

Data: Toruń, 15 lipca, 2023

L. Dz.

Prof. dr hab. Wiesław Nowak
Katedra Biofizyki
Instytut Fizyki
Uniwersytet M. Kopernika w Toruniu
87-100 Toruń, ul. Grudziądzka 5
(wiesiek@umk.pl)

RECENZJA rozprawy doktorskiej
Mgr. Carlosa Eduardo Sequiros Borja
pt.
“Ścieżki transportu ligandów w białkach”

Procesy życiowe są zjawiskami zachodzącymi w czasie i ich nieodłącznym elementem jest ruch molekuł. Komórki mają złożony metabolizm oparty na odpowiednio zsynchronizowanych przemieszczeniach biocząsteczek i mniejszych molekuł (lub jonów) zwanych często ligandami. Procesy te można badać doświadczalnie lub teoretycznie, np. rozwiązując komputerowo odpowiednie równania fizyczne. Transport ligandów jest zatem fundamentem w biologii i zasługuje na jak najpełniejsze zrozumienie. Praktycznie wszystkie skuteczne leki w swoim mechanizmie działania zawierają transport substancji aktywnych. Opis transportu w środowisku wodnym jest stosunkowo dobrze opracowany. Znacznie trudniej opisuje się ruch molekuł w heterogenicznych macierzach białkowych. Białka, zwłaszcza enzymatyczne, mają złożoną strukturę przestrzenną, co więcej, jest to struktura dynamiczna, czyli zmienia się w czasie. W białkach pojawiają się puste przestrzenie, dla prostoty zwane tutaj kanałami lub tunelami, które niekiedy pełnią krytyczne role funkcjonalne. Ma przykład, jeśli centrum katalityczne enzymu jest głęboko zanurzone w dużym białku globularnym, substrat musi pokonać niekiedy drogę kilku nanometrów by dotrzeć do tego centrum, „ulec” reakcji, a produkt musi opuścić białko, by enzym mógł uczestniczyć w kolejnym cyklu katalitycznym. Kinetyka tego transportu jest szczególnie ważna we właściwym funkcjonowaniu enzymu i podtrzymywaniu odpowiednich procesów życiowych.

Tematem przedłożonej mi do oceny rozprawy doktorskiej p. mgr. **Carlos* Eduardo Sequiros Borja** jest opracowanie efektywnych metod obliczeniowego wyznaczania wolnych przestrzeni (tuneli) w białkach i poznanie procesów transportowych ligandów (głównie wody) w kilku wybranych układach białkowych (głównie w hydrolazach). Temat rozprawy uważam za bardzo ważny, a cel doktoratu za trafnie wybrany. Nasza wiedza o tym jak przebiegają zjawiska transportowe jest wciąż mizerna, wiele narzędzi obliczeniowych opracowanych w rozlicznych laboratoria ma swoje ograniczenia, a pytania natury biologicznej w wielu wypadkach pozostają bez odpowiedzi. Znam, np. szereg enzymów, co do których nie ma pewności jak substraty i produkty „podróżują” wewnątrz białek. Problem ten jest jeszcze trudniejszy jeśli pomyślimy o kompleksach białko/RNA.

Chociaż nie ma w rozprawie tradycyjnej sekcji opisującej cele badań, po lekturze mogę stwierdzić, że Doktorant miał dwa cele:

1. Opracowanie i zbadanie lepszych wersji metod służących do badania pustych przestrzeni w białkach i ścieżek transportu.
2. Zastosowanie nowych metod obliczeniowych do zbadania kilku problemów biologicznych.

Tak rozumiane cele doktoratu zastały osiągnięte. Tytuł rozprawy jest niezwykle ogólny i tym samym mało „informatywny”, ale w zasadzie odpowiada jej zawartości.

Rozprawa została przygotowana na Wydziale Biologii, w Instytucie Biologii Molekularnej i Biotechnologii, pod kierunkiem dr hab. Jana Brezovsky’ego, prof. UAM. Praca doktorska ma formę cyklu publikacji (3 prace opublikowane w języku ang. w czasopismach międzynarodowych, jeden manuskrypt w formie preprintu opublikowanego w internecie) poprzedzonych Wprowadzeniem. Wprowadzenie, napisane samodzielnie przez doktoranta, składa się ze Wstępu i rozdziałów pt. „*Cavities in proteins and their study*” (7 str.), „*Ligand migration through molecular pathways*” (5 str) , Conclusions (1 str). Ta część zaopatrzona jest w 67 starannie i trafnie dobranych referencji literaturowych. Zasadnicze publikacje i Preprint zebrano w Dodatkach 1-4. Ostatnia część to komplet oświadczeń współautorów publikacji nt. ich wkładu. W świetle niedawnych interpretacji RDN nt. możliwości włączenia preprintu do treści rozprawy przedkładanej w formie cyklu prac, forma doktoratu nie budzi moich zastrzeżeń. Oświadczenia są jasne i pozwalają oszacować wkład Doktoranta w poszczególne artykuły.

Ocenę rozprawy ułatwi przytoczenie tych publikacji z dostępnej mi wersji elektronicznej:

P1. J. Brezovsky, A.S. Thirunavukarasu, B. Surpeta, **C.E. Sequeiros-Borja**, N. Mandal, D.K. Sarkar, C.J. Dongmo Fomthum, N. Agrawal, TransportTools: A library for highthroughput analyses of internal voids in biomolecules and ligand transport through them, *Bioinformatics*. 38 (2022) 1752–1753.

<https://doi.org/10.1093/bioinformatics/btab872>.

Impact Factor 2021: 6.931

MNiSW points 2022: 200

P2. **C. Sequeiros-Borja**, B. Surpeta, I. Marchlewski, J. Brezovsky, Divide-and-conquer approach to study protein tunnels in long molecular dynamics simulations, *MethodsX*. 10 (2023) 101968. <https://doi.org/10.1016/j.mex.2022.101968>.

Impact Factor 2023: N/A

MNiSW points 2023: 70

P3. K. Pakuła*, **C. Sequeiros-Borja***, W. Biała-Leonhard*, A. Paweła, J. Banasiak, A. Bailly, M. Radom, M. Geisler, J. Brezovsky, M. Jasiński, Restriction of access to the central cavity is a major contributor to substrate selectivity in plant ABCG transporters, *Cellular and Molecular Life Sciences*. 80 (2023) 105. <https://doi.org/10.1007/s00018-023-04751-6>.

Impact Factor 2021: 9.207

MNiSW points 2023: 140

P4. **C. Sequeiros-Borja**, A.S. Thirunavukarasu, C.J. Dongmo Fomthum, J. Brezovsky, Water will find a way: transport through narrow tunnels in hydrolases, *bioRxiv* 2023. doi: <https://www.biorxiv.org/content/10.1101/2023.05.24.542065v1>.

Impact Factor 2023: N/A

MNiSW points 2023: N/A

Jak widzimy, pan mgr **Carlos Eduardo Sequeiros Borja** jest pierwszym autorem w trzech artykułach, Jego wkład, zwłaszcza w pracach **P2** i **P4** jest wyraźnie dominujący. Chce tutaj podkreślić, że Doktorant nie tylko działał jako autor metody obliczeniowej, programista algorytmów czy skryptów, ale również jako biolog, interpretujący złożone procesy biofizyczne w świetle właściwości białek enzymatycznych i

ich mutantów. Publikacje **P1** i **P2** mają faktycznie charakter metodologiczny, ale **P3** i **P4** to są bardzo ładne przykłady wykorzystania metod obliczeniowych do interpretacji wyników pomiarów układów biologicznych. Artykuł **P1** ukazał się w czołowym czasopiśmie bioinformatycznym (krótka nota, ale z bardzo obszernymi suplementami), zaś artykuł **P3** w solidnym czasopiśmie międzynarodowym *Cellular and Molecular Life Sciences*. Warto zauważyć, że pan mgr Carlos Eduardo Sequiros Borja jest jeszcze współautorem czterech innych publikacji, (np. w *Brief. Bioinf.* z IF=14).

We Wprowadzeniu do rozprawy autor początkowo stara się zdefiniować główne pojęcia związane z pustymi przestrzeniami w białkach (Rys. 1) oraz z dwiema grupami metod pozwalających szacować położenie i rozmiary tych przestrzeni: (i) opartych na położeniu atomów białka oraz (ii) wykorzystujących śledzenie położenia małego obiektu sondującego, np. sfery o określonym promieniu (Rys. 2, str. 15). Podejście najczęściej używane w rozprawie to połączenie wyników CAVERA (typ (i)) oraz programu AQUA-DUCT (ii). Doktorant przyczynił się do końcowej fazy opracowania systemu zwanego „TransportTools” (publikacja **P1**). We Wprowadzeniu podane są dwa (słuszne) argumenty za celowością opracowania takiego narzędzia: 1. Automatyczne wyszukiwanie tuneli jest obiektywne, 2. Program dokonuje analiz znacznie szybciej niż „mędrca szkiełko i oko”, zostawiając tym samym czas badacza na inne pożyteczne aktywności. W programie TransportTools wykorzystano podejście „dziel-i-rządź”, tzn. zastąpiono długotrwałą analizę jednego wielkiego zbioru danych zestawem serii obliczeń wykonywanych dla małych podzbiorów. Wyniki można uzyskać kilka-kilkanaście razy szybciej, zaś ich jakość jest porównywalna, jeśli nie identyczna.

Kolejny rozdział wprowadzenia do rozprawy to opis badań transportu w białkach. Praca **P3** poświęcona jest badaniom doświadczalnym i obliczeniowym transportera ABC z *Medicago truncatula*. Białko to ułatwia transport przez błonę komórkową ważnych metabolitów, m.in., 4-coumarate i flawonoid likwiritigenina (ang. *liqiritigenin*). Pewne mutacje punktowe zasadniczo zmieniają zdolności transportowe tego białka. Doktorant opisuje główne wyniki modelowania wyjaśniające powody tej zmiany aktywności, szczegółowo zreferowane w publikacji **P3**, wróć do omawiania tych wyników później. Z kolei manuskrypt **P4** zawiera ciekawe analizy i rezultaty modelowania trzech enzymów z grupy halogenaz: dehalogenazę haloalkanową z *Rhodococcus rhodochrous* (Hal), hydrolazę epoksydową I z *Solanum tuberosum* (Epx), oraz lipazę z *Candida rugosa* (Lip). Obszerna analiza tuneli i ich parametrów doprowadziła autorów do ciekawego spostrzeżenia: postulują oni, że woda może pokonywać ścieżki w białkach charakteryzujące się niezwykle małym promieniem, przewężenia mniejsze niż typowa w modelowaniu średnica modelowej cząsteczki wody. Jeśli hipoteza ta znajdzie doświadczalne potwierdzenie, będzie to znakomity wkład do rozumienia transportu wody w białkach.

Moim zdaniem Wprowadzenie do publikacji P1-P4 dobrze spełnia swoją rolę. Odsyła do głównych pozycji literaturowych związanych z modelowaniem kanałów. Zwięźle prezentuje istotę nowo opracowanej metody obliczeniowej, zarysowuje problemy biologiczne analizowane dla transportera ABC i grupy hydrolaz. Doktorant podkreśla główne elementy nowości zawarte w publikacjach, w których jest współautorem. Wprowadzenie ułatwia zrozumienie istot rozprawy i dobrze spina w spójną całość cykl publikacyjny. Pewnym niedostatkiem jest wg mnie brak chociażby skrótowego opisu metod modelowania molekularnego. Opis funkcji biologicznych badanych w rozprawie białek też mógłby być obszerniejszy. Poza nazwami łacińskimi niczego nie dowiadujemy się o znaczeniu badanych organizmów. Ciekawy wątek mutacji E40G w ludzkiej hydrolazie epoksydowej, badany w **P4**, we Wprowadzeniu jest całkowicie pominięty. Te uwagi są oczywiście dyskusyjne i nie redukują znacząco mojej bardzo dobrej oceny tej części rozprawy. Tekst czyta się łatwo, język angielski nie jest może idealny, ale całkiem poprawny. Nieliczne rysunki (1-6) są dobrej jakości, a odnośniki podane poprawnie.

Omówię teraz kolejne artykuły składające się na rozprawę.

Publikacja P1 (w b. dobrym czasopiśmie *Bioinformatics*, 8 autorów, 2 strony), ma charakter softwarowej noty aplikacyjnej. Sygnalizuje powstanie programu TransportTools umieszczonego w domenie publicznej. Program ten, napisany w Pythonie3 pozwala na integrację danych uzyskiwanych w długich i licznych symulacjach. Podnosi na wyższy poziom możliwości statystycznej analizy wyników modelowania transportu ligandów w białkach. Chociaż program oparty jest głównie na dwóch standardowych narzędziach CAVER i Aqua-Duct, to jestem przekonany, że będzie on przydatny w wielu laboratoriach zajmujących się modelowaniem biologicznym. Cenną cechą programu jest jego wydajność, Autorzy pokazują, że możliwa jest analiza nawet 5 mln ścieżek obliczonych dla sporych enzymów złożonych z 500 aminokwasów. Publikacja jest oczywiście ograniczona objętościowo wymaganiami wydawcy, ale uzupełniona jest potężnym zbiorem dodatków, które pozwalają zrozumieć istotę pomysłu na tę metodę i zaimplementować lokalnie ów software. W **P1** autorzy przedstawiają też odwołanie do wyników analiz DEhA i Lin B (hydrolazy) oraz pewnych ich wariantów z zamkniętymi tunelami. W moim ocenie jest to bardzo potrzebne i przydatne narzędzie, stosowane będzie nawet w mojej grupie badawczej do analizy NHaz. Wkład Doktoranta do tej publikacji nie był dominujący, polegał na testowaniu systemu i implementacji algorytmów do nakładania cząsteczek oraz wizualizacji znalezionych klastrów tuneli.

Z kolei w pracy **P2** wkład Doktoranta był bardzo istotny, obejmował nie tylko pomysł metody ale także wykonanie i analizę obliczeń i wyciąganie wniosków z uzyskanych wyników. Istotą pracy było pokazanie, że wyniki analiz tuneli realizowane klasycznym programem Caver v3.02 i nowo opracowaną, szybką metodą typu *divide-and-conquer*, czyli przy pomocy biblioteki TransportTools v0.9.3, jest taki sam. Zademonstrowano to przy pomocy wyników wyznaczania tuneli dla 100 ns trajektorii dehalogenazy DhaA. W opisie postępowania w przypadku metody z TransportTools zaciekały mnie dwie sprawy:

- a) W sekcji „*Tunnel Calculations*” autorzy przytaczają szereg parametrów, np. „*clustering threshold 3.0*” przyjętych w obliczeniach, ale nie podają żadnego uzasadnienia dla takich czy innych wartości, nie dyskutują też zupełnie wpływu doboru tych parametrów na jakość otrzymywanych wyników. Skąd użytkownik TransportTools ma mieć pewność, że przy innym wyborze parametrów zgodność obu schematów obliczeniowych będzie równie dobra?
- b) Przy łączeniu wyników uzyskanych dla fragmentów trajektorii („*slices*”) autorzy proponują zastosowanie filtrowania połączonego z renumeracją plików. Nie znalazłem przekonujących danych, że ta dość arbitralna procedura filtrowania daje jakieś poważne oszczędności czasu działania programów, ani też argumentów, że filtrowanie nie wprowadza żadnych artefaktów do obliczanych klastrów z tunelami. Może w czasie obrony uda się przedyskutować te techniczne kwestie?

W publikacji **P3** stykamy się już z analizą poważnego problemu biologicznego. Artykuł wyjaśnia skąd się bierze selektywność względem substratów w roślinnym białku ABCG. Modelowym obiektem badań był transporter ABCG46 z rośliny *Medicago truncatula*. Roślina ta uprawiana w Australii na paszę dla zwierząt jest bliską krewną popularnej lucerny siewnej (*Medicago sativa*) i należy do rodziny bobowatych (dawniej rośliny motylkowe). Jak dotąd nie ma doświadczalnej struktury przestrzennej tego białka błonowego, zatem do pracy zaproszono sztuczną inteligencję. Autorzy użyli serwera AlphaFold2 by zbudować strukturę 3D, wykorzystali pewne informacje ze znanych struktur, np. ABCG2 Homo sapiens, aby zbudować w miarę realistyczny model pełnoatomowy. Następnie, stosując dość wyrafinowane metody inicjacji obliczeń uzyskali 5x400ns trajektorii MD. Trzeba przypomnieć, że

transportery ABC, z racji swojej funkcji, mają na ogół dwie odmienne konformacje OF (*outward facing*) oraz IF (*inward facing*). Uzyskanie modelu IF było możliwe dzięki żmudnemu zastosowaniu techniki „*umbrella sampling*”. W publikacji jest stwierdzenie, że obliczono profil energii swobodnej (tzw. *potential of mean force*) na przejście z formy OF do IF, ale nigdzie nie znalazłem stosownego wykresu ani danych liczbowych. Przy pomocy CAVERA, stosując dość złożony proces heurystyczny, wyznaczono najbardziej prawdopodobną ścieżkę wejścia ligandów do wnętrza białka transportera. Kolejnym krokiem było dokowanie czterech metabolitów, ich nazwy są tak skomplikowane, że je pominię w tej recenzji. W rezultacie obszernych obliczeń stwierdzono, iż badane białko ma dużą wewnętrzną wnękę połączoną z otoczeniem tylko przy pomocy kanałów, które są nietrwałe – powstają i znikają. Następnie zbadano sytuację kanałów w wariacie, gdzie F562 zastąpiono innym aminokwasem (Tyr, Leu Ala). W wyniku starannych analiz szerokości tuneli i innych parametrów strukturalnych stwierdzono, że możliwości transportu dużych ligandów są mocno ograniczone w wariantach F562Y i F562A. Zarówno nie omawiane we Wprowadzeniu dane doświadczalne, jak i wyniki modelowania doprowadziły do ciekawej konkluzji: adaptacja białek ABCG związana z presją ewolucyjną u roślin doprowadziła do zmienności aminokwasowej na pewnych kluczowych pozycjach (jak F562) co pozwala tym białkom transportować selektywnie ważne metabolity.

Praca **P3** jest ładnym przykładem owocnej współpracy eksperymentatorów (zespół prof. Michała Jasińskiego) i teoretyków – tj. bioinformatyków strukturalnych. Rola Doktoranta w przeprowadzeniu obszernych badań modelowych, czasochłonnych analiz i dyskusji wyników była bardzo duża. W publikacji przyjęto szereg arbitralnych założeń, to może budzić wątpliwości, ale warto wiedzieć, że takie założenia w tej branży są powszechne i bez nich nie mielibyśmy żadnych danych z obliczeń. Rezultaty mogą być krytykowane (wyjście ze struktury homologicznej, słaba struktura typu „open”, dlaczego nie zastosowano metody SMD lub TMD do otrzymania formy open?, itd.) i weryfikowane niezależnymi metodami, ale jest to niewątpliwie ważny i oryginalny wkład do badań biologicznych na poziomie molekularnym, dokonany dzięki pracy, m.in., p. mgr Carlosa Eduardo Sequiros Borja.

Dla mnie jako dla biofizyka najbardziej intrygujący był Dodatek 4, zawierający manuskrypt **P4**. Można sparafrazować główny wynik tego artykułu w żartobliwy sposób: „w biologii wszystko jest możliwe, nawet przejście wielbłąda przez ucho igielne”. Autorzy, znów przy dominującym udziale Doktoranta, przeanalizowali ogromną liczbę danych z 5 mikrosekund symulacji kilku znanych hydrolaz (haloalkane dehalogenase (Hal) z *Rhodococcus rhodochrous*, epoxide hydrolase I (Epx) z *Solanum tuberosum*, lipazy Lip z *Diutina rugosa* (previously *Candida rugosa*) oraz ludzkiej hydrolazy epoksydowej (hEpx) i doszli do wniosku, iż woda może się przeciskać (czasami) przez kanały białkowe o średnicy mniejszej niż zwykle zakładana średnica molekuly wody (2.8 Ang.). Jeśli to się potwierdzi, obserwacja ta będzie miała duże konsekwencje w opisie transportu wody w białkach i w rozumieniu mechanizmów reakcji enzymatycznych. Autorzy stawiają hipotezę, że możliwości pokonywania przez wodę barier energetycznych wywoływanych oddziaływaniami Van Der Waalsa wynika z tego, że molekuly wody znajdujące się w tych przewężeniach mają szczególnie dużo wiązań wodorowych z atomami szkieletu białka. Dane komputerowe uzyskane przy pomocy TransportTools potwierdzają to. Moje wątpliwości budzi podanie na Rys. 4b obrazka z molekułą wody tworzącą 5 wiązań wodorowych i dodanie takiego „boxu” w Rys. 3. Standardowa woda ma dwa atomy będące wodorem (tlen jako donator) i dwie pary elektronowe w tlenie (akceptor) więc oczekiwana liczba standardowych wiązań wodorowych nie powinna przekraczać 4. Oczywiście są też tzw. „*bifurcated H-bonds*”, czy może oddziaływania elektrostatyczne o charakterze przypominających wiązania wodorowe, jednak w omawianym manuskrypcie nie znalazłem komentarza na ten temat. W rozdziale tym nie ma też dyskusji alternatywnych wytłumaczeń tej obserwacji. A może fakt, że w przewężeniu mamy tyle wiązań

wodorowych oznacza, że molekula wody po prostu blokuje ten kanał? Na ile rozumiem procedurę odkrywania kanału, nie ma ona elementu sprawdzającego czy konkretna cząsteczka w czasie symulacji faktycznie „podróżuje kanałem” od wejścia do wyjścia, wydaje się, że program tylko sprawdza fakt obecności molekuł w danych współrzędnych ścieżki. Obecność wody w przewężeniu a transport tej wody przez to przewężenie, to jednak nie jest to samo. Dobrze byłoby zaproponować doświadczenia mutacyjne sprawdzając tę ciekawą hipotezę. Omawiając pracę **P4** pragnę podkreślić, że uzyskanie prezentowanych wyników (liczne suplementy) wymagało od Doktoranta b. dobrego opanowania warsztatu obliczeniowego i wiele żmudnej pracy nad zebraniem i redukcją danych. Artykuł prezentuje oryginalną hipotezę i może wywołać sporą dyskusję wśród specjalistów.

Strona redakcyjna rozprawy doktorskiej jest bez uwag z mojej strony, widać staranność i wysiłek by jak najklarowniej przedstawić treści.

KONKLUZJA

Stwierdzam, że przedłożona mi do oceny rozprawa doktorska mgr **Carlota Eduardo Sequiros Borja** stanowi rozwiązanie problemu naukowego związanego z naukami biologicznymi. Zarówno rozprawa jak i pozostały dorobek publikacyjny dowodzą, że doktorant potrafi samodzielnie prowadzić badania naukowe z wykorzystaniem istniejących technik komputerowych i wielkoskalowych symulacji biomolekuł oraz potrafi dać wkład do opracowywania nowych metod badawczych i interpretacji wyników doświadczeń na materiale biologicznym. Rozprawa dowodzi, że orientuje się On bardzo dobrze we współczesnej związanej z projektem doktorskim literaturze naukowej.

Recenzowana praca spełnia ustawowe (tj. Ustawa o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki, Dz. U. z 2022 poz. 574 z późn. zm.) **oraz zwyczajowe wymagania stawiane rozprawom doktorskim. Z przekonaniem wnoszę o dopuszczenie pana magistra Carlota Eduardo Sequiros Borja do dalszych etapów postępowania prowadzącego do uzyskania stopnia naukowego doktora nauk ścisłych i przyrodniczych w dyscyplinie nauki biologiczne.**



Wiesław Nowak, prof. zw.