

ZAŁĄCZNIK 3A

AUTOREFERAT

1. WYKSZTAŁCENIE

- X 2006 – VI 2010 Uniwersytet im. Adama Mickiewicza w Poznaniu. Studia doktoranckie na Wydziale Fizyki, specjalność biofizyka; praca doktorska zatytułowana 'Symulacje numeryczne wpływu dynamiki białka na pewne biomolekularne procesy przeniesienia elektronu'. **Stopień doktora nauk fizycznych w zakresie biofizyki.**
- X 2002 – II 2007 Uniwersytet im. Adama Mickiewicza w Poznaniu. Pięcioletnie studia magisterskie na Wydziale Fizyki, kierunek fizyka, specjalność fizyka teoretyczna; praca magisterska zatytułowana 'Badania numeryczne wpływu nierównowagi wibracyjnej na procesy przeniesienia elektronu', ukończone z wyróżnieniem. **Tytuł magistra fizyki.**
- X 2001 – II 2006 Uniwersytet im. Adama Mickiewicza w Poznaniu. Pięcioletnie studia magisterskie na Wydziale Chemii, kierunek chemia, specjalność chemia teoretyczna; od 2002 Indywidualne Studia Przyrodnicze, niektóre kursy dodatkowo ukończone na Wydziale Matematyki i Informatyki; praca magisterska zatytułowana 'Równoległe obliczenia kwantowo-chemiczne z użyciem funkcji jawnie skorelowanych na przykładzie atomu helu', ukończone z wyróżnieniem. **Tytuł magistra chemii.**
- IX 2001 – VI 2002 Policealne Studium Zawodowe im. Mikołaja Kopernika w Nowym Tomysłu, specjalność Marketing i Zarządzanie, ukończone z wyróżnieniem. **Tytuł Specjalisty ds. Marketingu i Zarządzania.**
- IX 1997 – VI 2001 Liceum Ogólnokształcące im. Mikołaja Kopernika w Nowym Tomysłu, klasa informatyczna, ukończone z wyróżnieniem.

2. ZATRUDNIENIE W INSTYTUCJACH NAUKOWYCH

- od XI 2015 Tessella Ltd., Stevenage, UK (**Analyst Programmer**), doradztwo naukowe m.in. dla przemysłu chemicznego i farmaceutycznego.
- X 2010 – X 2015 Cancer Research UK (The Francis Crick Institute od 1/04/2015), Londyn, UK (**Postdoctoral Research Fellow**). 1/04/2015 Cancer Research UK London Research Institute został włączony w skład nowo powstałego The Francis Crick Institute.
- 2008 – 2010 BioInfoBank Institute, Poznań, Polska (**Junior Researcher, Software Developer**).

3. AKTYWNOŚĆ NAUKOWA I INNE FORMY AKTYWNOŚCI (szczegóły w załączniku 4A)

Moje studia na Wydziale Chemii rozpocząłem w październiku 2001. Równoległe, w czasie weekendów, uczęszczałem do Policealnego Studium Zawodowego (Marketing i Zarządzanie) i Akademii Sieci Komputerowych CISCO. Będąc studentem drugiego roku chemii, dołączyłem do grupy badawczej Prof. Jacka Rychlewskiego jako student indywidualny i rozpocząłem

moją karierę naukową w dziedzinie chemii kwantowej. W tym samym czasie uzyskałem zgodę na kontynuację studiów w ramach Indywidualnych Studiów Przyrodniczych, specjalnego trybu studiów, który umożliwił mi studiowanie równoległe fizyki i otrzymanie drugiego tytułu magistra po spełnieniu odpowiednich kryteriów. Jako student indywidualny, byłem zaangażowany w część organizacyjną mojej grupy badawczej; pozwoliło mi to na zrozumienie struktury Uniwersytetu i uzyskanie nowych umiejętności.

Bardzo szybko zorientowałem się w kierunku nauk obliczeniowych. Studia na dwóch kierunkach były dużym wyzwaniem jednakże będąc zainteresowanym chemią kwantową, wiedziałem, że fizyka jest komplementarna do chemii. Aby uzyskać dodatkową wiedzę, ukończyłem także kilka kursów na Wydziale Matematyki i Informatyki.

Po nagłej śmierci Prof. Jacka Rychlewskiego w 2003 roku, pozostałem w tej samej grupie, a moim opiekunem został Prof. Jacek Komasa. Atmosfera w grupie i warunki rozwoju były wspaniałe; oprócz rozwijania projektów związanych z moją pracą magisterską, miałem wolność rozwijania mojej kariery w różnych kierunkach naukowych. Jako student trzeciego roku napisałem moją pierwszą publikację poświęconą zrównolegleniu obliczeń kwantowo-chemicznych i uczestniczyłem w mojej pierwszej konferencji międzynarodowej. Oprócz uczęszczania na obowiązkowe zajęcia, brałem udział w wielu różnorodnych szkoleniach organizowanych przez Poznańskie Centrum Superkomputerowo-Sieciowe. W lutym 2006 ukończyłem moje studia na Wydziale Chemii. W tym samym czasie rozpocząłem współpracę z Prof. Michałem Kurzyńskim z Wydziału Fizyki w kontekście mojej drugiej pracy magisterskiej w obszarze fizyki statystycznej. W październiku 2006 roku rozpocząłem moje studia doktoranckie pod opieką Prof. Michała Kurzyńskiego, w lutym 2007 roku ukończyłem moje studia magisterskie na Wydziale Fizyki, a w czerwcu 2010 roku otrzymałem stopień naukowy doktora nauk fizycznych w zakresie biofizyki. Moja praca doktorska była poświęcona rozwojowi i symulacjom modeli sieciowych opisujących dynamikę białka (jeszcze bez uwzględnienia struktury molekularnej). Podczas studiów odbyłem staże naukowe, zarówno akademickie (uniwersytety w Erlangen i w Gdańsku) jak i w przemyśle (WAVIN Metalplast). Podczas studiów doktoranckich prowadziłem zajęcia dydaktyczne (chemia teoretyczna, wstęp do informatyki, technologie informacyjne). Studiując chemię otrzymałem uprawnienia pedagogiczne do nauczania chemii w szkołach podstawowych i średnich. Te umiejętności przydały się też kiedy zacząłem nauczać na poziomie szkoły wyższej.

W czasie studiów doktoranckich dołączyłem do zespołu BioInfoBank Institute jako Junior Researcher i Software Developer. Rozwijałem oprogramowanie dla Instytutu i brałem udział w projektach naukowych, które były różne od tematyki mojej pracy doktorskiej, między innymi w projektowaniu leków przeciwbakteryjnych.

Od października 2010 do października 2015 byłem zatrudniony na stanowisku Postdoctoral Research Fellow w Cancer Research UK London Research Institute (Biomolecular Modelling Laboratory, grupa badawcza Paula Batesa; 1/04/2015 CRUK LRI stało się częścią The Francis Crick Institute), gdzie moje wcześniejsze umiejętności i doświadczenie zostały znacząco rozszerzone w kierunku biologii molekularnej. Praktycznie było to dla mnie przejście do nowej dziedziny, gdyż przenieśliśmy się ze studiowania modeli teoretycznych w kierunku rzeczywistych układów molekularnych. Dlatego też potrzebowałem prawie dwóch lat aby

rozwinąć nowe projekty i opublikować wyniki badań. Jednocześnie to właśnie prace z okresu podoktorskiego przyniosły mi praktycznie wszystkie cytowania oraz od nowa stworzyły moją wartość indeksu Hirsha.

Moje doświadczenie naukowe jest różnorodne, od obliczeń kwantowo chemicznych wysokiej precyzji, przez metody semiempiryczne chemii kwantowej, do fizyki statystycznej i biologii molekularnej. Są to m.in.: obliczenia wysokiej precyzji dla małych atomów, rozwój algorytmów równoległych, symulacje dynamiki stochastycznej w celu modelowania szybkości procesów transferu elektronu i przeniesienia energii swobodnej w białkach, projektowanie leków, obliczenia i zastosowania średnich czasów pierwszego przejścia, przewidywanie struktur kompleksów białko-białko, struktura i dynamika lejów energetycznych oddziałujących białek (ang. *protein-protein binding funnels*), rozwój strategii do określenia poprawności (ang. *scoring*) struktur kompleksów białko-białko i przewidywanie wpływu mutacji na oddziaływania białko-białko.

Za każdym razem zbierałem informacje niezbędne do wykonania projektu, co miało znaczący wpływ na końcowy wynik. Wszystkie projekty były powiązane zastosowaniem metod numerycznych do znalezienia odpowiedzi na postawiony problem naukowy. W wielu przypadkach pisałem własne oprogramowanie i tworzyłem sekwencyjne lub równoległe algorytmy w różnych językach takich jak Fortran, C++, Python i Java. Niektóre projekty wymagały rozwijania serwisów internetowych i zarządzania nimi. Byłem w stanie nawiązać współpracę lokalną jak i międzynarodową. Jako pierwszy autor większości publikacji, byłem odpowiedzialny za wysłanie manuskryptu, przygotowanie korekt i zarządzanie projektami.

Brałem udział w różnorodnych konferencjach, spotkaniach i warsztatach, na temat obliczeń wysokiej wydajności, fizyki statystycznej, chemii kwantowej, modelowania molekularnego i innych (Załącznik 4A: Dodatkowe Informacje). Jest to dowodem moich szerokich zainteresowań, co daje mi możliwość spojrzenia na problemy naukowe z różnej perspektywy. Udział w licznych kursach stacjonarnych i online podczas stażu podoktorskiego (Załącznik 4A: Dodatkowe Informacje) jest dowodem na to, że potrafię samodzielnie znaleźć drogę aby pozyskać nową interesującą dla mnie wiedzę.

W Cancer Research UK zajmowałem się dokowaniem molekularnym, badaniem oddziaływań między białkami, wpływem mutacji i dynamiką sieci złożonych. Jestem twórcą darmowego oprogramowania: SwarmDock Server oraz RaTrav (więcej szczegółów w punktach 5 i 7 poniżej). Brałem udział w międzynarodowym konkursie CAPRI (Critical Assessment of PRediction of Interactions; <http://www.ebi.ac.uk/msd-srv/capri/>) z dość dobrymi wynikami od kompleksu T50 (wyniki na: <http://web.mit.edu/sheny/capri.html> jako SwarmDock i Bates). W 2014 wzięłem udział w dwóch nowych dla mnie międzynarodowych konkursach – CASP11 (przewidywanie struktur białek; przygotowałem automatyczny serwer do analizy jakości rozwiązań) oraz CSAR (oddziaływania białko-mała cząsteczka; dwie fazy – scoring i docking).

W 2015, dzięki współpracy z grupą prof. dra hab. Janusza Raka z Uniwersytetu Gdańskiego, rozszerzyłem moją tematykę badawczą o badanie mechanizmów pękania nici DNA w

zależności od lokalnego otoczenia co ma znaczenie w kontekście tworzenia nowych terapii przeciwnowotworowych.

W listopadzie 2015 zacząłem pracować dla Tessella Ltd., brytyjskiej firmy doradztwa naukowego, w celu zdobycia doświadczenia w komercyjnym środowisku naukowym, co pozwoli mi lepiej zrozumieć oba naukowe światy, akademicki i komercyjny. Jednocześnie nadal rozwijam dwa główne dla mnie tematy tj. oddziaływania białko-białko oraz sensybilizatory biologiczne.

W kontekście pracy organizacyjnej, będąc studentem pierwszego roku chemii, przyłączyłem się na rok do Naukowego Koła Chemików i brałem udział w organizacji wydarzeń dla społeczeństwa takich jak pokazy chemiczne. Później brałem udział w przygotowaniu wielu innych wydarzeń (Drzwi Otwarte, Promocja Edukacyjna, Dni Kultury i Sztuki, egzaminy wstępne, wybory Dziekana) i pracowałem na rzecz studentów niepełnosprawnych Uniwersytetu im. Adama Mickiewicza. Stworzyłem i zarządzałem kilkoma stronami internetowymi (Naukowe Koło Chemików, Samorząd Studentów, XIV Krajowa Szkoła Nadprzewodnictwa). Podczas studiów doktoranckich pracowałem w Komisji Ekonomicznej i Sądzie Koleżeńskim dla doktorantów. Po przeniesieniu się do Londynu uczestniczyłem w zbiórce pieniędzy dla Cancer Research UK. Dodatkowo brałem udział w wydarzeniu uświadamiającemu ludziom szkodliwość promieniowania UV.

Poprawiam moje umiejętności zarządcze organizując wydarzenia dla Polaków w grupach kulturalnej i turystycznej. Ostatnio zebrałem polską grupę na wykład Michaela Levitta na Imperial College. Natomiast podczas wycieczek zadawano mi wielokrotnie różnorodne pytania naukowe. Tłumaczę też na język polski jeden z programów na platformę Android, co umożliwi mi pracę w ogromnej społeczności pochodzącej z różnych zakątków świata. Aby poprawić moje umiejętności biznesowe (przydatne w kontekście zarządzania grantami) zdałem pięć egzaminów (Management Accounting, Financial Accounting, Business Mathematics, Business Economics, Ethics, Corporate Governance and Business Law) w certyfikowanym centrum egzaminacyjnym i w październiku 2013 otrzymałem CIMA Certificate in Business Accounting.

Podsumowanie (szczegóły w załączniku 4A):

- opublikowałem 18 prac naukowych (analiza bibliometryczna w punkcie 6); w pracach gdzie jestem pierwszym autorem zajmowałem się wysłaniem manuskryptu oraz korespondencją z edytorem;
- po doktoracie stworzyłem od podstaw dwa pakiety oprogramowania naukowego;
- uczestniczyłem w 7 krajowych i międzynarodowych grantach;
- zostałem zaproszony jako recenzent dla *Bioinformatics* (IF 2014: 4.981);
- nawiązałem współpracę z naukowcami pochodzącymi z różnych kultur;
- biorę udział w sieciach badawczych, zarówno w ramach międzynarodowych konkursów jak i poza nimi (np. sieć stworzona w celu rozwijania i testowania algorytmów do przewidywania struktur kompleksów białko-białko i powinowactwa), szczegóły w punkcie 8 poniżej;
- posiadam ponad pięcioletnie podoktorskie międzynarodowe doświadczenie naukowe (Wielka Brytania); uczestniczyłem w trzech krótkoterminowych stażach naukowych;

- prezentowałem wyniki ustnie na krajowych i międzynarodowych konferencjach naukowych (również na zaproszenie) i w formie posteru, przygotowywałem wykłady popularnonaukowe, wielokrotnie przedstawiałem wyniki badań w instytucjach dla których pracowałem;
- prowadziłem zajęcia dydaktyczne na Uniwersytecie im. Adama Mickiewicza;
- otrzymałem stypendia i nagrody za pracę naukową i organizacyjną;
- wciąż doksztalcam się biorąc udział w samodzielnie wybieranych kursach;
- wspierałem doktorantów w grupie wiedzą i umiejętnościami programistycznymi.

4. LISTA OPUBLIKOWANYCH PRAC NAUKOWYCH NALEŻĄCYCH DO POWIĄZANEGO TEMATYCZNIE CYKLU: STRUKTURA I DYNAMIKA KOMPLEKSÓW BIAŁKO-BIAŁKO

	TYTUŁ	IF¹
H1	M. Torchala , P. Chelminiak, P. A. Bates, 'Mean first-passage time calculations: comparison of the deterministic Hill's algorithm with Monte Carlo simulations', <i>Eur. Phys. J. B</i> 85 , 116 (2012). doi: 10.1140/epjb/e2012-20760-8. Udział własny: 55%.	1.282
H2	M. Torchala , I. H. Moal, R. A. G. Chaleil, J. Fernandez-Recio, P. A. Bates, 'SwarmDock: a server for flexible protein-protein docking', <i>Bioinformatics</i> 29 , 807-809 (2013). doi: 10.1093/bioinformatics/btt038. Udział własny: 70%.	4.621
H3	M. Torchala , I. H. Moal, R. A. G. Chaleil, R. Agius, P. A. Bates, 'A Markov-chain model description of binding funnels to enhance the ranking of docked solutions', <i>Proteins: Structure, Function, and Bioinformatics</i> 81 , 2143-2149 (2013). <i>Fig. 1 has been chosen as a cover page image.</i> doi: 10.1002/prot.24369. Udział własny: 70%.	2.921
H4	R. Agius, M. Torchala , I. H. Moal, J. Fernandez-Recio, P. A. Bates, 'Characterizing changes in the rate of protein-protein dissociation upon interface mutation using hotspot energy and organization', <i>PLOS Comp. Biol.</i> 9 , e1003216 (2013). doi: 10.1371/journal.pcbi.1003216. Udział własny: 30%.	4.829
H5	M. Torchala , P. A. Bates, 'Predicting the structure of protein-protein complexes using the SwarmDock Web Server', w: <i>Protein Structure Prediction 3rd Edition (Methods in Molecular Biology Vol. 1137)</i> , Rozdział 13, Humana Press, Springer, 2014. ISBN: 978-1-4939-0365-8 doi: 10.1007/978-1-4939-0366-5_13. Udział własny: 90%.	Rozdział
H6	I. H. Moal, M. Torchala , P. A. Bates, J. Fernandez-Recio, 'The scoring of poses in protein-protein docking: current capabilities and future directions', <i>BMC Bioinformatics</i> 14 , 286 (2013). doi: 10.1186/1471-2105-14-286. Udział własny: 40%.	2.672
H7	M. Torchala , P. Chelminiak, M. Kurzynski, P. A. Bates, 'RaTrav: a tool for calculating mean first-passage times on biochemical networks', <i>BMC Syst. Biol.</i> 7 , 130 (2013). doi: 10.1186/1752-0509-7-130. Udział własny: 45%.	2.853
Suma IF¹:		19.178

¹ impact factors dla roku publikacji; www.webofknowledge.com/JCR
Oświadczenia wkładu współautorów znajdują się w załączniku 5.

5. LISTA STWORZONEGO DARMOWEGO OPROGRAMOWANIA NAUKOWEGO POWIĄZANA Z OPUBLIKOWANYM CYKLEM PRAC NAUKOWYCH

- S1** **SwarmDock Server**: web serwis do przewidywania struktury 3D kompleksów białko-białko. <http://bmm.crick.ac.uk/~SwarmDock/>
- S2** **RaTrav**: narzędzie do obliczania średnich czasów pierwszego przejścia i obsadzeń przy pomocy algorytmów Hilla i Monte Carlo. <http://sourceforge.net/projects/ratrav>

Oświadczenia wkładu deweloperów znajdują się w załączniku 5.

6. ANALIZA BIBLIOMETRYCZNA CYTOWAŃ NA PODSTAWIE WYNIKÓW Z WEB OF SCIENCE, SCOPUS I GOOGLE SCHOLAR CITATIONS (27/10/2015)

	Przed doktoratem (2003-2009)	Po doktoracie (2012-2015)	Razem
Liczba publikacji	7	11	18
Suma IF ¹	8.567	31.347	39.914
Cytowania Web of Science ²	11	87	98
Cytowania Scopus ²	11	81	92
Cytowania Google Scholar ³	20	118	138
h index	5 (Web of Science), 5 (Scopus), 6 (Google Scholar)		

¹wszystkie impact factors dla roku publikacji, IF 2014 dla publikacji z 2015

²bazy Web of Science i Scopus mają duże opóźnienia w aktualizacji cytowań

³moje cytowania Google Scholar: <http://scholar.google.co.uk/citations?user=G8YpCOgAAAAJ>

7. STRUKTURA I DYNAMIKA KOMPLEKSÓW BIAŁKO-BIAŁKO

Wprowadzenie

Podczas syntezy białka aminokwasy są łączone dzięki wiązaniu peptydowemu. Biorąc pod uwagę strukturę chemiczną, aminokwasy mogą wyglądać względnie prosto, gdyż zawierają centralny atom węgla oznaczany jako C α , atom wodoru, grupę aminową, grupę karboksylową oraz specyficzny łańcuch boczny. Jednakże ogromna liczba białek o różnorodnej strukturze trójwymiarowej może być utworzona ze zbioru 20 naturalnie występujących aminokwasów [Branden, 1999].

Białka są odpowiedzialne za różnorodne funkcje w molekularnej maszynie jaką jest życie [Branden, 1999]. Jednakże zdecydowana większość funkcji biologicznych jest sprawowana nie przez pojedyncze białka ale przez oddziaływania białko-białko. Dynamika tworzenia i dysocjacji kompleksów jest związana z pełnioną funkcją. Warto również zauważyć, że białka mogą wykonywać swoje funkcje nawet jeśli niektóre fragmenty pozostają nieuporządkowane [Chouard, 2011].

Oprócz dzikich typów (ang. *wild type*) kompleksów białko-białko, ważny problem, związany między innymi z projektowaniem leków, stanowi przewidywanie wpływu mutacji na stabilność kompleksów białko-białko [Moretti, 2013]. Zmiany w specyficzności (ang. *specificity*) i powinowactwie (ang. *affinity*) oddziałujących białek może prowadzić między innymi do niekontrolowanego wzrostu komórek i nowotworów [Nero, 2014].

W kontekście fizyki, każde dwa białka mają możliwość oddziaływania ze sobą; jednakże czy te oddziaływania mają sens biologiczny jest zależne od wielu czynników. Doświadczalnie rozwiązane struktury kompleksów białko-białko istnieją tylko dla małej części możliwych funkcjonalnie kompleksów tworzących interaktom. Z jednej strony, metody doświadczalne przewidywania struktury białek mają swoje ograniczenia. Z drugiej strony, modelowanie komputerowe trójwymiarowej struktury kompleksów białko-białko (dokowanie) jest wciąż uważane za złożony problem. Stąd też rozwój wydajnych metod obliczeniowych, rozpoczynając od przewidywania miejsc aktywnych (ang. *binding sites*) do określania poprawności struktur (ang. *scoring*) wciąż pozostaje aktywnym nurtem badań naukowych. Zadanie to jest wysoce interdyscyplinarne i obejmuje matematykę, chemię, fizykę, biologię i informatykę.

Warto zwrócić uwagę na rodzaj współdziałania pomiędzy eksperymentem i teorią. Na przykład w przypadku przewidywania struktury kompleksów białko-białko rozwój metod obliczeniowych jest wspomagany dzięki rozwiązaniom eksperymentalnie strukturalnym. Z drugiej strony poprawność wyników eksperymentalnych może być potwierdzona obliczeniami. Natomiast struktury kompleksów, które nie zostały rozwiązane eksperymentalnie mogą być przewidziane przez wcześniej sprawdzone metody obliczeniowe i tym samym pewne informacje mogą być uzyskane na temat mechanizmu tworzenia się kompleksu.

Aby utworzyć kompleks, receptor (większe białko) i ligand (mniejsze białko) muszą znaleźć siebie w zatłoczonym środowisku (rozpoznawanie molekularne). Jeden z możliwych mechanizmów tworzenia kompleksu (zob. Fig. 1 w [Blundell, 2006]) przewiduje tworzenie przejściowego kompleksu (ang. *transient encounter complex*) w wyniku niespecyficznych zderzeń kierowanych głównie przez oddziaływanie elektrostatyczne. Następnie wiele takich kompleksów bardzo szybko rozdzieli się, jednakże niektóre produktywnie kompleksy będą zmieniać orientację i zaczną zbliżać się do końcowej struktury w drugiej fazie kierowanej głównie przez usuwanie rozpuszczalnika (ang. *desolvation*). Dopasowanie oddziałujących powierzchni (interfejs) jest końcowym etapem procesu [Blundell, 2006].

Metody obliczeniowe dokowania molekularnego składają się z dwóch głównych części: przewidywania struktury trójwymiarowej kompleksu (ang. *sampling*) i tworzenia rankingu według używanej funkcji oceniającej (ang. *scoring*) [Vajda, 2013]. Problem ten jest wciąż nierozwiązany pomimo wielu podejść rozwijanych przez różne grupy badawcze [Saladin, 2010; Melquiond, 2010; London, 2010; Zacharias, 2010].

Problem dokowania można podzielić na część sztywną (ang. *rigid*) i giętką (ang. *flexible*). W dokowaniu sztywnym wykonywane jest tylko przeszukiwanie przy pomocy translacji i rotacji; ten typ dokowania może być opisywany przez mechanizm zamka i klucza (ang. *lock and key*). Jednakże wiele białek zmienia swój kształt w czasie tworzenia kompleksu, zatem by w pełni opisać ten proces potrzebujemy także przeszukiwania przestrzeni konformacyjnej co czyni problem dokowania jeszcze bardziej wymagającym obliczeniowo. Te zmiany mogą być lokalne lub kolektywne i mogą trwać od femtosekund do sekund [Wildman, 2007]. Dokowanie giętkie może być opisywane mechanizmami takimi jak wymuszone dopasowanie (ang. *induced fit*) i wybór konformacyjny (ang. *conformational selection*). Zostało

dowodzone, że nie istnieje jeden właściwy mechanizm tworzenia kompleksów; to czy tworzenie kierowane jest przez wymuszone dopasowanie czy wybór konformacyjny może być zależne od innych czynników, takich jak dynamika białka i stężenie ligandu [Greives, 2014]. Dwie główne drogi obliczeniowe radzenia sobie ze zmianami konformacyjnymi to dynamika molekularna i analiza modów normalnych [Andrusier, 2008]. Obie metody mogą prowadzić do podobnych rezultatów, jednakże dynamika molekularna jest metodą wymagającą większych mocy obliczeniowych [Ahmed, 2010].

Wartość otrzymana przy pomocy pewnej funkcji oceniającej (ang. *scoring function*) może być przyporządkowana do każdej struktury przewidzianej w procesie dokowania. Jednakże otrzymanie wiarygodnych funkcji oceniających oraz poprawne przewidywanie powinowactwa wciąż stanowią problem w obszarze dokowania molekularnego [Kastritis, 2013; Moal, 2013]. Jednym z powodów jest zależność od zbioru struktur używanych w czasie tworzenia takich funkcji [Vajda, 2013]. Dobrze działająca funkcja oceniająca powinna być w stanie odróżnić prawie natywny kompleks białko-białko od struktury niepoprawnej. Jednakże nie jest to zawsze możliwe więc inne czynniki mogą być przydatne w poszukiwaniu poprawnego miejsca aktywnego jak na przykład podobieństwo kształtów (ang. *shape complementarity*). Niestety okazuje się, że czasami interfejsy są płaskie i nie można odnaleźć żadnej oczywistej przestrzeni mogącej być miejscem aktywnym (ang. *binding pocket*) [Nero, 2014].

Oprócz poprawnej trójwymiarowej struktury kompleksu białko-białko zwróconej przez algorytm, równie ważna jest dynamika tworzenia się kompleksu, tzn. jak białka znajdują swoich partnerów w wielowymiarowej przestrzeni przejść konformacyjnych i jak w czasie tworzenia kompleksu powstają leje energetyczne oddziałujących białek (ang. *binding funnels*), tzn. jaka jest struktura przestrzeni stanów konformacyjnych i w jaki sposób białka zmieniają swoje stany konformacyjne kiedy przemierzają tą sieć. Idea leju energetycznego oddziałujących białek jest uogólnieniem leju tworzonego w czasie fałdowania pojedynczego białka [Dill, 1997; Tsai, 1999]. Jest to struktura energetyczna przypominająca lej, która staje się coraz bardziej wąska kiedy białka zbliżają się do siebie. Bardzo ważną cechą jest fakt istnienia wielu możliwych dróg, które prowadzą do dna (ang. „*funnel not tunnel*”); niektóre z nich mogą być częściej wybierane. Dynamika tworzenia kompleksów zależy od kształtu leja oraz entalpowych i entropowych wąskich gardeł (ang. *bottlenecks*). Rozmiar leja jest zależny od typu kompleksu [Hunjan, 2008].

Wciąż otwartym problemem jest jak generować leje energetyczne metodami obliczeniowymi. Pewne podejścia zostały stworzone by generować ścieżki struktur poprawnych stereochemicznie [Farrell, 2010; Ahmed, 2011] ale nie jest oczywiste jak uogólnić wiele ścieżek na lej i jaka jest pewność że wygenerowana struktura energetyczna jest pełna. Ostatnio dane eksperymentalne z *paramagnetic relaxation enhancement* wraz z dokowaniem sztywnym zostały użyte w celu badania kompleksów przejściowych [Kozakov, 2014]. Dokowanie sztywne pozwoliło znacząco zmniejszyć wymiarowość przestrzeni poszukiwań jednakże pełen problem wymaga uwzględnienia dokowania giętkiego.

Osobne pytanie, na które nie jest jeszcze znana odpowiedź to jak mutacje mogą wpływać na kształt leja; odpowiedź na to pytanie pozwoli na tworzenie wydajnych i zindywidualizowanych terapii biorących pod uwagę dynamikę oddziaływań.

7.1 Jak przewidzieć trójwymiarową strukturę kompleksów białko-białko? [H2, H5, S1]

Analiza modów normalnych to jedna z podstawowych metod służących modelowaniu giętkości (ang. *flexibility*) białek [Nicolai, 2014]. Metoda ta była rozlegle testowana w celu odpowiedzi na pytania odnośnie liczby modów normalnych potrzebnych do opisu przejścia oraz wielkości tej zmiany [Tama, 2001; Dobbins, 2008]. Warto zauważyć, że jedynie jednostkowe wektory przejścia są znane a nie znamy najlepszych amplitud. Jest to powodem, dla którego należy zaimplementować metodę wielowymiarowej optymalizacji, która wraz z funkcją oceniającą będzie kierowała przeszukiwanie przestrzeni w kierunku najbardziej obiecujących obszarów. Jedną z bardzo wydajnych metod jest Particle Swarm Optimization [Kennedy, 2001; Das, 2008], której idea pochodzi z obserwacji kolektywnego i społecznego zachowania organizmów takich jak mrówki czy pszczoły.

Kiedy rozpocząłem moją pracę dla Cancer Research UK, pierwsza wersja algorytmu dokowania giętkiego SwarmDock (część SwarmDock Server), została opublikowana [Moal, 2010]. Jednakże był to kod roboczy, nieprzygotowany do użytku przez inne osoby. Używając wiedzy zdobytej w nowej grupie badawczej rozpocząłem rozwijanie web serwisu który używa algorytmu SwarmDock i rozszerza go do narzędzia, które może być z łatwością używane przez użytkowników nawet z ograniczoną wiedzą na temat chemii obliczeniowej.

Głównym zadaniem wykonywanym przez SwarmDock Server jest znalezienie nieznannej związanej struktury kompleksu białko-białko na podstawie niezwiązanych struktur oddziałujących partnerów w formacie PDB i tym samym przewidywanie nieznanego miejsca aktywnego. Oprócz badań podstawowych nad własnościami oddziaływań białko-białko, web serwis może być też pomocny w projektowaniu leków, w celu określenia siły wiązania potencjalnych białkowych antagonistów do białkowego receptora o znaczeniu farmakologicznym.

Około dwa lata zajęło mi zaimplementowanie istniejących w grupie badawczej kodów, włączenie zewnętrznych narzędzi i rozwinięcie nowych programów (Python i C++), a następnie określenie jakości wyników (ang. *benchmark*) nowego web serwisu do dokowania białek. Oprócz stworzenia wieloetapowego mechanizmu obliczeń, równie ważne była kompatybilność z ciągle zmieniającym się środowiskiem obliczeniowym w Cancer Research UK i wspieranie użytkowników. Pliki PDB, w związku z ograniczeniami eksperymentalnymi, są często niekompletne. Czasami struktura danych wewnątrz pliku nie odpowiada formatowi PDB. Było wyzwaniem napisanie czytnika plików (ang. *parser*), który będzie w stanie poradzić sobie z wieloma problemami (brakujące atomy, brakujące residua, niestandardowe residua i inne) i poinformuje użytkownika jeśli przesłane pliki są tak mocno uszkodzone, że nie mogą być naprawione automatycznie. Zasoby obliczeniowe oferowane przez Cancer Research UK są w oczywisty sposób ograniczone stąd też jednym z zadań było napisanie mechanizmu umożliwiającego uczciwe współzawodnictwo pomiędzy zadaniami różnych użytkowników. Osobny techniczny problem stanowi zabezpieczenie web serwisu

przed potencjalnymi atakami hakerów, zapewniając możliwe najwyższy stopień bezawaryjności dla naukowców chcących używać serwisu.

W skrócie mechanizm działania SwarmDock Server obejmuje trzy fazy: faza wstępna (ang. *preprocessing*; naprawa i minimalizacja struktur, obliczanie modów normalnych dla receptora i ligandu przy pomocy ElNemo [Suhre, 2004]), właściwe dokowanie (generowanie około 120 punktów startowych dookoła receptora i uruchomienie mieszanej optymalizacji przy pomocy Particle Swarm Optimization i lokalnego przeszukiwania; optymalizacja pozycji i orientacji ligandu i współczynników przy modach normalnych dla obu partnerów) oraz faza końcowa (ang. *postprocessing*; minimalizacja, przyporządkowanie każdej ze struktur wartości potencjału rozwiniętego przez Tobi [Tobi, 2010], klastrowanie hierarchiczne w odległości 3Å; filtrowanie lejów energetycznych zostało dodane w ostatnim czasie).

SwarmDock Server pracuje w dwóch trybach: dokowanie w ciemno (ang. *full blind*) oraz z ograniczeniami (ang. *restrained*). W drugim trybie użytkownik może wybrać residua tworzące wg niego interfejs i tym samym ograniczyć przestrzeń przeszukiwania, a nawet zmusić serwer do wygenerowania rozwiązań w interesującej użytkownika części cząsteczki receptora.

Publikacja [H2] wprowadza nowe narzędzie jakim jest SwarmDock Server. W celu zmierzenia wydajności, testowałem serwis na Benchmark 4.0, zbiorze 176 par receptor/ligand w formie niezwiązanej (ang. *unbound*), o różnym stopniu trudności dokowania i funkcji, który był stworzony do testowania algorytmów dokowania [Hwang, 2010]. Wyniki ślepego dokowania używając SwarmDock Server były porównane z kompleksem białko-białko dostępnym w formie związanej (ang. *bound*), a współczynnik sukcesu (ang. *success rate*) rozumiany jako liczba kompleksów, dla których SwarmDock Server znalazł co najmniej jedno akceptowalne rozwiązanie (ang. *acceptable solution*) wyniósł 72% biorąc pod uwagę wszystkie rozwiązania i ok. 36% biorąc pod uwagę tylko 10 najlepszych rozwiązań (ang. *top 10*) wybranych wg procedury oceniającej (zob. Tabela 1 w [H2]). Dodatkowo SwarmDock Server został porównany z ClusPro [Comeau, 2004], obecnie najlepszym serwerem do dokowania molekularnego, na zbiorze wybranych w uczciwy sposób 29 par receptor/ligand. Wyniki porównania pokazały wysoką jakość narzędzia SwarmDock Server. W najnowszym Benchmark 5.0 okazał się najlepszym serwerem wśród testowanych [Vreven, 2015]. Dodatkowo od czerwca 2015 r. znacząco udało się podnieść liczbę poprawnych rozwiązań w top 10 dzięki implementacji nowego systemu demokratycznego oceniania opartego na algorytmie uczenia maszynowego (ang. *machine learning*) i jednym z systemów głosowania (praca w przygotowaniu).

Osobny sposób by sprawdzić algorytmy dokowania to wzięcie udziału w międzynarodowym konkursie dokowania białko-białko CAPRI (Critical Assessment of PRediction of Interactions), gdzie rozwiązane eksperymentalnie struktury związane kompleksów białko-białko są utrzymywane w tajemnicy, a algorytmy dokowania są testowane z danymi wejściowymi w postaci struktur niezwiązanych; czasami tylko sekwencja jest dana, więc należy dodatkowo wymodelować strukturę trójwymiarową niezwiązanego białka na podstawie sekwencji. Istnieją dwie drogi by brać udział: serwer automatyczny (zwykle 48 godzin by przesłać rozwiązania) i tryb manualny (zwykle 2 tygodnie by przesłać rozwiązania). Ponadto istnieją

dwa etapy: przewidywanie (ang. *prediction*) oraz ocenianie (ang. *scoring*). Uczestnicy etapu przewidywania przesyłają 10 modeli. Otrzymane modele są poddawane analizie i klasyfikowane jako wysokiej jakości (ang. *highly accurate*), średniej jakości (ang. *medium accuracy*), akceptowalne (ang. *acceptable*) lub niepoprawne (ang. *incorrect*) biorąc pod uwagę liczbę natywnych kontaktów (ang. *fraction of native contacts*), RMSD (ang. *root mean square deviation*) łańcucha głównego ligandu oraz RMSD łańcucha głównego dla reszduów tworzących interfejs.

SwarmDock Server rozpoczął uczestnictwo w automatycznym przewidywaniu struktur od czterech ostatnich kompleksów CAPRI 5th Assessment (2010-2013) [Lensink, 2013] i był jedynym serwerem, który stworzył co najmniej akceptowalne rozwiązanie dla każdego z czterech problemów (T53, T54, T57, T58); szczegóły można odnaleźć w Supplementary Table S11A w [Lensink, 2013]. Dla przykładu dla kompleksu T54 tylko cztery grupy przewidziały poprawne rozwiązanie – trzy manualne i SwarmDock Server. Po wszystkich rundach Biomolecular Modelling Laboratory otrzymało drugie miejsce na świecie, a SwarmDock Server siedemnaste (jednakże brał udział tylko w ostatnich czterech rundach w związku z momentem opublikowania). W najnowszych zestawieniach jest już w pierwszej piątce [Lensink, 2015].

W części dotyczącej oceniania struktur (ang. *scoring part*), każda osoba przewidująca struktury kompleksów jest proszona o przesłanie 100 modeli, a organizatorzy tworzą zbiór około 1000 modeli. Celem tej części jest wybór 10 najlepszych modeli z tego zbioru. W tej części CAPRI, Biomolecular Modelling Laboratory uzyskało trzecie miejsce (Supplementary Table S11B w [Lensink, 2013]). Niedawno zostały ogłoszone wyniki dla kompleksu T59 (pierwszy kompleks z nowego zbioru rund rozpoczętych w 2013 roku); SwarmDock nie był w stanie znaleźć akceptowalnej struktury automatycznie, jednakże używając SwarmDock Server i ręcznego wyboru najlepszych struktur akceptowalna struktura była wysłana do ręcznej części konkursu. Znanym problemem algorytmów dokowania jest fakt, że czasami poprawna struktura jest gdzieś w zbiorze rozwiązań ale nie w najlepszych 10 wg używanej funkcji oceniającej, stąd też wciąż jest potrzeba by poprawiać procedury oceniające jakość struktur. Poprawa do współczynnika sukcesu dla zbioru 10 najlepszych struktur dla SwarmDock Server jest tematem pracy [H3], opisaną w następnym podpunkcie.

Ostatnia runda CAPRI (30; T68 – T94; maj – sierpień 2014) była przygotowana w ramach CASP11. Większość kompleksów stanowiły homodimery, było kilka homotetramerów i heterodimerów. Runda ta była znacząco trudniejsza gdyż tylko sekwencja była dana dla obu partnerów, a sukces w dokowaniu zależał też od poprawnego modelowania struktury niezwiązanej (dla części automatycznej użyłem struktury niezwiązanej wygenerowanej przez serwer rozwijany w grupie badawczej, dla części ręcznej oceniłem dla każdego monomeru jakość 150 struktur przygotowanych przez uczestników CASP11 i wybrałem najlepszą do dokowania). W przypadku homodimerów, zaprojektowaliśmy nową funkcję oceniającą stworzoną przy pomocy operatorów symetrii. Wyniki zarówno dla części przewidywania jak i oceniania są dobre [Lensink, 2015].

Podręcznik opisujący jak używać SwarmDock Server został opublikowany jako rozdział w książce [H5]. O ile publikacja [H2] pokazuje głównie wyniki testów wydajności narzędzia

(ang. *benchmark results*), w publikacji [H5] czytelnik może odnaleźć dodatkowe szczegóły na temat etapów działania serwisu oraz trzy przykłady (w tym dwa układy przeciwciało/antygen), do których dołączone są dodatkowe pliki by pomóc w zrozumieniu sposobu użycia i wyników. Rozdział w książce zawiera szczegółowy opis dodatkowych funkcjonalności jak np. dokowanie z ograniczeniami, które używa idei RayTracing (punkt początkowy dokowania musi „wiedzieć” co najmniej jeden atom residuum wybranego jako ograniczenie); na obrazku 3 w [H5] można zobaczyć wizualizację wyników w trybie ślepych i z ograniczeniami. Dodatkowe informacje dla użytkowników umieszczone są w sekcji „Notes”.

Niektóre funkcjonalności, jak np. przewidywanie miejsca aktywnego (ang. *binding site prediction*) zaimplementowałem do serwera ale testy wydajności są dostępne tylko na stronie SwarmDock Server [S1].

Najlepszym uczuciem dla twórcy jest gdy inni ludzie używają oprogramowania. Ostatnio kilka publikacji eksperymentalnych zacytowało SwarmDock Server. Dodatkowo ponad 350 użytkowników z różnych krajów użyło serwisu od momentu opublikowania, wysyłając ponad 2000 zadań. Dodatkowo, jedna z amerykańskich spółek, DNASTAR Inc., kupiła prawa do stworzenia komercyjnej wersji SwarmDock Server (wspierałem transfer kodu).

7.2 Jak odróżnić prawdziwe (ang. *true positive*) leje energetyczne oddziałujących białek (ang. *protein-protein binding funnels*) od tych, które mogą sprawiać wrażenie prawdziwych (ang. *false positive*)? [H3, S1]

Rozwiązania z SwarmDock Server (dla ślepego dokowania około 500, około 120 punktów startowych i cztery uruchomienia z każdego punktu startowego) są klastrowane przez algorytm opisany w szczegółach w [H5], gdzie 10 najlepszych rozwiązań to najniżej energetyczni członkowie pierwszych 10 klastrów uporządkowanych przez potencjał punktowy opisany przez Tobi [Tobi, 2010]. Jak wspomniano powyżej, nawet jeśli poprawna struktura kompleksu białko-białko znajduje się gdzieś w zbiorze struktur wyprodukowanych przez algorytm, jest bardzo trudne by zawsze zwrócić najpoprawniejsze struktury w zbiorze 10 najlepszych rozwiązań. Jest to związane z faktem, że funkcje oceniające są tylko przybliżeniem a lej energetyczny z najniżej energetyczną strukturą może tylko przypominać prawdziwy lej (ang. *false positive*; dno prawdziwego leja może nie być osiągnięte więc nie istnieje najniżej energetyczny poprawny model pochodzący z poprawnego leja energetycznego). Nie ma też żadnego wzoru w wartościach funkcji oceniającej, który mógłby pomóc w znalezieniu prawdziwego regionu oddziaływania między białkami (widać to na obrazku 1 w [H3]).

W ogólności nie jest oczywiste jak budować przestrzeń stanów konformacyjnych i czy architektura takiej sieci jest kompletna. W publikacji [H3] zajmuję się badaniem lejów energetycznych oddziaływania między białkami w celu poprawienia współczynnika sukcesu (liczba poprawnych struktur) w zbiorze 10 „najlepszych” wg funkcji oceniającej struktur. Metodologia jest ogólna i może być użyta do zbioru rozwiązań stworzonego przez inne algorytmy. Używam wszystkich rozwiązań SwarmDock Server ze ślepego dokowania (obrazek 1 w [H3]) i traktuję je jako stany w sieci stanów konformacyjnych (obrazek 2 w [H3]), którą

rzędzi dynamika markowowska. Przydzielam prawdopodobieństwa przejść podobnie do podejścia *stochastic roadmap simulations* [Apaydin, 2002]. W ogólności sieć nie jest spójna w związku z wybraną zasadą do tworzenia połączeń między stanami – różnica między RMSD ligandu nie może przekraczać 6Å; każda podsieć tworzy osobny lej energetyczny. Następnie obliczane są obsadzenia stanów aby znaleźć maksymalną wartość dla każdego leja. W pracy [H3] pokazano, że usunięcie struktur energetycznych, które nie przypominają leja za pomocą filtrowania rozwiązań z niskim maksymalnym obsadzeniem (poniżej wcześniej znalezionej wartości), może poprawić ranking otrzymanych z dokowania modeli; współczynnik sukcesu w ramach 10 zwróconych modeli poprawił się o 22% w porównaniu do wyników przedstawionych w [H2].

Metodologia ta została zaimplementowana w SwarmDock Server [S1]. Dodatkowo przestrzeń stanów konformacyjnych jest tworzona w formacie RaTrav opisanym w [H7] i może być używana do dalszych analiz.

Podobna metodologia była przeze mnie używana do oceniania modeli kompleksów białko-białko dla kompleksu CAPRI T59. Wyniki części oceniającej zostały opublikowane (http://www.ebi.ac.uk/msd-srv/capri/round28/R28_T59/resultScorer.html) a moja automatyczna metodologia oceniająca OccuScore (id S12, Torchala and Bates) była w grupie 9 zwycięskich z 29 uczestniczących grup. Ponadto OccuScore był jedynym automatycznym serwerem, który zwrócił poprawne rozwiązanie w części oceniającej dla kompleksu T59.

7.3 Jak dobre są współcześnie dostępne funkcje oceniające (ang. *scoring functions*) i czy mogą być jeszcze lepsze jeśli połączy się je ze sobą? [H6]

Wiele metod zostało stworzonych w celu oceniania poprawności struktur otrzymanych w procesie dokowania molekularnego. W ogólności funkcja oceniająca powinna działać dla każdej struktury kompleksu białko-białko, nie tylko dla struktur podobnych do tych na jakich funkcja była nauczana. Jednakże nie istnieje żadna metoda, która działa idealnie dla dowolnego modelu dostarczając 100% pewności co do poprawności rozwiązania lub też zdolna uporządkować modele względem ich bliskości do struktury natywnej.

Metodologia opisana w [H3], związana z filtrowaniem lejów energetycznych, była oparta na jednym potencjale. W publikacji [H6] opisujemy wyniki testów wydajności 115 funkcji oceniających na zbiorze modeli obejmujących 118 kompleksów, dla których bliska natywnej struktura została znaleziona przez SwarmDock Server rozpoczynając ze struktur niezwiązanych. Byliśmy w stanie zidentyfikować pewne znane strategie służące do oceniania modeli jak i znaleźć nowe podejścia. Przeprowadziliśmy osobne analizy dla sztywnych i giętkich kompleksów białko-białko. Analizowaliśmy i klastrowaliśmy pojedyncze funkcje oceniające, policzyliśmy współczynniki sukcesu i prawdopodobieństwa warunkowe znalezienia poprawnej struktury o danej jakości lub lepszej. Następnie analizowaliśmy pary funkcji oceniających posługując się kilkoma pojęciami teorii zbiorów.

W części Supplementary Materials zamieściliśmy szczegóły na temat wydajności funkcji oceniających. Dla przykładu z Supplementary Table 2 jest widoczne, że potencjał, który używam do oceniania struktur w SwarmDock Server, CP_TSC [Tobi, 2010], radzi sobie bardzo

dobrze w porównaniu do innych funkcji oceniających. Dodatkowo możliwe jest połączenie go z innymi funkcjami oceniającymi by otrzymać jeszcze lepszą wydajność (Supplementary Table 3), szczególnie gdy różnorodne funkcje oceniające rozpoznają różne aspekty niekowalencyjnego wiązania między białkami. Jednakże wciąż nie jest oczywiste jaka jest najlepsza forma funkcyjna aby połączyć kilka funkcji oceniających. Z drugiej strony całkiem popularne obecnie jest używanie metod *machine learning* z różnymi parametrami zamiast szukać jednej zwięzłej formy funkcyjnej dla funkcji oceniającej.

7.4 Jak analizować leje energetyczne oddziałujących białek (ang. *protein-protein binding funnels*) przy użyciu średnich czasów pierwszego przejścia? [H1, H7, S1, S2]

Jak pokazano wcześniej, prawdopodobieństwa obsadzeń pomogły w odróżnianiu lejów energetycznych. W celu zrozumienia dynamiki stanów konformacyjnych w poprawnym leju energetycznym (ang. *true positive binding funnel*) w publikacji [H7] posłużyłem się średnimi czasami pierwszego przejścia (MFPT, ang. *mean first-passage time*).

Przed zastosowaniem do rzeczywistych układów biologicznych, metody służące do obliczania średnich czasów pierwszego przejścia zostały zaimplementowane w programach komputerowych i przetestowane na prostych sieciach. Jest to temat pracy [H1], która porównuje algorytmy Hilla i Monte Carlo na prostych sieciach w kontekście dokładności i złożoności czasowej obliczeń zależnych od typu i rozmiaru sieci. Pokazaliśmy, że przy zadanej odległości do pokonania, wynik obliczenia konkretnego średniego czasu pierwszego przejścia zależy od wyboru stanu początkowego i końcowego oraz od ich wzajemnego położenia na sieci. Pokazaliśmy też, że jeśli sieć zawiera małą liczbę cykli, wtedy metoda Hilla dostarcza dokładnych wyników szybciej niż stochastyczna metoda symulacji Monte Carlo. Ta praca była ważna w kontekście zrozumienia obecnych ograniczeń obliczeniowych.

Następnie napisaliśmy RaTrav, narzędzie do obliczenia średnich czasów pierwszego przejścia i prawdopodobieństw obsadzeń, które może działać na dowolnych sieciach, wspiera sieci niespójne oraz czasy lokalne (różne wagi krawędzi). Oczywiście jest, że program wymagał dodatkowych godzin rozwijania i testowania. Sieć jest zdefiniowana w pliku tekstowym przy użyciu specjalnych zaproponowanych przeze mnie słów kluczowych. Sposób używania i zastosowania RaTrav zostały zamieszczone w pracy [H7]. Faktem jest, że sieci stanów konformacyjnych dla układów biologicznych zawierają wiele cykli, stąd też metoda symulacji Monte Carlo stanowi rozsądny wybór.

Przykład 1 w [H7] jest poświęcony analizie poprawnego leju energetycznego. Wybrałem lej dla kompleksu białko wiążące witaminę D/aktyna (kod w bazie PDB: 1KXP), który wcześniej był wykorzystany w pracy [H3] w celu pokazania przykładu sieci stanów konformacyjnych otrzymanej przez SwarmDock Server. Obliczyłem średnie czasy pierwszego przejścia pomiędzy wszystkimi stanami sieci i następnie użyłem tych czasów jako wag krawędzi. Jako ostatni etap, wykorzystałem algorytm Dijkstry aby znaleźć najkrótszą ścieżkę w sieci stanów konformacyjnych ze stanu niepoprawnego do stanu najbliższego natywnej konformacji kompleksu. Przy pomocy średnich czasów pierwszego przejścia uprzywilejowana ścieżka została odnaleziona (obrazek 2 w [H7]), jak również przejścia które ograniczają przemieszczanie się w tej sieci. Warto zauważyć, że RaTrav jest ogólnym narzędziem i może być używany do projektów z różnych dziedzin, gdzie należy obliczyć średnie czasy pierwszego

przejścia. Ponadto pełny kod źródłowy jest dostępny co umożliwi innym osobom rozwój kodu według ich potrzeb [S2].

Metodologia generowania sieci stanów konformacyjnych opisana w [H3] i [H7] została już przeze mnie zaimplementowana w SwarmDock Server [S1]. Użytkownik może wygenerować przestrzeń stanów konformacyjnych (używany jest potencjał OPUS-PSP [Lu, 2008]) i wykonać obliczenia na tej sieci przy użyciu RaTrav. Publikacja [H7] zawiera Supplementary Files dla użytkowników w celu lepszego zrozumienia oprogramowania.

7.5 Jak mutacje wpływają na szybkość dysocjacji kompleksów białko-białko? [H4]

Mutacje powodują, że białka dysocjują szybciej lub wolniej, a tym samym zmieniają dynamikę molekularnej sieci oddziaływań. Destabilizujące mutacje w obszarze interfejsu są cechą charakterystyczną chorób takich jak nowotwory. Z drugiej strony, wykrywanie stabilizujących mutacji jest kluczowe dla projektowania białek i efektywnego projektowania leków. Szybkość z jaką białka łączą się ze sobą tworząc kompleks jest ograniczona przez szybkość dyfuzji i pewne czynniki farmakologiczne. Z drugiej strony, szybkość dysocjacji jest głównie zależna od oddziaływań krótkozasięgowych. Stąd też metody obliczenia i minimalizacji szybkości dysocjacji okazały się bardzo ważnym elementem w optymalizacji projektowania nowych leków. W czasie publikacji pracy [H4], nie istniała żadna metoda aby obliczeniowo oszacować zmiany w szybkości dysocjacji w związku z mutacją.

W publikacji [H4] byliśmy głównie zainteresowani opracowaniem nowej strategii obliczeniowej aby oszacować wpływ mutacji na szybkość dysocjacji oddziaływania między białkami. Oprócz numerycznego przewidywania wartości, dla którego osiągnęliśmy wysokie korelacje, skupiliśmy się również na wykrywaniu bardziej rzadkich mutacji stabilizujących, gdyż wykrywanie tego typu mutacji jest ważne w kontekście strategii projektowania leków. Pokazaliśmy, że wartość $\Delta\Delta G$ mutacji punktowej do alaniny może być użyta do oszacowania szybkości dysocjacji po mutacji do dowolnego typu residuum, a także po wielokrotnych mutacjach. Jest to nowa strategia opisywania wpływu mutacji, a ponadto biorąc pod uwagę, że większość baz danych mutacji zawiera głównie mutacje do alaniny (ang. *alanine-scanning experiments*), metodologia ta pozwala na zastosowanie tych danych. Innym ważnym elementem jest fakt, że dla danej mutacji nie tylko rozważaliśmy lokalne efekty ale także zmiany w globalnej architekturze i dystrybucji regionów silnego wiązania i efekty współpracy między tymi regionami. Nasze wyniki wskazują, że dystrybucja regionów krytycznych dla stabilności na interfejsie białko-białko jest funkcją wielkości kompleksu.

Podsumowanie

Najważniejsze wyniki są następujące:

1. Rozwój SwarmDock Server, darmowego narzędzia do giętkiego dokowania par białko-białko, który jest używany także przez zewnętrzne grupy badawcze. Wysokie wyniki w Critical Assessment of PRediction of Interactions (CAPRI) jako członek grupy Paula Batesa (ręczne przewidywanie) oraz jako SwarmDock Server (automatyczne przewidywanie). Zaimplementowanie metodologii generowania przestrzeni stanów konformacyjnych w SwarmDock Server co pozwala użytkownikom na generowanie

sieci stanów konformacyjnych dla dowolnej pary białko-białko i przeprowadzanie różnorodnych obliczeń na tej sieci.

2. Zastosowanie modelu łańcuchów Markowa do dokowania białko-białko i pokazanie, że filtrowanie struktur energetycznych nie przypominających prawdziwego leja energetycznego może znacząco poprawić ranking modeli kompleksu białko-białko.
3. Pokazanie, że połączenie tradycyjnych metod oceniających modele jest korzystne w celu poprawy pozycji poprawnych kompleksów białko-białko w rankingu wygenerowanych struktur.
4. Rozwój RaTrav, wolnego i otwartego (ang. *open source*) narzędzia do obliczania średnich czasów pierwszego przejścia i prawdopodobieństw obsadzeń. Pokazanie jak używać RaTrav w celu znalezienia uprzywilejowanych ścieżek i przejść ograniczających poruszanie się w sieci stanów konformacyjnych przy pomocy średnich czasów pierwszego przejścia.
5. Pokazanie, że wartość $\Delta\Delta G$ mutacji punktowej do alaniny może być użyta do oszacowania szybkości dysocjacji po mutacji do dowolnego residuum oraz mutacji wielokrotnych.

8. INNE OSIĄGNIĘCIA PO DOKTORACIE

1. Uczestniczyłem w rundzie CAPRI z kompleksami T55 i T56 poświęconej przewidywaniu wpływu mutacji na oddziaływania białko-białko, w której 22 grupy przewidywały efekt mutagenyzy dla dwóch zaprojektowanych receptorów hemaglutyniny wirusa grypy [Moretti, 2013]. Publikacja [Moretti, 2013] nie została włączona do cyklu ze względu na dużą liczbę autorów (publikacja społeczności dokowania molekularnego).
2. Uczestniczyłem we wspólnym konkursie CASP/CAPRI poświęconemu przewidywaniu struktur homo i heterokompleksów (dimerów, trimerów, tetramerów) oraz znajdowaniu najlepszych struktur w wygenerowanym zbiorze (scoring). W ramach konkursu testowałem podejście automatyczne (oznaczone jako SwarmDock) oraz manualne (oznaczone jako Bates) z dobrymi wynikami [Lensink, 2015]. Publikacja [Lensink, 2015] nie została włączona do cyklu ze względu na dużą liczbę autorów (publikacja społeczności dokowania molekularnego).
3. Uczestniczyłem w pracach sieci badawczej dotyczącej testowania możliwości algorytmów dokowania molekularnego oraz przewidywania powinowactwa. Każda z grup niezależnie testowała rozwijane narzędzia badawcze. SwarmDock Server okazał się najlepszym narzędziem w ramach testowanych metod [Vreven, 2015]. Publikacja [Vreven, 2015] nie została włączona do cyklu ze względu na dużą liczbę autorów (publikacja sieci badawczej).

BIBLIOGRAFIA

[H1]-[H7], [S1], [S2] punkty 4 i 5 powyżej.

[Ahmed, 2010] A. Ahmed, S. Villinger, H. Gohlke, 'Large-scale comparison of protein essential dynamics from molecular dynamics simulations and coarse-grained normal mode analysis', *Proteins: Structure, Function, and Bioinformatics* **78**, 3341-3352 (2010).

[Ahmed, 2011] A. Ahmed, F. Rippmann, G. Barnickel, H. Gohlke, 'A normal mode based geometric simulation approach for exploring biologically relevant conformational transitions in proteins', *J. Chem. Inf. Model.* **51**, 1604-1622 (2011).

- [Andrusier, 2008] N. Andrusier, E. Mashiach, R. Nussinov, H. J. Wolfson, 'Principles of flexible protein-protein docking', *Proteins: Structure, Function, and Bioinformatics* **73**, 271-289 (2008).
- [Apaydin, 2002] M. S. Apaydin, C. E. Guestrin, C. Varma, D. L. Brutlag, J-C. Latombe, 'Stochastic roadmap simulation for the study of ligand-protein interactions', *Bioinformatics* **18 Suppl. 2**, S18-S26 (2002).
- [Blundell, 2006] T. L. Blundell, J. Fernandez-Recio, 'Brief encounters bolster contacts', *Nature* **444**, 279-280 (2006).
- [Branden, 1999] C. Branden, J. Tooze, *Introduction to Protein Structure* 2nd Edition, Garland Publishing Inc., 1999.
- [Chouard, 2011] T. Chouard, 'Breaking the protein rules', *Nature* **471**, 151-153 (2011).
- [Comeau, 2004] S. R. Comeau, D. W. Gatchell, S. Vajda, C. J. Camacho, 'ClusPro: an automated docking and discrimination method for the prediction of protein complexes', *Bioinformatics* **20**, 45-50 (2004).
- [Das, 2008] S. Das, A. Abraham, A. Konar, 'Swarm Intelligence Algorithms in Bioinformatics', w: *Computational Intelligence in Bioinformatics (Studies in Computational Intelligence Vol. 94)*, Rozdział 4, Springer, 2008.
- [Dill, 1997] K. A. Dill, H. S. Chan, 'From Levinthal to pathways to funnels', *Nature Struct. Biol.* **4**, 10-19 (1997).
- [Dobbins, 2008] S. E. Dobbins, V. I. Lesk, M. J. E. Sternberg, 'Insights into protein flexibility: the relationship between normal modes and conformational change upon protein-protein docking', *Proc. Natl. Acad. Sci. USA* **105**, 10390-10395 (2008).
- [Farrell, 2010] D. W. Farrell, K. Speranskiy, M. F. Thorpe, 'Generating stereochemically acceptable protein pathways', *Proteins: Structure, Function, and Bioinformatics* **78**, 2908-2921 (2010).
- [Greives, 2014] N. Greives, H-X. Zhou, 'Both protein dynamics and ligand concentration can shift binding mechanism between conformational selection and induced fit', *Proc. Natl. Acad. Sci. USA* **111**, 10197-10202 (2014).
- [Hunjan, 2008] J. Hunjan, A. Tovchigrechko, Y. Gao, I. A. Vakser, 'The size of the intermolecular energy funnel in protein-protein interactions', *Proteins: Structure, Function, and Bioinformatics* **72**, 344-352 (2008).
- [Hwang, 2010] H. Hwang, T. Vreven, J. Janin, Z. Weng, 'Protein-protein docking benchmark version 4.0', *Proteins: Structure, Function, and Bioinformatics* **78**, 3111-3114 (2010).
- [Kastritis, 2013] P. L. Kastritis, A. M. J. J. Bonvin, 'On the binding affinity of macromolecular interactions: daring to ask why proteins interact', *J. R. Soc. Interface* **10**, 20120835 (2013).
- [Kennedy, 2001] J. Kennedy, R. C. Eberhart, *Swarm Intelligence*, Morgan Kaufmann Publishers, 2001.
- [Kozakov, 2014] D. Kozakov, K. Li, D. R. Hall, D. Beglov, J. Zheng, P. Vakili, O. Schueler-Furman, I. Ch. Paschalidis, G. Marius Clore, S. Vajda, 'Encounter complexes and dimensionality reduction in protein-protein association', *eLife* **3**, e01370 (2014).
- [Lensink, 2013] M. F. Lensink, S. J. Wodak, 'Docking, scoring and affinity prediction in CAPRI', *Proteins* **81**, 2082-2095 (2013).
- [Lensink, 2015] M. F. Lensink, S. Velankar, A. Kryshtafovych, S-Y. Huang, D. Schneidman, A. Sali, J. Segura, N. Fernandez-Fuentes, S. Viswanath, R. Elber, S. Grudin, P. Popov, E. Neveu, H. Lee, M. Baek, S. Park, L. Heo, G. Rie Lee, C. Seok, S. Qin, H-X. Zhou, D. W. Ritchie, B. Maigret, M-D. Devignes, A. Ghoorah, M. Torchała, R. A. G. Chaleil, P. A. Bates, E. Ben-Zeev, M. Eisenstein, S. S. Negi, Z. Weng, T. Vreven, B. G. Pierce, T. M. Borrmann, J. Yu,

F. Ochsenbein, R. Guerois, A. Vangone, J. P. G. L. M. Rodrigues, G. van Zundert, M. Nellen, L. Xue, E. Karaca, A. S. J. Melquiond, K. Visscher, P. L. Kastiris, A. M. J. J. Bonvin, X. Xu, L. Qiu, C. Yan, J. Li, Z. Ma, J. Cheng, X. Zou, Y. Shen, L. X. Peterson, H-R. Kim, A. Roy, X. Han, J. Esquivel-Rodriguez, D. Kihara, X. Yu, N. J. Bruce, J. C. Fuller, R. C. Wade, I. Anishchenko, P. J. Kundrotas, I. A. Vakser, K. Imai, K. Yamada, T. Oda, T. Nakamura, K. Tomii, C. Pallara, M. Romero-Durana, B. Jimenez-García, I. H. Moal, J. Fernandez-Recio, J. Young Joung, J. Yun Kim, K. Joo, J. Lee, D. Kozakov, S. Vajda, S. Mottarella, D. R. Hall, D. Beglov, A. Mamonov, B. Xia, T. Bohnuud, C. A. Del Carpio, E. Ichiishi, J. Gray, E. Chermak, L. Cavallo, R. Oliva, S. J. Wodak, 'Prediction of homo- and hetero-protein complexes by ab-initio and template-based docking: a CASP-CAPRI experiment', *Proteins: Structure, Function, and Bioinformatics*, 2015, in press.

[London, 2010] N. London, O. Schueler-Furman, 'High-resolution protein-protein docking', w: *Protein-Protein Complexes. Analysis, Modeling and Drug Design*, Rozdział 8, Imperial College Press, 2010.

[Lu, 2008] M. Lu, A. D. Dousis, J. Ma, 'OPUS-PSP: an orientation-dependent statistical all-atom potential derived from side-chain packing', *J. Mol. Biol.* **376**, 288-301 (2008).

[Melquiond, 2010] A. S. J. Melquiond, A. M. J. J. Bonvin, 'Data-driven docking. Using external information to spark the biomolecular rendez-vous', w: *Protein-Protein Complexes. Analysis, Modeling and Drug Design*, Rozdział 7, Imperial College Press, 2010.

[Moal, 2010] I. H. Moal, P. A. Bates, 'SwarmDock and the use of normal modes in protein-protein docking', *Int. J. Mol. Sci.* **11**, 3623-3648 (2010).

[Moal, 2013] I. H. Moal, R. Moretti, D. Baker, J. Fernandez-Recio, 'Scoring functions for protein-protein interactions', *Curr. Opinion Struct. Biol.* **23**, 862-867 (2013).

[Moretti, 2013] R. Moretti, S. J. Fleishman, R. Agius, M. Torchala, P. A. Bates, P. L. Kastiris, J. P. G. L. M. Rodrigues, M. Trellet, A. M. J. J. Bonvin, M. Cui, M. Rooman, D. Gillis, Y. Dehouck, I. Moal, M. Romero-Durana, L. Perez-Cano, C. Pallara, B. Jimenez, J. Fernandez-Recio, S. Flores, M. Pacella, K. Praneeth Kilambi, J. J. Gray, P. Popov, S. Grudin, J. Esquivel-Rodriguez, D. Kihara, N. Zhao, D. Korkin, X. Zhu, O. N. A. Demerdash, J. C. Mitchell, E. Kanamori, Y. Tsuchiya, H. Nakamura, H. Lee, H. Park, C. Seok, J. Sarmiento, S. Liang, S. Teraguchi, D. M. Standley, H. Shimoyama, G. Terashi, M. Takeda-Shitaka, M. Iwadate, H. Umeyama, D. Beglov, D. R. Hall, D. Kozakov, S. Vajda, B. G. Pierce, H. Hwang, T. Vreven, Z. Weng, Y. Huang, H. Li, X. Yang, X. Ji, S. Liu, Y. Xiao, M. Zacharias, S. Qin, H. Zhou, S. Huang, X. Zou, S. Velankar, J. Janin, S. J. Wodak and D. Baker, 'Community-wide evaluation of methods for predicting the effect of mutations on protein-protein interactions', *Proteins: Structure, Function, and Bioinformatics* **81**, 1980-1987 (2013).

[Nero, 2014] T. L. Nero, C. J. Morton, J. H. Holien, J. Wielens, M. W. Parker, 'Oncogenic protein interfaces: small molecules, big challenges', *Nature Rev. Cancer* **14**, 248-262 (2014).

[Nicolai, 2014] A. Nicolai, P. Delarue, P. Senet, 'Low-frequency, functional, modes of proteins: all-atom and course-grained normal modes analysis', w: *Computational Methods to Study the Structure and Dynamics of Biomolecules and Biomolecular Processes (Springer Series in Bio-/Neuroinformatics)*, Rozdział 15, Springer, 2014.

[Saladin, 2010] A. Saladin, C. Prevost, 'Protein-protein docking', w: *Protein-Protein Complexes. Analysis, Modeling and Drug Design*, Rozdział 6, Imperial College Press, 2010.

[Suhre, 2004] K. Suhre, Y.-H. Sanejouand, 'ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement', *Nucleic Acids Res.* **32** (Web Server issue), W610-W614 (2004).

[Tama, 2001] F. Tama, Y.-H. Sanejouand, 'Conformational change of proteins arising from normal mode calculations', *Prot. Eng.* **14**, 1-6 (2001).

- [Tobi, 2010] D. Tobi, 'Designing coarse grained-and atom based-potentials for protein-protein docking', *BMC Struct. Biol.*, **10**, 40 (2010).
- [Tsai, 1999] C-J. Tsai, S. Kumar, B. Ma, R. Nussinov, 'Folding funnels, binding funnels, and protein function', *Protein Science* **8**, 1181-1190 (1999).
- [Vajda, 2013] S. Vajda, D. R. Hall, D. Kozakov, 'Sampling and scoring: A marriage made in heaven', *Proteins: Structure, Function, and Bioinformatics* **81**, 1874-1884 (2013).
- [Vreven, 2015] T. Vreven, I. H. Moal, A. Vangone, B. G. Pierce, P. L. Kastiris, M. Torchala, R. Chaleil, B. Jimenez-Garcia, P. A. Bates, J. Fernandez-Recio, A. M. Bonvin, 'Updates to the integrated protein-protein interaction benchmarks: Docking benchmark version 5 and affinity benchmark version 2', *J. Mol. Biol.* **427**, 3031-3041 (2015).
- [Wildman, 2007] K. Henzler-Wildman, D. Kern, 'Dynamic personalities of proteins', *Nature* **450**, 964-972 (2007).
- [Zacharias, 2010] M. Zacharias, 'Scoring and refinement of predicted protein-protein complexes', w: *Protein-Protein Complexes. Analysis, Modeling and Drug Design*, Rozdział 9, Imperial College Press, 2010.

Mieczysław Torchala