Modelowanie struktury RNA ukierunkowane na motywy funkcjonalne

Marta Szachniuk

Instytut Informatyki, Politechnika Poznańska

Tematyka niniejszego projektu badawczego jest zanurzona w świecie bioinformatyki - nowatorskiej dziedziny, która łączy biologię z informatyką, aby rozwikłać tajemnice życia. Bioinformatyka dostarcza naukowcom narzędzi do tworzenia modeli teoretycznych zjawisk biologicznych oraz opracowywania metod obliczeniowych do ich symulacji i analizy. Te modele i metody mogą być wykorzystywane między innymi do wyjaśniania procesów życiowych na poziomie komórkowym lub projektowania konstruktów molekularnych dla biotechnologii czy biomedycyny. W bioinformatyce – podobnie jak w innych dziedzinach nauki – rozwinęły się różne specjalizacje. Na przykład bioinformatyka strukturalna, z którą związany jest nasz projekt, koncentruje się na badaniu białek, DNA i RNA, ich struktur, cech strukturalnych, charakterystycznych motywów oraz zależności między strukturą a funkcją. W centrum naszych badań leży RNA. Jest to niezwykle wszechstronna i potężna cząsteczka, która znajduje zastosowanie w terapii, diagnostyce oraz systemach biologicznych do przetwarzania informacji *in vivo*, takich jak sieci regulacji genów czy syntetyczne obwody biologiczne. RNA może wykonywać obliczenia, przechowywać dane i kontrolować procesy biologiczne w środowisku komórkowym. Funkcja RNA jest ściśle związana z jego strukturą. Szczególne znaczenie funkcjonalne przypisuje się motywom strukturalnym, które są przedmiotem wielu badań eksperymentalnych. Wzorce te nie są tylko powtarzającymi się sekwencjami kodu genetycznego. Niektóre dobrze znane motywy – takie jak trypleksy, kwadrupleksy, itp. – to skomplikowane trójwymiarowe struktury, które mogą być złożone z wielu nici utrzymywanych razem przez zarówno kanoniczne, jak i niekanoniczne interakcje bliskiego i dalekiego zasięgu. Cechą wyróżniającą inne wzorce, takie jak szpilki do włosów czy pętle wewnętrzne, jest topologia struktury drugorzędowej, która określa wzajemne ułożenie jedno- i dwuniciowych fragmentów cząsteczki RNA. Zrozumienie wzorców strukturalnych może doprowadzić do znaczących postępów w medycynie i biotechnologii, zwłaszcza jeśli będziemy potrafili projektować cząsteczki zawierające specyficzne motywy funkcjonalne. Przewidywanie i modelowanie struktur zawierających konkretne, wskazane przez eksperta motywy jest jednak trudnym zadaniem i nie radzą sobie z nim nawet najlepsze systemy oparte na AI, takie jak AlphaFold 3. W naszym projekcie podejmujemy się tego wyzwania i koncentrujemy się na modelowaniu oraz analizie funkcjonalnych cząsteczek RNA z uwzględnieniem motywów strukturalnych.

W ramach projektu opracujemy nowe biologicznie interpretowalne modele do wnioskowania o motywach funkcjonalnych w cząsteczkach RNA. Sformułujemy oparte na tych modelach problemy obliczeniowe związane z wyszukiwaniem motywów, ich identyfikacją, klasyfikacją, klastrowaniem, oceną podobieństwa, ewaluacją, oceną jakości oraz modelowaniem. Przebadamy złożoność obliczeniową tych problemów, aby dobrać odpowiednie techniki algorytmiczne i zasoby sprzętowe oraz oszacować prawdopodobieństwo uzyskania optymalnych rozwiązań w rozsądnym czasie. Zaprojektujemy wydajne algorytmy obliczeniowe oraz modele sztucznej inteligencji adresujące zdefiniowane problemy. Zaimplementujemy algorytmy w postaci przyjaznych dla użytkownika i środowiska narzędziach bioinformatycznych, a ich kody udostępnimy w otwartych repozytoriach typu GitHub. Dokonamy oceny stworzonych narzędzi pod względem ich dokładności, praktycznego zastosowania, trwałości oraz zrównoważenia środowiskowego. Przeprowadzimy pierwsze eksperymenty obliczeniowe z zastosowaniem opracowanych narzędzi do modelowania i analizy przykładowych cząsteczek RNA zawierających motywy reprezentujące różne klasy strukturalne i funkcjonalne. Uzyskane wyniki opublikujemy w artykułach naukowych o otwartym dostępie.

Realizując powyższe cele będziemy opierali się na dotychczasowych doświadczeniach zespołu w modelowaniu i analizie struktur kwasów nukleinowych. Nasza grupa badawcza opracowała już ponad 20 różnych obliczeniowych narzędzi analitycznych i baz danych dla bioinformatyki strukturalnej. Na arenie międzynarodowej jest również znana z eksperckiego modelowania struktur trójwymiarowych RNA o wysokiej precyzji i jakości - w konkursach RNA-Puzzles i CASP od lat uzyskuje jedne z najlepszych wyników. Te ostatnie zawdzięczamy przede wszystkim doskonałej współpracy członków naszego multidyscyplinarnego zespołu oraz umiejętności łączenia eksperymentu biologicznego z obliczeniowym. Cechy te zapewnią sprawną realizację niniejszego projektu. W jego wyniku praktycy będą mogli szybko, dokładnie i niezawodnie analizować, modelować i udoskonalać RNA z motywami funkcjonalnymi, co ułatwi postęp w naukach strukturalnych, nanotechnologii i biomedycynie zorientowanej na RNA (diagnostyka i terapia). Projekt jest także odpowiedzią na rosnące zapotrzebowanie społeczności na narzędzia oparte na wiedzy eksperckiej do oceny i przetwarzania modeli struktur molekularnych generowanych m.in. przez systemy głębokiego uczenia i wydobywania z nich nowej wiedzy.