

dr hab. Tadeusz Kuliński, prof. IChB PAN

Zespół Modelowania Biomolekularnego

Instytut Chemii Bioorganicznej PAN

ul. Noskowskiego 12/14, 61-704 Poznań

Recenzja pracy doktorskiej mgr Jakuba Pasia

p.t. „**Application and implementation of probabilistic profile- profile comparison methods for protein fold recognition.**”

(tytuł w języku polskim: „**Wdrożenie i zastosowania probabilistycznych metod porównawczych profil-profil w rozpoznawaniu pofałdowania białek**”.)

Przedstawiona do recenzji rozprawa doktorska mgr Jakuba Pasia ma formę spójnego tematycznie zbioru artykułów opublikowanych w czasopismach naukowych. Promotorem pracy jest dr hab. Marcin Hoffmann, prof. UAM, a promotorem pomocniczym dr Krystian Eitner. Przedmiotem mojej oceny, w myśl wymagań *Ustawy o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki* z dnia 14 marca 2003 r. (Dziennik Ustaw nr 65, poz. 595, art. 13) oraz *Rozporządzenia Ministra Nauki i Szkolnictwa Wyższego z dnia 22 września 2011 r. w sprawie szczegółowego trybu i warunków przeprowadzania czynności w przewodach doktorskich, w postępowaniu habilitacyjnym oraz w postępowaniu o nadanie tytułu profesora*, jest oryginalność rozwiązane problemu naukowego, ogólna wiedza teoretyczna Kandydata w dziedzinie nauk chemicznych, a także umiejętność samodzielnego prowadzenia pracy naukowej. Ponieważ rozprawę doktorską mgr Jakuba Pasia stanowi część pracy zbiorowej, moja recenzja zawiera ocenę indywidualnego wkładu kandydata w jej powstanie.

Pracę doktorską otwiera „Lista artykułów opublikowanych w czasopismach naukowych”. Na „Liście...” znalazło się 17 publikacji, które ukazały się w latach 2002-2011 (w roku 2002-1 pozycja, w 2003-4, w 2004-7, w 2005-1, w 2006-2, w 2010-1, w 2011-1). Autor określa swój wkład do każdej z nich. W trzech Jakub Paś jest pierwszym autorem. Publikacje nie są przedstawione chronologicznie. Kryterium kolejności na liście jest współczynnik *impact factor* oraz cytowalność. Autor niestety w żaden sposób nie wyróżnia tych prac, które stanowią spójny tematycznie zbiór artykułów i faktycznie wchodzi w skład dysertacji.

Po „Liście...” następuje streszczenie napisane w języku polskim, w którym autor w sposób lapidarny przedstawia cel pracy: „wskazanie iż metody porównywania profil-profil mogą przewyższać inne metody rozpoznawania foldów w analizie spokrewnionych białek, i że mogą być one stosowane nie tylko do rozpoznawania foldów, ale także do innych celów takich jak wykrywanie i identyfikacja genów, granic domen białkowych oraz modelowania złożonych struktur białkowych”.

Nagromadzenie ogromnej ilości informacji, uzyskanych przy sekwencjonowaniu genomu i zdeponowanych w bazach danych, spowodowało duże i wciąż rosnące zainteresowanie metodami rozpoznawania pofałdowania nowo odkrytych białek, które pozwoliłyby przewidywać struktury przestrzenne oraz funkcje biologiczne dla białek posiadających niską homologię sekwencyjną do białek o znanych strukturach trzeciorzędowych zdeponowanych w bazie danych struktur białkowych *Protein Data Bank*. Porównywanie profili pozwala uliniować sekwencje i przyporządkować je do wzorca strukturalnego bardziej precyzyjnie, dzięki czemu możliwe jest uzyskiwanie lepszych modeli strukturalnych. Metody te mają również zastosowanie do analizy szlaków metabolicznych, enzymatycznych, do analizy ewolucyjnej, przewidywania efektywności dokowania molekularnego oraz projektowania leków. Prace przedstawione w rozprawie doktorskiej mgr Jakuba Pasia dotyczą więc problematyki naukowej o dużym i aktualnym znaczeniu poznawczym i aplikacyjnym. Wskazują, że zastosowanie probabilistycznych metod porównawczych profil-profil do rozpoznania pofałdowania białek jest skuteczne i często przewyższa akuratnością wyniki uzyskiwane przy pomocy innych metod.

W Komentarzu do zbioru publikacji wchodzących w skład rozprawy Autor odwołuje się do 8 spośród 17 swoich publikacji z „Listy artykułów...” (pozycja 2, 5, 7, 8, 9, 13, 15 i 17). Komentarz napisany jest w języku angielskim i liczy 28 stron. Autor podzielił go na 8 części: I. *Summary* (powtórzenie w języku angielskim streszczenia w języku polskim), II. *Introduction* (zawiera opis różnych metod stosowanych do rozpoznawania pofałdowania białek), III. *Applications profile-profile comparison methods*, IV. *Implementations of profile-profile comparison methods*, V. *Conclusions*, VI. *Figures*, VII. *References*, w ostatniej VIII. znalazły się kopie wszystkich publikacji z „Listy artykułów opublikowanych w czasopismach naukowych”.

W części III., Doktorant omawia zastosowania probabilistycznych metod porównawczych profil-profil w rozpoznawaniu pofałdowania białek w 3 pracach eksperymentalnych (pozycja 7., 8. i 13. z „Listy artykułów...”).

W publikacji 13. zatytułowanej “Two sequences encoding chalcone synthase in yellow lupin (*Lupinus luteus* L.) may have evolved by gene duplication” (D Narożna, J Pas, J Schneider, CJ Mądrzak, **Cellular & molecular biology letters** 2004, 9: 95-105, IF 0,495, ilość cytowań 2 wg. ISI Web of SCIENCE, 5 wg. <http://scholar.google.com/citations>) mgr Jakub Paś zastosował metody porównywania typu profil-profil do wykrywania i identyfikacji genów syntazy chalkonowej (CHS) z *Lupinus luteus*. Dopasowanie sekwencji odległych homologów i badania filogenetyczne syntazy chalkonowej z 54 gatunków roślin oraz odległych homologów z *Deinococcus radiodurans* wykazały, że enzym CHS z łubinu jest prawdopodobnie zakodowany przez co najmniej dwa różne geny, które wyewoluowały już około 16 milionów lat temu. Zastosowanie Metaserwera (Structure Prediction Meta Server) zaimplementowanego w BioInfoBanku pozwoliło na uliniwienie 52 sekwencji białek z bazy danych NCBI. Poza wysoce zachowawczymi aminokwasami z kieszeni katalitycznej enzymu (co stanowi podstawowe kryterium zaliczenia nowo odkrytego białka do rodziny białek CHS), wykryto 21 wspólnych aminokwasów poza centrum katalitycznym, co potwierdziło, że nowo odkryte sekwencje są funkcjonalnymi syntazami.

W publikacji 7. zatytułowanej “ Analysis of structure and function of tenascin-C”, (J Pas, E Wyszko, K Rolle, L Rychlewski, S Nowak, R Żukiel, J Barciszewski, **The international journal of biochemistry & cell biology** 2006, 38 (9): 1594-1602, IF 4,804 , ilość cytowań 19 wg. ISI Web of SCIENCE, 35 wg. scholar.google.com) za pomocą technik biochemicznych autorzy wykazali wyższy poziom tenascyny-C (Tn-C) w tkankach ludzkich nowotworów mózgu, jelita i piersi, i wskazali na możliwość zastosowania tego białka jako potencjalnego markera nowotworowego oraz celu nowych terapii lekowych. Tenascyna-C jest wielodomenowym białkiem macierzy zewnątrzkomórkowej. Autorzy publikacji jako pierwsi zastosowali do badania tego białka i jego genu analizę bioinformatyczną oraz modelowanie molekularne. Metody porównywania sekwencji typu profil-profil pozwoliły określić funkcjonalne elementy w dużym wielodomenowym białku, wszystkie zmienne części białka oraz jego izoformy. Do określenia granic domen białkowych Doktorant wykorzystał oprogramowanie Domain-Split. Dla identyfikacji domen homologów tenascyny-C zastosował bazę danych GRDB (Gene Relational Sequence Database). W trakcie tworzenia modeli domen białka Tn-C, dla znalezienia homologów o znanej strukturze trzeciorzędowej, bazę PDB przeszukiwano stosując jako niezależne kwerendy sekwencje każdej zidentyfikowanej domeny.

W publikacji 8. z „Listy...” zatytułowanej „Structure prediction, evolution and ligand interaction of CHASE domain”, (J Pas, M von Grotthuss, LS Wyrwicz, L Rychlewski, J

Barciszewski, **FEBS Letters** 2004, 576 (3): 287-290, **IF 3,843** , ilość cytowań 21 wg. ISI Web of SCIENCE, 29 wg. scholar.google.com), przedstawiono wyniki analizy ewolucji białek wiążących cytokininy (hormony roślinne uczestniczące w procesach wzrostu) i mechanizmy ich oddziaływania z ligandami. Oprogramowanie bazy danych GRDB użyto do identyfikacji odległych sekwencyjnie homologów domeny CHASE w białkach oddziałujących z cytokininami, w szczególności białka CRE1 z *Arabidopsis thaliana*. Wykonano model białka CRE1 i dokowanie molekularne *trans*-zeatyny i kinetyny dla zbadania zdolności wiązania ligandów przez to białko.

Indywidualny wkład Doktoranta do tych prac zbiorowych dotyczył koncepcji, analizy filogenetycznej, modelowania molekularnego, przewidywania elementów strukturalnych tenascyny-C, przewidywania struktury trzeciorzędowej, analizy motywów strukturalnych, analizy topologicznej białek, dokowania molekularnego. Uczestniczył również w opracowaniu wyników i w przygotowaniu manuskryptów. W dwóch z nich, szczególnie ważnych i bardzo licznie cytowanych, jest pierwszym autorem. Choć prace te ukazały się już wiele lat temu, o ich aktualności świadczy fakt, że nawet w trakcie pisania tej recenzji wzrastała liczba ich cytowań.

Publikacja p.t. „ORFeus: detection of distant homology using sequence profiles and predicted secondary structure” (K Ginalski, **J Pas**, LS Wyrwicz, M Von Grotthuss, JM Bujnicki, L Rychlewski, **Nucleic Acids Research** 2003, 31 (13), 3804-3807, **IF 6,575** , ilość cytowań 84 wg. ISI Web of SCIENCE, 127 wg. scholar.google.com), która powstała w Laboratorium Bioinformatyki w BioInfoBanku, jest istotnym osiągnięciem mgr Jakuba Pasia w zakresie rozwoju i wdrażania probabilistycznych metod porównawczych profil-profil do rozpoznawania podobieństwa białek. Opisany w niej w pełni zautomatyzowany serwer ORFeus udostępniony w ramach portalowego Metaserwera: Structure Prediction Meta Server (<http://BioInfo.PL/Meta/>), umożliwia analizę podobieństwa sekwencji białek. Zastosowany w nim algorytm autorstwa mgr Jakuba Pasia pozwala na użycie dowolnej metody przewidywania struktury drugorzędowej, generującej opis możliwej prawdopodobnej lokalnej struktury białkowej w postaci prawdopodobieństwa wystąpienia trzech form: alfa-helisy, beta-karty i pętli. Informacja o przewidywanej strukturze jest przechowywana w postaci profilu prawdopodobieństwa. Implementacja wykorzystująca oprogramowanie PSIPRED pozwala wykonać złożone analizy sekwencji generując meta profile o czułości podwyższonej względem innych metodologii. Doktorant deklaruje, że jego wkład w tym projekcie dotyczył przygotowania algorytmu i baz danych, przeprowadzenia testów porównawczych, opracowywania wyników oraz przygotowania manuskryptu.

W części IV Komentarza mgr Jakub Paś odwołuje się do jeszcze kilku innych swoich publikacji dotyczących wdrażania probabilistycznych metod porównawczych profil-profil do rozpoznawania pofałdowania białek (pozycja 5., 9, 15. i 17. z „Listy artykułów...”).

W publikacji 5. z „Listy...” zatytułowanej „Application of 3D Jury, GRDB, and Verify3D in fold recognition” (M Grotthuss, **J Paś**, L Wyrwicz, K Ginalski, L Rychlewski, **Proteins: Structure, Function, and Bioinformatics 2003**, 53 (S6): 418-423, **IF 4,313**, ilość cytowań 21 wg. ISI Web of SCIENCE, 45 wg. scholar.google.com) opisane jest zastosowanie narzędzi bioinformatycznych wymienionych w tytule do przewidywania struktur białek w konkursie CASP5. Autorzy podkreślają, że najistotniejszym czynnikiem prowadzącym do prawidłowego rozpoznawania zwojów było stosowanie bazy GRDB opartej na programie ORFeus jako instrumentu dla odnajdywania odległych homologów analizowanych białek.

W pozostałych omawianych w IV. części Komentarza publikacjach mgr Jakub Paś przedstawia implementacje algorytmów w narzędzia dostępne w serwisach bioinformatycznych, które zostały zrealizowane z jego udziałem: 3D-Hit (pozycja 9), serwis PDB PREVIEW (pozycja 15.), baza danych GRDB (pozycja 17. z „Listy...”).

O znaczeniu i aktualności metod stosowanych i wdrażanych przez mgr Jakuba Pasia świadczy pierwsza publikacja z „Listy...” p.t. „ELM: the status of the 2010 eukaryotic linear motif resource”, (**Nucleic Acids Research 2010**, 38, Database Issue D167-D180, **IF 8,055**, ilość cytowań 97 wg. ISI Web of SCIENCE, 142 wg. scholar.google.com). Mgr Jakub Paś jest w niej jednym z 23 autorów. Praca jest wynikiem współpracy 8 doskonałych grup naukowych z 6 krajów europejskich. Doktorant definiuje swój wkład do powstania tej tak często cytowanej publikacji opisującej bazę ELM (Eukaryotic Linear Motif resource) jako udział „przy opracowywaniu serwisu PDB BLAST przewidującego strukturę białek na podstawie sekwencji oraz opracowanie algorytmu do oznaczania zachowawczych motywów aminokwasowych”.

Na „Liście artykułów opublikowanych w czasopismach naukowych” jest jeszcze szereg publikacji dotyczących analizy sekwencji i przewidywania struktury białek, jednak mgr Paś nie odnosi się do nich w swoim komentarzu.

Podsumowując stwierdzam, że praca doktorska mgr Jakuba Pasia jest wartościowym przykładem zastosowania i rozwoju metod bioinformatycznych do przewidywania i modelowania struktur białek. Praca przedstawia oryginalne rozwiązanie sformułowanego celu naukowego za pomocą nowoczesnego i wciąż rozwijanego warsztatu i właściwie dobranych metod. Autor legitymuje się znaczącym dorobkiem publikacyjnym w czasopismach naukowych o wysokim współczynniku *impact factor*. Uważam, że mgr Jakub Paś mógł

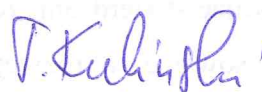
ubiegać się o nadanie stopnia doktora już osiem lat temu. Już wówczas jego dorobek publikacyjny był wystarczający, aby uzyskać ten stopień.

Oceniając stronę redakcyjną pracy muszę być bardzo krytyczny. Niestaranność polega nie tylko na licznych literówkach i niedbałości pod względem językowym. W opisach trzech zamieszczonych w Komentarzu rysunków pochodzących z publikacji Autora brakuje odwołania z jakich prac je zaczerpnął (rysunek 1. pochodzi z publikacji 7., rysunek 2 i 3 z publikacji 8. z „Listy artykułów...”). Opis rysunku 1. jest niepełny. W opisie rysunku 3. na stronie 18 autor odwołuje się do kolorów, tymczasem rysunek jest w odcieniach szarości. Ponadto rozdzielczość tego rysunku jest bardzo zła. Niestety kopia publikacji 8. zamieszczonej w części VIII. rozprawy jest również białą czarną. Występuje bałagan w odnośnikach referencyjnych w Komentarzu (np. na str. 15 referencje jak w publikacji 7., str. 1600).

Do Komentarza zakradły się też nieścisłości merytoryczne. Na przykład na str. 14 opisując zastosowanie metody porównywania typu profil-profil do wykrywania i identyfikacji genów syntazy chalkonowej, Autor pisze o uliniowieniu 57 sekwencji i wykryciu 21 wspólnych aminokwasów w centrum aktywnym, podczas gdy w publikacji źródłowej (7.) opisuje 52 sekwencje białkowe, które pozwoliły ustalić 21 wspólnych aminokwasów **poza** centrum aktywnym. Zachęcam Kandydata do większej staranności w przyszłości.

Opisane powyżej niedociągnięcia i błędy nie mają wpływu na moją pozytywną merytoryczną ocenę pracy. Chciałbym podkreślić, że mgr Jakub Paś realizował swoje projekty we współpracy z wieloma zespołami, w tym z zespołami eksperymentatorów. Wykazał, że potrafi samodzielnie rozwiązywać problemy i z sukcesem wykorzystywać swoją wiedzę w realizacji różnych projektów naukowych.

Wyrażam opinię, że przedstawiona mi do oceny rozprawa doktorska, ze wszech miar spełnia wszystkie wymogi określone w ustawie i stawiam wniosek o dopuszczenie mgr Jakuba Pasia do dalszych etapów przewodu doktorskiego.



Poznań, dnia 19 września 2013 r.

dr hab. Tadeusz Kuliński, prof. IChB PAN