

## **Eksploracja cech i modelowanie struktury kwadrupleksów**

**Marta Szachniuk**

*Instytut Informatyki, Politechnika Poznańska*

Tematyka projektu osadzona jest w bioinformatyce będącej jednym z najmłodszych obszarów badawczych nauk obliczeniowych. Bioinformatyka pomaga formułować teoretyczne modele zjawisk, które są przedmiotem badań w naukach o życiu oraz opracowuje metody obliczeniowe pozwalające na komputerowe symulowanie tych zjawisk i ich analizę. Wspomaga również badania eksperymentalne prowadzone w naukach biologicznych dostarczając narzędzi do ich przetwarzania.

Niniejszy projekt skupia się na kwadrupleksach, specyficznych 4-niciowych motywach, które formują się w cząsteczkach kwasów nukleinowych (DNA i RNA). Ich istnienie w komórkach organizmów żywych potwierdzono stosunkowo niedawno. Ostatnie dziesięciolecie zaowocowało intensyfikacją badań eksperymentalnych nad tymi motywami strukturalnymi. Potwierdzono ich duży potencjał w biomedycynie, m.in. w diagnostyce medycznej oraz terapii niektórych chorób, w tym także nowotworowych. Z dużym zaangażowaniem analizuje się obecnie możliwe funkcje kwadrupleksów w procesach komórkowych. Funkcje cząsteczek są z kolei ściśle związane z ich budową. Analiza funkcji i struktury molekuł stanowią więc badania komplementarne, które w dużym stopniu wykonywane są *in silico*. W przypadku kwadrupleksów, nie istnieje jednak jeszcze zbyt wiele metod komputerowych, które wspomagają badanie tych motywów strukturalnych. Dotychczas powstało ok. 10 algorytmów, które na podstawie sekwencji nukleotydowej przewidują potencjalne miejsca ich tworzenia oraz dwie bazy danych, które gromadzą wybiórcze informacje dotyczące tych motywów [1]. Niedawno pojawiła się też możliwość znajdowania kwadrupleksów w strukturach 3D określonych w drodze eksperymentu. Oferują ją dwa programy, DSSR [2] oraz RNApdbec [3]. Ten drugi program został opracowany przez nasz zespół i opublikowany w 2018 roku. Analiza kwadrupleksów oraz wchodzących w ich skład tetrad, którą przeprowadziliśmy z użyciem programu RNApdbec pozwoliła nam zaobserwować wzorce topologiczne w strukturach tych motywów. Zaowocowało to zdefiniowaniem przez nas nowej klasyfikacji ONZ dla tetrad i kwadrupleksów bazującej na cechach topologicznych ich struktur [4].

Kontynuując badania nad kwadrupleksami planujemy opracować metody obliczeniowe, które umożliwią kompleksową analizę ich struktury na wszystkich poziomach: pierwszo-, drugo- i trzeciorzędowym. W ramach projektu stworzymy narzędzie do wyznaczania różnych parametrów opisujących właściwości struktury kwadrupleksu. Przeanalizujemy wszystkie dostępne publicznie struktury kwasów nukleinowych zawierających kwadrupleksy pod kątem wartości tych parametrów. Przeprowadzimy analizy statystyczne oraz klastrowanie struktur na podstawie wybranego podzbioru cech strukturalnych. Przebadamy właściwości struktur dla różnych klas kwadrupleksów w nomenklaturze ONZ. Planujemy również zbadać relację między architekturą kwadrupleksu a jego potencjalną funkcją oraz otoczeniem kwadrupleksu w cząsteczce kwasu nukleinowego. W badaniach tych wykorzystamy sieci neuronowe i metody uczenia maszynowego. Stworzymy bazę danych, która będzie gromadziła kompleksową informację o strukturach tetrad i kwadrupleksów oraz pozwoli efektywnie wyszukiwać dane strukturalne. Planujemy również opracowanie metody modelowania struktur kwasów nukleinowych zawierających motyw kwadrupleksu. Będzie ona bazowała na modelach uzyskanych z uczenia maszynowego oraz na silniku obliczeniowym systemu RNAComposer [5], który opracowaliśmy dla przewidywania struktury 3D RNA.

Opracowane przez nas narzędzia bioinformatyczne zostaną udostępnione w ramach RNAPolis [6] - platformy obliczeniowej dedykowanej bioinformatyce strukturalnej kwasów nukleinowych. Przyczynią się one w sposób istotny do rozwoju badań relacji struktura - funkcja w zakresie kwasów nukleinowych. Wpłyną na rozwój nowych metod modelowania i projektowania cząsteczek kwasów nukleinowych z kwadrupleksami. Będzie to miało duże znaczenie dla projektowania nowych dróg diagnostyki i terapii zorientowanej na DNA i RNA czy generowania sztucznych cząsteczek kwasów nukleinowych na potrzeby nanotechnologii.

- [1] E. P. Lombardi, A. Londono-Vallejo, A guide to computational methods for G-quadruplex prediction, *Nucleic Acids Research*, gkz1097, 2019.
- [2] X.-J. Lu, H. J. Bussemaker, W. K. Olson, DSSR: an integrated software tool for dissecting the spatial structure of RNA, *Nucleic Acids Research* 43, 2015, 142.
- [3] T. Zok, M. Antczak, M. Zurkowski, M. Popena, J. Blazewicz, R.W. Adamiak, M. Szachniuk, RNApdbec 2.0: multifunctional tool for RNA structure annotation, *Nucleic Acids Research* 46(W1), 2018, W30-W35.
- [4] M. Popena, J. Miskiewicz, J. Sarzynska, T. Zok, M. Szachniuk, Topology-based classification of tetrad and quadruplex structures, *Bioinformatics*, 2019, in press (epub: 7.10.2019, doi:10.1093/bioinformatics/btz738).
- [5] Popena, M, Szachniuk, M, Antczak, M, Purzycka, KJ, Lukasiak, P, Bartol, N, Blazewicz, J, Adamiak, RW, Automated 3D structure composition for large RNAs. *Nucleic Acids Research* 40, 2012, e112.
- [6] M. Szachniuk, RNAPolis: computational platform for RNA structure analysis, *Foundations of Computing and Decision Sciences* 44(2), 2019, 241-257.