Recenzja rozprawy doktorskiej mgr Przemysława Porębskiego

Praca doktorska mgr Przemysława Porębskiego nosi tytuł ”Improvement of model building and refinement of macromolecular crystal structures”. Wyniki uzyskano w laboratorium prof. Władysława Minora na University of Virgina w USA.

Rozprawa dotyczy rentgenowskich badań krystalograficznych kompleksów białkowych i składa się z 5 głównych rozdziałów – Wstęp, Cele doktoratu, Materiały i metody, Wyniki, Dyskusja – oraz długiej listy odnośników. Doktorat jest napisany po angielsku (z akceptowalną dawką błędów), ale tytuły rozdziałów podaję w wersji polskiej. Praca poświęcona jest głównie rozwijaniu algorytmów rozwiązywania struktur białkowych na podstawie danych rentgenowskich.

Wstępny rozdział dostarcza kompetentnego przeglądu dostępnych metod rozwiązywania struktur białkowych i prowadzi do sformułowania celu pracy doktorskiej. Jest nim ulepszenie modelowania struktury grup bocznych w ramach programu nazwanego *Fitmunk* i napisanego w języku *Python*. Program korzysta z szeregu zewnętrznych bibliotek, których lista jest przedstawiona.

Algorytm programu *Fitmunk* ma trzy główne składniki. Pierwszym z nich jest heurystycznie wprowadzona efektywna energia, która zależy od mapy gęstości elektronowej oraz częstości występowania zadanej konformacji w bibliotece rotamerów. Energia ta uwzględnia też energię oddziaływania grupy bocznej z łańcuchem głównym (człony kulombowskie i van der waalsowskie). W doktoracie brakuje mi przedstawienia argumentów przemawiających za takim akurat wyborem defincji energii i wyjaśnień, co by się stało przyjmując jakąś inną definicję.

Drugim składnikiem jest utworzenie bibliotek konformerów na podstawie plików zdeponowanych w Protein Data Bank (PDB). Wybrano struktury o rozdzielczości lepszej niż 1.8 Å i wyeliminowano struktury zbyt bliskie sobie. Skorzystano również z dodatkowych kryteriów eliminacji struktur. W uzupełnieniu A po-

dano 2 075 dopuszczonych plików PDB, z których uzyskano 330 642 konformacji grup bocznych. Eliminacja powtórzeń na poziomie r.m.s.d=0.1 Å oraz sytuacji sprzecznych ze stereochemią doprowadziła do utworzenia biblioteki nazwanej cl0.1 i złożonej z 83 042 konformacji. Hierarchiczne klasterowane pozwala utworzyć pochodne uproszczone biblioteki o mniejszej liczbie elementów (do 5 762 w wypadku cl0.7). Biblioteki te służą do wybrania konformacji, które są najbardziej zgodne z pomiarami. Robi się to poprzez wybieranie różnych ścieżek szukania zgodności – dla całego białka.

Trzecim składnikiem jest zaproponowanie algorytmu eliminacji tych ścieżek poszukiwawczych, które nie prowadzą do żadnego rozwiązania. Robi się to poprzez przypisanie konformacjom efektywnej energii i tym samym narzucenie ścieżkom topologii drzewa oraz na wykorzystaniu kryteriów energetycznych. Kryteria te uwzględniają również zgodność z więzami sterycznymi. Metodologię tę przetestowano na zbiorze 36 struktur, które zostały rozwiązane niedawno przy użyciu innych metod.

Autor zajmuje się także problemem rozpoznawania sekwencji. Do rozwiązania go zaproponował metodę ukrytych modeli Markova, w której dokonuje się stochastycznych podstawień bądź usunięć aminokwasów przy wykorzystaniu innego wyrażenia na energię – zoptymalizowanego dla każdego reziduum.

W rozdziale poświęconym wynikom, mgr Porębski przedstawia testy działania programu *Fitmunk* i skuteczności wyboru bibliotek konformacji. Dla tradycyjnej biblioteki rotamerów *Fitmunk* prowadzi do dokładnosści rozwiązywania struktur tuż poniżej 60%. Jednak zastosowanie nowych bibliotek konformacji zwiększa dokładność nawet do powyżej 90%. Niezbyt rozumiem rysunek 9 przedstawiający te wyniki z dwóch powodów:
1. Nie jest jasne, jak zdefiniowane są porównywane biblioteki oznaczone symbolami cl0.2, cl0.7, cl1.5 i cl2.0;
2. Nie jest dla mnie intuicyjnie jasne, dlaczego biblioteka o – najwyraźniej – najmniejszej liczbie elementów (cl0.7) daje dokładność praktycznie taką samą,

jak pełna biblioteka (pozostałe biblioteki są gorsze).
Nie widzę, żeby to było dyskutowane.

Pozostałe testy, dla tego samego (lub prawie tego samego) zestawu bibliotek autor przedstawia dla szeregu innych parametrów programu *Fitmunk*. Są to wybór promienia uśredniania, wybory parametrów dotyczących efektywnej energii, usuwania kolizji sterycznych, itd. *Fitmunk* okazuje się zawsze wygrywać w porównaniu z dotychczasowymi podejściami.

Metodologia wypracowana przez mgr Porębskiego również zwiększa sukces rozpoznawania sekwencji. Zwykła metoda rozpoznawania reziduum przez dopasownia do rozkładu gęstości jest skuteczna co najwyżej 60-procentowo. Mgr Porębski wprowadził system ocen dopasowania, w którym są parametry wyznaczone drogą trenowania na przykładowych strukturach. W wyniku tych działań pozwala zidentyfikować poszczególne aminokwasy z sukcesem większym od 10 do 30 punktów procentowych niż w dotychczasowych podejściach, w zależności od aminokwasu.

Mgr Porębski również przetestował *Fitmunk* na tzw. prawdziwych przykładach struktur. Pozwoliło to ulepszyć rozpoznanie tych struktur i usunąć błędy w już rozwiązanych strukturach. Jako główne zadanie w dalszym rozwijaniu tej metodologii mgr Porębski uznaje ulepszenie wyrażeń na efektywne energie. Używanie *Fitmunka* powinno w zdecydowany sposób przyśpieszyć i zautomatyzować rozwiązywanie struktur krystalograficznych. *Fitmunk* ma więc duże szanse na na powszechne stosowanie w krystalografii. Jest to więc duże osiągnięcie.

W Web of Science można sprawdzic, że mgr Porębski jest współautorem 10 artykułów naukowych i kilku streszczeń konferencyjnych. Jeden z tych artykułów jest cytowany 40 razy. Jest to:
K. A. Majorek, P. Porębski, A. Dayal, M. D. Zimmerman. K. Jabłońska, A. J. Stewart, M. Chruszcz, i W. Minor, Structural and immunologic characterization of bovine, horse, and rabbit serum albumins, Molecular Immunology, 52, 174-182

(2012).

Całkowita liczba cytowań jego artykułów wynosi 73, co – jak na doktorat – jest wynikiem znakomitym. Mgr Porębski jest pierwszym autorem jednego z tych artykułów.

Doktorat mgr Porębskiego jest całkowicie poświęcony ulepszaniu metod rozwiązywania struktur krystalicznych. Jest w nim niewiele na temat zastosowań tych metod do białek, o których jest mowa w 10 artykułach, których jest współautorem. Są to białka takie jak albuminy, białka alergenów, syntetazy dethobiotyny, kinazy fosfogliceratów, acetylotransferazy, itd. Ubolewam nad tym pominięciem, gdyż jest to obszar, dla którego mgr Porębski mógł wykazać, do czego w sumie prowadzą jego metody i dlaczego ich użycie było tam istotne. Mógłby też w ten sposób wykazać, że rozumie biologię tych układów a nie tylko metodologię postępowania. Innymi słowy, że kontroluje również rozumienie wyników uzyskanych przy użyciu jego programu.

Uważam, że

1. Tematyka doktoratu jest bardzo ważna.
2. Tematyka doktoratu jest technicznie trudna jednak mgr Porębski zdołał wypracować podejście, które wydaje się być skuteczne.
3. Program numeryczny opracowany przez mgr Porębskiego powinien stanowić ważne narzędzie ułatwiającym interpretację doświadczeń krystalograficznych.

Mgr Przemysław Porębski zatem z pewnością spełnia warunki wymagane do nadania stopnia naukowego doktora z biofizyki. Wnoszę o dopuszczenie doktoratu do dalszych etapów przewodu doktorskiego.

Prof. dr hab. Marek Cieplak
Instytut Fizyki PAN
10 września 2014 r.