

Vizualizace sekundární struktury RNA s využitím existujících struktur

Richard Eliáš

Matematicko-fyzikální fakulta

richard.elias@matfyz.cz

Motivácia a ciele práce

Molekula RNA sa stáva predmetom mnohých štúdií, vďaka čomu rastie dopyt po nástrojoch pomáhajúcich pri jej analýze. Vlastnosti molekuly sú síce ovplyvnené primárnou štruktúrou (poradím nukleotidov v reťazci), no viac závisia na ich priestorovom usporiadaní (terciárna štruktúra). My sa v práci zaoberáme trochu zjednodušeným modelom - sekundárnou štruktúrou. Tú reprezentuje zoznam nukleotidov spojených väzbou. Tieto nukleotidy musia byť blízko aj v priestore a tak nám sekundárna štruktúra relatívne dobre aproximuje terciárnu, pre ktorú neexistujú spoľahlivé metódy zisťovania štruktúry už ani pre relatívne malé molekuly.

Prvým krokom pri analýze RNA molekuly je často rozbor obrázka jej sekundárnej štruktúry. Medzi základné kritéria ktoré musia obrázky molekúl spĺňať patrí rovinnosť nakreslenia, kreslenie loopov na kružnice a stemy na priamkách. Pri porovnávaní štruktúr sa využíva taktiež kreslenie častí majúcich podobnú funkciu a tvar na rovnaké miesta v obrázkoch, čo pomáha lepšej orientácii v molekule a napomáha nájsť konzervované časti v molekulách. V súčasných nástrojoch (mFold, RNAviz, RNAView, ...) sa toto posledné kritérium nedodržuje, čo má za následok ťažké nachádzanie konzervovaných častí v molekulách.

Kreslenie chýbajúcich častí v molekule

XYZ

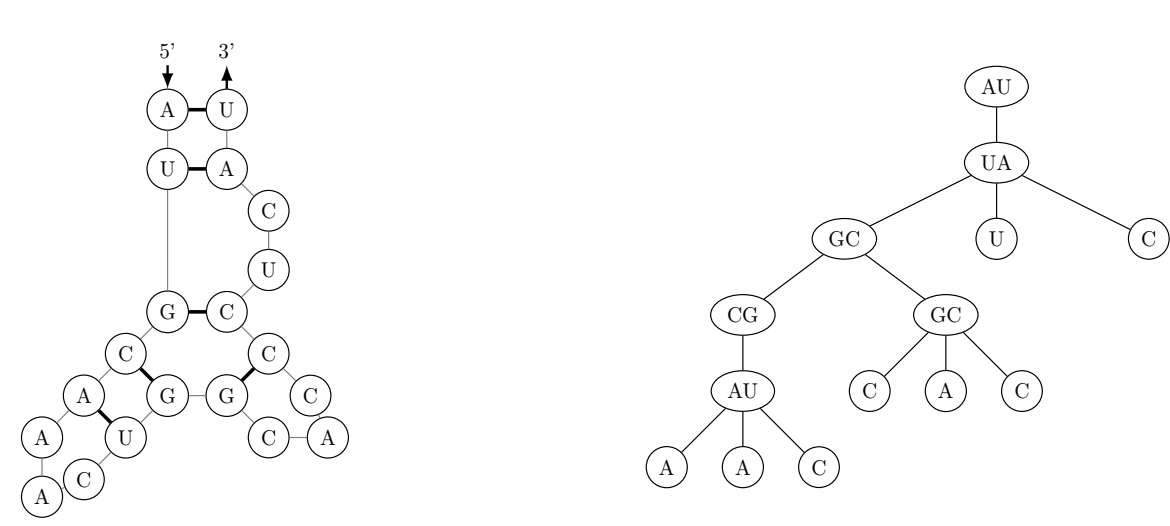
Obrazky, grafy, statistika

Grafl

XYZ

ABC

Stromová reprezentácia RNA a použitie tree-edit-distance algoritmu



$$\delta(\emptyset, \emptyset) = 0$$
$$\delta(F, \emptyset) = \delta(F - r_F, \emptyset) + c_{del}(r_F)$$
$$\delta(\emptyset, G) = \delta(\emptyset, G - r_G) + c_{ins}(r_G)$$
$$\delta(F, G) = \begin{cases} \delta(F - r_F, G) + c_{del}(r_F) \\ \delta(F, G - r_G) + c_{ins}(r_G) \\ \delta(F - r_F, G - r_G) + \delta(r_F - r_F, r_G - r_G) + c_{upd}(r_F, r_G) \end{cases}$$

Vďaka zanedbaniu existencie pseudouzlov, môžeme sekundárnu štruktúru reprezentovať ako usporiadaný zakorenený strom (pseudouzly sú páry vedúce medzi vetvami stromu). Transformácia sekundárnej štruktúry do stromovej podoby je na obrázku. Vnútorný vrchol stromu reprezentuje báзовý pár a listy nespárované nukleotidy.

Pôvodným nakreslením chceme hýbať čo najmenej a tak potrebujeme spôsob ako odlišiť časti, ktoré sa v oboch molekulách nemenia. K tomu využijeme algoritmus *tree-edit-distance*, ktorý nám dá návod ako určiť najmenší počet úprav, ktorými vieme transformovať šablónovú molekulu na cieľovú. Úpravami myslíme editačné operácie update, insert a delete (zmena bázy vo vrchole, vloženie lebo zmazanie vrcholu).

Základom algoritmu je rekurzívny vzorec, ktorý určí vzdialenosť medzi dvoma stromami (*r* označuje najpravejší, alebo najľavejší vrchol lesa *F* a *G*, *c_{del}*, *c_{ins}*, *c_{upd}* sú ceny mazania, vkladania a updatu vrcholu v strome).

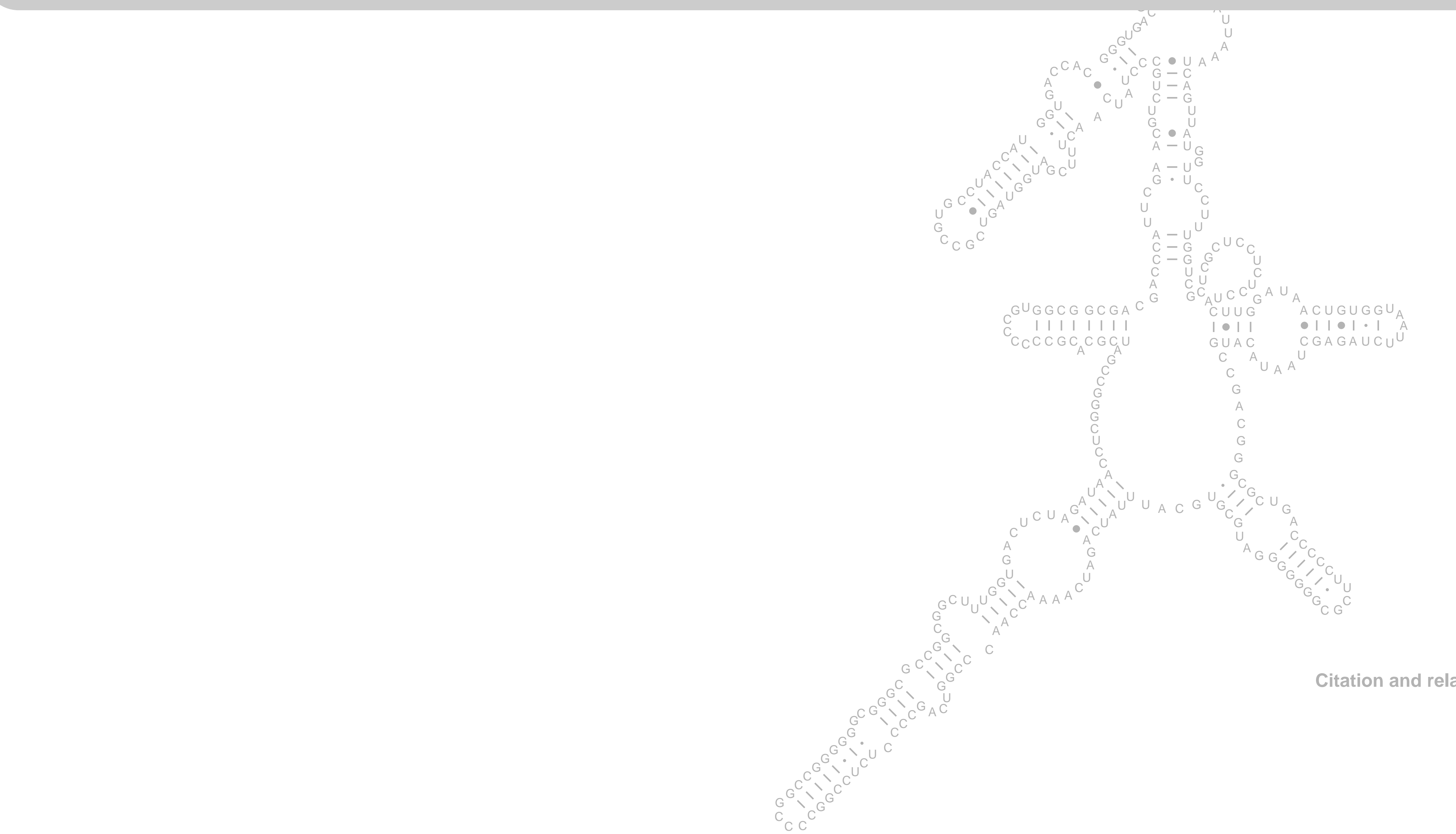
.. tu bude este nieco pokracovat

Nástroj TRAVeLer

XYZ

Výsledky

Tu budu zhrnute vysledky prace



- Homo sapiens* (K03432)
1. Eukaryota
 2. eukaryote crown group
 3. Fungi/Metazoa group
 4. Metazoa
 5. Eumetazoa
 6. Bilateria
 7. Coelomata
 8. Deuterostomia
 9. Chordata
 10. Craniata
 11. Vertebrata
 12. Gnathostomata
 13. bony vertebrates
 14. lobe-finned fish and tetrapod clade
 15. Tetrapoda
 16. Amniota
 17. Mammalia
 18. Theria
 19. Eutheria
 20. Primates
 21. Catarrhini
 22. Hominidae
 23. Homo
- June 2004