

UNIVERZITA KOMENSKÉHO
FAKULTA MATEMATIKY, FYZIKY A
INFORMATIKY

GRAFICKÉ PROSTREDIE NA EDITOVANIE
RNA MOTÍVOV

2010

Peter Herman

Univerzita Komenského
Fakulta Matematiky, Fyziky a Informatiky

Grafické prostredie na editovanie RNA motívov

Bakalárska práca

Študijný program: Informatika
Študijný odbor: 9.2.1 Informatika
Školiace pracovisko: Katedra Informatiky
Školiteľ: Mgr. Bronislava Brejová, PhD.

Bratislava, 2010

Peter Herman

Prehlásenie

Čestne prehlasujem, že som túto bakalársku prácu vypracoval samostatne s použitím citovaných zdrojov.

.....

Pod'akovanie

Chcel by som sa týmto poďakovať vedúcej mojej bakalárskej práce Mgr. Bronislave Brejovej, PhD. za rady, pomoc a hlavne trpezlivosť, ktorú so mnou mala.

Abstrakt

Vyhľadávanie RNA génov pomocou štruktúrálnych motívov sa stalo za posledné roky v bioinformatike bežným problémom a vzniklo niekoľko programov na jeho riešenie. Štruktúralne motívy sa podobajú na regulárne výrazy, ktoré navyše môžu špecifikovať pozície reťazca na ktorých musia byť navzájom komplementárne znaky.

V tejto bakalárskej práci podávame prehľad problematiky hľadania RNA štruktúrálnych motívov a prezentujeme nové grafické prostredie, v ktorom možno tieto štruktúralne motívy editovať a exportovať do formátu používaného programom RNABob.

Kľúčové slová: grafické prostredie, RNABob, RNA gény, sekundárna štruktúra, štruktúralne motívy

Obsah

1	Úvod	1
2	Biologické základy	2
2.1	DNA	2
2.2	RNA	3
2.2.1	Štruktúra RNA	3
2.2.2	Transkripcia	4
3	Hľadanie štrukturálnych motívov	5
3.1	RNAMot	5
3.2	RNAMotif	6
3.3	Locomotif	7
3.4	RNABob	7
3.5	Formát deskriptorov programu RNABob	8
4	Implementácia programu	11
4.1	Použité technológie	11
4.2	Návrh štruktúry	11
4.3	Kontrola orientácie reťazcov v hélixе	12
4.4	Detekcia cyklov	14
4.5	Porovnanie s existujúcimi prostrediami	14
5	Užívateľský manuál	16
5.1	Hlavné okno	16
5.2	Scéna	18

<i>OBSAH</i>	v
5.2.1 Insert mód	18
5.2.2 Connect mód	19
5.2.3 Move mód	19
5.3 Objekty	19
5.3.1 Farebné zobrazenie	21
6 Záver	22

Kapitola 1

Úvod

Biológovia si čoraz viac uvedomujú významný podiel RNA génov na procesoch v živých organizmoch. Funkcia týchto génov je silne závislá od ich štruktúry. To spôsobuje, že táto štruktúra sa medzi rôznymi živočíšnymi druhmi zachováva.

Z podoby takejto štruktúry vieme zistiť, ktoré časti reťazca RNA sú spárované, teda sú tvorené komplementárnymi stavebnými jednotkami a tým získavame obmedzenia na možnú podobu reťazca. Našou úlohou teda je, ak na vstupe dostaneme popis štruktúry RNA génu vytvorený na základe výskytov v známych organizmoch, nájsť v neznámom genóme také úseky, ktoré spĺňajú nami vytvorené obmedzenia, a teda existuje pravdepodobnosť, že kódujú tento RNA gén.

Programov riešiacich túto úlohu je niekoľko, pre väčšinu je však potrebné vytvárať opisy takejto štruktúry ručne v textovom editore. Cieľom tejto práce je vytvoriť grafické prostredie, ktoré umožní užívateľovi vytvárať takéto štruktúry pohodlne s vizuálnou spätnou väzbou a exportovať ich popis do formátu používaného jedným z týchto programov, RNABobom.

Kapitola 2

Biologické základy

Bioinformatika sa od svojho vzniku zaoberá predovšetkým skúmaním genetickej informácie obsiahnutej v bunkách živých organizmov. Genetická informácia je uložená v jadre bunky zakódovaná v takzvaných nukleových kyselinách. Sú to makromolekulové látky zložené z piatich základných nukleotidov (adenín, cytozín, guanín, tymín a uracil). Nukleové kyseliny majú podobu dlhých nukleotidových reťazcov, v ktorých je informácia kódovaná prostredníctvom poradia a typu nukleotidov.

2.1 DNA

Hlavnou úlohou deoxyribonukleovej kyseliny (DNA) je dlhodobé uchovávanie genetickej informácie. Vlákno DNA je tvorené dvomi komplementárnymi nukleotidovými reťazcami skladajúcich sa iba z adenínu, cytozínu, guanínu a tymínu, pričom komplementárne páry¹ tvoria adenín s tymínom a cytozín s guanínom (táto komplementarita funguje na základe chemickej štruktúry nukleotidov).

Vďaka tomu, že kostra vlákna, na ktorej sú nukleotidy uchytené, je tvorená striedaním fosfátov s naviazaným cukrom, sú konce DNA vlákna asymetrické a preto vieme určiť orientáciu reťazca. Tieto konce sa označujú 5' a 3' a dva reťazce, ktoré tvoria vlákno DNA, sú vždy orientované opačne (teda

¹Nazývané tiež Watson-Crickove páry.

ak jeden reťazec má na konci 3' druhý reťazec má na tomto konci 5').

Pre všetky naše účely je DNA reprezentovaná reťazcom nad abecedou {A,C,G,T}, ktorý sa tiež nazýva primárna štruktúra DNA.

2.2 RNA

Ribonukleová kyselina (RNA) je narozdiel od DNA tvorená iba jedným nukleotidovým reťazcom a tymín sa v ňom nevyskytuje, je nahradený uracilom, ktorý tvorí komplementárny pár s adenínom.

V organizmoch zastáva RNA celú radu dôležitých funkcií. U jednoduchších organizmov, napríklad u RNA vírusov môže niesť genetickú informáciu, a v zložitejších organizmoch napríklad pomáha pri syntéze bielkovín, či priamej regulácii chemických procesov.

Podobne ako pri DNA sú oba konce rozlíšiteľné a označujú sa tiež 5' a 3'.

2.2.1 Štruktúra RNA

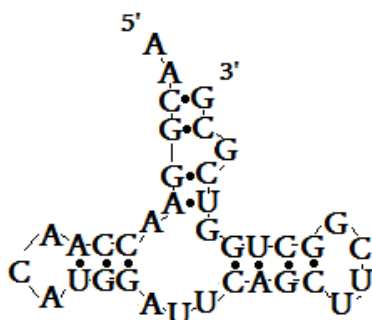
Primárna štruktúra

Pod primárnou štruktúrou RNA si môžeme predstaviť analogicky ako pri DNA poradie nukleotidov v reťazci, teda reťazec nad abecedou {A,C,G,U}.

Sekundárna štruktúra

Keďže nukleotidy nemajú komplementárne vlákno, s ktorým by vytvorili väzby, majú veľkú voľnú energiu a v dôsledku toho je štruktúra v podobe voľného nespárovaného vlákna veľmi nestabilná. Aby svoju energiu znížilo, vlákno RNA sa poohýba a vytvorí sa väzby medzi komplementárnymi úsekmi vlákna samotného, ako vidíme napríklad na obrázku 2.1. Informácia o tom, ktoré nukleotidy reťazca sú spojené väzbou, sa nazýva sekundárna štruktúra.

Väzby v sekundárnej štruktúre už nemusia vznikať iba medzi adenínom a uracilom resp. medzi cytozínom a guanínom. Môže sa stať, že dva "nekompatibilné" nukleotidy sa spárujú v dôsledku silných väzieb vytvorených z



Obr. 2.1: Príklad sekundárnej štruktúry.

oboch strán nukleotidu. Najlahšie spárovateľný, teda najčastejšie sa vyskytujúci pár mimo Watson-Crickových, je pár guanín s uracilom, ktorý má ešte porovnateľne nízku energiu s Watson-Crickovými pármami.

Je bežné nájsť v rôznych genómoch príklady RNA, ktoré majú rovnakú sekundárnu štruktúru, aj keď nezdediajú žiadnu významnú podobnosť v primárnej štruktúre. Veľké zmeny v primárnej štruktúre bývajú často tolerované v prípade, že k mutácii dôjde i u väzbového partnera.[1] Dôvodom pre toto je, že pri niektorých druhoch RNA hrá dôležitú úlohu vo funkčnosti sekundárna, resp. terciárna štruktúra².

2.2.2 Transkripcia

Funkčné RNA molekuly sú zakódované v genóme, teda molekulách DNA. Transkripcia je proces tvorby RNA kópie z DNA sekvencie. Jedná sa o prvý krok v expresii génov.[2] Počas transkripcie sa na vlákno DNA naviaže enzým RNA polymeráza, ktorý prechádza po vlákne a vytvára komplementárny reťazec RNA, ktorý je zhodný s komplementárnym vláknom DNA, avšak výskyty tymínu sú v ňom nahradené uracilom.[3] Takto preložený RNA gén môže buď vykonávať určitú funkciu, alebo ďalej poslúžiť na syntézu proteínov.

²Rozloženie atómov v priestore, ktoré býva sekundárnou štruktúrou silno ovplyvnené.

Kapitola 3

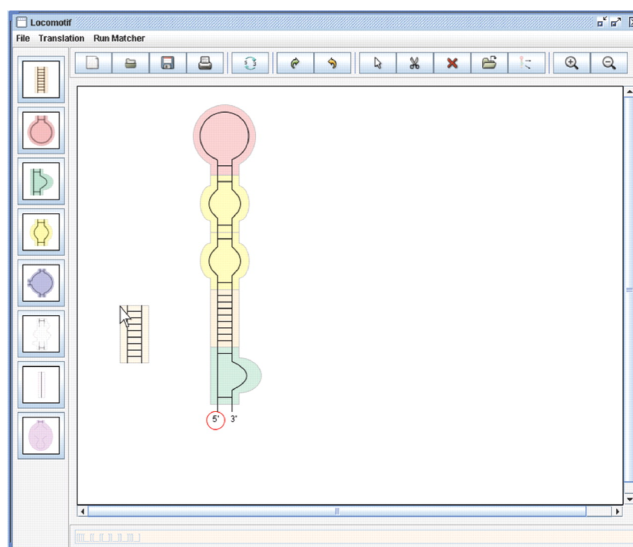
Hľadanie štrukturálnych motívov

Ako sme spomenuli v predchádzajúcej kapitole, RNA gény v jednotlivých druhoch sú často zachovávané hlavne pomocou sekundárnej štruktúry. Keď chceme hľadať RNA gény v novom genóme, môžeme sa pokúsiť nájsť také, ktoré sa podobajú svojou štruktúrou na už známe RNA gény z iných organizmov. Sekundárnu štruktúru, prípadne triedu podobných sekundárnych štruktúr, známych génov môžeme popísať pomocou špeciálneho motívu. Úlohou býva nájsť v DNA úseky, ktorých transkripciou získame RNA reťazec, ktorý má sekundárnu štruktúru zhodnú s hľadaným motívom, ktorý vyhľadávame. Motívy vo všeobecnosti vymenovávajú poradie spárovaných a nespárovaných úsekov v RNA a pre každý špecifikujú jeho dĺžku, prípadne niektoré zachované časti sekvencie.

Väčšina programov nám nedáva žiadnu garanciu toho, že nájdený reťazec RNA túto štruktúru naozaj nadobúda, iba nám hovorí, že je teoreticky schopný ju nadobudnúť.

3.1 RNAMot

RNAMot je prvý program, ktorý sa zaoberal vyhľadávaním motívov v databázach sekvencií. Program implementoval prehľadávanie všetkých možností s postupným orezávaním možností, teda pracuje najhoršie v exponenciálnom čase pre každý spárovaný úsek.[4]



Obr. 3.1: Prostredie Locomotif.

Algoritmus postupoval po jednotlivých spárovaných úsekoch a pre každý z nich v reťazci našiel vhodné vzdialené (možné rozsahy dĺžok všetkých štruktúrnych prvkov sú dané na vstupe) komplementárne úseky reťazca, o ktorých následne prehlásil, že sa spoja do hľadaného hélisu. Týmto získal ďalšie obmedzenia na polohu ostatných prvkov a pokračoval vo výstavbe štruktúry. Ak sa mu podarilo spojiť umiestniť celú štruktúru, vyhodnotil voľnú energiu takejto štruktúry a z prekrývajúcich sa štruktúr vypisoval najpravdepodobnejšiu. Navyše umožňoval užívateľovi špecifikovať poradie napasovávania štruktúr, čím sa dal čas behu programu podstatne skrátiť, ak sa postupovalo od najdlhších helixov, pre ktoré sa vyskytovalo menej náhodných výskytov.

3.2 RNAMotif

RNAMotif je ďalší z programov, ktorý umožňuje vyhľadávanie štruktúrnych motívov.[5] Podobne ako RNAMot si vybuduje zo štruktúr "vyhľadávací strom", ktorý sa postupne snaží napasovať na reťazec. Od ostatných programov sa líši hlavne tým, že umožňuje užívateľovi napísať si vlastnú skórovaciu funkciu, vďaka ktorej je schopný zohľadniť aspekty, ktoré ostatné programy nemôžu zohľadňovať, napríklad celkový počet nezhôd v hľadanej sekvencii.



Obr. 3.2: Jednoduchý príklad pseudouzla.

3.3 Locomotif

Locomotif sa tiež zaoberá vyhľadávaním motívov.[6] Od iných programov sa odlišuje tým, že používa efektívne dynamické programovanie založené na termodynamických modeloch sekundárnej štruktúry RNA. Na ich vytváranie používa grafické prostredie naprogramované v Jave (obr. 3.1) a používa client-server model, kde užívateľ si vytvorí motív u seba a program na vyhľadávanie sa vygeneruje a skompiluje na vzdialenom serveri.

Locomotif si môže dovoliť používať dynamické programovanie vďaka tomu, že neumožňuje vytvárať štruktúry nazývané pseudouzly. Ak vlákno RNA obsahuje medzi dvoma časťami tvoriacimi hélix úsek spárovaný s reťazcom ležiacim mimo túto oblasť, nazýva sa tento spárovaný úsek pseudouzol. Ak motív neobsahuje pseudouzly, dá sa popísať stochastickými bezkontextovými gramatikami a pre tieto vieme v čase $\Theta(N^3)$ nájsť dynamickým programovaním riešenie.

Nakoniec skontroluje, či sa reťazec skutočne s najväčšou pravdepodobnosťou poskladá do nami zadaného motívu. Výsledky Locomotifu teda môžeme považovať za naj dôveryhodnejšie. Bohužiaľ platíme za to tým, že nemôžeme vyhľadávať štruktúry s pseudouzlami.

3.4 RNABob

RNABob je implementácia RNAMOTu používajúca algoritmicky rozdielny prístup.[7] V pozadí je nedeterministický konečný automat, ktorý dynamicky generuje prechodovú funkciu pre niektoré stavy počas behu.

Ide o bežne používaný program, ktorý však podobne ako RNAMot a RNAMotif prichádza iba ako konzolová aplikácia, a deskriptory opisujúce sekundárnu štruktúru treba editovať ručne v textovom editore.

Keďže ručné opisovanie štruktúr môže byť zdĺhavé a náročné, rozhodli sme sa pre tento program vytvoriť grafické prostredie umožňujúce editáciu týchto štruktúrnych motívov.

3.5 Formát deskriptorov programu RNABob

Deskriptor programu RNABob opisuje sekundárnu štruktúru, ktorú chceme hľadať, a prípadné zachované sekvenčné motívy. Celú štruktúru si môžeme predstaviť ako reťazce nukleotidov patriace rôznym štruktúrnym elementom pospájané za sebou do radu. Štruktúrne elementy sú vo všeobecnosti dvoch druhov. Prvým druhom sú spárované úseky (nazývané "hélix", prípadne "double strand") a druhým druhom sú nespárované vlákna (nazývané "single strand"). Keďže hélix je tvorený dvoma vláknami pospájanými nukleotidovými väzbami, vyskytne sa v tejto štruktúre každý hélix pochopiteľne dvakrát, každé jeho vlákno raz. Okrem týchto základných štruktúr umožňuje RNABob ešte tretí druh, reláciu, ktorá je zovšeobecnením hélixu a umožňuje vyjadriť striktnjšie, alebo voľnejšie podmienky na podobu spárovaných úsekov. Napríklad nám umožní nedovoliť G-U páry, ktoré RNABob bežne dovoľuje.

Samotný deskriptor pozostáva z dvoch častí:

V prvom riadku je topologický popis reťazca, to znamená zoznam jednotlivých štruktúrnych prvkov prislúchajúcim jednotlivým častiam reťazca v poradí od 5' konca k 3' koncu. Každý štruktúrny prvok má určený jednoznačný identifikátor. Identifikátor začína písmenom **s**, ak ide o nespárovaný úsek, **h**, ak ide o hélix a **r**, ak ide o reláciu, a pokračuje unikátnym reťazcom (štandardne sa používajú kladné celé čísla). Druhý výskyt hélixu alebo relácie má rovnaký identifikátor, ale pridá sa k nemu na konci apostrof.

Potom nasleduje niekoľko riadkov, z ktorých každý opisuje vlastnosti jednej štruktúry. Na začiatku riadku je identifikátor prvku, potom nasleduje celé číslo vyjadrujúce počet nezhôd povolených oproti predpisu a predpis reťazca.

Predpis vymedzuje tvar nukleotidového reťazca. Okrem značiek jednotlivých nukleotidov (A,C,G,U resp. T) môže obsahovať i písmená z takzvaného IUPAC¹ single-letter kódu, čo je vlastne dohodnutý súbor písmen, ktoré vyjadrujú rôzne podmnožiny nukleotidov. Napríklad R znamená, že na tomto mieste môže byť adenín, alebo guanín. Pre označenie ľubovoľnej bázy sa používa písmeno N. Ak sa jedná o hélix resp. reláciu, sú to dva reťazce oddelené dvojbodkou. Ako prvý je napísaný ten reťazec, ktorý sa vyskytuje bližšie k 5', a teda skôr v topologickom opise. Oba reťazce sú napísané v smere od 5' k 3', a preto si treba uvedomiť, že väzby sú vytvorené zrkadlovo vzhľadom na dvojbodku.

Počet povolených nezhôd udáva, na koľkých pozíciách sa môže nájsť reťazec líšiť od zadaného predpisu. Ak ide o nespárovaný úsek, je vyjadrený jedným číslom, ak ide o inú štruktúru, sú tu dve čísla oddelené dvojbodkou, ktoré vyjadrujú počet povolených nezhôd pre každý reťazec zvlášť.

Pomerne často nie je možné presne určiť dĺžku reťazca v niektorom z týchto prvkov. Preto RNABob umožňuje do reťazca pridať hviezdičku, ktorá označuje, že na tejto pozícii je buď ľubovoľný nukleotid (N), alebo nič. Napríklad reťazec ***A*G* možno interpretovať ako AG, NAG, ANG, NANG, alebo NNANG.

Na príklade deskriptora (obr. 3.3) vidíme, že je možné, aby za sebou priamo nasledovali aj dva kúsky hélixu. Dva hélixu oddelené na jednej strane malým "výrastkom" (v angličtine nazývame tento vyčnievajúci kúsok bulge) sú bežný jav v sekundárnych štruktúrach. Na obrázku 3.4 vidíme, ako by vyzerala štruktúra zodpovedajúca tomuto deskriptoru.

¹International Union of Pure and Applied Chemistry


```

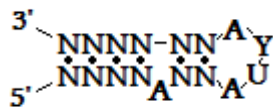
# r17.des
#
# descriptor of the consensus binding site for
# the phage R17 coat protein: a stem with
# a single-base A bulge and a 4-base loop
#

h1 s1 h2 s2 h2' h1'

h1 0:0 NNNN:NNNN
h2 0:0 NN:NN
s1 0   A
s2 0   AUYA

```

Obr. 3.3: Príklad platného deskriptora pre RNABob.



Obr. 3.4: Sekundárna štruktúra zodpovedajúca deskriptoru.

Kapitola 4

Implementácia programu

Hlavným cieľom tejto práce je vytvoriť program, ktorý umožňuje vytváranie deskriptorov pre RNABob. Chceli by sme užívateľsky čo najpríjemnejší systém editovania. Umožníme teda užívateľovi vytvárať základné štrukturálne prvky, spájať ich a editovať ich vlastnosti. V tejto kapitole rozoberieme použité technológie, návrh štruktúry a zaujímavé problémy, s ktorými som sa počas implementácie stretol.

4.1 Použité technológie

Prioritou pri vyberaní prostredia, v ktorom som vyvíjal grafické prostredie, bola ľahká prenositeľnosť medzi jednotlivými platformami. Nakoniec som sa rozhodol pre framework Qt[8], ktorý je vyvíjaný Nokiou, pretože podporuje širokú paletu platforiem, používa C++ a prichádza s užívateľsky veľmi prístupným dizajnérom formulárov.

4.2 Návrh štruktúry

Pri rozhodovaní ako navrhnuť triedy v mojom programe som sa z veľkej časti opieral o praktickosť pri grafickej manipulácii a editovaní. Rozhodol som sa nevytvoriť oddelenú štruktúru na pozadí a grafickú reprezentáciu objektov, ale spojiť obe funkčnosti do jednej triedy, keďže obe majú výrazne podobné

grafové štruktúry a všetky operácie sa dajú vykonávať priamo na grafických objektoch.

Pri analyzovaní štruktúry som sa rozhodol pre dva základné typy objektov, ktoré jasne plynú z formátu deskriptorov, a to hélixy a stemy. Ďalej som potreboval vytvoriť 3' a 5' ako významné body štruktúry. Nakoniec som ako základnú štruktúru použil jednoduchý graf s vrcholmi stupňa najviac dva. Základné objekty sú vrchol (trieda *Node*) a hrana (trieda *Arrow*), ktorá spája dva vrcholy. Trieda *Helix* funguje ako overhead nad dvomi hranami aj s ich koncovými vrcholmi, združuje ich ako ich rodičovský objekt, spravuje údaje pre vytváranie deskriptora a riadi ich geometriu.

Trieda *Helix* drží informácie potrebné pre vytvorenie deskriptora, čiže predpis oboch nukleotidových reťazcov (spolu s informáciou, ktorý je na hrane bližšie k 5' a ktorý na hrane bližšie k 3'), a počet dovolených nezhôd. Zároveň je možné zmeniť hélixu na reláciu (sú reprezentované jedným objektom, pretože správanie sa týchto prvkov je pre naše potreby úplne rovnaké).

Ak objekt z triedy *Arrow* neprislúcha žiadnemu hélixu, drží si všetky tieto informácie sám.

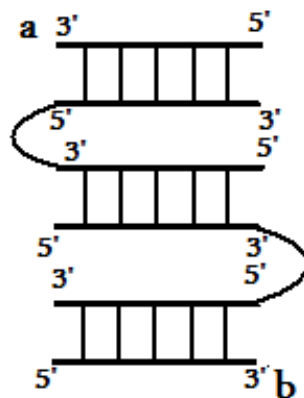
Tento prístup mi neumožňuje, aby dva hélixy zdieľali vrchol, a teda nemôžem vymodelovať štruktúru, kde za sebou nasledujú dva hélixy, preto bolo nutné vytvoriť epsilonové hrany, ktoré slúžia len na topologické usporiadanie štruktúry, ale nereprezentujú žiaden nukleotidový reťazec. Označenie hrany za epsilonovú sa rieši nastavením príslušného predpisu nukleotidového reťazca na 'e', táto hrana sa vo výslednom deskriptore nezobrazí.

4.3 Kontrola orientácie reťazcov v hélixe

Program má implementované niektoré podporné funkcie na priebežné kontrolovanie štruktúry, ktoré znemožňujú pridať hranum ktorá by vytvorila nezmyselnú štruktúru.

Ako sme už spomínali v kapitole 2, nemôžu byť obe vlákna hélixu orientované v rovnakom smere. Môj program preto neumožňuje prídanie hrany, ktorá by niečo takéto spôsobila. Problém nám spôsobuje, že pokiaľ nie je

komponent spojený s 5' alebo 3', nevieme ako bude hrana orientovaná. Niekedy však vieme už z relatívnej orientácie héličov povedať, že hranu pridať nemožno (obr. 4.1).



Obr. 4.1: Vidíme, že vrcholy a a b nejde spojiť hranou, pretože by sme ju nevedeli zorientovať.

Na overovanie orientovania sme implementovali algoritmus podobný algoritmu Union Find. Držíme si informáciu o komponentoch, ktoré sú nezávisle orientovateľné. Každému z týchto komponentov priradíme na začiatku náhodnú orientáciu. Niektoré tieto komponentov môžu byť prevrátiteľné (môžeme zameniť 5' konce za 3' konce a opačne).

Na začiatku je každý hélič samostatný, prevrátiteľný komponent a 5' a 3' tvoria samostatné komponenty, ktoré už preorientovať nejde. Pred pridaním každej hrany sa pozrieme na vrcholy, ktoré spája. Rozlišujeme niekoľko možností:

- *Oba vrcholy patria do jedného komponentu, potom vieme ľahko rozhodnúť, či bude hranu možné zorientovať správne.*
- *Vrcholy sú z rôznych komponentov, pričom aspoň v jednom je možné prevrátiť orientácie. Potom je hranu možné pridať a v prípade potreby v prevrátiteľnom komponente otočiť orientácie.*
- *Vrcholy sú z rôznych komponentov, ktoré už nemožno prevracať. Vieme*

rozhodnúť, či je hrana korektne orientovaná, a ak nie, nemožno ju pridať.

Ak je možné hranu pridať, pridáme ju a spojíme oba komponenty do jedného. Tento nový komponent je orientovateľný práve vtedy, ak boli oba spájané komponenty orientovateľné.

Ako vidíme, pridanie hrany do tejto štruktúry trvá $O(N)$, kde N je počet vrcholov. Keďže N je pomerne malé, je táto zložitosť pre nás úplne postačujúca, z teoretického hľadiska by však algoritmus riešiaci tento problém rýchlejšie mohol byť zaujímavý.

Problém môže nastať po tom, ako niektoré hrany, alebo hélixy boli zmazané. Preto vždy po odstránení nejakého objektu pred najbližším overovaním štruktúru vytvoríme nanovo.

4.4 Detekcia cyklov

Detekcia cyklov prebieha jednoducho. Ak spájame vrcholy A a B , majú oba tieto vrcholy stupeň najviac 1. Keďže náš štruktúrálny motív je graf stupňa najviac 2, vieme, že komponent, do ktorého patrí A , je cesta. Pozrieme sa na posledný vrchol tejto cesty a ak je ním B , vieme, že pridaním hrany by sme vytvorili cyklus. Ak je koncový vrchol cesty iný, môžeme hranu pridať.

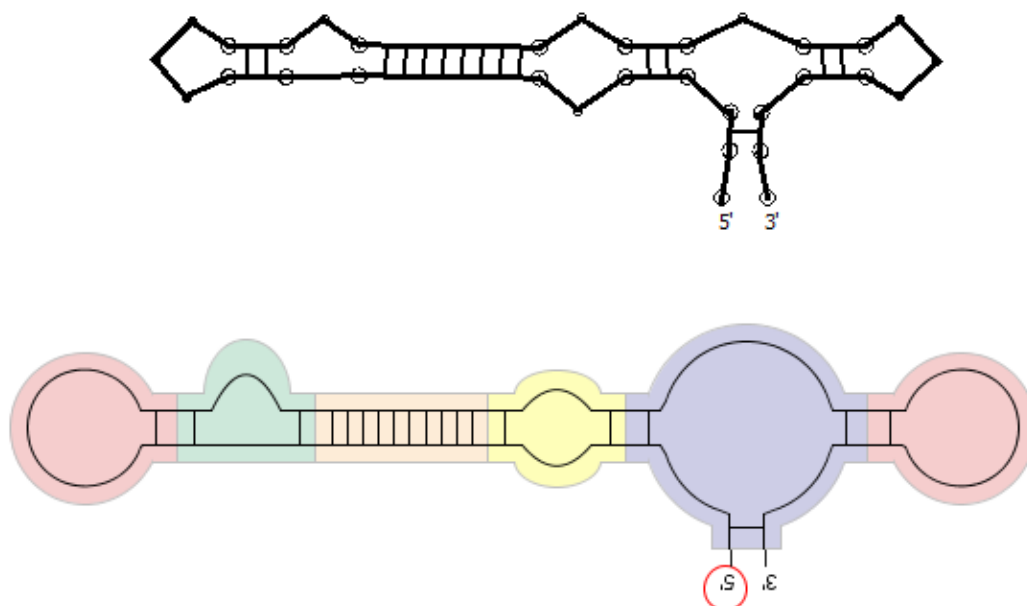
4.5 Porovnanie s existujúcimi prostrediami

Jediné nám známe grafické prostredie, ktoré sa zaoberá tvorbov motívov, je Locomotif [6]. Líši sa od nášho programu ako v rozsahu funkčnosti, tak v spôsobe editovania motívu.

Prvým rozdielom, ktorý treba vyzdvihnúť, je to, že Locomotif nemá podporu štruktúr obsahujúcich pseudouzly. Keďže RNABob, pre ktorý sme naše GUI primárne vyvíjali, umožňuje ľubovoľné štruktúry, prirodzene sme chceli vytváranie pseudouzlov v našom prostredí umožniť.

Ďalším rozdielom je štruktúralna úroveň, na ktorej funguje editovanie motívov. Kým Locomotif má bloky tvorené zoskupeniami, ktoré sa v štruktúrach

vyskytujú, my vytvárame motívy priamo zo stemov a hélixov (obr. 4.2). Aby Locomotif udržal iba jeden motív, umožňuje mazanie a pridávanie iba na koncoch reťazcov. Preto ak potrebujeme zmeniť štruktúru nachádzajúcu sa v strede, musíme zmazať všetky ostatné časti tak, aby sa ocitla na kraji. Náš program overí súvislosť štruktúry až pri exporte do deskriptorového súboru, čiže umožňuje editáciu v ľubovoľnej časti motívu.



Obr. 4.2: Príklad štruktúry vytvorenej v našom GUI (hore) a v Locomotife (dole).

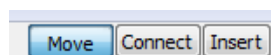
Kapitola 5

Užívateľský manuál

Táto kapitola obsahuje prehľadnú užívateľskú príručku, ktorá opisuje funkcie nášho GUI.

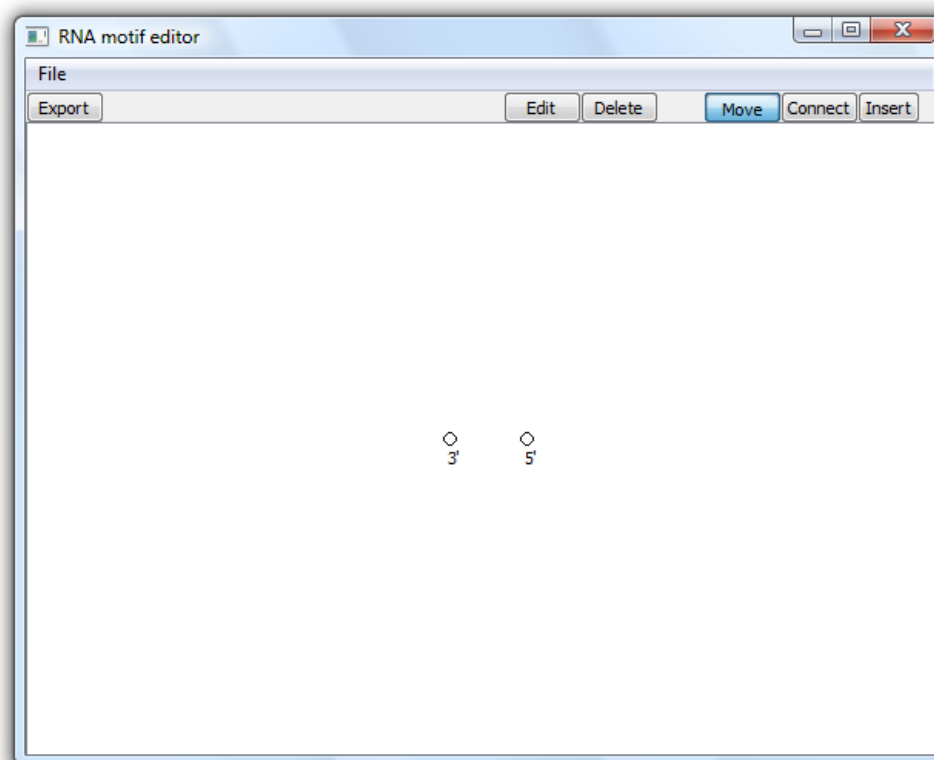
5.1 Hlavné okno

Po spustení sa objaví hlavné okno aplikácie (obr 5.2). Tvorí ho menu, niekoľko tlačidiel a grafická scéna, v ktorej prebieha editácia RNA motívu. Po spustení sa v nej nachádzajú iba vrcholy reprezentujúce konce 5' a 3'.



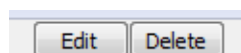
Obr. 5.1: Výber módu.

Skupina tlačidiel na pravom okraji (obr 5.1) slúži na výber módu, v ktorom bude grafická scéna pracovať. V závislosti od práve aktívneho módu sa grafická scéna rozhoduje ako interpretovať akcie spravené myšou. Tlačidlo symbolizujúce práve aktívny mód ostáva stlačené. O týchto módoch sa dozvieme viac neskôr.



Obr. 5.2: Hlavné okno aplikácie

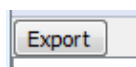
Pomocou ďalšej skupiny tlačidiel v strede (obr 5.3) môžeme robiť akcie s práve vybraným objektom. A to buď tento objekt zmazať tlačidlom Delete, alebo nastaviť jeho vlastnosti tlačidlom Edit.



Obr. 5.3: Zmazanie a editácia objektu.

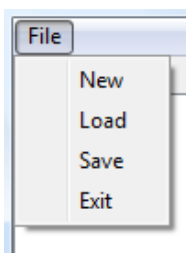
Tlačidlo Export (obr 5.4) v ľavom rohu slúži na vytvorenie deskriptora, po jeho zmačknutí sa objaví dialóg, v ktorom vyberieme súbor, do ktorého chceme deskriptor uložiť. Ak je štruktúra uzavretá (existuje cesta z 5' k 3'), program do zvoleného súboru exportuje deskriptor, inak užívateľovi oznámi, že došlo k chybe.

V menu(obr. 5.5) sa nachádzajú tlačidlá, ktoré nám umožňujú zmazať celú scénu (New), uložiť stav scény vrátane všetkých vlastností do súboru



Obr. 5.4: Export deskriptora do súboru.

(Save), načítať uložený stav scény a vlastností objektov zo súboru (Load) a ukončiť aplikáciu (Exit).



Obr. 5.5: Menu.

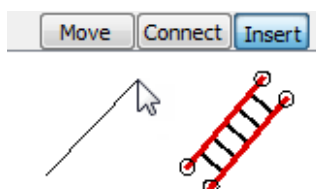
5.2 Scéna

V scéne prebieha všetka práca s RNA motívmi. Na začiatku obsahuje iba dva objekty, 5' a 3'.

5.2.1 Insert mód

Ak sa scéna nachádza v insert móde, slúži kliknutie myškou na vkladanie objektov. Máme dve možnosti, buď vkladáme hélixy, alebo pomocné vrcholy, ktoré slúžia na nepriame vedenie nespárovaných úsekov.

Ak stlačíme ľavé tlačidlo myši, začne sa z pozície ťahať čiara, ktorá symbolizuje miesto, kde sa po pustení tlačidla objaví hélixa (obr. 5.6).

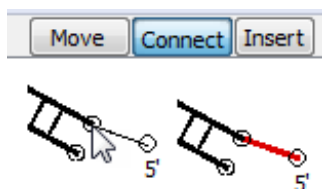


Obr. 5.6: Vkladanie hélixy.

Pravým kliknutím vložíme pomocný vrchol na pozíciu myši.

5.2.2 Connect mód

Connect mód slúži na vytváranie nespárovaných úsekov medzi vrcholmi hélíxov. Stlačíme ľavé tlačidlo myši na mieste, odkiaľ chceme viesť nespárovaný úsek, a natiahneme priamku k miestu, s ktorým ho chceme spojiť. Ak sa na oboch pozíciách nachádza vrchol, ktorému možno pridať hranu, a ak pridanie tejto hrany nevytvorí cyklus ani nepokazí relatívnu orientáciu, táto hrana sa pridá (obr 5.7). Okrem toho sa v tomto móde dá meniť dĺžka hélíxov: vyberieme hélix kliknutím ľavým tlačidlom a otočením kolieska myši môžeme zmeniť jeho dĺžku.



Obr. 5.7: Vkladanie nespárovaného úseku.

5.2.3 Move mód

Tento mód slúži na zmenu polohy objektov na scéne. Jednoduchým chytením ľavým tlačidlom a ťahaním môžeme objekt pod kurzorom presúvať. Navyše ak vyberieme hélix, môžeme ho otočením kolieska rotovať okolo jeho stredu.

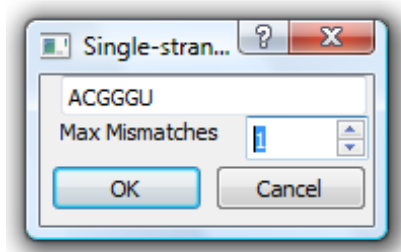
5.3 Objekty

Za vybrateľné objekty scéna považuje hélíxy a hrany, ktoré nie sú vnútornou súčasťou hélíxov. Každému z týchto objektov je po jeho vybratí možné nastaviť jeho vlastnosti (potrebné pre vytvorenie deskriptoru), prípadne ho zmazať. Nastavenie vlastností sa uskutočňuje prostredníctvom dialógu pre nespárovaný úsek (obr. 5.8), resp. pre hélix (obr. 5.9). Tu sa dá nastaviť

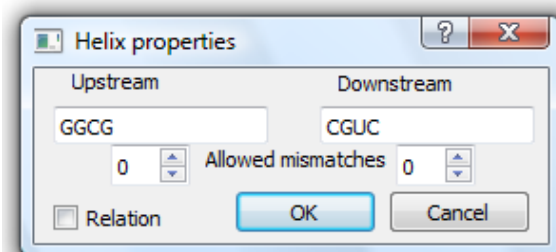
predpis nukleotidových reťazcov, počet nezhôd, prípadne či má ísť o hélix, alebo reláciu.

Epsilonové hrany

Ak má hrana nastavený ako predpis reťazca 'e', počíta sa ako neexistujúca a pri tvorbe deskriptora sa správa akoby oba jej koncové body boli jeden objekt. Tieto hrany sú vykreslené prerušovanou čiarou. Hrany susediace s pomocnými bodmi sa automaticky nastavujú na epsilonové, čo v prípade potreby môže užívateľ zmeniť v menu.



Obr. 5.8: Dialóg pre nespárovaný úsek.



Obr. 5.9: Dialóg pre hélix/reláciu.

Vybrané objekty je možné zmazať tlačidlom Delete. Ak je vymazávaný objekt hélix a sú naň napojené nejaké hrany, tie sú vymazané zo scény tiež. Pomocné body nie je možné zmazať priamo, ale zmiznú automaticky, ak sú odobrané všetky hrany, ktoré z nich viedli. Vrcholy 3' a 5' sa zmazať nedajú, pretože každá štruktúra ich musí obsahovať.

5.3.1 Farebné zobrazenie

Hrany na scéne vyjadrujú svojou farbou informáciu o svojom stave. Sú to nasledujúce farby:

- Červená: hrana alebo jej hélix ešte nemá nastavený predpis nukleotidov.
- Zelená: hrana alebo jej hélix je práve vybraný objekt
- Čierna: hrana alebo jej hélix nie je momentálne vybraná, ale má už nastavený svoj predpis pre deskriptor

Kapitola 6

Záver

Zostrojili sme grafické prostredie, ktoré umožňuje editáciu štruktúrnych RNA motívov a export do deskriptorového formátu používaného programom RNABob. Jediný spôsob, akým bolo možné doteraz vytvárať pre tento program deskripty, bol textový editor. Priniesli sme teda nový, intuitívny spôsob ako pristupovať k tomuto vytváraniu motívov.

Budúca práca na projekte

Jeden zo smerov, ktorým by sa mohol uberať ďalší vývoj programu, by bolo pridať možnosť importovať textové deskripty RNABobu a ich rozmiestnenie na grafickej scéne. Toto by umožnilo ľahkú úpravu štruktúr, pre ktoré už deskripty existujú, pretože by nebolo treba vytvárať ich v našom prostredí odznova.

Druhou cestou, po ktorej by sa mohla vydať ďalšia práca na projekte, by bolo rozoznávanie štruktúry podľa obrázkov. Sekundárna štruktúra RNA býva biológmi väčšinou kreslená jedným ťahom v smere od 5' k 3' s prípadnými spárovanými úsekmi dokreslenými dodatočne. Verím preto, že možnosť voľne kresliť štruktúru, ktorú by program automaticky rozoznával a transformoval na objekty, by bola pre ich pohodlnú prácu veľkým skokom vpred.

Literatúra

- [1] Durbin, R., Eddy, S., Krogh, A., and Mitchison, G. (1998). Biological sequence analysis. Cambridge University Press.
- [2] MedicineNet.com. "Transcription definition"
- [3] Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P.(2002). Molecular Biology of the Cell. Garland Science
- [4] Gautheret, D; F Major, R Cedergren. "Pattern searching/alignment with RNA primary and secondary structures: an effective descriptor for tRNA". Computer Applications in the Biosciences: CABIOS 6 (4): 325-331. ISSN 0266-7061
- [5] Macke TJ, Ecker DJ, Gutell RR, Gautheret D, Case DA, Sampath R (November 2001). "RNAMotif, an RNA secondary structure definition and search algorithm". Nucleic Acids Res. 29 (22): 4724–35. PMID 11713323
- [6] Reeder J, Reeder J, Giegerich R (July 2007). "Locomotif: from graphical motif description to RNA motif search". Bioinformatics 23 (13): i392–400. doi:10.1093/bioinformatics/btm179. PMID 17646322
- [7] S. R. Eddy. RNABOB: a program to search for RNA secondary structure motifs in sequence databases.
- [8] Qt - A cross-platform application and UI framework. <http://qt.nokia.com/>