

ul. Wita Stwosza 63, 80-308 Gdańsk  
tel. 58 523 5124, fax 58 523 5012, email: adam.liwo@ug.edu.pl

Gdańsk, 23 maja 2019

**Recenzja rozprawy doktorskiej  
pani magister Aleksandry Elżbiety Badaczewskiej-Dawid  
pt. „Modelowanie struktury i dynamiki białek z użyciem modeli  
gruboziarnistych o różnej rozdzielczości”**

Białka mają fundamentalne znaczenie w funkcjonowaniu organizmów żywych. Ich funkcje zależą od dwóch czynników (1) unikalnej trójwymiarowej struktury oraz (2) dynamiki zmian konformacyjnych. Cechy (1) i (2) tworzą zresztą jedynie punkty odniesienia; dynamiczne zmiany konformacyjne w toku zachodzenia procesów biologicznych są często na tyle duże jak chociażby w przypadku chaperonów molekularnych czy cytochromu, że mówi się o różnych formach tego samego białka. W przypadku białek plastycznych, kompletna struktura jest natomiast uzyskiwana dopiero w miejscu docelowym. Nawet w przypadku białek o określonej strukturze metody eksperymentalne nie nadążają za jej wyznaczeniem z uwagi na długi czas i duże koszty przede wszystkim pracy ludzkiej, co przekłada się na fakt, że struktury zostały eksperymentalnie określone jedynie dla ułamka procenta znanych sekwencji. Doskonalenie zaawansowanych metod teoretycznych modelowania struktur i dynamiki białek jest zatem niezbędne i to nie tylko w celach poznawczych lecz również jak najbardziej praktycznych, przede wszystkim związanych z badaniem molekularnych mechanizmów chorób (zarówno egzo- jak i endogennych) a dalej projektowania skutecznych leków i terapii. Z uwagi na duży rozmiar badanych układów dobrym rozwiązaniem jest modelowanie gruboziarniste, umożliwiające rozszerzenie skali rozmiaru i czasu analizowanych układów o kilka rzędów wielkości. Temu zagadnieniu, a w szczególności opracowaniu nowego "agresywnie" gruboziarnistego modelu SURPASS, jest poświęcona recenzowana rozprawa doktorska. Rozprawa została wykonana w Pracowni Teorii Biopolimerów Wydziału Chemii Uniwersytetu Warszawskiego pod kierunkiem dra hab. Dominika Gronta, który jest wychowankiem prof. Andrzeja Kolińskiego, pioniera symulacji struktur i dynamiki białek w Polsce.

Rozprawa doktorska p. mgr Aleksandry Badaczewskiej-Dawid składa się ze streszczenia w języku polskim i w języku angielskim (1 strona każde), słowniczka skrótów (3 strony), wstępu (3 strony), celów pracy (1 strona), przeglądu literaturowego (38 stron), opisu wyników pracy badawczej (38 stron), podsumowania (3 strony), bibliografii (212 pozycji) oraz dodatku złożonego dwóch tabel, wykazu prac stanowiących podstawę rozprawy oraz innych prac, których Kandydatka jest współautorką, z 2 prac oryginalnych gdzie Kandydatka jest pierwszą autorką oraz 2 artykułów przeglądowych (wszystkie ujęte we wspomnianych wykazie) oraz dwóch tabel stanowiących materiały pomocnicze. Wymieniony w wykazie publikacji dorobek Kandydatki składa się z 7 prac już opublikowanych i 3 wysłanych do redakcji. W przypadku

kandydatki do stopnia doktora nauk chemicznych jest to zdecydowanie wyróżniający dorobek naukowy.

Rozprawa doktorska p. Aleksandry Badaczewskiej-Dawid jest napisana bardzo dobrze i ciekawie. Czyta się ją prawie jak książkę przygodową, od której trudno się oderwać z ciekawości co będzie dalej. Piszę to ponieważ w tym roku mija 30 lat odkąd zacząłem się zajmować się modelowaniem gruboziarnistym białek i trudno mnie zaciekać czymkolwiek w tej dziedzinie. Kandydatce to się udało.

We wstępie Kandydatka krótko opisuje znaczenie modelowania struktur i dynamiki białek i ich kompleksów oraz wymienia i streszcza metody tego modelowania, zarówno oparte na wiedzy zaczerpniętej z baz danych strukturalnych, jak i na prawach fizyki. Zwraca również uwagę na kluczowe znaczenie efektywnego próbkowania przestrzeni konformacyjnej. Po wstępie definiuje cele pracy: opracowanie gruboziarnistego modelu SURPASS białek, przetestowanie go w odtwarzaniu okołonatywnych struktur znanych białek oraz ich dynamiki, jak również opracowanie algorytmu odbudowy wymodelowanych struktur do reprezentacji o większej rozdzielczości aż do pełnoatomowej włącznie. Po lekturze celu pracy nasuwa się uwaga, że tytuł jest co prawda adekwatny do treści rozprawy, wydaje się jednak zbyt szeroki i sugeruje, że Kandydatka jedynie wykorzystwała istniejące modele gruboziarniste do swoich badań podczas, gdy zrobiła znacznie więcej.

Wstęp literaturowy otwiera omówienie eksperymentalnych metod określania struktur białek: krystalografii rentgenowskiej, łącznie z najnowszym jej wariantem czyli krystalografią femtosekundową, spektroskopii magnetycznego rezonansu jądrowego (NMR), kriomikroskopii elektronowej (cryo-EM) oraz spektrometrii mas dla reakcji wymiany proton-deuter. Ta część jest napisana bardzo jasno i zwięźle i jest poparta adekwatnymi odsyłaczami literaturowymi. Tym niemniej brakuje tutaj wzmianki o metodach małokątowego rozpraszania promieni X (SAXS) i neutronów (SANS) oraz tak ostatnio rosnącej pod względem znaczenia i popularności metody sieciowania chemicznego połączonego ze spektrometrią mas (XLMS). Metody te nie dają co prawda samej pełnej struktury ale w wielu wypadkach strukturę można przy ich pomocy uzyskać w połączeniu z zaawansowanymi metodami modelowania molekularnego, co pokazano chociażby w ostatnich eksperymentach CASP. Ponadto, w przypadku metody NMR należałoby wspomnieć o poważnym problemie związanym z niejednoznacznością przypisaniem sygnałów co powoduje czasami możliwe przypisanie danego piksu NOE do więcej niż 100 alternatywnych par protonów. Na ten problem zwracał uwagę Gaetano Montelione w trzech ostatnich eksperymentach CASP.

Kolejne 2 podrozdziały wstępu literaturowego są poświęcone metodom przewidywania struktur białek, odpowiednio poprzez modelowanie porównawcze oraz "swobodnego modelowania", adekwatnego w przypadku białek o sekwencjach zbyt mało podobnych do tych odpowiadających białkom obecnym w bazach danych strukturalnych. Również w tym drugim wypadku Kandydatka skupia się wyłącznie na metodach opartych na wiedzy, głównie składania struktury z fragmentów z baz danych zainicjowanej wiele lat temu przez grupę Davida Bakera. Nie omawia metod opartych na fizyce oddziaływań co jest jednak zrozumiałe ponieważ jak na razie metody te, oprócz jednostkowych przypadków niezwykle trafionych przewidywań, nie awansowały do grona najbardziej skutecznych podejść. W kolejnym podrozdziale Kandydatka umieszcza również metody obliczeniowe w kontekście modelowania dynamiki białek, jeszcze raz podkreślając konieczność modelowania gruboziarnistego. Świetną ilustracją przedstawionych wywodów jest rys. 3 pokazujący wyraźnie, że modelowanie na poziomie rozdzielczości atomowej długo (a może i nigdy) nie osiągnie skal czasowych procesów biologicznych.

Kolejne podrozdziały traktują o reprezentacji łańcucha polipeptydowego: od pełnoatomowej po reprezentacje gruboziarniste o różnym stopniu uproszczenia. Kandydatka omawia pokrótce model CABS opracowany w macierzystej pracowni, UNRES, ROSETTA, jak również modele oparte na strukturze: sieci elastycznych i modele Go. Zwraca uwagę, że jakkolwiek modele takie jak CABS i UNRES mają już silnie zredukowaną reprezentację łańcucha polipeptydowego dalsza redukcja, a co za tym idzie zwiększenie skali czasu i wielkości, są możliwe. Na zakończenie tej części omawia popularne metody stosowane do odbudowy łańcucha polipeptydowego.

W kolejnych podrozdziałach Kandydatka omawia pola siłowe stosowane w modelowaniu białek, pracujące w reprezentacji pełnoatomowej i gruboziarnistej. Zwraca uwagę na konieczność uwzględnienia rozpuszczalnika (jawnie lub w ujęciu średniopolewym) oraz na to, że stabilność białek jest jedynie marginalna i dlatego należy włożyć bardzo dużo pracy i uwagi w opracowanie i parametryzację efektywnej funkcji energii. Autorka zwraca uwagę na dwa typy potencjałów: fizyczne i statystyczne (te ostatnie wyprowadzone w oparciu o zastosowanie zasady Boltzmanna do rozkładów parametrów geometrycznych z baz danych strukturalnych). Ta część nie jest moim zdaniem tak dopracowana jak poprzednie. Przede wszystkim nie jest podana zasada tworzenia fizycznych pól siłowych. W przypadku pól pełnoatomowych formalnie tworzy się je poprzez parametryzację powierzchni energii potencjalnej (PES) w przybliżeniu Borna-Oppenheimera molekuł odpowiadających fragmentom tworzących analizowane układy (o PES Autorka wspomina jednak wcześniej). Te powierzchnie odpowiadają uśrednieniu względem elektronowych stopni swobody. Należało by w rozdziale 2.4.1 pokrótce scharakteryzować sens fizyczny poszczególnych wkładów (np. wkłady torsyjne są faktycznie wkładami korelacyjnymi, które można otrzymać z rachunku zaburzeń drugiego rzędu). Autorka nie podaje w ogóle fizycznych podstaw tworzenia potencjałów gruboziarnistych. Tymczasem taka podstawa istnieje: potencjały gruboziarniste wywodzą się z potencjałów średniej siły badanych układów. Tak jak w modelach pełnoatomowych można te potencjały rozbić na wkłady pochodzące od fragmentów a następnie wyprowadzić odpowiednie wyrażenia i sparametryzować je. Jest to o wiele trudniejsze zadanie niż to dla potencjałów pełnoatomowych, tym niemniej do zrealizowania [zob. np. A. K. Sieradzan, M. Makowski, A. Augustynowicz, A. Liwo, J. Chem. Phys., 146, 124106 (2017) i G.S. Ayton, W.G. Noid, G.A. Voth, 32, 929 (2007)]. Tymczasem z wywodów przedstawionych przez Kandydatkę można odnieść wrażenie, że potencjały gruboziarniste są zawsze potencjałami statystycznymi. Następnie Autorka omawia metody przeszukiwania przestrzeni konformacyjnej białek, w szczególności dynamikę molekularną oraz dynamikę Monte Carlo, w tym dynamikę i dynamikę Monte Carlo z wymianą replik.

Po tej części ogólnej Kandydatka przechodzi do bardziej szczegółowego opisu modeli gruboziarnistych powstałych bądź używanych w macierzystej pracowni: CABS, ROSETTA oraz SURPASS. Moim zdaniem opis modelu SURPASS, powstałego przecież w ramach recenzowanej rozprawy doktorskiej, powinien znaleźć się w części badań własnych a nie w części literaturowej.

Opis badań własnych Autorki rozpoczyna się wyszczególnieniem zbiorów struktur eksperymentalnych i symulowanych używanych w testach (było ich 6) oraz zdefiniowaniem miar podobieństwa do struktury natywnej oraz fluktuacji, jak również innych używanych wielkości. W kolejnym rozdziale oraz załączonej pracy P1 Kandydatka opisuje wyniki symulacji białek z użyciem opracowanego modelu SURPASS. Model ten jest oparty na innej koncepcji niż CABS, UNRES, ROSETTA czy MARTINI gdzie jednostkami są ustalone fragmenty reszt aminokwasowych niezależnie od struktury drugorzędowej. W modelu SURPASS definicja jednostki zależy od struktury drugorzędowej, której są 3 rodzaje:

helikalna (H), nić (S) oraz nieuporządkowana (C). Centrum jest zlokalizowane mniej więcej w środku między 4 atomami węgla  $\alpha$ , przy czym jego współrzędna jest "ruchomą średnią" położenia tych atomów. Naturalnie, w tej sytuacji integralną częścią danych do symulacji jest struktura drugorzędowa, co wskutek wysokiej dokładności metod przewidywania tej struktury nie stanowi większego problemu. Nie występują tutaj centra łańcuchów bocznych jako samodzielne jednostki. Autorka wyprowadziła statystyczne potencjały krótko- i dalekozasięgowe, łącznie z potencjałami oddziaływania między centrami, które jednak nie zależą od sekwencji. Jedynie pomiędzy centrami S występują kierunkowe a przez to organizujące strukturę wiązania wodorowe. Pomimo tak drastycznych uproszczeń reprezentacji oddziaływań, model jest w stanie zlokalizować w symulacjach bez ograniczeń struktury bliskie natywnym z całkiem dobrą rozdzielczością. Model nie umożliwia jednak na ogół rozróżnienia w oparciu o energię struktur nienatycznych od bliskich natywnym, w szczególności struktur okołonatywnych od ich topologicznych odbić lustrzanych.

W kolejnej rozprawie oraz w załączonej pracy P2 Autorka zbadała wpływ założonej struktury drugorzędowej na jakość wyników. Struktura drugorzędowa była wzięta z plików PDB, wygenerowana algorytmem DSSP, przewidziana algorytmem PSIPRED lub przypadkowa (COIL). Jak należało się spodziewać, to ostateczne założenie nie doprowadziło w ogóle do pojawienia się struktur okołonatywnych w symulacjach, natomiast użycie struktury przypisanej jak w plikach PDB lub wyliczona algorytmem DSSP dawało dobre wyniki. Nieco gorsze ale cały czas bardzo dobre wyniki dawało użycie struktury drugorzędowej wyznaczonej algorytmem PSIPRED. Autorka zbadała również wpływ więzów na odległości z pomiarów NOE na symulowane struktury. Stwierdziła, że jakość symulowanych struktur znacznie wzrasta. Nadal jednak nie jest możliwe odróżnienie struktury okołonatywnej od topologicznego odbicia lustrzanego; aby to zrobić konieczne jest użycie modeli o większej rozdzielczości, np. CABS.

W kolejnej części Kandydatka porównała zdolność modelu SURPASS do modelowania dynamiki białek w otoczeniu struktur natywnych z innymi podejściami: dynamiki molekularnej pełnoatomowej, CABS-Flex oraz sieci elastycznych (ENM). Odniesieniem był zbiór konformacji wiązki wyznaczonej metodą NMR. Profile fluktuacji otrzymane z modelu SURPASS okazały się w przybliżeniu zgodne z profilami wyznaczonymi z wiązek konformacji wyznaczonych metodą NMR, jakkolwiek w gorszej zgodności niż w przypadku mającego wyższą rozdzielczość modelu CABS (lecz lepszej niż w przypadku metody sieci elastycznych).

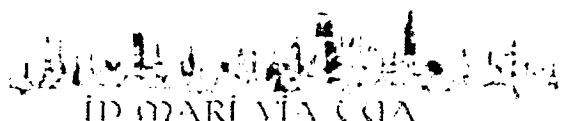
W ostatniej części rozprawy Kandydatka zajęła się odbudową łańcucha w reprezentacji SURPASS do śladu atomów węgla  $\alpha$ , z którego można już łatwo odbudować reprezentację pełnoatomową przy użyciu już opracowanych algorytmów. Odbudowa śladu  $C^\alpha$  z reprezentacji SURPASS jest zagadnieniem trudnym. Autorka stworzyła w tym celu bibliotekę fragmentów (SURELib), które wpasowuje się następnie w łańcuch w reprezentacji SURPASS. W przypadku konformacji w reprezentacji SURPASS wygenerowanych ze struktur PDB dokładność odbudowy wynosi poniżej 1 Å, natomiast odbudowa struktur otrzymanych w wyniku symulacji w modelu SURPASS wywołuje co prawda pogorszenie zgodności ze strukturą natywną o 0.45-1.16 Å, jednak najniższe RMSD wynosi dla większości przypadków poniżej 5 Å (dodatkowa tabela 6.8). Algorytm odbudowy odtwarza ponadto poprawnie lokalną geometrię łańcucha, w szczególności odległość między kolejnymi atomami  $C^\alpha$ . Algorytm odbudowy domyka wieloskalowość opracowanego przez Kandydatkę podejścia do modelowania struktur białek.

Na zakończenie dwie drobne uwagi odnośnie zauważonych błędów i niezręczności stylistycznych: w sekcji 2.3.2 jest cytowany odsyłacz 145, który miał być zapewne odsyłaczem do pionierskiej pracy Warshella i Levitta z roku 1976 (*J. Mol. Biol.*, **103**, 227) na temat ich gruboziarnistego modelu zwijania białek, natomiast jest odsyłaczem do pracy innych autorów na temat potencjałów typu Go, w którym nie ma wzmianki o pracach Levitta i Warshella. W całym tekście słowo "rysunek" i "tabela" przy odwołaniach do rysunków i tabel występuje w mianowniku co jest zapewne wynikiem stosowania automatycznego systemu odnośników. Nie ma w załącznikach pracy P3. Rozumiem, że ta praca została dopiero złożona w redakcji czasopisma a w załącznikach są tylko prace już opublikowane jednak należało albo dołączyć manuskrypt albo ograniczyć listę prac opublikowanych dołączonych do rozprawy do czterech. Są to oczywiście uwagi drobne niemające wpływu na moją bardzo pozytywną ocenę recenzowanej rozprawy.

Podsumowując, rozprawę doktorską p. mgr Aleksandry Badaczewskiej-Dawid uważam za wybitną. Autorka opracowała od podstaw, zaimplementowała, sparametryzowała i przetestowała z sukcesem unikalny i nietrywialny gruboziarnisty model białek, jak również opracowała procedurę odbudowy łańcucha pełnoatomowego z reprezentacji zredukowanej. Była to tytaniczna praca, wystarczająca moim zdaniem na bardzo dobrą habilitację. Ponadto opracowany model jest rozwojowy i po wprowadzeniu specyficznych oddziaływań między łańcuchami bocznymi i średniopoloowych wkładów opisujących chiralność może łatwo zyskać pełną zdolność predykcyjną. Cele rozprawy zostały w pełni zrealizowane. Wymienione powyżej uwagi krytyczne, z których większość ma charakter dyskusyjny, nie umniejszają w żadnym stopniu jej wartości. Rozprawa spełnia z należytą precyzją wymagania stawiane rozprawom doktorskim przez Ustawę z dnia 14 marca 2003 roku o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. nr 65, poz. 595 z późn. zm.), jak również zwyczajowe standardy stawiane rozprawom doktorskim w dziedzinie nauk przyrodniczych i ścisłych. Dlatego z pełnym przekonaniem zwracam się do Wysokiej Rady Wydziału Chemii Uniwersytetu Warszawskiego o dopuszczenie p. mgr Aleksandry Badaczewskiej-Dawid do dalszych etapów przewodu doktorskiego. Biorąc pod uwagę wysoki ładunek nowości naukowej, ogrom i wysoki profesjonalizm wykonanej pracy, znakomitą jakość wyników oraz ich znaczenie dla modelowania białek wnoszę o wyróżnienie rozprawy. Szczegółowe uzasadnienie wniosku o wyróżnienie stanowi załącznik do niniejszej recenzji.



prof. dr hab. Józef Adam Liwo



ul. Wita Stwosza 63, 80-308 Gdańsk  
tel. 58 523 5124, fax 58 523 5012, email: adam.liwo@ug.edu.pl

Gdańsk, 23 maja 2019

**Uzasadnienie wniosku o wyróżnienie rozprawy doktorskiej  
pani magister Aleksandry Elżbiety Badaczewskiej-Dawid  
pt. „Modelowanie struktury i dynamiki białek z użyciem modeli  
gruboziarnistych o różnej rozdzielczości”**

W ramach swojej rozprawy doktorskiej Kandydatka stworzyła, zaimplementowała i przetestowała unikalny model gruboziarnisty do symulacji białek, który może doprowadzić do kolejnego przełomu w zasięgu symulacji układów o znaczeniu biologicznym. Wykonała samodzielnie ogromną pracę podczas, gdy tworzenie, parametryzację i testowanie modelu gruboziarnistego zwykle realizuje sztab ludzi już co najmniej ze stopniem doktora. Badania do rozprawy doktorskiej przeprowadziła a wyniki opracowała bardzo rzetelnie i profesjonalnie. Jej rozprawę zaliczam do grona 5% najlepszych rozpraw doktorskich, które recenzowałem. Materiał jej rozprawy wystarczyłby moim zdaniem na bardzo dobrą habilitację.

O wysokiej jakości pracy Kandydatki świadczy Jej dotychczasowy dorobek naukowy, w którym zawierają się 4 dotychczas opublikowane prace oryginalne w bardzo dobrych czasopismach z listy filadelfijskiej oraz 1 praca wysłana z materiału samej rozprawy, jak również 3 inne prace opublikowane i 2 wysłane do czasopism z listy filadelfijskiej.

W świetle przedstawionych wywodów a także argumentów przytoczonych w mojej recenzji rozprawy doktorskiej p. mgr Aleksandry Elżbiety Badaczewskiej-Dawid, z pełnym przekonaniem wnoszę do Wysokiej Rady Wydziału Chemii Uniwersytetu Warszawskiego o wyróżnienie jej rozprawy doktorskiej.

  
prof. dr hab. Józef Adam Liwo