

Załącznik 2: Autoreferat

1. Imię i Nazwisko

Agnieszka Rybarczyk

2. Posiadane Stopnie i Tytuły Naukowe

1. Stopień doktora nauk technicznych - Politechnika Poznańska, Wydział Informatyki; dyscyplina informatyka, 21 grudnia 2010 r., „Algorytmiczne aspekty procesu degradacji RNA”, promotorzy: prof. dr hab. inż. Jacek Błażewicz, prof. dr hab. Marek Figlerowicz:
 - Doktorat wyróżniony przez Radę Wydziału Informatyki.
 - Nagroda Polskiego Towarzystwa Bioinformatycznego za najlepszą rozprawę doktorską z bioinformatyki obronioną w 2010 r.
2. Tytuł zawodowy magistra - Politechnika Poznańska, Wydział Elektryczny; kierunek - informatyka; specjalizacja – inżynieria oprogramowania, 14 września 2001 r., „Wyszukiwanie motywów w sekwencjach nukleotydowych”, promotor: prof. dr hab. inż. Jacek Błażewicz.
3. Tytuł zawodowy inżyniera - Politechnika Poznańska, Wydział Elektryczny; kierunek - informatyka, 21.09.1999 r., „Inteligentny klient bazy danych”, promotor: dr hab. inż. Jerzy Nawrocki:
 - Praca inżynierska realizowana we współpracy z Haraldem Kjellin (University of Stockholm and Royal Institute of Technology, Department of Computer and Systems Sciences, Szwecja).
 - Praca inżynierska została rozwinięta w projekt o nazwie “Gtest” będący narzędziem umożliwiającym przeprowadzanie testów przez Internet, nad którym dalsze prace wykonano w ramach praktyk letnich 30.06.2000 – 30.07.2000 pod nadzorem Haralda Kjellin (University of Stockholm and Royal Institute of Technology, Department of Computer and Systems Sciences, Szwecja).
4. Tytuł zawodowy magistra – Uniwersytet im. Adama Mickiewicza w Poznaniu, Wydział Biologii; kierunek - biologia; specjalizacja – biologia molekularna, 30 czerwca 2004 r., „Dziedziczenie chloroplastów u mszaków”, promotor: prof. dr hab. Zofia Szwejkowska-Kulińska.
5. Tytuł zawodowy licencjata - Uniwersytet im. Adama Mickiewicza w Poznaniu, Wydział Biologii; kierunek - biologia; specjalizacja – biologia molekularna, 6.09.2002 r., „Charakterystyka odwrotnych transkryptaz”, promotor: prof. dr hab. Zofia Szwejkowska-Kulińska.

3. Dotychczasowe zatrudnienie w jednostkach naukowych

1. Instytut Informatyki, Wydział Informatyki, Politechnika Poznańska, adiunkt, 1 etat, 1.03.2011 – obecnie.
2. Instytut Politechniczny Państwowej Wyższej Szkoły Zawodowej im. Jana Amosa Komeńskiego, Leszno, starszy wykładowca, ½ etatu, 1.10.2017 - obecnie
3. Instytut Chemii Bioorganicznej Polskiej Akademii Nauk, Poznań, adiunkt, ½ etatu, 1.05.2014 – obecnie.

4. Instytut Politechniczny Państwowej Wyższej Szkoły Zawodowej im. Jana Amosa Komeńskiego, Leszno, starszy wykładowca, 1 etat, 1.10.2012 – 30.09.2017.
5. Instytut Politechniczny Państwowej Wyższej Szkoły Zawodowej im. Jana Amosa Komeńskiego, Leszno, starszy wykładowca, ½ etatu, 1.10.2011 – 30.09.2012.
6. Instytut Informatyki, Wydział Informatyki, Politechnika Poznańska, asystent, 1 etat, 1.10.2008 – 28.02.2011.
7. Instytut Informatyki, Wydział Informatyki, Politechnika Poznańska, informatyk w projekcie FP6 UE "METAFUNCTIONS" (Environmental – and Meta-Genomics – a bioinformatic system to detect and assign functions to habitat specific gene patterns), 1 etat, 1.10.2005 – 30.09.2008.
8. Instytut Informatyki, Wydział Informatyki, Politechnika Poznańska, asystent, 1 etat, 1.10.2002 – 30.09.2005.
9. Instytut Informatyki, Wydział Informatyki, Politechnika Poznańska, asystent, ½ etatu, 1.10.2001 – 30.09.2002.

4. Wskazanie osiągnięcia naukowego wynikającego z ustawy o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki

4.1. Tytuł osiągnięcia naukowego

Kombinatoryczne algorytmy analizy i projektowania degradomu RNA

4.2. Lista prac wchodzących w skład osiągnięcia naukowego

Agnieszka Rybarczyk (2018), Kombinatoryczne algorytmy analizy i projektowania degradomu RNA, Poznan Monographs in Computing and Its Applications 21, Nakom, ISBN: 978-83-63919-30-6.

4.3. Omówienie celu naukowego ww. prac oraz osiągniętych wyników wraz z omówieniem ich ewentualnego wykorzystania

Niniejsza monografia przedstawia kompleksowy model procesu degradacji RNA obrazując złożoność tego procesu, a także akcentuje nieocenione znaczenie oraz wkład biologii obliczeniowej i informatyki w zakresie wspomaganie oraz rozwiązywania problemów zakorzenionych w biologii czy też medycynie. Zaproponowane sformułowania problemów kombinatorycznych, ich analiza oraz dedykowane im algorytmy, stanowią istotne wsparcie dla biologów, umożliwiając im efektywną analizę danych oraz ich interpretację, a także symulowanie eksperymentów biochemicznych bez konieczności ich wykonywania. Badania cząstkowe z wykorzystaniem wspomnianych powyżej algorytmów posłużyły dodatkowo do zbudowania nowych, złożonych modeli obrazujących przebieg zarówno degradacji enzymatycznej jak i nieenzymatycznej (dane nie publikowane wcześniej) oraz umożliwiły finalnie opracowanie nowego podejścia do projektowania sztucznych regulatorowych cząsteczek RNA. Zastosowane i opisane w rozprawie modele złożonych systemów biologicznych pozwoliły na sprawdzenia wielu scenariuszy zachowania się systemu w zadanych warunkach oraz doprowadziły do uzyskania bardzo interesujących wniosków biologicznych.

4.3.1. Wprowadzenie

Jednym z największych osiągnięć biologii II połowy XX wieku było poznanie podstaw procesu ekspresji informacji biologicznej. Powstał wówczas schemat (centralny dogmat) opisujący funkcjonowanie organizmów żywych koncentrujący się przede wszystkim na DNA jako nośniku informacji genetycznej oraz białkach będących zarówno składnikami budulcowymi komórek, wykonawcami wszystkich procesów oraz regulatorami procesu ekspresji genów. Schemat ten zakładał, że RNA pełni jedynie rolę drugoplanową, jest pośrednikiem w procesie syntezy białka (mRNA, tRNA), a także szkieletem spajającym multienzymatyczne kompleksy (rRNA).

Ważnym elementem tak stworzonego modelu było założenie, że proces ekspresji informacji genetycznej przebiega tak samo w komórkach roślinnych oraz zwierzęcych, a co za tym idzie istnieje przesta zależność między złożonością danego organizmu a ilością genów.

Rozwój technik związanych z zastosowaniem sekwencjonowania pokazał, że komórkowe mechanizmy sterujące są bardziej skomplikowane niż to przewidywał centralny dogmat. Okazało się, że zarówno u zwierząt jak i roślin rejony kodujące białka stanowią jedynie niewielki procent materiału genetycznego (np. u człowieka 1-3% [1]), a także pozwolił na odkrycie szerokiej gamy niekodujących cząsteczek RNA (ncRNA, ang. *noncoding RNA*), w tym szczególnie małych regulatorowych RNA (srRNA, ang. *small regulatory RNA*). Odkrycie to zainicjowało intensywne, trwające do dnia dzisiejszego badania dotyczące nowych funkcji jakie mogą pełnić cząsteczki RNA, a także trwale zmieniło pogląd na transkrypcję, który obecnie postrzegany jest jako dynamiczny zbiór cząsteczek RNA o różnej wielkości, strukturze i funkcji powstałych w procesie przepisania genów DNA na RNA, w procesie transkrypcji [2].

Regulatorowe RNA są 19–28 nukleotydowymi cząsteczkami, które mogą powstawać w wyniku cięcia enzymatycznego lub degradacji komórkowych niekodujących RNA, a ich funkcja jest niezależna od funkcji rodzicielskich RNA [2]–[4]. Pełnią one kluczową rolę m.in. w regulacji transkrypcji i translacji RNA, procesie degradacji RNA, a także biorą udział w składaniu i redagowaniu RNA, czy też wpływają na strukturę genomu [5].

Proces degradacji RNA jest jednym z podstawowych procesów zachodzących w organizmach żywych, dzięki któremu zachowywana jest równowaga (homeostaza) komórkowa. Związany jest nie tylko z usuwaniem zbędnych cząsteczek RNA, ale także m.in. z dojrzewaniem i przetwarzaniem RNA, kontrolą jakości RNA oraz obroną komórki przed infekcjami wirusowymi. Wyróżniamy w ogólności dwa rodzaje degradacji RNA: niezależną od czynników komórkowych, determinowaną jedynie strukturą przestrzenną RNA (tzw. degradację/hydrolizę nieenzymatyczną) oraz przeprowadzaną przez enzymy (tzw. degradację enzymatyczną). Badania nad nieenzymatyczną degradacją RNA rozpoczęły się wraz z odkryciem katalitycznych właściwości RNA, powiązanych bezpośrednio z jego strukturą przestrzenną. Skłoniło to badaczy do przyjrzenia się roli, jaką struktura RNA pełni w procesach komórkowych. Analiza tego typu zagadnień na poziomie tak złożonego układu jakim jest komórka, uwzględniając wszystkie, bardzo złożone mechanizmy oraz ich komponenty, nie była i nadal nie jest możliwa na poziomie eksperymentalnym. Zbadano więc najpierw proces rozpadu RNA, pomijając wszystkie czynniki komórkowe, które mogłyby na niego oddziaływać, skupiając się jedynie na wpływie struktury przestrzennej RNA. Badania te jednoznacznie wskazały, że degradacja RNA jest zależna od struktury RNA, a także, że wzór degradacji rozważanych cząsteczek RNA podobny jest do tego, który powstaje w wyniku działania enzymów tnących RNA w komórce (rybonukleaz). Wynika to najprawdopodobniej z tego, że jeżeli pierwotnie RNA cięty był w miejscach szczególnie na to podatnych, białka, które wyewoluowały później preferowały i ułatwiały rozpad RNA w tych samych miejscach [6]–[9].

Uważa się, że mechanizmy i rodziny białek zaangażowane w degradację RNA są podobne zarówno u bakterii jak i organizmów wyższych. Dodatkowo substratem dla nich może być dowolny RNA, bez względu na pełnioną przez siebie funkcję [10], [11]. Sugeruje to, że bardzo ważnym elementem decydującym o stabilności cząsteczek RNA w komórce jest ich struktura przestrzenna.

Mimo, że dotychczas opisano wiele mechanizmów degradacji RNA, nadal przebieg tego zjawiska jest mało zbadany i wiele jego aspektów pozostaje nieznanymi. Dodatkowo ciągle odkrywane są nowe stabilne pośrednie produkty degradacji RNA (degradanty RNA) i ich rola w funkcjonowaniu organizmów, a także ich udział w przebiegu wielu chorób. Dane na temat nowych funkcjonalnych cząsteczek RNA gromadzone są w wielu rozproszonych bazach danych lub istnieją w postaci osobnych załączników do artykułów co bardzo utrudnia ich połączoną analizę [2], [12]–[14].

Dynamiczny rozwój technologii badawczych w dziedzinie genomiki i biologii molekularnej, a co za tym idzie regularne generowanie ogromnej liczby danych, a także złożoność powiązań pomiędzy nimi, których żaden człowiek nie byłby w stanie manualnie przeanalizować, spowodowała, że rozwinęły się zupełnie nowe gałęzie nauki na styku nauk podstawowych takich jak matematyka, biologia i informatyka. Przykładem takiej nowej, szybko rozwijającej się nauki jest biologia obliczeniowa (inaczej bioinformatyka), która zajmuje się tworzeniem skomplikowanych modeli matematycznych wspomagających przetwarzanie danych, modelowaniem zachodzących w organizmach żywych procesów oraz poszukiwaniem nowych metod rozwiązywania problemów biologicznych i biochemicznych. W wielu przypadkach rozważane problemy okazują się być trudne obliczeniowo.

W niniejszej rozprawie podany został całościowy model procesu degradacji RNA. Przedstawiony został nowy model kombinatoryczny dla problemu mapowania miejsc cięć w RNA (problemu RNA PDP) na podstawie wyników przeprowadzonego eksperymentu biochemicznego oraz algorytm metaheurystyczny pozwalający rozwiązać ten problem z uwzględnieniem zarówno błędów pozytywnych jak i negatywnych w danych wejściowych. Laboratoryjne przebadanie wzoru rozpadu dowolnej cząsteczki RNA wymuszonego jej strukturą przestrzenną jest niestety zadaniem bardzo kosztownym, zarówno pod względem czasu, jak i środków, dlatego też w pracy zaproponowano dodatkowo algorytm symulujący przebieg procesu nieenzymatycznej degradacji RNA. Uwzględnia on strukturę przestrzenną RNA, a także reguły stabilności zaproponowane przez Kierka i współpracowników [6]–[9] i pozwala przewidzieć wzór degradacji dowolnej cząsteczki RNA bez przeprowadzania eksperymentu biochemicznego.

Badania cząstkowe z wykorzystaniem wspomnianych powyżej algorytmów posłużyły do zbudowania złożonych modeli obrazujących przebieg zarówno degradacji enzymatycznej jak i nieenzymatycznej oraz umożliwiły finalnie opracowanie nowego podejścia do projektowania sztucznych regulatorów cząsteczek RNA.

Modelowanie złożonych systemów biologicznych jest procesem trudnym ze względu na bardzo dużą liczbę wciąż odkrywanych nowych faktów biologicznych i zależności między nimi. Dodatkowo koniecznym jest uwzględnienie tego, że wiele procesów w komórkach organizmów żywych przebiega równolegle, a ich elementy (substraty, produkty) mogą być wspólne. Dużą pomoc stanowią tutaj matematyczne formalizmy, które umożliwiają opisanie rozważanego systemu za pomocą pewnego modelu oraz jego analizę, co pozwala na odkrycie jego nieznanymi wcześniej własności. Dodatkowo, w ramach tak opracowanego modelu biologicznego mogą zmieniać jego parametry, a następnie sprawdzać jak wpłynie to na zachowanie badanego systemu biologicznego. W niniejszej pracy zastosowano w tym celu sieci Petriego, które są intuicyjne dla biologa (mogą być reprezentowane graficznie), a także są łatwe do analizy matematycznej.

Zgłębianie zagadnień związanych ze stabilnością RNA, a także jego rozpadem zarówno enzymatycznym jak i nieenzymatycznym oraz powstawaniem i rolą funkcjonalnych degradantów RNA stanowi obecnie bardzo rozwijany kierunek badań. Badania tego typu mogą mieć kluczowe znaczenie dla zrozumienia podłoża wielu chorób oraz wielu procesów zachodzących w komórkach organizmów żywych, a także mogą przyczynić się do rozwoju takich dziedzin jak biologia molekularna czy medycyna. Biologia obliczeniowa umożliwia natomiast modelowanie tych procesów, formułowanie i analizę problemów kombinatorycznych oraz konstrukcję algorytmów rozwiązujących te problemy.

4.3.2. Cel i zakres pracy

Degradacja RNA stanowi jeden z podstawowych i bardzo istotnych procesów zachodzących w organizmach żywych. Głównym jej celem jest usuwanie niepotrzebnych lub nieprawidłowych cząstek RNA, zapobiegając w ten sposób ich akumulacji i zapewniając prawidłowe funkcjonowanie komórki. Początkowo uważano, że jej produktami są jedynie pojedyncze rybonukleotydy, ale najnowsze doniesienia pokazują, że degradacja zarówno enzymatyczna jak i nieenzymatyczna (co do której zachodzą podejrzenia, że również ma miejsce w komórkach) funkcjonalnych RNA takich jak np. rRNA, tRNA czy snoRNA może prowadzić do powstawania stabilnych fragmentów RNA, które pełnią różnorodne funkcje zarówno w warunkach fizjologicznych jak i związanych z przebiegiem różnych chorób. Dodatkowo wszystkie cząsteczki RNA w jednakowym lub w podobnym stopniu wystawione są w komórce na działanie maszynierii odpowiedzialnej za ich degradację, jednakże jedne rozpadają się szybciej a inne wolniej, jedne pozostają stabilne, a inne ulegają natychmiastowemu rozpadowi. Sugeruje to, że ważnym elementem decydującym o ich stabilności jest ich struktura pierwszo-, drugo- i trzeciorzędowa. Bardzo duży poziom złożoności procesu degradacji RNA powoduje, że nie jest możliwe analizowanie wszystkich jego aspektów z wykorzystaniem istniejących metod eksperymentalnych. Pociąga to za sobą konieczność wsparcia ze strony metod i narzędzi modelowania oferowanych przez biologię obliczeniową.

Podstawowym celem niniejszej rozprawy było:

- Opracowanie całościowego modelu procesu degradacji RNA zarówno od strony informatycznej jak i matematycznej.
- Opracowanie narzędzi informatycznych umożliwiających przewidywanie degradacji dowolnej cząsteczki RNA, a co za tym idzie pozwalających wnioskować o jej stabilności, a także projektować sztuczne cząsteczki małych regulatorowych RNA.
- Stworzenie metod informatycznych pozwalających na rekonstrukcję informacji o rejonach decydujących o niestabilności RNA.

W celu realizacji przedstawionych powyżej zadań, konieczne było wykonanie następujących badań szczegółowych:

- Opracowanie nowego sformułowania kombinatorycznego dla oryginalnego problemu RNA PDP (rekonstrukcji położenia miejsc cięcia dla zadanej cząsteczki RNA na podstawie eksperymentu biochemicznego) oraz dla jego nowej wersji uwzględniającej oprócz błędów negatywnych, też błędy pozytywne w instancji. Określenie stopnia złożoności obliczeniowej dla nowej wersji problemu oraz skonstruowanie algorytmu metaheurystycznego składającego się z dwóch współpracujących ze sobą metaheurystyk przeszukiwania tabu (ang. *tabu search*, *TS*).
- Opracowanie nowej metody przewidywania udokładnionej (rozszerzonej) struktury drugorzędowej RNA (uwzględniającej zarówno parowania kanoniczne jak i niekanoniczne).
- Opracowanie algorytmu dokładnego podziału i odcięć (ang. *branch and cut*) symulującego przebieg procesu nieenzymatycznej degradacji RNA z uwzględnieniem struktury pierwszo- oraz drugorzędowej kanonicznej i udokładnionej RNA, a także reguł stabilności zaproponowanych przez Kierzka oraz jego współpracowników [6]–[9]. Uzyskane wyniki zostały skonfrontowane z wynikami przeprowadzonego eksperymentu biochemicznego.
- Opracowanie nowego podejścia umożliwiającego oszacowanie wartości opóźnień dla tranzycji w sieciach czasowych DPN (ang. *Timed Petri Net*) oraz częstości uruchomień dla tranzycji reprezentujących procesy biologiczne w modelach opartych o stochastyczne sieci Petriego (SPN, ang. *Stochastic Petri Net*).
- Opracowanie modelu nieenzymatycznej degradacji RNA w oparciu o stochastyczne sieci Petriego, obrazującego przebieg procesu nieenzymatycznej degradacji RNA zgodnie z opisem eksperymentów biochemicznych przedstawionych przez Kierzka i współpracowników [6]–[9],

a także zgodnie z tym jak swoje obliczenia wykonuje opisany w niniejszej rozprawie algorytm dokładny dotyczący przewidywania stabilności cząsteczek RNA w oparciu o ich strukturę przestrzenną RNA oraz reguły stabilności wiązań internukleotydowych w łańcuchu RNA [6]–[9]. Wykonanie analizy strukturalnej oraz symulacyjnej modelu w celu przedstawienia wniosków.

- Opracowanie modelu enzymatycznej degradacji RNA w oparciu o stochastyczne sieci Petriego opisującego rolę degradantów powstających z YRNA w regulacji apoptozy i zapalenia w monocytach/makrofagach, a co za tym idzie w procesie pogłębiania się zmian miażdżycowych. Wykonanie analizy strukturalnej oraz symulacyjnej modelu prowadzącej do uzyskania interesujących konkluzji biologicznych.
- Opracowanie metody projektowania sztucznych małych regulatorowych RNA na przykładzie miRNA. Wynikowe cząsteczki z założenia będą mogły wpływać na proces ekspresji informacji genetycznej w komórkach organizmów żywych dla których zostały zaprojektowane, w tym m.in. na przebieg procesu degradacji dojrzałych RNA, co może prowadzić do powstawania i akumulacji stabilnych fragmentów RNA (degradantów).

Rozprawa została podzielona na następujące rozdziały. W rozdziale 1 zostało zawarte wprowadzenie do tematyki pracy wraz z uzasadnieniem podjętych badań oraz opisem ich znaczenia. Natomiast rozdział 2 dotyczy podstawowych zagadnień z zakresu informatyki teoretycznej.

Rozdział 3 zawiera opis podstawowych zagadnień biologii molekularnej, ze szczególnym uwzględnieniem procesu degradacji RNA. W rozdziale 4 opisano problem mapowania miejsc cięć w cząsteczce RNA (problem RNA PDP). Przedstawiono nowe sformułowania tego problemu zarówno dla wersji oryginalnej jak i dla wersji uwzględniającej równocześnie oba typy błędów (negatywnych i pozytywnych). Ze względu na to, że jest to problem silnie NP--zupełny zarówno dla przypadku bez błędów jak i z błędami negatywnymi lub/i pozytywnymi zaproponowano algorytm metaheurystyczny składający się z dwóch współpracujących ze sobą heurystyk przeszukiwania tabu (ang. *tabu search*). W kolejnym rozdziale przedstawiono algorytm dokładny podziału i odcięć (ang. *branch and cut*) pozwalający przewidywać rozpad dowolnej cząsteczki RNA z uwzględnieniem jej struktury przestrzennej oraz danych literaturowych dotyczących stabilności wiązań internukleotydowych łączących nukleotydy w łańcuchu RNA. Opisano również nowe podejście do przewidywania struktury udokładnionej RNA. W rozdziale 6 zaproponowano oraz przeanalizowano model nieenzymatycznej degradacji RNA oparty na stochastycznych sieciach Petriego, zgodny z opisem eksperymentów biochemicznych oraz z tym jak wykonuje swoje obliczenia algorytm opisany w rozdziale 5. W rozdziale 7 zaprezentowano oraz przeanalizowano model degradacji enzymatycznej na przykładzie degradantów powstających z YRNA i ich roli w procesie pogłębiania się zmian miażdżycowych. Rozdział 8 zawiera opis metody projektowania sztucznych cząsteczek małych regulatorowych RNA na przykładzie miRNA (mikroRNA). W rozdziale 9 zawarto podsumowanie rozprawy oraz wnioski i potencjalne kierunki badawcze.

4.3.3. Podstawowe zagadnienia z zakresu kombinatoryki

W rozdziale 2 monografii zostały wprowadzone podstawowe pojęcia z zakresu informatyki teoretycznej, które są niezbędne do zrozumienia omawianej pracy. Obejmują one zagadnienia związane z grafami, z pominięciem grafów nieskierowanych, gdyż rozważane w rozprawie grafy są wyłącznie grafami skierowanymi (tj. sieci Petriego). Dodatkowo, przedstawiona została szczegółowo teoria z zakresu złożoności obliczeniowej oraz omówione zostały następujące algorytmy: podziału i odcięć (ang. *branch-and-cut*), zachłanny (ang. *greedy algorithm*) oraz heurystyka przeszukiwania tabu (ang. *tabu search*). Ostatnia część rozdziału została poświęcona sieciom Petriego. Zaprezentowano podstawowe definicje, formy reprezentacji, metody analizy oraz podstawowe rodzaje sieci Petriego.

4.3.4. Podstawowe zagadnienia z zakresu biologii molekularnej

W rozdziale 3 monografii zawarto wprowadzenie do zagadnień związanych z budową kwasów nukleinowych, skupiając się na RNA (kwasie rybonukleinowym), który jest przedmiotem dalszych rozważań. Opisano jego strukturę pierwszo-, drugo- oraz trzeciorzędową, a także podstawowe jego rodzaje obecne w komórkach organizmów żywych. Przedstawiono również szczegółowo proces degradacji RNA, zarówno enzymatycznej jak i nieenzymatycznej, podkreślając jego wagę oraz złożoność.

4.3.5. Problem mapowania miejsc cięć RNA

Proces nieenzymatycznej degradacji cząsteczki RNA zależny od jej struktury można podsumować w następujący sposób. Wejściowa cząsteczka RNA (występująca w wielu kopiach) o zadanej długości, najpierw pęka w pierwotnych miejscach cięcia, w wyniku czego powstają fragmenty (długości) pierwotne. Następnie fragmenty te ulegają kolejnym pęknięciom w miejscach cięć wtórnych, co daje zbiór fragmentów wtórnych. Wynikiem procesu degradacji RNA jest więc zbiór fragmentów pierwotnych i wtórnych pochodzących z dwóch oddzielnych eksperymentów biochemicznych (które przeprowadzono w celu zbadania wpływu struktury RNA na proces jego nieenzymatycznej hydrolizy): degradacji RNA z wieloma znakowanymi nukleotydami oraz degradacji RNA z jednym znakowanym nukleotydem na końcu 5' cząsteczki RNA (jej "lewym" końcu). Każdy fragment pierwotny jest cięty tylko raz. Dla każdego fragmentu mamy jedynie informację o jego długości i o tym czy pochodzi z 5' końca. Nie wiadomo natomiast czy jest to fragment pierwotny czy wtórny. Celem jest zrekonstruowanie położenia pierwotnych i wtórnych miejsc cięcia w cząsteczce wejściowej, bazując na ograniczonych informacjach pochodzących z eksperymentów biochemicznych.

Pierwsze sformułowanie kombinatoryczne problemu mapowania miejsc cięć RNA (RNA PDP) zostało przedstawione w [15], [16]. Tam też zaproponowano dla niego algorytm dokładny podziału i odcięć oraz pokazano, że problem ten jest silnie NP-zupełny w wersji decyzyjnej bez błędów negatywnych.

Zanim zostanie przedstawione nowe sformułowanie tego problemu, konieczne jest wprowadzenie kilku podstawowych pojęć oraz definicji. Niech L oznacza długość wejściowej cząsteczki RNA, której wzór degradacji jest rekonstruowany. Natomiast D jest multizbiorem (gdzie poszczególne elementy mogą występować w wielu kopiach) zawierającym fragmenty, które zostały uzyskane z degradacji wejściowej cząsteczki RNA z wieloma znakowanymi nukleotydami (czyli zawiera wszystkie odległości pomiędzy miejscami cięć pierwotnych oraz końcami cząsteczki RNA razem z parami fragmentów uzyskanych w wyniku cięć wtórnych), a jego podzbiór $Z \subseteq D$ zawiera długości, które posiadają "lewy" koniec wejściowej cząsteczki RNA (pochodzące z degradacji RNA z jednym znakowanym nukleotydem, czyli wszystkie długości pierwotne oraz te spośród długości wtórnych, które posiadają znakowany koniec 5' cząsteczki RNA). W obu zbiorach dozwolone są błędy negatywne (brakujące fragmenty). Na razie nie są rozpatrywane błędy pozytywne (nadmiarowe fragmenty). Celem jest znalezienie dwóch rozłącznych zbiorów liczb całkowitych P_1 i P_2 , gdzie P_1 oznacza zbiór pierwotnych miejsc cięcia, a P_2 wtórnych miejsc cięcia.

Dla zbioru liczb całkowitych $P = \{1, \dots, L-1\}$, niech $R(P)$ oznacza zbiór par $(x, y) \neq (0, L)$, takich, że $x, y \in P \cup \{0, L\}$ i $x < y$. Jeżeli P jest zbiorem pierwotnych miejsc cięcia to $R(P)$ jest zbiorem fragmentów pierwotnych opisanych jako (x, y) , gdzie x odnosi się do położenia jego "lewego" końca, a y "prawego".

Definicja 1.

Niech L będzie liczbą całkowitą dodatnią, C liczbą całkowitą nieujemną oraz niech P_1 i P_2 będą dwoma zbiorami liczb całkowitych takimi, że $0 < x < L$ dla każdego $x \in P_1 \cup P_2$. Para (P_1, P_2) jest C -zgodna z D i Z jeżeli spełnione są poniższe ograniczenia [17]:

Dana jest funkcja $f: R' \rightarrow P_2$ pomiędzy podzbiorem

$$(1.1) \quad R' \subseteq R(P_1) \text{ i } P_2 \text{ taka, że } x < f(x, y) < y, \forall (x, y) \in R'$$

$$(1.2) \quad D \subseteq D' = \bigcup_{(x,y) \in R(P_1)} \{y - x\} \cup \bigcup_{(x,y) \in R'} \{y - f(x, y), f(x, y) - x\}$$

$$(1.3) \quad Z \subseteq Z' = P_1 \cup \bigcup_{(0,y) \in R'} f(0, y)$$

$$(1.4) \quad |D'| - |D| + |Z'| - |Z| \leq C$$

Zbiór R' (wyrażenie 1.1) zawiera te fragmenty pierwotne, które cięte są na wtórne. Dla każdego $(x, y) \in R'$ ($x, y, f(x, y) \in P_2$) jest położeniem wtórnego miejsca cięcia w ramach fragmentu, którego położenie określa (x, y) . Dodatkowo, funkcja f wymusza, aby każdy fragment pierwotny był cięty co najwyżej jeden raz. Multizbiór D' (wyrażenie 1.2) zawiera długości wszystkich fragmentów pierwotnych obecnych w $R(P_1)$ oraz wszystkich fragmentów wtórnych, których lokalizacja w ramach cząsteczki wejściowej opisana jest za pomocą $(x, f(x, y))$ i $(f(x, y), y)$ będących wynikiem wtórnego cięcia w ramach fragmentu $(x, y) \in R'$ na pozycji $f(x, y)$. W powyższej definicji zakładamy brak błędów pozytywnych, w związku z czym narzucone jest, że D' zawiera multizbiór D (długości fragmentów pochodzące z eksperymentu degradacji RNA). Natomiast zbiór Z' zawiera wszystkie fragmenty wtórne oraz pierwotne, które posiadają "lewy" koniec wejściowej cząsteczki RNA (czyli mają znakowany koniec 5', rozpoczynają się od pozycji 0 w wejściowej cząsteczce RNA). Wyrażenie 1.4 nakłada górne ograniczenie na liczbę brakujących fragmentów w zbiorach D i Z [17].

Ze względu na to, że dozwolone są jedynie błędy negatywne (braki) w zbiorach D i Z , celem jest minimalizacja całkowitej liczby błędów negatywnych w D i Z .

Na podstawie powyższej definicji, można otrzymać przypadek idealny problemu RNA PDP, po ograniczeniu instancji problemu do nie zawierających błędów negatywnych ($C = 0$).

Warto zwrócić uwagę, że jeżeli dany element jest obecny w zbiorze Z' , ale nie występuje w zbiorach D i Z , liczba uwzględnionych dotychczas błędów negatywnych jest liczona podwójnie (ponieważ element ten będzie znajdował się w obu zbiorach: D' i Z').

Zgodnie z powyższą definicją, jeżeli jakiś fragment pierwotny nie został wtórnie przecięty to nie jest to traktowane jako błąd. Jeżeli założymy jednak, że każdy fragment pierwotny zawsze musi ulec rozpadowi na dwa fragmenty wtórne (co ma często miejsce w rzeczywistości biologicznej), to wtedy modyfikacji musi ulec wyrażenie (1.1) w Definicji 1, którego nowa postać będzie następująca: $R' = R(P_1)$ [17].

Natomiast, uwzględnienie błędów pozytywnych (nadmiarowych fragmentów) w zbiorach wejściowych wymagałoby następujących modyfikacji w ramach Definicji 1 [17]:

- Z ograniczenia (1.2) należałoby usunąć wyrażenie $D \subseteq D'$.
- Z ograniczenia (1.3) należałoby usunąć wyrażenie $Z \subseteq Z'$.
- Ograniczenie (1.4) przyjęłoby postać:

$$(1.4') \quad |D'| + |D| - 2|D' \cap D| + |Z'| + |Z| - 2|Z' \cap Z| \leq C$$

Problem RNA PDP można teraz sformułować następująco:

Problem 1. Problem RNA PDP w wersji przeszukiwania [17]

Parametry: Liczba całkowita dodatnia L , nieujemna liczba całkowita C , multizbiór D , zbiór Z liczb całkowitych dodatnich takich, że $0 < x < L$ dla każdego $x \in D$ i $Z \subseteq D$.

Cel: Znajdź dwa zbiory P_1 i P_2 takie, że (P_1, P_2) jest C -zgodny z D i Z .

Problem 2. Problem RNA PDP w wersji decyzyjnej [17]

Parametry: Liczba całkowita dodatnia L , nieujemna liczba całkowita C , multyzbiór D , zbiór Z liczb całkowitych dodatnich takich, że $0 < x < L$ dla każdego $x \in D$ i $Z \subseteq D$.

Cel: Czy istnieje para zbiorów (P_1, P_2) , która jest C -zgodna z D i Z ?

Silna NP-zupełność problemu RNA PDP dla $C=0$ (przypadek idealny) została udowodniona poprzez transformację pseudowielomianową wychodzącą z problemu podzbioru o zadanej sumie (ang. *Numerical Matching With Target Sums*) [18] i zaprezentowana w pracy [15].

Problem RNA PDP (mapowania miejsc cięć w RNA) w wersji decyzyjnej z uwzględnieniem błędów negatywnych (gdzie $C>0$) jest oczywiście co najmniej tak trudny jak idealny problem RNA PDP ($C=0$) [15], [17]. Złożoność zmodyfikowanego problemu RNA PDP, gdzie $R'=R(P_1)$ nie została formalnie udowodniona, ale prawdopodobnie pozostaje silnie NP-zupełna nawet dla $C=0$ [17].

Aby rozwiązać problem RNA PDP (mapowania miejsc cięć RNA) z błędami negatywnymi i pozytywnymi (w wersji przeszukiwania) zaproponowano algorytm metaheurystyczny, co wynika oczywiście z silnej NP-trudności rozważanego problemu [15]. Zdecydowano się na algorytm przeszukiwania tabu, ponieważ heurystyka tego typu, stosowana jest z powodzeniem w przypadku rozwiązywania wielu trudnych problemów kombinatorycznych [18].

Algorytm dokładny rozwiązujący problem RNA PDP z błędami negatywnymi istnieje i został opublikowany w [15]. Nie radzi on sobie jednak z większymi instancjami, stąd też propozycja algorytmu heurystycznego. W algorytmie przyjęto, że każdy fragment pierwotny uległ wtórnemu cięciu, przy czym modyfikacja tego założenia z punktu widzenia konstrukcji algorytmu jest bardzo łatwa (tj. przyjęcie, że nie każdy fragment pierwotny został przecięty).

Struktura algorytmu jest następująca. Niech S będzie rozwiązaniem problemu RNA PDP, gdzie P_1 jest zbiorem pierwotnych a P_2 wtórnych miejsc cięcia spełniających ograniczenia dla problemu. Aby ocenić jakość rozwiązania S , rozważane są dwie funkcje kryterialne [17]: $F(S)$ będąca liczbą fragmentów, które występują w D ale są nieobecne w D_S oraz liczbą fragmentów, które występują w Z ale są nieobecne w Z_S (nadmiarowe fragmenty), a także $G(S)$ będąca liczbą fragmentów, które występują w D_S ale są nieobecne w D oraz liczbą fragmentów, które występują w Z_S ale są nieobecne w Z (brakujące fragmenty). Wartości tych funkcji są minimalizowane, a ich suma jest globalną funkcją kryterialną algorytmu.

Odnosząc się do wyrażenia (1.4') można zauważyć, że dla rozwiązania S , para (P_1^S, P_2^S) jest $(F(S)+G(S))$ -zgodna z D i Z .

Liczba pierwotnych miejsc cięcia nie jest z góry znana, a więc zadaniem algorytmu jest również ustalenie tej wartości. Zaproponowany algorytm przeszukiwania tabu dla problemu RNA PDP uruchamiany jest kilka razy z różną liczbą cięć pierwotnych (obliczanych na bazie zbiorów wejściowych D i Z) w celu znalezienia najlepszego rozwiązania.

Algorytm bazuje na dwóch współpracujących ze sobą metaheurystykach przeszukiwania tabu nazywanych TS_{primary} oraz $TS_{\text{secondary}}$. Obie metaheurystyki uruchamiane są ponownie dla każdej rozważanej wartości pierwotnych cięć, dostarczając za każdym razem nowe rozwiązanie dla problemu RNA PDP (różniące się od poprzednich liczbą pierwotnych miejsc cięcia). Pierwszy z rozpatrywanych algorytmów TS_{primary} odpowiedzialny jest za rekonstrukcję pierwotnych miejsc cięcia i nie bierze pod uwagę wtórnych miejsc cięcia. Drugi natomiast, $TS_{\text{secondary}}$ rekonstruuje wtórne miejsca cięcia i zakłada, że pierwotne miejsca cięcia podczas jego pracy nie mogą ulec zmianie [17]. Ogólny schemat algorytmu został przedstawiony poniżej.

Algorytm 1 Ogólna struktura algorytmu

Wejście: D, Z, L , zakres wartości v Wyjście: S_{best}

- 1: $F_{best} \leftarrow \infty, G_{best} \leftarrow \infty$
 - 2: **for** każdej rozważanej liczby pierwotnych miejsc cięcia v **do**
 - 3: Wygeneruj początkowy zbiór zawierający pierwotne miejsca cięcia \mathcal{P}_1 ($IS_{primary}$)
 - 4: Spróbuj znaleźć lepszy zbiór pierwotnych miejsc cięcia \mathcal{P}_1^* ($TS_{primary}$)
 - 5: Wygeneruj początkowy zbiór zawierający wtórne miejsca cięcia \mathcal{P}_2 , pierwotne miejsca cięcia \mathcal{P}_1^* uznawane są za ustalone i niezmiennie ($IS_{secondary}$)
 - 6: Spróbuj znaleźć lepszy zbiór wtórnych miejsc cięcia \mathcal{P}_2^* , nie modyfikując pierwotnych miejsc cięcia ($TS_{secondary}$), oraz niech $S_v^* = (\mathcal{P}_1^*, \mathcal{P}_2^*)$ będzie końcowym rozwiązaniem
 - 7: **if** $F(S_v^*) + G(S_v^*) < F_{best} + G_{best}$ **then**
 - 8: $F_{best} \leftarrow F(S_v^*), G_{best} \leftarrow G(S_v^*), S_{best} \leftarrow S_v^*$
 - 9: **end if**
 - 10: **end for**
-

Lista zabronionych ruchów dla algorytmu $TS_{primary}$ znajduje się w macierzy M , gdzie liczba wierszy równa jest liczbie cięć pierwotnych, a liczba kolumn wynosi $L-1$ (możliwe miejsca cięcia). Pojedynczy element macierzy $M_{i,j}$ przechowuje informację o ilości iteracji, podczas których zabronione jest ustalenie i -tego miejsca cięcia pierwotnego na pozycji j . Początkowo wszystkie wartości $M_{i,j}$ ustawione są na 0. $TS_{primary}$ kończy swoje działanie po $|D|$ iteracjach. Natomiast lista tabu dla algorytmu $TS_{secondary}$ ma postać macierzy M' o $|D|$ wierszach i $|D|$ kolumnach, które odpowiadają długościom wtórnym. Pojedynczy element macierzy M'_{ij} przechowuje informację o ilości iteracji, podczas których zabronione jest połączenie dwóch fragmentów wtórnych o takich długościach, że ich suma odpowiadałaby jednemu z fragmentów pierwotnych. Początkowo wszystkie komórki M'_{ij} mają przypisaną wartość 0. $TS_{secondary}$ kończy swoje działanie po $2|D|$ iteracjach.

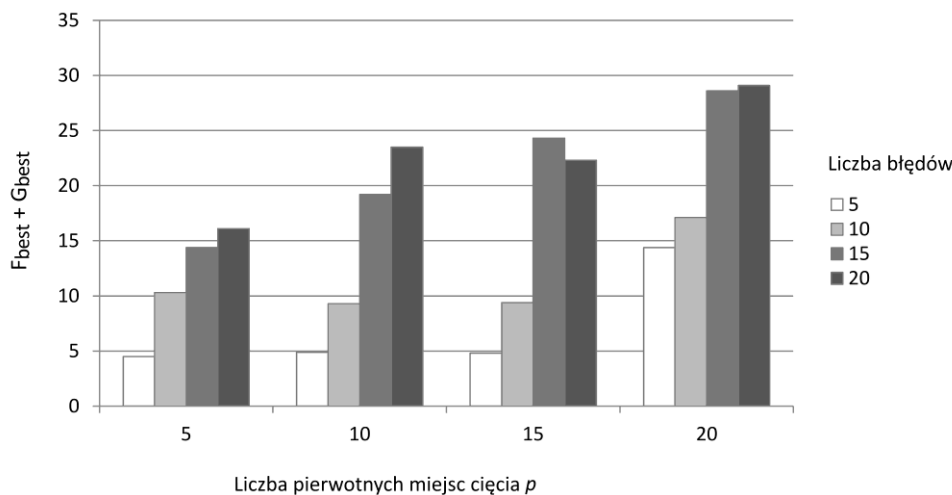
Przedstawiony powyżej algorytm został przetestowany w obszernym eksperymencie obliczeniowym. Użyto w tym celu danych wygenerowanych losowo. Metaheurystykę przeszukiwania tabu zaimplementowano w języku C++, a wszystkie obliczenia przeprowadzono na komputerze z procesorem Intel Xeon E5-2670, zegarem 2.60 GHz, pamięcią operacyjną 16 GB oraz z systemem operacyjnym Linux.

Dane do eksperymentu zostały przygotowane w następujący sposób. Wygenerowano cząsteczki RNA o długości $L=5000$, które podobnie jak przypadku rzeczywistych cząsteczek [15], [19], posiadały liczbę pierwotnych miejsc cięcia $p=5, 10, 15$ lub 20 . W pierwszej kolejności losowane były pierwotne miejsca cięcia używając generatora liczb losowych z jednorodnym rozkładem prawdopodobieństwa, z przedziału $[1, 4999]$. Następnie, dla każdej instancji oraz dla każdego fragmentu pierwotnego (x,y) , wylosowano wtórne miejsca cięcia przy wykorzystaniu generatora liczb losowych z jednorodnym rozkładem prawdopodobieństwa, z przedziału $[x+1,y-1]$.

Dane losowe podzielono na 4 zbiory. Pierwszy zbiór zawierał instancje bez błędów w zbiorach wejściowych D i Z . Drugi zawierał instancje z $5,10,15$ lub 20 błędami negatywnymi, które otrzymano poprzez losowe usuwanie elementów z $D \cup Z$. W zbiorze trzecim znajdowały się instancje z $5, 10, 15$ lub 20 błędami pozytywnymi, które dodano do $D \cup Z$ używając generatora liczb losowych z jednorodnym rozkładem prawdopodobieństwa z przedziału $[1, 4999]$. Ostatni zbiór zawierał instancje z $e = 5, 10, 15$ lub 20 błędami negatywnymi oraz z taką samą liczbą błędów pozytywnych, co daje sumarycznie

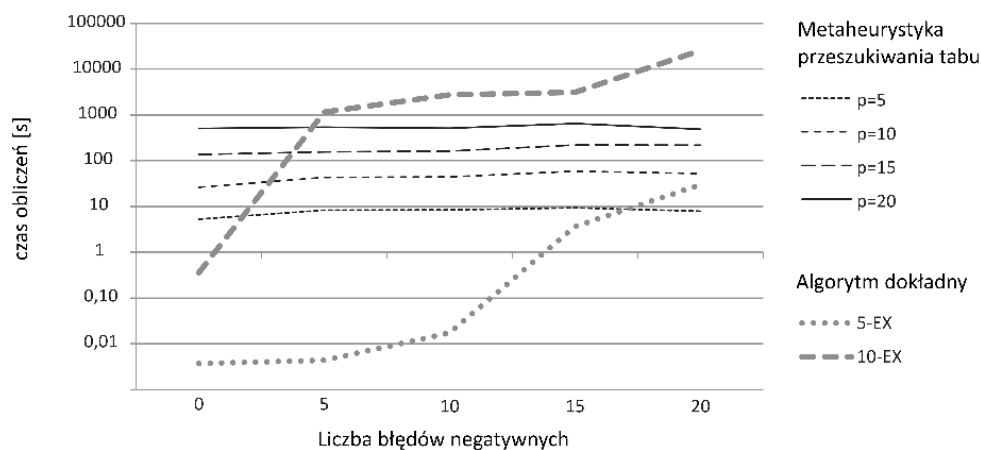
liczbę błędów równą $2e$. Dla każdego zbioru parametrów (liczba pierwotnych miejsc cięcia, liczba błędów negatywnych, liczba błędów pozytywnych) wygenerowano losowo i rozwiązano po 10 instancji. Zaprezentowane wyniki są wartościami średnimi ze wszystkich uruchomień algorytmu.

Przykładowe wyniki dla instancji z błędami negatywnymi (bez błędów pozytywnych) umieszczone zostały na Rysunku 1. Pokazuje on, że wartości globalnej funkcji kryterialnej $F_{best}+G_{best}$ są niemal równe liczbie błędów dla instancji do 15 pierwotnych miejsc cięcia i do 10 błędów negatywnych. Instancje z większą liczbą pierwotnych miejsc cięcia lub z większą liczbą błędów negatywnych wydają się być większym wyzwaniem dla algorytmu [17].



RYСУNEK 1 WARTOŚCI GLOBALNEJ FUNKCJI KRYTERIALNEJ $F_{BEST}+G_{BEST}$ DLA INSTANCJI Z BŁĘDAMI NEGATYWNYMI.

Dodatkowo wykonano również porównanie w zakresie czasu obliczeń dla algorytmu dokładnego oraz zaproponowanej metaheurystyki poszukiwania tabu dla problemu RNA PDP. Algorytm dokładny opisany w [15] ma złożoność obliczeniową wykładniczą i uwzględnia jedynie błędy negatywne oraz nie bierze pod uwagę błędów pozytywnych. Czasy obliczeń dla algorytmu dokładnego (w skali logarytmicznej), dla liczby pierwotnych miejsc cięcia równej $p = 5$ (krzywa 5-Ex) i $p = 10$ (krzywa 10-Ex) pokazane są na Rysunku 2. Dla porównania pokazano również wyniki dla metaheurystyki przeszukiwania tabu [17].



RYСУNEK 2 ŚREDNIE CZASY OBLICZEŃ DLA OBU ALGORYTMÓW DLA INSTANCJI Z BŁĘDAMI NEGATYWNYMI.

Podsumowując, przedstawione powyżej nowe sformułowanie kombinatoryczne dla oryginalnego problemu RNA PDP oraz dla wersji tego problemu uwzględniającego błędy pozytywne oraz negatywne w instancji wraz z zaproponowanym dla tego problemu algorytmem metaheurystycznym zostały szczegółowo opisane w Rozdziale 4 omawianej monografii. Algorytm składa się z dwóch współpracujących ze sobą heurystyk przeszukiwania tabu i uwzględnia zarówno błędy negatywne jak i pozytywne. Rozwiązania przez niego generowane są dobrej jakości, a liczba zrekonstruowanych cięć jest zbliżona do tych, które obecne są w rozwiązaniach optymalnych. Parametry (liczba błędów oraz miejsc cięcia) wykorzystane do wygenerowania instancji testowych, zostały dobrane w taki sposób, aby odzwierciedlały rzeczywiste przypadki. W związku z tym, zaprezentowany algorytm będzie dobrze sprawdzał się w zastosowaniach praktycznych, przy wspomaganiu analizy danych pochodzących z eksperymentów biochemicznych.

Algorytm dokładny, który został wcześniej zaproponowany dla problemu RNA PDP, uwzględnia jedynie błędy negatywne i radzi sobie tylko z instancjami zawierającymi do 10 miejsc cięć pierwotnych. Przedstawiona metaheurystyka przeszukiwania tabu jest jedynym znanym autorem algorytmem, który umożliwi rozwiązanie tego problemu dla dużej liczby miejsc cięcia oraz przy obecności błędów pozytywnych w danych wejściowych.

4.3.6. Algorytm symulujący proces nieenzymatycznej degradacji RNA w oparciu o przewidywanie udokładnionej struktury drugorzędowej

W rozdziale 5 monografii przedstawiono nowe podejście do przewidywania udokładnionej struktury RNA (określane jako potok przetwarzania o nazwie RC/Rp (obecnie zaimplementowane jako RNAvista)) oraz nowy algorytm podziału i odcięć (ang. *branch and cut*) dla problemu przewidywania stabilności cząsteczek RNA w oparciu o ich sekwencję, strukturę przestrzenną, a także reguły stabilności opracowane przez Kierzka i współpracowników [6]–[9]. Wyniki działania algorytmu zostały porównane z wynikami uzyskanymi dla przykładowych cząsteczek RNA w eksperymencie biochemicznym [19]. Algorytm wykorzystuje zewnętrzne programy do przewidywania struktury drugorzędowej kanonicznej (RNAfold, RNAstructure, CentroidFold, ContextFold, IPknot, CONTRAFold) oraz udokładnionej (RC/Rp (RNAvista)).

Nowe podejście do przewidywania udokładnionej struktury drugorzędowej RNA

Najnowsze badania pokazują, że około 40% wszystkich par zasad w ustrukturyzowanych cząsteczkach RNA to pary niekanoniczne (nie GU i nie typu Watson-Crick) [20]. Udowodniono również, że struktura drugorzędowa uwzględniająca jedynie pary kanoniczne jest niewystarczająca do prawidłowej predykcji struktury trzeciorzędowej RNA oraz do wykonania dopasowania homologicznych sekwencji RNA [21]. Dlatego też bardzo istotne jest uzyskanie udokładnionej (rozszerzonej) struktury drugorzędowej RNA, uwzględniającej zarówno parowania kanoniczne jak i niekanoniczne.

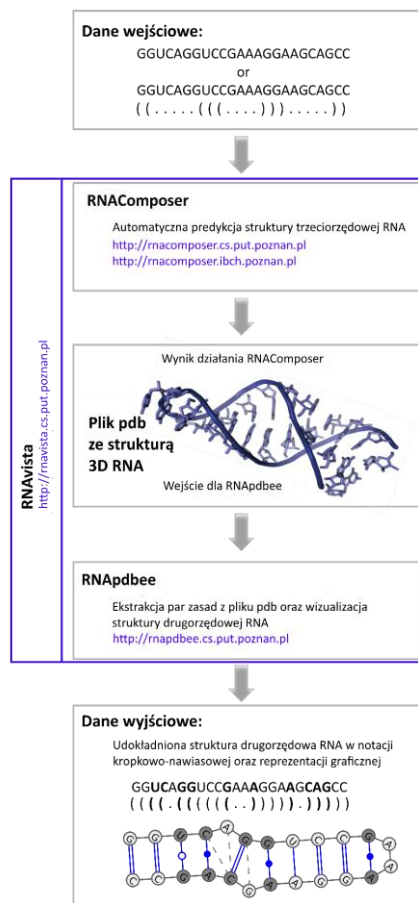
Spośród ponad 50 dostępnych programów dedykowanych do predykcji drugorzędowej struktury RNA, tylko 7 jest w stanie poradzić sobie z przewidywaniem zarówno kanonicznych jak i niekanonicznych par zasad, przy czym każdy z nich posiada pewne ograniczenia.

Dostęp do informacji związanych z niekanonicznymi oddziaływaniami oraz parami wielokrotnymi w cząsteczce RNA jest zdecydowanie łatwiejszy w sytuacji, gdy struktura drugorzędowa uzyskiwana jest poprzez ekstrakcję par zasad ze struktury trzeciorzędowej RNA. Programami, które to umożliwiają oraz dodatkowo klasyfikują uzyskane w ten sposób pary zgodnie z notacją LW (Leontis Westhof) [20] są: RNAView [22], MC-Annotate [23], 3DNA/DSSR [24] oraz RNApdbee [25]. RNApdbee dodatkowo oferuje też klasyfikację zgodną z notacją Saengera [26].

Podejście opisane w niniejszym rozdziale bazuje na przewidywaniu struktury trzeciorzędowej (wychodząc z sekwencji lub struktury kanonicznej 2D (drugorzędowej) RNA czyli zawierającej pary typu Watson-Crick i GU), a następnie ekstrakcji par zasad ze współrzędnych położenia atomów struktury

RNA z pliku PDB (opisującego strukturę 3D (trzeciorzędową) RNA), w celu uzyskania rozszerzonej struktury drugorzędowej. Jest ono bardzo ogólne, co oznacza, że do predykcji struktury 3D można wykorzystać dowolne, istniejące oprogramowanie do tego dedykowane (np. FARFAR [27], DMD [28] czy też RNAComposer [29]), podobnie do uzyskania wtórnie struktury 2D z 3D (np. RNAView, MC-Annotate, 3DNA/DSSR, RNApdbee). Z punktu widzenia użytkownika, bardzo istotnym elementem decydującym o wyborze programu do predykcji struktury 3D RNA jest szybkość oraz jak największa automatyzacja procesu obliczeń. W zakresie efektywnego przewidywania struktur 3D dużych cząsteczek RNA (do 500 nt), jedyną metodą, która osiągnęła stadium pełnej automatyzacji przy jednoczesnej optymalizacji czasu przetwarzania jest RNAComposer. Z tego też powodu, przedstawione podejście zostało zaimplementowane poprzez połączenie funkcjonalności dwóch narzędzi: RNAComposer oraz RNApdbee (określane dalej jako potok przetwarzania RC/Rp), które obecnie funkcjonują w postaci zintegrowanej jako RNAvista [30] (patrz Rysunek 3). Aby uzyskać udokładnioną strukturę drugorzędową dla zadanej cząsteczki RNA, użytkownik może więc skorzystać bezpośrednio z narzędzia RNAvista, lub też osobno, najpierw z RNAComposer, a potem RNApdbee. W obu przypadkach uzyskane wyniki będą identyczne. W celu oszacowania jakości i precyzji wyników generowanych przez RC/Rp (RNAvista) przeprowadzono obszerny eksperyment obliczeniowy z wykorzystaniem struktur drugorzędowych RNA zdeponowanych w bazie danych RNA STRAND [31]. Uwzględniono zarówno struktury potwierdzone drogą eksperymentalną jak i uzyskane z analizy porównawczej sekwencji. Zbiór ten następnie podzielono na 4 podzbiory, biorąc pod uwagę jako kryterium długość sekwencji. Następnie wykorzystano je w dwóch eksperymentach: w pierwszym wejściem była tylko sekwencja, w drugim była nim dodatkowo też kanoniczna struktura 2D RNA. Wyniki pokazały, że w ogólności RC/Rp (dla całego zbioru testowego) przewidział poprawnie 30.64% - 57.57% niekanonicznych par zasad obecnych w strukturach referencyjnych w przypadku, gdy wejście stanowiła tylko sekwencja RNA, natomiast 49.91% - 70.51% dla sytuacji, gdy na wejściu podana była zarówno sekwencja jak i kanoniczna struktura 2D RNA.

Dodatkowo przeprowadzono też testy w zakresie dokładności predykcji oraz adnotacji par zasad dla motywów rekurencyjnych, które posiadają jedynie oddziaływania niekanoniczne. Uruchomiono RC/Rp (RNAvista) dla siedmiu takich motywów pobranych z bazy danych 3D Motif Atlas [32], przy czym wejście stanowiła jedynie sekwencja RNA. W celu uzyskania kanonicznej struktury 2D wykorzystano każdy z obecnie dostępnych w ramach RNAComposer programów do tego dedykowanych. Uzyskane wyniki jednoznacznie wskazują, że pierwszy krok wykonywany przez RC/Rp (RNAvista) ma bardzo istotny wpływ na wyniki końcowe. Szczególnie na uwagę zasługuje CentroidFold [33], który osiągnął najlepsze rezultaty na tym etapie. Właściwe rozpoznanie kanonicznych par zasad znacząco przekłada się na końcową rozszerzoną strukturę 2D RNA.



RYSUNEK 3 SCHEMAT DZIAŁANIA POTOKU PRZETWARZANIA RC/RP (RNAVISTA) PROWADZĄCEGO DO OTRZYMANIA ROZSZERZONEJ STRUKTURY DRUGORZĘDOWEJ NA PODSTAWIE SEKWENCJI I/LUB STRUKTURY RNA.

Algorytm symulujący proces nieenzymatycznej degradacji RNA

Badania prowadzone nad procesem degradacji RNA w warunkach charakteryzujących się nieobecnością czynników komórkowych (czyli tzw. hydrolizą nieenzymatyczną), pokazały, że jest to proces silnie zależny od struktury przestrzennej RNA (zarówno pierwszo-, drugo- jak i trzeciorzędowej). W pierwszych eksperymentach biochemicznych, przeprowadzonych na krótkich oligorybonukleotydach zauważono, że cięcia występowały jedynie w rejonach jednoniciowych, natomiast rejony dwuniciowe pozostawały stabilne. Dodatkowo zaobserwowano, że poszczególne wiązania internukleotydowe w łańcuchu RNA wykazują różną podatność na cięcie, i tak najbardziej niestabilnym okazało się być wiązanie między UA, które było hydrolizowane 1.5-2 razy szybciej niż między CA. Wiązanie YC (UC, CC) było 3-5 razy bardziej stabilne niż YA (UA, CA), a wiązanie pomiędzy YG (UG, CG) i YU (UU, CU) rozpadało się 20-50 razy wolniej niż pomiędzy YC (UC, CC) i YA (UA, CA). Pozostałe wiązania RR (AA, AG, GG, GA) i RY (AU, AC, GU, GC) wykazywały stabilność w badanych warunkach cięcia. Analiza omawianych wiązań pozwoliła na uszeregowanie ich począwszy od najmniej do najbardziej stabilnych w następujący sposób: UA>CA>YC>YG>YU, gdzie Y oznacza C lub U, a R oznacza A lub G [6]–[9]. Powoduje to, że każda cząsteczka RNA ma trochę inny wzór degradacji (listę fragmentów na jakie ulega rozpadowi).

Analiza wzoru degradacji cząsteczki RNA dostarcza wielu informacji, m.in.: pozwala wnioskować o jej stabilności, przewidywać czy powstałe z niej degradanty mają szansę przetrwać (nie ulec dalszej degradacji) i pełnić określone funkcje, czy też projektować sztuczne cząsteczki RNA, tak, aby rozpadały się generując zadane, stabilne degradanty.

Laboratoryjne zbadanie wzoru degradacji dowolnej cząsteczki RNA jest żmudne i czasochłonne, dlatego też zaproponowano algorytm podziału i odcięć symulujący przebieg procesu nieenzymatycznej degradacji RNA. Uwzględnia on sekwencję oraz strukturę przestrzenną zadanej cząsteczki RNA, a także reguły stabilności opracowane przez Kierzka i współpracowników [6]–[9], tak aby na bazie tych informacji przewidzieć jej wzór degradacji. Natomiast sam problem przewidywania rozpadu (degradacji) dowolnej cząsteczki RNA w wersji przeszukiwania został zdefiniowany w [16].

Zanim zostanie przedstawiony zaproponowany w pracy algorytm konieczne jest wprowadzenie kilku definicji.

Niech $P=\{UA, CA, UC, CC, UG, CG, UU, CU\}$ będzie zbiorem wszystkich dopuszczalnych miejsc cięcia w cząsteczce RNA zgodnie z danymi pochodzącymi z eksperymentów wykonanych przez Kierzka i współpracowników [6]–[9].

Na podstawie zależności zaobserwowanych pomiędzy szybkością rozpadu poszczególnych wiązań internukleotydowych w cząsteczce RNA, opracowano następujące przedziały wartości $[c, d]$ dla każdego dopuszczalnego miejsca cięcia w zbiorze P [16], [19]:

$$\text{Dla wiązań pomiędzy YG oraz YU:} \quad c_{(YG, YU)} = 0 \quad d_{(YG, YU)} = 0.1$$

$$\text{Dla wiązań pomiędzy YC:} \quad c_{(YC)} = 20 \quad d_{(YC)} = 50$$

$$\text{Dla wiązania pomiędzy CA:} \quad c_{(CA)} = 3 * c_{(YC)} \quad d_{(CA)} = 5 * d_{(YC)}$$

$$\text{Dla wiązania pomiędzy UA:} \quad c_{(UA)} = 1.5 * c_{(CA)} \quad d_{(UA)} = 2 * d_{(CA)}$$

Dla każdego miejsca cięcia ze zbioru P , uwzględniając błędy pomiarowe [34], [35], zdefiniowane zostały następujące miary degradacji (przynależności) [19]:

$$\mu(UA) = \frac{\left(d_{(UA)} - c_{(UA)}\right)^{\frac{1}{2}}}{1 + \left(d_{(UA)} - c_{(UA)}\right)^{\frac{1}{2}}} = 0.953, \quad \mu(CA) = \frac{\left(d_{(CA)} - c_{(CA)}\right)^{\frac{1}{2}}}{1 + \left(d_{(CA)} - c_{(CA)}\right)^{\frac{1}{2}}} = 0.932,$$

$$\mu(YC) = \frac{\left(d_{(YC)} - c_{(YC)}\right)^{\frac{1}{2}}}{1 + \left(d_{(YC)} - c_{(YC)}\right)^{\frac{1}{2}}} = 0.846, \quad \mu(YG, YU) = 0.1$$

W celu przewidzenia wzoru degradacji dowolnej cząsteczki RNA, zaprojektowany został algorytm podziału i odcięć (ang. *branch-and-cut*), który zaimplementowano w języku C++. Algorytm symuluje proces nieenzymatycznej degradacji RNA podanej na jego wejściu cząsteczki RNA, na bazie jej sekwencji oraz reguł stabilności zaproponowanych przez Kierzka i współpracowników. Zastosowanie tych reguł, wymaga, aby została wykonana predykcja struktury drugorzędowej, do czego wykorzystywane jest zewnętrzne narzędzie. W uzyskanej strukturze 2D wyszukiwane są odcinki jednoniciowe (o długości nie mniejszej niż dwa), po czym rozpoznawane są w nich miejsca cięcia obecne w zbiorze P i przypisywane są do nich odpowiadające im miary degradacji. Spośród nich, tylko te miejsca cięcia, które mają najwyższą wartość miary degradacji są brane pod uwagę. W każdym, spośród wybranych miejsc na-

stępnie wykonywane jest cięcie, którego wynikiem jest zawsze para fragmentów RNA. Ważnym założeniem jest, że zaraz po hydrolizie, oba wynikowe fragmenty oddysocjują i każdy z nich przybiera nową strukturę drugorzędową. W związku z tym, w celu przewidzenia dalszych produktów spontanicznej degradacji RNA, konieczne jest ponowne wykorzystanie zewnętrznego programu służącego do predykcji struktury 2D RNA. Z uwagi na to, że dane wyjściowe poszczególnych programów mają inny format, stworzono parser w języku Perl.

W sytuacji, gdy dla rozważanego fragmentu RNA nie udało się znaleźć dopuszczalnego miejsca cięcia (zgodnie ze zbiorem P) lub gdy iloczyn miar degradacji miał wartość mniejszą niż próg odcięcia (który jest parametrem podanym na wejściu algorytmu przez użytkownika), taki fragment uznawany jest za końcowy produkt degradacji i dodawany do zbioru wynikowego [19].

Algorytm kończy pracę, gdy nie istnieje już żaden fragment, który zawierałby dopuszczalne miejsce cięcia. Wyniki działania algorytmu (wzór degradacji cząsteczki RNA) prezentowany jest w postaci zbioru fragmentów będących wynikiem nieenzymatycznej degradacji.

Powyżej opisany algorytm został przetestowany w eksperymencie obliczeniowym z wykorzystaniem dwóch modelowych, sztucznych cząsteczek: RNA-A₁₀₈ (o długości 108 nt) oraz RNA-B₆₆ (o długości 66 nt). Cząsteczki te zaprojektowano tak, aby charakteryzowały się niską stabilnością w rozważanych warunkach eksperymentalnych oraz z uwzględnieniem reguł stabilności wiązań zaproponowanych przez Kierzka i współpracowników. Struktury drugorzędowe tych cząsteczek zostały przewidziane programem RNAstructure [36]. Następnie cząsteczki te poddano eksperymentowi biochemicznemu degradacji RNA z jednym znakowanym nukleotydem na 5' końcu (wykonanemu przez autorkę), którego celem było uzyskanie ich wzoru degradacji. Podstawowe kroki eksperymentu podsumowano w monografii, a dokładniejszy ich opis można znaleźć w [19]. Większość miejsc cięcia wystąpiła, tak jak tego oczekiwano, pomiędzy nukleotydami Y i A, ale pojawiły się też takie, które nie były zgodne z regułami stabilności opracowanymi przez Kierzka i współpracowników, mianowicie te na następujących pozycjach: 28, 30 i 31, w przypadku RNA-A₁₀₈ oraz 21, 23, 60 dla RNA-B₆₆.

Jednocześnie przewidziano miejsca cięcia dla tych cząsteczek *in silico*, z wykorzystaniem omawianego algorytmu podziału i odcięć, a następnie porównano z wynikami otrzymanymi z eksperymentu biochemicznego.

Ze względu na to, że predykcja struktury 2D wykonywana jest wielokrotnie przez algorytm i, że cięcia występują jedynie w rejonach jednoniciowych RNA, wybór odpowiedniego narzędzia do przewidywania struktury 2D jest bardzo istotny. Dlatego też, algorytm został przetestowany z wykorzystaniem różnych metod dedykowanych do predykcji struktury 2D. Zastosowano: RNAstructure, RNAfold, CentroidFold, CONTRAFold, ContextFold, IPknot i RC/Rp w wersjach RC/Rp-1, RC/Rp-2, RC/Rp-3 (zależnych od wyboru narzędzia do ekstrakcji par zasad ze struktury trzeciorzędowej RNA, odpowiednio RNAView, MC-Annotate, 3DNA/DSSR). RNAstructure, RNAfold, CentroidFold, CONTRAFold, ContextFold i IPknot są wbudowane w RNAComposer i użytkownik ma możliwość dokonania wyboru w zakresie tego, który z nich zostanie wykorzystany do predykcji kanonicznej struktury 2D. Za każdym razem, gdy uruchamiano RC/Rp (RNAvista), RNAComposer uruchamiany był z jednym, wskazanym przez użytkownika programem do przewidywania kanonicznej struktury 2D RNA (jeżeli był to RNAstructure to oznaczenie RC/Rp w wynikach było następujące: RC_{RNAstructure}/Rp-1, RC_{RNAstructure}/Rp-2 lub RC_{RNAstructure}/Rp-3, etc.).

W poniższej tabeli (patrz Tabela 1) przedstawiono wyniki uzyskane drogą eksperymentu biochemicznego oraz przykładowe, wygenerowane przez algorytm.

TABELA 1 PRZYKŁADOWE WYNIKI UZYSKANE PRZEZ ALGORYTM DLA SZTUCZNYCH CZĄSTECZEK: RNA-B₆₆ I RNA-A₁₀₈. MIEJSCA CIĘCIA PRZEWIDZIANE PRZEZ ALGORYTM, KTÓRE SĄ ZGODNE Z WYNIKAMI EKSPERYMENTU BIOCHEMICZNEGO ZOSTAŁY WYTŁUSZCZONE.

Programy do przewidywania struktury drugorzędowej	Fragmenty RNA zidentyfikowane w eksperymencie biochemicznym oraz przez algorytm	
	RNA-B ₆₆	RNA-A ₁₀₈
Wyniki eksperymentu biochemicznego związanego z nieenzymatyczną degradacją RNA	14, 18, 20, 21, 23, 24, 37, 39, 43, 45, 53, 60	12, 27, 28, 29, 30, 31, 61, 84
RNAfold	14, 18, 20, 24, 26, 35, 37, 39, 43, 45	27, 29, 61, 84
RC _{RNAfold} /Rp-1	14, 18, 20, 24, 35, 37, 43, 45	12, 23, 27, 29, 61, 84
RC _{RNAfold} /Rp-2	18, 20, 24, 35, 37, 43, 45, 53	12, 23, 27, 29, 61, 84
RC _{RNAfold} /Rp-3	14, 18, 20, 24, 35, 37, 39, 43, 45, 53	27, 29, 61, 84
RNAstructure	24, 26, 35, 37, 39, 43, 45, 53	27, 29, 61, 84
RC _{RNAstructure} /Rp-1	14, 18, 20, 24, 35, 37, 43, 45	12, 23, 27, 29, 61, 84
RC _{RNAstructure} /Rp-2	18, 20, 24, 35, 37, 43, 45, 53	12, 23, 27, 29, 61, 84
RC _{RNAstructure} /Rp-3	14, 18, 20, 24, 35, 37, 39, 43, 45, 53	27, 29, 61, 84

Analiza uzyskanych wyników pokazała, że algorytm przewiduje wszystkie możliwe fragmenty będące wynikiem degradacji wejściowego RNA (przy odpowiednim doborze narzędzi do predykcji kanonicznej struktury 2D). Wyniki działania algorytmu zostały porównane tylko z fragmentami uzyskanymi z eksperymentu biochemicznego degradacji RNA ze znakowanym nukleotydem na 5' końcu (czyli podzbiorem wszystkich degradantów). Pokazano wcześniej w [15], że jeśli jesteśmy w stanie dobrze przewidzieć fragmenty z 5' końca jako pierwotne lub wtórne, to jesteśmy też w stanie dobrze przewidzieć resztę fragmentów. Stąd też, jeśli dobrze przewidziane zostały przez algorytm te fragmenty, które można było zweryfikować z eksperymentem biochemicznym degradacji RNA z jednym znakowanym nukleotydem na 5' końcu, to reszta przewidzianych przez niego fragmentów też jest poprawna.

Analiza uzyskanych wyników wykazała również, że algorytm wykonuje swoją pracę szybko i efektywnie, nawet wtedy, gdy do wielokrotnych predykcji struktury 2D (refoldowania) wykorzystywany jest RC/Rp (RNAvista).

Algorytm był w stanie zidentyfikować poprawnie większość produktów nieenzymatycznej degradacji w przypadku obu badanych cząsteczek RNA. Wszystkie miejsca cięcia, jakie nie zostały przez algorytm rozpoznane to miejsca niezgodne z regułami opracowanymi przez Kierzka i współpracowników. Niestety, zgodnie z posiadaną przez autorkę wiedzą, jedynie Kierzek i współpracownicy analizowali jakościowo i ilościowo stabilność wiązań internukleotydowych w łańcuchach RNA oraz zaproponowali reguły opisujące ich stabilność.

4.3.7. Model nieenzymatycznej degradacji RNA w oparciu o stochastyczne sieci Petriego

W rozdziale 6 został przedstawiony oraz przeanalizowany model matematyczny oparty na stochastycznych sieciach Petriego przedstawiający przebieg procesu nieenzymatycznej degradacji RNA z uwzględnieniem struktury przestrzennej RNA, a także reguł stabilności zaproponowanych przez Kierzka i współpracowników (analogicznie działa algorytm przedstawiony w rozdziale 5) oraz eksperymentu biochemicznego (patrz rozdział 5). Zaprezentowano również nowe podejście umożliwiające

oszacowanie wartości opóźnień dla tranzycji w sieciach czasowych DPN oraz częstości uruchomień dla tranzycji reprezentujących procesy biologiczne w modelach opartych o stochastyczne sieci Petriego.

Heurystyka pozwalająca na oszacowanie wartości opóźnień dla tranzycji w sieciach czasowych DPN oraz częstości uruchomień dla tranzycji reprezentujących procesy biologiczne w modelach opartych o stochastyczne sieci Petriego

Modele zbudowane w oparciu o klasyczne sieci Petriego są modelami jakościowymi, co oznacza, że opisują one jedynie strukturę rozważanego systemu. W przypadku systemów biologicznych nie stanowi to dużego ograniczenia, ponieważ struktura w dużej mierze determinuje ich funkcję i w wielu przypadkach pozwala na odkrycie interesujących własności badanego układu. Dodatkowo, dysponując różnego rodzaju danymi ilościowymi, można je uwzględnić w modelu posługując się rozszerzeniami klasycznych sieci Petriego, np. czasowymi (istnieje wiele ich odmian) czy stochastycznymi. Sieci takie zwiększają możliwości zastosowania teorii sieci Petriego do badania złożonych systemów biologicznych.

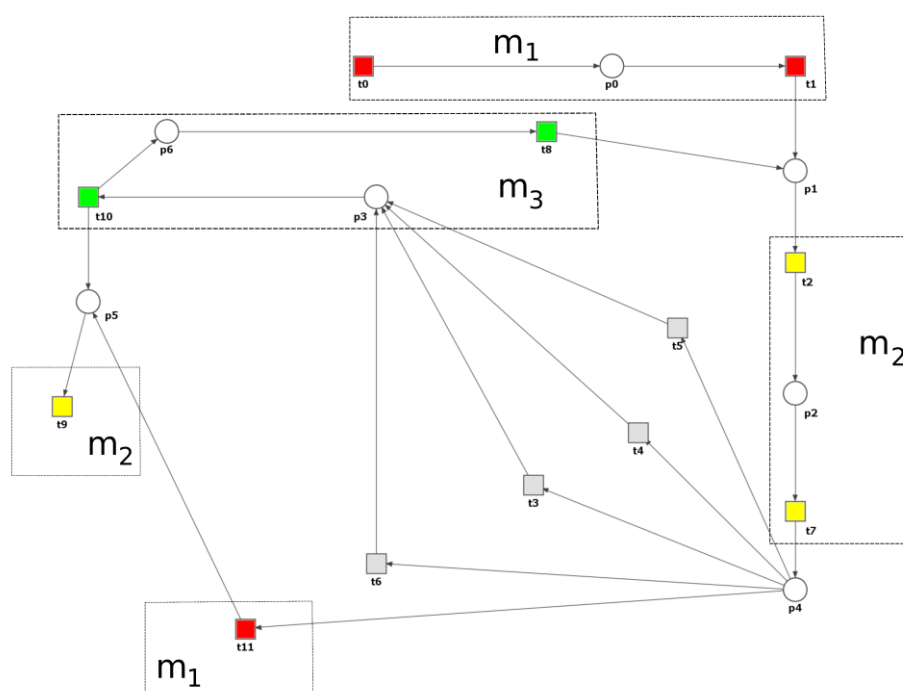
Z badaniami dotyczącymi sieci czasowych oraz stochastycznych bardzo często związany jest problem doboru prawidłowych i precyzyjnych parametrów czasowych czy też kinetycznych. Parametry te powinny odzwierciedlać rzeczywistość biologiczną i gwarantować, że sieć w trakcie symulacji zachowywać będzie się zgodnie z dostępną wiedzą biologiczną w zakresie badanego procesu. Niestety, dla większości reakcji chemicznych i procesów brak jest dokładnych danych eksperymentalnych dotyczących czasu ich trwania lub informacje te są bardzo nieprecyzyjne. Dlatego też, opracowano nowe podejście pozwalające na oszacowanie wartości opóźnień dla tranzycji w sieciach czasowych DPN oraz częstości uruchomień dla tranzycji reprezentujących procesy biologiczne w modelach opartych o stochastyczne sieci Petriego na bazie nieprecyzyjnych danych pochodzących z danych literaturowych oraz wiedzy eksperta. Podejście zaprezentowano poniżej [37].

1. Oszacowanie wartości opóźnienia dla każdej tranzycji (czasu jaki musi upłynąć, aby uruchomiona tranzycja wyprodukowała tokeny). Uzyskujemy w ten sposób wartości opóźnień dla sieci DPN.
 - a. W ramach rozważanego modelu klasycznej sieci Petriego należy zidentyfikować te procesy, które trwają najkrócej oraz te, które trwają najdłużej.
 - b. Następnie na bazie danych literaturowych (czasy przybliżone np. około godziny, kilka minut) oraz korzystając z wiedzy eksperta (o pewnych zależnościach np. synteza trwa dłużej niż wiązanie liganda do receptora, tworzenie złożonego kompleksu i wiązanie do receptora trwa krócej niż oddysocjowanie od receptora etc.), należy opracować własną skalę, np. o zakresie wartości od 1 do 500.
 - c. Wszystkim tranzycjom obecnym w rozważanym modelu, trzeba przypisać wartości z zakresu opracowanej skali.
 - d. Następnie należy przeprowadzić analizę symulacyjną modelu w celu sprawdzenia, czy każda tranzycja ma szansę się uruchomić. Jeżeli w trakcie symulacji okaże się, że pewne tranzycje są nieaktywne, koniecznym jest powrót do kroku 1b celem modyfikacji skali lub/i przeanalizowania zależności czasowych pomiędzy modelowanymi procesami.
2. Oszacowanie częstości uruchomienia dla każdej tranzycji. Uzyskujemy w ten sposób (na bazie obliczonych w kroku 1 wartości opóźnień dla sieci DPN) częstości uruchomień dla tranzycji w sieci SPN (stochastycznej sieci Petriego).
 - a. Częstości uruchomień dla każdej tranzycji oblicz jako odwrotność wartości opóźnienia obliczonego dla niej w kroku 1.

W celu uzyskania parametrów czasowych dla sieci czasowej DPN, wystarczy ograniczyć się jedynie do wykonania kroku 1 w ramach omawianego podejścia. Natomiast, aby uzyskać częstości uruchomień dla tranzycji w stochastycznej sieci Petriego, konieczne jest wykonanie obu kroków opisanej powyżej heurystyki.

Model nieenzymatycznej degradacji RNA w oparciu o stochastyczne sieci Petriego

Zaproponowany model został wykonany za pomocą narzędzia Holmes [38], a jego struktura została przedstawiona na Rysunku 4. Sieć składa się z 7 miejsc oraz 12 tranzycji, których numery wraz z przypisanymi im nazwami zostały opisane w monografii.



RYСУNEK 4 MODEL SIECI PETRIEGO PRZEDSTAWIAJĄCY PRZEBIEG PROCESU NIEENZYMATYCZNEJ DEGRADACJI RNA. NIE-TRYWIALNE ZBIORY MCT ZOSTAŁY OTOCZONE RAMKĄ ORAZ PODPISANE. TRANZYCCJE W RAMACH DANYCH ZBIORÓW MCT ZAZNACZONE SĄ ZA POMOCĄ PROSTOKĄTÓW WYPEŁNIONYCH KOLOREM (DLA KAŻDEGO ZBIORU INNYM).

Zaprezentowany powyżej model nie uwzględnia informacji ilościowych, a jedynie opisuje strukturę rozważanego systemu, jest więc modelem jakościowym. Ze względu na dostępność przybliżonych informacji ilościowych dotyczących stabilności wiązań internukleotydowych w cząsteczce RNA [6]–[9], wykorzystano stochastyczną sieć Petriego jako rozszerzenie omawianego modelu. W tym celu, dla każdej tranzycji należało oszacować jej częstość uruchomienia. Tranzycjom, które odpowiadały cięciom pomiędzy poszczególnymi wiązaniami internukleotydowymi przypisano częstości uruchomień na bazie miar degradacji (przynależności) zdefiniowanych i obliczonych w rozdziale 5. W przypadku pozostałych tranzycji, wykorzystano heurystykę opisaną w rozdziale 6 w celu oszacowania ich częstości uruchomień.

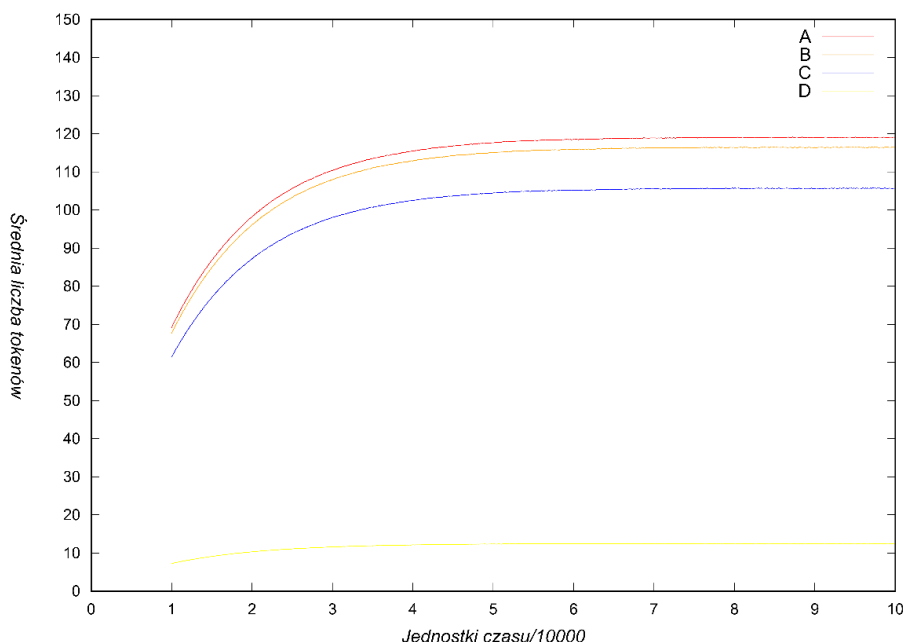
Dla modelu przeprowadzono analizę strukturalną za pomocą oprogramowania Holmes [38]. Sieć pokryta jest 5 t-niezmiennikami oraz nie zawiera minimalnych p-niezmienników. W przedstawionym modelu można wyróżnić 5 zbiorów MCT.

W celu dokonania pełniejszego opisu zależności występujących w rozważanym modelu, przeprowadzono analizę typu *knockout* w oparciu o zbiór t-niezmienników (patrz rozdział 2). Do przeprowadzenia analizy tego typu wykorzystano oprogramowanie MonaLisa [39]. Dla każdego zbioru MCT podlegającym wyłączeniu obliczono, które tranzycje w wyniku tego staną się nieaktywne.

Dodatkowo, przeprowadzono analizę symulacyjną z wykorzystaniem algorytmu symulacji stochastycznej Gillespiego (SSA) [40] zaimplementowanego w oprogramowaniu Snoopy [41]. W ramach przeprowadzonej symulacji ustalono liczbę kroków algorytmu na 100000, wyniki są natomiast średnią z 50000 jego uruchomień [37].

Wyniki przeprowadzonej symulacji, które zostały zaprezentowane na Rysunku 5 pokazują, że zależności dotyczące podatności na hydrolizę nieenzymatyczna wiązań łączących poszczególne nukleotydy w łańcuchu RNA zostały zachowane (UA>CA>YC>YG>YU).

Szczegółowy opis przeprowadzonych analiz został zamieszczony w omawianej monografii, w rozdziale 6, a ich teoretyczne podstawy w rozdziale 2.



RYСУNEK 5 WYNIKI SYMULACJI DLA TRANZYCJI MODELU: A. PRZEBIEG SYMULACJI DLA TRANZYCJI T_3 (CIĘCIE MIĘDZY NUKLEOTYDAMI UA). B. PRZEBIEG SYMULACJI DLA TRANZYCJI T_4 (CIĘCIE MIĘDZY NUKLEOTYDAMI CA), C. PRZEBIEG SYMULACJI DLA TRANZYCJI T_5 (CIĘCIE MIĘDZY NUKLEOTYDAMI YC), D. PRZEBIEG SYMULACJI DLA TRANZYCJI T_6 (CIĘCIE MIĘDZY NUKLEOTYDAMI YG, YU).

4.3.8. Model enzymatycznej degradacji RNA w oparciu o stochastyczne sieci Petriego

W rozdziale 7 monografii przedstawiono oraz przeanalizowano model matematyczny oparty na stochastycznych sieciach Petriego, przedstawiający proces enzymatycznej degradacji RNA na przykładzie YRNA oraz wywodzących się z niego degradantów sRNY i ich roli w regulacji apoptozy i zapalenia w monocytach/makrofagach. Sieć została wykonana za pomocą narzędzia Holmes [38] i pokazana jest na Rysunku 6. Miejsca reprezentowane w postaci podwójnych okręgów na Rysunku 6 nazywane są miejscami logicznymi i stanowią graficzną reprezentację tego samego miejsca w sieci (np. miejsce p_{34} ma 3 różne lokalizacje).

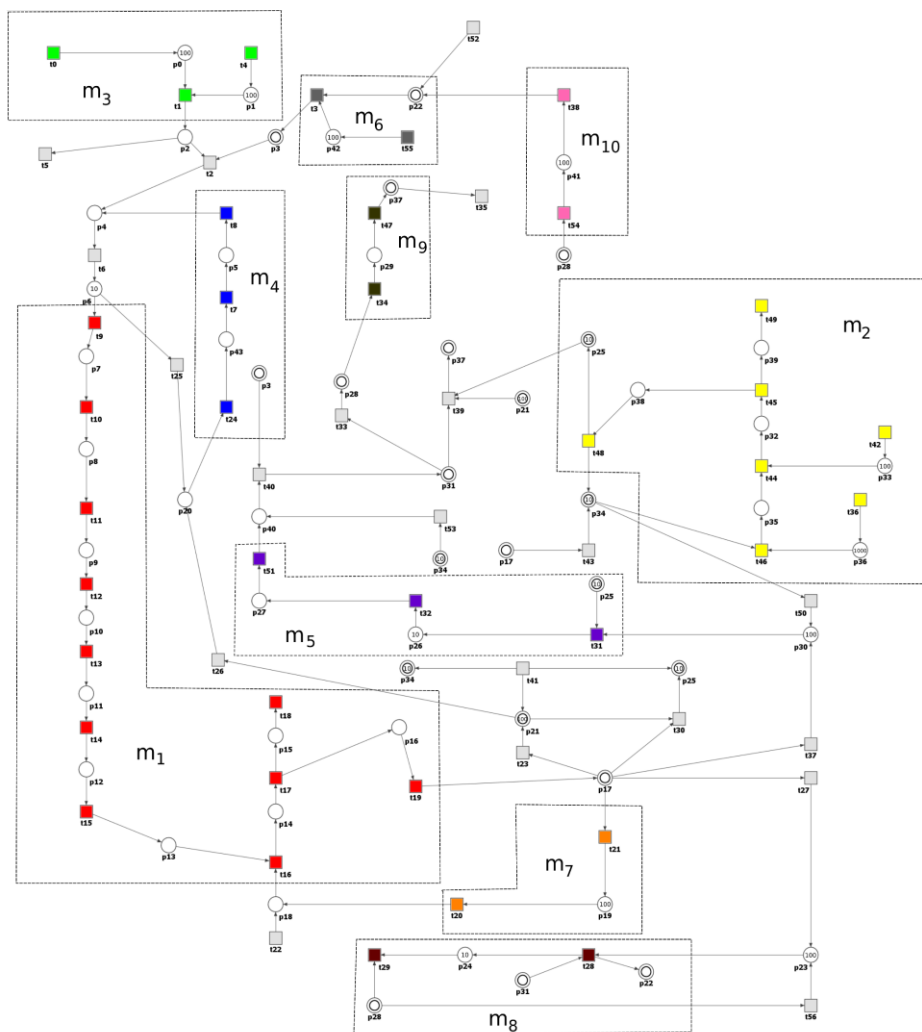
Sieć składa się z 44 miejsc oraz 57 tranzycji (których numery wraz z przypisanymi im nazwami zostały opisane w monografii) oraz znakowania początkowego określającego ilość tokenów dla każdego miejsca (zero lub więcej). Dla każdej tranzycji w sieci oszacowano jej częstość uruchomienia wykorzystując opisaną w rozdziale 6 heurystykę.

Dla modelu przeprowadzono analizę strukturalną za pomocą oprogramowania Holmes [38]. Sieć w całości pokryta jest przez 82 t-niezmienniki i nie zawiera minimalnych p-niezmienników. W związku z tym, dalsza analiza skoncentrowana była na t-niezmiennikach. Następnym krokiem analizy było wygenerowanie zbiorów MCT na bazie zbioru t-niezmienników. Dzielą one strukturę sieci na oddzielne obszary, które mają pewne określone funkcje biologiczne w ramach modelowanego systemu. W sieci wyróżniono 10 nietrywialnych zbiorów MCT (tj. składających się z dwóch lub więcej tranzycji). Następnie wykonano klastrowanie algorytmem UPGMA (ang. *unweighted pair group method with arithmetic mean*) opierającym się na metodzie średnich połączeń oraz zastosowano miarę korelacji Pearson jako miarę podobieństwa, w wyniku czego otrzymano 7 klastrów t-niezmienników. Liczbę klastrów wybrano przeprowadzając test porównawczy dla liczby klastrów od 2 do 20, gdzie do oceny jakości podziału na klastry użyto konturowej miary jakości (ang. *mean split silhouette*, MSS). Analiza biologiczna wyników klastrowania pozwoliła na sformułowanie następujących wniosków. Zauważono, że stymulacja receptora TLR7 poprzez kompleksy sRNY-Ro60 (zarówno zewnątrz- jak i wewnątrzkomórkowe), a co za tym idzie uruchomienie dwóch ścieżek sygnałnych prowadzących do aktywacji stanu zapalnego (ścieżka sygnałna zależna od NF- κ B) i apoptozy (ścieżka sygnałna zależna od aktywacji kaspazy 3), ma bardzo istotny wpływ na proces pogłębiania się zmian miażdżycowych.

Wyniki klastrowania nie zawsze okazują się być w pełni satysfakcjonujące. Jako wynik powyższej analizy otrzymano dwa duże klastry obejmujące praktycznie całą sieć (94% wszystkich t-niezmienników) oraz 5 klastrów posiadających po jednym t-niezmienniku (dokładny opis znajduje się w omawianej monografii). Uzyskany wynik jest cechą unikalną rozważanego systemu, sugerującą, że modelowane w ramach niego procesy są silnie od siebie zależne i ze sobą powiązane.

W celu dokonania pełnej i właściwej interpretacji biologicznej rozpatrywanej sieci, konieczne było przeprowadzenie dodatkowych analiz. Z tego też powodu przeprowadzono dwa rodzaje analizy typu *knockout*: w oparciu o zbiór t-niezmienników oraz symulacyjną.

Celem pierwszej analizy było określenie wpływu wyłączenia poszczególnych tranzycji w sieci na zbiór t-niezmienników oraz pośrednio na inne tranzycje (reakcje) w modelowanym systemie. Do jej przeprowadzenia wykorzystano oprogramowanie MonaLisa [39]. Dla każdej tranzycji, która została wyłączona obliczono, które tranzycje w wyniku tego staną się nieaktywne oraz które podprocesy (określone przez t-niezmienniki) nie będą mogły być wykonane. Dodatkowo, jak to zdefiniowano w rozdziale 2, wyłączane tranzycje, które występują zawsze razem w tych samych nietrywialnych zbiorach MCT, mają taki sam wpływ na modelowany system. W związku z tym ich wpływ na zachowanie systemu biologicznego został przedstawiony wspólnie. Dla każdego zbioru MCT podlegającemu *knockoutowi* pokazano wyniki w postaci procentu t-niezmienników oraz tranzycji na które ma on wpływ. Wyniki zostały przedstawione w monografii w postaci tabelarycznej a ich analiza pozwoliła na przedstawienie wniosków w zakresie wpływu kompleksów sRNY-Ro60 (zarówno zewnątrz- jak i wewnątrzkomórkowych), działających poprzez aktywację receptora TLR7, na proces pogłębiania się zmian miażdżycowych.



RYSUNEK 6 MODEL SIECI PETRIEGO WRAZ Z POCZĄTKOWYM ZNAKOWANIEM, PRZEDSTAWIAJĄCY UDZIAŁ s-RNY POWSTAJĄCYCH Z YRNA W REGULACJI APOPTOZY I ZAPALENIA W MONOCYTACH/MAKROFAGACH. NIETRZYWIALNE ZBIORY MCT ZOSTAŁY OTOCZONE RAMKĄ ORAZ PODPISANE. TRANZycJE W RAMACH DANYCH ZBIORÓW MCT ZAZNACZONE SĄ ZA POMOCĄ PROSTOKĄTÓW WYPEŁNIONYCH KOLOREM (DLA KAŻDEGO ZBIORU INNYM). MIEJSCA LOGICZNE REPREZENTOWANE SĄ JAKO PODWÓJNE OKRĘGI.

Analiza symulacyjna rozważanej sieci stochastycznej została wykonana z wykorzystaniem algorytmu symulacji stochastycznej Gillespiego (SSA) [40] zaimplementowanego w oprogramowaniu Snoopy [41]. W każdej symulacji ustalono liczbę kroków algorytmu na 100000, wyniki są średnią z 50000 uruchomień algorytmu. Przeprowadzona analiza typu *knockout* polegała na wybraniu tranzycji, które następnie zostały wyłączone poprzez ustalenie przypisanej im częstości uruchomienia na równą 0, a ich wpływ na modelowany system został przeanalizowany.

W ramach analizy sprawdzono następujące scenariusze:

- Wpływ kompleksu s-RNY/Ro60 na proces apoptozy i pogłębienia się zmian miażdżycowych.
- Wpływ aktywacji receptora TLR7 na proces apoptozy i pogłębienia się zmian miażdżycowych.
- Wpływ aktywacji receptora TLR7 oraz reaktywnych form tlenu powstających w wyniku różnych procesów patofizjologicznych na proces apoptozy i pogłębienia się zmian miażdżycowych.
- Wpływ Interleukiny-6 (Il-6) na proces apoptozy i pogłębienia się zmian miażdżycowych.
- Wpływ peroksydacji lipidów na procesy prowadzące do apoptozy i pogłębienia się zmian miażdżycowych.

Szczegółowy opis przeprowadzonych analiz został zamieszczony w omawianej monografii, w rozdziale 7, a ich teoretyczne podstawy w rozdziale 2.

W efekcie przeprowadzonej symulacyjnej analizy *knockout* modelu, sformułowano następujące wnioski biologiczne. Uzyskane wyniki potwierdzają, że zablokowanie receptora TLR7 (co może zostać zrealizowane np. za pomocą chlorochiny) lub s-RNY (co można osiągnąć np. poprzez komplementarne oligonukleotydy) powoduje, że nie dochodzi do uruchomienia ścieżek sygnałowych: zależnej od NF- κ B oraz zależnej od aktywacji kaspazy 3 i prowadzącej do apoptozy. Jednocześnie, mimo, iż ścieżka sygnałowa NF- κ B jest nieaktywna, stan zapalny ulega jedynie pewnemu obniżeniu, podobnie jak proces pogłębiania się zmian miażdżycowych. Wynika to z tego, że miażdżycy jest chorobą złożoną i wieloaspektową i inne czynniki (np. RFT i RFA) mają też wpływ na jej przebieg i rozwój.

Jeśli dodatkowo zostaną wyeliminowane RFT (reaktywne formy tlenu) i RFA (reaktywne formy azotu) powstające w wyniku różnych procesów patofizjologicznych, to dopiero wtedy obserwujemy zatrzymanie procesu zapalnego i pogłębiania się zmian miażdżycowych.

Natomiast zablokowanie reakcji zapalnej bezpośrednio, poprzez inhibicję cytokiny prozapalnej Il-6, powoduje z jednej strony zatrzymanie pogłębiania się zmian miażdżycowych, a z drugiej niewielki wzrost apoptozy, co jest zgodne z danymi dostępnymi w literaturze [42]–[44].

W sytuacji gdy zablokowany zostanie proces peroksydacji lipidów, wyniki są bardzo zbliżone do tych, zaprezentowanych dla inhibicji cytokiny prozapalnej Il-6. Zaobserwować można z jednej strony całkowite zahamowanie progresji miażdżycy i procesu zapalnego, a z drugiej niewielki wzrost apoptozy, co jest również zgodne z informacjami dostępnymi w literaturze [45], [46].

4.3.9. Metoda projektowania sztucznych cząsteczek małych regulatorowych RNA

W rozdziale 8 omawianej monografii zostało opisane podejście do projektowania małych regulatorowych RNA na przykładzie miRNA, zaimplementowane jako narzędzie o nazwie AmiRNA Designer.

MikroRNA (miRNA, ang. *microRNA*) są najlepiej poznanymi małymi regulatorowymi RNA (srRNA, ang. *small regulatory RNA*), najpowszechniej występującymi w komórkach roślinnych i zwierzęcych, których długość wynosi 19–25 nt (przeciętnie 22 nt). Geny dla miRNA ulegają w procesie transkrypcji przepisaniu przez RNA Polimerazę II na pri-miRNA (pierwotny prekursor) o wielkości do kilku kb, który przyjmuje strukturę połączonych ze sobą szpilek do włosów. W przypadku zwierząt i roślin występują różnice w biogenezie miRNA oraz strukturze ich prekursorów. U zwierząt w dojrzewanie pri-miRNA zaangażowane są dwa enzymy z rodziny rybonukleaz III: Drosha i Dicer. Enzym Drosha na terenie jądra komórkowego dokonuje wstępnej obróbki pri-miRNA, skracając go na końcach 5' i 3' i generując produkt określany jako pre-miRNA o długości około 70 nukleotydów. Pre-miRNA posiadają strukturę spinki do włosów składającej się z pętli i dwuniciowego trzonu, w ramach którego na jego ramieniu 5' i/lub 3' znajduje się miRNA o długości około 21 nukleotydów. Pre-miRNA transportowany jest do cytoplazmy, gdzie enzym Dicer wycina z niego dupleks miRNA/miRNA*. U roślin prekursor miRNA są zdecydowanie dłuższe, osiągając rozmiary przekraczające nawet 600 nukleotydów. Dodatkowo charakteryzują się dużym zróżnicowaniem strukturalnym, a cały proces wycięcia dupleksu miRNA/miRNA* o długości ok. 21 nt realizowany jest na terenie jądra komórkowego przez jeden enzym - endorybonukleazę DCL1 (ang. *Dicer-like 1 protein*) [47], [48].

W kolejnym kroku (podobnie u roślin jak i u zwierząt) dojrzałe miRNA (nić przewodnia dupleksu RNA) zostaje włączone w kompleks RISC, podczas gdy miRNA* (nić towarzysząca) ulega degradacji. Rozpoznanie i wybór nici przewodniej miRNA przez kompleks RISC następuje w oparciu o własności termodynamiczne dupleksu miRNA/miRNA*. MiRNA wraz z kompleksem RISC może rozpoznawać mRNA genu docelowego i wpływać na poziom jego ekspresji. Istotą tego mechanizmu regulacji jest komplementarność pomiędzy miRNA a obszarem 3' UTR (ang. *3' Untranslated Region*) transkryptu: jeżeli komplementarność ta jest całkowita to wtedy ma miejsce degradacja mRNA, natomiast jeżeli jest niepełna, to następuje zatrzymanie procesu translacji, przy czym ilość transkryptu pozostaje bez zmian.

W ten sposób dokonuje się zazwyczaj regulacja ekspresji genów zwierzęcych. W przypadku modeli roślinnych ta korelacja nie jest już tak oczywista, a mechanizm według którego działa miRNA pozostaje nadal nie do końca wyjaśniony [48], [49].

W ostatnich latach został położony duży nacisk na rozwój technik regulacji ekspresji genów w oparciu o endo- i egzogenne miRNA. Główną ideą stojącą za metodologią opartą na sztucznych miRNA (zwanymi amiRNA, ang. *artificial miRNA*) jest zaprojektowanie takich sekwencji RNA o długości 21 nt, które regulowałyby ekspresję wybranego genu. Tak opracowana sekwencja, wprowadzana jest następnie do prekursora miRNA, zastępując w nim natywne miRNA. Powstała w ten sposób struktura przybiera postać zbliżoną do naturalnego prekursora miRNA i podlega identycznym procesom przetwarzania, jakie zachodzą w ramach biogenezy miRNA, mimo, że amiRNA nie było pierwotnie obecne w rozważanym organizmie. Stąd też, tak przygotowana sekwencja amiRNA nazywana jest sztuczną. Technologia oparta o sztuczne miRNA postrzegana jest jako bardzo obiecująca, nie tylko w genomowych badaniach funkcjonalnych, ale także w zakresie odporności roślin na wirusy czy też w zakresie inżynierii genetycznej, której celem jest m.in. poprawa jakości plonów [48], [50].

Obecnie dostępnych jest kilka narzędzi umożliwiających projektowanie sztucznych miRNA dla różnych grup organizmów. Ze względu jednak na to, że biogeneza miRNA różni się dość znacząco w przypadku zwierząt i roślin, powstające narzędzia są wyspecjalizowane i dedykowane do konkretnych organizmów lub ich grup. Najczęściej stosowanym narzędziem jest oprogramowanie Web MicroRNA Designer (WMD3) [51].

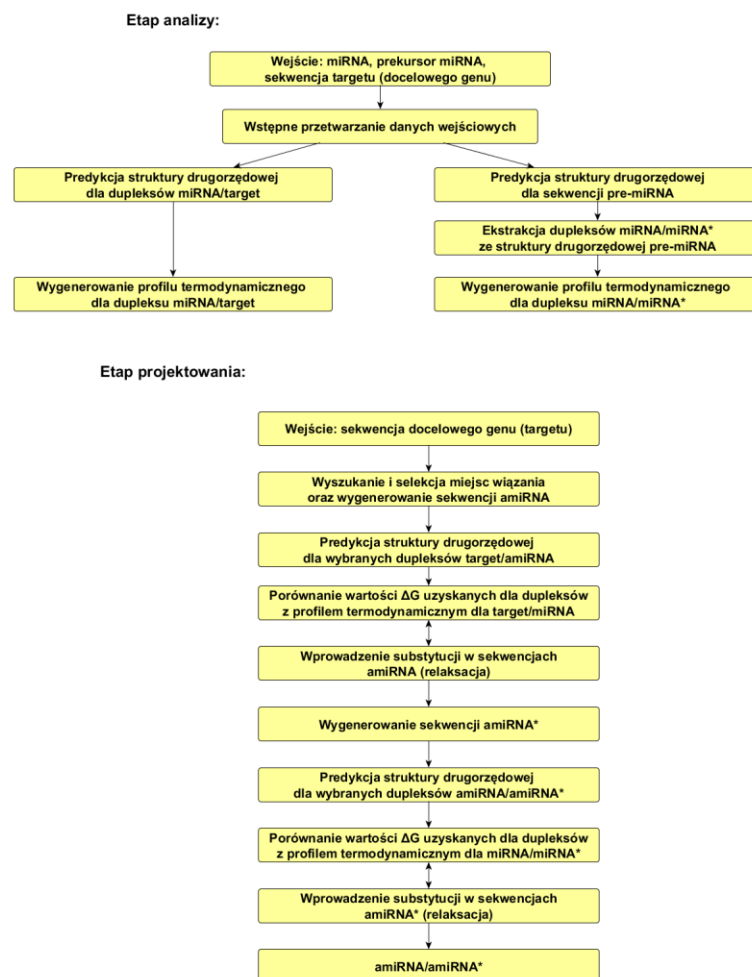
Przeprowadzono również eksperymentalne analizy przydatności wygenerowanych amiRNA w procesie wyciszania wybranych genów. Eksperyment biochemiczny wykonano dla amiRNA proponowanych przez narzędzie WMD3. Wykazał on, że sekwencje amiRNA wskazane przez WMD3 jako rekomendowane charakteryzowały się mniejszą efektywnością w zakresie wyciszania wybranych genów, niż te, które narzędzie to zaklasyfikowało jako zdecydowanie mniej polecane (osiągnęły efektywność w granicach 90%). Oznacza to, że zasady decydujące o tym, które amiRNA mają szansę zadziałać efektywnie po wprowadzeniu do danego organizmu nie są jeszcze w pełni poznane i zrozumiane. Dodatkowo, większość dostępnych narzędzi bazuje jedynie na komplementarności sekwencji oraz innych parametrach typu: obliczanie całościowej energii swobodnej hybrydyzacji dupleksu miRNA/gen docelowy (target) i/lub strukturze drugorzędowej prekursora miRNA [52]. Analiza statystyczna profili termodynamicznych opublikowanych i znanych miRNA pokazała, że wykazują one specyficzne termodynamiczne właściwości, które mają ogromny wpływ na funkcjonalność tych cząsteczek i są one odmienne od tych, które posiadają niefunkcjonalne miRNA [53], [54]. W odpowiedzi na to, zostało opracowane nowe podejście do projektowania sztucznych miRNA, opisane w niniejszej rozprawie. W ramach proponowanej metody wykonywana jest analiza rozkładu energii swobodnej dla dupleksów znanych miRNA/miRNA* oraz miRNA/target i na tej podstawie budowane są dla nich profile termodynamiczne. Następnie dla wybranego genu proponowane są potencjalne amiRNA, które oceniane są poprzez porównanie z utworzonymi wcześniej profilami. Pod uwagę brane są zarówno wartości energii swobodnej na każdej pozycji potencjalnego amiRNA jak i jej ogólny trend. Dla wyselekcjonowanych kandydatów wykonywana jest też analiza off-target, której celem jest wyeliminowanie tych amiRNA, które mogłyby dodatkowo wyciszać inne, przypadkowe geny danego organizmu [48].

Podejście do projektowania małych regulatorowych RNA na przykładzie miRNA, opisane w rozdziale 8 monografii, zostało zaimplementowane w języku C# pod systemy z rodziny Windows, jako narzędzie o nazwie AmiRNA Designer. Dostępne jest do pobrania, wraz z przykładowymi danymi wejściowymi oraz podręcznikiem użytkownika na stronie <http://www.cs.put.poznan.pl/arybarczyk/AmiRNA/>.

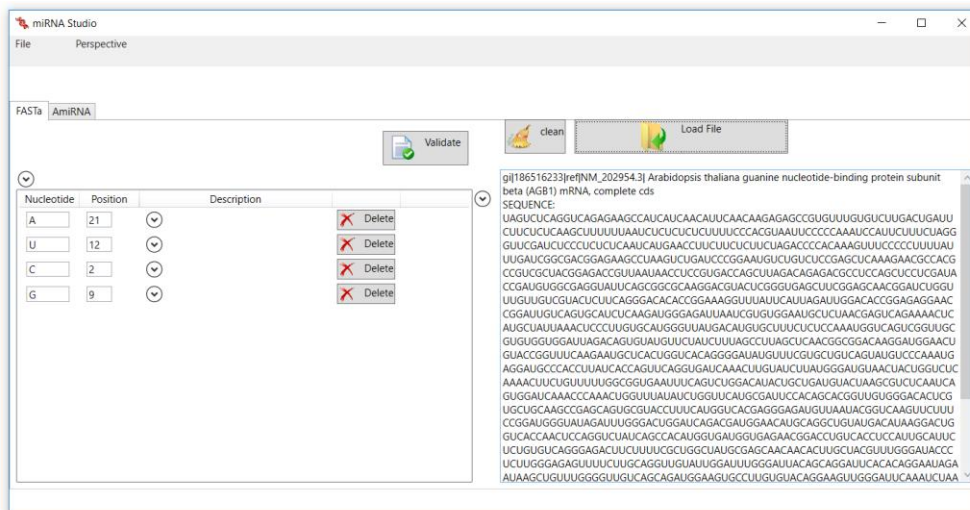
AmiRNA Designer umożliwia projektowanie sztucznych cząsteczek miRNA (amiRNA) ukierunkowanych na określony mRNA (target, który będzie wyciszany), bazując na właściwościach naturalnych

miRNA. Potencjalne jego zastosowania obejmują m.in. genomikę funkcjonalną czy też inżynierię genetyczną. W przeciwieństwie do istniejącego oprogramowania AmiRNA Designer pozwala na: przeanalizowanie naturalnych miRNA dla dowolnie wybranego organizmu, stworzenie profili termodynamicznych miRNA/miRNA* oraz miRNA/target na podstawie danych termodynamicznych wygenerowanych przez UNAFold [55] podczas predykcji struktury drugorzędowej oraz stworzenie własnych reguł filtracji na bazie sekwencji miRNA. Pozwala także na filtrowanie wyników amiRNA (analiza off-target), czyli sprawdzenie czy zaprojektowany amiRNA nie wchodzi w interakcję z mRNA, które nie są jego targetem. Ten etap wykonywany jest za pomocą stworzonych w tym celu skryptów w języku Perl.

Projektowanie amiRNA, zaimplementowane w oprogramowaniu AmiRNA Designer składa się z dwóch etapów: analizy, która obejmuje przetworzenie endogennych miRNA danego organizmu w celu uzyskania parametrów wymaganych do oceny potencjalnych amiRNA na kolejnym etapie oraz projektowania, w ramach którego zaproponowane zostaną amiRNA (oraz amiRNA*) i poddane filtracji na podstawie informacji pochodzących z poprzedniego etapu. Etap analizy dla danego organizmu wykonywany jest tylko raz, natomiast etap projektowania może być powtarzany wielokrotnie dla różnych genów docelowych. Ogólny schemat podejścia zobrazowany jest na Rysunku 7, natomiast Rysunek 8 przedstawia przykładowe okno oprogramowania AmiRNA Designer.



RYSUNEK 7 SCHEMAT DZIAŁANIA PODEJŚCIA DO PROJEKTOWANIA SZTUCZNYCH MI RNA ZAIMPLEMENTOWANY JAKO NARZĘDZIE AMI RNA DESIGNER. SCHEMAT PREZENTUJE NAJWAŻNIEJSZE KROKI WYKONYWANE PRZEZ AMI RNA DESIGNER, PODZIELONE NA DWA ETAPY: ANALIZY I PROJEKTOWANIA, KTÓRE MOGĄ BYĆ WYKONYWANE NIEZALEŻNIE OD SIEBIE. W WYNIKU PIERWSZEGO ETAPU OTRZYMYWANE SĄ PROFILE TERMODYNAMICZNE. DRUGI KROK UMOŻLIWIA ZAPROJEKTOWANIE SEKWENCJI AMI RNA ORAZ AMI RNA* DLA WYBRANEGO GENU DOCELOWEGO, BAZUJĄC NA WYNIKACH UZYSKANYCH NA ETAPIE ANALIZY



RYSUNEK 8 PRZYKŁADOWE OKNO OPROGRAMOWANIA AMIRNA DESIGNER

AmiRNA Designer został przetestowany na danych pochodzących z *Arabidopsis thaliana*. Sekwencje 157 prekursorów pre-miRNA pobrano z bazy miRBase rel18 [56], natomiast sekwencje 210 miRNA oraz ich genów docelowych (targetów) zaimportowano z bazy danych ASRP [57]. W celu wykonania analizy off-target, wykorzystano plik z transkryptami z bazy danych PlantGDB [58]. Geny, które wybrano jako docelowe celem ich wyciszenia, pobrano z bazy danych TAIR (<http://www.arabidopsis.org/>).

W ramach eksperymentu obliczeniowego wygenerowano potencjalne amiRNA dla następujących genów: AGB1 (ID: AT4G34460), oraz Immutans (ID: AT4G22260). Analiza przeprowadzona za pomocą oprogramowania psRNATarget [59] wykazała, że dla rozważanych genów nie istnieją naturalne miRNA, które mogłyby je wyciszać. Stanowi to dowód na to, że wśród miRNA wykorzystanych do stworzenia profili termodynamicznych nie był obecny żaden miRNA, który byłby dedykowany do regulacji badanych genów [48].

Faza projektowania została przeprowadzona z następującymi filtrami sekwencyjnymi: nukleotyd U na pozycji 12 (liczony względem sekwencji targetu) oraz opcją wymuszającą U na pierwszej pozycji amiRNA. Dla genu AGB1, AmiRNA Designer zaproponował 22 dupлексы amiRNA/amiRNA*, natomiast dla Immutans 20 kandydatów amiRNA/amiRNA*.

W celu przeprowadzenia dodatkowej weryfikacji działania narzędzia AmiRNA Designer, wykorzystano oprogramowanie psRNATarget z domyślnymi parametrami do znalezienia genów docelowych dla zaprojektowanych amiRNA. psRNATarget znalazł listę potencjalnych sekwencji targetowych, ale gen dla którego zaprojektowano amiRNA, miał zdecydowanie lepszą punktację niż pozostałe [48].

Wyniki uzyskane z wykorzystaniem narzędzia AmiRNA Designer w fazie projektowania zostały następnie poddane analizie off-target realizowanej przez zbiór specjalnie w tym celu opracowanych skryptów. Analiza przeprowadzona dla genu AGB1 pokazała, że tylko jeden amiRNA może oddziaływać z innym genem niż docelowy oraz, że żaden amiRNA* nie był komplementarny do żadnego transkryptu pochodzącego z *Arabidopsis thaliana*. W przypadku genu Immutans, zidentyfikowano 3 dupлексы amiRNA/amiRNA*, które nie wykazywały efektu off-target. Pozostałe amiRNA były komplementarne do różnych genów hipotetycznych, domniemanych lub o nieznannej funkcji białek [48].

4.3.10. Podsumowanie

Mechanizm degradacji RNA jest podstawowym i bardzo istotnym dla funkcjonowania organizmów żywych procesem, który jest gwarantem zachowania równowagi komórkowej. Badania z nim związane obejmują zarówno studia w zakresie degradacji warunkowanej samą strukturą RNA, która pozwala lepiej zrozumieć funkcję oraz rolę jaką RNA w niej odgrywa, a która cały czas jest słabo poznana, a także degradację enzymatyczną, rozważaną w kontekście zdecydowanie bardziej złożonego układu jakim jest komórka wraz ze wszystkimi czynnikami, które są w ten proces zaangażowane. W związku z tym analiza procesu degradacji RNA stanowi bardzo obiecujący kierunek badań, a niniejsza praca wpisuje się w ten zakres tematyczny.

W niniejszej rozprawie analizowano całościowy model procesu degradacji RNA od strony biologicznej, matematycznej i informatycznej. W szczególności zaproponowano nowy model kombinatoryczny dla znanego, oryginalnego problemu RNA PDP (mapowania miejsc cięć RNA), dla którego pokazano wcześniej, że jest on silnie NP-zupełny. Zaproponowano również sformułowanie tego problemu w wersji decyzyjnej oraz przeszukiwania, które uwzględnia, oprócz błędów negatywnych także błędy pozytywne w instancji. Dla tego problemu opracowano algorytm metaheurystyczny składający się z dwóch współpracujących ze sobą heurystyk przeszukiwania tabu. Algorytm bierze pod uwagę zarówno błędy negatywne jak i pozytywne, a rozwiązania przez niego generowane są dobrej jakości, przy liczbie zrekonstruowanych cięć zbliżonych do tych, które obecne są w rozwiązaniach optymalnych. Parametry (liczba błędów oraz miejsc cięcia) wykorzystane do wygenerowania instancji testowych zostały dobrane w taki sposób, aby odzwierciedlały rzeczywiste przypadki. W związku z tym zaprezentowany algorytm będzie dobrze sprawdzał się w zastosowaniach praktycznych przy wspomaganie analizy danych pochodzących z eksperymentów biochemicznych. Dla problemu RNA PDP zaproponowany został wcześniej algorytm dokładny podziału i odcięć, który uwzględnia jedynie błędy negatywne i radzi sobie tylko z instancjami zawierającymi do 10 miejsc cięć pierwotnych. Przedstawiona metaheurystyka przeszukiwania tabu jest jedynym znanym autorom algorytmem, który umożliwia rozwiązanie tego problemu dla dużej liczby miejsc cięcia oraz przy obecności błędów pozytywnych w danych wejściowych.

Z punktu widzenia biologów poznanie wzoru degradacji danej cząsteczki RNA niesie ze sobą dużo informacji, m.in. o stabilności danego RNA oraz pozwala przewidzieć, czy może być on być źródłem stabilnych, pośrednich produktów niepełnej degradacji (degradantów RNA), które mogą pełnić funkcje np. w zakresie regulacji genów w komórce. Laboratoryjne przebadanie wzoru rozpadu dowolnej cząsteczki RNA wymuszonego jej strukturą przestrzenną jest niestety zadaniem bardzo kosztownym, zarówno pod względem czasu, jak i środków, dlatego też w pracy zaproponowano dodatkowo algorytm dokładny podziału i odcięć, który symuluje przebieg procesu nieenzymatycznej degradacji RNA. Uwzględnia on strukturę przestrzenną RNA, a także reguły stabilności zaproponowane przez Kierzka i współpracowników i pozwala przewidzieć wzór degradacji dowolnej cząsteczki RNA bez przeprowadzania czasochłonnego eksperymentu biochemicznego. W ramach algorytmu przetestowano następujące narzędzia do przewidywania struktury drugorzędowej kanonicznej: RNAfold, RNAstructure, CentroidFold, CONTRAFold, IPknot oraz ContextFold, a także zaproponowane w rozprawie nowe podejście do predykcji struktury 2D udokładnionej (uwzględniające zarówno parowania kanoniczne jak i niekanoniczne) łączące funkcjonalność istniejących narzędzi RNAComposer oraz RNApdbee i określane w pracy jako potok przetwarzania RC/Rp (zaimplementowane obecnie jako RNAvista). Wyniki działania algorytmu porównano z wynikami eksperymentu biochemicznego. Analiza uzyskanych wyników wykazała, że algorytm wykonuje swoją pracę szybko i efektywnie, nawet wtedy, gdy do wielokrotnych predykcji struktury 2D (refoldowania) wykorzystywany jest RC/Rp (RNAvista). Algorytm był w stanie zidentyfikować poprawnie większość produktów nieenzymatycznej degradacji w przypadku obu badanych cząsteczek RNA. Wszystkie miejsca cięcia, jakie nie zostały przez algorytm rozpoznane to miejsca nie-

zgodne z regułami opracowanymi przez Kierzka i współpracowników. Wyniki wskazały również, że najlepszym narzędziem do predykcji struktury 2D okazał się być RC/Rp, gdyż w przypadku obu cząsteczek jako jedyny był w stanie przewidzieć wszystkie miejsca cięcia, które były zgodne z regułami stabilności Kierzka i współpracowników. Algorytm podczas swojej pracy wielokrotnie wykonuje ponowne predykcje struktury 2D dla kolejnych pośrednich produktów degradacji uzyskiwanych na kolejnych etapach, w związku z tym poprawność oraz dokładność przewidzianej struktury 2D przekłada się na jakość uzyskanych wyników. Czas obliczeń dla RC/Rp jest co prawda dłuższy niż dla pozostałych narzędzi dedykowanych do predykcji kanonicznej struktury 2D, ale jest on i tak zdecydowanie krótszy niż dla innych programów oferujących przewidywanie udokładnionej struktury 2D RNA, a generowane przez niego wyniki charakteryzują się wysoką jakością. Dokonane w pracy porównanie jest też cenną wskazówką dla biologów w kontekście wyboru narzędzi do predykcji struktury 2D.

W rozprawie zaproponowano również model matematyczny oparty na teorii stochastycznych sieci Petriego przedstawiający przebieg procesu nieenzymatycznej degradacji RNA z uwzględnieniem struktury przestrzennej RNA, a także reguł stabilności zaproponowanych przez Kierzka i współpracowników, w oparciu o działanie powyżej wspomnianego algorytmu podziału i odcięć oraz eksperymentu biochemicznego. Model przeanalizowano w pierwszej kolejności z wykorzystaniem podejścia typu *knockout* w oparciu o zbiór t-niezmienników, co pozwoliło na określenie wpływu poszczególnych procesów reprezentowanych w modelu, na przebieg procesu nieenzymatycznej degradacji RNA. Następnie wykonano analizę symulacyjną rozważanej sieci stochastycznej. Tranzycjom, które odpowiadały cięciom pomiędzy poszczególnymi nukleotydami w łańcuchu RNA przypisano częstości uruchomień na bazie miar degradacji (przynależności) zdefiniowanych i obliczonych dla algorytmu dokładnego symulującego przebieg procesu nieenzymatycznej degradacji RNA. Wyniki przeprowadzonej symulacji stochastycznej pokazały, że zależności dotyczące podatności na hydrolizę nieenzymatyczną wiązań internukleotydowych w cząsteczce RNA zaproponowane przez Kierzka i współpracowników zostały zachowane (UA>CA>YC>YG>YU). Do oszacowania częstości uruchomień dla tranzycji reprezentujących procesy biologiczne w rozważanym modelu opartym o stochastyczne sieci Petriego wykorzystano zaproponowane w pracy nowe podejście do tego zagadnienia.

W ramach niniejszej rozprawy opracowano również model oparty na teorii stochastycznych sieci Petriego przedstawiający proces enzymatycznej degradacji RNA na przykładzie YRNA i roli wywodzących się z niego stabilnych degradantów, w regulacji apoptozy i zapalenia w monocytach/makrofagach. Wprowadzenie tego modelu pozwoliło zaakcentować złożoność zjawiska degradacji RNA zachodzącej w komórkach, w ramach której uwzględnić należy wszystkie czynniki zaangażowane w analizowany proces. Przedstawiony system stanowi jedynie niewielki wycinek mechanizmów w jakich biorą udział cząsteczki pośrednich produktów degradacji RNA oraz ich wpływu na przebieg różnych chorób. Model przeanalizowano podobnie jak opisano to powyżej, wykorzystując podejście typu *knockout* w oparciu o zbiór t-niezmienników oraz *knockout* symulacyjny. Pozwoliło to na wyciągnięcie bardzo interesujących wniosków biologicznych.

Badania cząstkowe z wykorzystaniem wspomnianych wcześniej algorytmów posłużyły do zbudowania opisanych powyżej złożonych modeli obrazujących przebieg zarówno degradacji enzymatycznej jak i nieenzymatycznej RNA oraz umożliwiły finalnie opracowanie nowego podejścia do projektowania sztucznych regulatorowych cząsteczek RNA na przykładzie miRNA (mikroRNA). Zaproponowano nowe podejście do projektowania sztucznych miRNA ukierunkowanych na określony gen, który będzie wyciszony (wyłączony), bazując na właściwościach naturalnych miRNA. Zostało ono zaimplementowane jako narzędzie o nazwie AmiRNA Designer. Jego potencjalne zastosowania obejmują m.in. genomikę funkcjonalną czy też inżynierię genetyczną. W przeciwieństwie do istniejącego oprogramowania AmiRNA Designer jest bardzo elastyczny i pozwala na: przeanalizowanie naturalnych miRNA dla dowolnie wybranego organizmu, stworzenie profili termodynamicznych miRNA/miRNA* oraz miRNA/tar-get na podstawie danych termodynamicznych wygenerowanych przez UNAFold podczas predykcji

struktury drugorzędowej oraz stworzenie własnych reguł filtracji na bazie sekwencji w części dotyczącej projektowania sztucznego miRNA. Pozwala także na filtrowanie wyników amiRNA (analiza off-target) czyli sprawdzenie czy zaprojektowany amiRNA nie wchodzi w interakcję z innymi przypadkowymi mRNA danego organizmu, powodując ich wyłączenie.

W zakresie przyszłych kierunków badawczych dotyczących procesu degradacji RNA można wymienić:

- Opracowanie algorytmu symulującego degradację RNA z uwzględnieniem reguł stabilności podanych przez Kierzka i współpracowników oraz jego struktury przestrzennej, a także czynników komórkowych mających na nią wpływ, np. białek wiążących RNA i chroniących go przed degradacją, miejsc cięcia rozpoznawanych przez rybonukleazy. Wymagałoby to dodatkowo opracowania bazy danych zbierających informacje o motywach strukturalnych rozpoznawanych w RNA przez enzymy tnące oraz białka chroniące je przed cięciem.
- Opracowanie podejścia pozwalającego odróżnić produkty pośrednie degradacji, które mogą być funkcjonalne od niefunkcjonalnych.
- Opracowanie modeli opartych o teorię sieci Petriego przedstawiających proces enzymatycznej degradacji RNA dla innych, znanych stabilnych produktów pośrednich degradacji RNA i ich wpływ na przebieg różnych chorób.
- Opracowanie modelu opartego o teorię sieci Petriego obrazującego ścieżki powstawania stabilnych produktów degradacji z dojrzałych funkcjonalnych RNA takich jak: tRNA, rRNA, snRNA, snoRNA czy też mRNA.
- Sformułowanie problemu mapowania miejsc cięć RNA (RNA PDP) w sytuacji, gdy dozwolone byłyby cięcia wtórne dalszych rzędów (nie tylko w ramach fragmentów pierwotnych, ale też wtórnych) oraz opracowanie algorytmu dokładnego i heurystycznego dla takiego przypadku.

Rozwiązanie powyższych problemów oraz zrealizowanie zaproponowanych zadań miałyby na pewno znaczący wkład w badania, których celem jest lepsze poznanie oraz zrozumienie mechanizmów rządzących procesem degradacji RNA oraz w zakresie wykorzystania metod matematycznych oraz informatycznych w biologii molekularnej, genomice funkcjonalnej oraz medycynie.

5. Omówienie pozostałych osiągnięć naukowo-badawczych

5.1. Lista prac – pozostałe osiągnięcia

[O1] M. Antczak, M. Zablocki, T. Zok, A. Rybarczyk, J. Blazewicz, M. Szachniuk, RNAvista: a webserver to assess RNA secondary structures with non-canonical base pairs, *Bioinformatics*, 2018, doi: 10.1093/bioinformatics/bty609.

[O2] D. Formanowicz, M. Radom, A. Rybarczyk, P. Formanowicz, The role of Fenton reaction in ROS-induced toxicity underlying atherosclerosis – modeled and analyzed using a Petri net-based approach, *Biosystems*, 165, pp. 71-87, 2018, doi: 10.1016/j.biosystems.2018.01.002.

[O3] D. Formanowicz, A. Rybarczyk, P. Formanowicz, Factors influencing essential hypertension and cardiovascular disease modeled and analyzed using stochastic Petri nets, *Fundamenta Informaticae*, 160 (1-2), pp. 143-165, 2018, doi: 10.3233/FI-2018-1678.

[O4] M. Radom, A. Rybarczyk, B. Szawulak, H. Andrzejewski, P. Chabelski, A. Kozak, P. Formanowicz, Holmes: a graphical tool for development, simulation and analysis of Petri net based models of complex biological systems, *Bioinformatics*, 33(23), pp. 3822-3823, 2017, doi: 10.1093/bioinformatics/btx492.

- [O5] A. Rybarczyk, A. Hertz, M. Kasprzak, J. Blazewicz, Tabu Search for the RNA Partial Degradation Problem, *International Journal of Applied Mathematics Computational Science*, 27(2), pp. 401-415, 2017, doi: 10.1515/amcs-2017-0028.
- [O6] A. Mickiewicz, J. Sarzyńska, M. Miłostan, A. Kurzyńska-Kokorniak, A. Rybarczyk, P. Łukasiak, T. Kuliński, M. Figlerowicz, J. Błażewicz, Modeling of the catalytic core of Arabidopsis thaliana Dicer-like 4 protein and its complex with double-stranded RNA, *Computational Biology and Chemistry*, 66, pp. 44-56, 2017, doi: 10.1016/j.compbiolchem.2016.11.003.
- [O7] A. Rybarczyk, P. Jackowiak, M. Figlerowicz, J. Blazewicz, Computational prediction of non-enzymatic RNA degradation patterns, *Acta Biochimica Polonica*, 63(4), pp. 745-751, 2016, doi: 10.18388/abp.2016_1331.
- [O8] A. Hojka-Osinska, L. Budzko, A. Zmienko, A. Rybarczyk, P. Maillard, A. Budkowska, M. Figlerowicz, P. Jackowiak, RNA-Seq-based analysis of differential gene expression associated with hepatitis C virus infection in a cell culture, *Acta Biochimica Polonica*, 63(4), pp. 789-798, 2016, doi: 10.18388/abp.2016_1343.
- [O9] A. Mickiewicz, A. Rybarczyk, J. Sarzynska, M. Figlerowicz, J. Blazewicz, AmiRNA Designer - new method of artificial miRNA design, *Acta Biochimica Polonica*, 63, pp. 71-77, 2016, doi: 10.18388/abp.2015_989.
- [O10] A. Rybarczyk, N. Szostak, M. Antczak, T. Zok, M. Popenda, R.W. Adamiak, J. Blazewicz, M. Szachniuk, New in silico approach to assess RNA secondary structures with non-canonical base pairs, *BMC Bioinformatics*, 16, pp. 276, 2015, doi: 10.1186/s12859-015-0718-6.
- [O11] N. Szostak, F. Royo, A. Rybarczyk, M. Szachniuk, J. Blazewicz, A. del Sol, J.M. Falcon-Perez, Sorting signal targeting mRNA into hepatic extracellular vesicles, *RNA Biology*, 11(7), pp. 836-844, 2014, doi: 10.4161/rna.29305.
- [O12] M. Radom, A. Rybarczyk, R. Kottmann, P. Formanowicz, M. Szachniuk, F.O. Gloeckner, D. Rebholz-Schuhmann, J. Blazewicz, Poseidon: information retrieval and extraction system for meta-genomic marine science, *Ecological Informatics*, 12, pp. 10-15, 2012, doi: 10.1016/j.ecoinf.2012.07.003.
- [O13] M. Nowacka, P. Jackowiak, A. Rybarczyk, T. Magacz, P.M. Strozycycki, J. Barciszewski, M. Figlerowicz, 2D-PAGE as an effective method of RNA degradome analysis, *Molecular Biology Reports*, 39(1), pp. 139-146, 2012, doi: 10.1007/s11033-011-0718-1.
- [O14] J. Blazewicz, M. Figlerowicz, M. Kasprzak, M. Nowacka, A. Rybarczyk, RNA Partial Degradation Problem: motivation, complexity, algorithm, *Journal of Computational Biology*, 18(6), pp. 821-834, 2011, doi: 10.1089/cmb.2010.0153.
- [O15] J. Podkowinski, A. Zmienko, B. Florek, P. Wojciechowski, A. Rybarczyk, J. Wrzesinski, J. Ciesiolka, J. Blazewicz, A. Kondorosi, M. Crespi, A. Legocki, Translational and structural analysis of the shortest legume enod40 gene in *Lupinus luteus*, *Acta Biochimica Polonica*, 56(1), pp. 89-102, 2009.
- [O16] K. Jankowiak, A. Rybarczyk, R. Wyatt, I. Odrzykoski, A. Pacak, Z. Szweykowska-Kulinska, Organellar inheritance in the allopolyploid moss *Rhizomnium pseudopunctatum*, *Taxon*, 54(2), pp. 383-388, 2005.
- [O17] K. Jankowiak, J. Lesicka, A. Pacak, A. Rybarczyk, Z. Szweykowska-Kulinska, A comparison of group II introns of plastid tRNALys UUU genes encoding maturase protein, *Cellular and Molecular Biology Letters*, 9(2), pp. 239-251, 2004.

5.2. Opis pozostałych osiągnięć

Wiele z wymienionych w poprzedniej sekcji prac związana jest z podstawową tematyką badawczą opisaną w monografii oraz przedstawioną pokrótce w ramach sekcji 4.3. Pozostałe publikacje plasują się tematycznie w obszarze szeroko rozumianej bioinformatyki, biologii obliczeniowej oraz biologii molekularnej.

Analiza procesu degradacji RNA jest obecnie bardzo rozwijającym kierunkiem badań i powiązana jest bezpośrednio z niniejszą monografią. Dotychczasowe doniesienia wskazują, że stabilne produkty przejściowe degradacji RNA (tzw. fragmenty RNA, degradanty RNA) pełnią istotne funkcje regulacyjne w komórkach organizmów żywych, a także mogą być cennym źródłem biomarkerów. Analiza degradantów RNA jest niezwykle skomplikowana zarówno w kontekście badań laboratoryjnych jak i informatycznych. W pracy [O13] przedstawiono efektywną metodę analizy produktów przejściowych degradacji RNA techniką dwukierunkowej elektroforezy w żelu poliakryloamidowym (2D-PAGE). Natomiast w [O8] wykorzystano technikę analityczną związaną z technologią wysokoprzepustowego sekwencjonowania (NGS, ang. *next generation sequencing*), dokładniej RNA-seq, pozwalającą na identyfikowanie (na wysoką skalę) milionów cząsteczek RNA, która pozwoliła na wykrycie i przeanalizowanie tych procesów i szlaków komórkowych, które uległy zaburzeniu w wyniku infekcji wirusem HCV. W pracy dokonano porównania transkryptomów pochodzących z komórek zakażonych wirusem i zdrowych w celu określenia zmian w zakresie poziomu ekspresji poszczególnych genów.

Bardzo istotnym elementem rozpatrywanym w kontekście procesu degradacji RNA jest stabilność struktury RNA, która bezpośrednio wiąże się m.in. z przewidywaniem aktywności oraz czasu przebywania danej cząsteczki RNA w komórce. W celu przeanalizowania wpływu struktury RNA na proces jego degradacji oraz uzyskania wzoru degradacji zadanej cząsteczki RNA przeprowadzono eksperyment biochemiczny opisany w [O14]. W pracy sformułowano również po raz pierwszy problem rekonstrukcji położenia miejsc cięcia (pęknięć) dla zadanej cząsteczki RNA na podstawie eksperymentu biochemicznego (tylko na bazie długości, bez znajomości sekwencji - problem RNA PDP), zbadano jego złożoność obliczeniową oraz zaproponowano algorytm dokładny podziału i odcięć (ang. *branch-and-cut*) rozwiązujący ten problem.

W publikacji [O6] zaproponowano model roślinnej rybonukleazy DCL-4 (AtDCL-4) pochodzącej z *Arabidopsis thaliana*, w kompleksie z dwuniciowym RNA (dsRNA). Wykorzystując metody komputerowe zamodelowano rdzeń katalityczny DCL-4 obejmujący konserwatywną domenę PAZ wiążącą dsRNA, domenę Platformy, konektor oraz dwie domeny typu RNazy III o aktywności endonukleolitycznej na bazie białka Dicer z *Giardia*. DCL-4 pełni kluczową rolę w procesie biogenezy małych regulatorowych RNA, dokładniej siRNA.

Natomiast w [O11] badano motywy kierujące mRNA do egzosomów (ang. *extracellular vesicles, EV*) występujących w wątrobie. Przy wykorzystaniu metod bioinformatycznych zidentyfikowano 12 takich motywów, z czego 10 było zlokalizowanych w sekwencjach kodujących transkryptów, a 2 w ich rejonach 3' UTR. Określono również miRNA, które potencjalnie mogły się z nimi wiązać. Wykonano eksperyment biochemiczny, w ramach którego dla jednego wybranego motywu (obecnego w rejonie 3' UTR) potwierdzono, że kieruje on, posiadające go mRNA do egzosomów.

W publikacji [O15] korzystając z metod bioinformatycznych oraz eksperymentalnych, zaproponowano strukturę drugorzędową oraz zidentyfikowano cztery konserwatywne domeny dla transkryptu genu ENOD40-1 pochodzącego z *Lupinus luteus*, uczestniczącego w regulacji podziałów komórkowych towarzyszących nodulacji. Uzyskaną strukturę 2D dla *L. luteus* porównano ze strukturą drugorzędową opublikowaną dla transkryptu ENOD40-1 dla *Glycine max*, wskazując, że może ona przyjmować dwie postaci, zależne od warunków panujących w komórce i pełnić w zależności od tego różne funkcje.

W ramach swojej pracy badawczej habilitantka zajmowała się również wykorzystaniem sieci Petriego do modelowania i analizy złożonych systemów biologicznych. Publikacja [O4] dotyczy nowego oprogramowania Holmes dedykowanego do tego celu. Umożliwia ono tworzenie modeli klasycznych oraz czasowych sieci Petriego, ich rozbudowaną analizę oraz bogatą wizualizację wyników generowanych przez różne moduły analityczne. Oprogramowanie to, zostało wykorzystane przy budowie oraz analizie modeli opublikowanych w [O2], [O3], a także zaproponowanych w ramach niniejszej monografii.

W ramach realizacji europejskiego projektu METAFUNCTIONS opracowano system o nazwie Poseidon wspomagający analizę oraz ekstrakcję danych dotyczących metagenomów morskich, z publikacji naukowych. Dokonano implementacji metody służącej do automatycznej, wstępnej klasyfikacji artykułów, zaprojektowano odpowiednie słowniki wykorzystywane przez program Whatizit (wykorzystywany przez system Poseidon), interfejsu www systemu Poseidon, statystyk, systemu pomocy dla kuratora systemu oraz komunikacji z bazą danych [O12].

Habilitantka współpracowała również z zespołem kierowanym przez prof. dr hab. Zofię Szwejkowską-Kulińską z Instytutu Biologii Molekularnej i Biotechnologii UAM. W ramach swojej pracy magisterskiej na kierunku biologia habilitantka wykonywała analizę informatyczną oraz badania laboratoryjne. Celem pracy było określenie sposobu dziedziczenia organelli u mszaków na przykładzie *Rhizomnium pseudopunctatum*. Wyniki badań zostały przedstawione w następujących publikacjach: [O16] i [O17].

Referencje

- [1] J. Wang, D. C. Samuels, S. Zhao, Y. Xiang, Y. Y. Zhao, i Y. Guo, „Current Research on Non-Coding Ribonucleic Acid (RNA)”, *Genes*, t. 8, nr 12, s. E366, 2017.
- [2] P. Jackowiak, M. Nowacka, P. M. Strozycki, i M. Figlerowicz, „RNA degradome—its biogenesis and functions”, *Nucleic Acids Res.*, t. 39, nr 17, s. 7361–7370, 2011.
- [3] Z. Li, C. Ender, G. Meister, P. S. Moore, Y. Chang, i B. John, „Extensive terminal and asymmetric processing of small RNAs from rRNAs, snoRNAs, snRNAs, and tRNAs.”, *Nucleic Acids Res.*, t. 40, nr 14, s. 6787–6799, 2012.
- [4] M. Zywicki, K. Bakowska-Zywicka, i N. Polacek, „Revealing stable processing products from ribosome-associated small RNAs by deep-sequencing data analysis.”, *Nucleic Acids Res.*, t. 40, nr 9, s. 4013–4024, 2012.
- [5] M. Kasprzyk, A. M. Mleczko, P. Celichowski, i K. Bakowska-Zywicka, „Przetwarzanie RNA - niezwykły mechanizm powstawania nowych klas niekodujących RNA z funkcjonalnych RNA”, *Postepy Biochem.*, t. 60, nr 3, s. 295–304, 2014.
- [6] A. Bibillo, M. Figlerowicz, i R. Kierzek, „The non-enzymatic hydrolysis of oligoribonucleotides. VI. The role of biogenic polyamines”, *Nucleic Acids Res.*, t. 27, s. 3931–3937, 1999.
- [7] A. Bibillo, M. Figlerowicz, K. Ziomek, i R. Kierzek, „The nonenzymatic hydrolysis of oligoribonucleotides. VII. Structural elements affecting hydrolysis”, *Nucleosides Nucleotides Nucleic Acids*, t. 19, s. 977–994, 2000.
- [8] R. Kierzek, „Hydrolysis of oligoribonucleotides: influence of sequence and length”, *Nucleic Acids Res.*, t. 20, s. 5073–5077, 1992.
- [9] R. Kierzek, „Nonenzymatic Cleavage of Oligoribonucleotides”, *Methods Enzymol.*, t. 341, s. 657–675, 2001.
- [10] J. Houseley i D. Tollervey, „The Many Pathways of RNA Degradation.”, *Cell*, t. 136, nr 4, s. 763–776, 2009.
- [11] M. P. Deutscher, „Degradation of Stable RNA in Bacteria.”, *J. Biol. Chem.*, t. 278, nr 46, s. 45041–45044, 2003.

- [12] Z. Hizir, S. Bottini, V. Grandjean, M. Trabucchi, i E. Repetto, „RNY (YRNA)-derived small RNAs regulate cell death and inflammation in monocytes/macrophages.”, *Cell Death Dis.*, t. 8, nr 1, s. e2530, 2017.
- [13] C. Cole *i in.*, „Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs.”, *RNA*, t. 15, nr 12, s. 2147–2160, 2009.
- [14] P. Kumar, S. B. Mudunuri, J. Anaya, i A. Dutta, „tRFdb: a database for transfer RNA fragments.”, *Nucleic Acids Res.*, t. 43, s. D141–D145, 2015.
- [15] J. Blazewicz, M. Figlerowicz, M. Kasprzak, M. Nowacka, i A. Rybarczyk, „RNA Partial Degradation Problem: motivation, complexity, algorithm.”, *J. Comput. Biol.*, t. 18, nr 6, s. 821–834, 2011.
- [16] A. Rybarczyk, *Algorytmiczne aspekty procesu degradacji RNA*. Wydawnictwo Nakom, 2010.
- [17] A. Rybarczyk, A. Hertz, M. Kasprzak, i J. Blazewicz, „Tabu search for the RNA partial degradation problem.”, *Int. J. Appl. Math. Comput. Sci.*, t. 27, nr 2, s. 401–415, 2017.
- [18] M. R. Garey i D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*. San Francisco: W.H. Freeman and Company, 1997.
- [19] A. Rybarczyk, P. Jackowiak, M. Figlerowicz, i J. Blazewicz, „Computational prediction of non-enzymatic RNA degradation patterns”, *Acta Biochim. Pol.*, t. 63, nr 4, s. 745–751, 2016.
- [20] N. B. Leontis i E. Westhof, „Geometric nomenclature and classification of RNA base pairs.”, *RNA*, t. 7, nr 4, s. 499–512, 2001.
- [21] M. Parisien i F. Major, „The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data.”, *Nature*, t. 452, nr 7183, s. 51–55, 2008.
- [22] H. Yang *i in.*, „Tools for the automatic identification and classification of RNA base pairs”, *Nucleic Acids Res.*, t. 31, s. 3450–3460, 2003.
- [23] P. Gendron, S. Lemieux, i F. Major, „Quantitative analysis of nucleic acid three-dimensional structures”, *J Mol Biol*, t. 308, s. 919–936, 2001.
- [24] X. J. Lu i W. K. Olson, „3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures”, *Nucleic Acids Res.*, t. 31, s. 5108–5121, 2003.
- [25] M. Antczak *i in.*, „RNAPdbee—a webserver to derive secondary structures from pdb files of knotted and unknotted RNAs”, *Nucleic Acids Res.*, t. 42, s. W368–W372, 2014.
- [26] W. Saenger, *Principles of Nucleic Acid Structure*. Springer New York, 1984.
- [27] R. Das, J. Karanicolas, i D. Baker, „Atomic accuracy in predicting and designing noncanonical RNA structure.”, *Nat. Methods*, t. 7, nr 4, s. 291–294, 2010.
- [28] N. V. Dokholyan, S. V. Buldyrev, H. E. Stanley, i E. I. Shakhnovich, „Discrete molecular dynamics studies of the folding of a protein-like model.”, *Fold. Des.*, t. 3, nr 6, s. 577–587, 1998.
- [29] M. Popena *i in.*, „Automated 3D structure composition for large RNAs”, *Nucleic Acids Res.*, t. 40, s. e112–e112, 2012.
- [30] M. Antczak, M. Zablocki, T. Zok, A. Rybarczyk, J. Blazewicz, i M. Szachniuk, „RNAvista: a web-server to assess RNA secondary structures with non-canonical base pairs”, *Bioinformatics*, 2018.
- [31] M. Andronescu, V. Bereg, H. H. Hoos, i A. Condon, „RNA STRAND: the RNA secondary structure and statistical analysis database”, *BMC Bioinformatics*, t. 13, s. 340, 2008.
- [32] A. I. Petrov, C. L. Zirbel, i N. B. Leontis, „Automated classification of RNA 3D motifs and the RNA 3D Motif Atlas”, *RNA*, t. 19, s. 1327–1340, 2013.
- [33] K. Sato, M. Hamada, K. Asai, i T. Mituyama, „CENTROIDFOLD: a web server for RNA secondary structure prediction”, *Nucleic Acids Res.*, t. 37, s. W277–W280, 2009.
- [34] G. Taguchi, *System of Experimental Design: Engineering Methods to Optimize Quality and Minimize Costs*. UNIPUB/Kraus International Publications, 1987.
- [35] J. C. Zhang i M. A. Styblinski, *Yield and Variability Optimization of Integrated Circuits*. Springer Science, New York, 1995.
- [36] J. S. Reuter i D. H. Mathews, „RNAstructure: software for RNA secondary structure prediction and analysis”, *BMC Bioinformatics*, t. 5, s. 129, 2010.
- [37] D. Formanowicz, A. Rybarczyk, i P. Formanowicz, „Factors influencing essential hypertension and cardiovascular disease modeled and analyzed using stochastic Petri nets”, *Fundam. Informaticae*, t. 160, s. 143–165, 2018.

- [38] Marcin Radom, „Holmes: a graphical tool for development, simulation and analysis of Petri net based models of complex biological systems”, *Bioinformatics*, accepted for publication 2017.
- [39] J. Einloft J. an. Ackermann, J. Nöthen, i I. Koch, „Monalisa – visualization and analysis of functional modules in biochemical networks”, *Bioinformatics*, t. 29, nr 11, s. 1469–1470, 2013.
- [40] D. T. Gillespie, „Exact stochastic simulation of coupled chemical reactions”, *J. Phys. Chem.*, t. 81, s. 2340–2361, 1977.
- [41] M. Heiner, S. Lehrack, D. Gilbert, i W. Marwan, „Extended Stochastic Petri Nets for Model-Based Design of Wetlab Experiments”, *Trans. Comput. Syst. Biol. XI*, t. 5750, s. 138–163, 2009.
- [42] Y. Liu, P. K. Li, C. Li, i J. Lin, „Inhibition of STAT3 signaling blocks the anti-apoptotic activity of IL-6 in human liver cancer cells”, *J. Biol. Chem.*, t. 285, nr 35, s. 27429–27439, 2010.
- [43] J. O. Jin, X. Han, i Q. Yu, „Interleukin-6 induces the generation of IL-10-producing Tr1 cells and suppresses autoimmune tissue inflammation”, *J. Autoimmun.*, t. 40, s. 28–44, 2013.
- [44] T. Yoshimoto i T. Yoshimoto, *Cytokine Frontiers: Regulation of Immune Responses in Health and Disease*. Springer, 2014.
- [45] R. S. Hundal, B. S. Salh, J. W. Schrader, A. Gomez-Munoz, V. Duronio, i U. P. Steinbrecher, „Oxidized low density lipoprotein inhibits macrophage apoptosis through activation of the PI 3-kinase/PKB pathway”, *J. Lipid Res.*, t. 42, s. 1483–1491, 2001.
- [46] M. Riazy, J. H. Chen, Y. Yamamoto, H. Yamamoto, V. Duronio, i U. P. Steinbrecher, „OxLDL-mediated survival of macrophages does not require LDL internalization or signalling by major pattern recognition receptors”, *Biochem. Cell Biol.*, t. 89, nr 4, s. 387–395, 2011.
- [47] N. G. Bologna *i in.*, „Multiple RNA recognition patterns during microRNA biogenesis in plants.”, *Genome Res.*, t. 23, s. 1675–1689, 2013.
- [48] A. Mickiewicz, A. Rybarczyk, J. Sarzynska, M. Figlerowicz, i J. Blazewicz, „AmiRNA Designer - new method of artificial miRNA design.”, *Acta Biochim. Pol.*, t. 63, nr 1, s. 71–77, 2016.
- [49] P. Brodersen i O. Voinnet, „Revisiting the principles of microRNA target recognition and mode of action.”, *Nat. Rev. Mol. Cell Biol.*, t. 10, s. 141–148, 2009.
- [50] S. Petri i G. Meister, „siRNA design principles and off-target effects.”, *Methods Mol. Biol.*, t. 986, s. 59–71, 2013.
- [51] R. Schwab, S. Ossowski, M. Riester, N. Warthmann, i D. Weigel, „Highly specific gene silencing by artificial microRNAs in Arabidopsis.”, *Plant Cell*, t. 18, nr 5, s. 1121–1133, 2006.
- [52] S. Yu i G. Pilot, „Testing the efficiency of plant artificial microRNAs by transient expression in *Nicotiana benthamiana* reveals additional action at the translational level.”, *Front. Plant Sci.*, t. 5, s. 622, 2014.
- [53] A. Khvorova, A. Reynolds, i S. D. Jayasena, „Functional siRNAs and miRNAs exhibit strand bias.”, *Cell*, t. 115, s. 209–216, 2003.
- [54] J. Krol *i in.*, „Structural features of microRNA (miRNA) precursors and their relevance to miRNA biogenesis and small interfering RNA/short hairpin RNA design.”, *J. Biol. Chem.*, t. 279, s. 42230–42239, 2004.
- [55] N. R. Markham i M. Zuker, „UNAFold: software for nucleic acid folding and hybridization.”, *Bioinformatics*, t. 453, s. 3–31, 2008.
- [56] S. Griffiths-Jones, H. K. Saini, S. van Dongen, i A. J. Enright, „miRBase: tools for microRNA genomics.”, *Nucleic Acids Res.*, t. 36, s. D154–D158, 2008.
- [57] A. M. Gustafson, E. Allen, S. Givan, D. Smith, J. C. Carrington, i K. D. Kasschau, „ASRP: the Arabidopsis small RNA project database.”, *Nucleic Acids Res.*, t. 33, s. D637–D640, 2005.
- [58] Q. Dong, S. D. Schlueter, i V. Brendel, „PlantGDB, plant genome database and analysis tools.”, *Nucleic Acids Res.*, t. 32, s. D354–D359, 2004.
- [59] X. Dai i P. X. Zhao, „psRNATarget: a plant small RNA target analysis server.”, *Nucleic Acids Res.*, t. 39, s. W155–W159, 2011.


Agnieszka Rybarczyk
(podpis)