_{i2} jednotkový vektoru n je $n_1 = -n/|n|$. Potom pozícia stredu kružnice s je $s = (n_1 \cdot c) + v_i$. Teda pozícií ťažiska vrcholu v_i sa pripočíta zväčšený vektor n_1 o hodnotu c .

Ak je známy polomer a stred kružnice je už možné hľadať pozície jednotlivých znakov predpisu. V prvom rade je potrebné určiť riadiaci bod X na kružnici a pomocou jeho rotácií okolo bodu s nájsť všetky hľadané pozície. Bod X je vlastne priesečník polpriamky

smerujúcej z ťažiska vrcholu kolmej na úsečku $a_{i1}a_{i2}$, ktorá je vzhľadom na túto úsečku opačne orientovaná ako bod s s kružnicou k (viď. Obrázok 3.31). Rotáciou tohto bodu rovnako kladným a záporným smerom okolo bodu s v počte zodpovedajúcom počtu znakov v predpise sa nájdu hľadané pozície. Pre hľadané súradnice x a y sú vzorce rotácie bodu X okolo bodu s o uhol λ takéto

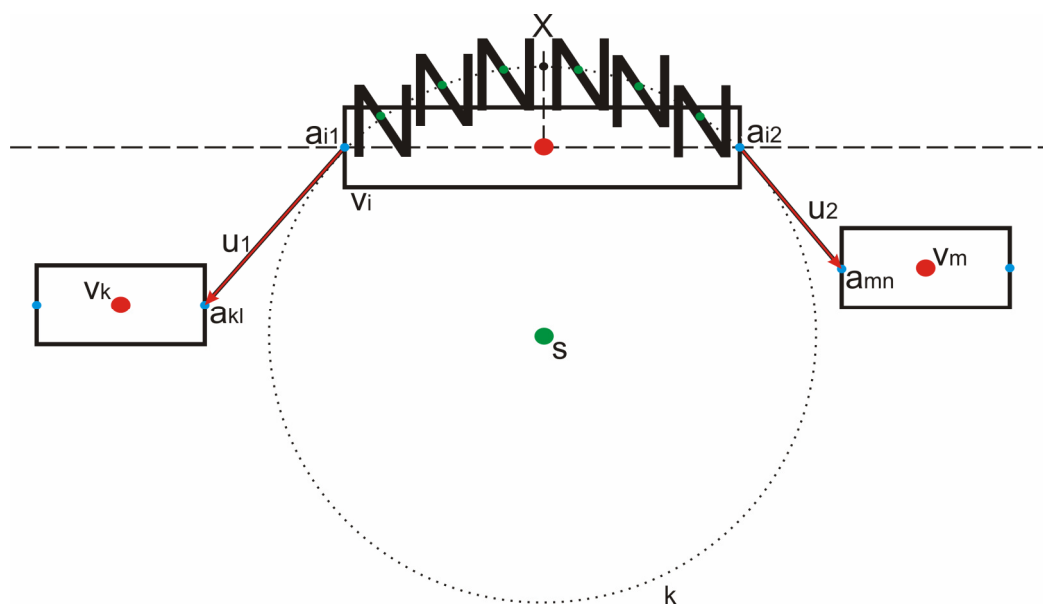
$$x = s_x + ((X_x - s_x) \cdot \cos(\text{rad}(\lambda)) - (X_y - s_y) \cdot \sin(\text{rad}(\lambda)))$$

$$y = s_y + ((X_y - s_y) \cdot \cos(\text{rad}(\lambda)) + (X_x - s_x) \cdot \sin(\text{rad}(\lambda)))$$

Aby bolo možné nájsť všetky pozície je nutné uhol λ pre každý znak meniť, pričom nesmie dohádzať k ich prekryvaniu. Uhol rotácie medzi dvoma pozíciami kedy šírka jedného znaku je m je

$$\beta = \frac{360 \cdot m}{2\pi r}$$

Výpočet pozícií znakov prebieha postupne. V prípade že je nepárny počet znakov, znak nachádzajúci sa v strede predpisu má pozíciu bodu X . Následne pre body nachádzajúce sa napravo od stredového znaku je bod X rotovaný najprv o uhol $\lambda = \beta$ a potom postupne pre každý ďalší znak sa λ zvýši o β . Pre znaky naľavo od stredového znaku je bod X rotovaný najprv o uhol $\lambda = 360 - \beta$ a potom postupne pre každý ďalší znak sa λ zníži o β . Ak je párný počet znakov v predpise sú stredové znaky dva. Pravému znaku prislúcha rotácia bodu X o $\lambda = \beta/2$ a postupne pre ďalšie znaky napravo od neho sa λ zvýši o β . Ľavému znaku prislúcha rotácia bodu X o $\lambda = 360 - \beta/2$ a postupne pre ďalšie znaky naľavo od neho sa λ zníži o β (viď. Obrázok 3.31).

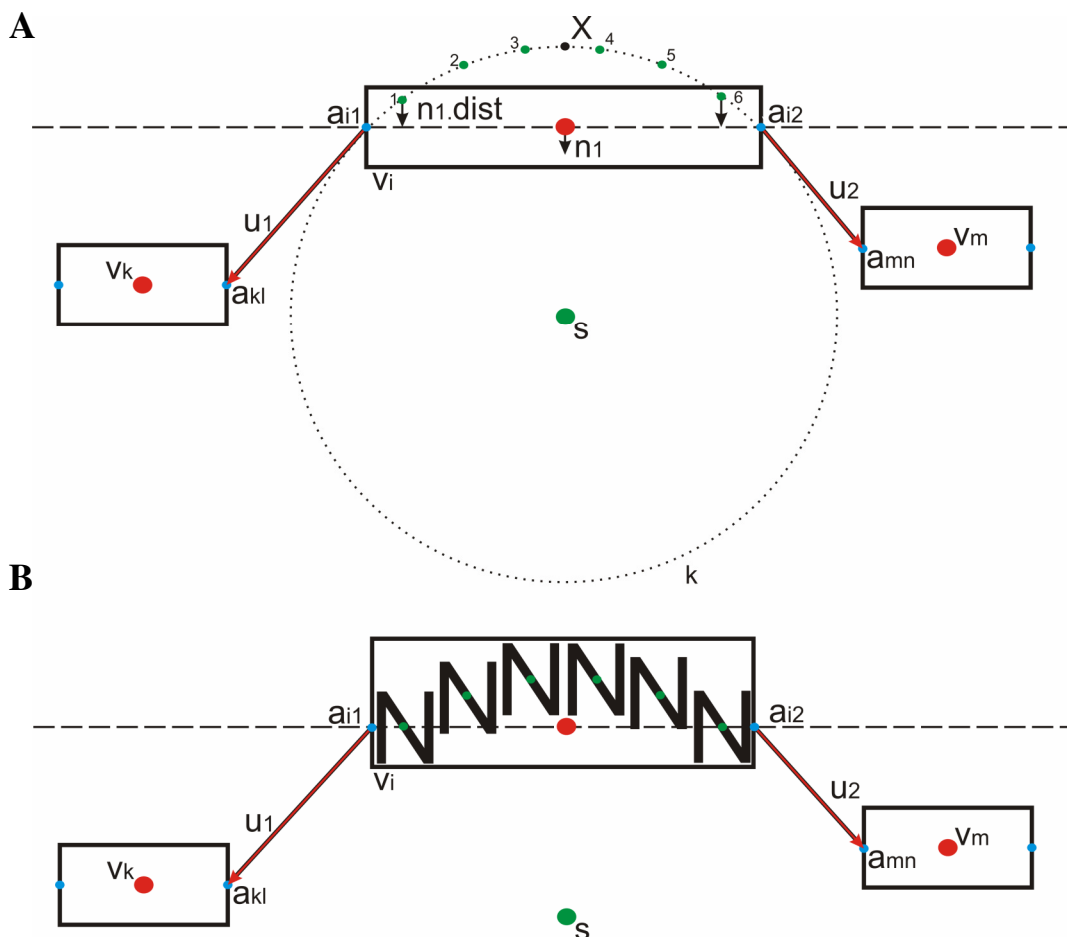


Obrázok 3.31: Príklad nájdenia riadiaceho bodu X a všetkých pozícií znakov predpisu. Bod X je opačne orientovaný ako stred kružnice k a je jej priesečníkom s polpriamkou kolmou na priamku na ktorej ležia body a_{i1} a a_{i2} . Následne pozície všetkých znakov sú rotáciami bodu X okolo bodu s .

Na Obrázok 3.31 je si možné všimnúť, že sa pozície znakov nachádzajú mimo vrcholu. Pretože je nutné minimalizovať pohyb asociačných bodov vo výpočte funkcií FDn , FDe a FDe z dôvodu aby posun vrcholov podľa síl nebol týmto pohybom narúšaný je nutné posunúť smerom k vrcholu vypočítané pozície znakov predpisu. Smerový vektor posunu je práve jednotkový vektor n_1 ktorým je kolmý na úsečku $a_{i1}a_{i2}$ a je orientovaný smerom k stredu kružnice k . Tento vektor je predĺžený o vzdialenosť o ktorú sa musia pozície znakov posunúť (viď. Obrázok 3.32 A). Je to vzdialenosť najbližšej pozície k úsečke $a_{i1}a_{i2}$. Ide o prvý alebo posledný znak predpisu (viď. Obrázok 3.32 A). Vzdialenosť je potom počítaná ako vzdialenosť pozície prvého respektíve posledného znaku predpisu z_1 od priamky na ktorej ležia body a_{i1} a a_{i2} . Rovnica priamky na ktorej ležia body a_{i1} a a_{i2} je $n_{11}x + n_{12}y + c = 0$, potom vzdialenosť bodu z_1 od priamky $n_{11}x + n_{12}y + c = 0$ je

$$dist = \frac{|n_{11}z_{1x} + n_{12}z_{1y} + c|}{\sqrt{n_{11}^2 + n_{12}^2}}$$

Súradniciam všetkých pozícií znakov je následne pripočítaný vektor $dist \cdot n_1$ (viď. Obrázok 3.32 A).



Obrázok 3.32: Príklad posunu pozícií znakov do vnútra vrcholu a prispôsobenie jeho rozmerov predpisu.
 (A) Nájdenie smeru posunu pozícií $n1$ ktorý je kolmý na úsečku $ai1ai2$ a zároveň je orientovaný smerom k stredu kružnice k . Každá pozícia znaku je potom posunutá podľa $n1$ o vzdialenosť ktorá sa rovná vzdialenosti prvého alebo posledného znaku predpisu od priamky na ktorej ležia body $ai1ai2$. (B) Prispôsobenie rozmerov vrcholu podľa predpisu.

Dôležitou súčasťou transformácie vrcholu SSE je prispôsobenie jeho výšky a šírky predpisu, ktorý je už kreslený podľa kružnice. Šírka vrcholu w_n je počítaná ako vzdialenosť medzi prvým a posledným znakom. Teda ak je pozícia prvého znaku z_1 a druhého znaku z_2 , vzdialenosť medzi nimi je $w_n = |z_2 - z_1|$. Keďže pozícia vlastne znamená ťažisko v tomto prípade je nutné pripočítať pravej aj ľavej strane vrcholu polovicu šírky jedného znaku. Potom šírka vrcholu je $w_n = |z_2 - z_1| + m$. Výška vrcholu je vzdialenosť bodu X od priamky na ktorej ležia body z_1 a z_2 . Ak je normálový vektor rovnice priamky na ktorej ležia body z_1 a z_2 vektor n_z , rovnica priamky je $n_z x + n_z y + c = 0$, potom vzdialenosť bodu X od priamky $n_z x + n_z y + c = 0$ je

$$distX = \frac{|n_{z1} X_x + n_{z2} X_y + c|}{\sqrt{n_{z1}^2 + n_{z2}^2}}$$

Tu je potrebné pripočítať výške polovičnú výšku znaku h pre hornú aj dolnú časť vrcholu. Potom výsledná výška vrcholu je $distX + h$ (viď. Obrázok 3.32 B).

Tento prístup zabezpečí, to že sa každý vrchol typu SSE prispôsobí svojim susedným vrcholom. Teda v závislosti od uhlov, ktoré navzájom zvierajú je element ohýbaný. V celkovom zobrazení grafu RNA motívu vrcholy SSE vytvárajú kruhovitú útvary a na miestach kde sú v jednej línii ako ich susedia ostanú v pôvodnom tvare ako pri inicializácii.

3.2.8. Beh algoritmu

Každá z funkcií FDn , FDa , FDe , rotácie, zrkadlové otočenie a udržiavanie vrcholov typu SSE v kružniciach sú v predchádzajúcich kapitolách opísané ako samostatné jednotky, ktoré riadia vykresľovanie RNA motívu. Na to aby boli dodržané požiadavky pre správne vykreslený RNA motív, musia byť tieto funkcie postupne aplikované na graf v správnom čase. Riadenie behu algoritmu, kedy sa graf transformuje až do výslednej podoby zabezpečujú hodnoty výsledných energií En , Ea a Ee silových funkcií FDn , FDa a FDe . Postupne v každej iterácii sa kontrolujú ich aktuálne hodnoty pričom sa uchováva informácia o minimálnej hodnote akú každá z energií dovtedy dosiahla. Beh algoritmu končí, keď už istý počet iterácií minimálne hodnoty energií neklesajú.

Po načítaní a vytvorení grafovej štruktúry zo vstupného deskriptoru je nutné priradiť jednotlivým vrcholom počiatočné pozície. V metóde kreslenia grafov pomocou pružín (viď. Sekcia 2.1). sú vrcholom na začiatku behu algoritmu priradené náhodné pozície pretože na ne pôsobí iba jeden typ sily a to sila pružín. Vtedy nezáleží na prvotnom rozmiestení vrcholov na kresliacej ploche. No ak v grafe pôsobí viacero odlišných síl, podľa ktorých sú počas behu algoritmu vrcholy posúvané, môže sa stať, že tieto sily budú navzájom brániť vrcholom v posune na správne miesta a tým môže dôjsť k zauzleniu štruktúry. Prvotné rozmiestnenie preto nesmie byť náhodné.

Počiatočné pozície vrcholov sú priraďované postupne podľa toho, ako sú za sebou radené elementy v mape deskriptoru. Vzdialenosti medzi vrcholmi musia byť čo najväčšie aby mali počas behu algoritmu dostatočný priestor na pohyb k správnym pozíciám. Následne je aplikovaná funkcia udržiavania vrcholov typu SSE v kružniciach. Aplikuje sa aj počas behu algoritmu neustále po každom pohybe vrcholu na susedné vrcholy typu SSE a to ešte pred aplikáciou silových funkcií. Ešte pred spustením behu algoritmu sú vypočítané výsledné sily pôsobiace na ťažiská vrcholov funkcií FDn , FDa a FDe . Ak je graf zostrojený, vrcholy majú prvotné pozície, funkcia udržiavania vrcholov typu SSE

v kružniciach je aplikovaná na každý prislúchajúci typ a ťažiská vrcholov majú vyrátané výsledné sily funkcií FD_n , FD_a a FDe môže byť spustený cyklus, v ktorom sa bude graf transformovať až do výslednej podoby.

Jedna iterácia cyklu znamená, že sa vykoná operácia, ktorá buď posunie vrcholy istým smerom alebo rotuje náhodne vybraný element. Následne sa aplikujú funkcie FD_n , FD_a a FDe na tie vrcholy, ktoré touto akciou mohli byť ovplyvnené a podľa hodnôt En , Ea a Ee sa rozhodne či bude cyklus zastavený a tým graf potvrdený ako výsledný, alebo bude cyklus pokračovať v ďalších iteráciách. Operácia vykonaná v jednej iterácii môže byť dvoch typov. Ide o rotovanie alebo posúvanie vrcholov. V každej iterácii sa náhodne vyberie jedna z nich. Aby bolo možné náhodný výber kontrolovať sú týmto operáciám priradené pravdepodobnostné hodnoty, ktoré vyjadrujú z akou pravdepodobnosťou sa v jednej iterácii vykoná buď rotácia alebo posun. Ak ide o rotáciu vyberie sa náhodný vrchol z grafu a aplikuje sa príslušná funkcia automatickej rotácie. Pri elementoch hélíx a relácie je aplikovaná metóda, kedy sú s vrcholom otáčané aj celé časti grafu, ktoré svojimi buď ľavými alebo pravými asociačnými bodmi element uzatvára.

Ak je v iterácii vybraná funkcia posunu, tak kontrola síl a prípadná zmena pozície bude prebiehať postupne vo všetkých vrcholoch. Po každom posune vrcholu je nutné aplikovať funkciu udržiavania vrcholov typu SSE v kružniciach a silové funkcie FD_n , FD_a a FDe na vrcholy, ktoré mohli byť týmto ovplyvnené. Aby bola zmena pozícií plynulá pohyb je kontrolovaný tým, že sa vrchol pohne o vzdialenosť, ktorá sa rovná napríklad štvrtej odmocnine dĺžky prislúchajúceho smerového vektora. Tento prepočet sa mení v závislosti na veľkosti síl nastavených v silových funkciách. Prvý spôsob ako nájsť tento vektor je súčet výsledných síl teda smerových vektorov funkcií FD_n , FD_a a FDe ktoré pôsobia na ťažisko vrcholu. Vplyv jednotlivých síl je možné kontrolovať im prislúchajúcimi konštantami respektíve škálami S_n , S_e a S_a , ktorými budú tieto vektory násobené.

Druhý spôsob je náhodný výber funkcie FD_n , FD_a a FDe podľa ktorého výslednej sily bude vrchol posunutý. Tu je taktiež potrebné kontrolovať výskyt jednotlivých funkcií a tým aj ich vplyv počas behu programu a to tak, že každej je priradená pravdepodobnostná premenná, ktorá vyjadruje z akou pravdepodobnosťou sa vyberie v prípade posunu jedna z nich. Ako dobrý prístup sa však ukázalo použitie práve prvého spôsobu.

4. Implementácia a vyhodnotenie

Našu metódu silovo riadeného vykresľovania RNA motívov sme implementovali v programe RNA viewer. Metóda obsahuje viacero konštánt, ktoré sme naladili tak, aby nakreslené motívy čo najlepšie zodpovedali požiadavkám, ktoré sme sformulovali v úvodných kapitolách. Okrem algoritmu je v programe k dispozícii viacero nástrojov, na konečnú úpravu nakreslenia a na zmenu štruktúry v deskriptore. V tejto kapitole popíšeme implementačné detaily a ukážeme nakreslené výsledky vyprodukované našim programom.

4.1. Implementačné detaily

Program RNA viewer je implementovaný v jazyku C++ v prostredí QT. Vývojové prostredie QT C++ je vhodným variantom na implementáciu, pretože má všetky svoje knižnice optimalizované tak, aby boli kompatibilné s viacerými operačnými systémami. Z toho dôvodu je možné pomerne jednoducho a bez zásahov do zdrojového kódu programu vygenerovať spustiteľný súbor ktorý je možné spustiť pod zvoleným operačným systémom.

4.2. Ladenie konštánt

Jednotlivé funkcie metódy silovo riadeného vykresľovania motívov majú svoje vlastné konštanty, ktoré kontrolujú ich vplyv. Ich nastavovaním je možné zabezpečiť dodržiavanie požiadaviek správneho vykresľovania RNA motívu. Ide o konštanty jednotlivých silových funkcií FDn , FDe a FDa . Vzhľadom na to, že beh algoritmu je založený na vektorovom súčte výsledných hodnôt silových funkcií pôsobiacich vo vrchole, je potrebné nastaviť škálovanie konštanty (viď. Sekcia 3.2.8). Základná konštanta, od ktorej sa ostatné odvíjajú, je veľkosť jedného znaku Z_w . Práve od veľkosti znaku závisia vzájomné vzdialenosti jednotlivých vrcholov.

Funkcia FDn môže byť nastavovaná konštantou l , ktorá vyjadruje pokojovú dĺžku pružiny (viď. Sekcia 3.2.1). To znamená, že jej hodnota udáva vzájomné vzdialenosti asociovaných vrcholov. Jednou z požiadaviek správneho vykreslenia motívu je dodržiavanie rovnakých vzdialeností medzi nukleotidmi, preto $l = Z_w/3$. Tretina veľkosti znaku udáva primeranú vzdialenosť medzi znakmi v predpise a aj medzi vrcholmi.

Funkcii FDe môžu byť nastavované dve konštanty: maximálna hodnota g sily, ktorú môže FDe nadobudnúť a veľkosť poľa h v okolí vrcholu v ktorom dochádza ku kolízií (viď. Sekcia 3.2.2). Hodnota konštanty g v podstate nie je až tak dôležitá pretože o vplyve

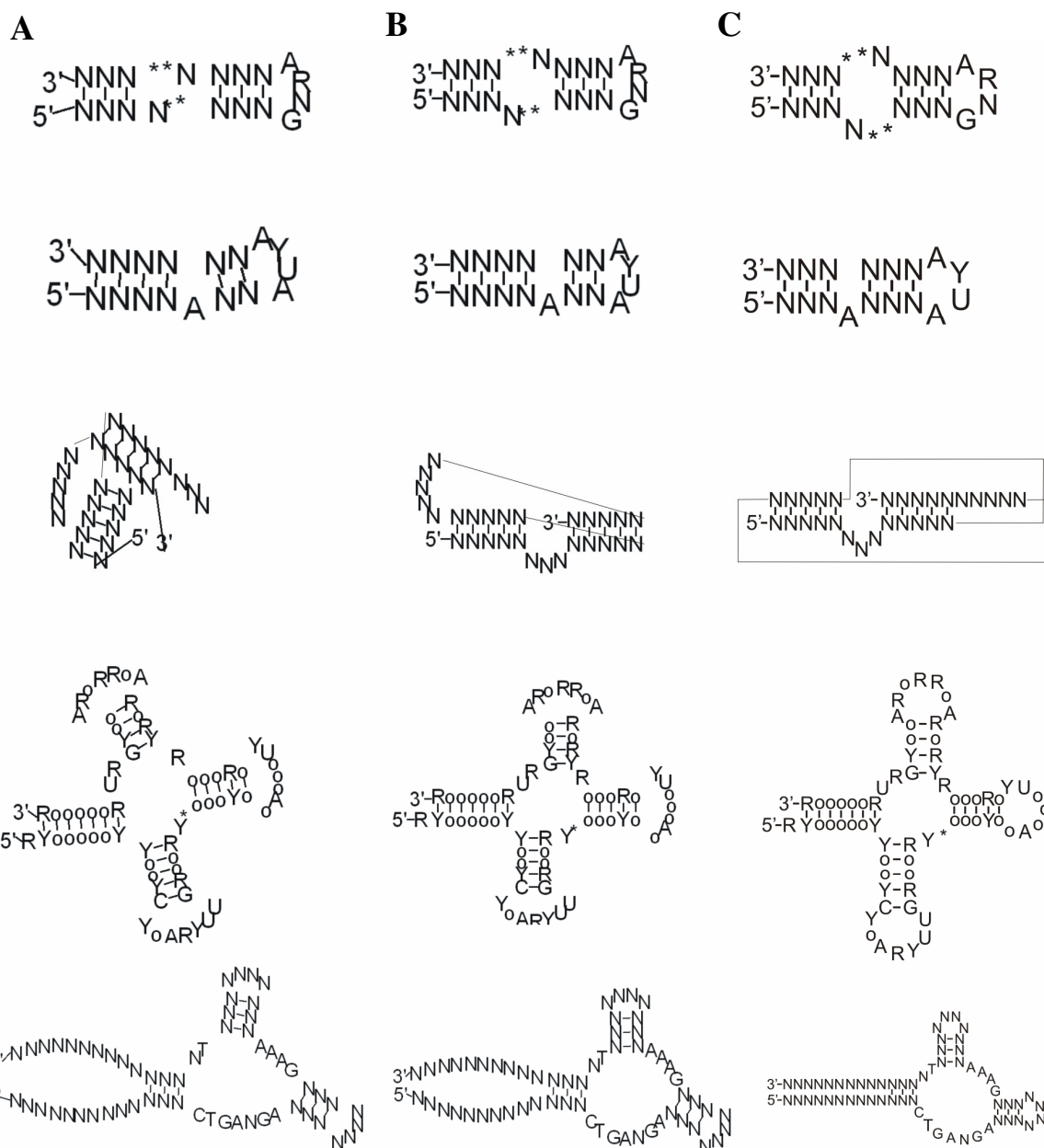
funkcie FDe rozhoduje jej prislúchajúca škála Se . Preto $g = 200$. Keďže každý vrchol má okolo seba pole veľkosti h nesmie táto konštanta v súčte dvoch vrcholov zásadne presahovať hodnotu konštanty l . Preto $h = l/2$.

Funkcii FDa je možné taktiež nastavovať konštantu, ktorá vyjadruje maximálnu možnú hodnotu sily funkcie FDa (viď. Sekcia 3.2.3). Ide o konštantu m , ktorej hodnota taktiež ako pri funkcii FDe nie je dôležitá, pretože o vplyv funkcie FDa rozhoduje jej prislúchajúca škála Sa . Preto $m = 200$.

Konečné nastavenie škál vplyvu silových funkcií je postupným ladením ustálené na hodnotách $Sn = 1, Sa = 1$ a $Se = 4$, kde vplyv funkcie pre vyhýbanie sa kolíziám FDe musí byť oproti ostatným funkciám štvornásobne zväčšený. Toto ladenie prebiehalo testovaním rôznych typov RNA štruktúr a ich porovnávaním s preferovanými ručnými nakresleniami.

4.3. Príklady vykreslených motívov

Vykresľovanie RNA motívov v programe RNA viewer prebieha v dvoch fázach. Načítanie vstupného deskriptora a jeho automatické vykreslenie a následná úprava motívu do požadovaného tvaru. Naša metóda najlepšie vykresľuje štruktúrou jednoduché motívy, kde nie je nutné vykonávať veľa zmien aby bol dosiahnutý požadovaný. Najlepšie výsledky mala naša metóda pri vykresľovaní motívov s pravidelnými štruktúrami kde automaticky vykreslený motív bol veľmi podobný s motívom kresleným ručne. V menej pravidelných motívoch a v motívoch, kde bolo potrebné do väčšej kružnice zasadiť helixy sa vyskytol nedostatok funkcie, ktorá podľa susedných vrcholov ohýba predpis SSE. Funkcia v týchto prípadoch neohne predpis SSE dostatočne na to aby tvoril kružnicu. V prípade, že motív obsahuje pseudouzol výsledný tvar po automatickom vykreslení podstatne menej znázorňuje požadovaný výsledok. Je to spôsobené tým, že sa body pseudouzlu nachádzajú na opačných koncoch motívu. Keďže každá hrana je zároveň aj pružina tieto konce sú rovnako ako u iných bodoch spojených hranou priťahované. Tento problém by mohol byť odstránený dočasným vylúčením hrán spôsobujúcich pseudouzly (viď. Sekcia 4.4).



Obrázok 4.1: Príklad vykreslenia RNA motívov programom RNA viewer. (A) Automatické vykreslenie. (B) Finálny výsledok po úprave pomocou nástrojov RNA viewer. (C) Ručne nakreslená RNA motív.

4.4. Rozšírenia programu

Prvým nedostatkom programu RNA viewer je to, že motívy, kde sa nachádzajú slučky nie sú správne vykresľované, pretože sú zväčša priťahované k sebe vzdialené konce motívu. Možným rozšírením by mohlo byť vyhľadávanie takýchto úsekov a vyradenie asociácií, ktoré zapríčiňujú tento nežiaduci stav. Po skončení behu algoritmu by boli asociácie vrátené do motívu tak, aby neprekrižovali žiaden element. Ďalším rozšírením je lepšie udržiavanie SSE v kružniciach pretože ak je jeden SSE zložený s väčšieho počtu znakov

kruhovitý útvar v ktorom sa vykresľuje je väčší ako by bolo potrebné. Možným používateľským rozšírením je to aby si program uchovával zmeny po ukončení algoritmu, ktoré používateľ robí na konečnom dokreslení motívu. Tieto zmeny by potom po ďalšom automatickom vykreslení boli automaticky aplikované na motív.

Záver

Cieľom tejto práce bolo navrhnúť a implementovať metódu, ktorá dokáže automaticky vykresľovať RNA motívy, tak aby boli dodržané požiadavky správneho vykreslenia motívu. Snahou bolo dosiahnuť to aby automatické vykresľovanie motívov pomocou našej metódy založenej na silovo riadenom vykresľovaní vytváralo nakreslenia, ktoré by sa čo najviac podobali preferovaným ručným nakresleniam motívov.

Hlavnou časťou práce bolo navrhnutie algoritmu, ktorý automaticky vykreslí RNA motív tak, aby bolo dodržaných čo najviac požiadaviek správneho vykreslenia motívu. Inšpirovaním sa metódou kreslenia grafov a grafových štruktúr pomocou síl bolo v tomto prípade dobrým riešením, pretože postupným implementovaním a testovaním rôznych typov síl a ich kombinácií sme dospeli ku konečnému stavu, kedy vieme kontrolovať silami požiadavky pre správne vykreslený RNA motív. Zároveň je možnosť doplniť do algoritmu nové typy síl na kontrolu požiadaviek, ktoré by mohli postupom času pribudnúť. Táto metóda je preto otvorená rozšíreniam.

Algoritmus, ktorý automaticky vykresľuje motív okrem jednoduchých štruktúr zložených s desiatich elementov prehľadne vykresľuje aj komplexnejšie štruktúry, kde používateľ následne len drobnými zmenami dosiahne požadovaný tvar. Existujú ale aj také, ktorým algoritmus nedokáže zabezpečiť dodržanie podmienok správneho vykreslenia. Ide o motívy, ktoré obsahujú pseudouzly. V týchto prípadoch sú asociačné body, ktoré majú byť v susedstve spravidla na opačných koncoch štruktúry to spôsobuje problémy pri postupnej transformácii grafu motívu do výsledného nakreslenia.

Myslíme si, že program RNA viewer má potenciál na to aby bol v praxi využívaný, pretože motívy vykreslené pomocou našej metódy sú veľmi podobné preferovaným ručným nakresleniam a preto čas na ich konečnú úpravu do výsledného tvaru je podstatne kratší, ako je to v iných nástrojoch. Ide o jednoduché úpravy v RNA viewer, ktoré si nie je potrebné prvotné naštudovanie z používateľského manuálu.

V priebehu tejto práce sme sa bližšie zoznámili s problematikou vykresľovania nielen RNA motívov, ale aj samotných grafov pričom sme základnú metódu vykresľovania pomocou síl pružín rozšírili o sily, ktoré doposiaľ v týchto metódach použité neboli. Ukázali sme, že je možné rôzne typy podmienok kontrolovať pomocou správne navrhnutých síl a tým dosiahnuť požadovaný výsledok vo vykreslenom grafe. Dúfame, že

tento prístup poslúži niekomu ako inšpirácia v problematike kreslenia RNA motívov alebo aj v teórií kreslenia grafov a grafových štruktúr.

Zdroje

- [1] D. AUBER, M. DELEST, J. DOMENGER and S. DULUCQ. Efficient drawing of RNA secondary structure. *Journal of Graph Algorithms and Applications*, 10(2):329–351, 2006.
- [2] Y. Byun and K. Han. PseudoViewer3: generating planar drawings of large-scale RNA structures with pseudoknots. *Bioinformatics*, 25(11):1435-1437, 2009.
- [3] K. Darty, A. Denise and Y. Ponty. VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, 25(15):1974-1975, 2009.
- [4] R. Davidson and D. Harel. Drawing graphs nicely using simulated annealing. *ACM Trans Graph.*, 15(4):301-331, 1996.
- [5] S.R. Eddy. RNABob: a program to search for RNA secondary structure motifs in sequence databases. unpublished, 1996.
- [6] D. Gautheret, F. Major and R. Cedergren. Pattern searching/alignment with RNA primary and secondary structures: an effective descriptor for tRNA. *Computer Applications in the Biosciences*, 6(4):325-31, 1990. ISSN 0266-7061.
- [7] E. Glen and A. Vasudevan. jViz.Rna-a java tool for RNA secondary structure visualization. *NanoBioscience*, 4(3):212-218, 2005.
- [8] E. G. Golshtein and N. V. Tretyakov. Modified Lagrangians and monotone maps in optimization. *New York: Wiley*, pages 6, 1996. ISBN 0471548219.
- [9] D. Harel and Y. A Fast Multi-Scale Method for Drawing Large Graphs. In *Graph Drawing*, 1984:235-287 Springer, 2001.

- [10] T Kamada and S. Kawai. An algorithm for drawing general undirected graphs. *Information Processing Letters*, 31(1):7–15, 1989.
- [11] S. Kirkpatrick, C. D. Gelatt, M. P. Vecchi. Optimization by Simulated Annealing. *Science*, 220(4598):671-680, 1983.
- [12] T. J. Macke, D. J. Ecker, R. R. Gutell, D. Gautheret, D. A. Case, and R. Sampath. RNAMotif, an RNA secondary structure denition and search algorithm. *Nucl. Acids Res.*, 29(22):4724-4735, 2001.
- [13] B. A .Shapiro, J. Maizel, L. E. Lipkin, K. Currey, and C. Whitney. Generating non-overlapping displays of nucleic acid secondary structure. *Nucleic Acids Res*, 12(1 Pt 1): 75–88, 1984.
- [14] K. Shimizu, Y. Ishizuka and F. Jonathan. Nondifferentiable and two-level mathematical programming. *Boston: Kluwer Academic Publishers*, pages 19, 1997. ISBN 0792398211.
- [15] M. THOMAS, J. FRUCHTERMAN and E. REINGOLD. Graph Drawing by Force-directed Placement. *Software:ractice and experience*, 21(11):1129-1164, 1991.
- [16] Z. WEINBERG and R. BREAKER, R2R - software to speed the depiction of aesthetic consensus RNA secondary structures. *Weinberg and Breaker BMC Bioinformatics*, 12(3):3, 2011.

Prílohy

Používateľské rozhranie programu RNA viewer

Program RNA viewer ponúka rozhranie pomocou ktorého môže používateľ kresliť RNA motívy popísané v deskriptore. Po načítaní deskriptoru program navrhne výzor RNA motívu. Následne si používateľ môže pomocou grafických nástrojov ešte motív prispôbiť svojim požiadavkám. K dispozícii má grafické nástroje, ale aj možnosť zmeniť štruktúru motívu priamo v deskriptore. Výsledný obrázok je potom možné exportovať do niekoľkých grafických formátov.

Načítanie a vykreslenie RNA motívu

Po spustení programu RNA viewer je používateľovi ponúknuté prostredie zložené z dvoch hlavných častí. Prvá časť je textové pole umiestnené napravo okna programu. V tomto poli je zobrazený deskriptor. Druhá časť je kresliace pole, kde je podľa deskriptoru v textovom poli RNA motív vykresľovaný.

Prvá možnosť ako s programom pracovať je načítanie RNA motívu z už vytvoreného deskriptoru. Po načítaní je podľa algoritmu popísaného v metodike vykresľovania RNA motív v kresliacom poli vykreslený. Následne je možné pomocou grafických nástrojov motív prispôbovať požiadavkám používateľa. Okrem toho je možné v textovom poli meniť deskriptor a tým aj štruktúru motívu.

Druhá možnosť ako s programom pracovať je postupné vytváranie deskriptoru v textovom poli. Teda používateľ môže písaním do textového poľa postupne vytvárať vlastný deskriptor, na ktorý je následne aplikovaný algoritmus na vykresľovanie RNA motívov. Tento motív si taktiež používateľ môže pomocou grafických nástrojov prispôbovať svojim požiadavkám.

Používateľské nástroje

Používateľské nástroje, ktoré ponúka program RNA viewer sú umiestnené podľa typu do vlastných pracovných líšt, ktoré si môže používateľ premiestňovať v rámci okna programu na miesta kde mu to najviac vyhovuje.

Lišta s nástrojmi pre textové pole obsahuje tlačidlo EDIT ktoré aktivuje možnosť prepisovania textového poľa teda povolí zmeny v deskriptore, tlačidlo RESIZE, ktoré aktualizuje vykreslený RNA motív podľa zmien v textovom poli a tlačidlo POS, ktoré

zapíše do textového poľa ako komentár deskriptoru aktuálne vykreslený motív, ktorý sa môže po uložení do súboru .des môže neskôr načítať.

Lišta s grafickými nástrojmi obsahuje tlačidlo Select kedy je povolený výber viacerých elementov zo štruktúry RNA motívu. Po opätovnom stlačení tohto tlačidla teda jeho deaktivácií je vybraný iba jeden element. Stlačený kláves Shift nahradzuje aktiváciu tlačidla Select a jeho pustením sa Select deaktivuje. Výber elementu je zabezpečovaný kliknutím myši do oblasti jeho výskytu. Ak je aktivované tlačidlo MOVE používateľ môže hýbať všetkými vybranými elementmi. Ak je aktivované tlačidlo ROT, používateľ môže rotovať všetky vybrané elementy. Ak je aktivované tlačidlo HAND, používateľ môže celý RNA motív posúvať na nové pozície. Ak je stlačené tlačidlo MIRR, aplikuje sa funkcia zrkadlového otočenia na všetky vybrané elementy typu hélix alebo relácia. Tlačidlom Center používateľ umiestni RNA motív do stredu kresliacej plochy a tlačidlom Color je možné meniť farbu všetkým vybraným elementom.

Samostatná lišta, ktorá ale nie je bežnému používateľovi zobrazená obsahuje rôzne nastavenia algoritmu pre vykresľovanie RNA motívu. Je možné ju zobrazit' v prípade rozširovania a úpravy programu.

Výstupné súbory

Výstupný súbor programu RNA viewer je v podstate obrázok vykresleného RNA motívu. Kresliacu plochu programu je preto možné exportovať do obrázkových formátov. Požívateľ má na výber formáty BMP, JPG, JPEG a PNG. Okrem toho je možné kresliacu plochu uložit' do PDF súboru veľkosti A4. Zároveň je možnosť uložit' si do deskriptoru aktuálne textové pole.

CD nosič: