

Data: Toruń/Kozi Gród, 15 maja, 2019

L. Dz.

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

**RECENZJA rozprawy doktorskiej
Mgr. Aleksandry E. Badaczewskiej-Dawid pt.**

“Modelowanie struktury i dynamiki białek z użyciem modeli gruboziarnistych o różnej skali rozdzielczości”

Biopolimery, a spośród nich białka, zajmują szczególną pozycję w chemii, biologii i biofizyce. W laboratoriach na całym świecie od dziesiątek lat trwają intensywne badania doświadczalne i teoretyczne tych fascynujących układów. Zgromadzono ogromną wiedzę na temat budowy, struktury i funkcji białek, jednak wiele zagadek i problemów pozostaje nierozwiązanych. Duży wkład, mierzony np. liczbą publikacji w czołowych periodykach naukowych, w poznanie białek wnoszą metody komputerowe. Dzięki rozwojowi technologii obliczenia są teraz coraz szybsze i coraz tańsze, zatem można efektywnie modelować nie tylko struktury, ale i dynamikę wielu białek. Niestety, mimo stosowania superkomputerów i wydajnych algorytmów dostępne badaniom teoretycznym skale czasowe są wciąż b. małe – rutynowo lokują się w zakresie dziesiątek, może setek nanosekund. Oczywiście im większe białko tym więcej kosztuje jego badanie klasycznymi metodami pełnoatomowej dynamiki molekularnej. Znaczny postęp w eksplorowaniu dużych układów, i to na długich skalach czasowych, można uzyskać stosując w modelowaniu tzw. metody gruboziarniste, w których analizowanymi obiektami modelowymi są ziarna imitujące nie pojedyncze atomy, ale całe aminokwasy lub nawet grupy aminokwasów.

Przedstawiona mi do recenzji rozprawa doktorska p. mgr Aleksandry E. Badaczewskiej-Dawid dotyczy właśnie takich zagadnień. Doktorantka postawiła sobie za cel opracowanie i przetestowanie nowej metody reprezentacji struktury i dynamiki białek charakteryzującej się wysokim poziomem gruboziarnistości. Korzyści z opracowania takiej metody są oczywiste: daje ona możliwości sięgnięcia w symulacjach komputerowych w nowe, niedostępne dotąd, obszary szeroko rozumianej przestrzeni konformacyjnej dużych białek. W szczególności model gruboziarnisty może być pomocny w badaniu problemu zwijania białek metodami zwanymi czasem „ab initio” czy „de novo”, a więc opierającymi się jedynie na potencjalach fizycznych czy statystycznych. Ponieważ jesteśmy świadkami dynamicznego postępu w liczbie poznanych struktur dużych kompleksów białkowych, konieczność realistycznego modelowania takich systemów jest paląca, zatem wybór zagadnienia rozwiązywanego w doktoracie pani Badaczewskiej-Dawid uważam za niezwykle trafny, właściwy i aktualny. Nie jest to problem czysto akademicki – zagadnienie ma poważne uzasadnienie praktyczne i cała tematyka lokuje się w ambitnym nurcie rozwoju nowych metod obliczeniowych chemii.

W literaturze są oczywiście znane i rozwijane od wielu lat podejścia gruboziarniste, dowodem na to jest publikacja P5 w sławnym Chemical Reviews (2016), której doktoranta jest współautorką, jednak to, co jest zaprezentowane w cyklu prac P1-P4 oraz w samej rozprawie jest podejściem nowym i oryginalnym.

Rozprawa została napisana na Wydziale Chemii Uniwersytetu Warszawskiego, w Pracowni Teorii Polimerów, pod kierunkiem Pana dr hab. Dominika Gronta. Autorka przedstawiła cele pracy we wstępie: jej głównym zadaniem było stworzenie nowego modelu gruboziarnistego białek, nazwanego SURPASS< wychodzącego poza pewne ograniczenia istniejących podejść. Drugim celem było dokonanie oceny tego modelu w szczególności w aspekcie wydajności i skuteczności modelowania zarówno struktur jak i dynamiki białek. **W mojej ocenie oba cele w pełni osiągnięto.**

Rozprawa doktorska Aleksandry Badaczewskiej-Dawid, napisana w języku polskim, składa się ze Wstępu (3 strony), dwóch zasadniczych rozdziałów („Przegląd literatury” – 38 stron, „Wyniki pracy badawczej” – 38 stron), podsumowania (3 str.) spisu bibliografii (22 strony – 212 pozycji), materiałów dodatkowych do dwóch sekcji, spisu innych prac Autorki. Dołączono też kopie 5 publikacji wieloautorskich zawierających zasadniczą część wyników prezentowanych w rozprawie. Jak rozumiem (brak oświadczeń współautorów, brak informacji we wstępie) mojej ocenie podlega treść przedstawiona w formie klasycznej rozprawy, zaś publikacje mają charakter pomocniczy, ułatwiający zapoznanie się z dokonaniami doktorantki.

Po logicznym wprowadzeniu w tematykę (Rozdział 1 – Wstęp) następuje obszerny Rozdział 2.

Tytuł tego rozdziału „Przegląd literatury” jest nieco mylący, ponieważ zawiera on w swojej strukturze w zasadzie przegląd metod związanych z badaniem struktury białek i gruboziarnistym modelowaniem. Oczywiście są w tej części odniesienia do najważniejszych pozycji literaturowych. Przegląd czyta się bardzo dobrze, napisany jest zwięźle i jasno, sposób prezentacji nie budzi moich zastrzeżeń. Są drobne wady stylistyczne czy wyrażenia żargonowe, np. „konformacje zdenaturowane” str. 17, „predykcje” (zamiast przewidywania), rys. 5 na str. 25 ma dziwnie urwaną strukturę tripeptydu (b. „niechemiczną”), Autorka nie podała definicji ważnego pojęcia „geometrycznego centrum wiązania peptydowego” (?środek linii C α -N?), na str. 29 jest zwrot „z powierzchnią struktury” a powinno być z „powierzchnią białka”. Pojawiają się dziwne zbitki „kooperatywne oddziaływania hydrofobowe są siłą napędzającą elastyczność konformacyjną białka” lub skróty myślowe: str. 30: „wiązania i kąty płaskie są zwykle przybliżone oscylatorem harmonicznym”. Inne uwagi stylistyczne:

Rys.7 str. 30 – w odróżnieniu od często stosowanego określenia „oddziaływania niewiążące”, nie jest ścisły termin „energia niewiążąca” podany na rysunku.

Str. 31 – „koszt obliczeniowy jest kwadratowy”

Str. 31 – „zaletą pól siłowych.. jest przenośność na inne klasy..”

Str. 34 „temperatura zdefiniowana jako średnia po energiach kinetycznych” – nie znam takiej definicji temperatury.

Pewne moje wątpliwości budzi stosowanie w tym rozdziale pojęcia „dynamika”. Nie jest jasno co należy rozumieć pod tym słowem. W szczególności za nieprecyzyjne uważam stwierdzenie (str. 35) iż „*dobrze zaprojektowane typy lokalnych modyfikacji łańcucha polipeptydowego, realizowane w odpowiednio długich symulacjach MC, mogą zapewnić całkiem realistyczny obraz dynamiki, porównywalny z trajektoriami MD*”. Istotą problemu wg mnie jest to, iż owszem, jeśli chodzi o reprezentatywność zespołów statystycznych generowanych w długich symulacji MC względem długich symulacji MD – to

są one podobne i pozwalają na obliczanie wielkości termodynamicznych. Jeśli natomiast interesuje nas ewolucja czasowa układu w „czasie rzeczywistym” – wg mnie MC i MD dają całkowicie odmienne dane z fundamentalnych powodów.

Poza kompetentnym przeglądem głównych metod modelowania struktury białek, zwłaszcza gruboziarnistych, najciekawszą i najważniejszą częścią omawianego rozdziału jest sekcja 2.6.3 gdzie opisano własny ciągły model gruboziarnisty SURPASS (Single United Residue per Pre-Averged Secondary Structure fragment), opublikowany w pracy Dawid A.E., Gront D., Kolinski A., *J. Chem Theory and Comp.* 2017 (13). Istotnym założeniem tej metody jest zastąpienie każdego aminokwasu (residuum) łańcucha białkowego poprzez punktową masę, ulokowaną w uśrednionej pozycji 4 sąsiednich atomów C α i oddziałującą empirycznymi potencjałami z innymi modelowymi aminokwasami. Liczba stopni swobody spada o ok 2 rzędy wielkości w stosunku do modelu pełnoatomowego. Potencjały SURPASS są oryginalne i tylko jakościowo są zbliżone do innych podobnych podejść (SICHO, MARTINI, CABS, UNRES). Istotne jest włączenie do potencjałów wiedzy empirycznej – mają one w pewnej części charakter potencjałów statystycznych opartych na danych strukturalnych dotyczących białek oraz zawierają informacje o strukturze drugorzędowej odpowiednich części łańcucha białkowego. W obecnym wariantcie metoda ma tylko kilka typów atomów i nie wprowadzono parametryzacji opartej na lokalnej specyfice oddziaływań danego aminokwasu. Zbyt mało to nie dziwi – chociaż koncept jest słuszny by uzależnić „typy” centrów SURPASS od bezpośredniego otoczenia - ponieważ technicznie zagadnienie to jest złożone. Prosta analiza kombinatoryczna, uwzględniająca fakt włączenia informacji z 4 kolejnych aminokwasów do jednego centrum daje (przy założeniu 20 standardowych aminokwasów) ok. 1600 rodzajów centrów metody SURPASS, które należałoby sparametryzować. W każdym razie, nawet w obecnej, nieco uproszczonej, wersji metoda SURPASS jest bardzo ciekawą opcją dla modelowania białek i zarówno sam pomysł metody, przyjęte założenia upraszczające, fakt jej zaprogramowania i dość szerokiego przetestowania tego podejścia oceniam bardzo wysoko.

Autorka wykonała szereg badań wstępnych pozwalających ocenić wady i zalety metody SURPASS. Jak słusznie pisze, symulacje Monte Carlo z wymianą replik, bardzo wydajne, nie prowadzą do super dobrych zwiniętych białek, tylko do struktur „białkopodobnych” Zatem SURPASS nie nadaje się bezpośrednio do zwijania nieznanymi białek techniką ”de novo”. Jednak próbkowanie przestrzeni konformacyjnej biopolimerów SURPASS realizuje bardzo dobrze, wyniki można b. zgrabnie wizualizować i porównywać z istniejącymi strukturami, zaś wyprodukowane struktury wcale nie są takie tragiczne i często zbliżone do natywnych. W przyszłości pojawią się znaczne możliwości zastosowania podejść opartych właśnie na SURPASS do modelowania wielkich kompleksów białkowych, co jest obecnie niemal niemożliwe technikami pełnoatomowymi, zaś dla biologii XXI wieku kluczowe.

Opis metody zawarty w sekcji 2.6.3 nasunął mi kilka pytań i uwag:

1. Str. 45; Czy na pewno uśrednianie fragmentów 4-resztowych prowadzi do „wygładzania fragmentów nieustrukturalizowanych” – skoro liczba elementów opisu maleje, to spodziewałbym się, że pętle będą bardziej, a nie mniej ”kanciaste”?
2. Nie jest precyzyjnie opisane co to znaczy, że atomy C α mają orientację. Spora część tego fragmentu dotyczy właśnie orientacji atomów, oczywiście te o kształtach (?potencjał anizotropowy?) elipsoidy obrotowej mają jakieś właściwości kierunkowe, ale te sferycznie symetryczne ...?

3. Byłbym ciekaw (może na obronie się dowiem...) jak technicznie uwzględniono wpływ rozpuszczalnika (wody?) na struktury SURPASS. Czy dałoby się wprowadzić jakiś niejawni opis fragmentów zanurzonych w błony komórkowe?
4. Doktorantka pisze (str. 4), iż wagi członów w wyrażeniach na potencjał zoptymalizowano w serii długich obliczeń, ale nie dyskutuje w jaki sposób wykonano tę optymalizację, jaką metodą i w oparciu o jakie kryteria. (*trial-and-error?*)
5. W zasadzie żaden z potencjałów pokazanych na rys. 14 nie ma fizycznego charakteru, bowiem ich asymptotyka nie pozwala na zrywanie wiązania. Jest to oczywiście cecha większości powszechnie stosowanych pól siłowych, jestem jednak ciekaw, czy metoda SURPASS z tego powodu nie generuje jakichś kontrolowanych „artefaktów”, czy też nie prowadzi to do ograniczonego zakresu stosowalności metody.

Główny Rozdział 3 zawiera wyniki pracy badawczej doktorantki, w znakomitej części opublikowane w najlepszych czasopismach specjalistycznych (publikacje P1-P5). Otwiera ten rozdział pożyteczny spis 6 starannie wybranych baz danych białkowych, służących do tworzenia i testowania metody SURPASS. Jest też zestaw podstawowych wzorów i definicji pojęć służących do oceny wyników, brakuje w nim definicji RG czyli (jak pisze autorka) promienia bezwładności. Sekcja 3.1 oparta na pracy P1 stanowi prezentację metody SURPASS i przedstawia wyniki modelowania serii białek z bazy GLOBULE_195. Niektóre wyniki, np. przedstawione na rys. 16 są spektakularne, Klatki z pseudotrajektorii MC są b. podobne do PDB-owskich struktur niewielkich białek (50-100 reszt) na poziomie rmsd ok. 1.6 do 2.9 Ang. W tej części zaciekały mnie dwie niedopowiedziane kwestie: (1) skąd wiadomo kiedy skończyć symulację MC? A może coś lepszego się pojawi za 1000 kroków? (2) Czy jest jakaś funkcja pokazująca, że zbliżamy się do struktury natywnej, czy też „sukces” polega na tym, iż w mrowiu konformacji proponowanych przez SURPASS udaje się wyłuskać kilka struktur, które są bliskie strukturze (znanej) wzorcowej? Praca P1 jest doskonale napisana, informatywna, a spostrzeżenia i cechy nowej metody są dobrze poparte doświadczeniami obliczeniowymi.

W kolejnej sekcji (3.1.2) autorka bada bardzo ważny problem, do jakiego stopnia więzy nakładane w metodzie SURPASS, a wynikające z narzuconej z góry struktury drugorzędowej modelu mają wpływ na uzyskane struktury. Czy można z nich zrezygnować, czy rezultaty są czułe na jakość zadanej SS2 (helikalna, beat – kartka, pętla). Pomysł badania jest b. dobry i realizacja też. Dane przedstawione na rys. 21 pięknie pokazują, że w tego typu metodach więzy SS2 są kluczowe. Rys. 22 (str. 62) początkowo mi się bardzo spobał, ale po chwili zauważyłem, że przy białku alfa+beta 2vh nie ma wcale modeli o zwoju podobnym do struktury natywnej (są ułamki procenta) – zatem albo jest to błąd na rysunku, albo jest to może źle dobrany przykład.

Sekcja 3.1.3 z kolei opisuje wyniki badań wpływu więzów uzyskanych z wielowymiarowego NMR dla białek. Z widm NMR można uzyskać tzw. odległości oparte o NOE. Doktorantka pomysłowo przeniosła więzy związane z parami atomów na więzy pomiędzy ziarnami metody SURPASS (rys. 24) i następnie systematycznie przebadła jaki rodzaj więzów jaką daje poprawę w wynikach modelowania. Jest to moim zdaniem cenne badanie metodologiczne. Było dla mnie zaskoczeniem odkrycie w serii wyników prezentowanych na rys. 25 kilku białek, dla których więzy nic nie poprawiały. Doktorantka w zasadzie nie podała żadnego fizycznego wyjaśnienia tej obserwacji, poza uwagą, iż widocznie model SURPASS sobie tak dobrze radzi w tych kilku przypadkach, że żadne nowe więzy już nic nie ulepszą. Zgoda, to jest stwierdzenie faktu, ale czy jest jakiś konkretny powód (niekoniecznie czysty przypadek) – np. jakies

cechy wspólne tych układów, że więzy już nic nie wnoszą?? Wg mnie, jest to ciekawa zagadka, może warta głębszego zbadania.

Kolejne zagadnienie to próba oceny jakości wyników „dynamicznych” metody SURPASS. Zwykle interesują nas takie cechy jak ruchliwość poszczególnych fragmentów białka mierzona miarą RMSF (fluktuacje). W modelach pełnoatomowych nie ma żadnego problemu poza pieniędzmi (czas komputera, energia elektryczna). Tutaj powstaje poważna kwestia, czy bardzo gruboziarnisty charakter modelu pozwala mieć nadzieję na uzyskanie jakichś wartościowych danych o białku w aspekcie np. zdolności do wiązania innych partnerów w kompleksie. Przetestowano zbiór 140 białek (FLEX_1400) zawierających dane doświadczalne z NMR. Co więcej, porównano wyniki z coraz popularniejszym podejściem (nawet w mojej grupie stosowanym) – Elastic Network Model (ENM). No i faktycznie zarówno ENM jak i testowany CABS dały najlepszą zgodność rezultatów z doświadczeniem. Optymistyczne jest to, że jak pisze autorka „dostarczone [przez SURPASS] profile fluktuacji są dość realistyczne”. Dla 30% jest nawet całkiem ok, dla reszty nieco gorzej. Badania z tego fragmentu rozprawy były żmudne ale potrzebne. To że wyniki nie są „super”, nie dyskwalifikuje metody, po prostu nakazuje zachować ostrożność przy interpretacji wyników. Poza tym, można słusznie oczekiwać, że dodanie potencjałów lokalnych do SURPASS polepszy jakość metody w tym aspekcie. Ogólnie badanie to uznaję za ważny postęp metodologiczny, dobrze, że zebrano te dane.

Ponieważ jako recenzent krytyczny muszę się do „czegoś przyczepić”, to wyrażam lekkie zdziwienie zastosowaniem regresji liniowej do danych z wykresów podanych na rys. 28 – czy spodziewamy się faktycznie takiej korelacji? Czy z tych R^2 na tak niskim poziomie możemy się czegoś nauczyć?

Powszechną słabością modeli gruboziarnistych (CG) jest to, że cząsteczki chemiczne nie są zbudowane z ziaren, ale jak się powszechnie wierzy, z atomów. Zatem porządna dyskusja wyników modelowania wymaga operowania położeniami atomów i najlepiej struktur pełnoatomowych (AA). Opracowano wiele metod estymowania takich struktur w oparciu o struktury gruboziarniste. Porządna metoda CG powinna dać wskazówki badaczom, jak odtworzyć strukturę AA. W sekcji 3.3 doktorantka opisała starannie nowy algorytm odbudowy położenia atomów $C\alpha$ w oparciu o bibliotekę fragmentów SURELib (300 kompletów unikalnych 5 lub 6-resztowych fragmentów łańcucha głównego). Praca P3, w JCTC, pisana tylko przez doktorantkę i prof. A. Kolinskiego, jest przykładem b. dobrej pracy programistycznej i analitycznej. Są w niej fajne pomysły dobrze opisane, zaś samo podejście jest sprawdzone i przetestowane. Algorytm działa dobrze, daje ładne wyniki odbudowy położenia atomów $C\alpha$ (0.5- 0.2 Ang. RMSD dla 99% i 92 % białek z fragmentami α i β). Najlepsze jest to, że nowy algorytm też dobrze (choć oczywiście gorzej) odbudowuje położenia atomów $C\alpha$ dla klasek z symulacji. Tutaj też odbudowane struktury mają stosunkowo niskie RMSD (3 Ang.) od struktur uzyskanych znacznie droższymi metodami. Takie wyniki z pewnością zachęcą część badaczy do próbowania metody SURPASS w swoich analizach białek. Mnie np. marzyłoby się „zaatakowanie” całego spliceomu, może to nastąpić (w pierwszej kolejności) w obszarze metod gruboziarnistych.

Strona techniczna - Literówki :

Str 29 „ta samą” -> tą samą

Str. 58 podpis, „koku” → kroku

Str. 105, poz lit [13], czy na pewno jest to dobre cytowanie do metody PRIMO (Gopal, Feig 2010, Proteins)?

Str. 109, poz. [11] dane niepełne

Drobne uwagi redakcyjne:

Tytuł napisano kapitalikami, wg mnie niepotrzebnie, w polskich tekstach tak się nie robi. Dlaczego wszędzie jest stosowany zwrot „autor” a nie autorka??

Praca doktorska napisana jest na ogół poprawnym językiem. Redakcja jest bardzo staranna, niemal nie ma literówek. Mocną stroną doktoratu są dobrej jakości rysunki i ilustracje. Rozdział przeglądowy stanowi w moim przekonaniu cenne kompendium metod gruboziarnistych i może być z powodzeniem polecany początkującym doktorantom czy studentom, zwłaszcza, że zagadnienia dynamiki gruboziarnistej nie były, i ile wiem, opisywane dotąd w języku polskim .

KONKLUZJA

Pani mgr A.E. Badaczewska-Dawid przygotowując swoją rozprawę doktorską opracowała wraz z zespołem współautorów oryginalną i efektywną nową metodą badania struktury i dynamiki białek metodą modelowania gruboziarnistego. Realizacja tego zadania wymagała wiele pomysłowości i pokonania wielu trudności. Metoda może być wykorzystana w nowych badaniach trudnych problemów biologicznych i chemicznych. Z pewnością przyczyniała się do rozwoju metod chemii teoretycznej i teorii biopolimerów. Doktorantka uzyskała szereg wartościowych szczegółowych wyników naukowych opublikowanych w czołowych międzynarodowych czasopismach specjalistycznych. Jak rozumiem, metoda może być wkrótce udostępniona ogółowi badaczy. W trakcie realizacji pracy doktorskiej przyczyniała się do przygotowania poważnego artykułu przeglądowego do Chemical Reviews (IF>20, cytowany już ok. 200 razy).

Stwierdzam, zatem z pełnym przekonaniem, że przedłożona mi do oceny rozprawa doktorska mgr. Aleksandry E. Badaczewskiej-Dawid pt **“Modelowanie struktury i dynamiki białek z użyciem modeli gruboziarnistych o różnej skali rozdzielczości”** stanowi oryginalne rozwiązanie problemu naukowego. Dzieło to demonstrowa wiedzę doktorantki w naukach chemicznych, zwłaszcza w zakresie biologicznej chemii fizycznej, czy z pogranicza biofizyki molekularnej. Lektura prac i rozprawy utwierdza mnie w przekonaniu, że doktorantka posiadała umiejętność samodzielnego prowadzenia pracy naukowej, a doktorat tj. spełnia wymagania ustawowe [Ustawa z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki ze zmianami późniejszymi (Dz. U. Nr 65, poz. 595, ze zm. w Dz. U. z 2005 r. Nr 164, poz. 1365 oraz w Dz. U. z 2011 r. Nr 84, poz. 455] oraz zwyczajowe stawiane rozprawom doktorskim. **Wobec powyższego wnoszę o dopuszczenie Panią magister Aleksandrę Badaczewską-Dawid do dalszych etapów przewodu doktorskiego.**

W osobnym piśmie przedstawię argumenty za przedstawieniem rozprawy do wyróżnienia, zgodnie z panującymi na Wydziale zasadami.



Wiesław Nowak, prof. zw.