

Recenzja rozprawy doktorskiej

mgr. Michała Pawła Żurkowskiego

pt. „Algorithms for feature exploration and modeling of quadruplex structures”

1. Problem badawczy i jego znaczenie

Praca doktorska Pana mgr. Michała Żurkowskiego została wykonana w Instytucie Informatyki Politechniki Poznańskiej pod opieką naukową prof. dr hab. Marty Szachniuk. Rozprawa poświęcona jest opracowaniu nowych algorytmów i narzędzi bioinformatycznych służących lepszemu poznaniu i zrozumieniu struktur G-kwadrupeksów występujących w cząsteczkach kwasów rybonukleinowych, DNA oraz RNA. G-kwadrupeksy to niekanoniczne struktury tworzone przez bogate w ciągi guanin cząsteczki DNA lub RNA, w których cztery guaniny, tzw. tetrady guaninowe (tetrady G), są stabilizowane przez wiązania wodorowe Hoogsteena i jednowartościowy kation, najczęściej kation potasowy (K^+). G-kwadrupeksy pełnią ważną rolę regulacyjną w podstawowych procesach biologicznych, takich jak replikacja, transkrypcja czy translacja, podczas gdy nieprawidłowe tworzenie tych struktur jest związane z niestabilnością genomu i powstawaniem nowotworów oraz schorzeń neurodegeneracyjnych. Struktury G-kwadrupeksu występują w rejonach promotorowych genów, na końcach chromosomów, w rejonach regulatorowych transkryptów, czy też w obrębie pętli R, czyli trójniciowych struktur kwasu nukleinowego, składających się z hybrydy DNA:RNA i wybrzuszony jednoniciowego DNA, w obrębie którego mogą powstawać struktury G-kwadrupeksu.

Zrozumienie funkcji biologicznych odgrywanych przez G-kwadrupeksy wymaga szczegółowej wiedzy na temat ich struktury. Opracowane przez Doktoranta narzędzia bioinformatyczne mogą mieć ważne znaczenie dla szerokiego grona badaczy skupiających się na problemie zaburzonej regulacji procesów komórkowych, wynikającym z nieprawidłowego tworzenia się G-kwadrupeksów; na przykład niestabilności genomu, która może być przyczyną pojawiania się wielu chorób, w tym chorób o podłożu nowotworowym czy neurodegeneracyjnym. Rozwiązania zaprezentowane w rozprawie doktorskiej mogą mieć także wymiar praktyczny. Bazy danych oraz narzędzia do analizy i wizualizacji motywów G-kwadrupeksowych mogą stanowić wsparcie dla osób zajmujących się empirycznymi badaniami tych niekanonicznych motywów strukturalnych.

2. Wkład autora

Przedłożona rozprawa doktorska jest oparta na zbiorze pięciu publikacji o charakterze naukowym. Pierwsza praca: T. Zok, N. Kraszewska, J. Miskiewicz, P. Pielacinska, **M. Żurkowski**, M. Szachniuk (2021), ONQUADRO: a database of experimentally determined quadruplex structures; *Nucleic Acids Research* (IF₂₀₂₁ 19.2) [A1], dotyczy opracowania repozytorium ONQUADRO dedykowanego kwadrupeksom. Repozytorium ONQUADRO umożliwia gromadzenie i przetwarzanie danych

o strukturach kwadrupleksów otrzymanych drogą eksperymentalną, w tym danych dotyczących sekwencji, struktur drugo- i trzeciorzędowych, tetrad oraz helis. ONQUADRO to samoaktualizujące się repozytorium, a dane pozyskiwane są z bazy Protein Data Bank (PDB).

Druga praca: **M. Zurkowski**, T. Zok, M. Szachniuk (2022), DrawTetrado to create layer diagrams of G4 structures; *Bioinformatics*, (IF₂₀₂₂ 5.8) [A2], opisuje opracowanie narzędzia DrawTetrado, które automatyzuje tworzenie diagramów warstwowych w grafice 2.5-wymiarowej. DrawTetrado stwarza możliwość wizualizacji motywów G-kwadrupleksowych, w tym wizualizacji wszystkich typów pętli, co w znacznym stopniu ułatwia zrozumienie budowy strukturalnej analizowanych G-kwadrupleksów. DrawTetrado jest użytecznym i niezwykle pomocnym narzędziem, w szczególności może zostać wykorzystane podczas przygotowywania publikacji naukowych, w których niezbędne jest umieszczenie rysunków poglądowych dotyczących omawianych G-kwadrupleksów. Narzędzie DrawTetrado zostało zintegrowane z bazą danych ONQUADRO i aplikacją internetową WebTetrado. Na podkreślenie zasługuje fakt, iż Pan Michał Żurkowski jest pierwszym autorem omawianej publikacji, czyli jego wkład w powstanie narzędzia DrawTetrado był kluczowy.

Aplikacja WebTetrado została opisana w trzeciej pracy, wchodzącej w skład rozprawy doktorskiej: B. Adamczyk, **M. Zurkowski**, M. Szachniuk, T. Zok (2023), WebTetrado: a webserver to explore quadruplexes in nucleic acid 3D structures; *Nucleic Acids Research* (IF₂₀₂₃ 16.6) [A3]. WebTetrado służy do analizy struktur G-kwadrupleksów, które zostały otrzymane *in silico* lub eksperymentalnie, a pozyskane dane nie zostały jeszcze przesłane do bazy PDB. WebTetrado jest uzupełnieniem bazy danych ONQUADRO i umożliwia zautomatyzowaną identyfikację, klasyfikację i analizę strukturalną tetrad DNA i RNA, kwadrupleksów i helis G4.

Kolejny problem badawczy poruszony w rozprawie doktorskiej dotyczy dopasowania motywów struktur trzeciorzędowych przyjmowanych przez G-kwadrupleksy i wykrywania podobnych „podstruktur” w porównywanych cząsteczkach. Dopasowywanie struktur cząsteczek kwasów nukleinowych jest niezbędne podczas analizy ich cech strukturalnych i podobieństw między nimi. W celu elastycznego dopasowania struktur trzeciorzędowych RNA i DNA, Doktorant opracował dwa nowe algorytmy, GEOS i GENS. Algorytmy te pozwalają na efektywne dopasowanie wybranych fragmentów kwasów nukleinowych. Ponadto, algorytmy zostały sparametryzowane tak, aby działać w dwóch trybach, niezależnym i zależnym od sekwencji. Wspomniane algorytmy zostały opisane w czwartej publikacji: **M. Zurkowski**, M. Antczak, M. Szachniuk (2023), High-quality, customizable heuristics for RNA 3D structure alignment; *Bioinformatics*, (IF₂₀₂₃ 4.4) [A4]. Pan Michał Żurkowski jest autorem wiodącym w niniejszej publikacji.

W piątej pracy: **M. Zurkowski**, M. Swiercz, F. Wozny, M. Antczak, M. Szachniuk (2024), RNAhugs web server for customized 3D RNA structure alignment; *Nucleic Acids Research* (IF₂₀₂₄ 16.6) [A5], przedstawiony został system o nazwie RNAhugs, który umożliwia uruchomienie algorytmów GEOS i GENS zarówno zależnie, jak i niezależnie od sekwencji, z możliwością dostosowania wartości parametru RMSD (RMSD to miara odległości między dwoma zbiorami atomów o tej samej liczności, wyrażona w Å). W pracy tej, Doktorant zajmuje pozycję wiodącą (pierwszy autor).

Ponadto, rozprawa doktorska zawiera opis algorytmu LinkTetrado (dane nieopublikowane). LinkTetrado to automatyczna metoda detekcji motywów multimerycznych opartych na strukturach G-kwadrupleksu. Jest to narzędzie, które pozwala na odnalezienie w cząsteczkach DNA i RNA nukleotydów wchodzących w interakcje z tetradami. W wyniku tych oddziaływań tworzą się pentady, heksady, heptady, itd. Zastosowanie LinkTetrado pozwala na rozszerzenie katalogu znanych motywów strukturalnych i umożliwia analizę ich właściwości. Do tej pory algorytm LinkTetrado został

zweryfikowany w kontekście danych eksperymentalnych pochodzących z analiz wykonanych metodą magnetycznego rezonansu jądrowego (NMR).

Podsumowując, Pan Michał Żurkowski jest pierwszym autorem trzech prac o charakterze naukowym wchodzących w skład rozprawy doktorskiej (A2, A4 i A5), czyli jego wkład w powstanie niniejszych publikacji był kluczowy, wiodący. W dwóch pozostałych pracach (A1 i A3), Doktorant znajduje się na, odpowiednio, piątej i drugiej pozycji listy Współautorów. Wkład intelektualny Doktoranta w powstanie pięciu omawianych publikacji został potwierdzony przez Panią Promotor i pozostałych Współautorów prac. Na podstawie załączonych do rozprawy doktorskiej oświadczeń Współautorów, przedruków publikacji i deklaracji Kandydata, nie mam wątpliwości, iż wkład mgr. M. Żurkowskiego w powstanie dzieła zatytułowanego „*Algorithms for feature exploration and modeling of quadruplex structures*”(Algotrymy do badania cech i modelowania struktur kwadrupleksowych) był dominujący.

Na uwagę zasługuje również fakt, iż wszystkie ze wspomnianych prac zostały opublikowane w prestiżowych, szeroko rozpoznawanych czasopismach naukowych, trzy prace w *Nucleic Acids Research* (5-IF = 16.1) i dwie w *Bioinformatics* (5-IF = 7.6). Oba te czasopisma są przypisane m.in. do dyscypliny naukowej: Informatyka techniczna i telekomunikacja i przyjmują najwyższą wartość punktową, tj. 200 pkt, zgodnie z Wykazem czasopism naukowych i recenzowanych materiałów z konferencji międzynarodowych (MNiSW/MEiN).

3. Poprawność

Rozprawa doktorska Pana Michała Żurkowskiego została napisana w języku angielskim. Praca zawiera streszczenia (w języku angielskim i polskim), listę publikacji wchodzących w skład rozprawy doktorskiej (zwieńczoną opisem wkładu Doktoranta w prowadzone badania), wstęp teoretyczny (*Introduction*), opis uzyskanych wyników (*Results*), wnioski (*Conclusions*), listę osiągnięć naukowych i dydaktycznych Doktoranta (*Scientific achievements*), spis literatury (*Bibliography*) cytowanej w rozdziałach: *Introduction* oraz *Results*, przedruki publikacji (*Publication reprints*) oraz oświadczenia współautorów dotyczące ich wkładu w powstałe publikacje (*Co-author declarations*). Wszystkie prace wchodzące w skład rozprawy doktorskiej zostały opublikowane w bardzo dobrych, recenzowanych czasopismach o zasięgu międzynarodowym (czasopisma z tzw. listy JCR, ang. *Journal Citation Reports*), co świadczy o bardzo wysokiej jakości naukowej prowadzonych badań i analiz. Pan Michał Żurkowski pełni rolę autora wiodącego (pierwszego) w trzech z pięciu publikacji wchodzących w skład rozprawy doktorskiej.

Ogólnie, praca została przygotowana starannie i przejrzysto, nie mam do niej większych uwag. Moje drobne uwagi dotyczą braku podania tytułu rozprawy doktorskiej w języku polskim (np. tytuł taki mógłby znaleźć się w rozdziale *Streszczenie*). Ponadto, w podrozdziale 1.2. (Quadruplex motifs), we fragmencie opisującym występowanie i rolę G-kwadrupleksów, zabrakło mi odniesień literaturowych. Z kolei rozdział *Conclusions* zawiera w głównej mierze podsumowanie działań podjętych w celu realizacji pracy doktorskiej, a także uzyskanych rezultatów. Wnioski stanowią niewielką część wspomnianego rozdziału. Powyższe nieliczne uwagi w żadnym stopniu nie wpływają na pozytywną ocenę całej rozprawy doktorskiej.

4. Wiedza kandydata

O wiedzy Doktoranta możemy wnioskować bezpośrednio na podstawie zawartości rozdziałów *Introduction* i *Results*. Wprowadzenie teoretyczne (*Introduction*) podzielone jest na cztery podrozdziały opisujące odpowiednio: strukturę kwasów nukleinowych, motywy G-kwadrupeksowe, modelowanie struktury kwasów nukleinowych, podstawowe pojęcia w teorii algorytmów. Rozdział ten dobrze wprowadza czytelnika w problematykę rozprawy doktorskiej. Odnośnie podrozdziału 1.2. (Quadruplex motifs), uważam, iż *Wstęp* korzystnie wzbogaciłoby dodanie informacji na temat struktur G-kwadrupeksów występujących w pętlach R. To bardzo ważny aspekt dotyczący stabilności genomowego DNA, w odniesieniu do regulacji procesów replikacji i transkrypcji. Ponadto, tak jak wspomniałam powyżej, w podrozdziale tym zabrakło mi odniesień literaturowych do pozycji opisujących występowanie i biologiczną rolę G-kwadrupeksów. Uważam jednak, że uchybienie to nie świadczy o braku wiedzy ze strony Kandydata w zakresie omawianej problematyki badawczej (ważne informacje na temat G-kwadrupeksów zostały podane), lecz wynika z drobnego przeoczenia.

Kolejny rozdział (*Results*) pokrótce opisuje wyniki badań i analiz przeprowadzonych przez Doktoranta, z podziałem na podrozdziały: (i) badanie cech kwadrupeksów (w odniesieniu do publikacji A1 i A3), (ii) wizualizacja struktur G-kwadrupeksów (w odniesieniu do publikacji A2), (iii) zestawienie/porównanie (ang. *alignment*) struktur trzeciorzędowych (w odniesieniu do publikacji A4 i A5) oraz (iv) wykrywanie motywów multimerycznych (dane nieopublikowane, dotyczące metody LinkTetrado do detekcji motywów multimerycznych opartych na strukturach G-kwadrupeksu). Rozdział ten jest zwięzły i treściwy, stosunkowo prosty w odbiorze, nawet dla czytelnika nieposiadającego doświadczenia w zakresie tworzenia algorytmów informatycznych. Umiejętność prostego przekazu informacji świadczy o dojrzałości naukowej Doktoranta i dobrej orientacji w zagadnieniach naukowych przedstawionych w przedłożonej rozprawie doktorskiej. Spis literatury (*Bibliography*) został sporządzony w kolejności alfabetycznej i zawiera pozycje cytowane w rozdziałach *Introduction* i *Main Results*. Pozycje zawarte w spisie literatury częściowo pokrywają się z pozycjami zawartymi w spisie literatury dotyczącej poszczególnych prac wchodzących w skład rozprawy doktorskiej. Cytowane są prace najnowsze, opublikowane w roku 2024, jak i te starsze, odnoszące się np. do pierwszych badań nad strukturą i funkcją biomolekuł.

Podsumowując, na podstawie informacji przedstawionych w rozprawie doktorskiej, pozytywnie oceniam ogólną wiedzę Kandydata w omawianym obszarze badawczym, dotyczącym projektowania narzędzi bioinformatycznych służących lepszemu poznaniu i zrozumieniu struktur G-kwadrupeksów występujących w cząsteczkach kwasów rybonukleinowych.

5. Inne uwagi

Doktorant wspomniał w pracy o motywach typu i (i-motifs), czy podczas obrony mogłabym poprosić o rozwinięcie tego tematu, czy znane są obecnie narzędzia do analizy tego typu motywów strukturalnych?

6. Podsumowanie

Biorąc pod uwagę opinie zaprezentowane w poprzednich punktach i wymagania zdefiniowane przez art. 187 Ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (z późniejszymi zmianami)¹ moja ocena rozprawy pod względem trzech podstawowych kryteriów jest następująca:

A. Czy rozprawa zawiera oryginalne rozwiązanie problemu naukowego? (wybierz jedną opcję stawiając znak **X**)

<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Zdecydowanie TAK	Raczej TAK	Trudno powiedzieć	Raczej NIE	Zdecydowanie NIE

B. Czy po przeczytaniu rozprawy zgadzasz się, że kandydat posiada ogólną wiedzę teoretyczną w dyscyplinie Informatyka techniczna i telekomunikacja?

<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Zdecydowanie TAK	Raczej TAK	Trudno powiedzieć	Raczej NIE	Zdecydowanie NIE

C. Czy kandydat posiada umiejętność samodzielnego prowadzenia pracy naukowej?

<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Zdecydowanie TAK	Raczej TAK	Trudno powiedzieć	Raczej NIE	Zdecydowanie NIE

Podpis

¹ <http://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU20190000276>