



UNIwersytet WarsZawski
WYDZIAŁ CHEMII



**Rozwój wieloskalowych metod molekularnego modelowania
białek oraz ich zastosowanie w badaniach mechanizmów
funkcjonowania białek**

dr Sebastian Kmiecik

Warszawa, październik 2014 r.

**Autoreferat – załącznik nr 1 do wniosku o nadanie stopnia naukowego doktora
habilitowanego**

Autoreferat

1. Imię i Nazwisko

Sebastian Kmiecik

2. Posiadane dyplomy, stopnie naukowe/ artystyczne – z podaniem nazwy, miejsca i roku ich uzyskania oraz tytułu rozprawy doktorskiej.

09.2007 – Doktor nauk chemicznych, specjalność: chemia teoretyczna. Wydział Chemii, Uniwersytet Warszawski. Tytuł rozprawy doktorskiej: "Wieloskalowe modelowanie mechanizmów zwijania białek globularnych", promotor: prof. dr hab. Andrzej Koliński.

09.2002 – Magister chemii, Wydział Chemii, Uniwersytet Warszawski.

3. Informacje o dotychczasowym zatrudnieniu w jednostkach naukowych/ artystycznych.

od 03.2010 do dziś – Wydział Chemii, Uniwersytet Warszawski

Stanowisko: specjalista naukowy

od 10.2007 do 02.2010 – Selvita, Kraków.

Stanowisko: Kierownik działu chemii obliczeniowej (od 02.2009 do 02.2010); Kierownik projektów (od 02.2008 do 01.2009); Architekt systemów (od 10.2007 do 01.2008).

Selvita jest firmą badawczo-rozwojową zajmującą się dostarczaniem kompleksowych rozwiązań z zakresu chemii, biotechnologii oraz bioinformatyki dla jednostek badawczo-rozwojowych i różnych gałęzi przemysłu. W ramach pracy w firmie Selvita zajmowałem się działalnością naukową oraz wdrażaniem innowacyjnych rozwiązań w zakresie komputerowego wspomagania badań farmaceutycznych i biotechnologicznych:

- jako kierownik projektów badawczo-wdrożeniowych (projekty pozytywnie zweryfikowane i dofinansowane przez Program Operacyjny Innowacyjna Gospodarka Unii Europejskiej oraz przez Naczelną Organizację Techniczną – [omówione w załączniku nr 2 do wniosku](#)) oraz projektów badawczych w dziedzinie komputerowego wspomagania projektowania leków.
- jako uczestnik (wykładowca lub prezentujący poster) międzynarodowych konferencji naukowych, warsztatów i szkoleń – [omówione w załączniku nr 2 do wniosku](#).

4. Wskazanie osiągnięcia¹ wynikającego z art. 16 ust. 2 ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. nr 65, poz. 595 ze zm.):
- a) tytuł osiągnięcia naukowego,

„Rozwój wieloskalowych metod molekularnego modelowania białek oraz ich zastosowanie w badaniach mechanizmów funkcjonowania białek”

b) publikacje wchodzące w skład osiągnięcia naukowego

* – autor korespondencyjny

H1. Michał Jamróz, Modesto Orozco, Andrzej Koliński i **Sebastian Kmiecik***. *Consistent View of Protein Fluctuations from All-Atom Molecular Dynamics and Coarse-Grained Dynamics with Knowledge-Based Force-Field*. **Journal of Chemical Theory and Computation**, 9: 119-125, 2013.

H2. Michał Jamróz, Andrzej Koliński i **Sebastian Kmiecik***. *CABS-flex: server for fast simulation of protein structure fluctuations*. **Nucleic Acids Research**, 41: W427-W431, 2013.

H3. Michał Jamróz, Andrzej Koliński i **Sebastian Kmiecik***. *CABS-flex predictions of protein flexibility compared with NMR ensembles*. **Bioinformatics**, 30(15):2150-4, 2014.

H4. Maciej Błaszczuk, Michał Jamróz, **Sebastian Kmiecik*** i Andrzej Koliński*. *CABS-fold: server for the de novo and consensus-based prediction of protein structure*. **Nucleic Acids Research**, 41:W406-W411, 2013.

H5. **Sebastian Kmiecik**, Michał Jamróz i Michał Koliński*. *Structure prediction of the second extracellular loop in G-protein-coupled receptors*. **Biophysical Journal**, 106(11):2408-16, 2014.

H6. **Sebastian Kmiecik*** i Andrzej Koliński*. *Simulation of chaperonin effect on protein folding: a shift from nucleation-condensation to framework mechanism*. **Journal of the American Chemical Society**, 133: 10283–9, 2011.

H7. **Sebastian Kmiecik***, Dominik Gront, Maksim Kouza i Andrzej Koliński*. *From coarse-grained to atomic-level characterization of protein dynamics: transition state for the folding of B domain of protein A*. **The Journal of Physical Chemistry B**, 116: 7026–32, 2012.

¹ w przypadku, gdy osiągnięciem tym jest praca/ prace wspólne, należy przedstawić oświadczenia wszystkich jej współautorów, określające indywidualny wkład każdego z nich w jej powstanie

H8. Mateusz Kurciński, Andrzej Koliński i **Sebastian Kmiecik***. *Mechanism of Folding and Binding of an Intrinsically Disordered Protein as Revealed by Ab Initio Simulations*. **Journal of Chemical Theory and Computation**, 10 (6): 2224–31, 2014.

H9. Jacek Wabik, **Sebastian Kmiecik**, Dominik Gront, Maksim Kouza* i Andrzej Koliński*. *Combining Coarse-Grained Protein Models with Replica-Exchange All-Atom Molecular Dynamics*. **International Journal of Molecular Sciences**, 14: 9893–9905, 2013.

H10. Michał Jamroz, Andrzej Koliński i **Sebastian Kmiecik***. *Protocols for efficient simulations of long time protein dynamics using coarse-grained CABS model*. **Methods in Molecular Biology** (Clifton, N.J.), 1137:235-250, 2014.

H11. **Sebastian Kmiecik**, Michał Jamroz i Andrzej Koliński*. *Multiscale approach to protein folding dynamics*. Rozdział w książce: **Multiscale Approaches to Protein Modeling**, wyd. Springer Berlin Heidelberg, ed. Andrzej Koliński, 281-294, 2011.

H12. **Sebastian Kmiecik***, Jacek Wabik, Michał Koliński, Maksim Kouza i Andrzej Koliński. *Coarse-Grained Modeling of Protein Dynamics*. Rozdział w książce: **Computational Methods to Study the Structure and Dynamics of Biomolecules and Biomolecular Processes** – Springer Series in Bio-/Neuroinformatics, wyd. Springer Berlin Heidelberg, ed. Adam Liwo, 1:55-79, 2014.

c) omówienie celu naukowego/artystycznego ww. pracy/prac i osiągniętych wyników wraz z omówieniem ich ewentualnego wykorzystania.

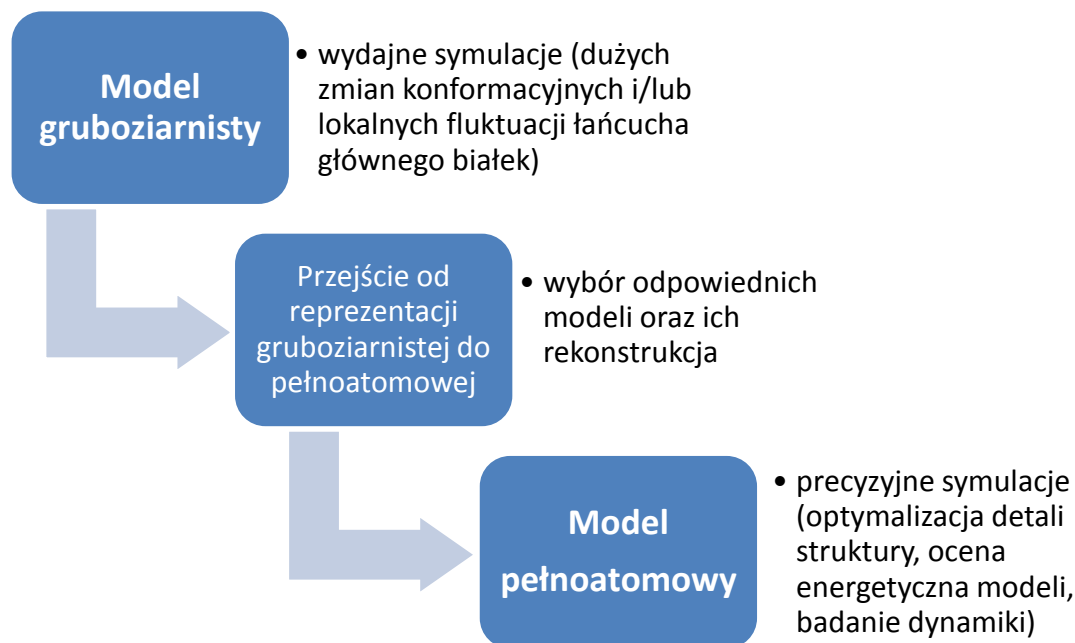
Ogólne przedstawienie tematyki prac oraz głównego celu naukowego

Białka należą do najważniejszych molekuł odpowiedzialnych za funkcjonowanie organizmów żywych. Zrozumienie molekularnych podstaw funkcjonowania białek wymaga poznania ich struktury i dynamiki. W organizmach żywych, stopień ruchliwości struktury białek (odpowiadający za funkcje) zależy od białka i obejmuje kontinuum pomiędzy dwoma ekstremami: od sztywnych lub prawie sztywnych struktur (np. struktury wielu enzymów) do bardzo ruchliwych białek o nieustalanej strukturze (Chouard 2011). Do tej pory najlepiej poznano molekularne podstawy funkcji białek wykazujących dużą sztywność struktury. Najmniej poznane pozostają molekularne mechanizmy w których białka wykazują dużą zmienność strukturalną, a poziom trudności badań, zarówno eksperymentalnych jak i teoretycznych, wzrasta wraz ze stopniem ruchliwości i wielkością badanych białek.

Głównym celem prac wchodzących w skład osiągnięcia naukowego był rozwój nowych i wydajnych metod wieloskalowego modelowania białek oraz ich wykorzystanie w poznaniu mechanizmów funkcjonowania białek. Prace te dotyczyły przede wszystkim badań i rozwoju nowych metodologii modelowania gruboziarnistego w zastosowaniu do wydajnych symulacji

dynamiki białek, ale również zastosowania i optymalizacji różnych strategii modelowania pełnoatomowego w połączeniu z modelem gruboziarnistym.

Zarówno struktura jak i dynamika białek mogą być scharakteryzowane w sposób hierarchiczny (Henzler-Wildman, Lei et al. 2007), dlatego komputerowe modelowanie struktury i dynamiki białek oraz ich kompleksów, może być realizowane wieloskalowo, to znaczy na różnych poziomach skali rozdzielczości jak i skali czasowej (patrz prace przeglądowe **H11** i **H12**). W ostatnich dziesięcioleciach nastąpił intensywny rozwój technik modelowania, zarówno w rozdzielczości pełnoatomowej jak i gruboziarnistej. Technika pełnoatomowej dynamiki molekularnej od wielu lat jest uważana za wzorzec w dziedzinie komputerowego modelowania dynamiki białek (Karplus and Kuriyan 2005). Jednak, pomimo stałego wzrostu mocy obliczeniowej komputerów, technika ta pozostaje wciąż zbyt mało wydajna. Symulacje dynamiki molekularnej są obecnie ograniczone do krótkich skal czasowych (rzędu mikrosekund przy użyciu konwencjonalnych jednostek obliczeniowych, lub milisekund przy użyciu dedykowanego superkomputera dla małych białek (Shaw, Maragakis et al. 2010; Shaw 2013)). Większość interesujących zmian konformacyjnych białek (odpowiedzialnych za ważne funkcje w organizmach żywych) pozostaje niedostępna dla pełnoatomowej dynamiki molekularnej ze względu na zbyt dużą skalę czasową i/lub zbyt duży rozmiar systemu (Vendruscolo and Dobson 2011). Wydajne symulacje takich procesów, o dużych skalach czasowych i dla dużych systemów biomolekularnych, mogą być wykonane przy użyciu modeli gruboziarnistych (patrz prace przeglądowe **H11**, **H12** oraz (Liwo, He et al. 2011)). W ostatnich latach, liczba badań naukowych opartych na wynikach modelowania gruboziarnistego wzrosła wielokrotnie (Takada 2012). Wzrost ten został spowodowany między innymi intensyfikacją badań doświadczalnych, które zwykle dostarczają jedynie szczątkowych informacji o strukturze i dynamice badanych układów. W związku z tym, obecnie główną rolą metod modelowania jest wspomaganie interpretacji badań eksperymentalnych poprzez dostarczanie pełnoatomowych struktur opisujących badane procesy. Przejście, od wydajnej symulacji w reprezentacji gruboziarnistej do wyników modelowania w postaci pełnoatomowych modeli jest możliwe dzięki podejściu modelowania wieloskalowego (patrz Rysunek 1). Ważna rola wieloskalowego modelowania we współczesnych badaniach biochemicznych została wyróżniona w ubiegłym roku przez Komitet Noblowski. Zeszlóroczna Nagroda Nobla z Chemii (2013) została przyznana za "rozwój wieloskalowych modeli złożonych systemów chemicznych". W uzasadnieniu tej nagrody, Komitet Noblowski uznał pionierskie prace Michaela Levitta i Arieha Warshela w dziedzinie gruboziarnistego modelowania białek (Levitt and Warshel 1975; Warshel and Levitt 1976) za ważny krok w kierunku lepszego zrozumienia funkcjonowania dużych układów biomolekularnych.



Rysunek 1. Ogólny schemat wieloskalowego modelowania białek wykorzystany w pracach wchodzących w skład osiągnięcia naukowego.

Szczegółowe cele naukowe oraz przedstawienie wyników prac

Szczegółowe cele naukowe obejmowały:

I. Opracowanie metody modelowania fluktuacji struktury białek globularnych:

- (i) opracowanie gruboziarnistej metody przewidywania fluktuacji białek globularnych: porównanie z wynikami symulacji metodą pełnoatomowej dynamiki molekularnej (**praca H1**);
- (ii) opracowanie wieloskalowej metody do przewidywania fluktuacji białek globularnych: serwer CABS-flex (**praca H2**);
- (iii) porównanie wyników modelowania metodą CABS-flex z fluktuacjami obserwowanymi w zespołach struktur rozwiązanych metodą NMR (**praca H3**).

II. Opracowanie metod przewidywania struktury i dynamiki białek:

- (i) opracowanie wieloskalowej metody do przewidywania struktury i dynamiki białek globularnych: serwer CABS-fold (**praca H4**);
- (ii) opracowanie wieloskalowej metody symulacji pętli zewnątrzkomórkowych receptorów sprzężonych z białkami G (**praca H5**).

III. Opracowanie metod modelowania oraz ich zastosowanie do poznania mechanizmów funkcjonowania białek:

- (i) badanie mechanizmu funkcjonowania białka opiekuńczego (**praca H6**);
- (ii) badanie mechanizmu zwijania białka A – analiza struktury stanu przejściowego (**praca H6 i H7**);
- (iii) badanie mechanizmu zwijania i wiązania białka nieustrukturyzowanego (**praca H8**).

IV. Opracowanie protokołów wieloskalowego modelowania dynamiki białek:

- (i) opracowanie protokołu modelowania wieloskalowego z wykorzystaniem szybkiej optymalizacji struktur pełnoatomowych (**praca H7**);
- (ii) opracowanie protokołu modelowania wieloskalowego z wykorzystaniem symulacji metodą dynamiki molekularnej z wymianą replik (**praca H9**);
- (iii) opracowanie schematów symulacji dynamiki zwijania białek (**praca H10**).

V. Podsumowanie stanu wiedzy, oraz prac własnych, w zakresie wieloskalowego modelowania dynamiki białek:

- (i) opracowanie artykułu przeglądowego pt. „Wieloskalowe podejście do modelowania dynamiki białek” (**praca H11**);
- (ii) opracowanie artykułu przeglądowego pt. „Gruboziarniste modelowanie dynamiki białek” (**praca H12**).

W ciągu ostatnich kilkunastu lat, technika pełnoatomowej dynamiki molekularnej (MD) stała się powszechnym narzędziem używanym do symulacji fluktuacji struktury białek (dynamiki blisko stanu natywnego). Takie symulacje są najczęściej wykonywane poprzez: (i) ich ukierunkowanie danymi doświadczalnymi o dynamice białek, (ii) start z pojedynczej natywnej struktury bez ukierunkowania zewnętrznymi danymi (Vendruscolo 2007; Fisette, Lague et al. 2012). W 2007 roku, powszechny pogląd o poprawności nieukierunkowanych symulacji MD został wsparty poprzez badania grupy Modesta Orozco (Rueda, Ferrer-Costa et al. 2007). Orozco wraz ze współpracownikami wykazał, że niezależnie od tego jakie pole siłowe zostanie użyte do pełnoatomowych symulacji (startujących z natywnych struktur, o długości 10 nanosekund, w jawnym rozpuszczalniku), otrzymany obraz fluktuacji jest spójny (Rueda, Ferrer-Costa et al. 2007). W zastosowaniu do symulacji białek globularnych o różnych zwojach, Orozco wykorzystał cztery najbardziej popularne pola siłowe: AMBER, CHARMM, GROMOS i OPLS. Symulacje te wykonano przy użyciu wydajnej infrastruktury superkomputerowej:

- w przeliczeniu na jeden standardowy procesor wykonane obliczenia zajęły około 50 lat;
- wygenerowano ponad 1.5 terabajta danych trajektorii białek.

W naszych badaniach, używając danych trajektorii MD z badań Orozco, pokazaliśmy, że charakterystyka struktury i dynamiki białek w trajektoriach MD jest zgodna z wynikami gruboziarnistych symulacji przeprowadzonych modelem CABS (**praca H1**). Co ważne, koszt obliczeniowy dynamiki gruboziarnistej okazał się znacznie (około 6 tysięcy razy) niższy niż

symulacji MD. W toku badań, przetestowaliśmy szereg ustawień modelu CABS tak aby otrzymać zestaw parametrów (długość symulacji, temperatura, parametry skalujące pole siłowe, różne wersje więzów stabilizujące strukturę białek) pozwalający uzyskać jak największe podobieństwo do trajektorii MD z badań Orozco. Opracowany schemat symulacji blisko-natywnych fluktuacji modelem CABS, został następnie wykorzystany jako podstawowy element automatycznej i wieloskalowej metody do przewidywania fluktuacji białek globularnych: serwera CABS-flex (**praca H2**) (serwer jest dostępny pod adresem <http://biocomp.chem.uw.edu.pl/CABSflex>). W serwerze CABS-flex zintegrowano następujące etapy modelowania: (i) symulacje blisko-natywnych fluktuacji modelem CABS, (ii) analizę trajektorii oraz wybór modeli reprezentatywnych (metodą analizy skupień), (iii) rekonstrukcję wybranych modeli do rozdzielczości pełnoatomowej oraz ich optymalizację przy użyciu metody BBQ (Gront, Kmiecik et al. 2007) oraz, w dalszym kroku, metody ModRefiner (Xu and Zhang 2011). Jediną wejściową informacją wymaganą przez serwer CABS-flex jest plik zawierający współrzędne pojedynczej struktury białka (w formacie PDB). Jednym z głównych wyników modelowania serwera CABS-flex jest zestaw struktur (w rozdzielczości pełnoatomowej), obrazujący fluktuacje wejściowej konformacji białka. Taki zestaw struktur może mieć zastosowanie w badaniach nad funkcją białka (np. w dokowaniu molekularnym lub w analizie wpływu fluktuacji struktury na wyeksponowanie poszczególnych reszt aminokwasowych do rozpuszczalnika²). Nasze dalsze badania skupiły się na porównaniu przewidywań fluktuacji białek, metodą CABS-flex oraz MD, ze zmiennością struktur w zespołach struktur NMR (ang. *NMR ensembles*) (**praca H3**). Takie porównanie wykonano dla zestawu 140 białek (w badaniach wykorzystano białka, których zespoły struktur zostały otrzymane techniką NMR, oraz dla których trajektorie MD zdeponowano w bazie MoDEL (Meyer, D'Abramo et al. 2010)). Porównanie wyników modelowania metodą CABS-flex z fluktuacjami obserwowanymi w zespołach struktur NMR wykazało ich wysoką korelację (korelacja pomiędzy tzw. profilami fluktuacji wyrażającymi średnią ruchliwość poszczególnych reszt aminokwasowych). Dla całego zestawu białek, średnia korelacja pomiędzy CABS-flex a strukturami NMR okazała się być nieco wyższa, niż ta pomiędzy symulacjami MD a strukturami NMR. Wyniki tego porównania potwierdziły, że opracowana przez nas metoda może być obliczeniowo tanią alternatywą dla symulacji MD w zastosowaniu do przewidywania rejonów struktury białek, które ulegają zmianom konformacyjnym, oraz do przewidywania stopnia takich zmian (**praca H3**).

Jak pokazano w badaniach mechanizmów zwijania białek (Kmiecik and Kolinski 2007; Kmiecik and Kolinski 2008), oraz w badaniach ukierunkowanych na przewidywanie struktury białek (Kolinski and Bujnicki 2005; Jamroz and Kolinski 2010), model CABS umożliwia również przewidywanie struktury białek, bądź fragmentów struktury o znacznej długości, wyłącznie na podstawie sekwencji, tj. *de novo*. Przewidywanie *de novo* jest najtrudniejszym zadaniem przewidywania struktury białek, możliwym do zrealizowania przez nieliczne

² Taka analiza może mieć zastosowanie w metodach wykrywających miejsca oddziaływania białka (np. przewidywających skłonność do agregacji). Potencjalne zastosowania metody CABS-flex zostały szerzej opisane w dalszej części autoreferatu.

techniki modelowania teoretycznego. Aby umożliwić społeczności naukowej łatwy dostęp do unikalnej metodologii modelowania *de novo* przy użyciu modelu CABS, opracowaliśmy automatyczną i wieloskalową metodę do przewidywania struktury i dynamiki białek globularnych: serwer CABS-fold (**praca H4**) (serwer jest dostępny pod adresem <http://biocomp.chem.uw.edu.pl/CABSfold>). Oprócz schematów modelowania *de novo*, serwer CABS-fold został wyposażony w procedury umożliwiające przeprowadzenie modelowania porównawczego (przy użyciu szablonów). W serwerze CABS-fold zintegrowano ze sobą następujące procedury modelowania: (i) procedury przewidywania struktury wyłącznie na podstawie sekwencji, oraz przy użyciu modeli – szablonów, w oparciu o model CABS; (ii) analizę trajektorii oraz wybór modeli reprezentatywnych (metodą analizy skupień); (iii) rekonstrukcję wybranych modeli do rozdzielczości pełnoatomowej oraz ich optymalizację (przy użyciu metody BBQ (Gront, Kmiecik et al. 2007) oraz, w dalszym kroku, metody ModRefiner (Xu and Zhang 2011)). Serwer CABS-fold dostarcza w wynikach modelowania od kilku – do kilkunastu wynikowych modeli (reprezentujących najczęściej powtarzające się konformacje w trajektorii modelu CABS). Tak więc, analiza wszystkich wynikowych modeli może dostarczyć informacji nie tylko o strukturze blisko-natywnej ale również o pośrednich etapach zwiłania.³

Jak wspomniano wyżej, jedną z unikalnych cech modelu CABS, jest zdolność do przewidywania konformacji dużych (w porównaniu do innych metod) fragmentów struktury wyłącznie na podstawie sekwencji. Postanowiliśmy to wykorzystać w opracowaniu metody przewidywania struktury drugiej zewnątrzkomórkowej pętli (*ang. the second extracellular loop, ECL2*) w receptorach sprzężonych z białkami G (GPCR) (**praca H5**). Receptory GPCR są białkami błonowymi, które pełnią kluczowe role w funkcjonowaniu organizmów żywych. Ich funkcjonalnie ważnym fragmentem jest pętla ECL2. Otrzymanie dokładnej struktury pętli ECL2 jest niewralgicznym elementem budowania modeli receptorów GPCR, przeznaczonych do praktycznego zastosowania w projektowaniu leków (Peeters, van Westen et al. 2011; Wheatley, Wootten et al. 2012). Używając różnych procedur wieloskalowego modelowania opartych na modelu CABS, przeprowadziliśmy symulacje modelowania struktury pętli ECL2 (o długości od 13 do 34 reszt aminokwasowych) dla 13 różnych receptorów GPCR. Symulacje przeprowadzono uwzględniając pełną giętkość struktury dla wszystkich pętli zewnątrzkomórkowych obecnych w modelowanych receptorach, podczas gdy część trans-membranowa była „zamrożona” w konformacji znanej z eksperymentu. Dodatkowo, nasza procedura modelowania wykorzystywała przewidywania struktury drugorzędowej (otrzymane z niezależnych teoretycznych metod przeznaczonych do tego celu) oraz informacje o mostkach dwusiarczkowych. W wyniku modelowania otrzymaliśmy zestawy: (i) konformacji charakteryzujących się niską energią potencjalną (wg pełnoatomowego pola siłowego) lub (ii) konformacji najczęściej powtarzających się w symulacji. Zestawy te zawierały modele pętli ECL2 podobne do modeli otrzymanych metodami doświadczalnymi, a poziom podobieństwa okazał się być porównywalny do wyników otrzymanych przez inne uznane metody

³ Potencjalne zastosowania metody CABS-fold zostały szerzej opisane w dalszej części autoreferatu.

modelowania. Dzięki licznym uproszczeniom w procedurze modelowania⁴ koszt obliczeniowy naszej metody okazał się bardzo niski.

Ponieważ wiele badań eksperymentalnych wskazuje na dużą giętkość pętli ECL2 i na ważną rolę dynamiki pętli ECL2 w funkcjonowaniu receptorów GPCR (Fraser 1989; Nanevich, Wang et al. 1996; Zhao, Hwa et al. 1996; Ott, Troskie et al. 2002; Shi and Javitch 2002; Klco, Wiegand et al. 2005; Conner, Hawtin et al. 2007; Dror, Arlow et al. 2011; Wheatley, Wootten et al. 2012; Seibt, Schiedel et al. 2013) celem badań teoretycznych powinno być przewidywanie nie tylko pojedynczych modeli ale zestawów modeli obrazujących dynamikę pętli. Rezultatem naszych przewidywań są zestawy 10 lub 100 modeli o najniższej energii potencjalnej, lub modeli reprezentujących najczęściej powtarzające się konformacje. Analiza otrzymanych zestawów modeli wskazała, przynajmniej dla niektórych receptorów, że pętle ECL2 mogą podlegać znacznym przemieszczeniom względem części transbłonowej. Otrzymane zbiory modeli zostały zweryfikowane poprzez porównanie z pojedynczymi strukturami otrzymanymi za pomocą rentgenografii strukturalnej (niestety dokładne dane na temat dynamiki pętli ECL2 nie są dostępne). Aby umożliwić późniejszą eksperymentalną weryfikację, oraz porównanie z innymi metodami przewidywania, otrzymane przez nas zestawy modeli zostały udostępnione na stronie: <http://biocomp.chem.uw.edu.pl/GPCR-loop-modeling>.

Oprócz symulacji dynamiki pojedynczych białek, metody wieloskalowego modelowania molekularnego mogą służyć również do symulacji bardziej złożonych mechanizmów, takich jak funkcjonowanie białka opiekuńczego (**praca H6**). Tzw. białka opiekuńcze (*ang. chaperones*) są odpowiedzialne za wspomaganie procesu zwijania białek (Hartl and Hayer-Hartl 2009). Szczególnym przypadkiem białek opiekuńczych są białka o kształcie pustego cylindra (*ang. chaperonin*), które wspomagają zwijanie białek poprzez ich „kapsułkowanie” wewnątrz swojej cylindrycznej struktury. Dokładny mechanizm ich działania wciąż pozostaje nieznanym – obserwacja dynamiki białek w obecności lub wewnątrz cylindra białka opiekuńczego za pomocą technik eksperymentalnych jest trudna i nie umożliwia dokładnego wglądu w mechanizm zachodzącego procesu. Do tej pory zaproponowano wiele mechanizmów funkcjonowania tych białek opiekuńczych, które można podzielić na: pasywne (wspomaganie zwijania poprzez zapobieganie agregacji, bez wpływu na mechanizm zwijania białek) i aktywne (wspomaganie zwijania poprzez wpływ na mechanizm zwijania białek) (Hartl and Hayer-Hartl 2009; Jewett and Shea ; Lucent, England et al. 2009). Jedną z historycznie pierwszych i najbardziej popularnych teorii jest mechanizm iteracyjnego wyżarzania (*ang. iterative annealing mechanism, IAM*), który skupia się na badaniu efektu cyklicznych oddziaływań białek z hydrofobowymi ścianami białka opiekuńczego. W **pracy H6** przedstawiliśmy symulacje mechanizmu iteracyjnego używając wydajnej procedury modelowania gruboziarnistego z wykorzystaniem modelu CABS. Symulacje przeprowadziliśmy:

⁴ Niski koszt obliczeniowy modelowanie gruboziarniste, bez jawnej obecności lipidów trans-membranowych, w połączeniu z szybką optymalizacją pełnoatomową, bez uwzględnienia rozpuszczalnika

- dla dwóch małych białek, które są jednymi z najlepiej przebadanych pod kątem mechanizmów zwijania: domeny B białka A (BdpA) i domeny B1 białka G (GB1);
- w dwóch wersjach: (1) uwzględniając wpływ białka opiekuńczego oraz (2) bez uwzględnienia wpływu białka opiekuńczego – w celu porównania wyników symulacji obu wersji.

Do tej pory wyniki badań mechanizmu iteracyjnego wyzarzania sugerowały, że dzięki wiązaniu białek do hydrofobowych ścian białka opiekuńczego możliwe jest rozwijanie nieprawidłowo zwiniętych konformacji (Lucent, England et al. 2009). Wyniki naszych badań pokazały, że cykliczne rozrywanie białka (poprzez oddziaływania ze ścianami białka opiekuńczego) może mieć także wpływ na mechanizm zwijania białka poprzez jego modulację na bardziej efektywną eksplorację przestrzeni konformacyjnej i na unikanie nieprawidłowo zwiniętych konformacji. Zaobserwowana modulacja mechanizmu zwijania, polegająca na rozluźnieniu sieci oddziaływań trzeciorzędowej struktury i na stabilizacji elementów struktury drugorzędowej może być opisana jako przesunięcie od mechanizmu nukleacji-kondensacji (*ang. nucleation-condensation*) do mechanizmu szkieletowego (*ang. framework*). Mechanizmy szkieletowy i nukleacji-kondensacji to klasyczne modele opisujące proces zwijania białka. Są one uważane za ekstrema jednego, ogólnego mechanizmu zwijania białek globularnych (Gianni, Gwydosz et al. 2003; White, Gianni et al. 2005). Mechanizm szkieletowy jest faworyzowany przez wysoce stabilną strukturę drugorzędową (której tworzenie jest pierwszym etapem zwijania), mechanizm nukleacji-kondensacji zakłada jednoczesne formowanie struktury drugo- i trzeciorzędowej. Co ciekawe, przejście jednego mechanizmu w drugi zostało zaobserwowane na przykład w wyniku zmiany właściwości chemicznych środowiska lub w wyniku mutacji pojedynczej reszty aminokwasowej (Gianni, Gwydosz et al. 2003; White, Gianni et al. 2005).

W **pracy H6**, oprócz analizy mechanizmu działania białka opiekuńczego przedstawiliśmy także analizę mechanizmu zwijania białka BdpA. Przedstawiona charakterystyka stanu przejścia (stanu odpowiadającego maksimum energii swobodnej na ścieżce zwijania białka) – w formie map kontaktów i wykresów mobilności reszt aminokwasowych białka – okazała się być zgodna z dostępnymi danymi eksperymentalnymi (z analizą wartości ϕ , czyli punktową mutacją białek w połączeniu z pomiarami kinetycznymi). Charakterystyka stanu przejścia BdpA została następnie rozszerzona w **pracy H7** o prezentację zestawu struktur pełnoatomowych.

W toku dalszych prac badawczych, opracowana została wydajna strategia modelowania gruboziarnistego do badania jednoczesnego zwijania i wiązania białka nieustrukturyzowanego (**praca H8**). Mechanizm takich procesów wciąż nie jest w pełni zrozumiany. W badaniach użyliśmy kompleksu ufosforylowanej formy białka indukowanego przez kinazę (pKID) z domeną białkową KIX, który jest modelowym przykładem kompleksu białkowego do badań nad molekularnymi mechanizmami funkcjonowania białek nieustrukturyzowanych. Co

ważne, symulacje przeprowadziliśmy bez użycia informacji o eksperymentalnej strukturze pKID związanej z domeną KIX:

- symulacje rozpoczęto z niezwiązanej, losowo ułożonej i nieustrukturyzowanej konformacji pKID;
- podczas symulacji konformacja łańcucha pKID, oraz jego pozycja, nie były ukierunkowane ani ograniczone żadnymi więzami (z kolei dynamika domeny KIX była ograniczona do blisko natywnych fluktuacji łańcucha głównego).

Symulacje tak dużych zmian konformacyjnych, bez używania wiedzy o strukturze związanego białka, są niedostępne dla klasycznych metod modelowania (i według naszej wiedzy nigdy nie były przeprowadzone dla badanego systemu pKID/KIX). W wyniku przeprowadzonych symulacji dokowania molekularnego otrzymaliśmy struktury białka pKID zadokowane w natywnym miejscu i w konformacji blisko-natywnej (patrz film obrazujący przykładową trajektorię: <http://youtu.be/WhS8UFaEodk>). Analiza trajektorii dokowania molekularnego pozwoliła na otrzymanie charakterystyki szerokiego zestawu przejściowych kompleksów, na drodze do blisko-natywnego kompleksu, która dobrze odpowiadała danym eksperymentalnym (Sugase, Dyson et al. 2007). Oprócz tego, w **pracy H8** wykonaliśmy analizę mechanizmu jednoczesnego zwijania i wiązania się białka pKID, która pokazała, że kluczowym etapem tego procesu jest utworzenie słabych natywnych oddziaływań pomiędzy odpowiednim fragmentem uformowanej helisy pKID a powierzchnią białka KIX. Te oddziaływania tworzą swoisty zarodek zwijania i wiązania, po utworzeniu którego pozostałe fragmenty struktury pKID mogą ulec kondensacji od wysoce nieorganizowanej do zorganizowanej i bliskonatywnej struktury. Zaobserwowany mechanizm przypomina mechanizm nukleacji-kondensacji powszechnie obserwowany dla zwijania się białek globularnych (Fersht 1995; Itzhaki, Otzen et al. 1995; Daggett and Fersht 2003). Co ciekawe, podobny scenariusz zwijania i wiązania się białek nieustrukturyzowanych zaobserwowano ostatnio dla innych białek na podstawie badań doświadczalnych: inżynierii białek i pomiarów kinetycznych dla systemu ACTR/NCBD (Dogan, Mu et al. 2013) oraz spektroskopią zatrzymanego przepływu dla domen PDZ (Haq, Chi et al. 2012).

Oprócz opracowania metod wieloskalowych symulacji dostosowanych do badania poszczególnych procesów molekularnych (opisanych powyżej), zaproponowaliśmy również uniwersalne procedury umożliwiające połączenie modelowania gruboziarnistego z podejściem pełnoatomowym. W ramach optymalizacji i testowania różnych schematów modelowania wieloskalowego opracowaliśmy:

- procedurę modelowania wieloskalowego z wykorzystaniem szybkiej optymalizacji struktur pełnoatomowych (**praca H7**);
- procedurę modelowania wieloskalowego z wykorzystaniem symulacji metodą dynamiki molekularnej z wymianą replik (**praca H9**);

- oraz procedury symulacji dynamiki zwijania białek przy użyciu modelu CABS – od stanu zdenaturowanego do stanu blisko-natywnego (**praca H10**).

Metoda przedstawiona w **pracy H7** pozwala na wydajną i szybką rekonstrukcję trajektorii dynamiki w reprezentacji gruboziarnistej (wyłącznie na podstawie pozycji węgla alfa; np. otrzymaną modelem CABS) do reprezentacji pełnoatomowej. Metoda przedstawiona w **pracy H9** umożliwia wykonanie wydajnych symulacji pełnoatomowej dynamiki molekularnej z wymianą replik w sposób bardziej wydajny niż rozpoczynając symulacje ze zdenaturowanych struktur. Zwiększenie wydajności (szybsze zrównoważenie symulacji) okazało się możliwe dzięki rozpoczęciu symulacji z odpowiednio dobranych struktur początkowych otrzymanych w symulacjach modelem gruboziarnistym. Z kolei metody symulacji opisane w **pracy H10** bazują na opracowanych i przetestowanych wcześniej przeze mnie protokołach badania mechanizmów zwijania białek: barnazy oraz inhibitora chymotrypsyny (Kmieciak and Kolinski 2007) oraz domeny B1 białka G (Kmieciak and Kolinski 2008). Przedstawiona metodologia pozwala na łatwe zastosowanie protokołów symulacji i analizy danych do badania mechanizmów zwijania nowych systemów białkowych.

Wyniki przedstawione w pracach wchodzących w skład osiągnięcia naukowego ilustrują unikalne możliwości opracowanych metod w wykonaniu symulacji różnych systemów makromolekularnych. Podstawową zaletą opracowanych sposobów symulacji, w porównaniu do innych istniejących technik modelowania, jest ich większa wydajność pozwalająca na: (i) znacznie szybsze przeprowadzenie symulacji (umożliwiając tym samym przeprowadzenie większej liczby symulacji i dla większej liczby obiektów), (ii) przeprowadzenie symulacji procesów, których skale czasowe są nieosiągalne dla klasycznych metod modelowania.

Zaproponowane metody modelowania molekularnego stanowią podstawę do dalszego rozwoju podejścia wieloskalowego łączącego modelowanie gruboziarniste i pełnoatomowe. Obecnie przewiduje się, że taki właśnie kierunek badań pozwoli w przyszłości na modelowanie jeszcze bardziej złożonych maszyn makromolekularnych (np. kompleksów białkowych o dużym rozmiarze), które są w centrum zainteresowania nauk biomedycznych (patrz także prace przeglądowe **H11**, **H12** oraz (Scheraga, Khalili et al. 2007; Russel, Lasker et al. 2009; Whitford, Noel et al. 2009)).

Omówienie wykorzystania wyników

1. Wykorzystanie metody CABS-flex

Metoda CABS-flex (podstawy metody oraz jej testy opisano w **pracach H1, H2, H3**) została udostępniona w formie darmowego serwera umożliwiającego łatwe wykonanie symulacji dla białka podanego przez użytkownika. Przez około półtora roku (od maja 2013 do września 2014), zanotowaliśmy około 31,4 tysięcy wyświetleń stron internetowych serwera CABS-flex, w tym około 5,9 tysięcy unikalnych użytkowników. Do września 2014, metoda

CABS-flex została użyta w opublikowanych badaniach dotyczących zależności struktura-dynamika-funkcja białek odpowiedzialnych za ważne procesy życiowe:

- przez Newo (Newo 2014) w analizie dynamiki struktury homologów helikaz (wykonana analiza może mieć zastosowanie w projektowaniu nowych leków przeciw pasożytniczych, których celami będą badane helikazy).
- przez Liu i innych (Liu, Werner et al. 2014) w analizie dynamiki struktury serpin (białek należących do rodziny inhibitorów proteazy serynowej) jako narzędzie wspomagające interpretację danych eksperymentalnych.
- przez Fraga i innych (Fraga, Grana-Montes et al. 2014) w analizie dynamiki struktury małych domen bogatych w mostki di-siarczkowe jako narzędzie uzupełniające wyniki innych teoretycznych przewidywań dynamiki badanych białek.

Metoda CABS-flex została również wykorzystana jako element nowej strategii do przewidywania miejsc na powierzchni białka odpowiadających za agregację, opracowanej w ramach pracy magisterskiej na Uniwersytecie Autonomicznym w Barcelonie pod kierunkiem prof. Salvadora Ventury (Pujagut 2013). W ramach tej pracy opracowano nowe narzędzie - Aggrescan3D – służące do przewidywania miejsc odpowiedzialnych za agregację. Metoda Aggrescan3D wykorzystuje następujące dane:

- informację o strukturze białka (o stopniu wyeksponowania reszt aminokwasowych do rozpuszczalnika oraz ich wzajemnym rozmieszczeniu) w połączeniu z informacją o giętkości strukturalnej białka (wykorzystywane są modele obrazujące fluktuacje białka przewidziane metodą CABS-flex);
- zdolność poszczególnych aminokwasów do agregacji – przewidziana metodą Aggrescan (metoda opracowana przez grupę prof. Ventury, jest popularnym narzędziem do przewidywania miejsc odpowiedzialnych za agregację wyłącznie w oparciu o sekwencję białka (Conchillo-Sole, de Groot et al. 2007; de Groot, Castillo et al. 2012)).

Testy metody Aggrescan3D w dwóch wersjach: (1) z wykorzystaniem modeli obrazujących giętkość białka przewidzianych metodą CABS-flex, (2) bez wykorzystania wiedzy o giętkości białek, pokazały, że wykorzystanie informacji o giętkości strukturalnej otrzymanej metodą CABS-flex jest wymagane do otrzymania poprawnych predykcji (Pujagut 2013). Badania te są obecnie kontynuowane w ramach współpracy naukowej pomiędzy grupą prof. Ventury a mną i moimi współpracownikami. Celem współpracy jest optymalizacja i dalsza weryfikacja metody Aggrescan3D oraz jej udostępnienie społeczności naukowej.

Oprócz wyżej wymienionych, do najbardziej obiecujących zastosowań metody CABS-flex należy jej wykorzystanie jako części składowej różnych procedur modelowania molekularnego:

- metod przewidujących zdolność do agregacji białek – na przykład według schematu technologii SAP (*spatial aggregation propensity*) opracowanej przez Chennamsetty i

in. (Chennamsetty, Voynov et al. 2009). Chennamsetty i in. wykorzystali pełnoatomowe symulacje dynamiki molekularnej do symulacji zmienności wyeksponowania do rozpuszczalnika reszt aminokwasowych odpowiedzialnych za agregację. Zastąpienie symulacji dynamiki molekularnej metodą CABS-flex, może potencjalnie: (i) znacznie przyspieszyć i poprawić dokładność metody SAP; (ii) umożliwić zastosowanie SAP dla przeciwciał o dużym rozmiarze (obecnie ze względu na zbyt duży koszt obliczeniowy dynamiki molekularnej nie jest to możliwe);

- metod przewidujących miejsca na powierzchni białka, które odpowiadają za oddziaływania: (i) białek z innymi białkami, (ii) białek z kwasami nukleinowymi, (iii) białek z ligandami małowcząsteczkowymi. Obecnie większość metod przewidujących miejsca oddziaływań białek nie wykorzystuje informacji o zmienności wyeksponowania poszczególnych reszt aminokwasowych do rozpuszczalnika. Istnieje wiele różnego rodzaju metod przewidujących miejsca oddziaływań typu (i), (ii) lub (iii). Analogicznie do wyżej opisanej metody Aggrescan^{3D}, metody te mogą być łatwo rozszerzone o wykorzystanie predykcji CABS-flex;
- metod dokowania molekularnego (np. według schematu tzw. *ensemble docking* (Korb, Olsson et al. 2012));
- metod projektowania leków lub projektowania białek (np. do analizy wpływu mutacji lub stabilności białek o różnej sekwencji) a w szczególności do wstępnych badań przesiewowych obejmujących wiele różnych wariantów białek (do takich badań klasyczne metody modelowania – np. pełnoatomowa MD – są zbyt kosztowne);
- metod umożliwiających interpretację danych eksperymentalnych o giętkości strukturalnej białek (np. w interpretacji danych NMR, lub krystalografii białek – w metodach *molecular replacement*).

2. Wykorzystanie metody CABS-fold

Metoda CABS-fold (**praca H4**) została udostępniona w formie darmowego serwera umożliwiającego łatwe wykonanie procesu przewidywania struktury białek. Przez około półtora roku (od maja 2013 do września 2014), zanotowaliśmy około 15,0 tysięcy wyświetleń stron internetowych serwera CABS-fold, w tym około 3,3 tysięcy unikalnych użytkowników. Do września 2014, metoda CABS-fold została użyta w opublikowanych badaniach dotyczących zależności struktura-funkcja białek odpowiedzialnych za ważne procesy życiowe:

- przez Cong i innych w przewidywaniu struktury białka prionowego (Cong, Casiraghi et al. 2013);
- przez Mosalaganti (Mosalaganti 2014) w przewidywaniu fragmentów struktury białka odpowiedzialnego za podział chromosomów podczas podziału komórkowego.

W porównaniu z innymi metodami do przewidywania struktury białek, unikalną⁵ cechą modelu CABS jest przewidywanie struktury wyłącznie na podstawie sekwencji (tzw. modelowanie *de novo*) dla: (i) białek do długości około stu reszt aminokwasowych lub (ii)

⁵ Model CABS należy do bardzo wąskiej grupy, dwóch-czterech stosowanych obecnie i dobrze opisanych w literaturze wydajnych metod modelowania gruboziarnistego umożliwiających modelowanie *de novo* dla białek lub fragmentów białek o znacznej długości.

długich fragmentów białek (Jamroz and Kolinski 2010). Unikalne możliwości modelowania *de novo* w oparciu o model CABS zostały pozytywnie zweryfikowane podczas 6-go ogólnoświatowego eksperymentu CASP (*The Critical Assessment of protein Structure Prediction* (Kolinski and Bujnicki 2005)). Ponadto, należy wspomnieć, że metody CAS and TOUCHSTONE są oparte na metodologii modelu CABS w zakresie modelowania *de novo* i są aktualnie jednymi z najlepszych narzędzi do tego celu (Zhang, Kolinski et al. 2003; Roy, Kucukural et al. 2010; Xu, Zhang et al. 2011). Możliwości modelowania *de novo* przy użyciu metody CABS-fold zostały zilustrowane przykładami w opisie wyników modelowania na http://biocomp.chem.uw.edu.pl/CABSfold/example_results.php.

Serwer CABS-fold jest przystosowany do wykonania następujących zadań modelowania molekularnego:

- przewidywanie struktury tylko na podstawie sekwencji aminokwasowej (użytkownik może również określić spodziewaną strukturę drugorzędową oraz dowolne więzy odległości pomiędzy wybranymi resztami aminokwasowymi, które będą użyte w symulacji);
- przewidywanie struktury na podstawie szablon(u/ów) – w przypadku dostarczenia wielu alternatywnych szablonów, głównym kryterium wiarygodności szablonów jest zgodność ich struktury (tzw. *consensus modeling*, zgodne fragmenty szablonów są traktowane jako wiarygodne i stanowią ścisły wzorzec dla modelowania porównawczego, podczas gdy niezgodne są traktowane z niskim poziomem ufności);
- przewidywanie struktury pętli lub brakujących fragmentów białek – w przypadku dostarczenia jednego szablonu z brakującym fragmentem białka (np. w miejscu występowania pętli) serwer CABS-fold umożliwia wymodelowanie brakującego fragmentu, podczas gdy reszta białka pozostaje „zamrożona” w konformacji podanej przez użytkownika;
- przewidywanie struktury białka z użyciem fragmentarycznych danych eksperymentalnych. Dane takie są przekazywane w postaci więzów odległości pomiędzy wybranymi węglami alfa;
- przewidywanie struktury stanów pośrednich (występujących na ścieżce zwijania białek globularnych) lub przewidywanie struktury białek nie wykazujących stabilnej struktury natywnej (np. białek nieustrukturyzowanych) – w celu przewidywania mało stabilnych form struktury białek należy przeanalizować wszystkie modele reprezentujące wyniki klastrowania lub/i trajektorię modelowania w reprezentacji gruboziarnistej (dostępne w wynikach modelowania serwera CABS-fold).

3. Wykorzystanie pozostałych opracowanych metod oraz wyników modelowania:

- zaproponowany sposób przewidywania pętli zewnątrz komórkowych białek błonowych (praca **H5**) stanowi wydajną alternatywę dla innych metod modelowania. Ponadto, wyniki przedstawione w pracy **H5** (oraz zestaw modeli

wynikowych upublicznionych w ramach tej pracy) stanowią punkt odniesienia dla innych prób modelowania tych ważnych celów terapeutycznych (przedstawiono wyniki modelowania dla 13 receptorów GPCR i dla większości z nich po raz pierwszy). Przedstawione wyniki sugerują, że możliwe jest dalsze udoskonalenie metody.

- wyniki badania mechanizmu funkcjonowania białka opiekuńczego (**praca H6**) pozwoliły na zaproponowanie nowej hipotezy, która może być przetestowana przez odpowiednio zaprojektowane prace doświadczalne. Ponadto, w **pracy H6** oraz **H7** przedstawiono szczegółową interpretację mechanizmu zwijania białka BdpA, która również pozwala na dalszą szczegółową weryfikację metodami eksperymentalnymi lub teoretycznymi.
- nowa strategia symulacji zwijania i wiązania białek nieustrukturyzowanych, zaproponowana w **pracy H8**, rozszerzyła w sposób jakościowy możliwości dokowania molekularnego i wieloskalowego modelowania białek. Wykonane symulacje były unikalne pod względem uwzględnienia pełnej giętkości dokowanego peptydu oraz nie korzystania z wiedzy o natywnej formie dokowanego białka, ani o miejscu wiązania. Otrzymane wyniki symulacji pozwoliły na interpretację zgrubnych danych eksperymentalnych oraz zaproponowanie hipotezy odnośnie mechanizmu jednoczesnego zwijania i wiązania się białek nieustrukturyzowanych. Ze względu na ogromne zapotrzebowanie nauk biomedycznych w zakresie wiarygodnych teoretycznych symulacji procesów jednoczesnego dokowania i zwijania białek oraz brak narzędzi do tego celu, dalszy rozwój metody jednoczesnego dokowania i zwijania białek (przedstawionej w pracy **H8**) jest obecnie kontynuowany przez mnie oraz współpracowników. Kolejnym krokiem rozwoju metody będzie jej przystosowanie do udostępnienia w formie powszechnie dostępnego serwera obliczeniowego.
- procedura przejścia od wyników modelowania gruboziarnistego do rozdzielczości pełnoatomowej oraz dalszej optymalizacji struktury w rozdzielczości pełnoatomowej, przedstawiona w **pracy H7**, znalazła zastosowanie w metodach modelowania wieloskalowego (m. in. w **pracach H5, H8** i (Kouza, Hu et al. 2013)), oraz jest obecnie podstawą do rozwijania nowych, ulepszonych metod modelowania prowadzonych w Pracowni Teorii Biopolimerów na Wydziale Chemii, UW. Procedura powstała w oparciu o publicznie dostępne narzędzia modelowania i może być użyta również do rekonstrukcji i optymalizacji modeli pochodzących z modelu gruboziarnistego innego niż model CABS.
- procedura modelowania wieloskalowego przedstawiona w **pracy H9** jest obecnie używana w pracach badawczych prowadzonych w Pracowni Teorii Biopolimerów na Wydziale Chemii, UW (do symulacji większych układów niż ten przedstawiony w **pracy H9** oraz do rozwijania nowych, ulepszonych metod modelowania).
- procedury badania mechanizmu formowania się struktury białek globularnych przedstawione w pracy **H10** pozwalają na zautomatyzowane przeprowadzenie

szeregu symulacji przy użyciu modelu CABS oraz na ich analizę według schematów zaprojektowanych i przetestowanych w pracach (Kmieciak and Kolinski 2007; Kmieciak and Kolinski 2008).

W pracach wchodzących w skład osiągnięcia naukowego zaproponowałem, wraz ze współpracownikami, łatwo dostępne i proste w obsłudze narzędzia modelowania molekularnego: serwer CABS-flex (prace **H1**, **H2**, **H3**), oraz serwer CABS-fold (praca **H4**). Tego typu serwery są bardzo przydatne w popularyzowaniu nowych metod modelowania komputerowego, oraz – co bardzo ważne – umożliwiają łatwe wykorzystanie zaawansowanych metod modelowania przez badaczy niewyspecjalizowanych w dziedzinie modelowania np.: przez biologów i chemików doświadczalnych. Ponadto, przedstawione prace opisują nowatorskie narzędzia i techniki modelowania wieloskalowego (**H7**, **H8**, **H9**, **H10**), których elementy są możliwe do pobrania (ze strony <http://biocomp.chem.uw.edu.pl/tools> lub też z innych stron jeśli wykorzystano narzędzia opracowane przez inne grupy badawcze). Oprócz udostępnienia narzędzi, w ramach prac wchodzących w skład osiągnięcia naukowego, opublikowano również szereg szczegółowych danych wynikowych, oraz trójwymiarowych modeli, umożliwiających ich wykorzystanie w innych teoretycznych lub eksperymentalnych studiach badanych problemów (prace **H5**, **H6**, **H7**, **H8**).

Referencje

- Chennamsetty, N., V. Voynov, et al. (2009). "Design of therapeutic proteins with enhanced stability." *Proc Natl Acad Sci U S A* **106**(29): 11937-11942.
- Chouard, T. (2011). "Structural biology: Breaking the protein rules." *Nature* **471**(7337): 151-153.
- Conchillo-Sole, O., N. S. de Groot, et al. (2007). "AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides." *Bmc Bioinformatics* **8**: 65.
- Cong, X., N. Casiraghi, et al. (2013). "Role of Prion Disease-Linked Mutations in the Intrinsically Disordered N-Terminal Domain of the Prion Protein." *Journal of Chemical Theory and Computation* **9**(11): 5158-5167.
- Conner, M., S. R. Hawtin, et al. (2007). "Systematic analysis of the entire second extracellular loop of the V(1a) vasopressin receptor: key residues, conserved throughout a G-protein-coupled receptor family, identified." *J Biol Chem* **282**(24): 17405-17412.
- Daggett, V. and A. R. Fersht (2003). "Is there a unifying mechanism for protein folding?" *Trends Biochem Sci* **28**(1): 18-25.
- de Groot, N. S., V. Castillo, et al. (2012). "AGGRESCAN: method, application, and perspectives for drug design." *Methods Mol Biol* **819**: 199-220.
- Dogan, J., X. Mu, et al. (2013). "The transition state structure for coupled binding and folding of disordered protein domains." *Sci Rep* **3**: 2076.
- Dror, R. O., D. H. Arlow, et al. (2011). "Activation mechanism of the beta2-adrenergic receptor." *Proc Natl Acad Sci U S A* **108**(46): 18684-18689.
- Fersht, A. R. (1995). "Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications." *Proc Natl Acad Sci U S A* **92**(24): 10869-10873.
- Fisette, O., P. Lague, et al. (2012). "Synergistic applications of MD and NMR for the study of biological systems." *J Biomed Biotechnol* **2012**: 254208.
- Fraga, H., R. Grana-Montes, et al. (2014). "Association Between Foldability and Aggregation Propensity in Small Disulfide-Rich Proteins." *Antioxidants & Redox Signaling* **21**(3): 368-383.
- Fraser, C. M. (1989). "Site-directed mutagenesis of beta-adrenergic receptors. Identification of conserved cysteine residues that independently affect ligand binding and receptor activation." *J Biol Chem* **264**(16): 9266-9270.
- Gianni, S., N. R. Guydosh, et al. (2003). "Unifying features in protein-folding mechanisms." *Proc Natl Acad Sci U S A* **100**(23): 13286-13291.
- Gront, D., S. Kmiecik, et al. (2007). "Backbone building from quadrilaterals: a fast and accurate algorithm for protein backbone reconstruction from alpha carbon coordinates." *J Comput Chem* **28**(9): 1593-1597.
- Haq, S. R., C. N. Chi, et al. (2012). "Side-chain interactions form late and cooperatively in the binding reaction between disordered peptides and PDZ domains." *Journal of the American Chemical Society* **134**(1): 599-605.
- Hartl, F. U. and M. Hayer-Hartl (2009). "Converging concepts of protein folding in vitro and in vivo." *Nat Struct Mol Biol* **16**(6): 574-581.
- Henzler-Wildman, K. A., M. Lei, et al. (2007). "A hierarchy of timescales in protein dynamics is linked to enzyme catalysis." *Nature* **450**(7171): 913-916.
- Itzhaki, L. S., D. E. Otzen, et al. (1995). "The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding." *J Mol Biol* **254**(2): 260-288.
- Jamroz, M. and A. Kolinski (2010). "Modeling of loops in proteins: a multi-method approach." *BMC Struct Biol* **10**: 5.

- Jewett, A. I. and J. E. Shea (2009). "Reconciling theories of chaperonin accelerated folding with experimental evidence." *Cell Mol Life Sci* **67**(2): 255-276.
- Karplus, M. and J. Kuriyan (2005). "Molecular dynamics and protein function." *Proceedings of the National Academy of Sciences of the United States of America* **102**(19): 6679-6685.
- Klco, J. M., C. B. Wiegand, et al. (2005). "Essential role for the second extracellular loop in C5a receptor activation." *Nat Struct Mol Biol* **12**(4): 320-326.
- Kmiecik, S. and A. Kolinski (2007). "Characterization of protein-folding pathways by reduced-space modeling." *Proc Natl Acad Sci U S A* **104**(30): 12330-12335.
- Kmiecik, S. and A. Kolinski (2008). "Folding pathway of the b1 domain of protein G explored by multiscale modeling." *Biophysical Journal* **94**(3): 726-736.
- Kolinski, A. and J. M. Bujnicki (2005). "Generalized protein structure prediction based on combination of fold-recognition with de novo folding and evaluation of models." *Proteins-Structure Function and Bioinformatics* **61 Suppl 7**: 84-90.
- Korb, O., T. S. G. Olsson, et al. (2012). "Potential and Limitations of Ensemble Docking." *J Chem Inf Model* **52**(5): 1262-1274.
- Kouza, M., C.-K. Hu, et al. (2013). "A structure-based model fails to probe the mechanical unfolding pathways of the titin I27 domain." *J Chem Phys* **139**(6): -.
- Levitt, M. and A. Warshel (1975). "Computer-Simulation of Protein Folding." *Nature* **253**(5494): 694-698.
- Liu, L., M. Werner, et al. (2014). "Collapse of a Long Axis: Single-Molecule Förster Resonance Energy Transfer and Serpin Equilibrium Unfolding." *Biochemistry* **53**(18): 2903-2914.
- Liwo, A., Y. He, et al. (2011). "Coarse-grained force field: general folding theory." *Phys Chem Chem Phys* **13**(38): 16890-16901.
- Lucent, D., J. England, et al. (2009). "Inside the chaperonin toolbox: theoretical and computational models for chaperonin mechanism." *Phys Biol* **6**(1): 015003.
- Meyer, T., M. D'Abamo, et al. (2010). "MoDEL (Molecular Dynamics Extended Library): a database of atomistic molecular dynamics trajectories." *Structure* **18**(11): 1399-1409.
- Mosalaganti, S. (2014). *Structural studies on human RZZ and Cln3p*. PhD, Max-Planck-Institut für Molekulare Physiologie.
- Nanevicz, T., L. Wang, et al. (1996). "Thrombin receptor activating mutations. Alteration of an extracellular agonist recognition domain causes constitutive signaling." *J Biol Chem* **271**(2): 702-706.
- Newo, A. S. (2014). "Molecular Modeling of the Plasmodium falciparum Pre-mRNA Splicing and Nuclear Export Factor PfU52." *The Protein Journal* **33**(4): 354-368.
- Ott, T. R., B. E. Troskie, et al. (2002). "Two mutations in extracellular loop 2 of the human GnRH receptor convert an antagonist to an agonist." *Mol Endocrinol* **16**(5): 1079-1088.
- Peeters, M. C., G. J. van Westen, et al. (2011). "Importance of the extracellular loops in G protein-coupled receptors for ligand recognition and receptor activation." *Trends Pharmacol Sci* **32**(1): 35-42.
- Pujagut, R. (2013). *AGGREGSCAN3D: A tool to predict aggregation propensity in protein surfaces*. MSc, Univesitat Autònoma de Barcelona.
- Roy, A., A. Kucukural, et al. (2010). "I-TASSER: a unified platform for automated protein structure and function prediction." *Nat Protoc* **5**(4): 725-738.
- Rueda, M., C. Ferrer-Costa, et al. (2007). "A consensus view of protein dynamics." *Proc Natl Acad Sci U S A* **104**(3): 796-801.
- Russel, D., K. Lasker, et al. (2009). "The structural dynamics of macromolecular processes." *Curr Opin Cell Biol* **21**(1): 97-108.
- Scheraga, H. A., M. Khalili, et al. (2007). "Protein-folding dynamics: overview of molecular simulation techniques." *Annu Rev Phys Chem* **58**: 57-83.

- Seibt, B. F., A. C. Schiedel, et al. (2013). "The second extracellular loop of GPCRs determines subtype-selectivity and controls efficacy as evidenced by loop exchange study at A2 adenosine receptors." Biochem Pharmacol **85**(9): 1317-1329.
- Shaw, D. E. (2013). "Millisecond-Long Molecular Dynamics Simulations of Proteins on a Special-Purpose Machine." Biophysical Journal **104**(2): 45a-45a.
- Shaw, D. E., P. Maragakis, et al. (2010). "Atomic-Level Characterization of the Structural Dynamics of Proteins." Science **330**(6002): 341-346.
- Shi, L. and J. A. Javitch (2002). "The binding site of aminergic G protein-coupled receptors: the transmembrane segments and second extracellular loop." Annu Rev Pharmacol Toxicol **42**: 437-467.
- Sugase, K., H. J. Dyson, et al. (2007). "Mechanism of coupled folding and binding of an intrinsically disordered protein." Nature **447**(7147): 1021-1025.
- Takada, S. (2012). "Coarse-grained molecular simulations of large biomolecules." Curr Opin Struct Biol **22**(2): 130-137.
- Vendruscolo, M. (2007). "Determination of conformationally heterogeneous states of proteins." Curr Opin Struct Biol **17**(1): 15-20.
- Vendruscolo, M. and C. M. Dobson (2011). "Protein Dynamics: Moore's Law in Molecular Biology." Current Biology **21**(2): R68-R70.
- Warshel, A. and M. Levitt (1976). "Theoretical Studies of Enzymic Reactions - Dielectric, Electrostatic and Steric Stabilization of Carbonium-Ion in Reaction of Lysozyme." J Mol Biol **103**(2): 227-249.
- Wheatley, M., D. Wootten, et al. (2012). "Lifting the lid on GPCRs: the role of extracellular loops." Br J Pharmacol **165**(6): 1688-1703.
- White, G. W., S. Gianni, et al. (2005). "Simulation and experiment conspire to reveal cryptic intermediates and a slide from the nucleation-condensation to framework mechanism of folding." J Mol Biol **350**(4): 757-775.
- Whitford, P. C., J. K. Noel, et al. (2009). "An all-atom structure-based potential for proteins: bridging minimal models with all-atom empirical forcefields." Proteins **75**(2): 430-441.
- Xu, D., J. Zhang, et al. (2011). "Automated protein structure modeling in CASP9 by I-TASSER pipeline combined with QUARK-based ab initio folding and FG-MD-based structure refinement." Proteins-Structure Function and Bioinformatics **79 Suppl 10**: 147-160.
- Xu, D. and Y. Zhang (2011). "Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization." Biophysical Journal **101**(10): 2525-2534.
- Zhang, Y., A. Kolinski, et al. (2003). "TOUCHSTONE II: a new approach to ab initio protein structure prediction." Biophysical Journal **85**(2): 1145-1164.
- Zhao, M. M., J. Hwa, et al. (1996). "Identification of critical extracellular loop residues involved in alpha 1-adrenergic receptor subtype-selective antagonist binding." Mol Pharmacol **50**(5): 1118-1126.

Sebastian Kurecik