

**dr Dorota Latek**  
**Uniwersytet Warszawski**  
**Wydział Chemii**

## **AUTOREFERAT**

**Wyznaczenie struktur i mechanizmów działania wybranych  
białek transbłonowych przy uwzględnieniu ich  
różnorodności ewolucyjnej**

*Warszawa, kwiecień 2019*

## Spis treści

1. Imię i Nazwisko.....	3
2. Informacje o posiadanych dyplomach i uzyskanych stopniach naukowych.....	3
3. Informacje o dotychczasowym zatrudnieniu w jednostkach naukowych .....	3
4. Wskazanie osiągnięcia 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. z 2018 r. poz. 1789):	
a) Tytuł osiągnięcia naukowego .....	4
b) Wykaz publikacji naukowych stanowiących podstawę osiągnięcia naukowego .....	4
c) Omówienie celu naukowego/artystycznego ww. prac i osiągniętych wyników wraz z omówieniem ich ewentualnego wykorzystania.....	5
I. Cel przeprowadzonych badań .....	5
II. Stan obecnej wiedzy.....	7
III. Opracowana metodologia badań .....	12
IV. Szczegółowe omówienie osiągnięcia i jego wykorzystania.....	18
V. Podsumowanie .....	31
VI. Literatura .....	32
5. Omówienie pozostałych osiągnięć naukowo-badawczych .....	39

## 1. Imię i nazwisko

Dorota Latek

## 2. Informacje o posiadanych dyplomach i uzyskanych stopniach naukowych

2009	<b>doktor nauk chemicznych</b> Rozprawa doktorska: “Modelowanie struktury białek na podstawie fragmentarycznych i niedokładnych danych doświadczalnych” – promotor pracy: prof. dr hab. Andrzej Koliński (Wydział Chemii Uniwersytetu Warszawskiego)
2004	<b>magister chemii</b> Praca magisterska: “Przewidywanie struktury białek globularnych w oparciu o fragmentaryczne dane NMR” – promotor: prof. dr hab. Andrzej Koliński (Wydział Chemii Uniwersytetu Warszawskiego)

## 3. Informacje o dotychczasowym zatrudnieniu w jednostkach naukowych

2004-2009	Wydział Chemii Uniwersytetu Warszawskiego Studia doktoranckie
2009-2010	Instytut Chemii Fizycznej Polskiej Akademii Nauk Staż podoktorski (forma zatrudnienia: umowa o dzieło)
2010-2013	Międzynarodowy Instytut Biologii Molekularnej i Komórkowej w Warszawie Staż podoktorski (forma zatrudnienia: umowa o pracę)
2011	European Molecular Biology Laboratory, Heidelberg, Niemcy Staż podoktorski (5 miesięcy, forma zatrudnienia: umowa o pracę)
2014	University of California San Francisco, Stany Zjednoczone Staż podoktorski (3 miesiące, forma zatrudnienia: stypendium)
Od 2013 r.	Wydział Chemii Uniwersytetu Warszawskiego adiunkt (forma zatrudnienia: umowa o pracę)

4. Wskazanie osiągnięcia 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. z 2018 r. poz. 1789):

a) Tytuł osiągnięcia naukowego

*Wyznaczenie struktur i mechanizmów działania wybranych białek transbłonowych przy uwzględnieniu ich różnorodności ewolucyjnej*

b) Wykaz publikacji naukowych stanowiących podstawę osiągnięcia naukowego

Lp.	Autor/autorzy, rok wydania, tytuł publikacji, czasopismo, tom, strony	Impact Factor <sup>1</sup>	Liczba cytowań <sup>2</sup>
H1	Latek D, Kolinski M, Ghoshdastider U, Debinski A, Bombolewski R, Plazinska A, Jozwiak K & Filipek S* (2011) <i>Modeling of ligand binding to G protein coupled receptors: cannabinoid CB1, CB2 and adrenergic beta 2 AR.</i> <b>J Mol Model</b> , 17, 2353-2366	1.797	14
H2	Dreisigacker S, Latek D, Bockelmann S, Huss M, Wiczorek H, Filipek S, Gohlke H, Menche D & Carlomagno T* (2012) <i>Understanding the inhibitory effect of highly potent and selective archazolides binding to the vacuolar ATPase.</i> <b>J Chem Inf Model</b> , 52, 2265-2272	4.304	11
H3	Latek D, Modzelewska A, Trzaskowski B, Palczewski K & Filipek S* (2012) <i>G protein-coupled receptors--recent advances.</i> <b>Acta Biochim Pol</b> , 59, 515-529	1.185	48
H4	Trzaskowski B, Latek D, Yuan S, Ghoshdastider U, Debinski A & Filipek S* (2012) <i>Action of molecular switches in GPCRs--theoretical and experimental studies.</i> <b>Curr Med Chem</b> , 19, 1090-1109	4.07	157
H5	Latek D*, Pasznic P, Carlomagno T & Filipek S* (2013) <i>Towards Improved Quality of GPCR Models by Usage of Multiple Templates and Profile-Profile Comparison.</i> <b>PLOS ONE</b> , 8, e56742	3.534	34
H6	Latek D*, Bajda M, Filipek S* (2016) <i>A Hybrid Approach to Structure and Function Modeling of G Protein-Coupled Receptors.</i> <b>J Chem Inf Model</b> , 56(4), 630-641	3.76	10
H7	Latek D* (2017) <i>Rosetta Broker for membrane protein structure prediction: concentrative nucleoside transporter 3 and corticotropin-releasing factor receptor 1 test cases.</i>	1.308	2

	<b>BMC Structural Biology</b> , 17(1), 8		
<b>H8</b>	Miszta P, Pasznik P, Jakowiecki J, Szttyler A, <b>Latek D</b> , Filipek S* (2018) <i>GPCRM – a homology modeling web service with triple membrane-fitted quality assessment of GPCR models</i> . <b>Nucleic Acid Research</b> , 46(W1), W387-W395	11.561 <sup>3</sup>	1
<b>H9</b>	Pasznik P, Rutkowska E, Niewieczeral S, Cielecka-Piontek J, <b>Latek D*</b> (2019) <i>Potential off-target effects of beta-blockers on gut hormone receptors: in silico study including GUT-DOCK – a web service for small-molecule docking</i> . <b>PLOS ONE</b> , 14(1), e0210705	2.766 <sup>3</sup>	1
<b>H10</b>	<b>Latek D*</b> , Rutkowska E, Niewieczeral S, Cielecka-Piontek J (2019) <i>Drug-induced diabetes type 2: In silico study involving class B GPCRs</i> . <b>PLOS ONE</b> , 14(1), e0208892	2.766 <sup>3</sup>	0

\* - autor korespondencyjny

<sup>1</sup>zgodnie z rokiem publikacji

<sup>2</sup>wg Web of Science z dnia 08.04.2019

<sup>3</sup>brak danych z roku publikacji, impact factor z roku 2017

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

### I. Cel przeprowadzonych badań

Struktura oraz funkcja białek danego organizmu jest kształtowana w toku ewolucji przez szereg czynników środowiskowych prowadzących do wystąpienia określonego fenotypu, często celowo indukowanego [1]. W mikroskali zmiany te są obserwowane na podstawie zmienności sekwencji nukleotydowej i aminokwasowej [2]. Celem moich badań przeprowadzonych przy uwzględnieniu różnorodności ewolucyjnej białek transbłonowych opisanych w niniejszym autoreferacie było:

- stworzenie kompletnej procedury modelowania ukierunkowanej zarówno na strukturę, jak i funkcję wybranych rodzin białek transbłonowych: receptorów sprzężonych z białkiem G (ang. G protein-coupled receptors, GPCR) [3], transporterów (inaczej: przenośników) substancji rozpuszczonych (ang. solute carriers, SLCs) [4] oraz adenylozotryfosfataz (ATPaz) [5], co opisano m.in. w podrozdziałach: **GPCRM — aplikacja internetowa ukierunkowana na strukturę receptora GPCR**, **GUT-DOCK — aplikacja internetowa ukierunkowana na funkcję receptora GPCR** oraz

## **Charakterystyka strukturalna wybranych białek transbłonowych** (dwie ostatnie rodziny białkowe)

- wyznaczenie istotnych funkcjonalnie cech struktury wybranych reprezentantów wymienionych powyżej rodzin białkowych (receptory GPCR podobne do rodopsyny, sekretynowe, receptory GPCR z klasy F, wakuolarna V-ATPaza, transporter koncentrujący nukleozydy CNT3), co opisano w podrozdziale: **Charakterystyka strukturalna wybranych białek transbłonowych**, za pomocą stworzonej procedury oraz narzędzi, w których została zaimplementowana
- opisanie molekularnych mechanizmów działania wybranych białek transbłonowych związanych z ich naturalnym procesem aktywacji oraz lekowo lub chorobowo indukowanym zahamowaniem ich funkcji, co opisano w podrozdziałach: **Badania nad mechanizmem aktywacji receptorów GPCR, Projektowanie inhibitorów białek transbłonowych** oraz **Molekularne podstawy reakcji polekowych**
- zbadanie wpływu występowania polimorfizmów jednego nukleotydu na strukturę oraz funkcję wybranych białek transbłonowych oraz wyjaśnienie związanych z nimi określonych fenotypów obserwowanych w badaniach klinicznych, co opisano w podrozdziale: **Molekularne podstawy istotnych farmakologicznie polimorfizmów**

Informacja ewolucyjna, w postaci oceny zmienności sekwencji (m.in. profil sekwencyjny, motywy sekwencyjne) [2], podobieństwa struktur oraz kofaktorów i substratów homologów sekwencyjnych [6] razem z potencjałami statystycznymi [7] została przeze mnie użyta przede wszystkim w modelowaniu struktur białkowych [H1-2] oraz [H5-10]. Informacja ewolucyjna w postaci podobieństwa funkcji białek homologicznych [H3-4] została przeze mnie użyta w badaniach nad molekularnymi mechanizmami aktywacji białek transbłonowych o nierozwiązanych do tej pory strukturach. Zbudowane przeze mnie modele białek transbłonowych uwzględniające odpowiadające ich sekwencji warianty genetyczne zidentyfikowane u pacjentów [8], znalazły również zastosowanie biomedyczne [H9-10], w badaniach z dziedziny medycyny spersonalizowanej [9].

Jako cel badań wybrałam integralne białka transbłonowe ze względu na ich istotną rolę nie tylko w ścieżkach sygnałowych komórki, ale również w transporcie i metabolizmie substancji endo i egzogennych [10]. Ponadto, wybrane przeze mnie rodziny białek są niezwykle trudne do badania metodami eksperymentalnymi, nie tylko w przypadku odpowiedniej substytucji ich natywnego niepolarnego środowiska podczas eksperymentu [11], ale również w przypadku rejestracji i analizy otrzymanych danych [3, 4, 5].

Opracowanie przeze mnie spójnej i kompletnej procedury modelowania dla tych białek wymagało osiągnięcia równowagi między dążeniem do jak największego podobieństwa całego modelu białka do struktury natywnej [12] a optymalizacją procedury w kierunku dokładnego odbudowania samego rejonu miejsca aktywnego przy jednoczesnym osiągnięciu powtarzalności wyników, co jest niezbędne podczas projektowania nowych leków [13, 14]. Implementacja procedury modelowania w postaci aplikacji internetowych dostarczyła prostych narzędzi do przeprowadzenia podstawowych badań teoretycznych dotyczących dowolnego receptora GPCR o nieznannej strukturze [H5-6] [H8-10]. Dzięki temu stała się możliwa nie tylko ich dokładna charakterystyka strukturalna, stanowiąca podstawę do badań eksperymentalnych, ale również szczegółowe badanie mechanizmów ich aktywacji, co opisałam w publikacjach [H1] [H3-4]. Moim celem było również porównanie cech strukturalnych i mechanizmów aktywacji receptorów GPCR pochodzących z dwóch, różniących się podrodzin: receptorów podobnych do rodopsyny (wymienne stosowana nazwa: klasa A receptorów GPCR) oraz receptorów sekretynowych (klasa B). Receptorom z klasy A były poświęcone moje publikacje [H1] [H3-6] [H8], receptorom klasy B prace [H9-10], reprezentantowi klasy F (receptor SMO) część publikacji [H6], a dokładnemu omówieniu wszystkich klas receptorów GPCR publikacje przeglądowe [H3-4].

Istotna część obecnie stosowanych w farmakoterapii leków jest ukierunkowana na cele biologiczne o nierozwiązanej jeszcze strukturze [15]. Znajomość struktury celu biologicznego leku pozwala na poprawienie jego selektywności, a przede wszystkim zmniejsza prawdopodobieństwo wystąpienia niepożądanych reakcji polekowych. Dlatego celem moich badań było wyznaczenie molekularnych podstaw polekowej cukrzycy typu II i możliwości jej uniknięcia [H9-10] oraz dodatkowo określenie roli różnorodności genetycznej pacjentów w przebiegu farmakoterapii analogami nukleozydów [H7].

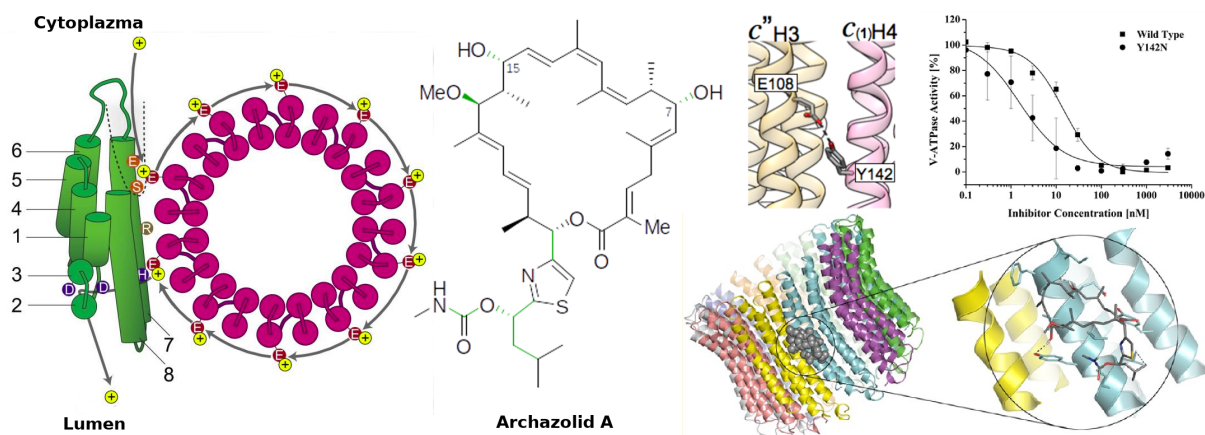
## II. Stan obecnej wiedzy

Szczególnie istotną rolę metod teoretycznych w badaniach dotyczących białek transbłonowych jest dostarczanie wstępnego modelu białka (najczęściej homologicznego) podczas rozwiązywania jego struktury na podstawie danych pochodzących z krystalografii rentgenowskiej i kriomikroskopii elektronowej [16]. W drugim przypadku dokładne odczytanie danych strukturalnych odzwierciedlonych na zdjęciu mikroskopowym wymaga zbudowania startowej konformacji białka, która następnie jest optymalizowana na podstawie tych danych [17]. Podobne podejście jest od wielu lat stosowane w krystalografii rentgenowskiej, w której

do mapy gęstości elektronowej atomów dopasowywany jest model białka [18]. Różnice w stosunku do białek globularnych polegają na stworzeniu podczas eksperymentu szczególnego, niepolarnego środowiska imitującego błonę komórkową, np. za pomocą miceli [19] lub nanodysków lipidowych [11]. W przypadku badania metodami krystalografii białek błonowych podczas ich działania, np. w konformacji aktywnej w przypadku receptorów GPCR, konieczne jest dodatkowo ustabilizowanie danego stanu konformacyjnego białka. Stało się to możliwe w ostatnich latach dzięki odkryciu stabilizujących receptor nanocząstek (rekombinowanych antygenów) [20]. Brian Kobilka wraz ze współpracownikami ze Stanów Zjednoczonych i Europy w roku 2011 po raz pierwszy otrzymał strukturę krystaliczną receptora GPCR ( $\beta$ 2-adrenergicznego) w konformacji aktywnej w oddziaływaniu z białkiem G [21], dopiero po 11 latach od rozwiązania pierwszej struktury receptora GPCR (rodopsyny) [19] oraz jej dimerów [22] przez Krzysztofa Palczewskiego i współpracowników (Sławomira Filipka). Inne rozwiązanie zastosowane dla receptora adenozynowego A2A zostało zaproponowane przez Heptares Therapeutics i polega na termostabilizacji stanu aktywnego receptora GPCR przez mutacje punktowe (technologia StaR®) [23]. Ostatnio opublikowana struktura aktywnego receptora GPCR (kalcytoninowego CT-R z podrodziny receptorów sekretynowych) o rozdzielczości 3.3 Å została otrzymana metodą kriomikroskopii elektronowej przy zastosowaniu szczególnej metody rejestrowania danych umożliwiającej otrzymanie struktury o wysokiej rozdzielczości (płytko fazowa Volta) [24]. O trudnościach eksperymentalnych dotyczących stabilizacji stanu aktywnego receptorów GPCR świadczy fakt, że część dotychczas rozwiązanych struktur krystalicznych reprezentuje konformację tylko stanu pośredniego, czyli nie w pełni aktywnego, pomimo związania całkowitego lub częściowego agonisty [25].

Podobne trudności eksperymentalne dotyczą pozostałych dwóch rodzin białek transbłonowych, które badałam. Struktury rotora V-ATPazy ilustrujące jego działanie zostały rozwiązane metodą kriomikroskopii elektronowej przy odtworzeniu niepolarnego środowiska błony komórkowej za pomocą detergentu DDM ( $\beta$ -D-maltozyd n-dodecyłu) [26] oraz nanodysków lipidowych [27] dopiero po, odpowiednio, 3 i 6 latach po moich badaniach opisanych w publikacji [H2]. M.in. dopiero w [26] pokazano stadia pośrednie ilustrujące mechanizm działania rotora (patrz Rys. 1).





**Rysunek 1.** Mechanizm działania domeny transbłonowej maszyny molekularnej - rotora V-ATPazy (lewy panel). Indukowana przez hydrolizę ATP rotacja pierścienia helikalnego umożliwia przenoszenie jonów wodorowych (żółte) przez błonę komórkową za pośrednictwem m.in. reszt kwasu glutaminowego (czerwone). Środkowy panel - wzór strukturalny jej inhibitora — archazolidu A (blokującego m.in. resztę Y142) o aktywności oraz sposobie wiązania do białka (prawy panel) potwierdzonych doświadczalnie w publikacji [H2]. Na podstawie: Shep DG et al. PNAS 2016, 113 (12) 3245-3250, Roh SH et al. Mol Cell 2018, 69(6):993-1004.e3. oraz [H2].

Struktura bakteryjnego transportera koncentrującego nukleozydy (ang. concentrative nucleoside transporter) vCNT3 z rodziny SLC28 w pojedynczej konformacji zamkniętej (wiążącej ligand) skierowanej do wewnątrz komórki (ang. inward occluded) została otrzymana metodami krystalografii rentgenowskiej, również przy zastosowaniu detergentu DDM jako solubilizatora [28]. Otrzymana struktura vCNT3 posłużyła w moich badaniach jako szablon strukturalny w modelowaniu przez homologię. Niedawno, w roku 2017, za pomocą krystalografii rentgenowskiej, otrzymano struktury odpowiadające pełnemu cyklowi działania (od konformacji otwartej na zewnątrz komórki do konformacji otwartej do wewnątrz, po przeniesieniu substratu) innego transportera z rodziny SLC29 [29].

Niewielka liczba dotychczas rozwiązanych metodami eksperymentalnymi struktur białek transbłonowych ilustruje trudności i jednocześnie jest uzasadnieniem moich badań teoretycznych opisanych w niniejszym autoreferacie. Trudności eksperymentalne dotyczą również prawidłowego rozpoznania potencjalnych celów biologicznych dla nowych substancji aktywnych farmakologicznie, co w konsekwencji prowadzi do występowania niepożądanych reakcji polekowych [30], np. z powodu modulacji innych ścieżek sygnałowych w komórce, co opisałam w publikacji [H9-10]. Testy funkcjonalne (ang. functional assays) masowo stosowane do oceny faktycznej (nie tylko powinowactwa) aktywności biologicznej ligandu [30], mają ograniczenia wynikające z tego, że eksperyment nie jest przeprowadzany w warunkach czynności biologicznej danego białka *in vivo* [31]. Możliwość aktywowania przez ligand kilku

ścieżek sygnałowych w komórce została potwierdzona np. dla uważanych za antagonistów receptora  $\beta$ -adrenergicznego leków beta-adrenolitycznych (carvedilol i nebivolol), które w rzeczywistości aktywują jako agoniści ścieżkę sygnałową  $\beta$ -arrestyny [30]. Nawet skринing fenotypowy [30], ze względu na standardową procedurę, nie jest w stanie wykryć słabych oddziaływań ligandów z receptorami związanych z efektami ubocznymi farmakoterapii obserwowanymi dopiero już w często trwających wiele lat badaniach klinicznych. Jest to spowodowane faktem, że zautomatyzowane protokoły stosowane np. w testach ADME-Tox pozwalają tylko wykluczyć związki o wysokiej cytotoksyczności, pozostawiając jednak związki o relatywnie małej toksyczności [32]. Przykładem jest tu choćby dokonany przeze mnie przegląd efektów ubocznych różnych klas leków dotyczący samej tylko cukrzycy, opisany w [H9-10]. Cukrzyca typu II należy do długofalowych skutków ubocznych farmakoterapii w odróżnieniu od natychmiastowych niepożądanych reakcji polekowych (ang. adverse drug reactions).

Dopiero w ostatnich latach, w wyniku istotnej liczby zgłaszanych przypadków niepożądanych reakcji polekowych [33], obserwowany jest wzrost zainteresowania molekularnymi mechanizmami powstawania takich reakcji i podejmowane są próby opracowywania protokołów przewidujących ich wystąpienie [34]. Badania nad mechanizmami efektów ubocznych są jednak w dużym stopniu utrudnione przez niekompletne, a często sprzeczne [35] wyniki badań klinicznych, co opisałam w mojej publikacji [H10] podając przykład dwóch niezależnych badań klinicznych dotyczących statyn. Sytuację komplikują ponadto coraz częstsze doniesienia związane z indywidualną odpowiedzią pacjenta na leczenie związaną z jego profilem genetycznym [36]. Podejmowane są jednak próby ustandaryzowania oraz upubliczniania informacji o niepożądanych reakcjach polekowych, np. w formie ogólnodostępnych baz danych jak SIDER (ang. Side Effect Resources) udostępnionych przez European Molecular Biology Laboratory [36]. Należy tu jednak wspomnieć, że niektóre zaobserwowane efekty uboczne leków mogą być wykorzystywane w repozycjonowaniu leków, które jest alternatywą do projektowania nowych substancji farmakologicznie aktywnych wymagających długotrwałych testów klinicznych przed zastosowaniem w leczeniu.

Przedstawione powyżej trudności eksperymentalne związane z otrzymaniem struktur i badaniem białek transbłonowych były przewyżczone na przestrzeni lat przez szereg różnych metod teoretycznych. Należą do nich: modelowanie porównawcze (w tym również przez homologię), modelowanie *de novo* (bez szablonu strukturalnego z białek homologicznych), budowania białka z fragmentów (ang. fragment assembly) informacja ewolucyjna w postaci skorelowanych mutacji aminokwasowych, metody przewlekania (ang. threading),

rozpoznawania topologii białka (ang. fold recognition) oraz wiele innych o mniejszym zasięgu stosowania. W dziedzinie projektowania nowych substancji aktywnych biologicznie standardowe metody teoretyczne obejmują przede wszystkim skryning wirtualny (oparty na strukturze liganda lub strukturze receptora) oraz optymalizację związku wiodącego w celu poprawy jego specyficzności i selektywności. Wymienione powyżej metody zostały przede mnie szczegółowo opisane we fragmentach wszystkich publikacji stanowiących podstawę niniejszego autoreferatu [H1-10] oraz dodatkowo w publikacjach [B1-9] oraz w 8 rozdziałach książkowych. Na największą uwagę zasługuje praca badawcza Andrzeja Kolińskiego w dziedzinie przewidywania struktur białkowych *de novo* [37] oraz innych polskich badaczy w dziedzinie modelowania białek uwzględniającego informację ewolucyjną [38, 39]. Opracowanie algorytmu Rosetta do budowania modelu białka z krótkich fragmentów aminokwasowych (ang. fragment assembly) [40] przez Davida Bakera i współpracowników okazało się przełomowe nie tylko dla metod obliczeniowych, ale dla całej dziedziny wiedzy dotyczącej struktury i funkcji białek. Opracowanie optymalnej metody konwersji informacji strukturalnej z szablonów białek homologicznych do postaci więzów opisujących białko o nieznaną strukturę przez Andreja Sali, twórcy MODELLER-a [41] umożliwiło charakteryzację strukturalną olbrzymiej liczby białek (baza danych MODBASE), podobnie jak inne narzędzie do modelowania porównawczego: SWISS-MODEL, opisane w pracy [42].

Zbyt mała populacja struktur białek transbłonowych w bazie Protein Data Bank (PDB) przez wiele lat była przeszkodą w opracowaniu potencjałów statystycznych do modelowania *de novo* (bez stosowania szablonu strukturalnego), a nawet w próbach modelowania porównawczego dla tych białek. W związku z tym niezwykle istotne wydają się publikacje z kilku ostatnich lat, dotyczące użycia w modelowaniu struktur białek transbłonowych informacji ewolucyjnej w postaci skorelowanych mutacji do przewidywania kontaktów [43, 44] oraz profili sekwencyjnych do trafnego wyznaczenia struktury drugorzędowej i topologii tych białek [45] uzupełnionej o pierwsze wyprowadzenia efektywnych potencjałów statystycznych [46]. Jednym z kierunków rozwoju metod obliczeniowych w obszarze badań nad białkami transbłonowymi jest modyfikacja już istniejących narzędzi (np. Rosetta-MP [47]) i pól siłowych (np. CHARMM [48], AMBER [49]) tak, aby uwzględniały błonę komórkową w sposób jawny (parametryzacja cząsteczek lipidów) lub niejawny (pośrednio) przez wprowadzenie do pola siłowego dodatkowych potencjałów charakteryzujących w przybliżony sposób oddziaływanie białka i ligandów z lipidami (np. metoda IMM1 [49]). Często stosowanym rozwiązaniem, również w przypadku moich badań opisanych w publikacjach [H1-2] [H7] jest użycie pełnoatomowej dynamiki molekularnej (GROMACS [51], NAMD [52]) do

poprawienia modelu białka transbłonowego i obserwacji mechanizmów jego działania, już w skali mikrosekundowej [53].

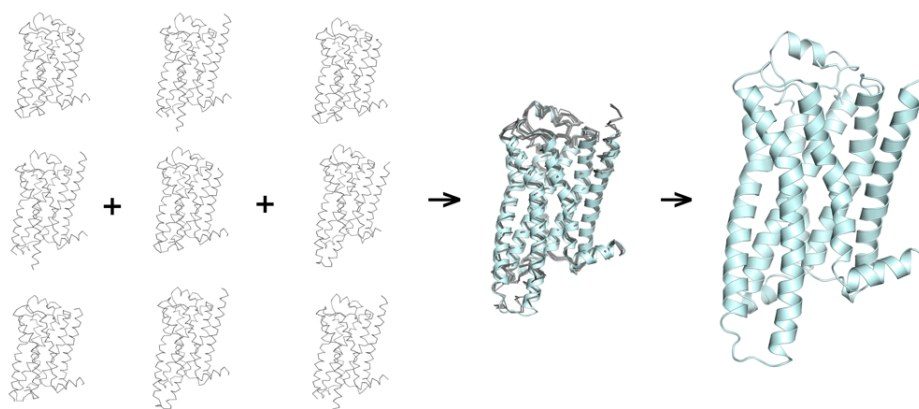
Niezwykle interesujące są rozpowszechnione w ostatnich latach metody obliczeniowe naśladujące dobór ewolucyjny organizmów (algorytmy genetyczne) lub sieci neuronowe (metody uczenia maszynowego), zastosowane przeze mnie w dokowaniu ligandów (Autodock [H1-2] oraz GOLD [H6]). W tych metodach właściwą odpowiedź na rozwiązanie problemu biologicznego (np. właściwej struktury białka lub sposobu wiązania się do niego ligandu) znajduje się, naśladując sposób, w jaki działa ewolucja, tzn. wybierając najkorzystniejszą z punktu widzenia wydatku energetycznego ścieżkę rozwoju dla danego organizmu w określonym środowisku zewnętrznym (np. algorytm genetyczny w programie do dokowania GOLD [54]).

### III. Opracowana metodologia badań

Stworzenie, optymalizacja i zastosowanie w badaniach nowej procedury modelowania struktur białek błonowych ze szczególnym uwzględnieniem receptorów GPCR zostało przeze mnie opisane w publikacjach: [H1] [H3-6] [H8-10] (receptory GPCR) oraz w publikacjach: [H2] [H7] (inne białka transbłonowe). Zastosowałam w niej m.in. poprawianie modelu za pomocą symulacji pełnoatomowej dynamiki molekularnej [H2] [H7] oraz modelowanie w oparciu o wiele uśrednionych szablonów [H5], które było nowatorskim rozwiązaniem problemu braku odpowiedniego szablonu strukturalnego, którego można by było użyć w tzw. „trudnym” modelowaniu przez homologię receptorów GPCR, gdy identyczność sekwencji celu i szablonu wynosi znacznie poniżej 30%. Wstępem do moich badań była opisana przeze mnie w publikacji [H3] analiza podobieństwa sekwencyjnego receptorów GPCR podobnych do rodopsyny i jej czterech gałęzi:  $\alpha$ ,  $\beta$ ,  $\gamma$  i  $\delta$ . Wykazała ona, że receptory z tej klasy są w większości niepodobne sekwencyjnie do siebie nawzajem (identyczność sekwencji poniżej 20%). Wyjątek stanowi wyraźnie heterogeniczna gałąź  $\delta$ , której dużą część stanowią receptory węchowe (ang. olfactory receptors), o identyczności sekwencji rzędu 40%. Na opisanym powyżej przykładzie receptorów podobnych do rodopsyny można wyraźnie zaobserwować zachowywanie w trakcie ewolucji struktury (siedmiohelikalna topologia) i funkcji białka (przekazywanie sygnału do wnętrza komórki) bez zachowywania sekwencji.

Rozwiązanie problemu „trudnego” modelowania przez homologię receptorów GPCR, opisałam szczegółowo w publikacji [H5]. Polega ono na ważonym uśrednieniu, zależnym od stopnia podobieństwa między całymi sekwencjami lub ich fragmentami, dowolnej liczby

dostępnych w bazie PDB struktur receptorów GPCR w procedurze budowania modelu białka za pomocą zmodyfikowanej metody gradientów sprzężonych z elementami dynamiki molekularnej (MODELLER [41]). Rozwiązanie to, umożliwiające zbudowanie modelu dowolnego receptora GPCR o nieznannej strukturze przy użyciu wielu szablonów (ang. multiple templates modeling; patrz Rys. 2), zaimplementowałam w aplikacji internetowej GPCRM opisanej w publikacjach [H5-6] oraz [H8]. Szablony strukturalne, czyli trójwymiarowe struktury białek z bazy PDB, są konwertowane na zbiór więzów odległości międzyatomowych i kątów dwuściennych między wiązaniami. Dodatkowo model białka jest charakteryzowany przez zbiór więzów stereochemicznych (długości wiązań, kąty między wiązaniami) otrzymywany z obliczeń mechaniki molekularnej (CHARMM-22 [48]) oraz przez wartości kątów dwuściennych i odległości międzyatomowych otrzymane na podstawie statystyki znanych struktur białkowych w bazie PDB. Wymienione powyżej dane strukturalne w kolejnym kroku są wykorzystywane do zbudowania modelu białka przez dopasowanie jego łańcucha aminokwasowego do wygenerowanych wcześniej więzów z prawdopodobieństwem uwzględniającym podobieństwo sekwencyjne i rozkład Boltzmana, co zostało opisane w oryginalnej publikacji autora MODELLERA (Sali et al. [55]).



**Rysunek 2.** Budowanie modelu receptora GPCR przy użyciu wielu szablonów strukturalnych pochodzących z białek homologicznych.

Ze względu na to, że wszystkie receptory GPCR są zbudowane z siedmiu helis transbłonowych (mają tę samą topologię) możliwe jest użycie więcej niż 2-3 szablonów strukturalnych bez straty rozdzielczości końcowego modelu [H5]. W przypadku białek globularnych o znacznie różniących się topologiach przekroczenie granicy 2-3 szablonów w modelowaniu przez homologię powoduje zahamowanie zwiększania się dokładności końcowego modelu z powodu użycia zbioru zbyt różniących się od siebie więzów

strukturalnych, niemożliwych do spełnienia jednocześnie [56]. Dzięki opisanej powyżej procedurze modelowania w oparciu o wiele szablonów strukturalnych możliwe stało się dokładne odwzorowanie pewnych typowych zniekształceń helis (załamań, wypukłości, bruzd), które są cechą charakterystyczną białek błonowych. Zostało to opisane w publikacji [H5] na przykładzie receptorów: adenozynowego A<sub>2A</sub> oraz opioidowego κ (w suplemencie).

W trakcie moich badań nad receptorami GPCR z klasy B opracowałam również procedurę modelowania ich allosterycznego miejsca aktywnego na przykładzie receptora hormonu uwalniającego kortykotropinę CRF<sub>1</sub> [H7]. W tym przypadku zniekształcenie helis transbłonowych (głównie helisy VI) obserwowane w strukturze krystalicznej było spowodowane szczególnym oddziaływaniem z antagonistą CP-376395 niemożliwym do skopiowania z żadnej innej dostępnej w PDB struktury receptora GPCR. Do modelowania allosterycznego miejsca aktywnego receptora CRF<sub>1</sub> zastosowałam algorytm modelowania białek błonowych *de novo* wykorzystujący potencjały statystyczne (Rosetta Broker [57]), który umożliwił znaczne odsunięcie helisy VI od miejsca aktywnego, pozwalając na prawidłowe zadokowanie antagonisty, co opisałam w publikacji [H7].

Budowanie modelu białka przy użyciu wielu szablonów strukturalnych jest poprzedzone opracowaną przeze mnie i opisaną w publikacji [H5] procedurą generowania dopasowania (inaczej: uliniowienia) sekwencji aminokwasowych celu i szablonów. Oprócz powszechnie stosowanych metod (dopasowanie dwóch sekwencji, dopasowanie wielu sekwencji) zaimplementowano dodatkowo dwie metody, wcześniej nieużywane w automatycznych procedurach modelowania receptorów GPCR. Pierwsza z nich, opisana w [H5], polega na dopasowaniu sekwencji celu i szablonu na podstawie dopasowania ich profili sekwencyjnych zawierających informację ewolucyjną na temat częstości występowania danego aminokwasu na danej pozycji w grupie sekwencji homologicznych (Gribskov et al. [58]). Profile sekwencyjne są generowane w tym przypadku za pomocą narzędzia BLAST (Altschul et al. [59]) na podstawie dostępnych w bazie NR (non-redundant) sekwencji homologicznych. Porównywanie profili sekwencyjnych było stosowane do poprawiania dopasowania sekwencyjnego w modelowaniu białek globularnych, m.in. przez polskich badaczy [60], jednak nie w szczególnym przypadku receptorów GPCR. W moich badaniach zaobserwowałam, że dopasowanie sekwencji aminokwasowych na podstawie dopasowania ich profili sekwencyjnych pozwoliło na detekcję przerw w dopasowaniach sekwencyjnych, które są związane z występowaniem bruzd (przerwa w sekwencji celu) i wypukłości (przerwa w sekwencji szablonu) w helisach transbłonowych receptorów GPCR, co opisałam w publikacji [H5] (również w suplemencie do tej publikacji).

Druga z metod generowania dopasowania sekwencyjnego, opracowana i opisana przeze mnie w publikacji [H8], opiera się na wstępnym dopasowaniu strukturalnym wszystkich użytych do modelowania szablonów. Na podstawie dopasowania strukturalnego generowane jest w kolejnym kroku dopasowanie sekwencji wielu szablonów do sekwencji celu. Metoda ta jest szczególnie przydatna w „trudnym” modelowaniu przez homologię, np. receptora GPCR z klasy F na szablonie strukturalnym receptora z klasy A (bardzo niskie podobieństwo sekwencyjne między celem i szablonem) [H6], w którym zawodzą wszystkie standardowe metody do automatycznego generowania dopasowania. Okazała się ona trafna w przypadku modelowania przeze mnie struktury receptora SMO (receptor z klasy F) podczas konkursu GPCR Dock 2013 [61] (patrz pkt 5 niniejszego autoreferatu oraz [B1]), co szczegółowo opisałam w publikacji [H6].

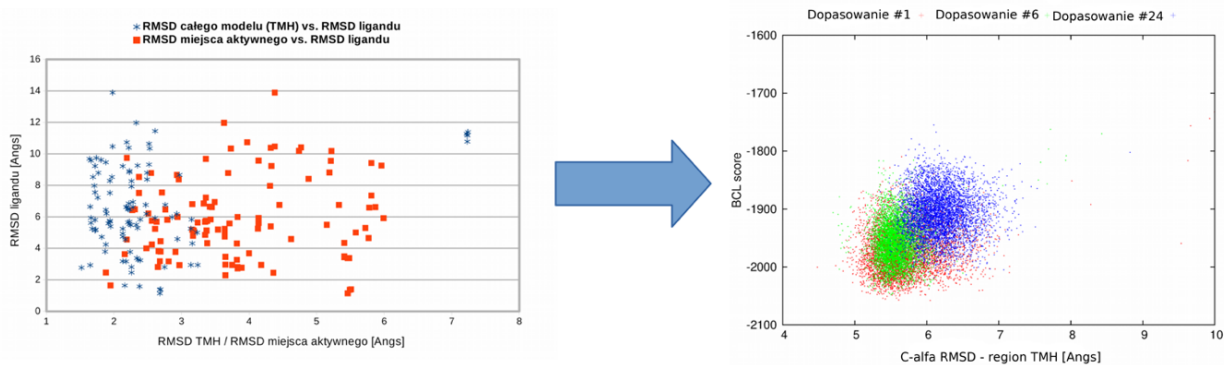
Modele receptorów GPCR zbudowane procedurą ważonego uśredniania wielu szablonów strukturalnych opisaną przeze mnie w [H5] w kolejnym kroku były oceniane za pomocą dwóch osobnych kryteriów w zależności od końcowego celu modelowania (badanie oddziaływań z innymi podjednostkami białkowymi lub projektowanie leków). W pierwszym przypadku wymagana była taka optymalizacja procedury modelowania, aby globalna (dla całego białka) wartość RMSD względem struktury natywnej z PDB lub innej miary oceniającej dokładność modelu białka (np. TM-score) była jak najniższa. W tym celu najkorzystniejsze okazało się użycie potencjałów statystycznych, co opisałam w publikacjach [H6] i [H8]. Potencjały statystyczne dotyczące oddziaływań krótko i dalekozasięgowych zostały wyprowadzone na podstawie dostępnych w PDB struktur białek błonowych. Zastosowałam tu potencjały zawarte w narzędziu BCL::Score [46], co opisałam w publikacji [H6]. Ten sposób oceny jakości modeli (ang. model quality assessment) został przeze mnie zaimplementowany w aplikacji internetowej GPCRM, co opisałam w publikacji [H8].

Drugie opracowane przeze mnie kryterium oceny wygenerowanego modelu receptora GPCR przydatne w projektowaniu leków lub badaniu mechanizmów oddziaływań z małowcząsteczkowymi ligandami zostało opisane w publikacji [H6] oraz [H9-10]. Polegało ono na optymalizacji procedury modelowania receptora w kierunku jak najlepszego odwzorowania jego oddziaływań z ligandem w rejonie miejsca aktywnego. W publikacji [H6] opisałam kryterium wyboru najlepszego modelu receptora GPCR na podstawie dwustopniowego dokowania ligandów. W pierwszym etapie ligand był dokowany za pomocą narzędzia GOLD [54]. Do kolejnego etapu – dokowania liganda ze zwiększoną precyzją za pomocą narzędzia Glide były wybierane te modele receptorów, dla których otrzymano najbardziej zbieżne predykcje sposobu wiązania liganda w pierwszym etapie. W badaniach opisanych w [H9-10]

zastosowałam inne podejście, oparte na współczynnikach wzbogacenia EF (ang. enrichment factors). W tym celu dla kilku wybranych receptorów GPCR z klasy B, ich ligandów o potwierdzonej eksperymentalnie aktywności oraz wygenerowanego za pomocą narzędzia DUD-E [13] zbioru niepodobnych chemicznie ligandów wyznaczyłam współczynniki EF (oraz inne miary: ROC, BEDROC, AUC). Modele receptorów GPCR o najwyższych współczynnikach EF były wybierane do dokowania ligandów. Dzięki temu kryterium wyboru uzyskałam modele receptorów o potwierdzonej skuteczności w wirtualnym skryningu, rozpoznające typowe dla danego receptora ligandy aktywne wśród nieaktywnych. Wyniki moich badań oddałam do użytku publicznego w postaci aplikacji internetowej GUT-DOCK [H9]. GUT-DOCK umożliwia dokowanie do wyżej wymienionych modeli receptorów (wiążących hormony jelitowe) dowolnych ligandów małocząsteczkowych oraz krótkich peptydów za pomocą narzędzia Autodock VINA [62]. Spośród dostępnych obecnie funkcji oceniających używanych w dokowaniu ligandów (opartych na polu siłowym, empirycznych, opartych na potencjałach statystycznych lub uczeniu maszynowym), wybrałam funkcję empiryczną zakodowaną w programie Autodock VINA i wyprowadzoną na podstawie bazy PDBbind zawierającą dane eksperymentalne na temat powinowactwa ligandów zawartych w kompleksach zdeponowanych w PDB.

W niektórych przypadkach oba opisane powyżej kryteria oceny modeli receptorów GPCR mogą dawać zbieżne wyniki (np. wyniki grupy Stockholm-Carlsson oraz Copenhagen-Gloriam, ale tylko dla jednego receptora 5-HT<sub>1B</sub>, w konkursie GPCR Dock 2013 [61] – patrz pkt 5 autoreferatu), jednak w przeważającej liczbie przypadków (patrz Rys. 3, lewy panel, niebieskie punkty) modele białek o najlepszych wynikach w wirtualnym skryningu nie są strukturami najbliższymi natywnych (czyli o najmniejszym globalnym RMSD względem natywnej), co zostało opisane szerzej w mojej publikacji [H6] i we wcześniejszej publikacji dotyczącej białek globularnych [14]. Nieco lepszą korelację obserwuje się między dokładnością predykcji samego miejsca aktywnego i wyników dokowania ligandów (patrz Rys. 3, lewy panel, czerwone punkty). Choć wyniki konkursu GPCR Dock 2013 stanowią trudny do interpretacji heterogeniczny zbiór danych, co opisałam w publikacji [H6] przeprowadzając zamiast tego obliczenia na homogenicznych zbiorach danych wygenerowanych w czasie konkursu przez naszą grupę (Warsaw), to dają jednak wyobrażenie co do obecnych trudności w modelowaniu kompleksów receptorów GPCR.





**Rysunek 3.** Brak zależności między wartością RMSD względem struktury natywnej dla całego modelu a wynikami dokowania ligandu (w postaci RMSD ligandu) jest istotnym utrudnieniem badań teoretycznych receptorów GPCR (lewy panel rysunku). Zmiana położenia jednej, istotnej reszty aminokwasowej, np. wskutek błędu w dopasowaniu sekwencyjnym celu i szablonu, może kompletnie zmienić wynik dokowania ligandu. Dlatego jeden z opracowanych przeze mnie modułów GPCRM oparty na potencjałach statystycznych umożliwia poprawienie rejonu miejsca aktywnego, m.in. przez poprawienie trafności dopasowania sekwencji celu i szablonu podczas modelowania struktury receptora (prawy panel rysunku). Rysunek na podstawie [H6] (prawy panel) oraz wyników wszystkich grup badawczych biorących udział w GPCR Dock 2013 [61] w kategorii kompleksu receptora 5-HT<sub>2B</sub> (lewy panel).

Opracowana przeze mnie procedura modelowania receptorów GPCR (GPCRM) [H5-6] [H8] oraz ich kompleksów z ligandami małowcząsteczkowymi (GUT-DOCK) [H9-10] umożliwia przeprowadzenie kompletnych badań *in silico* nad dowolnym receptorem GPCR z klasy A lub B. Na przykład w publikacji [H1] oraz [H3-4] zostały opisane badania nad mikroprzełącznikowym mechanizmem aktywacji receptorów kannabinoidowych oraz innych receptorów GPCR z klasy A wykorzystujące, jeszcze wówczas niezautomatyzowaną, opisaną powyżej procedurę. Z kolei w końcowym etapie moich dotychczasowych badań została ona wykorzystana do oceny pobocznych oddziaływań leków z receptorami GPCR poza ich głównym celem molekularnym (ang. off-target) [H9-10]. W tym przypadku zbadano mechanizmy molekularne reakcji polekowych dotyczących wystąpienia zaburzeń metabolizmu glukozy prowadzących do cukrzycy typu II.

Opisana powyżej metodologia dotycząca badań nad receptorami GPCR została uzupełniona o symulacje dynamiki molekularnej, które z jednej strony stanowiły walidację otrzymanych modeli homologicznych, a z drugiej strony pozwoliły na obserwację działania wybranych białek. W badaniach opisanych w publikacji [H2] zastosowano symulacje pełnoatomowej dynamiki molekularnej (CHARMM [48]) do poprawienia dokładności modelu homologicznego rotora V-ATPazy (w wersji dzikiej oraz mutantu), a następnie do badania stabilności kompleksów V-ATPazy z pochodnymi archazolidu A w celu wytypowania najsilniej

wiążącego się związku (YASARA [63], AMBER[49]). W publikacjach [H9-10] opisałam użycie dynamiki molekularnej (AMBER) do wygenerowania zbioru zbliżonych konformacji receptorów glukagonowych (ang. conformational ensemble), które wykazały większą skuteczność w wirtualnym skriningu niż startowe struktury krystaliczne GCGR i GLP1R, szczególnie w przypadku receptora GLP1R [H9].

W przypadku homotrimeru transportera koncentrującego nukleozydy CNT3 (ang. concentrative nucleoside transporter) z rodziny SLC28 zastosowałam dynamikę molekularną do w poprawienia modelu (ang. refinement), co opisałam w publikacji [H7]. Dynamika molekularna poprawiła wzajemne dopasowanie podjednostek białkowych w homotrimerze oraz posłużyła jako narzędzie do oceny jakości modelu podjednostki, który został zbudowany metodą hybrydową (modelowanie przez homologię połączone z *de novo*). W przypadku transportera CNT3 informacja ewolucyjna została użyta w kilku aspektach. Przede wszystkim, struktura krystaliczna vCNT, bakteryjnego homologa hCNT3, umożliwiła zbudowanie znacznej części samej podjednostki oraz całego homotrimeru hCNT3 metodą modelowania porównawczego. Ponadto, wykorzystując potencjały statystyczne wyprowadzone na podstawie znanych struktur białek błonowych i zaimplementowane w narzędziu Rosetta Broker [57] oraz jednowymiarową predykcję regionów transbłonowych (TOPCONS [64]) zbudowałam brakujący, N-terminalny fragment hCNT3. Dzięki temu uzyskałam pełny i funkcjonalny (w symulacji dynamiki molekularnej) model homotrimeru CNT3. Transporter CNT3 posłużył mi w tym przypadku za układ modelowy do opracowania procedury przewidywania *de novo* struktury białek transbłonowych uwzględniającej informację ewolucyjną.

#### IV. Szczegółowe omówienie osiągnięcia i jego wykorzystania

##### ***GPCRM — aplikacja internetowa ukierunkowana na strukturę receptora GPCR***

Procedura modelowania przez homologię receptorów GPCR, opracowana i opisana przeze mnie po raz pierwszy w publikacji [H1], została w pełni zautomatyzowana w postaci ogólnie dostępnej aplikacji internetowej GPCRM, opisanej przeze mnie w publikacji [H5] oraz w późniejszych [H6] i [H8]. Aplikacja jest przeznaczona do generowania modeli receptorów GPCR nie tylko z klasy A, ale także z innych klas (np. B [H9-10]) oraz również w przypadku niskiego podobieństwa między sekwencją celu a szablonu (identyczność sekwencji poniżej 30%), co opisałam w publikacji [H8]. Budowanie modelu homologicznego receptora GPCR rozpoczyna się w niej od wyboru odpowiednich szablonów na podstawie dostarczonej przez użytkownika sekwencji celu. Na podstawie dopasowania sekwencji celu i dostępnych w bazie

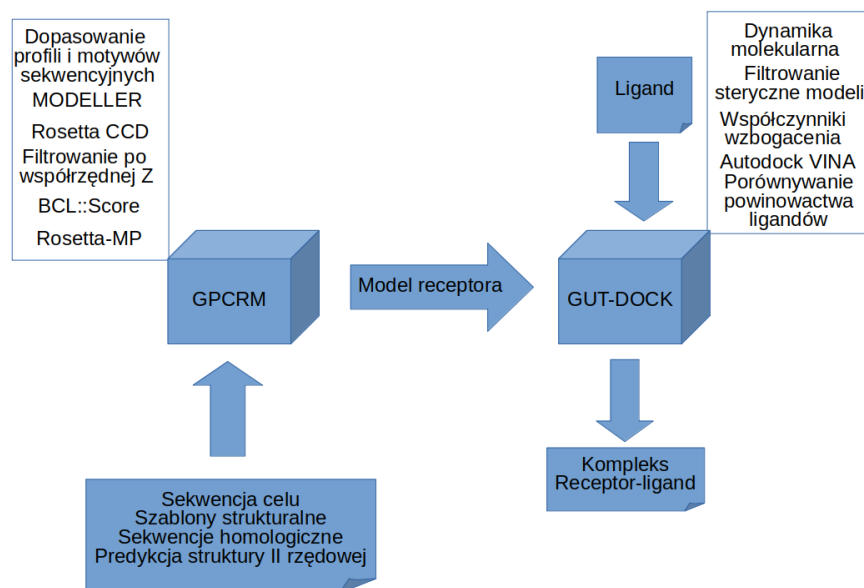
danych aplikacji sekwencji szablonów za pomocą narzędzia CLUSTALW2 [65] i obliczonych miar podobieństwa sekwencyjnego (identyczność sekwencji i podobieństwo sekwencji) wybierane są najbardziej podobne szablony. W przypadku trybu modelowania dla zaawansowanych użytkowników ('Advanced Mode'), opracowanego i opisanego przeze mnie w publikacji [H5], jest możliwość wybrania dowolnej liczby szablonów, biorąc pod uwagę nie tylko samo podobieństwo sekwencyjne, ale również profile aktywności biologicznej receptora i szablону. W przypadku automatycznego trybu modelowania, jeśli identyczność sekwencji przekracza 40%, wybierany jest tylko jeden szablon. W innym przypadku wybierane są dwa, najbardziej podobne sekwencyjnie szablony. Budowanie modelu receptora GPCR przy użyciu kilku szablonów strukturalnych pozwala na uniknięcie błędów związanych z wyborem pojedynczego niewłaściwego szablónu, ponieważ końcowy model nie zależy od tego jednego szablónu, ale jest średnią ważoną wszystkich użytych w modelowaniu szablonów. W rezultacie końcowy model o najniższej energii może być np. sumą elementów, które są wybrane jako najlepsze dla danego fragmentu sekwencji, pochodzących z różnych szablonów, co opisałam w publikacji [H5] oraz [H6] (receptor SMO), a co zostało wcześniej opisane w [66].

W kolejnym kroku modelowania, za pomocą narzędzia BLAST [59], generowany jest profil sekwencyjny dla celu, który jest następnie porównywany z profilami sekwencyjnymi szablonów (ang. profile-profile alignment) za pomocą MUSCLE [67] (algorytm dopasowania progresywnego). Wygenerowane w ten sposób dopasowanie sekwencyjne okazało się znacznie trafniejsze w przewidywaniu przerw w dopasowywanych sekwencjach niż odpowiadające mu proste dopasowanie wielu sekwencji (ang. multiple sequence alignment) lub proste dopasowania dwóch sekwencji (ang. pairwise sequence alignment) w przypadku receptorów GPCR co wykazałam w badaniach opisanych w [H5]. Wygenerowane dopasowanie sekwencji jest następnie optymalizowane w celu poprawnego dopasowania do siebie typowych dla receptorów GPCR motywów sekwencyjnych [H4], które są silnie zachowywane w trakcie ewolucji. Typowe dla receptorów GPCR różnych klas motywy sekwencyjne opisałam szczegółowo m.in. w publikacji [H4], uwzględniając ich rolę w procesie aktywacji tych receptorów. Dopasowane sekwencje aminokwasowe celu i szablonów są w kolejnym kroku użyte do zbudowania modelu za pomocą narzędzia MODELLER [41] w sposób opisany powyżej, w procedurze ważonego uśredniania wielu szablonów strukturalnych. Pętle między helisami transbłonowymi w modelu podlegają dodatkowemu modelowaniu za pomocą dwóch metod: MODELLER DOPE oraz iteracyjnej CCD (Rosetta, metoda spadku względem współrzędnych cyklicznych). Druga z metod polega na budowaniu pętli z krótkich fragmentów aminokwasowych (ang. fragment assembly) i jest przydatna w przypadku dłuższych pętli

(powyżej 20 aminokwasów), np. pętli zewnątrzkomórkowej EC2 (ang. extracellular loop 2), przy ograniczonych zasobach obliczeniowych aplikacji internetowej. W przypadku krótszych pętli dokładniejsze przeszukiwanie przestrzeni konformacyjnej metodą *ab initio* (np. MODELLER DOPE), czyli budowanie pętli od nowa, może dawać lepsze rezultaty, o czym wspomniałam w publikacji [H5] i co opisano dokładniej w [68]. Pętla zewnątrzkomórkowa EC2 między helisą V a VI w dużej części receptorów GPCR zawiera mostek dwusiarczkowy, który dzieli tę pętlę na dwie części podlegające osobnemu modelowaniu w opracowanym przeze mnie module w GPCRM [H5]. Dodatkowo wprowadziłam również potencjał minimalizujący przenikanie przez błonę pętli zewnątrz i wewnątrzkomórkowych oraz N i C-końców białka.

Wygenerowany model białka podlega w ostatnim kroku ocenie za pomocą trzech funkcji oceniających (Rosetta Total Score, Rosetta-MP, BCL::Score), co opisałam w publikacji [H8]. Ze względu na udział potencjałów statystycznych wyprowadzonych w oparciu o rozkład Boltzmann na podstawie struktur tylko białek błonowych zdeponowanych w PDB, dwie ostatnie funkcje dają znacznie lepsze wyniki w modelowaniu receptorów GPCR niż pierwsza, co wykazałam w publikacji [H6].

Oprócz warstwy obliczeniowej aplikacja GPCRM dostarcza użytkownikowi wyspecjalizowany interfejs, pozwalający na manualną modyfikację dopasowań sekwencyjnych oraz trójwymiarową wizualizację wygenerowanego modelu białka. Aplikacja GPCRM została wykorzystana z sukcesem w dwóch konkursach obliczeniowych o zasięgu światowym: GPCR Dock 2010 [B6] oraz GPCR Dock 2013 [61] [B1]. Stała się również podstawą do badań nad mechanizmami aktywacji receptorów GPCR, które opisałam w publikacjach [H1] oraz [H3-4], a także w innych publikacjach mojego współautorstwa (patrz pkt 5 niniejszego autoreferatu). Aplikacja GPCRM stała się również punktem wyjścia dla kolejnej aplikacji internetowej dotyczącej oddziaływań receptorów GPCR z ligandami (GUT-DOCK), którą opisałam w publikacjach: [H9-10]. Liczba cytowań oryginalnej publikacji opisującej algorytm GPCRM wynosi 46 (według Google Scholar), liczba użytkowników na dzień 22.03.2019 wynosi 7788 z 64 krajów, (odpowiadające dane dla aplikacji GUT-DOCK, opublikowanej w styczniu 2019r.: 164 użytkowników z 5 krajów). Wśród publikacji cytujących [H5], znalazła się m.in. publikacja grupy badawczej J. Selent na temat nowych agonistów 5-HT<sub>2A</sub> [69]. Powyższe dane świadczą o zasięgu stosowania opracowanej przeze mnie metody GPCRM, pomimo bardzo małej liczby stałych kontrybutorów (kierownik grupy badawczej S. Filipek, D. Latek oraz programista P. Pasznik). Dopiero przy okazji ostatniej publikacji [H8] z 2018 r. ich liczba znacząco się zwiększyła przekraczając trzy osoby.



**Rysunek 4.** Uproszczony schemat aplikacji internetowych GPCRМ i GUT-DOCK ukierunkowanych odpowiednio na strukturę i na funkcję receptorów GPCR.

Komentarza wymaga opracowanie aplikacji GPCRМ na tle innych działających wcześniej i później serwisów internetowych. GPCRМ został upubliczniony już w czerwcu 2012 jako wersja beta (GUT-DOCK w grudniu 2017). W 2012 roku w dziedzinie automatycznego modelowania receptorów GPCR były udostępnione: baza danych GPCRDB [70], serwis internetowy (aplikacja internetowa) GPCR-SSFE [71] oraz serwis internetowy GPCR-ModSim [72]. Dlatego badacze korzystali również z serwisów do modelowania wszystkich białek (również globularnych), np. Robetty (wersja automatyczna Rosetty), I-TASSER-a [73] lub SWISS-MODEL-a. Wymienione metody zostały porównane przeze mnie do GPCRМ w suplemencie do publikacji [H5]. W ówczesnych serwisach internetowych do modelowania tylko receptorów GPCR stosowany był MODELLER, a nowością od strony metodologicznej było budowanie modelu receptora z helis transbłonowych (ekstrahowanych w całości) pochodzących z różnych szablonów, ale bez ich uśredniania, a jedynie na zasadzie kopiowania współrzędnych i minimalizacji energii oddziaływań zbudowanego w ten sposób modelu o siedmiohelikalnej topologii (GPCR-SSFE [71]). Podobna metoda w dziedzinie modelowania receptorów GPCR, również opierająca się na minimalizacji oddziaływań między helisami, ale należąca do klasy metod *de novo*, została opublikowana w 2012r. (BiHelix [74]). Polega ona na budowaniu modelu receptora z siedmiu typowych helis transbłonowych, a następnie na ich obrocie (każdej z osobno) w celu minimalizacji energii wzajemnych oddziaływań. Wymienione powyżej metody różnią się zatem w sposób zasadniczy od aplikacji

GPCRM, która jest jednym z elementów opisanego w niniejszym autoreferacie osiągnięcia. Późniejsze metody (głównie kolejne wersje GPCRDB [75] i GPCR-SSFE [76] oraz dostosowanie I-TASSER-a do receptorów GPCR [77]), w dużym stopniu zaczynają już uwzględniać koncepcję uśrednienia wielu szablonów w modelowaniu receptorów GPCR, opublikowaną przeze mnie w 2013r. [H5]. Również od strony technicznej, pierwsza wersja GPCRM została napisana przy zastosowaniu Biopythona i Django [H5]. Te technologie zostały włączone, np. do bazy danych GPCRDB, ale dopiero po roku 2013r., a zatem GPCRM w momencie pierwszej publikacji był nowością nie tylko pod kątem algorytmów obliczeniowych, ale również pod kątem technologicznym. Obecnie włączenie potencjałów statystycznych wyprowadzonych na podstawie struktur białek błonowych z PDB (BCL::Score, Rosetta-MP) do oceny modeli receptorów GPCR nie jest realizowane ani w GPCRDB, ani w GPCR-SSFE, a jedynie w GPCR-I-TASSER [77], ale na zupełnie innej zasadzie (jest to z założenia program do przewlekania, ang. threading).

#### ***GUT-DOCK – aplikacja internetowa ukierunkowana na funkcję receptora GPCR***

Opisana powyżej procedura modelowania receptorów GPCR znalazła zastosowanie w drugiej, opublikowanej niedawno w [H9-10] aplikacji internetowej GUT-DOCK dotyczącej receptorów GPCR wiążących hormony jelitowe (ang. gut hormone receptors). Poprzednia aplikacja internetowa GPCRM była dla niej punktem wyjścia, tzn. posłużyła do wygenerowania modeli homologicznych wybranych receptorów GPCR. W tym przypadku wybrano tylko te modele receptorów, które dawały najlepsze rezultaty w wirtualnym skriningu. Centralną częścią aplikacji GUT-DOCK jest giętkie dokowanie małowzrostkowych ligandów do sztywnych struktur receptorów GPCR realizowane przez równoległy algorytm Autodock VINA [62]. Sposób wiązania liganda do receptora oceniony jako najkorzystniejszy energetycznie jest prezentowany użytkownikowi jako wynik, w postaci plików PDB oraz wizualizacji oddziaływań ligand-receptor (Ligplot). Na podstawie moich badań nad molekularną podstawą zaburzeń metabolizmu glukozy spowodowanych przez farmakoterapię opisanych w publikacjach [H9-10] oceniłam, że przydatne będzie również porównanie otrzymanej w dokowaniu teoretycznej energii wiązania z podobnie otrzymanymi wartościami dla wybranej klasy leków ( $\beta$ -adrenolityki). W ten sposób umożliwiłam użytkownikowi projektowanie nowych, aktywnych farmakologicznie substancji o potencjalnie mniejszej szkodliwości dla gospodarki węglowodanowej niż dotychczas stosowane leki. Pomimo oczywistych ograniczeń wynikających z konieczności użycia w aplikacji narzędzi dostępnych pod licencją open-source (Autodock VINA, OpenBabel), wynik jej działania okazał się porównywalny z możliwościami

programów licencyjnych (Glide, Schrodinger LLC), co opisałam w publikacji [H9]. Potwierdza to skuteczność opracowanej przeze mnie procedury modelowania kompleksów receptorów GPCR.

### *Charakterystyka strukturalna wybranych białek transbłonowych*

W 2012 roku, w momencie publikacji pracy [H2] była znana tylko struktura krystaliczna bakteryjnego homologa wakuolarniej (typ V) V-ATPazy opublikowana w 2005 roku [78]. Model homologiczny V-ATPazy wyizolowanej z drożdży, który zbudowałam na jego podstawie, zawierał dwie istotne dla mechanizmu działania rotora i równocześnie dla wiązania inhibitora - archazolidu A (i jego pochodnych) reszty: Glu137 i Tyr66, powtórzone (m.in. Glu108 i Tyr142 – patrz Rys. 1) w każdej podjednostce części transbłonowej rotora. Rotamery tyrozyny w pierścieniu rotora w moim modelu homologicznym zostały całkowicie potwierdzone przez obie struktury eksperymentalne V-ATPazy z roku 2015 [79] i 2018 [80], rozwiązane metodą kriomikroskopii elektronowej. Grupa boczna kwasu glutaminowego (m.in. Glu137) w moim początkowym modelu homologicznym była skierowana w stronę cytoplazmatyczną, jednak w trakcie symulacji pełnoatomowej dynamiki molekularnej, co opisano szczegółowo w suplemencie publikacji [H2], zmieniła konformację kierując się w stronę lumen, w odległości paru Å od sąsiadującej tyrozyny (m.in. Tyr66), tak jak jest to przedstawione na Rys. 1. Inhibitor – archazolid A wiązał się do pierścienia rotora, w której reszta Glu137 była w konformacji odpowiadającej właśnie temu drugiemu rotamerowi, co zostało przeze mnie przedstawione w publikacji [H2]. Struktura eksperymentalna (cryo-EM) rotora V-ATPazy opublikowana w 2015, opisująca trzy stany konformacyjne tego białka, zawierała ten drugi rotamer Glu tylko w kilku podjednostkach pierścienia i tylko w jednym z trzech opisanych stanów konformacyjnych (kod PDB: 3J9V), który faktycznie jest związany z nieco mniejszą stechiometrią przenoszonych protonów względem białka (3:1 zamiast 4:1), a więc z mniejszą aktywnością enzymu. Jednak struktura z 2018, zawierała już reszty kwasu glutaminowego w pierścieniu w dokładnie w tej samej konformacji otwartej, która jest związana z autoinhibicją tego enzymu [80], jaką przewidziałam na podstawie symulacji dynamiki molekularnej. Ten wybrany przykład ilustruje dokładność przeprowadzonych przeze mnie i opisanych w publikacji [H2] obliczeń.

Struktura krystaliczna receptorów kannabinoidowych CB<sub>1</sub> i CB<sub>2</sub> ukazała się dopiero w roku CB<sub>1</sub>: 2016 (konformacja nieaktywna) i 2017 (aktywna) [81] oraz CB<sub>2</sub>: w 2019 (nieaktywna) [82], choć już wyniki badań eksperymentalnych z początku wieku świadczyły o istnieniu szczególnego mikroprzełącznika aminokwasowego [83]. W pracy [H1] opisałam

przede wszystkim rejon miejsca aktywnego, zawierający mikroprzełącznik Trp/Phe, którego ustawienie może być modulowane przez ligandy. Modele receptorów zostały zbudowane jeszcze przed opracowaniem aplikacji GPCRM przy wykorzystaniu standardowej procedury modelowania przez homologię (MODELLER) i szablonu strukturalnego receptora adenozyнового. Późniejsze publikacje na temat receptorów kannabinoidowych [84], cytujące [H1], już korzystające z metody budowania modelu z wielu szablonów oraz z później otrzymanych danych z mutagenyzy [85], potwierdzają rolę tego mikroprzełącznika i jego ustawienie w formie nieaktywnej (Trp6.48 – rotamer gauche+, czyli  $-60^\circ$  oraz Phe3.36 trans, czyli  $-180^\circ$ ) i aktywnej (Trp6.48 rotamer trans oraz Phe3.36 gauche+) receptora. Struktury CB<sub>1</sub> z 2016 i 2017 roku (np. 5TGZ – nieaktywna oraz 5XR8 – aktywna) pokazują te same ustawienia mikroprzełącznika Trp/Phe w obecności antagonisty i agonisty, które pokazałam 5 lat wcześniej w publikacji [H1]. Między innymi opublikowana w tym roku praca R. J. Doerksena [86], cytująca moją publikację [H1], podkreśla rolę pokazanego przeze mnie 8 lat temu mikroprzełącznika w receptorze CB<sub>2</sub>. Niemniej jednak struktura krystaliczna CB<sub>2</sub> wiążącego antagonistę z 2019 (kod PDB 5ZTY) [82] pokazuje ustawienie mikroprzełącznika zaskakująco przypominające CB<sub>1</sub> wiążącego agonistę. W mojej publikacji [H1], która w dużym stopniu opierała się tylko na dokowaniu, a nie na dynamice molekularnej, nie stwierdziłam jednoznacznie tej różnicy między ustawieniem mikroprzełącznika w CB<sub>1</sub> i CB<sub>2</sub>. Niemniej jednak, w wynikach, które umieściłam razem ze współpracownikami w [H1], były pewne przesłanki mogące wskazywać na różnice w profilu funkcjonalnym CB<sub>1</sub> i CB<sub>2</sub>. Mianowicie, w suplemencie publikacji [H1] (Fig. S2) przedstawione są wyniki symulacji dynamiki molekularnej dla obu receptorów w postaci wykresów kątów  $\chi_1$  dla dwóch najlepszych modeli receptorów (a) i (b), osobno dla obu reszt mikroprzełącznika - Trp i Phe. W przypadku CB<sub>1</sub> wartości obu kątów definiujące rotamery grup bocznych aminokwasów są w przybliżeniu stałe przez całą symulację (z chwilowym przeskokiem w przypadku drugiego modelu). Natomiast w przypadku CB<sub>2</sub>, w trakcie symulacji z udziałem drugiego modelu receptora (d), można było zaobserwować wyraźną zmianę kąta  $\chi_1$  dla Phe i Trp o około 120 stopni, czyli spontaniczną zmianę ustawienia mikroprzełącznika do pozycji aktywnej. Były to jednak zbyt krótkie symulacje (1ns), abym mogła na ich podstawie sformułować dodatkowe wnioski o różnicach między CB<sub>1</sub> a CB<sub>2</sub>, tak jak zostało to opisane w publikacji [82].

Przykład receptorów kannabinoidowych pokazuje problem badawczy dziedziny receptorów GPCR. Każda nowa struktura receptora GPCR ukazująca się w PDB (obecnie już kilka receptorów na rok) umożliwia coraz dokładniejsze badanie mechanizmów ich działania w porównaniu do opublikowanych badań teoretycznych bazujących na modelach



homologicznych zbudowanych na wcześniejszych strukturach krystalicznych. Ten problem pojawił się w moich badaniach przy okazji receptorów glukagonowych [H9-10]. Model receptora GIPR został przeze mnie zbudowany na podstawie struktury apo receptora glukagonowego GCGR z 2013 i 2016 roku. Późniejsze struktury holo receptora GCGR z 2017 lub struktury holo receptora GLP1R z 2017 (drugiego bliskiego homologa sekwencyjnego GIPR), obie z antagonistami w allosterycznych miejscach aktywnych, różnią się od wcześniejszych struktur, choć nieznacznie pomimo związania ligandu (odpowiednio: 1.2Å i 0.8Å), głównie w rejonie fragmentu łączącego domenę zewnątrzkomórkową z domeną transbłonową [87]. W moich badaniach opisanych w publikacjach [H9-10], nie był badany rejon struktury między dwiema domenami. Zatem modelowanie oparte na nowych strukturach z 2017 r. nie wniosłoby żadnej istotnej zmiany do badań nad receptorem GIPR. Model receptora GIPR, podobnie jak modele dwóch pozostałych receptorów klasy B: PAC1R oraz VIPR1, oraz struktury krystaliczne GCGR i GLP1R (po symulacjach dynamiki molekularnej) zostały przeze mnie umieszczone w ogólnie dostępnej aplikacji GUT-DOCK.

Również w przypadku transportera CNT3, moje badania opisane w publikacji [H7] zostały przeprowadzone na podstawie wcześniejszej struktury z 2012 roku (kod PDB 3TIJ). Niemniej jednak nowa struktura (a właściwie seria struktur z różnymi ligandami) z 2014 [88], została otrzymana przy wykorzystaniu wcześniejszej struktury jako inputu do programu PHENIX. Wszystkie obserwacje dotyczące wcześniejszej struktury zostały potwierdzone przez nową, a różnice obserwowane między nimi obejmują jedynie drobne, lokalne zmiany konformacji wynikające z oddziaływania transportera z różnego rodzaju ligandami nukleozydowymi m.in. z analogami cytydyny. Jednak publikacja [H7] dotyczy przede wszystkim struktury całego białka, zatem szablon strukturalny vCNT z 2012 r. mógł zostać użyty do budowy homotrimeru CNT3 za pomocą opracowanej przeze mnie procedury modelowania białek transbłonowych. Procedura modelowania transportera CNT3, opisana w [H7], polega na połączeniu modelowania przez homologię uwzględniającego zachowywane w ewolucji ligandy (MODELLER) oraz modelowania *de novo* helis transbłonowych (TOPCONS oraz Rosetta Broker). Transporter CNT3 jest reprezentantem rodziny białek transbłonowych SLC28, które stanowią znacznie bardziej skomplikowany układ biologiczny niż receptory sprzężone z białkiem G ze względu na różnorodność topologii oraz różnorodność mechanizmów działania. W przypadku receptorów GPCR dotychczas poznany mechanizm aktywacji polega na lokalnej zmianie konformacyjnej w miejscu aktywnym po związaniu ligandu, która indukuje globalną zmianę konformacji (głównie helisy VI) jednak przy zachowaniu tej samej topologii. W przypadku transporterów SLC przekazanie ligandu ze

środowiska zewnątrzkomórkowego do wewnątrz wymaga tak istotnej zmiany konformacyjnej białka, że obie konformacje (np.: otwarta na zewnątrz i do wewnątrz komórki) zupełnie różnią się topologią zachowując jedynie helikalną postać białka. Ze względu na niedostateczny zbiór krystalicznych struktur transporterów w konformacji otwartej i zamkniętej nie jest obecnie możliwe wyprowadzenie dla nich dalekozasięgowych potencjałów statystycznych na podstawie informacji ewolucyjnej i strukturalnej zawartej w bazie PDB. Do niedawna nie było również możliwe zbudowanie modelu homologicznego transportera np. w konformacji otwartej na zewnątrz na podstawie szablonu strukturalnego konformacji otwartej do wewnątrz komórki. Badania L. R. Forrest [89] wykazały jednak możliwość obejścia problemu kompletnej zmiany topologii białka w procesie transportu ligandu za pomocą wykorzystania częściowej symetrii między obiema konformacjami transportera. Jest to koncepcja odwróconej topologii (ang. *inverted topology*). Rozwiązanie zaproponowane przez L. R. Forrest w oczywisty sposób nie uwzględnia jednak sytuacji, w której nie ma danych krystalograficznych dla żadnej z konformacji dla całego białka lub dla jego fragmentu, jak to miało miejsce w przypadku hCNT3, którego bliski homolog w PDB (vCNT) nie zawierał N-terminalnego fragmentu sekwencji.

W publikacji [H7] zaproponowałam procedurę modelowania fragmentów transporterów SLC, które nie są dostępne w postaci danych krystalograficznych. Rozwiązanie to oparte jest na budowaniu modelu białka *de novo* przy wykorzystaniu krótkich fragmentów białek błonowych z PDB (Rosetta) oraz protokołu modelowania dużych systemów białkowych, w którym podjednostka jest traktowana jako bryła sztywna (Broker) [90]. W ten sposób udało mi się rozwiązać strukturę ponad 100-aminokwasowego N-terminalnego fragmentu transportera hCNT3, który zawiera istotny dla funkcji tego białka polimorfizm pojedynczego nukleotydu (SNP). Modelowanie zostało przeprowadzone przy uwzględnieniu obecności (a więc również oddziaływań) całego homotrimeru. Model homotrimeru został wcześniej przygotowany przeze mnie na podstawie struktury bakteryjnego homologa vCNT przy uwzględnieniu podczas modelowania zachowanych w ewolucji ligandów (kationu sodu i urydyny). Walidacja modelu transportera hCNT3 została przeprowadzona za pomocą 100 ns symulacji dynamiki molekularnej [H7]. Przeprowadzona przeze mnie symulacja dynamiki molekularnej pozwoliła na obserwację: interfejsu podjednostek homotrimeru, miejsca aktywnego hCNT3 i sąsiadujących helis (wpływ jonu  $\text{Na}^+$  na konformację białka w trakcie symportu) oraz ruchliwości aminokwasów związanych z SNP [H7].

## ***Badania nad mechanizmem aktywacji receptorów GPCR***

Wygenerowane za pomocą GPCRM modele posłużyły do przeprowadzenia przeze mnie badań nad mechanizmem aktywacji receptorów GPCR opisanych w publikacjach: [H1] [H3-4] oraz w innych publikacjach wymienionych w punkcie 5 niniejszego autoreferatu. W [H1] opisałam istnienie mikroprzełącznika aminokwasowego Trp/Phe, którego zmiana konformacyjna jest uważana za początek lokalnych zmian konformacyjnych receptorów kannabinoidowych powodujących w kolejnych etapach globalne zmiany konformacyjne prowadzące do w pełni funkcjonalnych w oddziaływaniu z białkiem G konformacji tych receptorów. Moja publikacja uzupełniła wcześniejsze badania eksperymentalne tych receptorów [83, 91]. Wzrost liczby struktur receptorów GPCR w PDB umożliwił w kolejnych latach po mojej publikacji [H5] budowanie przy użyciu GPCRM coraz dokładniejszych modeli receptorów kannabinoidowych pozwalających, np. na badania drogi wejścia ligandu do wnętrza receptora [84].

W wysokocytowanej publikacji przeglądowej [H4] opisałam podrodziny GPCR pod kątem ich zachowywanych w ewolucji motywów sekwencyjnych i związanych z nimi mikroprzełączników aminokwasowych, których zmiana położenia, np. wskutek oddziaływania z agonistą, powoduje obniżenie energii wewnętrznej białka i w rezultacie zwiększenie populacji stanów aktywnych receptora [20]. W tabeli 1 przedstawiłam reprezentatywny dla publikacji [H4] opis samej tylko podrodziny receptorów podobnych do rodopsyny pod kątem mechanizmów aktywacji.

**Tabela 1.** Mikroprzełączniki aminokwasowe receptorów GPCR z klasy A, opisane przeze mnie w publikacjach [H1] oraz [H4].

<b>Etap aktywacji<sup>1</sup></b>	<b>Mikroprzełączniki i odpowiadające im motywy sekwencyjne</b>	<b>Indukowana globalna zmiana konformacyjna receptora</b>
1	Mikroprzełącznik - zamek 3-7 (ang. 3-7 lock switch) Glu3.28 (Asp3.32), Lys7.43 (Tyr7.43)	Zerwanie połączenia między helisą III a VII poprzedzone zerwaniem mostku solnego między Lys7.43 a zasadą Schiffa tworzoną przez retinal i Glu3.28.
2	Mikroprzełącznik transmisji (poprzednia nazwa: dźwigniowy [H4]) (ang. transmission switch) Trp6.48, P5.50, P6.50 CwxP (helisa VI)	Helisa VI przesuwa się w stronę helisy V i ulega niewielkiej rotacji, a następnie jej część cytoplazmatyczna ulega odchyleniu.
3	Mikroprzełącznik dźwigniowy (ang. toggle switch) Tyr7.53 nPxxxy (helisa VII)	Przegrupowanie w cytoplazmatycznym końcu helisy VII poprzedzone rotacją grupy bocznej tyrozyny w celu reorganizacji sieci wiązań wodorowych wewnątrz receptora umożliwiających w następstwie oddziaływanie z białkiem G.

4	Mikroprzełącznik - zamek jonowy (ang. ionic lock switch) Glu3.49/Arg3.50 – Glu6.30/Thr6.34 (d/e)Ry (helisa III)	Zerwanie mostku solnego argininy i kwasu glutaminowego, po którym następuje zbliżenie helisy III z V oraz VI z V.
---	--	---

<sup>1</sup>Kolejność etapów aktywacji receptorów klasy A uszeregowana wg Wescott et al. (PNAS 2016, 113(35): 9928-9933), od zmian konformacyjnych w obrębie części zewnątrzkomórkowej po związaniu ligandu, przez środek receptora, kończąc na części cytoplazmatycznej oddziałującej z białkiem G.

### ***Projektowanie inhibitorów białek transbłonowych***

Jak wspomniałam powyżej, GUT-DOCK opisany przeze mnie w publikacjach [H9-10] może być również wykorzystany w projektowaniu nowych leków. Oparłam się w tym przypadku na testach autorów programu Autodock VINA (liczba cytowań od 2010 r. - ponad 7600, wg Google Scholar) zależności między aproksymowaną teoretyczną energią wiązania a eksperymentalnie potwierdzonym faktem wiązania danego liganda przez receptor. Jak opisałam w [H9-10], cytując również inne publikacje, efektywność niskokosztowego wirtualnego skryningu od wielu lat dorównuje, a nawet przewyższa poziom sukcesu skryningu eksperymentalnego. Niemniej jednak ostateczne potwierdzenie biologicznej aktywności danego liganda w modulacji ścieżki sygnałowej receptora wymaga szeregu badań biochemicznych, a nawet klinicznych [30]. Sprawę komplikuje dodatkowo możliwość wystąpienia pobocznych interakcji leku z innymi celami biologicznymi w komórce (ang. off-target), których efekty są obserwowane dopiero na przestrzeni lat w badaniach klinicznych. Część z takich pobocznych interakcji można wykorzystać w repozycjonowaniu leków (ang. drug repositioning), dzięki któremu można uniknąć kosztownych badań nad toksycznością nowowprowadzanego leku. W publikacji [H10] zaproponowałam nowy zestaw związków do stosowania w leczeniu cukrzycy właśnie na zasadzie repozycjonowania. Związki przeze mnie zaproponowane, stosowane już w leczeniu innych chorób, mogą wykazywać efekt terapeutyczny dzięki wpływowi na szlak sygnałowy hormonów inkretynowych regulujących gospodarkę węglowodanową. Przewidywany efekt zaproponowanych związków może dotyczyć zarówno hamowania aktywności receptora glukagonowego GCGR (zmniejszenie produkcji glukozy), jak i aktywacji receptorów GLP1R i GIPR (zwiększenie produkcji insuliny). Rozróżnienie obu efektów wymaga jednak dalszych badań teoretycznych i eksperymentalnych, co opisałam w publikacji [H10].

Efekt hamujący działanie białka transbłonowego obserwowałam również na przykładzie kompleksu rotora V-ATP-azy z pochodnymi archazolidu A wykazującymi aktywność przeciwnowotworową [92], co opisałam w publikacji [H2]. W tym przypadku cykliczny ligand działa cytotoksycznie na komórki nowotworowe, zatrzymując rotację domeny

transbłonowej odpowiedzialnej za wystąpienie gradientu stężenia  $H^+$  (Rys. 1). Specyficzność adhezji inhibitora na powierzchni rotora jest w tym przypadku zapewniana nie tylko przez wiązania wodorowe, ale również przez komplementarność elektrostatyki, co zostało opisane przeze mnie w suplemencie publikacji [H2]. Mój udział dotyczył w tym przypadku badań *in silico*, czyli przygotowania modelu homologicznego domeny błonowej V-ATPazy, dokowania pochodnych archazolidu A, potwierdzenia ich sposobu wiązania za pomocą symulacji dynamiki molekularnej oraz wytypowania najsilniej wiążących się ligandów do badań biochemicznych. Wyniki teoretyczne opisane w publikacji [H2], zostały potwierdzone przez opisane w tej samej publikacji wyniki doświadczalne (patrz Rys. 1). Badania nad aktywnością przeciwnowotworową archazolidów opisaną m.in. w publikacji [H2] trwają w dalszym ciągu, właściwie nieprzerwanie od 2011 r. i wcześniej, w laboratorium D. Menche [93].

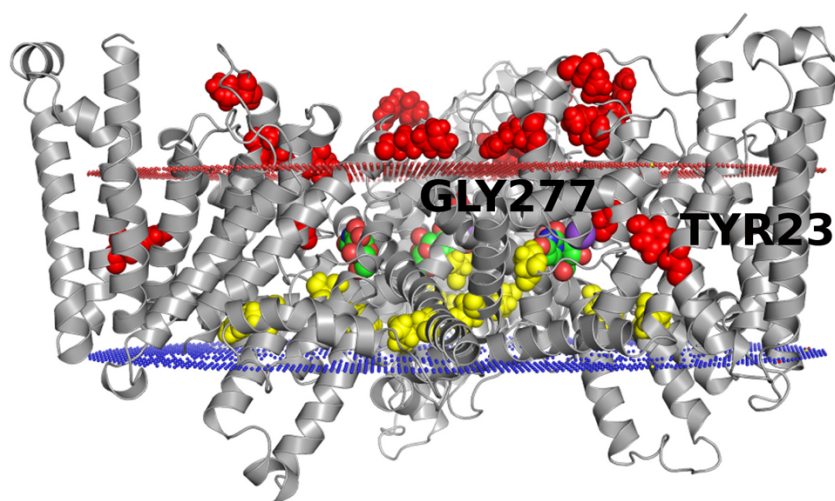
### ***Molekularne podstawy reakcji polekowych***

Jedną z funkcjonalności aplikacji GUT-DOCK polegającą na porównywaniu powinowactwa różnych związków względem wybranych receptorów GPCR, co opisałam w dwóch publikacjach [H9] oraz [H10]. W tym przypadku przebrałam główne klasy powszechnie stosowanych leków zaburzających metabolizm glukozy przy dłuższym stosowaniu u istotnej liczby pacjentów. Największe powinowactwo do receptorów glukagonowych wykazały, jak opisałam w publikacji [H10], statyny, leki  $\beta$ -adrenolityczne oraz glukokortykosteroidy, w nieco mniejszym stopniu: mineralokortykosteroidy i wprowadzone niedawno do farmakoterapii analogi tiazydów, a najsłabsze: neurosteroidy, diuretyki i tiazidy. W przypadku statyn oraz leków  $\beta$ -adrenolitycznych zauważyłam, że leki nowszej generacji w każdej z tych klas, mniej zaburzające gospodarkę węglowodanową, jednocześnie wykazują wysokie powinowactwo do receptorów glukagonowych, regulujących poziom glukozy we krwi. Podobną zależność zaobserwowałam dla mniej liczniejszego zbioru leków z różnych klas zdeponowanego w bazie efektów ubocznych leków SIDER (ang. Side Effect Resources) [36] [H10]. Na podstawie zaobserwowanej zależności zaproponowałam nowy sposób uniknięcia zaburzeń metabolizmu glukozy podczas farmakoterapii, polegający na takiej optymalizacji struktury chemicznej związków wiodących, która zwiększa ich powinowactwo do receptorów glukagonowych. W ten sposób, oprócz podstawowego efektu terapeutycznego (ang. on-target) substancja lecznicza mogłaby powodować również efekt inkretynowy regulujący poziom glukozy u pacjenta na zasadzie efektu ubocznego (ang. off-target). Tego typu rozwiązania są obecnie stosowane tylko jako politerapia, np. w leczeniu nadciśnienia (jednoczesna terapia hydrochlorotiazydami i sartanami [94]). Opisane w publikacjach [H9-10] badania teoretyczne

będą w kolejnych etapach mojej pracy badawczej potwierdzone badaniami biochemicznymi i klinicznymi.

### ***Molekularne podstawy istotnych farmakologicznie polimorfizmów***

Strukturalna interpretacja rzadkich, ze względu na silną negatywną selekcję transportera CNT3, niesynonimicznych mutacji pozwoliła mi na odkrycie molekularnych podstaw różnej odpowiedzi na farmakoterapię obserwowanej w badaniach klinicznych dotyczących leczenia zakażeń wirusowych i nowotworów, co opisałam w samodzielnie opublikowanej pracy [H7]. Np. mutacja Gly↔Arg (patrz Rys. 4) w obrębie miejsca aktywnego wiążącego nukleozyd może powodować utratę kluczowego oddziaływania sąsiadującej Gln z ligandem ze względu na utworzenie wiązania wodorowego z Arg. W efekcie zdolność koncentrowania (ang. concentrative capacity) cząsteczek nukleozydu przez transporter jest zmniejszona. Maleje w związku z tym absorpcja leku nukleozydowego (np. ribawiryny) i w efekcie jego niekorzystne cytotoksyczne działanie na komórki gospodarza. Dla pacjentów z mutacją Gly↔Arg zaobserwowano np. zmniejszone prawdopodobieństwo anemii podczas leczenia ribawiryną zakażenia wirusem zapalenia wątroby typu C. Efekt biologiczny drugiego, opisanego przeze mnie w publikacji [H7], polimorfizmu (Tyr↔Cys) zlokalizowanego w obrębie fragmentu białka zbudowanego przeze mnie *de novo*, pomimo już rozpoczętych badań eksperymentalnych (Badagnani et al. 2005) pozostaje jeszcze dokładnie nieopisany.

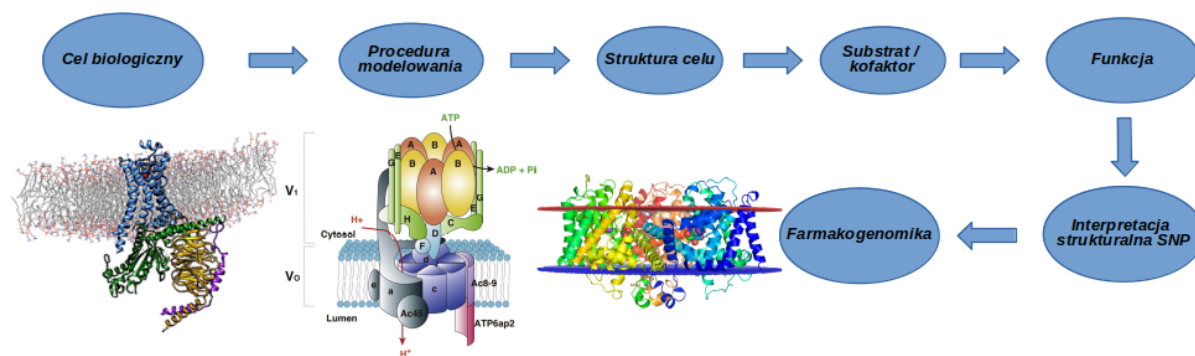


**Rysunek 4.** Polimorfizmy jednego nukleotydu (SNP) zaobserwowane dotychczas w białku hCNT3 (czerwone – odpowiadające mutacjom niesynonimicznym, żółte - synonimicznym). Na rysunku wyróżniono dwa z nich (jeden z nich – dotyczący reszty Gly277 jest w miejscu aktywnym) oraz kation sodu (fioletowy) i ligand (urydynę).

W publikacji [H7] zbadalam również kilka polimorfizmów zlokalizowanych w helisie transbłonowej VI receptora CRF<sub>1</sub>, która była modelowana *de novo*. Podobnie jak w przypadku hCNT3, są one związane z osłabieniem siły wiązania liganda z receptorem co w konsekwencji prowadzi do różnej odpowiedzi pacjentów na leczenie (np. astmy za pomocą wdychanych kortykosteroidów). W tym przypadku mutacje aminokwasów w rejonie helisy VI najprawdopodobniej powoduje osłabienie wiązania liganda z powodu utraty istotnych oddziaływań a także zmniejszenie dostępnej dla allosterycznego liganda przestrzeni z powodu zahamowania koniecznej do jego związania deformacji helisy VI, co opisałam w publikacji [H7].

### V. Podsumowanie

Przedstawione osiągnięcie naukowe dotyczące wyznaczenia struktur i mechanizmów działania wybranych białek transbłonowych przy uwzględnieniu ich różnorodności ewolucyjnej w sposób kompletny umożliwiło wyjaśnienie na poziomie molekularnym szeregu obserwowanych zjawisk biologiczno-chemicznych. Schemat przedstawiony na Rys. 5 ilustruje główne obszary przeprowadzonych badań.



**Rysunek 5.** Schemat osiągnięcia naukowego przedstawionego w niniejszym autoreferacie. W moich badaniach dokonałam kompletnego opisu struktur i mechanizmów działania reprezentantów trzech wybranych rodzin białek transbłonowych (od lewej): receptorów GPCR, V-ATPaz oraz rodziny transporterów SLC28 uwzględniając ich różnorodność ewolucyjną. Na podstawie: Clark et al. Beilstein J. Org. Chem. 2017, 13, 1071–1078, Sun-Wada et al. Biochimica et Biophysica Acta (BBA) – Bioenergetic 2015, 1847(10):1166-72 oraz [H7].

Przedstawione osiągnięcie umożliwiło również rozwój metod obliczeniowych używanych do badania struktur białek transbłonowych. Dzięki temu możliwe stało się skonstruowanie aplikacji, pozwalających na testowanie związków w sposób podobny jak jest obecnie stosowany w testach laboratoryjnych i osiągnięcie w obliczeniach poziomu dużego

zbioru danych. Szczególnym osiągnięciem moich badań było zapewnienie pewnego standardu powtarzalności niezbędnego w ogólnie dostępnych aplikacjach internetowych, a z założenia trudnego do zrealizowania w przypadku metod heurystycznych stanowiących istotną część metod obliczeniowych. Uważam, że przeprowadzone przeze mnie badania nad strukturami i mechanizmem działania białek transbłonowych były niezbędne do dokładnego zrozumienia przyczyn ich różnorodności ewolucyjnej. W moich badaniach w większości przyjął rolę obserwatora względem badanych obiektów, mimo trwających już badań np. nad ukierunkowaną ewolucją receptorów GPCR w celu modyfikacji ich funkcji [95].

## VI. Literatura

- [1] Buller AR, van Roye P, Cahn JKB, Scheele RA, Herger M, Arnold FH (2018) Directed Evolution Mimics Allosteric Activation by Stepwise Tuning of the Conformational Ensemble. *JACS* 140 (23), 7256-7266
- [2] Jones DT (2007) Improving the accuracy of transmembrane protein topology prediction using evolutionary information. *Bioinformatics* 23(5): 538–544
- [3] Kufareva I, Katritch V, Participants of GPCR Dock 2013, Stevens RC, Abagyan R (2014) Advances in GPCR modeling evaluated by the GPCR Dock 2013 assessment: meeting new challenges. *Structure* 22(8):1120-1139
- [4] César-Razquin A, Snijder B, Frappier-Brinton T, Isserlin R, Gyimesi G, Bai X, Reithmeier RA, Hepworth D, Hediger MA, Edwards AM, Superti-Furga GA (2015) Call for Systematic Research on Solute Carriers. *Cell* 62(3):478-87
- [5] Hinton A, Bond S, Forgac M (2009) V-ATPase functions in normal and disease processes. *Pflugers Arch.* 457(3):589-98
- [6] Martí-Renom MA, Stuart AC, Fiser A, Sánchez R, Melo F, Sali A. (2000) Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct* 29:291-325.
- [7] Woetzel N, Karakaş M, Staritzbichler R, Müller R, Weiner BE, Meiler J (2012) BCL::Score—Knowledge Based Energy Potentials for Ranking Protein Models Represented by Idealized Secondary Structure Elements. *PLOS ONE* 7(11): e49242.
- [8] Wang DG et al. (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 280(5366):1077-82.
- [9] Evans WE, McLeod HL (2003) Pharmacogenomics--drug disposition, drug targets, and side effects. *N Engl J Med* 348(6):538-49.



- [10] Lin L, Yee SW, Kim RB, Giacomini KM (2015) SLC transporters as therapeutic targets: emerging opportunities. *Nat Rev Drug Discov* 14(8):543-60
- [11] Denisov IG, Sligar SG (2016) Nanodiscs for Structural and Functional Studies of Membrane Proteins. *Nature Structural & Molecular Biology* 23(6), 481-486
- [12] Barth P, Schonbrun J, Baker D (2007) Toward high-resolution prediction and design of transmembrane helical protein structures. *PNAS* 104(40):15682-7
- [13] Mysinger MM, Carchia M, Irwin JJ, Shoichet BK (2012) Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J Med Chem* 55(14):6582-94
- [14] Fan H, Irwin JJ, Webb BM, Klebe G, Shoichet BK, Sali A (2009) Molecular docking screens using comparative models of proteins. *J Chem Inf Model* 49(11):2512-27
- [15] Hauser AS, Attwood MM, Rask-Andersen M, Schiöth HB, Gloriam DE (2017) Trends in GPCR drug discovery: new agents, targets and indications. *Nat Rev Drug Discov* 6(12):829-842.
- [16] Hendrickson WA (2016) Atomic-level analysis of membrane-protein structure. *Nat Struct Mol Biol* 23(6):464-7
- [17] Cheng Y (2015) Single-Particle Cryo-EM at Crystallographic Resolution. *Cell* 161(3):450-457
- [18] Rossmann MG, Blow DM (1962) The detection of sub-units within the crystallographic asymmetric unit. *Acta Cryst.* 15, 24-31.
- [19] Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M (2000) Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* 289(5480):739-45
- [20] Manglik A, Kobilka BK, Steyaert J (2017) Nanobodies to Study G Protein-Coupled Receptor Structure and Function. *Annu Rev Pharmacol Toxicol.* 57:19-37
- [21] Rasmussen SG et al. (2011) Crystal structure of the  $\beta$ 2 adrenergic receptor-Gs protein complex. *Nature* 477(7366):549-55
- [22] Fotiadis D, Liang Y, Filipek S, Saperstein DA, Engel A, Palczewski K (2003) Atomic-force microscopy: Rhodopsin dimers in native disc membranes. *Nature* 421(6919):127-8.
- [23] Robertson N, Jazayeri A, Errey J, Baig A, Hurrell E, Zhukov A, Langmead CJ, Weir M, Marshall FH (2011) The properties of thermostabilised G protein-coupled receptors (StaRs) and their use in drug discovery. *Neuropharmacology* 60(1):36-44
- [24] Liang YL et al. (2018) Cryo-EM structure of the active, Gsprotein complexed, human CGRP receptor. *Nature* 561(7724):492-497

- [25] Warne T, Moukhametzianov R, Baker JG, Nehmé R, Edwards PC, Leslie AG, Schertler GF, Tate CG (2011) The structural basis for agonist and partial agonist action on a  $\beta(1)$ -adrenergic receptor. *Nature* 469(7329):241-4.
- [26] Zhao J, Benlekbir S, Rubinstein JL (2015) Electron cryomicroscopy observation of rotational states in a eukaryotic V-ATPase. *Nature* 521(7551):241-5.
- [27] Roh SH, Stam NJ, Hryc CF, Couoh-Cardel S, Pintilie G, Chiu W, Wilkens S (2018) The 3.5-Å CryoEM Structure of Nanodisc-Reconstituted Yeast Vacuolar ATPase Vo Proton Channel. *Mol Cell* 69(6):993-1004.e3.
- [28] Johnson, Zachary Lee et al. (2014) Structural basis of nucleoside and nucleoside drug selectivity by concentrative nucleoside transporters. *eLife* vol. 3 e03604.
- [29] Hirschi M, Johnson ZL, Lee SY (2017) Visualizing multistep elevator-like transitions of a nucleoside transporter. *Nature* 545(7652):66-70.
- [30] Jacobson KA (2015) New paradigms in GPCR drug discovery. *Biochem Pharmacol* 98(4):541-55.
- [31] Merkel JS, Michaud GA, Salcius M, Schweitzer B, Predki PF (2005) Functional protein microarrays: just how functional are they? *Curr Opin Biotechnol* 16(4):447-52.
- [32] Meyer UA (2000) Pharmacogenetics and adverse drug reactions. *Lancet* 356(9242):1667-71
- [33] Yang M, Kiang M, Shang W (2015) Filtering big data from social media--Building an early warning system for adverse drug reactions. *J Biomed Inform* 54:230-40
- [34] Zhou H, Gao M, Skolnick J (2015) Comprehensive prediction of drug-protein interactions and side effects for the human proteome. *Sci Rep* 5:11090.
- [35] Obermeyer Z, Emanuel EJ (2016) Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. *N Engl J Med* 375(13):1216-9.
- [36] Kuhn M, Letunic I, Jensen LJ, Bork P (2015) The SIDER database of drugs and side effects. *Nucleic Acids Res.* doi: 10.1093/nar/gkv1075
- [37] Kolinski A (2004) Protein modeling and structure prediction with a reduced representation. *Acta Biochim Pol* 51(2):349-71
- [38] K Ginalski, A Elofsson, D Fischer, L Rychlewski (2003) 3D-Jury: a simple approach to improve protein structure predictions. *Bioinformatics* 19 (8), 1015-1018
- [39] Kurowski MA, Bujnicki JM (2003) GeneSilico protein structure prediction meta-server. *Nucleic Acids Research* 31 (13), 3305-3307
- [40] Rohl CA, Strauss CE, Misura KM, Baker D (2004) Protein structure prediction using Rosetta. *Methods Enzymol* 383:66-93

- [41] Eswar N, Webb B, Marti-Renom MA, et al. (2006) Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics*. Chapter 5:Unit-5.6.
- [42] Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 18(15):2714-23.
- [43] Hopf TA, Colwell LJ, Sheridan R, Rost B, Sander C, Marks DS (2012) Three-dimensional structures of membrane proteins from genomic sequencing. *Cell* 22;149(7):1607-21
- [44] Nugent T, Jones DT (2012) Accurate de novo structure prediction of large transmembrane protein domains using fragment-assembly and correlated mutation analysis. *PNAS* 109(24):E1540-7.
- [45] Khafizov K, Staritzbichler R, Stamm M, Forrest LR (2010) A Study of the Evolution of Inverted-Topology Repeats from LeuT-Fold Transporters Using AlignMe. *Biochemistry* 49 (50), 10702–10713
- [46] Woetzel N, Karakaş M, Staritzbichler R, Müller R, Weiner BE, Meiler J (2012) BCL::Score--knowledge based energy potentials for ranking protein models represented by idealized secondary structure elements. *PLOS One* 7(11):e49242
- [47] Yarov Yarovoy V, Schonbrun J, Baker D (2006) Multipass membrane protein structure prediction using Rosetta. *Proteins: Structure, Function, and Bioinformatics* 62 (4), 1010-1025
- [48] Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. and Karplus, M (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, 4: 187-217
- [49] Case DA, Cheatham TE, Darden T, et al. (2005) The Amber biomolecular simulation programs. *J Comput Chem* 26(16):1668-88
- [50] Lazaridis T (2003) Effective energy function for proteins in lipid membranes. *Proteins* 52(2):176-92.
- [51] Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ (2005) GROMACS: fast, flexible, and free *J Comput Chem* 26(16):1701-18.
- [52] Phillips JC, Braun R, Wang W, et al. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26(16):1781-802
- [53] Miao Y, McCammon JA (2016) G-protein coupled receptors: advances in simulation and drug discovery. *Curr Opin Struct Biol* 41:83-89
- [54] Jones G, Willett P, Glen RC, Leach AR, Taylor R (1997) Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* 267(3):727-48
- [55] Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234(3):779-815

- [56] Larsson P, Wallner B, Lindahl E, Elofsson A (2008) Using multiple templates to improve quality of homology models in automated homology modeling. *Protein science: a publication of the Protein Society* 17: 990–1002.
- [57] Porter JR, Weitzner BD, Lange OF (2015) A framework to simplify combined sampling strategies in Rosetta. *PLOS One* 10(9):e0138220.
- [58] Gribskov M, McLachlan AD, Eisenberg D (1987) Profile analysis: detection of distantly related proteins. *PNAS* 84(13):4355-8
- [59] Rychlewski L, Jaroszewski L, Li W, Godzik A (2000) Comparison of sequence profiles. Strategies for structural predictions using sequence information. *Protein Sci* 9(2):232-41
- [60] SF Altschul, W Gish, W Miller, EW Myers, DJ Lipman (1990) Basic local alignment search tool. *Journal of Molecular Biology* 215 (3), 403-410
- [61] Kufareva I, Katritch V, Participants of GPCR Dock 2013, Stevens RCS, Abagyan R (2014) Advances in GPCR Modeling Evaluated by the GPCR Dock 2013 Assessment: Meeting New Challenges. *Structure* 22(8), 1120 – 1139
- [62] Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 31(2):455-61
- [63] Krieger E, Vriend G (2014) YASARA View - molecular graphics for all devices - from smartphones to workstations. *Bioinformatics* 30(20):2981-2.
- [64] Bernsel A, Viklund H, Hennerdal A, Elofsson A (2009) TOPCONS: consensus prediction of membrane protein topology. *NAR* 37(Web Server issue):W465-8.
- [65] Larkin MA et al. (2007) Clustal W and Clustal X version 2.0, *Bioinformatics* 23 (21) 2947–2948
- [66] Srinivasan N, Blundell TL (1993) An evaluation of the performance of an automated procedure for comparative modelling of protein tertiary structure. *Protein Engineering, Design and Selection* 6(5), 501–512
- [67] Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5):1792-7
- [68] Messih MA, Lepore R, Tramontano A (2015) LoopIng: a template-based tool for predicting the structure of protein loops. *Bioinformatics* 31(23):3767-72
- [69] Martí-Solano M, Iglesias A, de Fabritiis G, Sanz F, Brea J, Loza MI, Pastor M, Selent J (2015) Detection of New Biased Agonists for the Serotonin 5-HT<sub>2A</sub> Receptor: Modeling and Experimental Validation. *Mol Pharm* 87(4), 740-746

- [70] Vroling B, Sanders M, Baakman C, et al. (2010) GPCRDB: information system for G protein-coupled receptors. *Nucleic Acids Res* 39(Database issue):D309-19
- [71] Worth CL, Kreuchwig A, Kleinau G, Krause G (2011) GPCR-SSFE: a comprehensive database of G-protein-coupled receptor template predictions and homology models. *BMC Bioinformatics* 12:185
- [72] Rodríguez, D.; Bello, X; Gutiérrez-de-Terán H (2012) Modeling G Protein-Coupled Receptors Through the Web. *Mol Inf* 31:334–41
- [73] Zhang Y (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 9:40
- [74] Abrol R, Bray JK, Goddard WA (2011) Bihelix: Towards de novo structure prediction of an ensemble of G-protein coupled receptor conformations. *Proteins* 80(2):505-18
- [75] Pándy-Szekeres G, Munk C, Tsonkov TM, et al. (2017) GPCRdb in 2018: adding GPCR structure models and ligands. *Nucleic Acids Res* 46(D1):D440-D446
- [76] Worth CL, Kreuchwig F, Tiemann JKS, et al. (2017) GPCR-SSFE 2.0-a fragment-based molecular modeling web tool for Class A G-protein coupled receptors. *Nucleic Acids Res* 45(W1):W408-W415
- [77] Zhang J, Yang J, Jang R, Zhang Y (2015) GPCR-I-TASSER: A Hybrid Approach to G Protein-Coupled Receptor Structure Modeling and the Application to the Human Genome. *Structure* 23(8):1538-1549
- [78] Murata T, Yamato I, Kakinuma Y, Leslie AG, Walker JE (2005) Structure of the rotor of the V-Type Na<sup>+</sup>-ATPase from *Enterococcus hirae*. *Science* 308(5722):654-9
- [79] Zhao J, Benlekbir S, Rubinstein JL (2015) Electron cryomicroscopy observation of rotational states in a eukaryotic V-ATPase. *Nature* 521(7551):241-5
- [80] Roh SH, Stam NJ, Hryc CF, Couoh-Cardel S, Pintilie G, Chiu W, Wilkens S (2018) The 3.5-Å CryoEM Structure of Nanodisc-Reconstituted Yeast Vacuolar ATPase Vo Proton Channel. *Mol Cell* 69(6):993-1004.e3
- [81] Hua T, Vemuri K, Nikas SP, et al. (2017) Crystal structures of agonist-bound human cannabinoid receptor CB1. *Nature* 547(7664):468-471.
- [82] Li X et al. (2019) Crystal Structure of the Human Cannabinoid Receptor CB2. *Cell* 176(3):459-467.e13
- [83] McAllister SD et al. (2003) An aromatic microdomain at the cannabinoid CB(1) receptor constitutes an agonist/inverse agonist binding region. *J Med Chem* 46(24):5139-52
- [84] Jakowiecki J, Filipek S (2016) Hydrophobic Ligand Entry and Exit Pathways of the CB1 Cannabinoid Receptor. *J Chem Inf Model* 56(12):2457-2466

- [85] Shim JY, Bertalovitz AC, Kendall DA (2011) Identification of essential cannabinoid-binding domains: structural insights into early dynamic events in receptor activation. *J Biol Chem* 286(38):33422-35
- [86] Pandey P, Roy KK, Doerksen RJ (2019) Negative allosteric modulators of cannabinoid receptor 2: protein modeling, binding site identification and molecular dynamics simulations in the presence of an orthosteric agonist. *J Biomol Struct Dyn* 5:1-16
- [87] Zhang H, Qiao A, Yang D, et al. (2017) Structure of the full-length glucagon class B G-protein-coupled receptor. *Nature* 546(7657):259-264
- [88] Johnson ZL, Lee JH, Lee K, et al. (2014) Structural basis of nucleoside and nucleoside drug selectivity by concentrative nucleoside transporters. *Elife* 3:e03604
- [89] Forrest LR (2013) Structural biology. (Pseudo-)symmetrical transport. *Science* 339(6118):399-401
- [90] DiMaio F, Leaver-Fay A, Bradley P, Baker D, André I (2011) Modeling Symmetric Macromolecular Structures in Rosetta3. *PLoS ONE* 6(6): e20450
- [91] McAllister SD, Hurst DP, Barnett-Norris J, Lynch D, Reggio PH, Abood ME (2004) Structural mimicry in class A G protein-coupled receptor rotamer toggle switches: the importance of the F3.36(201)/W6.48(357) interaction in cannabinoid CB1 receptor activation. *J Biol Chem* 279(46):48024-37
- [92] Pérez-Sayáns M1, Somoza-Martín JM, Barros-Angueira F, Rey JM, García-García A (2009) V-ATPase inhibitors and implication in cancer treatment. *Cancer Treat Rev* 35(8):707-13
- [93] Scheeff S, Menche D (2017) Total syntheses of the archazolids: an emerging class of novel anticancer drugs. *Beilstein J Org Chem* 13:1085-1098
- [94] Sowers JR, Raij L, Jialal I, et al. (2010) Angiotensin receptor blocker/diuretic combination preserves insulin responses in obese hypertensives. *J Hypertens* 28(8):1761-9
- [95] Keri D, Barth P (2018) Reprogramming G protein coupled receptor structure and function. *Curr Opin Struct Biol* 51:187-194

## 5. Omówienie pozostałych osiągnięć naukowo – badawczych.

### a) Podsumowanie dorobku naukowego

Liczba publikacji	<b>19</b>
Sumaryczny impact factor wg listy Journal Citation Reports zgodnie z rokiem opublikowania	<b>81.443</b>
Liczba cytowań publikacji według bazy Web of Science: (na dzień 08.04.2019)	<b>659</b>
Liczba cytowań publikacji według bazy Web of Science: (bez autocytowań, na dzień 08.04.2019)	<b>621</b>
Indeks Hirscha uwzględniający 19 publikacji wg bazy Web of Science: (na dzień 08.04.2019)	<b>11</b>

W dotychczasowych badaniach, poza tymi opisanymi w punkcie 4, zajmowałam się modelowaniem struktur białek transbłonowych i ich kompleksów ze związkami małowcząsteczkowymi i peptydami, co opisałam w publikacjach [B1-4] oraz [B6] i [B8]. Publikacje opracowane w okresie studiów doktoranckich [B7-9] i krótko po nich [B5], poświęcone tylko białkom globularnym stanowiły dla mnie podstawę metodologiczną do późniejszych badań nad białkami błonowymi.

Główne osiągnięcie naukowe stanowiące podstawę mojej pracy doktorskiej, dotyczyło stworzenia algorytmu do przewidywania struktur dużych (powyżej 150 aminokwasów) białek globularnych przy wykorzystaniu fragmentarycznych danych pochodzących z eksperymentu NMR, opisane przeze mnie w publikacji [B5] oraz [B8-9]. Ponadto, opracowałam algorytm wykorzystania w modelowaniu struktur białek globularnych informacji ewolucyjnej w postaci dalekozasięgowych kontaktów między resztami aminokwasowymi, który opisałam w publikacji [B7].

Najważniejszym osiągnięciem po okresie studiów doktoranckich, poza opisanymi w punkcie 4 niniejszego autoreferatu, uważam badania nad receptorami serotoninowymi i receptorem SMO w 2013 r. oraz nad receptorem dopaminowym D3 i chemokinowym CXCR4 w 2010 r., które zostały opisane w publikacjach [B1] i [B6]. Osiągnięcie to dotyczyło opracowania modelu oddziaływań wyżej wymienionych receptorów z małowcząsteczkowymi ligandami: ergotaminą (5-HT<sub>1B</sub> oraz 5-HT<sub>2B</sub>), SANT-1 oraz LY-2940680 (SMO), eticlopride (D3) oraz pochodną izotiomocznika IT1t oraz cyklicznym analogiem peptydu CVX15 (CXCR4). Powyższe badania obejmowały badania filogenetyczne receptorów i ich homologów, modelowanie struktury receptorów i ich kompleksów z ligandami, wyznaczenia

konformacji ligandów w stanie wolnym i w kompleksie z receptorem. Badania oddziaływań obu receptorów serotoninowych z ergotaminą pozwoliły na określenie molekularnych podstaw ich różnego profilu selektywności funkcjonalnej [B1]. Powyżej opisane badania pozwoliły mi na osiągnięcie sukcesu w ogólnoświatowym konkursie GPCR Dock 2013 [B1] (1 miejsce naszej grupy wśród 44 grup badawczych z całego świata) oraz zajęcie czołowego miejsca we wcześniejszym konkursie GPCR Dock 2010 [B6] (6 miejsce wśród 34 grup badawczych). Oba konkursy polegały na wyznaczeniu struktury wybranych receptorów GPCR z klasy A oraz F oraz sposobu wiązania przez nie małowcząsteczkowych ligandów bez jakiegokolwiek wcześniejszej informacji z badań krystalograficznych lub mikroskopowych, a zatem wykorzystując tylko sekwencję aminokwasową receptora i wzór chemiczny ligandu. Wymienione konkursy, podobnie jak organizowane od wielu lat: CASP (wyznaczanie struktur różnych rodzin białek, głównie globularnych) lub CAPRI (wyznaczanie struktur kompleksów białko-białko), bezstronnie oceniają trafność metod obliczeniowych stosowanych przez różne grupy badawcze. Oba konkursy GPCR Dock 2010 i GPCR Dock 2013, przeprowadzone na początku i w końcowym etapie mojej pracy badawczej na temat modelowania receptorów GPCR (uwieńczonej implementacją w postaci 2 aplikacji internetowych) podsumowują rozwój moich metod obliczeniowych opisanych w niniejszym autoreferacie.

Pozostałe osiągnięcia, opisane w publikacjach mojego współautorstwa [B2] i [B4] dotyczą określenia mechanizmów aktywacji wybranych receptorów GPCR z klasy A: receptora formylowego (FPR1) oraz lipidowego (S1P1). W przypadku FPR1 [B4] opisano mechanizm łączący mikroprzełącznik dźwigniowy (patrz Tab. 1), a w przypadku S1P1 [B2] mikroprzełącznik transmisji z przekierowaniem ruchu molekuł wody pośredniczących w wiązaniach wodorowych wewnątrz receptora.

Ostatnie, choć niezwykle ważne dla mnie osiągnięcie opisuje publikacja [B3]. Razem z moją wcześniejszą publikacją [B5] opisuje ona badania białek globularnych metodą rezonansu magnetycznego przy wykorzystaniu metod obliczeniowych. W obu przypadkach były to metody modelowania *de novo*, a więc niekorzystające z szablonów strukturalnych z bazy PDB. Zatem, obie prace stanowią uzupełnienie mojego warsztatu badawczego o inne, nieopisane w punkcie 4, metody obliczeniowe. Opublikowana w Journal of American Chemical Society publikacja [B3], zrealizowana podczas mojego krótkiego (5-miesięcznego) stażu podoktorskiego w EMBL Heidelberg opisuje zastosowanie danych NMR w wyznaczeniu sposobu wiązania liganda do białka (INPHARMA). Mój udział w badaniach polegał na stworzeniu prototypu algorytmu do przeszukiwania przestrzeni konformacyjnej kompleksu ligand-białko, który stanowił główny moduł do analizy danych w metodzie INPHARMA.



Metoda ta umożliwia wyznaczenie metodami NMR sposobu wiązania liganda bez konieczności rozwiązywania struktury całej jednostki białkowej.

**b) Publikacje stanowiące dorobek naukowy (poza cyklem publikacji wymienionych punkcie 4), opublikowane w czasopismach z bazy Journal Citation Reports)**

- po uzyskaniu stopnia naukowego doktora nauk chemicznych

Lp.	Autor/autorzy, tytuł publikacji, rok wydania, czasopismo, tom, strony	Impact Factor <sup>1</sup>	Liczba cytowań <sup>2</sup>
B1	Kufareva I, Katritch V, <b>Participants of GPCR Dock 2013</b> , Stevens RCS*, Abagyan R* (2014) <i>Advances in GPCR Modeling Evaluated by the GPCR Dock 2013 Assessment: Meeting New Challenges</i> . <b>Structure</b> , 22(8), 1120 – 1139	5.618	92
B2	Yuan S*, Wu R, <b>Latek D</b> , Trzaskowski B, Filipek S* (2013) <i>Lipid Receptor SIP1 Activation Scheme Concluded from Microsecond All-Atom Molecular Dynamics Simulations</i> . <b>PLOS Comput Biol</b> , 9(10), e1003261	4.829	17
B3	Codutti L, Skjaerven L, Angelini A, Grimaldi M, <b>Latek D</b> , Monecke P, Dreyer M & Carlomagno T* (2013) <i>Accounting for conformational variability in protein-ligand docking with NMR-guided rescoring</i> . <b>J Am Chem Soc</b> , 135(15), 5819-27	11.444	17
B4	Yuan S, Ghoshdastider U, Trzaskowski B, <b>Latek D</b> , Debinski A, Pulawski W, Wu R, Gerke V & Filipek S* (2012) <i>The role of water in activation mechanism of human N-formyl peptide receptor 1 (FPR1) based on molecular dynamics simulations</i> . <b>PLOS ONE</b> , 7, e47114	3.730	15
B5	<b>Latek D*</b> & Kolinski A (2011) <i>CABS-NMR--De novo tool for rapid global fold determination from chemical shifts, residual dipolar couplings and sparse methyl-methyl NOEs</i> . <b>J Comput Chem</b> , 32, 536-544	4.583	7
B6	Kufareva I, Rueda M, Katritch V, <b>GPCR Dock 2010 participants</b> , Stevens RCS*, Abagyan R* (2011) <i>Status of GPCR Modeling and Docking as Reflected by Community-wide GPCR Dock 2010 Assessment</i> . <b>Structure</b> , 19, 1108-1126	6.347	195

\* - autor korespondencyjny

<sup>1</sup>zgodnie z rokiem publikacji

<sup>2</sup>wg Web of Science z dnia 08.04.2019

- przed uzyskaniem stopnia naukowego doktora nauk chemicznych

Lp.	Autor/autorzy, tytuł/tytuły publikacji, rok wydania, nazwa wydawnictwa, recenzenci wydawniczy	Impact Factor <sup>1</sup>	Liczba cytowań <sup>2</sup>
B7	<b>Latek D*</b> & Kolinski A (2008) <i>Contact prediction in protein modeling: scoring, folding and refinement of coarse-grained models.</i> <b>BMC Struct Biol</b> , 8, 36	2.000	10
B8	<b>Latek D*</b> , Ekonomiuk D & Kolinski A (2007) <i>Protein structure prediction: combining de novo modeling with sparse experimental data.</i> <b>J Comput Chem</b> , 28, 1668-1676	4.297	20
B9	<b>Plewczyńska D</b> & Kolinski A* (2005) <i>Protein Folding with a Reduced Model and Inaccurate Short-Range Restraints.</i> <b>Macromol Theory Sim</b> , 14, 444-451	1.544	8

\* - autor korespondencyjny

<sup>1</sup>zgodnie z rokiem publikacji

<sup>2</sup>wg Web of Science z dnia 08.04.2019

### c) Monografie naukowe

Lp.	Autor/autorzy, rok wydania, tytuł, tytuł serii wydawniczej, edytor, nazwa wydawnictwa, tom, strony
1	<b>Latek D</b> , Trzaskowski B, Niewieczerzal S, Miszta P, Mlynarczyk K, Debinski A, Pulawski W, Yuan S, Szt Tyler A, Orzeł U, Jakowiecki J, Filipek S (2019) „ <i>Modeling of Membrane Proteins: From Bioinformatics to Molecular Quantum Mechanics.</i> ” w „ <i>Computational Methods to Study the Structure and Dynamics of Biomolecules and Biomolecular Processes From Bioinformatics to Molecular Quantum Mechanics</i> ”, Liwo A, Springer, vol. 8, 357-431.
2	Miszta P, Jakowiecki J, Rutkowska E, Turant M, <b>Latek D</b> , Filipek S (2018) „ <i>Approaches for Differentiation and Interconverting GPCR Agonists and Antagonists</i> ” w “ <i>Computational Methods for GPCR Drug Discovery. Methods in Molecular Biology</i> ”, Heifetz A, Humana Press, New York, NY, vol. 1705, 265-296
3	Rutkowska E, Miszta P, Mlynarczyk K, Jakowiecki J, Pasznic P, Filipek S, <b>Latek D</b> (2017) „ <i>Application of a Membrane Protein Structure Prediction Web Service GPCRM to a Gastric Inhibitory Polypeptide Receptor Model</i> ” w „ <i>Bioinformatics</i> ”

	<p>and Biomedical Engineering. IWBBIO 2017. Lecture Notes in Computer Science”, Rojas I, Ortuño F, Springer, Cham, vol. 10209, 151-162.</p>
4	<p><b>Latek D</b>, Trzaskowski B, Niewieczeral S, Miszta P, Mlynarczyk K, Debinski A, Pulawski W, Yuan S, Filipek S (2014) „<i>Modeling of Membrane Proteins</i>” w „<i>Computational methods to study the structure and dynamics of biomolecules and biomolecular processes – from bioinformatics to molecular quantum mechanics</i>”, Liwo A, Springer Series in Bio-/Neuroinformatics, vol. 1, 357–431.</p>
5	<p><b>Latek D &amp; Filipek S</b> (2012) „<i>Automatic template-based model generation of G-protein coupled receptors</i>” w „<i>Workshop 2012: From Computational Biophysics to Systems Biology</i>”, Carloni P et al., Publication Series of the Institute for Advanced Simulation, IAS Series, vol. 8, 111-114.</p>
6	<p>Gront D, <b>Latek D</b>, Kurcinski M, Kolinski A (2008) „<i>Template-Free Predictions of Three-Dimensional Protein Structures: From First Principles to Knowledge-Based Potentials</i>” w „<i>Prediction of Protein Structures, Functions, and Interactions</i>”, Bujnicki J, John Wiley &amp; Sons, Ltd., 117-141.</p>
7	<p>Gront D, <b>Latek D</b>, Kurcinski M, Kolinski A (2008) „<i>Wieloskalowe modelowanie bialek</i>” w „<i>Na pograniczu chemii i biologii</i>”, Koroniak H, Barciszewski J, Wydawnictwo Naukowe Uniwersytetu Adama Mickiewicza, vol. 17, 239-256.</p>
8	<p>Kolinski A, Gront D, Kmiecik K, Kurciński M &amp; <b>Latek D</b> (2006) „<i>Modeling protein structure, dynamics and thermodynamics with reduced representation of conformational space</i>” w „<i>NIC Workshop 2006: From Computational Biophysics to System Biology</i>”, Meinke J et al., John von Neumann Institute for Computing, Juelich, NIC Series, vol. 34, 21-28.</p>