

dr hab. Katarzyna Pachulska-Wieczorek,

Poznań, 22.11.2024

prof. ICHB PAN

Instytut Chemii Bioorganicznej PAN

Zakład Struktury i Funkcji RNA

kasiapw@ibch.poznan.pl

Recenzja rozprawy doktorskiej

mgr Marianny Plucinskiej-Jankowskiej

**pt. “Computational analysis of RNA secondary structures in viral genomes
using high-throughput structure probing methods”**

„Obliczeniowa analiza struktur drugorzędowych RNA w genomach wirusowych w oparciu o dane z metod wysokoprzepustowego próbkowania strukturalnego”

1. Problem badawczy i jego znaczenie

Przedstawiona do oceny rozprawa doktorska mgr Marianny Plucinskiej-Jankowskiej dotyczy bardzo ważnego zagadnienia, struktury cząsteczek RNA i konstruowania narzędzi obliczeniowych wspomagających analizę uzyskanych eksperymentalnie danych strukturalnych. Bez wątplenia podjęty problem badawczy wpisuje się w aktualne trendy w badaniach RNA, gdyż dokładne poznanie związku pomiędzy strukturą a funkcją cząsteczek RNA jest niezbędne dla pełnego zrozumienia zależnych od RNA procesów komórkowych oraz wykorzystania potencjału terapeutycznego RNA. Analizy oparte na połączeniu zależnej od struktury chemicznej modyfikacji RNA i sekwencjonowania nowej generacji, zrewolucjonizowały badania struktury drugorzędowej RNA. Jednak różnorodność stosowanych podejść eksperymentalnych i metod obliczeniowych stwarza duży problem w analizie i porównaniu danych uzyskanych przez różne grupy badawcze. Zgadzam się z Autorką, że dla rozwiązania tego problemu konieczne jest stworzenie uniwersalnych narzędzi obliczeniowych, które mogłyby być stosowane niezależnie od zastosowanego w sondowaniu struktury odczynnika chemicznego czy metody detekcji miejsc modyfikacji w RNA. Tego ambitnego zadania mgr Marianna Plucińska-Jankowska podjęła się w swojej pracy doktorskiej.

Rozprawa została wykonana w Zakładzie Biologii Obliczeniowej Instytutu Biologii Molekularnej i Biotechnologii Uniwersytetu im. Adama Mickiewicza w Poznaniu, pod kierunkiem prof. UAM dr hab. Marka Żywickiego, posiadającego duże doświadczenie i osiągnięcia w obliczeniowych analizach struktury RNA. Głównym celem rozprawy pracy było przeprowadzenie kompleksowej analizy porównawczej struktur drugorzędowych RNA w regionach UTR izolatów wirusa Zika (ZIKV), i wirusa dengi (DENV) z zastosowaniem opracowanych przez Autorkę nowych narzędzi obliczeniowych, probNORM i rnaCARD. Według informacji zawartych w pracy probNORM jest

uniwersalną metodą analizy danych z eksperymentalnego sondowania struktur drugorzędowych RNA o wysokiej przepustowości. Natomiast, rnaCARD umożliwia porównywanie alternatywnych struktur drugorzędowych RNA. Opracowane metody Autorka zastosowała do analizy danych z chemicznego sondowania struktury RNA wirusów Zika i dengi, zdeponowanych w ogólnodostępnych repozytoriach. Była ona ukierunkowana na identyfikację nowych konserwowanych motywów strukturalnych RNA, analizę struktur alternatywnych i porównanie ze znanymi strukturami modelowymi.

2. Ocena formalna i merytoryczna pracy. Wkład Autorki.

Rozprawa została przygotowana w języku angielskim, ma klasyczny układ i liczy 130 stron, w tym 30 stron bibliografii, spisów tabel i rysunków oraz materiałów uzupełniających. Rozdział *Wstęp (Introduction)* został poprawnie skonstruowany i dobrze wprowadza w tematykę badań. Świadczy on o dobrym ogólnym stanie wiedzy Autorki w zakresie struktury RNA, metod jej badania, analizy danych z eksperymentów sondowania struktury RNA, a także wirusów RNA. W jego kolejnych podrozdziałach Autorka opisuje strukturę RNA, sposoby jej wizualizacji, czynniki wpływające na proces zwijania się RNA, obliczeniowe, enzymatyczne i chemiczne metody analizy struktury drugorzędowej RNA oraz wykorzystanie danych z eksperymentów sondowania struktury RNA do predykcji modeli struktury 2D RNA. Ostatni podrozdział poświęciła wirusom z rodziny Flaviviridae, skupiając się na charakterystyce strukturalnej i funkcjonalnej RNA wirusa Zika, który stanowi obiekt badań Autorki. Mimo ogólnego pozytywnego wrażenia, rozdział ten zawiera pewne błędy merytoryczne oraz liczne błędy edytorskie, gramatyczne, interpunkcyjne oraz w cytowaniu prac innych autorów. Proszę Autorkę o odniesienie się do kilku szczególnie intrygujących mnie kwestii.

- Str. 11, podpis rysunku nr 1 jest niezgodny z oznaczeniami na rysunku.
- Str. 18, opisując wpływ modyfikacji potranskrypcyjnych na strukturę RNA Autorka odnosi się do Tabeli 2, jednak nie wydaje się ona tematycznie związana, gdyż przedstawia spis enzymów stosowanych w sondowaniu struktury drugorzędowej RNA.
- *Chemical probing* stanowi podrozdział w rozdziale *Enzymatic probing*, zakładam, że jest to tylko błąd edytorski przy konstruowaniu spisu treści.
- Str. 24, przy 1M7 powinna zostać zacytowana praca wprowadzająca ten odczynnik czyli Mortimer i Weeks 2007; prace cytowane przy 1M6 [71], FAI [66] i BzCN [72] w ogóle nie dotyczą tych odczynników.
- Str. 24, Autorka pisze ... *with FAI being particularly useful for probing RNA in complex environments due to its high solubility and reactivity [66], while NAI is favored for its ability to provide high-resolution structural information [55]*. Trudno mi się zgodzić z tym stwierdzeniem, gdyż FAI nie jest bardziej reaktywny niż NAI, zresztą wynika to już z oryginalnej pracy (Spitale et al. 2013). Ponadto, nieprawidłowe jest cytowanie pracy [66] w

tym miejscu, gdyż nie ma w niej mowy o FAI. Jest to praca z grupy K. Weeks'a, która z zasady nie stosuje odczynników innych niż własne.

- Str. 24, niestety nie zgadzam się ze stwierdzeniem Autorki *Ability to modify RNA with minimal perturbation to the native structure is characteristic of BzCN, making it ideal for in vivo studies* [72]. Jestem ciekawa skąd Autorka zaczerpnęła tę informację, bo cytuje w tym zdaniu pracę Smola 2015, gdzie nie ma mowy o BzCN. Niestety ja nie znam żadnej pracy, gdzie z sukcesem zastosowano BzCN *in vivo*. Według mojej wiedzy BzCN wcale nie jest idealny do badań *in vivo*, co zapewne wynika z jego bardzo krótkiego czasu działania.
- Str. 26, opisując metody chemicznego mapowania struktury drugorzędowej RNA Autorka wspomina o *hydroxyl radical probing* (nie *radicals*, jak użyto w pracy), według mojej wiedzy metoda ta pozwala na sondowanie struktury trzeciorzędowej RNA, a nie drugorzędowej, oraz miejsc wiązania RBP.
- Str. 27, w Tabeli nr 4 przy metodzie PARTE, brak organizmu, podpowiadam, były to drożdże. Natomiast w pracy [74] cytowanej przy SHAPE-MaP nie stosowano NAI, tylko 1M7. Poza tym, Tabela nr 4 nie jest wcale cytowana w tekście pracy.
- Str. 27, Autorka stwierdza, że w eksperymentach MaP, dochodzi do substytucji zasad w syntetyzowanej nici cDNA, według mnie są to również mutacje innego typu.

Rozdział *Materials and Methods* liczy zaledwie pięć stron, co jednak wydaje się normą w rozprawach o charakterze głównie obliczeniowym. Zabrakło w nim jednak informacji na temat pochodzenia danych SHAPE-MaP użytych w podrozdziale 4.1.2. Jeśli ją przeoczyłam, przepraszam.

Opis przeprowadzonych przez Autorkę badań i ich wyników został przedstawiony w trzech podrozdziałach. Pierwszy z nich opisuje zasadę działania metody probNORM, przedstawia jej szczegółową walidację dla danych typu *RT-stop*, porównanie z innymi dostępnymi metodami, zastosowanie dla analizy danych typu *RT-mutate* oraz utworzony przez Autorkę serwer www.probnorm.org. Ogólnie uważam, że metoda probNORM ma bardzo duży potencjał. Z uwagi na istniejące niespójności w zakresie analizy danych z chemicznego sondowania struktury RNA, szczególnie tych pochodzących z metod typu *RT-stop*, jest narzędziem bardzo potrzebnym. Doceniam również utworzenie własnego skryptu do obliczenia reaktywności poszczególnych reszt nukleotydowych w RNA (dla danych *RT-stop*). Autorka przedstawia liczne porównania danych bez normalizacji oraz przeanalizowanych z użyciem probNORM, które w przekonujący sposób wskazują, że zastosowanie probNORM skutkuje większym pokryciem niesparowanych zasad o reaktywności większej niż 0 i zapobiega tzw. utracie sygnału. Mam jednak kilka krytycznych uwag i prosiłabym Autorkę o odniesienie się do nich podczas publicznej obrony.

- Str. 40 i 63, podany link odnosi do strony programu rnaNORM. Czy dobrze rozumiem, że jest to tylko zmiana nazwy? Niemniej taka informacja powinna znaleźć się w pracy.

- Str. 43, referencja [91] cytowana przy programie ShapeMapper nie dotyczy tego programu.
- Dlaczego do walidacji programu probNORM wybrano tylko dane z eksperymentów typu *RT-stop* (Mod-seq i SHAPE-seq 2.0), skoro finalnie program został użyty do normalizacji danych z eksperymentów SHAPE-MaP?
- Str. 48, analizując korelację danych typu *RT-stop* w próbce modyfikowanej i niemodyfikowanej Autorka odnosi się do rysunku 9a, który przedstawia wyniki z eksperymentów SHAPE-MaP uzyskane w różnych warunkach.
- Str. 49 – 51, w tym rysunki 9, 10 i 11 zawierają wyniki badań opublikowane w pracach Grzywacz 2023 i Plucińska 2016, których Doktorantka jest współautorką. Nie rozumiem jednak ich umieszczenia w rozdziale przedstawiającym opis metody probNORM. Wprowadza to niepotrzebne zamieszanie. Jeśli faktycznie były one motywacją do stworzenia tej metody, ich opis powinien poprzedzać rozdziały o probNORM.
- Str. 50, stwierdzenie na podstawie danych uzyskanych dla tRNA, że *DMS-probing data [...] contribution to improving prediction accuracy for small RNAs is limited*, uważam za bardzo ryzykowne. Analogiczne analizy należałoby przeprowadzić dla innych klas krótkich RNA, bowiem tRNA to bardzo specyficzne cząsteczki, bogate w różnorodne modyfikacje.
- Str. 51, podpisy rysunków 10 i 11 zawierają istotne błędy edytorskie.
- Str. 53, jaki RNA kryje się pod określeniem *SRP transcript* ?
- Str. 57 – 61, Autorka porównuje probNORM z trzema innymi metodami (BUM-HMM, structure-seq i CMH). Co ciekawe, w tym porównaniu najgorzej wypada BUM-HMM. Jednak uzyskane wyniki są zupełnie inne, niż te w oryginalnej pracy wprowadzającej metodę BUM-HMM (Selega 2016). Proszę Autorkę o porównanie i przedyskutowanie uzyskanych przez nią danych dla 18S rRNA z danymi zaprezentowanymi na Rys. 4 w pracy Selega 2016.

Kolejny podrozdział części wynikowej opisuje metodę rnaCARD, przykłady jej zastosowania oraz serwer www.rnaCARD. Metoda umożliwia automatyczne, kompleksowe porównanie dwóch alternatywnych struktur tej samej cząsteczki RNA i według Autorki może być stosowana nie tylko dla pojedynczych transkryptów, ale również do jednoczesnej analizy wielu RNA, w tym całych transkryptomów. Jako jej zaletę Autorka wylicza możliwość identyfikacji podobnych a nie tylko identycznych motywów strukturalnych RNA. Jako przykład porównała dwie suboptymalne struktury dwóch ludzkich transkryptów SLIRP-201 i SCT-201. Szersze wykorzystanie rnaCARD przedstawione jest w kolejnych podrozdziałach, gdzie zaprezentowano porównanie rejonów UTR ZIKV i DENV. Podczas lektury tej części pracy nasunęło mi się kilka pytań i uwag, dlatego proszę Autorkę o odpowiedni komentarz.

- Na stronie 71 Autorka informuje, że wyniki rnaCARD można przeglądać w dwóch trybach, ale wymienia tylko jeden z nich. Domyślam się, że ten drugi to *differential mode*. Jak włączyć ten tryb? Na stronie rnaCARD nie widzę tej opcji.

- Czy i ewentualnie jaki jest limit długości cząsteczek RNA, które mogą zostać porównane z użyciem rnaCARD?
- Autorka twierdzi, że rnaCARD może zostać zastosowany do analiz w skali transkryptomowej. W pracy przedstawiono wyniki dla pojedynczych transkryptów, co przyjmuję ze zrozumieniem ze względu na wyznaczone cele. Jednak powyższe twierdzenie sugeruje, że testowano metodę na dużych zbiorach transkryptów. Jak dużych? Pytam, gdyż wielkość transkryptomów różnych organizmów jest bardzo zróżnicowana.
- Czy rozważano możliwość porównania więcej niż dwóch struktur danego transkryptu?
- Jak wypadają zalety i wady rnaCARD w odniesieniu do innych narzędzi automatycznego porównania struktur RNA?
- W opisie rysunku 26 powinno być *rnaCARD Web Server* a nie *probNORM Web Server*.

Ostatni podrozdział części wynikowej Autorka poświęciła opisowi analiz strukturalnych przeprowadzonych dla rejonów UTR RNA ZIKA i DENV. Wykorzystała dane SHAPE-MaP z pracy Huber 2019. W tym przypadku nie obliczała samodzielnie reaktywności (częstotliwości mutacji) lecz wykorzystała program ShapeMapper. Do normalizacji pozyskanych danych zastosowała probNORM, a następnie wykorzystując oprogramowanie RNAstructure przewidziała wsparte danymi eksperymentalnymi modele struktury drugorzędowej dla rejonów 5' i 3'UTR czterech izolatów każdego z wirusów. W kolejnym etapie wykorzystała metodę rnaCARD do przeprowadzenia ich kompleksowego porównania ze sobą oraz strukturami modelowymi. Intryguje mnie jednak kilka kwestii i proszę Autorkę, aby podczas publicznej obrony podjęła próbę ich bardziej dogłębnego wyjaśnienia i skomentowania.

- Jak już wspomniałam, w pracy nie pokazano walidacji metody probNORM dla danych typu *RT-mutate* z zastosowaniem krótkich RNA o znanej strukturze. Jednocześnie brakuje również porównania danych *RT-mutate* przed i po normalizacji metodą probNORM. Bardzo cenne byłoby również zestawienie częstotliwości mutacji nukleotydów znormalizowanych z zastosowaniem programu ShapeMapper (czyli tak jak w oryginalnej pracy, Huber 2019) z tymi znormalizowanymi przy użyciu probNORM.
- Kolejnym brakującym elementem jest dla mnie bardziej precyzyjna ocena dopasowania reaktywności znormalizowanych metodą probNORM do uzyskanych modeli struktury drugorzędowej RNA, np. analiza AUC ROC. To dopasowanie można w pewnym stopniu odczytać z reaktywności naniesionych na modele struktury drugorzędowej np. rys. 27 - 30, lecz jest to tylko „ocena na oko”. Bardziej precyzyjne podejście pozwoliłoby ocenić, co zmienia nowa metoda normalizacji i czy jest to zmiana na lepsze.
- Jakie kryterium zastosowano przy wyborze aktualnych struktur modelowych dla rejonów UTR RNA ZIKV i DENV? Z referencji podanych w rozdziale *Materials and Methods* (str. 45) wynika, że wybrany model 5'UTR ZIKV został otrzymany na podstawie danych

eksperymentalnych (SHAPE-CE *in vitro*), modele 3'UTR ZIKV i DENV pochodzą z różnych analiz *in silico*, nie wiem niestety skąd pochodzi model DENV 5'UTR, gdyż cytowana jest przy nim praca przeglądowa. Prace będące źródłem tych modeli nie są nawet raz cytowane w części wynikowej dysertacji. Stąd wynika moje kolejne pytanie. Czy modele w ramkach na rysunkach 31, 32, 33 i 34 pochodzą z prac cytowanych na stronie 45?

- Do którego z UTRów ZIKV odnoszą się dane przedstawione na rys. 27? Zakładam, że jest to 5'UTR, lecz nie wynika to z podpisu rysunku.
- Modele 5'UTR izolatów BRA i PF przedstawione na rys. 27 różnią się od tych z rys. 2 suplementu. Inny jest również rozkład reaktywności mimo, że do normalizacji wszędzie zastosowano probNORM. Z czego to wynika?
- W podpisach rysunków 27 do 30 głównego tekstu oraz rysunków 2, 3 i 4 suplementu Autorka umieściła informację, że reaktywności zostały obliczone metodą probNORM (*reactivity calculated with probNORM*). Jednak jest to zbyt daleko idące uproszczenie, gdyż według informacji przedstawionych w pracy reaktywności obliczono stosując program ShapeMapper, a jedynie do ich normalizacji zastosowano probNORM.
- Dostępne dane literaturowe sugerują występowanie czterech motywów typu pseudowęzła w 3'UTR RNA ZIKV i DENV. Dane znormalizowane metodą probNORM wspierają tworzenie trzech z nich. Wniosek ten Autorka wysunęła na podstawie analizy znormalizowanych reaktywności w rejonach potencjalnie zaangażowanych w tworzenie tych motywów. Zastanawiam się, dlaczego Autorka nie spróbowała przeprowadzić analizy z użyciem ShapeKnot wykorzystując dane znormalizowane za pomocą probNORM. ShapeKnot stanowi część pakietu RNAstructure, z którego Autorka korzystała w swoich analizach i jestem pewna, że jest biegła w jego stosowaniu. Tym samym nie do końca zgadzam się ze zdaniem ze strony 96 *Methods I used for secondary structure prediction do not enable pseudoknot prediction*.

Kolejny rozdział rozprawy stanowi obszerna Dyskusja, w której Autorka w ciekawy i umiejętny sposób podsumowuje wyniki swoich badań i zestawia je z innymi, zarówno w zakresie utworzonych metod, jak i analizy rejonów UTR ZIKV i DENV. Zgadzam się z wieloma stwierdzeniami Autorki dotyczącymi zalet i możliwości zastosowania obu utworzonych metod. Mam jednak pewne uwagi odnośnie zdania dotyczącego probNORM (str. 94) *To my knowledge, no other method for transcriptome-wide reactivity calculation is available as an online tool*. Po pierwsze, w przypadku danych typu *RT-mutate* jako plik wejściowy trzeba wprowadzić obliczone wcześniej za pomocą innego programu częstotliwości mutacji dla poszczególnych nukleotydów. Po drugie obliczenie reaktywności dla danych typu *RT-stop* wymaga pracy w *command line*, co zapewne jest proste dla bioinformatyka, ale nie dla innych badaczy. Myślę, że wykluczenie wymienionych powyżej ograniczeń stanowi dobry kierunek dla dalszego rozwoju wersji online metody probNORM.

Niestety od pewnego momentu Dyskusji ciężko podążać za tokiem myślenia Autorki, gdyż część odnośników literaturowych z tekstu nie ma swoich odpowiedników w spisie bibliografii. Domyślam się, że odnosząc się na str. 94 do pracy Li et al., Autorka ma na myśli pracę Li et al. 2018 z Cell Host Microbe (DOI: 10.1016/j.chom.2018.10.011), bo to właśnie w niej zastosowano metodę icSHAPE do analizy strukturalnej ZIKV RNA. Tej pracy nie ma niestety w bibliografii. Jeśli się mylę, proszę o korektę. Na stronie 96, Autorka pisze, że wymodelowana przez nią struktura motywu trójramiennego rozgałęzienia w 3'UTR ZIKV perfekcyjnie odzwierciedla tą uzyskaną w badaniach krystalograficznych. Jest to bardzo wartościowa obserwacja, jednak cytowanie w tym miejscu pracy o numerze [100] wydaje się niewłaściwe.

Bibliografia pracy liczy 133 pozycje. Jak wspomniałam powyżej odnośników literaturowych w pracy jest więcej, dokładnie 143. Zatem na liście brakuje dziesięciu pozycji. Poza tym zauważyłam, że przykładowo praca E. Mailler, 2019 DOI: 10.1002/wrna.1518, ma przypisane dwa numery referencyjne, [53] i [56]. Brakuje również konsekwencji w wymienianiu autorów poszczególnych publikacji lub źródła pracy, np. [59], [64], [128]. Niestety, brak rzetelności w cytowaniu prac innych autorów, a nawet całkowity brak odnośników literaturowych w miejscach, które tego wymagają jest ogólnym problemem całej dysertacji. W recenzji wymieniłam tylko część błędnych odnośników literaturowych, pominęłam wyliczanie miejsc w których ich brakuje.

Rzetelności zabrakło również przy sporządzaniu wykazu skrótów zawartych w pracy, który jest bardzo ograniczony i mam wrażenie, że wybór objaśnianych skrótów był zupełnie losowy. Ze względu na mnogość skrótów użytych w dysertacji dobrze przygotowany wykaz byłby niezwykle przydatny.

3. Podsumowanie

Mimo wymienionych powyżej krytycznych uwag pozytywnie oceniam wyniki badań Autorki uzyskane w ramach realizacji pracy doktorskiej. Mam świadomość, że osiągnięcie postawionych celów nie było łatwe i wymagało ogromu pracy. Warto podkreślić, że stanowią one próbę rozwiązania realnych problemów naukowców zajmujących się badaniami strukturalnymi RNA. Wartościowy jest również wybór modelu badawczego, czyli RNA wirusów Zika i dengi. Doceniam również stworzenie serwerów www.probnorm.org i [rnacard.org](http://www.rnacard.org), co czyni utworzone metody bardziej dostępne dla szerokiego grona badaczy RNA nie będących informatykami. Metody probNORM i rnaCARD nie zostały jeszcze opublikowane, ale według informacji zawartych w pracy opisujące je manuskrypty są w przygotowaniu. W obu przypadkach mgr Marianna Plucinska-Jankowska ma być pierwszą autorką. Tym bardziej szkoda, że sama dysertacja nie została przygotowana w sposób bardziej profesjonalny.

W końcowej konkluzji, biorąc pod uwagę wymagania zdefiniowane przez ustawę z dnia 14 marca 2003 roku o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. z 2017 r. poz. 1789) stwierdzam, że rozprawa doktorska mgr Marianny Plucinskiej-Jankowskiej zawiera oryginalne rozwiązanie problemu naukowego, świadczy o dobrej ogólnej wiedzy teoretycznej

Kandydatki w uprawianej dyscyplinie oraz umiejętności samodzielnego prowadzenia pracy naukowej.
W związku z powyższym wnioskuję do Rady Naukowej Dyscypliny Nauki Biologiczne Uniwersytetu im. Adama Mickiewicza w Poznaniu o dopuszczenie Pani mgr Marianny Plucinskiej-Jankowskiej do dalszych etapów przewodu doktorskiego.

W. Pachulska-Wierzowa