



UNIwersytet
Warszawski

CeNT CENTRUM
NOWYCH
TECHNOLOGII

2 września, 2021

Prof. dr hab. Joanna Trylska
e-mail: joanna@cent.uw.edu.pl
telefon (22) 55 43 683
<https://bionano.cent.uw.edu.pl>

Rada Naukowa Instytutu Chemii Bioorganicznej
Polskiej Akademii Nauk
Ul. Z. Noskowskiego 12/14
61-704 Poznań

Recenzja rozprawy doktorskiej mgr Agnieszki Mickiewicz-Piotrowskiej

Przedstawiona mi do recenzji rozprawa doktorska mgr Agnieszki Mickiewicz-Piotrowskiej zatytułowana „Analiza strukturalnych uwarunkowań wybranych etapów biogenezy małych regulatorowych RNA u *Arabidopsis thaliana*” została wykonana w Zakładzie Bioinformatyki Strukturalnej Instytutu Chemii Bioorganicznej Polskiej Akademii Nauk pod kierunkiem prof. dr hab. inż. Marty Szachniuk oraz promotora pomocniczego dr Joanny Sarzyńskiej.

Temat badawczy mgr Mickiewicz-Piotrowskiej dotyczył regulacji ekspresji genów na poziomie translacji za pomocą krótkich regulatorowych RNA. RNA zyskało w czasie obecnej pandemii ogromną popularność ze względu na jego użycie w szczepionkach. Dekady temu badano głównie mRNA oraz rybosomowe RNA ze względu na to, że bezpośrednio uczestniczą w translacji, czyli ostatnim etapie produkcji białek w komórkach. Jednak od wielu lat wiadomo, że niekodujące fragmenty RNA też mają ogromne znaczenie w komórce, w tym funkcję regulatorową, którą zajmowała się autorka rozprawy. Takie krótkie fragmenty RNA, zwane małymi regulatorowymi RNA (ang. *short regulatory RNA*, srRNA), oddziałujące komplementarnie z mRNA mogą regulować poziom translacji. Utworzenie komplementarnego dupleksu krótkiego RNA z fragmentem mRNA powoduje blokowanie właściwego oddziaływania mRNA z rybosomem lub prowadzi do degradacji mRNA. Jest to jednak ostatni etap bardziej skomplikowanego procesu, gdyż uruchomienie takiego potranskrypcyjnego wyciszania ekspresji genów wymaga oprócz RNA także innych białek. Małe regulatorowe RNA najpierw muszą zostać wycięte jako dupleksy, za co odpowiada rybonukleaza o nazwie Dicer, a następnie tworzą kompleks z innym białkiem o nazwie Argonaute (AGO), aby dupleks został rozpleciony i białko AGO „skierowało” jedną z nici w kierunku komplementarnego oddziaływania z odpowiednim fragmentem mRNA jako celem. Poznanie strukturalnych aspektów tego procesu jest istotne ze względu na zastosowanie biotechnologiczne, a także projektowanie leków, więc tematyka badawcza mgr. Mickiewicz-Piotrowskiej jest ważna i bardzo aktualna.

Rozprawa doktorska mgr Agnieszki Mickiewicz-Piotrowskiej została napisana w języku polskim i jest napisana poprawnie językowo. Wraz z załącznikami zawiera 150 stron. Rozprawa składa się z pięciu głównych rozdziałów, pięciu załączników oraz obszernej bibliografii, w której autorka odwołuje czytelnika do około 250 pozycji literaturowych.

W rozdziale 1 zatytułowanym Wstęp autorka opisuje podstawy i historię odkrycia zjawiska interferencji RNA. Ten sposób wyciszania genów odbywa się poprzez małe regulatorowe RNA. W tym rozdziale jest opisany mechanizm działania regulatorowych RNA, głównie u roślin. Autorka rozprawy podaje przykłady różnych regulatorowych RNA (np., microRNA, siRNA), a także opisuje białka, które uczestniczą w biogenezie roślinnych krótkich regulatorowych RNA (m. in., RNazy, białka typu Dicer i AGO). Autorka porusza także temat sztucznych microRNA (miRNA), które projektuje się w celu skutecznego wyciszenia konkretnego mRNA, ale na bazie naturalnego mechanizmu naturalnego miRNA. Sztuczne miRNA (w skrócie amiRNA) są istotnym narzędziem do badania ekspresji genów, zwłaszcza u roślin.

Rozdział 2 zawiera opis celu pracy. Celem pracy autorki była charakterystyka strukturalna biogenezy małych regulatorowych RNA dla białek i RNA roślinnych, na przykładzie *Arabidopsis thaliana*. Badania zawierały charakterystykę stabilności dupleksów zawierających mikroRNA, stworzenie narzędzia do projektowania sekwencji miRNA (AmiRNA Designer) służących do wyciszania konkretnych genów, porównanie sekwencji białek roślinnych typu Dicer i stworzenie modelu struktury trójwymiarowej białka roślinnego DCL4 w kompleksie z dupleksem RNA. Cele pracy zostały zrealizowane.

W rozdziale 3 autorka opisuje metody oraz oprogramowanie użyte do badań. W tej części wymienione są bazy danych, z których pobierano sekwencje RNA czy białek oraz struktury białek wraz z ich oznaczeniami w bazie Protein Data Bank. Wymieniono także wszystkie programy, które były wykorzystane w badaniach. Dalej autorka opisała metody określania drugorzędowej struktury RNA, typu model najbliższego sąsiedztwa (ang. *nearest neighbour model*) oraz metody stosowane do porównywania i uliniawiania sekwencji aminokwasowych. Opisała także metody i programy, którymi posługiwała się do generowania trójwymiarowych struktur białek i ich kompleksów z RNA. Dalej oceniała modele i wykorzystywała także technikę dynamiki molekularnej do zweryfikowania stabilności otrzymanych struktur.

W rozdziale 4 opisane są i przedyskutowane wyniki badań. Rozdział ten został podzielony na dwie części; jedna część dotyczy stworzonego przez autorkę oprogramowania amiRNA Design służącego do projektowania sekwencji miRNA, a druga dotyczy modelowania enzymu typu Dicer (o nazwie DCL4) oraz jego kompleksu z dupleksem miRNA/miRNA*.

W oprogramowaniu stworzonym przez autorkę rozprawy „sztuczne” sekwencje miRNA są projektowane na podstawie natywnych miRNA, ich prekursorów (czyli spinek RNA) i fragmentów mRNA, czyli celów z którymi oddziałują w komórce. Oprócz zaprojektowania sekwencji, które najlepiej oddziaływałyby z celem, sprawdzane są też niespecyficzne efekty oddziaływania tzw. *off-target effects*. Narzędzie AmiRNA Design jest darmowo dostępne dla użytkowników do pobrania i instalacji ze strony internetowej. Narzędzie to wykorzystuje zewnętrzny program UNAFold, do przewidywania struktury drugorzędowej dupleksów RNA. W podrozdziale 4.1 autorka opisuje krok po kroku etapy działania narzędzia amiRNA Designer przedstawiając też fragmenty przykładowych

plików wynikowych. Przykłady dotyczyły sekwencji *A. thaliana*, ale użytkownik może też stosować dane z innych organizmów. Dopuszczalne i często wskazane jest też stosowanie par GU i UG w duplesie. Generalnie narzędzie to bazuje na wprowadzonych danych sekwencyjnych, określaniu struktur drugorzędowych i szacowaniu profili energetycznych. Wynikiem programu jest lista sekwencji amiRNA, które dodatkowo zostały sprawdzone pod kątem tego, czy nie będą zdolne do wyciszenia też innego genu niż zadany. Określanie energii swobodnej struktur RNA dla par zasad zostało zaimplementowane na podstawie programu UNAFold. Jednak dla niesparowanych nieregularnych fragmentów opracowano własny sposób szacowania energii swobodnej oddziaływania w zależności od długości niesparowania, charakteru nieregularności czy sposobu wyrzucenia. W dalszych częściach tego rozdziału przedstawione są wyniki użycia AmiRNA Designer do zaprojektowania sztucznego miRNA dla dwóch wybranych docelowych genów. Następnie opisany jest plan eksperymentu, który należy przeprowadzić z użyciem zaprojektowanych sekwencji, aby zweryfikować działanie programu.

Druga część wyników opisuje sposób wyznaczenia struktury trójwymiarowej istotnych części wielodomenowego enzymu AtDCL4 w kompleksie z dupleksem RNA. AtDCL4 to roślinne białko typu Dicer z rośliny *A. thaliana*, o budowie domenowej, które wycina 21-nukleotydowe dupлексы RNA, czyli cząsteczki miRNA/miRNA*. Aby wyznaczyć strukturę AtDCL4 autorka rozprawy najpierw przeprowadziła analizę porównawczą sekwencji białek roślinnych typu Dicer. Ze względu na niskie podobieństwo sekwencji całych białek poszukiwała też szablonów strukturalnych do modelowania poszczególnych domen przez analizę porównawczą z białkami innych organizmów. W rezultacie modelowała pięć domen tego białka, na podstawie najlepszych wybranych szablonów z różnych organizmów. Co ważne zidentyfikowała też prawdopodobny region białka nieuporządkowany strukturalnie, który przyjmuje strukturę uporządkowaną dopiero po związaniu z RNA. Z powodu wielu domen i ich połączeń oraz niewielkich informacji strukturalnych dotyczących tej rodziny białek wiele etapów modelowania musiało być wykonanych ręcznie, na podstawie prób i błędów, a następnie udokładnianych. Dodatkowo, istotne było to, żeby dobrze przewidzieć strukturę miejsca katalitycznego w kompleksie z dupleksem RNA. Fragmenty łączące domeny, np. nieustrukturyzowane pętle, także stanowiły wyzwanie, gdyż są to fragmenty białka wysoce dynamiczne. Aby wyznaczyć prawdopodobne ułożenie domen względem siebie autorka używała metody analizy drgań normalnych (ang. *normal mode analysis*, NMA). Do udokładnienia fragmentów struktury, zwłaszcza w miejscu katalitycznym, autorka używała metody klasycznej dynamiki molekularnej. Stwierdziła, że w strukturze występują dwie istotne pętle, które najprawdopodobniej ułatwiają związanie substratu i jego oddysocjowanie po hydrolizie.

W krótkim Rozdziale 5 autorka podsumowuje najważniejsze wyniki rozprawy. Dalej przedstawia Streszczenie w języku polskim i angielskim, bibliografię oraz załączniki, zawierające między innymi sekwencje genów czy porównania sekwencji.

Po przeczytaniu rozprawy ciekawa jestem zdania autorki dotyczącego kilku kwestii przedstawionych poniżej.

Do modelowania struktur białek używano oprogramowania Modeller i I-Tasser. Na stronie 39 pojawia się stwierdzenie dot. narzędzia I-Tasser, że „... wyniki modelowania struktury całego białka ATDCL4 tą

metodą ... nie były zadowalające”. Jak to oceniono i co oznacza, że uzyskany model nie był zadowalający?

Czy próbowano może sprawdzić, czy podobne modele są do uzyskania w AlphaFold? Jestem ciekawa zdania autorki na temat zastosowania uczenia maszynowego do przewidywania struktury białek.

Jaka jest efektywność dostarczania amiRNA do komórek roślinnych? Czy korzyści z zaprojektowanej skutecznej sekwencji nie są niwelowane trudnością w dostarczaniu do komórek?

Ciekawa jestem na jakiej podstawie szacowano ΔG dla niesparowanych fragmentów (podrozdział 4.1.5.)? Czy sposób przypisania połowy wartości wynikał z prób i błędów, czy był oszacowany na podstawie temperatur topnienia dupleksów RNA zawierających wyrzuszenia?

Czy doświadczenia biologiczne z jakąś zaprojektowaną przez AmiRNA Designer sekwencją zostały przeprowadzone przez kogoś w laboratorium i skuteczność oprogramowania potwierdzona doświadczalnie?

Czy od czasu stworzenia modelu białka atDCL4 w kompleksie z dupleksem RNA pojawiło się więcej danych doświadczalnych mogących potwierdzić słuszność modelu autorki?

Jak długie były symulacje dynamiki molekularnej prowadzone w celu potwierdzenia stabilności kompleksów domena:RNA?

Zauważyłam w rozprawie kilka drobnych błędów, nieścisłości, uchybień czy literówek. Z obowiązku recenzenta przytaczam poniżej niektóre z nich.

Rys 1.1 w podpisie nie jest napisane, że kolorem pomarańczowym zaznaczono rybosomy

Str. 18 – powinno być metylotransferazę

Str. 22 – brakuje przecinka przed „a także”

Str. 22 – powinno być RNazy III

Str. 25 - brakuje spacji przed jednostką Å

Str. 31 – powinno być *A. thaliana*

Str. 38 – powinno być restraints

Str. 40 – przydałyby się odnośniki literaturowe do wymienionych pól siłowych

Str 44, 45 itd. – nie do końca rozumiem, dlaczego autorka zdecydowała się używać angielskiego słowa target w rozprawie zamiast polskiego cel. Rys. 4,1 Sekwencja Targetu trochę razi...

Załącznik nr 1 – kolumna „długość” nie jest potrzebna, gdyż zawiera takie same liczby. Wystarczyło pewnie napisać w podpisie jak długie są fragmenty

Niektóre rysunki są słabo czytelne np. Rys. 4.4 i 4.6

Str. 58 – powinno być „do profilu”

str. 95 – brakuje odstępów między 2 a ns oraz 1 a ns

str. 95 – określenie „model miał RMSD” jest niezręczne

str. 95 – pierwsze zdanie drugiego akapitu jest niezrozumiałe

str. 95 – brakuje kropki w skrócie ok.

str. 95 i 96 – nie wyjaśniono w rozprawie co oznacza LG score oraz β -faktor (ten ostatni zwykle określa się jako czynnik β)

rys. 4.28 i 4.30 – podpisy aminokwasów są niewyraźne i za małe

str. 102 – „tworzy interakcje” jest kalką z języka angielskiego, lepiej byłoby używać słowa „oddziałuje” lub „sposoby oddziaływania”. Uwaga to dotyczy też innych zdań rozprawy.

str. 102 – powinno być z 5' końcem

str. 103 – powinien być odstęp po ok. 60 oraz 5' koniec

Rys. 4.29 – potencjał raczej określamy jako dodatni i ujemny a nie pozytywny i negatywny

Pomimo powyższych uwag uważam, że rozprawa doktorska mgr Mickiewicz-Piotrowskiej porusza i rozwiązuje ważny temat wyciszania ekspresji genów przez microRNA u roślin i moje uwagi nie umniejszają dokonań autorki dotyczących meritum. Zrozumienie i opanowanie techniczne zjawiska interferencji RNA może umożliwić liczne zastosowania biotechnologiczne, a także pomóc w przyszłości w projektowaniu leków. Procesy te, jak to w biologii, są jednak skomplikowane, gdyż wymagają rozpoznawania RNA przez białka, transportu z jądra do cytoplazmy, odpowiedniego wycinania miRNA z ich prekursorów. Zaprojektowanie odpowiednich sekwencji miRNA nie jest oczywiste, gdyż nie jest to oparte jedynie na prostej komplementarności odpowiednich fragmentów RNA. Oprócz wiedzy na temat możliwych roślinnych prekursorów miRNA, trzeba uwzględnić dopasowanie miRNA do drugiej nici w dupleksie spinki RNA (miRNA/miRNA*) oraz do celu w transkrypcie, który ma być wygaszany. Badania doświadczalne pokazały, że skuteczne sekwencje miRNA nie są oparte tylko na prostej komplementarności nici. Dotychczasowe programy nie uwzględniały parametrów termodynamicznych oddziaływania dla różnych możliwych dopasowań nici RNA. Dodatkowo, okazuje się, że jeśli oddziaływanie w niektórych fragmentach jest zbyt silne, to wcale nie poprawia to skuteczności wyciszania. Dlatego też autorka wprowadziła możliwość zmiany sposobu parowania nukleotydów. Użytkownik może też wgrać do oprogramowania własny zestaw danych, na podstawie którego chce projektować sekwencje amiRNA. Doceniam ilość metod i programów, którymi autorka rozprawy się posługiwała, które musiała poznać i zrozumieć ich działanie, aby stworzyć oprogramowanie amiRNA Design. Dodatkowo, problem biologiczny jest na tyle skomplikowany i nie do końca poznany, że stworzenie algorytmu wymagało sporo pracy i zrozumienia aspektów biologicznych.

Druga część badań, w której autorka starała się zrozumieć mechanizm tworzenia kompleksu roślinnego białka typu Dicer z 21-nukleotydowym dupleksem RNA, również stanowiła trudny problem do rozwiązania. Wymagało to od autorki opanowania wielu technik przewidywania struktury trójwymiarowej oraz jej udokładniania. Białko to jest wielodomenowe, co dodatkowo utrudniało zadanie, gdyż przestrzenna orientacja domen nie jest łatwa do przewidzenia i zwykle domeny są dynamiczne. Nie było też wielu danych dla tej rodziny białek, a białka roślinne są różne od białek zwierzęcych. Metody automatyczne nie radzą sobie z przewidywaniem struktur kompleksów wielu domen, a dodatkowo pozycjonowanie dupleksu RNA nie było sprawą trywialną. Uważam, że model autorki może być wykorzystany do strukturalnego zrozumienia działania białek typu Dicer oraz do projektowania doświadczeń, na przykład przewidując, które mutanty poprawią bądź obniżą efektywność reakcji.

Wyniki rozprawy zostały już opublikowane w pięciu publikacjach z listy filadelfijskiej. W dwóch pracach mgr Mickiewicz-Piotrowska jest pierwszym autorem. Jedna z nich dotyczy programu AmiRNA

Designer, która od roku 2016 była już cytowana przez innych naukowców kilkanaście razy. Druga praca z pierwszym autorstwem mgr Mickiewicz-Piotrowska dotyczy stworzenia modelu kompleksu białka typu Dicer z dwuniciowym RNA.

Podsumowując stwierdzam, że rozprawa doktorska mgr Mickiewicz-Piotrowskiej stanowi oryginalne rozwiązanie problemu naukowego, potwierdza jego wiedzę teoretyczną w dziedzinie nauk ścisłych i przyrodniczych, dyscyplinie nauki chemiczne oraz umiejętność samodzielnego prowadzenia pracy naukowej. Rozprawa doktorska mgr Agnieszki Mickiewicz-Piotrowskiej spełnia więc warunki ustawy o stopniach naukowych i tytule naukowym. W związku z tym wnoszę o dopuszczenie mgr Mickiewicz-Piotrowskiej do dalszych etapów przewodu doktorskiego.

J. Trzeńska