



---

**POZNAN UNIVERSITY OF TECHNOLOGY**

---

FACULTY OF COMPUTING AND TELECOMMUNICATIONS

Institute of Computing Science

Streszczenie rozprawy doktorskiej

**Combinatorial Analysis  
of RNA Tertiary Structures  
with the Use of Angular  
Representations**

**Jakub Wiedemann, M.Sc.**

Supervisor: **Maciej Antczak, Ph.D., Dr Habil.**

Co-Supervisor: **Maciej Miłostan, Ph.D.**

Poznań, 2022

Niniejsza praca doktorska poświęcona jest badaniom nad strukturą cząsteczek RNA i tworzeniu metod, głównie kombinatorycznych, pozwalających ją analizować. W badaniach skoncentrowano się na analizie charakterystyk kątowych struktur przestrzennych (3D) RNA w kontekście porównywania struktur realizowanego w ramach ewaluacji modeli uzyskiwanych z wykorzystaniem metod obliczeniowych, a także identyfikacji specyficznych motywów strukturalnych zwanych pętlami wieloramiennymi. Prezentowane w pracy badania rozpoczęły się od analizy związku pomiędzy sekwencją cząsteczki RNA, a jej strukturą trzeciorzędową. Uzyskane wyniki wykazały, że stosunkowo wysokie podobieństwo sekwencyjne nie zawsze gwarantuje zachowanie kształtu przestrzennego cząsteczek. Ponadto badania przeprowadzone podczas opracowywania narzędzia *StructAnalyzer* pozwoliły zaobserwować, że struktury, które w ujęciu globalnym różnią się znacząco wykazują często wyraźne podobieństwo w ujęciu lokalnym. Potwierdziło to istotność problemu wyszukiwania lokalnych, konserwatywnych motywów pomiędzy pozornie różniącymi się od siebie strukturami. Konserwatywne motywy wskazują zwykle, że struktury, które je zawierają prawdopodobnie pełnią zbliżoną funkcję biologiczną.

Scharakteryzowane krótko wyniki zainspirowały opracowanie nowej miary, która ocenia podobieństwo struktur 3D RNA z lokalnej perspektywy. Mnogość struktur oraz złożoność procesu porównywania wielu struktur skierowały nas w kierunku wykorzystania reprezentacji kątowej do opisywania struktur przestrzennych, gdyż ta reprezentacja pozwala na pominięcie procesu uliniawiania struktur przestrzennych RNA. Opracowany algorytm *LCS-TA* pozwala na identyfikację najdłuższych ciągłych segmentów charakteryzujących się określonym współczynnikiem podobieństwa (bazującym na mierze *MCQ*) nie przekraczającym oczekiwanego przez użytkownika odcięcia pomiędzy porównywaną parą struktur 3D RNA. W rezultacie zwracana jest lokalizacja dopasowanych do siebie segmentów i ich dłu-

gość (będącą dość intuicyjną miarą podobieństwa analizowanych struktur). Zaproponowany algorytm został wykorzystany do oceny modeli 3D RNA zgłoszonych w rundzie IV konkursu RNA-Puzzles oraz udostępniony w ramach zestawu narzędzi wykorzystywanych przez społeczność RNA-Puzzles. Analizy wyników uzyskiwanych w ramach konkursu RNA-Puzzles wskazują jednoznacznie na istnienie motywów, które są szczególnie trudne do przewidzenia obliczeniowo. Jednym z nich są pętle wieloramienne (ang. *N*-way junctions). Motywy te, które często obejmują niekanoniczne pary zasad, znacząco wpływają na proces fałdowania cząsteczek RNA i ich ostateczny kształt. Ponadto analiza dostępnych źródeł wskazywała na brak na bieżąco uaktualnianych repozytoriów udostępniających znane konformacje obserwowane w eksperymentalnie określonych strukturach 3D RNA, co sprawia, że ich wykorzystanie chociażby podczas modelowania homologicznego często następuje trudności. W celu wypełnienia zidentyfikowanej luki opracowano bazę danych *RNAloops*, aby w sposób w pełni zautomatyzowany gromadzić i udostępniać w ramach jednego repozytorium informacje o multipętlach zidentyfikowanych w eksperymentalnie określonych strukturach przestrzennych cząsteczek RNA. Baza wyposażona w przyjazny użytkownikowi interfejs udostępnia następujące informacje o multipętlach: sekwencja, struktury 2D i 3D, oraz relacje zachodzące pomiędzy kolejnymi parami sąsiadujących rozgałęzień z wykorzystaniem trzech kątów Eulera oraz jednego kąta płaskiego. Tego typu przestrzenne zależności mogą zostać wykorzystane podczas eksperckiego modelowania 3D RNA, które zakłada wyszukiwanie obiecujących multipętli je spełniających. Korzyścią z wykorzystania kątowej reprezentacji w tym zastosowaniu jest większa uniwersalność procedury wyszukiwania obiecujących motywów. Platforma ponadto umożliwia użytkownikom elastyczne wyszukiwanie interesujących rekordów na podstawie szeregu kryteriów, m.in., sekwencji, struktury drugorzędowej, czy liczby rozgałęzień poszukiwanej pętli.

Funkcjonalności te wspierają, m.in., ekstrakcję motywów o określonych cechach, ich analizę porównawczą czy modelowanie struktur przestrzennych charakteryzujących się określonymi własnościami np. podczas projektowania rozwiązań terapeutycznych.