



Wydział Chemii Uniwersytetu Gdańskiego

Karolina Zięba

**BADANIE ODDZIAŁYWAŃ HYDROFOBOWYCH I LOKALNYCH
W MODELOWYCH UKŁADACH ORAZ BIAŁKACH**

Promotor:

prof. dr hab. Cezary Czaplewski

Promotor pomocniczy:

dr Magdalena Ślusarz

*Praca doktorska wykonana
w Pracowni Symulacji Polimerów,
w Katedrze Chemii Teoretycznej*

Gdańsk, 2022

SPIS TREŚCI

1. PODZIĘKOWANIA	4
2. WYKAZ SKRÓTÓW	6
3. WYKAZ UŻYTEGO OPROGRAMOWANIA I SERWERÓW OBLICZENIOWYCH	8
4. CEL PRACY.....	9
5. WSTĘP TEORETYCZNY	10
5. 1. Struktura białek	10
5. 1. 1. Podział struktur białek.....	11
5. 1. 2. Miary podobieństwa struktur białek.....	15
5. 2. Oddziaływania stabilizujące strukturę białek.....	16
5. 2. 1. Efekt hydrofobowy.....	17
5. 2. 5 . Oddziaływania lokalne	20
5. 3. Metody obliczeniowe	21
5. 3. 1. Metody <i>ab initio</i>	22
5. 3. 2. Metody półempiryczne.....	22
5. 4. Metody empiryczne mechaniki molekularnej	23
5. 4. 1. Pełnoatomowe pola sił	25
5. 4. 2. Gruboziarniste pole sił UNRES	27
5. 5. Dynamika molekularna	33
5. 5. 1. Potencjał oddziaływań Lennarda-Jonesa vs pseudotwardych kul Mie	36
5. 6. Metody poprawy próbkowania w symulacjach dynamiki molekularnej.....	38
5. 7. Metody analizy danych	40
5. 7. 1. Metoda ważonych histogramów.....	41
5. 7. 2. Metody analizy skupień	42
5. 8. Metody obliczania energii swobodnej układu.....	44
5. 8. 1. Metoda próbkowania parasolowego.....	44
5. 8. 2. Metoda zaburzenia energii swobodnej	46
5. 8. 3. Metoda Widom'a wprowadzania cząstki	47
5. 9. Teoretyczne metody przewidywania struktur białek	48
5. 9. 1. Modelowanie homologiczne	49
5. 9. 2. Metody <i>de novo</i> przewidywania struktur białek	50
5. 9. 3. Połączenie modelowania homologicznego z polem sił UNRES	51
5. 9. 4. Metoda AlphaFold.....	53
5. 9. 4. Eksperyment CASP.....	53

6. BADANIA WŁASNE	56
6. 1. Hydratacja hydrofobowa i oddziaływania hydrofobowe cząstek Lennarda-Jonesa oraz pseudotwardych kul Mie w różnych modelach wody.....	56
6. 2. Modelowanie molekularne zależności energii swobodnej hydratacji od rozmiaru niepolarniej cząstki	69
6. 3. Rozszerzanie gruboziarnistego pola sił UNRES do białek membranowych w dwuwarstwie lipidowej	81
6. 4. Rozszerzanie parametryzacji pola sił UNRES na 5 lokalnych typów reszt aminokwasowych białek.....	93
6. 5. Przewidywanie struktur białek przy użyciu gruboziarnistego pola sił UNRES oraz informacji z baz danych w 13-tej edycji eksperymentu naukowego CASP.....	102
7. PODSUMOWANIE I WNIOSKI	107
8. LISTA PUBLIKACJI STANOWIĄCYCH ZAŁĄCZNIKI DO PRACY	109
9. LISTA POZOSTAŁYCH PUBLIKACJI	110
10. WYKAZ RYSUNKÓW	111
11. LITERATURA	116

1. PODZIĘKOWANIA

Chciałabym serdecznie podziękować wszystkim osobom z Pracowni Chemii Polimerów oraz z Pracowni Modelowania Molekularnego, których miałam okazję poznać podczas trwania swoich studiów doktoranckich. Dziękuję Wam za bardzo miłą i przyjazną atmosferę pracy oraz owocną współpracę podczas realizacji wspólnych projektów badawczych.

Szczególne podziękowania pragnę przy tym skierować do moich promotorów – Profesora Cezarego Czaplewskiego i Doktor Magdaleny Ślusarz, a także do Profesora Adama Liwo za pomoc, cierpliwość, możliwość uczestniczenia w pracach badawczych rozwijanego od wielu lat w Naszym Zespole gruboziarnistego pola sił UNRES oraz za ciekawe inspiracje naukowe, dzięki którym stałam się skromną częścią świata nauki, który zawsze tak bardzo mnie fascynował. Wasza wiedza, dokonania naukowe, dociekliwość w dążeniu do celu, pasja i duże zaangażowanie świadczą o Waszym wysokim profesjonalizmie i stanowią dla mnie godny wzór do naśladowania.

W tym miejscu chcę szczególnie podziękować moim rodzicom oraz narzeczonemu – zawsze trzymaliście za mnie kciuki i mocno we mnie wierzyliście bardziej niż ja sama będę kiedykolwiek w stanie.

Przeprowadzone przeze mnie badania zostały sfinansowane ze środków Narodowego Centrum Nauki z grantu OPUS 13 (2017/25/B/ST4/01026), którego byłam podwykonawcą, OPUS 14 (2017/27/B/ST4/00926) oraz HARMONIA 9 (2017/26/M/ST4/00044), a także ze środków Ministerstwa Nauki i Szkolnictwa Wyższego w formie grantów Badań Naukowych Służących Rozwojowi Młodych Naukowców oraz Doktorantów przyznanych mi w latach 2018 (538-8370-B814-17 oraz 538-8370-B814-18), 2019 (538-8370-B336-18 oraz 528-8370-B336-18) oraz w 2021 roku (539-T110-B919-21).

Dziękuję również Centrum Modelowania Matematycznego i Komputerowego w Warszawie, Akademickiemu Centrum Komputerowemu Cyfronet Akademii Górniczo-Hutniczej oraz Trójmiejskiej Akademickiej Sieci Komputerowej za udostępnienie mocy obliczeniowej na swoich klastrach do wykonywanych przeze mnie badań.

*Pracę tą dedykuję
moim rodzicom oraz narzeczonemu Piotrowi.*

2. WYKAZ SKRÓTÓW

- CASP – z ang.: *Critical Assessment of Techniques for Protein Structure Prediction*. Światowej klasy eksperyment naukowy polegający na ocenie różnych metod teoretycznych służących do przewidywania III- i IV-rzędowych struktur białek na podstawie ich sekwencji aminokwasowych.
- FEP – z ang.: *Free Energy Perturbation* – metoda zaburzenia energii swobodnej. Jedna z metod chemii obliczeniowej, która służy do obliczania względnych wartości energii swobodnych badanego układu.
- GDT-TS – z ang.: *Global Distance Test-Total Score* – globalny test odległości – jedna z miar porównawczych stosowanych w chemii obliczeniowej, która służy do określenia względnego podobieństwa dwóch porównywanych ze sobą struktur. W przypadku białek jest ona niezależna od długości łańcucha polipeptydowego.
- H-REMD – z ang.: *Hamiltonian-Replica Exchange Molecular Dynamics* – dynamika molekularna z hamiltonowską wymianą replik, w której do danej temperatury i parametrów funkcji energii jest przydzielona większa liczba trajektorii.
- MAXLIKE – z ang.: *Maximum Likelihood* – metoda maksymalnego podobieństwa.
- MD – z ang.: *Molecular Dynamics* – dynamika molekularna.
- MREMD – z ang.: *Multiplexed Replica Exchange Molecular Dynamics* – dynamika molekularna ze zwielokrotnioną wymianą replik, w której do danej temperatury jest przydzielona większa liczba trajektorii.
- PDB – z ang.: *Protein Data Bank* – baza danych zawierająca m. in. struktury eksperymentalne białek i ich kompleksy, które są zapisane i dostępne do ściągnięcia w formacie cyfrowym z rozszerzeniem „.pdb”.
- PI – z ang.: *Particle Insertion* – metoda wprowadzania cząstki – jedna z metod chemii obliczeniowej, która służy do określania bezpośredniej wartości energii swobodnej badanego układu.
- PMF – z ang.: *Potential of Mean Force* – potencjał średniej siły.
- REMD – z ang.: *Replica Exchange Molecular Dynamics* – dynamika molekularna z wymianą replik, w której do danej trajektorii przydzielona jest tylko jedna temperatura.
- RFE – z ang.: *Restricted Free Energy* – ograniczona energia swobodna.

- RMSD – z ang.: *Root Mean Square Deviation* – jedna z miar porównawczych używanych w chemii obliczeniowej, która opisuje średnią odległość między atomami w porównywanych i nałożonych na siebie strukturach białek.
- TYYYY – symboliczne oznaczenie białka w eksperymencie CASP, gdzie T to „Target”, a YYYY to liczba porządkowa.
- TM-score – z ang.: *Template Modeling - score* – jedna z miar stosowanych w chemii obliczeniowej, która służy do porównywania ze sobą dwóch struktur białek. Miara ta w porównaniu do GDT-TS jest bardziej niezależna od długości łańcucha głównego białek.
- UNRES – z ang.: *United RESidue* – rodzaj gruboziarnistego pola sił opracowanego i rozwijanego w grupie Profesora Adama Liwo i Profesora Harolda Scheragi.
- WHAM – z ang.: *Weighted Histogram Analysis Method* – metoda analizy ważonych histogramów stosowana w chemii obliczeniowej m. in. do interpretacji wyników symulacji MREMD lub H-REMD z próbkowaniem parasolowym (*umbrella sampling*).

3. WYKAZ UŻYTEGO OPROGRAMOWANIA I SERWERÓW OBLICZENIOWYCH

Oprogramowanie:

- ChemDraw – program przeznaczony do rysowania wzorów strukturalnych związków chemicznych.
- Gnuplot – program przeznaczony do tworzenia dwu- i trójwymiarowych wykresów funkcji.
- Lammps – *Large-scale Atomic/Molecular Massively Parallel Simulator* – program przeznaczony do symulacji układów molekularnych przy pomocy metody dynamiki molekularnej.
- LaTeX – *Lamport TeX* – program do tworzenia grafik, plików tekstowych i prezentacji multimedialnych.
- PyMOL – *Python MOlecular* – program do trójwymiarowej wizualizacji struktur związków chemicznych.
- PSIpred – *PSI-blast based secondary structure PREDiction* – program i serwer obliczeniowy przeznaczony do przewidywania struktury drugorzędowej białek
- UNRES – *United RESidue* – program, serwer obliczeniowy oraz gruboziarniste pole sił przeznaczone do badania struktur białek.
- Chimera – *UCSF Chimera* – program służący do wizualizacji i analizy układów molekularnych.
- VMD – *Visual Molecular Dynamics* – program do modelowania i wizualizacji przestrzennej układów molekularnych.

Serwery obliczeniowe:

- Okeanos – klaster obliczeniowy Interdyscyplinarnego Centrum Modelowania Matematycznego i Komputerowego (ICM) w Warszawie (okeanos.icm.edu.pl).
- Piasek – klaster obliczeniowy Wydziału Chemii Uniwersytetu Gdańskiego używany głównie w grupie badawczej Prof. Adama Liwo i Prof. Cezarego Czapliewskiego (piasek4.chem.univ.gda.pl).
- Prometheus – klaster obliczeniowy Akademickiego Centrum Komputerowego Cyfronet Akademii Górniczo-Hutniczej (prometheus.cyfronet.pl).
- Tryton – klaster obliczeniowy Centrum Informatycznego Trójmiejskiej Akademickiej Sieci Komputerowej (CI TASK) (tryton.task.gda.pl).

4. CEL PRACY

Białka stanowią podstawowy budulec wszystkich organizmów żywych, a znajomość ich struktury i dynamiki odgrywa duże znaczenie w prawidłowym zrozumieniu mechanizmów biochemicznych opisujących procesy funkcjonowania żywych komórek. Zrozumienie zależności między strukturą i właściwościami białek jest niezwykle pomocne w diagnozowaniu przyczyn występowania wielu jednostek chorobowych (np. zmian nowotworowych, zaćmy czy choroby Parkinsona) oraz w projektowaniu nowych leków.

Obecnie znanych jest 185 728 (stan na 22.04.2022 r.) struktur eksperymentalnych białek, jednak stanowi to stosunkowo niewielki ułamek liczby sekwencji aminokwasowych zdeponowanych dotychczas w bazie UniProt¹ (230 895 644, stan na 22.04.2022 r.). Dlatego coraz większy prym w przewidywaniu struktur przestrzennych białek wiodą metody bioinformatyczne, które są zależne od strukturalnych i sekwencyjnych baz danych. Pełnoatomowe symulacje komputerowe białek są bardzo trudne do zrealizowania i wymagają dostępu do komputerów o bardzo wysokiej mocy obliczeniowej takich jak ANTON². Do tego celu zostały opracowane gruboziarniste pola sił, które poprzez znaczny stopień uproszczenia reprezentacji struktur białek umożliwiają skrócenie czasu trwania takich symulacji na standardowych klastrach obliczeniowych.

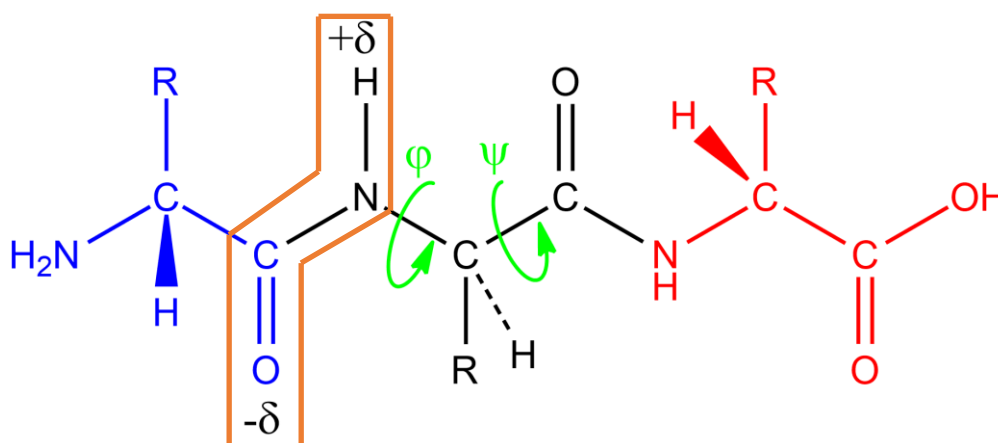
Struktura przestrzenna białek zależy od wielu czynników spośród których jednymi z najważniejszych są oddziaływania lokalne i hydrofobowe. Oddziaływania te mają kluczowy wpływ na stabilność i dynamikę struktur białek w badanym środowisku. Pierwszym celem mojej pracy było zbadanie obydwu typów oddziaływań w białkach. Do tego celu wykorzystałam gruboziarniste pole sił UNRES, w którym występują dwa centra oddziaływań – zjednoczone łańcuchy boczne oraz zjednoczone grupy peptydowe. Pole to jest domyślnie przeznaczone do badania białek w wodzie, dlatego w swoich badaniach staram się je rozszerzyć do środowiska lipidowego. W celu zwiększenia zdolności predykcyjnych pola UNRES dokonałam jego sparametryzowania z uwzględnieniem podziału reszt aminokwasowych białek na 5 grup. Do przewidywania struktur białek wykorzystałam również metodę hybrydową stanowiącą połączenie pola sił UNRES z modelowaniem homologicznym oraz wysokiej jakości modelami serwerowymi dla zadanej sekwencji aminokwasowej białka. Kolejnym celem mojej pracy było zbadanie hydratacji i oddziaływań hydrofobowych w prostych modelowych układach na sposób pełnoatomowy. Wyniki tych symulacji mogą być pomocne w zrozumieniu molekularnych podstaw efektu hydrofobowego w dużych układach biologicznych takich jak białka.

5. WSTĘP TEORETYCZNY

5. 1. Struktura białek

Białka stanowią jedną z najbardziej istotnych klas wielkocząsteczkowych związków chemicznych. Są podstawowym materiałem budulcowym wszystkich organizmów żywych, gdyż stanowią blisko 15 % całkowitej masy komórki oraz około 50 % jej składu chemicznego. Związki te odgrywają znaczącą rolę w wielu procesach biochemicznych, przykładowo takich jak biosynteza molekularna, ekspresja genowa, podziały i wzrost komórkowy, transport substancji chemicznych przez błony komórkowe oraz wiele innych^{3,4}.

Związki te są liniowymi polimerami zbudowanymi zazwyczaj z kombinacji 20 różnych reszt aminokwasowych. Każde białko posiada właściwą dla siebie, unikalną i precyzyjnie zdefiniowaną sekwencję aminokwasową określającą kolejność rozmieszczenia poszczególnych reszt w jej łańcuchu polipeptydowym⁵. Kolejność rozmieszczenia aminokwasów w białku zależy od informacji genetycznej kodowanej w kodonach cząsteczek kwasów nukleinowych DNA lub RNA⁶.



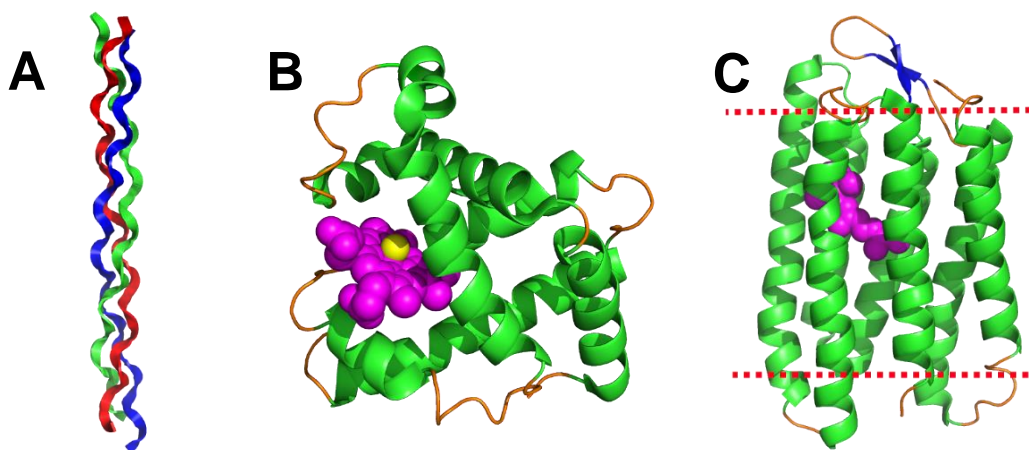
Rys. 1. Schematyczne przedstawienie łańcucha polipeptydowego białek z zaznaczonymi kątami dwuściennymi φ oraz ψ . Kolorem pomarańczowym oznaczono wiązanie peptydowe, niebieskim N-koniec, a czerwonym C-koniec łańcucha polipeptydowego.

Z powodu ogromnego stopnia złożoności białek, jak również szeregu funkcji jaką odgrywają one w komórkach organizmów żywych, od wielu lat trwają intensywne prace zmierzające do ustalenia budowy i właściwości wielu z nich. Eksperymentalne zbadanie struktury białek jest zadaniem bardzo trudnym i czasochłonnym, jednak dzięki rozwojowi technik komputerowych opracowano do tej pory wiele metod służących do teoretycznego wyznaczania struktur białek w krótszym czasie rzeczywistym. Kolejne rozdziały mojej rozprawy zostaną zatem poświęcone szczegółowemu omówieniu tych zagadnień.

5. 1. 1. Podział struktur białek


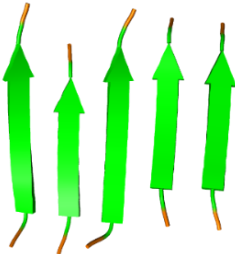
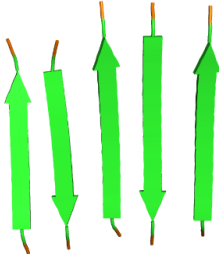
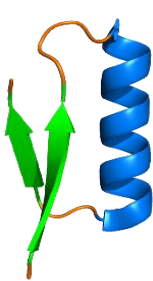

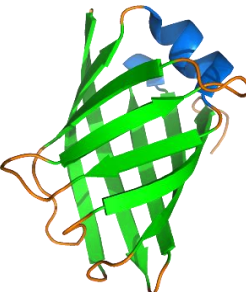
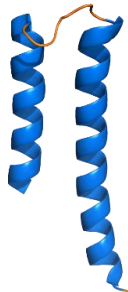
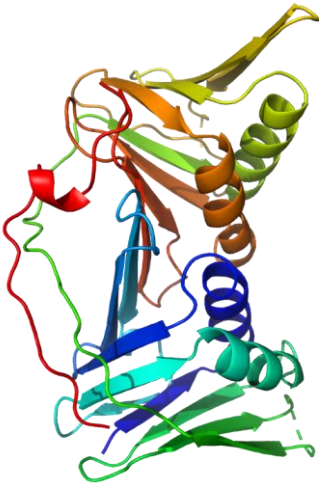
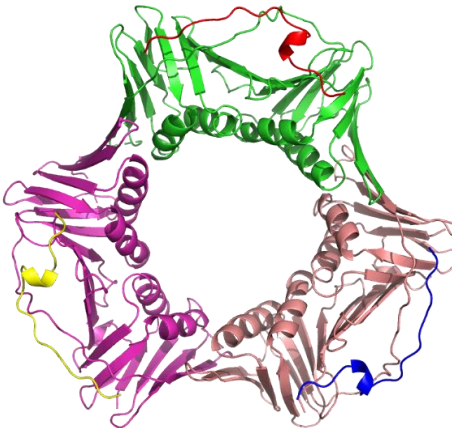
Jednym z podstawowych kryteriów podziału struktur białek jest ich kształt oraz zdolność do rozpuszczania się w zależności od środowiska. W tym zakresie białka są zazwyczaj klasyfikowane na trzy grupy, tj. fibrylarne, globularne oraz membranowe⁷.

Białka fibrylarne takie jak np. kolagen (Rys. 2A) tworzą regularne struktury liniowe, a także są nierozpuszczalne w wodzie oraz rozcieńczonych roztworach soli. W przeciwieństwie do nich białka globularne przyjmują sferyczny kształt. Jest to spowodowane dużym stopniem upakowania ich łańcucha polipeptydowego, w którym hydrofilowe łańcuchy boczne aminokwasów są ustawione na zewnątrz cząsteczki białka i tym samym wyeksponowane na działanie rozpuszczalnika. Efektem takiej budowy białek globularnych jest zazwyczaj ich bardzo dobra rozpuszczalność w roztworach wodnych. Przykładowym reprezentantem tego typu białek jest mioglobina (Rys. 2B). Odmienną grupę stanowią białka membranowe np. bakteriorodopsyna (Rys. 2C), które są powiązane z licznymi systemami błon komórkowych. W tym przypadku hydrofobowe łańcuchy boczne aminokwasów są skierowane na zewnątrz cząsteczki białka i dzięki temu wykazują one zdolność do oddziaływania z fazą niepolarną błon komórkowych. Skutkuje to dobrą rozpuszczalnością białek membranowych w roztworach surfaktantów oraz brakiem ich rozpuszczalności w roztworach wodnych.



Rys. 2. Przykładowe reprezentacje przestrzenne białka fibrylarnego – kolagenu (A), globularnego - mioglobiny (B) oraz membranowego – bakteriorodopsyny (C) (membrana została oznaczona kolorem czerwonym).

Każda z wyżej omówionych grup białek może tworzyć struktury odznaczające się różnym stopniem rzędowości. Parametr ten charakteryzuje poziom zaawansowania budowy białek i na jego podstawie dokonywany jest podział tworzonych przez nie struktur na I-, II-, III- oraz IV-rzędowe. Poniższy rysunek przedstawia przykładowe reprezentacje struktur białek o różnym stopniu rzędowości.

A	<p style="text-align: center;">Struktura I-rzędowa</p> <p style="text-align: center;">DHYNCSVSSGGQCLYSACPIFTKIQGTCYRGAACKCK</p>	
B	<p style="text-align: center;">Struktura II-rzędowa</p> <div style="display: flex; justify-content: space-around;"> <div style="text-align: center;"> <p>α-helisa</p>  </div> <div style="text-align: center;"> <p>równoległa β-kartka</p>  </div> <div style="text-align: center;"> <p>antyrownoległa β-kartka</p>  </div> </div>	
B-1	<p style="text-align: center;">Motywy strukturalne</p> <div style="display: flex; justify-content: space-around;"> <div style="text-align: center;"> <p>$\beta\alpha\beta$</p>  </div> <div style="text-align: center;"> <p>β-spinka</p>  </div> <div style="text-align: center;"> <p>klucz grecki</p>  </div> <div style="text-align: center;"> <p>α-helisa-pętla-α-helisa</p>  </div> </div>	
C	<p style="text-align: center;">Struktura III-rzędowa</p> 	<p style="text-align: center;">Struktura IV-rzędowa</p> 

Rys. 3. Przykładowe reprezentacje przestrzenne struktur białek o różnym stopniu rzędowości.

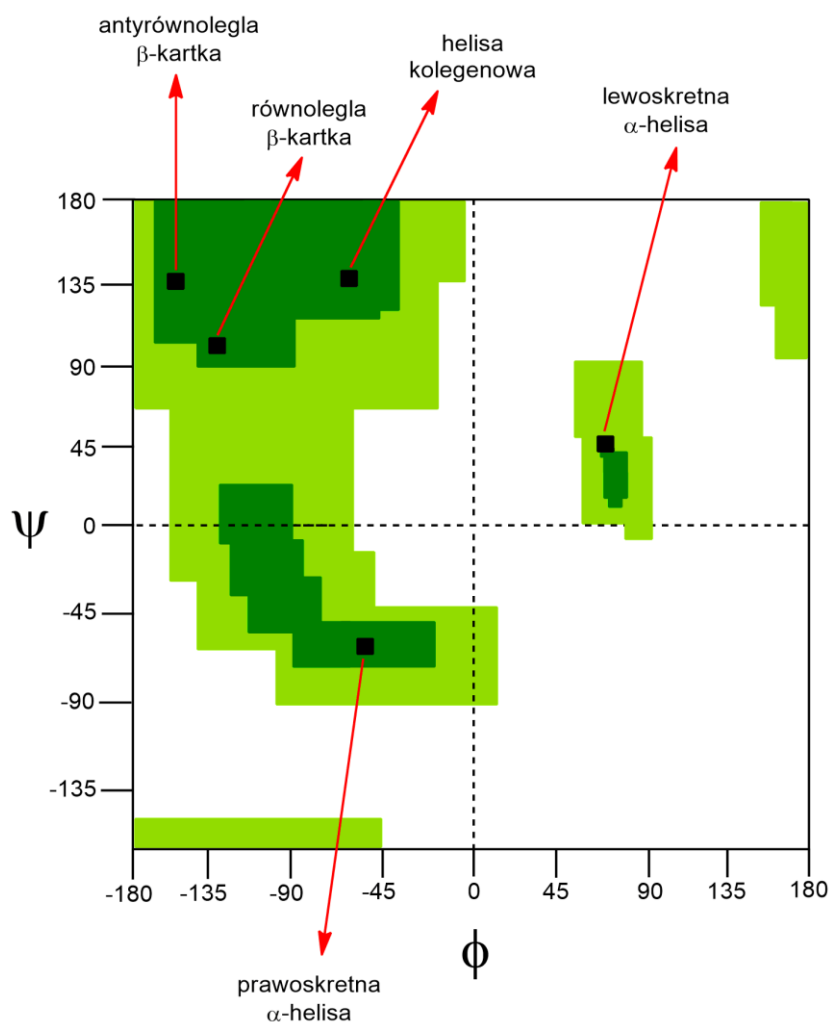
Struktura I-rzędowa (Rys. 3A) przedstawia liczbę oraz kolejność rozmieszczenia reszt aminokwasowych w łańcuchu głównym białka.

Struktura II-rzędowa (Rys. 3B) opisuje uporządkowaną konformację przestrzenną łańcucha polipeptydowego białka. Powszechnymi motywami struktur II-rzędowych są α -helisa, β -kartka, pętle oraz zgięcia. Struktury helikalne mają kształt zbliżony do cylindrycznego przypominający prawoskrętną śrubę stabilizowaną wewnątrzcząsteczkowymi wiązaniami wodorowymi. Występują one w kilku niezależnych od siebie formach opisywanych przez parametr n_m , w którym n to liczba reszt aminokwasowych przypadających na jeden pełny skręt helisy wokół własnej osi, natomiast m określa liczbę atomów zlokalizowanych w obrębie pojedynczego wiązania wodorowego^{8,9}. Zgodnie z tak przyjętą nomenklaturą standardowo wyróżnia się podział na helisę: 3.6₁₃ (α -helix), 3.6₁₀ oraz 4.4₁₆ (π -helix). Pierwsza z nich – zwyczajowo określana jako α -helisa (Rys. 3B) – jest dominującą formą konformacyjną struktury II-rzędowej w białkach globularnych¹⁰. Odmienne typy helisy jest obserwowany w kolagenie (Rys. 2A) – białku zlokalizowanym w tkance łącznej ssaków, który przyjmuje strukturę potrójnej helisy. Strukturę tego białka tworzą trzy ciasno skompresowane w przestrzeni łańcuchy polipeptydowe składające się z proliny, glicyny, hydroksylizyny i hydroksyproliny^{6,11}. Kolejna ze struktur II-rzędowych, tj. β -kartka ma formę rozciągniętego łańcucha polipeptydowego, który stabilizowany jest za pomocą wiązań wodorowych między ugrupowaniami NH i CO aminokwasów rozmieszczonych w tej samej płaszczyźnie. W tego typu strukturze sąsiednie łańcuchy białka mogą być rozmieszczone w tym samym kierunku i tworzyć w ten sposób konformację równoległą β -kartki lub w kierunkach przeciwnych, gdzie tworzą bardziej korzystną energetycznie konformację antyrównoległą β -kartki (Rys. 3B)⁵. W wielu białkach często spotykane są także struktury nadrzędne zwane popularnie motywami strukturalnymi. Są to określone konfiguracje odpowiednio rozmieszczonych względem siebie w przestrzeni struktur II-rzędowych białek³. Do najpopularniejszych motywów strukturalnych zaliczane są ugrupowania takie jak: β -kartka- α -helisa- β -kartka ($\beta\alpha\beta$), β -kartka-zwrot- β -kartka (β -spinka), α -helisa-pętla- α -helisa oraz tzw. klucz grecki⁶ (Rys. 3B-1). Charakterystycznym elementem budowy strukturalnej białek są również zgięcia łańcuchów polipeptydowych. W tym zakresie wyodrębniamy podział na zgięcia typu α , β , π oraz γ , w których uczestniczy odpowiednio pięć, cztery, sześć i trzy reszty aminokwasowe¹². Dozwolone stereochemicznie typy zgięć zależą od wartości kątów torsyjnych ψ oraz ϕ tworzonych przez atomy reszt aminokwasowych łańcucha głównego białek. Wzajemne zależności wartości tych kątów są przedstawiane przy pomocy dwuwymiarowego wykresu Ramachandrana, który obrazuje dozwolone i niedozwolone sterycznie konformacje łańcucha głównego białek (Rys. 4). Wykres ten został opracowany przez Ramachandrana i służy do oceny stereochemicznej struktur białek¹³.

Struktura III-rzędowa^{3,5} (Rys. 3C) opisuje sposób zwinięcia łańcucha głównego białka, który jest wynikiem wzajemnych oddziaływań między resztami aminokwasowymi jego łańcuchów bocznych. Do tych oddziaływań zaliczane są wiązania wodorowe, mostki disulfidowe i solne, oddziaływania lokalne, elektrostatyczne i hydrofobowe.

Struktura IV-rzędowa⁶ (Rys. 3D) opisuje wzajemne ułożenie przestrzenne dwóch lub większej liczby łańcuchów polipeptydowych w kompleksach białek.

Cechą szczególną II-, III- oraz IV-rzędowych struktur białek jest występowanie wśród nich tzw. domen. Mianem domeny określa się oddzielny fragment łańcucha polipeptydowego, który jest zdolny do samodzielnej egzystencji i zachowania swojej struktury przestrzennej niezależnie od pozostałej reszty cząsteczki białka¹⁴.



Rys. 4. Schematyczny wykres Ramachandrana przedstawiający dozwolone i niedozwolone sterycznie typy zgięć w łańcuchu głównym białka¹⁵.

5. 1. 2. Miary podobieństwa struktur białek

Przewidywanie struktur przestrzennych białek wymaga użycia odpowiednich miar opisujących podobieństwo w zakresie dwóch porównywanych ze sobą struktur białek. W chemii obliczeniowej znanych jest wiele takich miar, a wybór każdej z nich powinien być dopasowany indywidualnie do potrzeb analizowanego problemu badawczego. Umożliwiają one dokonanie ilościowej oceny modelowanej struktury, są także pomocne w jej identyfikacji oraz przewidywaniu potencjalnych zmian konformacyjnych w jej obrębie. W niniejszym rozdziale ograniczę się do przedstawienia jedynie tych miar podobieństwa, które wykorzystywałam w toku prowadzonych przez siebie badań.

Jedną z najczęściej stosowanych w bioinformatyce miar porównawczych jest średnie kwadratowe odchylenie pozycji atomów (ang.: *Root Mean Square Deviation*, RMSD)^{16,17}. Koncepcja tej miary opiera się na analizie średnich odległości między ustalonymi atomami w dwóch porównywanych ze sobą strukturach białek po uprzednim nałożeniu na siebie obydwu ze struktur. Opisywaną miarę przedstawia poniższa zależność matematyczna:

$$\text{RMSD} = \sqrt{\frac{1}{N} \sum_{i=1}^N (r_i^X - r_i^Y)^2} \quad (5.1.2a)$$

gdzie N oznacza liczbę porównywanych par atomów, a wyrażenie $(r_i^X - r_i^Y)$ reprezentuje odległość między i -tym atomem struktury X oraz i -tym atomem struktury Y. Im mniejsza jest wartość RMSD między dwiema porównywanymi ze sobą strukturami białek tym są one bardziej do siebie podobne. Wadą opisywanej miary jest jej ścisła zależność od długości łańcucha białkowego oraz nadmierna wrażliwość na niewielkie różnice między porównywanymi ze sobą strukturami białek.

Bardziej dokładną miarą do określania globalnych podobieństw dla porównywanych ze sobą struktur białek jest GDT-TS (ang.: *Global Distance Test - Total Score*)^{18,19}. Miara ta opisuje liczbę atomów C α w modelowanej strukturze białka, które mieszczą się w zdefiniowanej odległości odcięcia względem ich pozycji w strukturze porównawczej. Obliczenia te wykonywane są po optymalnej superpozycji obydwu ze struktur. Uzyskany wynik GDT-TS jest średnią arytmetyczną po wszystkich progach odcięcia wynoszących odpowiednio 1, 2, 4 i 8 Å. Wynik ten przedstawiany jest w formie procentowej w zakresie od 0 do 100% lub liczbowej w zakresie od 0 do 1, przy czym wyższe wartości opisywanej miary świadczą o większym podobieństwie porównywanych ze sobą struktur.

Sposób wyliczania opisywanej miary przedstawia poniższe równanie:

$$\text{GDT-TS} = \frac{(\text{GDT}_{p1} + \text{GDT}_{p2} + \text{GDT}_{p4} + \text{GDT}_{p8})}{4} \quad (5.1.2b)$$

gdzie GDT_{pd} to procent reszt aminokwasowych zlokalizowanych w odpowiadających im obszarach struktury referencyjnej ograniczonych przez odległości mniejsze lub równe wartości odległości odcięcia. Opisywana miara ma znaczną przewagę nad RMSD, ponieważ jest niezależna od długości łańcucha polipeptydowego i dzięki temu lepiej odzwierciedla globalne podobieństwa porównywanych ze sobą struktur.

Kolejną miarą opisującą względne podobieństwa w zakresie topologii porównywanych ze sobą struktur białek jest TM-score (ang.: *Template Modeling - score*)^{20,21}. Miara ta stanowi rozszerzenie metody MaxSub-score²² oraz GDT-TS^{18,19}, jest także rozszerzoną modyfikacją metody Levitta-Gersteina²³. TM-score jest miarą niezależną od długości łańcucha polipeptydowego białek, jednak w przeciwieństwie do GDT-TS jest ona mniej wrażliwa na występowanie lokalnych błędów powstałych podczas modelowania pętli oraz *N*- i *C*-końców białek. Wartości TM-score liczone są po optymalnej superpozycji porównywanych ze sobą struktur białek i plasują się w zakresie od 0 do 1, przy czym wartość 1 reprezentuje idealne dopasowanie strukturalne. Funkcję TM-score opisuje następujące równanie:

$$\text{TM - score} = \max \left[\frac{1}{L_T} \sum_{i=1}^{L_A} \frac{1}{1 + (D_i/D_0(L_T))^2} \right] \quad (5.1.2c)$$

gdzie L_T i L_A to liczba reszt aminokwasowych w referencyjnej oraz badanej strukturze białka, D_i opisuje odległość między i -tą parą reszt aminokwasowych w obydwu strukturach, „max” oznacza maksymalną wartość po optymalnym nałożeniu na siebie każdej ze struktur, natomiast D_0 jest skalą służącą do normalizacji odległości pary reszt aminokwasowych w porównywanych strukturach i jest ona dana jako:

$$D_0 = 1.24 \sqrt[3]{L_T - 15} - 1.8 \quad (5.1.2d)$$

5. 2. Oddziaływania stabilizujące strukturę białek

Do kluczowych czynników mających wpływ na stabilizację omawianych w poprzednim rozdziale struktur białek należą oddziaływania elektrostatyczne i van der Waalsa, wiązania wodorowe, mostki siarczkowe i solne oraz oddziaływania lokalne i efekt hydrofobowy.

Krótkozasięgowe oddziaływania elektrostatyczne występują między trwałym i indukowanym dipolem oraz dwoma trwałymi dipolami atomów reszt aminokwasowych białek. Z kolei dużo od nich słabsze oddziaływania van der Waalsa występują między dwoma indukowanymi dipolami atomów reszt aminokwasowych białek^{6,24,25}.

Wiązania wodorowe^{6,24,26} są oddziaływaniami o charakterze głównie elektrostatycznym i warunkują powstanie struktury II-rzędowej każdego białka, dzięki czemu są one kluczowym wyznacznikiem ich konformacji przestrzennych. Wiązania te występują między wolną parą elektronową atomu tlenu grupy karboksylowej C=O oraz atomem wodoru silnie spolaryzowanej grupy aminowej N-H. W cząsteczkach białek obydwie grupy funkcyjne są rozmieszczone względem siebie w sposób umożliwiający powstanie jak największej liczby wiązań wodorowych. Są to oddziaływania o charakterze głównie lokalnym, ponieważ blisko 70% z nich jest współtworzone przez atomy należące do łańcucha głównego białek.

Mostki siarczkowe²⁷ są kowalencyjnymi wiązaniami chemicznymi, które przyczyniają się do stabilizacji III-rzędowej struktury białek. Wiązania te występują między łańcuchami bocznymi cysteinowych reszt aminokwasowych białek położonych blisko siebie w przestrzeni.

Mostki solne^{6,24,28} mają duże znaczenie w stabilizowaniu struktury III-rzędowej białek. Występują one najczęściej między blisko rozmieszczonymi w przestrzeni dodatnio naładowanymi łańcuchami bocznymi lizyny, argininy i histydyny oraz ujemnie naładowanymi łańcuchami bocznymi kwasu glutaminowego lub asparaginowego.

5. 2. 1. Efekt hydrofobowy

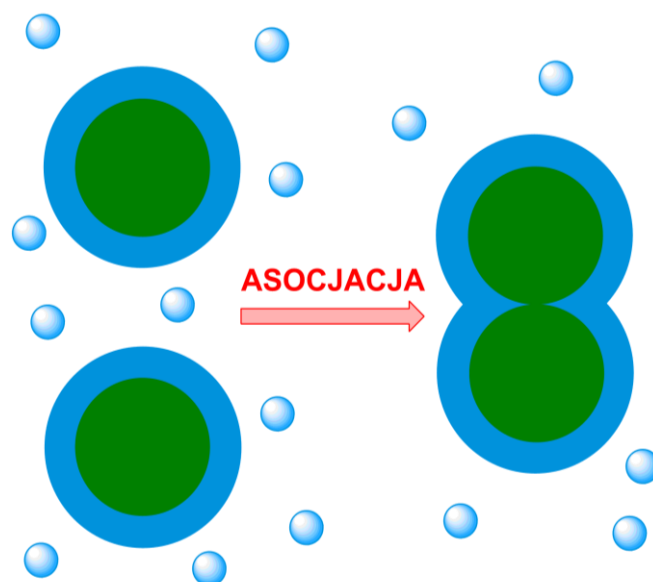
Efekt hydrofobowy^{6,29,30} ma duży wpływ na procesy tworzenia komórek, stabilizowanie struktur białek, wiązanie ligandów w białkach i ich kompleksach, agregację błon biologicznych w wodzie i wiele innych zjawisk. Termin ten odnosi się do procesów związanych z niską rozpuszczalnością substancji niepolarnych w wodzie oraz ich asocjacją w tym środowisku. Występowanie efektu hydrofobowego wynika z charakterystycznej wartości stałej dielektrycznej wody ϵ równej 80.2 w temperaturze 20°C przy ciśnieniu 1013 hPa³¹. Efekt hydrofobowy jest ilościowo wyrażany za pomocą zmiany wartości energii swobodnej Gibbsa ΔG , która dla procesów zachodzących w stałych warunkach ciśnienia i temperatury przedstawiana jest w następujący sposób:

$$\Delta G = \Delta H - T\Delta S \quad (5.2.1a)$$

gdzie ΔH to entalpia opisująca efekty energetyczne w badanym układzie, a $T\Delta S$ to człon entropowy charakteryzujący stopień uporządkowania tego układu w zależności od panującej

w nim temperatury. Frank i Evans już w 1945 r. dowiedli, że obserwowane w efekcie hydrofobowym zmiany entropowe układu są wynikiem zmian w sposobie uporządkowania struktury wody³². Wprowadzenie substancji niepolarniej do wody prowadzi do zaburzenia jej struktury. Jest to spowodowane zerwaniem części wiązań wodorowych między cząsteczkami wody w celu utworzenia hydrofobowej wnęki poprzez którą substancja niepolarna mogłaby wniknąć do wnętrza roztworu. Proces ten jest istotnie zależny od wielkości hydrofobowej wnęki, której średnica powinna być adekwatna do rozmiaru niepolarniej cząstki. Rozpuszczanie małych cząstek substancji niepolarniej o średnicy poniżej 1 nanometra powoduje zmianę gęstości wody i sposobu uporządkowania wiązań wodorowych między cząsteczkami wody w pierwszej warstwie solwatacyjnej takich substancji. W przypadku większych cząstek dochodzi do zerwania wiązań wodorowych między sąsiadującymi z nimi cząsteczkami wody. Skutkuje to tym, że cząsteczki wody orientują się na powierzchni dużej hydrofobowej wnęki, gdzie tworzą rozległą sieć wiązań wodorowych przyczyniających się do powstania międzyfazowej powierzchni typu woda – substancja niepolarna³³.

Substancje niepolarne podczas rozpuszczania w wodzie wykazują sporą tendencję do agregacji, która wynika z występowania silnych wiązań wodorowych między cząsteczkami wody oraz słabych oddziaływań przyciągających van der Waalsa między wodą i substancją niepolarną. Procesem tłumaczącym stosunkowo słabą rozpuszczalność tych substancji w wodzie jest hydratacja hydrofobowa³⁴. Jest to proces, w którym cząstki substancji niepolarniej są otaczane przez wysoce uporządkowaną strukturę wody. Konsekwencją tego procesu jest całkowite wykluczenie substancji niepolarniej z fazy wodnej. Zgodnie z teorią hydrofobowości Pratta i Chandlera³⁵ procesowi temu towarzyszą odmienne wielkości termodynamiczne względem temperatury, które zależą od rozmiaru niepolarniej cząstki wprowadzanej do wody. Wzrost temperatury układu w procesie hydratacji cząstek o średnicy dużo mniejszej niż 1 nanometr pociąga za sobą wzrost wartości ΔG , ΔH oraz $T\Delta S$. Natomiast w przypadku hydratacji większych cząstek rosnąca temperatura układu pociąga za sobą wzrost wartości ΔH i $T\Delta S$ oraz spadek wartości ΔG . Wykluczanie substancji niepolarnych z fazy wodnej jest ściśle skorelowane i sterowane za pośrednictwem oddziaływań hydrofobowych³⁶ występujących wówczas, gdy dwie cząstki substancji niepolarniej znajdują się dostatecznie blisko siebie w środowisku wodnym. Przejawem tych oddziaływań jest asocjacja cząstek niepolarnych w wodzie, która prowadzi do zmniejszenia ogólnej powierzchni hydrofobowej wnęki. Ma to na celu zminimalizowanie wartości ΔG po wprowadzeniu do wody drugiej cząstki substancji niepolarniej. Wraz ze wzrostem temperatury układu procesowi asocjacji towarzyszy spadek wartości entalpii ΔH oraz energii swobodnej ΔG , a także wzrost wartości członu entropowego $T\Delta S$ ³⁴.

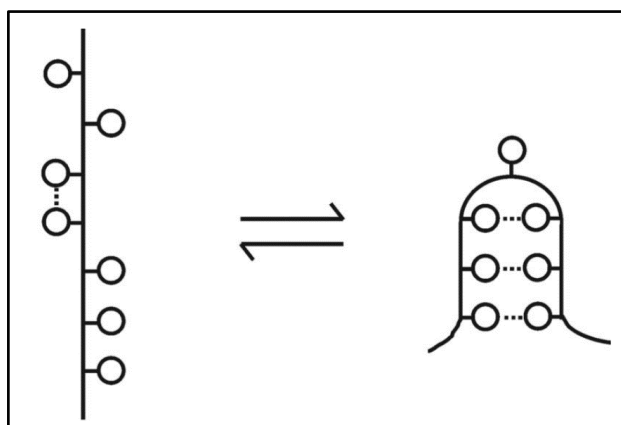


Rys. 5. Schematyczne przedstawienie efektu hydrofobowego dla dwóch niepolarnych cząstek wprowadzonych do wody. Kolor zielony przedstawia solwatowane cząstki, a niebieski sąsiadujące z nimi cząsteczki wody.

W celu scharakteryzowania oraz wyjaśnienia efektu hydrofobowego zastosowano do tej pory szeroki wachlarz metod eksperymentalnych, np. metodę dyfrakcji neutronów (ang.: *Neutron Diffraction*) oraz spektroskopię jądrowego rezonansu magnetycznego (ang.: *Nucleic Magnetic Resonance*, NMR), które umożliwiają określenie właściwości termodynamicznych substancji niepolarnych w wodzie. Niemniej jednak z uwagi na niską rozpuszczalność tych substancji w wodzie eksperymentalne zbadanie opisywanego efektu jest bardzo trudne w realizacji³⁴. Dlatego bardzo pomocne w tym zakresie są metody symulacji komputerowych do omówienia których przystąpię w dalszej części pracy.

Oddziaływania hydrofobowe są postrzegane jako jeden z najistotniejszych czynników stabilizujących strukturę białek. Cykl stabilizacyjny przebiega przez wiele etapów pośrednich i jest opisywany za pomocą mechanizmu zwijania i składania struktur białek. W etapie inicjującym ten mechanizm uczestniczą oddziaływania hydrofobowe między niepolarnymi resztami aminokwasowymi umieszczonymi blisko siebie w I-rzędowej strukturze białka. Efektem tych oddziaływań jest utworzenie tzw. kieszeni hydrofobowych postrzeganych za miejsca inicjacji fałdowania struktury białka. Miejscem inicjacji jest konformacja stanowiąca fragment cząsteczki białka, której występowanie zwiększa tempo jego przemian konformacyjnych³⁷. Kieszenie hydrofobowe posiadają charakterystyczne zgięcia, które utrzymują się do momentu osiągnięcia przez badane białko struktury stanu natywnego. Po utworzeniu kieszeni hydrofobowych w rozwiniętym łańcuchu białka poszukiwane są lokalne miejsca inicjacji. Odznaczają się one przejściową stabilnością, która może wzrosnąć podczas

oddziaływań z pobliskimi kieszeniami hydrofobowymi. Efektem tego struktura przestrzenna badanego białka staje się dużo bardziej uporządkowana³⁷.



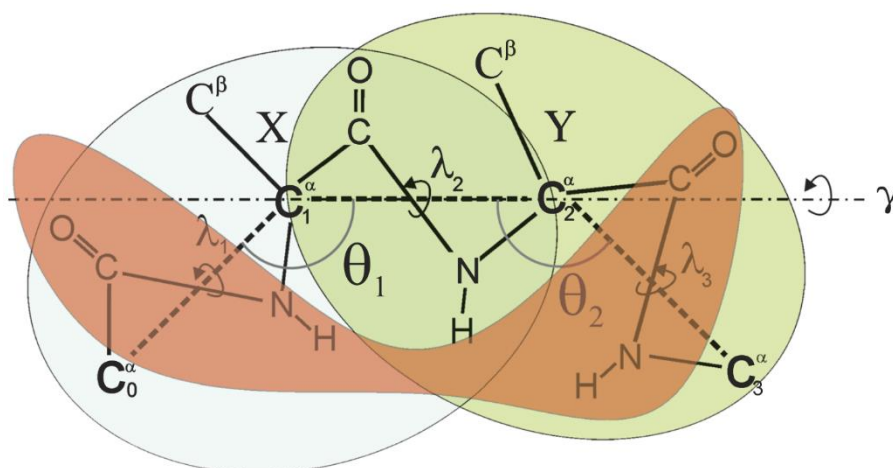
Rys. 6. Schematyczne przedstawienie etapu inicjacji faldowania struktur białek. Oddziaływania hydrofobowe między parą niepolarnych łańcuchów bocznych zostały oznaczone linią przerywaną. Struktury niezwinęta (po lewej) i zwinięta (po prawej) istnieją w stanie równowagi chemicznej³⁷.

5. 2. 5 . Oddziaływania lokalne

Oddziaływania lokalne określają II-rzędową strukturę białek oraz miejsca, w których występuje nieprawidłowy sposób ich uporządkowania strukturalnego. W ujęciu pełnoatomowym występują one pomiędzy i -tą aż do $(i+4)$ -tej reszty aminokwasowej włącznie³⁸. Natomiast w przypadku białek symulowanych z użyciem gruboziarnistego pola sił UNRES oddziaływania te występują pomiędzy trzema kolejnymi atomami C^α w łańcuchu głównym białek i atomem C^β każdej z reszt aminokwasowych, a także między zjednoczonymi grupami peptydowymi oddalonymi od siebie o jedną jednostkę peptydową. Oddziaływania lokalne w polu sił UNRES zależą od wartości wirtualnych kątów walencyjnych θ oraz torsyjnych γ zawartych odpowiednio pomiędzy trzema i czterema kolejnymi atomami C^α w łańcuchu głównym białka, a także od wartości kątów λ opisujących możliwe rotacje zjednoczonych grup peptydowych wokół osi wirtualnych wiązań $C^\alpha \cdots C^\alpha$ w łańcuchu głównym białka (Rys. 7)³⁹.

Oddziaływania lokalne odgrywają kluczową rolę w procesie zwijania struktur białek^{38,40}. W procesie tym powstaje wiele konformacji białek, a rola opisywanych oddziaływań polega na wzmocnieniu stopnia stabilizacji jedynie tych konformacji, które odznaczają się prawidłowymi kątami zgięć φ oraz ψ w łańcuchu głównym białka. Wzrost siły tych oddziaływań do wspomnianego mechanizmu przyczynia się do spadku bariery energetycznej między wydłużoną i zwiniętą konformacją białka. Minimum energetyczne stanu rozwiniętego staje się wówczas niższe i przesuwa się w kierunku konformacji odpowiadającej stanowi zwiniętemu. Wzrost siły oddziaływań lokalnych prowadzi do wzrostu szybkości zwijania się

struktur białek, czemu towarzyszy zmniejszenie energii konformacyjnej wszystkich stanów nienatywnych. Badania eksperymentalne przeprowadzone dla krótkich monomerycznych polipeptydów w wodzie takich jak np. poli-L-alanina dowodzą, że oddziaływania lokalne mają kluczowe znaczenie jedynie w tych fragmentach białek, których konformacje są podobne do struktury natywnej. Dowodzą ponadto, że mutacje polarnych reszt aminokwasowych są powodem wzrostu liczby oddziaływań lokalnych w mechanizmie zwijania białek⁴¹.



Rys. 7. Schemat obrazujący występowanie oddziaływań lokalnych w łańcuchu głównym białek w gruboziarnistym polu sił UNRES³⁹.

5. 3. Metody obliczeniowe

Metody chemii obliczeniowej wymagają mniejszych nakładów czasowych w porównaniu do tradycyjnych metod eksperymentalnych i często stanowią ich doskonałe uzupełnienie. Oferują one możliwość komputerowego zbadania układów molekularnych, których analiza eksperymentalna jest bardzo trudna lub wręcz niemożliwa w realizacji. Znajdują one szerokie zastosowanie w wielu gałęziach nauki, przemysłu i medycyny. Umożliwiają przewidywanie właściwości oraz śledzenie dynamiki wielu układów molekularnych. Są pomocne w analizie danych eksperymentalnych, opracowywaniu mechanizmów wielu reakcji chemicznych, a także w projektowaniu nowych materiałów oraz farmaceutyków. W zależności od rodzaju badanego układu molekularnego, sposobu jego reprezentacji oraz rodzaju stosowanych w nim przybliżeń metody obliczeniowe dzielą się na trzy grupy. Pierwsza z nich obejmuje kwantowo-chemiczne metody *ab initio*. Kolejną grupę stanowią metody półempiryczne, które bazują na prawach i założeniach mechaniki kwantowej, lecz w obliczeniach wykorzystują dodatkowo dane eksperymentalne. Ostatnią grupę stanowią

metody empiryczne opierające się na formalizmie mechaniki molekularnej, które do obliczeń wykorzystują różne rodzaje pól siłowych.

5. 3. 1. Metody *ab initio*

Metody *ab initio*^{42,43} opierają się na prawach i postulatach mechaniki kwantowej, a do obliczeń wykorzystują uniwersalne stałe fizyczne. Termin *ab initio* w dosłownym tłumaczeniu oznacza „od początku”. Podstawowym założeniem wszystkich metod *ab initio* jest rozwiązanie równania Schrödingera opisywanego poniższym wzorem:

$$\hat{H}\Psi = E\Psi \quad (5.3.1a)$$

gdzie \hat{H} to niezależny od czasu hamiltonian (operator energii całkowitej układu), E oznacza energię całkowitą układu, natomiast ψ jest molekularną funkcją falową zależną od współrzędnych położenia wszystkich jąder i elektronów obecnych w badanym układzie. Równanie to można rozpisać dla dowolnego układu wieloelektronowego, jednak poza atomem wodoru i kilkoma innymi modelowymi układami kwantowo-mechanicznymi jest ono nierozwiązywalne w sposób analityczny. Ma to związek z występującym w hamiltonianie członem opisującym oddziaływania odpychające między elektronami⁴⁴. W celu poradzenia sobie z tą sytuacją opracowano numeryczne rozwiązania bazujące na różnego typu przybliżeniach kwantowo-mechanicznych. Do najpopularniejszych z nich zaliczane są przybliżenie stanu stacjonarnego⁴⁵, adiatyczne przybliżenie Borna-Oppenheimera⁴⁶ oraz przybliżenie jednoelektronowe⁴⁷. Przykładową metodą wykorzystującą wszystkie te trzy typy przybliżeń jest metoda Hartree-Focka zwana także metodą pola samouzgodnionego (ang. *Self-Consistent Field, SCF*)⁴⁷. Metoda ta polega na iteracyjnym minimalizowaniu energii całkowitej układu na drodze odpowiedniej modyfikacji jednoelektronowych funkcji spinorbitalnych. Jej rozszerzeniem są metody post-Hartree-Focka⁴⁸, które bazują na podstawowym formalizmie obliczeniowym metody SCF, lecz w przeciwieństwie do niej uwzględniają także efekt korelacji elektronowej, który znacząco zwiększa precyzję uzyskiwanych wyników obliczeń.

5. 3. 2. Metody półempiryczne

Metody półempiryczne⁴⁴⁻⁴⁹ opierają się na formalizmie obliczeniowym metody SCF⁴⁷, lecz do obliczeń wykorzystują także dane eksperymentalne. Wykorzystanie danych eksperymentalnych umożliwia skorygowanie głównej wady metody SCF związanej z brakiem uwzględnienia efektu korelacji elektronowej. Idea tego typu metod opiera się na całkowitym

zaniebdywaniu całek dwuelektronowych lub zastąpieniu ich odpowiednim zestawem danych eksperymentalnych.

Poszczególne metody półempiryczne różnią się między sobą sposobem parametryzacji oraz możliwością potencjalnych zastosowań. Najnowszą spośród nich jest metoda PM7⁵⁰ (ang.: *Parametric Method 7*), która jest parametryzowana pod kątem jak najlepszej reprodukcji struktur krystalograficznych małych układów molekularnych oraz ich standardowych entalpii tworzenia. Do optymalizacji struktur takich układów wykorzystywana jest bardzo duża baza danych składająca się z wielu związków chemicznych i dostępnych dla nich danych eksperymentalnych. Dzięki temu metoda PM7 odznacza się dużo większą dokładnością i dużo mniejszymi kosztami obliczeniowymi w porównaniu do innych metod półempirycznych oraz kwantowo-chemicznych metod *ab initio*.

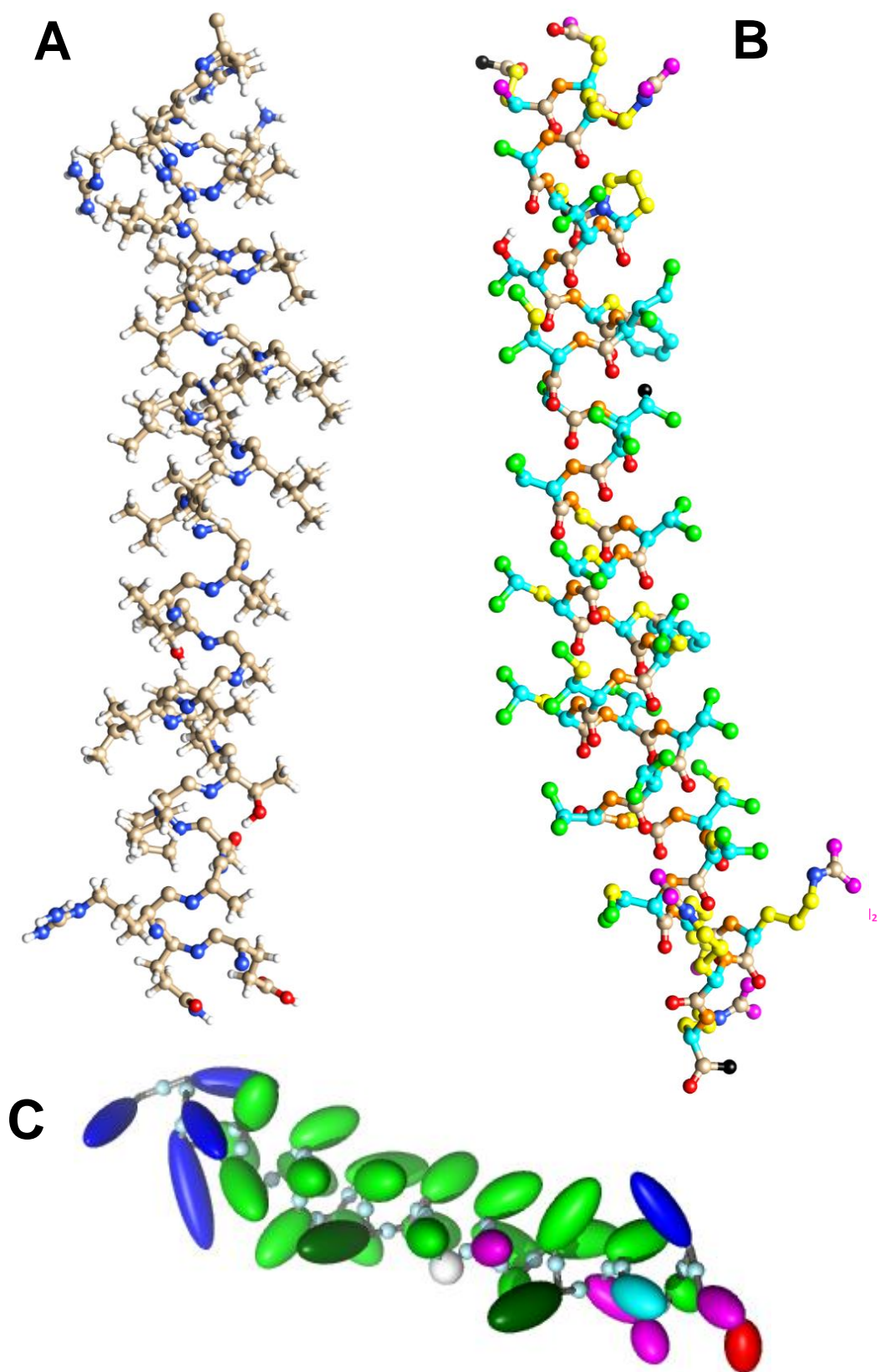
Metody półempiryczne umożliwiają przewidywanie struktur i właściwości wielu małych układów molekularnych⁵¹. Dokładność obliczeniowa metod półempirycznych zależy od stopnia podobieństwa badanych układów do związków stosowanych w parametryzacji każdej z tych metod i jest ona tym większa im mniejszy jest rozmiar badanego układu. Metody półempiryczne są dużo szybsze i mniej kosztowne obliczeniowo od metod *ab initio*, a ich głównym ograniczeniem jest dostępność danych eksperymentalnych dla związków stosowanych w ich parametryzacji.

5. 4. Metody empiryczne mechaniki molekularnej

Metody empiryczne mechaniki molekularnej^{52,53} opierają się na adyabatycznym przybliżeniu Borna-Oppenheimera⁴⁶, a do obliczeń wykorzystują różne rodzaje pól sił. W fizyce mianem pola sił określa się zbiór parametrów i funkcji matematycznych służących do przybliżonego opisu powierzchni energii potencjalnej badanego układu. Parametry te są pozyskiwane na drodze obliczeń kwantowo-mechanicznych (*ab initio* lub półempirycznych) lub na drodze ich dopasowania do danych doświadczalnych.

Pola sił dzielą się na trzy grupy spośród których każda w odmienny sposób opisuje reprezentację przestrzenną badanego układu. Pierwszą z tych grup stanowią pełnoatomowe pola sił (Rys. 8A) takie jak AMBER⁵⁴ oraz CHARMM⁵⁵, w których każdy atom obecny w badanym układzie postrzegany jest jako odrębne centrum oddziaływań. Kolejną grupę stanowią pola zjednoczonych atomów⁵⁶ (Rys. 8B), w których grupa metylowa (-CH₃), metylenowa (-CH₂-) oraz aminowa (-NH₂) są osobnymi centrami oddziaływań. Ostatnią grupę stanowią gruboziarniste pola sił (Rys. 8C), w których rolę centrów oddziaływań mogą pełnić wszystkie atomy reszt aminokwasowych białek, ich łańcuchy boczne,

struktury II-rzędowe lub domeny białek. Popularnymi przykładami tego typu pól sił są UNRES⁵⁷, MARTINI⁵⁸ oraz CABS⁵⁹.



Rys. 8. Projektja przestrzenna białka o kodzie PDB: 1IIJ w reprezentacji pełnoatomowej (A), zjednoczonych atomów (B) oraz gruboziarnistej na przykładzie pola UNRES (C).

Niezależnie od rodzaju użytego do obliczeń pola sił parametry każdego z nich są optymalizowane pod kątem jak najlepszego dopasowania właściwości badanego układu do danych eksperymentalnych takich jak np. geometria równowagowa, energie konformacyjne czy standardowe entalpie tworzenia. Wybór pola sił powinien być dostosowany indywidualnie do badanego układu i oczekiwanej dokładności obliczeniowej. Do symulacji małych cząsteczek związków chemicznych bardziej zasadne jest użycie pełnoatomowych pól sił. Koszty obliczeniowe oraz czas trwania takiej symulacji są wówczas relatywnie niskie w stosunku do wysokiej dokładności obliczeniowej dostarczającej wgląd we względnie niewielkie zmiany strukturalne i energetyczne badanych związków. Do dłuższych symulacji układów biologicznych takich jak białka zalecane jest użycie gruboziarnistych pól sił. Umożliwiają one duże wygładzenie hiperpowierzchni energii potencjalnej badanych związków. Ich niewątpliwym atutem są niskie koszty obliczeniowe, a wadą gorsza dokładność obliczeniowa w stosunku do pełnoatomowych pól sił, która wynika z dużego stopnia uproszczenia sposobu reprezentacji badanych związków.

5. 4. 1. Pełnoatomowe pola sił

Pierwszymi historycznie opracowanymi modelami empirycznymi mechaniki molekularnej były pełnoatomowe pola sił^{60,61}. Cechuje je duża dokładność obliczeniowa, dzięki czemu są one szeroko stosowane w większości prowadzonych obecnie symulacji komputerowych dla małych i prostych związków chemicznych. Wyrażenie na energię potencjalną w pełnoatomowych polach sił takich jak np. AMBER⁵⁴ jest przedstawiane w następujący sposób:

$$\begin{aligned}
 U_{pot} = & \sum_{\text{wi\u0105zan}} k_r (r - r_{eq})^2 + \sum_{\substack{\text{katow} \\ \text{wal.}}} k_\theta (\theta - \theta_{eq})^2 + \sum_{\substack{\text{katow} \\ \text{tors.}}} \frac{1}{2} k_\phi [1 + \cos(n\phi - \phi_0)] \\
 & + \sum_{\substack{\text{van} \\ \text{der} \\ \text{Waals}}} 4\epsilon \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \sum_{\substack{\text{oddz.} \\ \text{elektrost.}}} \frac{q_i q_j}{4\pi\epsilon_v\epsilon_0 r_{ij}}
 \end{aligned} \tag{5.4.1a}$$

Pierwszy człon powyższego równania opisuje energię wynikającą ze zmiany długości wiązań pomiędzy dwoma sąsiadującymi ze sobą atomami względem odpowiadającej im długości równowagowej. Wartość tej energii jest obliczana na podstawie modelu oscylatora harmonicznego jako suma iloczynu stałej siłowej k_r i kwadratu różnicy między równowagową (r_{eq}) i aktualną (r) długością wiązań⁶⁰. W analogiczny sposób obliczana jest energia związana z odchyleniem kątów walencyjnych z ich położenia równowagi – gdzie k_θ jest stałą siłową kąta walencyjnego, natomiast oznaczenia θ_{eq} oraz θ opisują równowagową i aktualną wartość tego

kąta. Trzeci człon powyższego równania opisuje zmiany energii powstałe na skutek lokalnych odkształceń kątów torsyjnych. Wyrażenie to ma postać funkcji okresowej, w której zmienne φ oraz φ_0 odpowiadają aktualnym i referencyjnym wartościom kątów torsyjnych, parametr n określa periodyczność tej funkcji (tj. liczbę lokalnych minimów energetycznych na powierzchni energii potencjalnej badanego związku w jednym pełnym cyklu optymalizacyjnym), natomiast k_φ to stała określana mianem bariery torsyjnej⁶⁰.

Ostatnie dwa człony równania (5.4.1a) reprezentują dalekozasięgowe oddziaływania niewiążące. Pierwszym z nich są oddziaływania van der Waalsa opisywane najczęściej za pomocą potencjału oddziaływań Lennarda-Jonesa⁶², który zostanie szczegółowo omówiony w rozdziale 5.5.1. Kolejny wkład do energii niewiążącej stanowią oddziaływania elektrostatyczne, w których wielkość r_{ij} reprezentuje odległość między atomami i oraz j , q_i oraz q_j są ładunkami elektrycznymi tych atomów, natomiast ε jest parametrem charakteryzującym stałą dielektryczną ośrodka^{60,61}.

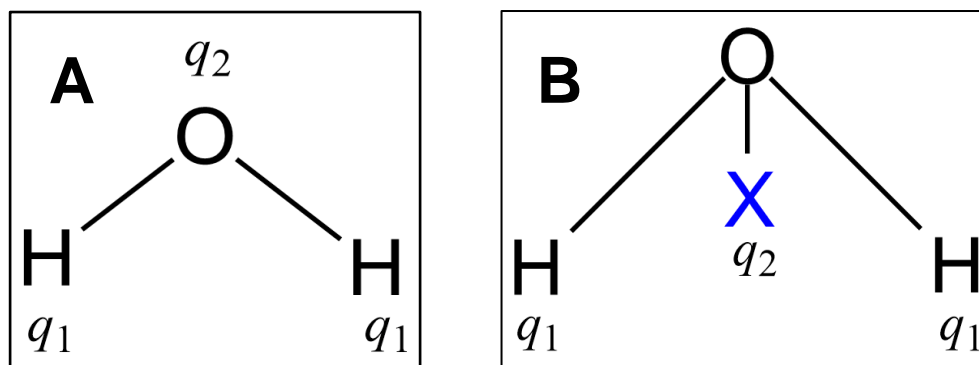
5. 4. 1. 1. Modele wody w pełnoatomowych polach sił

Modele wody^{63,64} stosowane w pełnoatomowych polach sił mają na celu jak najlepsze odtworzenie struktury i właściwości fizykochemicznych prawdziwej wody w symulowanych układach. Rodzaj użytego modelu wody ma duży wpływ na dokładność uzyskiwanych wyników obliczeń. Ze względu na liczbę występujących w nich centrów oddziaływań modele te dzielą się na: dwu-, trój-, cztero-, pięcio- oraz sześciopunktowe. Do najczęściej używanych należą modele trój- oraz czteropunktowe, ponieważ generują stosunkowo niskie koszty obliczeniowe oraz najlepiej spośród wszystkich odtwarzają strukturę i właściwości prawdziwej wody.

Modele trójpunktowe (ang.: *3-site*) (Rys. 9A) są zbudowane z jednego atomu tlenu oraz dwóch atomów wodoru na których rozmieszczone są cząstkowe ładunki elektryczne – jeden ujemny q_2 na dwusiecznej kąta walencyjnego H – O – H oraz po jednym dodatnim q_1 na każdym atomie wodoru. Większość z tych modeli posiada sztywną geometrię dopasowaną parametrycznie do geometrii prawdziwej cząsteczki wody. W swoich badaniach korzystałam z dwóch modeli trójpunktowych takich jak SPC/E⁶⁵ oraz TIP3P-Ew⁶⁶, których szczegółowe właściwości zostały przedstawione w Tab. 1.

Modele czteropunktowe (ang.: *4-site*) (Rys. 9B) są zbudowane z dwóch atomów wodoru, jednego atomu tlenu oraz wirtualnego atomu X zlokalizowanego na dwusiecznej kąta walencyjnego H – O – H. Ten wirtualny atom nie posiada masy i jest obdarzony jedynie cząstkowym ładunkiem ujemnym q_2 , a dwa pozostałe dodatnie ładunki q_1 są rozmieszczone na

każdym z atomów wodoru. Geometria tych modeli jest starannie dopasowana do geometrii prawdziwej cząsteczki wody. Jednym z ich przedstawicieli jest model TIP4P-2005⁶⁷ z którego również korzystałam podczas prowadzonych przez siebie badań; jego szczegółowe właściwości zostały przedstawione w Tab. 1.



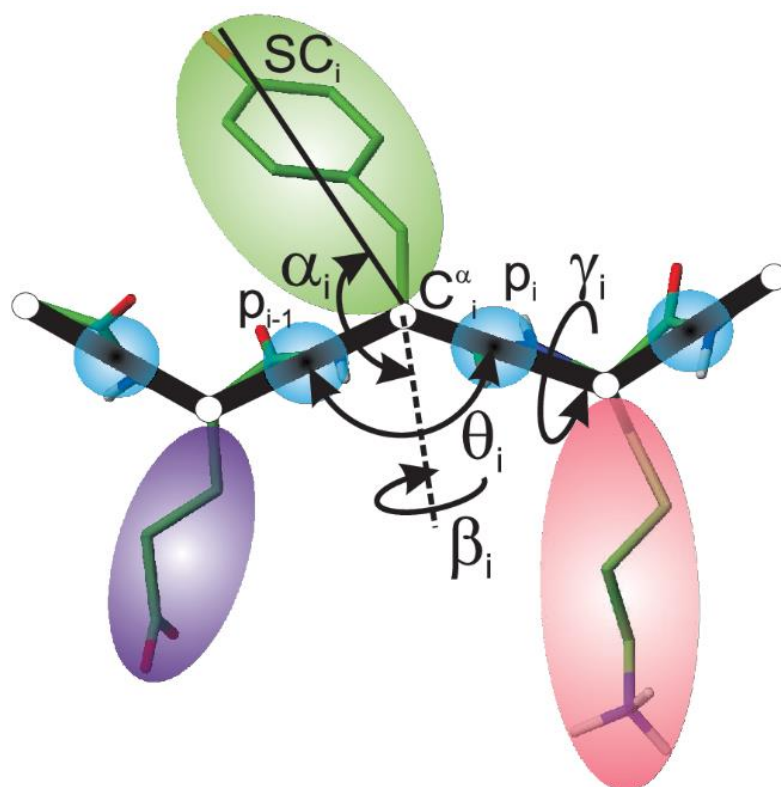
Rys. 9. Schematyczne przedstawienie A) trójpunktowego oraz B) czteropunktowego modelu wody.

Tab. 1. Parametry pola sił dla trójpunktowych modeli SPC/E⁶⁵ i TIP3P-Ew⁶⁶ oraz czteropunktowego modelu TIP4P-2005⁶⁷.

Model	σ [Å]	ϵ [kcal/mol]	$r_{\text{O-H}}$ [Å]	Kąt H-O-H [°]	q_1 [e]	q_2 [e]	$r_{\text{O-X}}$ [Å]
SPC/E	3.166	0.15535	1.0000	109.47	0.4238	-0.8476	-
TIP3P-Ew	3.188	0.10200	0.9572	104.52	0.4150	-0.8300	-
TIP4P-2005	3.1589	0.185201	0.9572	104.52	0.5564	-1.1128	0.1546

5. 4. 2. Gruboziarniste pole sił UNRES

UNRES (ang.: *UNited RESidue*)⁶⁸ jest przykładem gruboziarnistego pola sił opierającego się w całości na fizyce oddziaływań, które zostało opracowane w grupie badawczej Prof. Adama Liwo i Prof. Harolda Scheragi i służy do badania struktur białek w środowisku wodnym. W polu sił UNRES łańcuch polipeptydowy białka jest opisywany w postaci sekwencji atomów C^α połączonych ze zjednoczonymi łańcuchami bocznymi (ang.: *Side-Chain*, SC) oraz zjednoczonymi grupami peptydowymi (ang.: *peptide group*, p) zlokalizowanymi w połowie odległości między dwoma kolejnymi atomami C^α. Każda reszta aminokwasowa tego łańcucha jest reprezentowana przez dwa centra oddziaływań, tj. zjednoczone łańcuchy boczne (SC) oraz zjednoczone grupy peptydowe (p). Atomy C^α są punktami odniesienia służącymi do prawidłowego zdefiniowania geometrii przestrzennej łańcucha głównego białek (Rys. 10).



Rys. 10. Schematyczne przedstawienie łańcucha polipeptydowego białek w gruboziarnistym polu sił UNRES. Kolorem niebieskim oznaczono zjednoczone grupy peptydowe (p), białym atomy C^α , a elipsy oznaczone kolorem różowym, fioletowym i zielonym reprezentują zjednoczone łańcuchy boczne (SC). Zmienna θ_i to kąt walencyjny zawarty między trzema kolejnymi atomami C^α , zmienna γ_i to kąt torsyjny zawarty między czterema kolejnymi atomami C^α , natomiast kąty oznaczone jako α_i oraz β_i opisują położenie zjednoczonych łańcuchów bocznych względem łańcucha głównego białek⁶⁸.

W swoich badaniach korzystałam z dwóch wersji gruboziarnistego pola sił UNRES – starszej OPT-WTFSA-1⁶⁹ oraz najnowszej znanej jako NEWCT-9P³⁹. Najnowsza wersja tego pola sił definiuje wyrażenie na ograniczoną energię swobodną (ang.: *Restricted Free Energy*, RFE) w postaci sumy analitycznych przybliżeń do potencjałów średnich sił (ang.: *Potential of Mean Force*, PMF) łańcuchów polipeptydowych w wodzie zgodnie z następującym równaniem:

$$\begin{aligned}
 U = & w_{SC} \sum_{i < j} U_{SC_i SC_j} + w_{SCp} \sum_{i \neq j} U_{SC_i p_j} + w_{pp}^{VDW} \sum_{i < j-1} U_{p_i p_j}^{VDW} \\
 & + w_{pp}^{el} f_2(T) \sum_{i < j-1} U_{p_i p_j}^{el} + w_{bond} \sum_i U_{bond}(d_i) + w_b \sum_i U_b(\theta_i) \\
 & + w_{tor} f_2(T) \sum_i U_{tor}(\gamma_i, \theta_i, \theta_{i+1}) + w_{corr}^{(3)} f_3(T) \sum_{i < j} U_{corr;ij}^{(3)} \\
 & + w_{turn}^{(3)} f_3(T) \sum_i U_{turn;i}^{(3)} + w_{rot} \sum_i U_{rot}(\theta_i, \alpha_{SC_i}, \beta_{SC_i}) \\
 & + w_{ssbond} \sum_i U_{ssbond;i}
 \end{aligned} \tag{5.4.1a}$$

Natomiast w wersji OPT-WTFSA-1 wyrażenie na ograniczoną energię swobodną RFE jest zdefiniowane w następujący sposób:

$$\begin{aligned}
U = & w_{SC} \sum_{i < j} U_{SC_i SC_j} + w_{SCp} \sum_{i \neq j} U_{SC_i p_j} + w_{pp}^{VDW} \sum_{i < j-1} U_{p_i p_j}^{VDW} \\
& + w_{pp}^{el} f_2(T) \sum_{i < j-1} U_{p_i p_j}^{el} + w_{bond} \sum_i U_{bond}(d_i) + w_b \sum_i U_b(\theta_i) \\
& + w_{corr}^{(3)} f_3(T) \sum_{i < j} U_{corr;ij}^{(3)} + w_{corr}^{(4)} f_4(T) \sum_{i < j} U_{corr;ij}^{(4)} \\
& + w_{turn}^{(3)} f_3(T) \sum_i U_{turn;i}^{(3)} + w_{turn}^{(4)} f_4(T) \sum_i U_{turn;i}^{(4)} \\
& + w_{rot} \sum_i U_{rot}(\alpha_{SC_i}, \beta_{SC_i}) + w_{ssbond} \sum_i U_{ssbond;i} \\
& + w_{tor} f_2(T) \sum_i U_{tor}(\gamma_i) + w_{tord} f_3(T) \sum_i U_{tord}(\gamma_i, \gamma_{i+1}) \\
& + w_{SC-corr} f_2(T) \sum_{m=1}^3 \sum_i U_{SC-corr}(\tau_i^{(m)})
\end{aligned} \tag{5.4.2b}$$

W obydwu powyższych równaniach każdy człon energetyczny U_x wnosi odpowiedni wkład do efektywnej funkcji energii i jest mnożony przez wagi w_x wyznaczone na drodze odpowiedniej kalibracji pola sił UNRES:

- $U_{SC_i SC_j}$ to potencjał opisujący średnią energię swobodną oddziaływań hydrofobowych (hydrofilowych) między zjednoczonymi łańcuchami bocznymi. Potencjał ten domyślnie zawiera także wkłady wynikające z oddziaływań łańcucha bocznego z wodą,
- $U_{SC_i p_j}$ to potencjał wykluczonej objętości dla oddziaływań występujących między zjednoczonymi łańcuchami bocznymi i zjednoczoną grupą peptydową,
- $U_{p_i p_j}^{VDW}$ to potencjał oddziaływań van der Waalsa między dipolami elektrycznymi zjednoczonych grup peptydowych,
- $U_{p_i p_j}^{el}$ to potencjał oddziaływań elektrostatycznych między dipolami elektrycznymi zjednoczonych grup peptydowych,
- $U_{bond}(d_i)$ to potencjał opisujący zmiany energetyczne powstałe w wyniku odkształceń wirtualnych długości wiązań d_i między atomem C^α oraz zjednoczonym łańcuchem bocznym,
- $U_b(\theta_i)$ to potencjał opisujący geometrię łańcucha głównego białek, która wynika z odkształceń wirtualnych kątów walencyjnych θ_i ,

- $U_{tor}(\gamma_i, \Theta_i, \Theta_{i+1})$ to potencjał opisujący geometrię łańcucha głównego białek, który wynika z odkształceń wirtualnych kątów torsyjnych γ_i oraz wirtualnych kątów walencyjnych Θ_i . W wersji OPT-WTFSA-1 potencjał ten zależał wyłącznie od zmiany wartości kątów torsyjnych γ_i między czterema kolejnymi atomami C^α w łańcuchu głównym białek,
- $U_{corr;ij}^{(3)}$ oraz $U_{corr;ij}^{(4)}$ to potencjały opisujące wzajemną korelację oddziaływań elektrostatycznych i lokalnych między grupami peptydowymi oraz lokalnymi stanami konformacyjnymi reszt aminokwasowych w obrębie łańcucha głównego białek,
- $U_{turn;i}^{(3)}$ oraz $U_{turn;i}^{(4)}$ to potencjały skrętne, które opisują wzajemną korelację oddziaływań elektrostatycznych między grupami peptydowymi oddalonymi o trzy i cztery reszty aminokwasowe w łańcuchu głównym białek,
- $U_{rot}(\Theta_i, \alpha_{SCi}, \beta_{SCi})$ to potencjał opisujący stany rotameryczne zjednoczonych łańcuchów bocznych, które zależą od wartości wirtualnych kątów walencyjnych Θ_i oraz kątów α_{SCi} i β_{SCi} opisujących położenie środka zjednoczonych łańcuchów bocznych względem płaszczyzny zawierającej sąsiadujące ze sobą atomy C_i^α . W wersji OPT-WTFSA-1 potencjał ten był zależny jedynie od wartości kątów α_{SCi} oraz β_{SCi} ,
- $U_{ssbond;i}$ to potencjał opisujący energię mostków siarczkowych w funkcji odległości (d_i^{ssbond}) oddziałujących ze sobą zjednoczonych łańcuchów bocznych reszt cystein,
- $U_{tord}(\gamma_i, \gamma_{i+1})$ to potencjał występujący w wersji OPT-WTFSA-1 pola sił UNRES. Opisuje on zmiany energii łańcucha głównego białek będące wynikiem lokalnych odkształceń dwóch kolejnych kątów torsyjnych między wiązaniami wirtualnymi,
- $U_{SC-corr}(\tau_i^{(m)})$ to potencjał występujący w wersji OPT-WTFSA-1 pola sił UNRES. Opisuje on sprzężenie łańcucha głównego białek z lokalnymi łańcuchami bocznymi. Sprzężenie to opisywane jest za pomocą kątów torsyjnych $\tau_i^{(m)}$, gdzie indeks m oznacza typ kąta dwuściennego zawartego między atomami C^α w łańcuchu głównym i sprzęgniętym z nim łańcuchem bocznym białek⁷⁰.

Mnożniki $f_m(T)$ występujące w obu powyższych równaniach charakteryzują temperaturową zależność poszczególnych wkładów energetycznych U_x do ograniczonej energii swobodnej. Są one opisywane za pomocą poniższej zależności matematycznej:

$$f_m = \frac{\ln[\exp(1) + \exp(-1)]}{\ln\left[\exp\left(\left(\frac{T}{T_0}\right)^{n-1}\right) + \exp\left(-\left(\frac{T}{T_0}\right)^{n-1}\right)\right]} \quad (5.4.2c)$$

gdzie m jest rzędem skalowanego wkładu do energii RFE, T jest temperaturą bezwzględną, a T_0 przyjmuje wartość równą 300 K. Dla temperatury T_0 wartość czynnika skalującego $f_m(T)$ jest równa 1 i rośnie wraz ze spadkiem temperatury.

Gruboziarniste pole sił UNRES umożliwia znaczne skrócenie czasu trwania symulacji dla białek i peptydów w porównaniu do symulacji z użyciem pełnoatomowych pól sił stosujących jawną postać rozpuszczalnika. Zostało ono wykorzystane m. in. do zbadania dynamiki zwijania włókien β -amyloidowych⁷¹, peptydu 1E0L oraz 1BDD⁷², przewidywania struktury białek oligomerycznych⁷³, przestrzennego ułożenia składowych domen w białku PICK1⁷⁴, dynamiki przejścia konformacyjnego białka opiekuńczego Hsp70⁷⁵, a także przewidywania struktury białek membranowych w środowisku lipidowym⁷⁶.

5. 4. 2. 1. Optymalizacja pola sił UNRES

Optymalizacja pola sił UNRES polega na zwiększeniu jego zdolności do przewidywania struktur białek w środowisku wodnym³⁹. Do optymalizacji wykorzystuje się zbiór kilku różnych białek o małej liczbie reszt aminokwasowych oraz różnych typach struktur II-rzędowych. Pełny cykl optymalizacji składa się z wielu następujących po sobie iteracji. W każdej z nich dla zadanych parametrów pola sił symulowane są nowe zespoły konformacyjne badanych białek. Parametry te są kalibrowane przy pomocy metody H-REMD⁷⁷ oraz WHAM^{78,79}, a następnie optymalizowane przy pomocy metody maksymalnego podobieństwa (ang.: *Maximum Likelihood*, MAXLIKE)⁸⁰. Zadaniem metody MAXLIKE jest jak najlepsze dopasowanie do siebie rozkładów współrzędnych przestrzennych pochodzących ze zbioru symulowanych i eksperymentalnych zespołów konformacyjnych badanych białek. Podobieństwa obydwu tych rozkładów są opisywane za pomocą funkcji optymalizacji maksymalnego prawdopodobieństwa l^* danej jako:

$$l^* = \sum_p \sum_i W_{ip} \left[\ln Q_p(T_{ip}) - \sum_j a_{jip} \left(\omega_{jp} - \frac{U_{jp}}{RT_{ip}} \right) \right] \quad (5.4.2.1a)$$

gdzie R to stała gazowa, zmienna i to i -ta temperatura j -tej konformacji symulowanego białka p , człon W_{ip} to wkład wagowy tego białka w temperaturze T_{ip} do funkcji

optymalizacji l^* , U_{jp} to energia UNRES dla j -tej konformacji białka p . Wielkości ω_{jp} oraz Q_p są wyznaczone na drodze odpowiedniego przetworzenia wyników symulacji H-REMD przy pomocy metody WHAM opisanej w dalszej części pracy. Pierwsza z tych wielkości opisuje częstość zliczeń j -tej konformacji białka p w temperaturze T_{ip} , natomiast Q_p to konfiguracyjna funkcja podziału dla tego białka.

Podobieństwo j -tej symulowanej konformacji białka p do k -tej eksperymentalnej konformacji tego białka w temperaturze T_{ip} jest opisywane przez wielkość a_{jip} w następujący sposób:

$$a_{jip} = \frac{1}{A_{ip}} \sum_k^{n_{ip}} \frac{\exp \left[-\frac{d(\mathbf{X}_{kip}, \mathbf{Y}_{jp})^2}{2s_p^2} \right]}{\sum_l \exp \left[-\frac{d(\mathbf{Y}_{jp}, \mathbf{Y}_{lp})^2}{2s_p^2} \right]} \omega_{kip} \quad (5.4.2.1b)$$

gdzie n_{ip} to liczba eksperymentalnych konformacji białka p , zmienna k to k -ta eksperymentalna konformacja białka p , ω_{kip} to częstość zliczeń tej konformacji w temperaturze T_{ip} , natomiast A_{ip} to czynnik normalizacyjny białka p :

$$A_{ip} = \sum_j \sum_k \frac{\exp \left[-\frac{d(\mathbf{X}_{kip}, \mathbf{Y}_{jp})^2}{2s_p^2} \right]}{\sum_k \exp \left[-\frac{d(\mathbf{Y}_{jp}, \mathbf{Y}_{kp})^2}{2s_p^2} \right]} \omega_{kip} \quad (5.4.2.1c)$$

$$\omega_{kip} = \frac{1}{Q_p(T_{ip})} \exp \left[\sum_j a_{jip} \left(\omega_{jp} - \frac{U_{jp}}{RT_{ip}} \right) \right] \quad (5.4.2.1d)$$

Wielkość $d(\mathbf{X}, \mathbf{Y})$ w powyższych równaniach opisuje odległość między symulowaną i eksperymentalną konformacją białka p o współrzędnych przestrzennych zebranych w wektorach \mathbf{X} i \mathbf{Y} . Odległości te są liczone przy pomocy miary RMSD dla atomów C^α oraz środków zjednoczonych łańcuchów bocznych w porównywanych ze sobą konformacjach. Parametr s_p opisuje odchylenie standardowe funkcji Gaussa dla białka p . Wartość parametru s_p zależy od wielkości badanego białka i zazwyczaj plasuje się w zakresie od 1 do 2 Å. Jeśli wartość s_p jest mniejsza od miary RMSD dla porównywanych ze sobą struktur badanego białka to prawdopodobieństwo, że są one do siebie podobne nie przekracza 33%.

W całym cyklu optymalizacyjnym pola sił UNRES wartość funkcji l^* (równanie 5.4.2.1a) powinna się systematycznie zmniejszać ze wzrostem liczby kolejnych iteracji.

Dopiero wówczas, gdy wartości funkcji l^* przestaną się zmniejszać i zaczną osiągać względnie stałe wartości to proces optymalizacji pola sił UNRES ulega zakończeniu.

5. 5. Dynamika molekularna

Metoda dynamiki molekularnej (ang.: *Molecular Dynamics*, MD)^{81,82} polega na prowadzeniu zależnych od czasu symulacji właściwości termodynamicznych badanego układu. Jej rezultatem jest seria danych zapisywanych w jednostce czasu zwana trajektorią. Dane te są następnie konwertowane przy pomocy mechaniki statystycznej do mierzalnych wielkości makroskopowych. Dzięki temu możliwe jest porównanie wyników symulacji MD z danymi eksperymentalnymi. Symulacje MD w połączeniu z gruboziarnistymi polami sił umożliwiają badanie ścieżek zwijania białek⁸³, a także są wykorzystywane do nowoczesnego projektowania leków⁸⁴ i nanoobjektów⁸⁵ oraz do przewidywania właściwości fizykochemicznych badanych układów (np. współczynników dyfuzji, przewodności elektrycznej oraz lepkości dynamicznej⁸⁶).

W metodzie MD powierzchnia energii potencjalnej badanego układu jest wyznaczana na drodze numerycznego całkowania równań ruchu opierających się na II prawie dynamiki Newtona:

$$F_i = m_i \frac{d^2 y_i}{dt^2} \quad (5.5a)$$

gdzie F_i to składowe sił działające na i -tą cząstkę (centrum oddziaływań) o masie m_i , natomiast $d^2 y_i / dt^2$ to drugie pochodne położenia y_i i -tej cząstki po czasie t . Analityczne rozwiązanie powyższego równania jest bardzo trudne (o ile w ogóle możliwe) z uwagi na sprzężoną naturę dość chaotycznych ruchów wszystkich cząstek w symulowanym układzie. Dlatego w praktyce obliczeniowej powyższe równanie jest rozwiązywane na sposób numeryczny przy pomocy tzw. algorytmów integracyjnych.

Podstawową metodą numeryczną do całkowania równań ruchu w symulacjach MD jest klasyczny algorytm Verleta (ang.: *The Basic Verlet Algorithm*), który wykorzystuje przyspieszenie a_i dla i -tej cząstki (centrum oddziaływań) w czasie t oraz sumuje jej położenia w czasie $t - \Delta t$ oraz $t + \Delta t$ w następujący sposób:

$$y_i(t + \Delta t) = 2y_i(t) - y_i(t - \Delta t) + a(t)\Delta t^2 \quad (5.5b)$$

Zaletą tego algorytmu jest duża prostota obliczeniowa i szybkość, a jego główną wadą umiarkowana precyzja oraz brak możliwości bezpośredniego wyznaczenia prędkości cząstek (centrów oddziaływań) w aktualnym kroku czasowym. Popularnym algorytmem eliminującym wadę tej metody numerycznej jest prędkościowy algorytm Verleta (ang.: *The Velocity Verlet*

Algorithm). Jest on stosowany m. in. do całkowania uogólnionych równań ruchu dla i -tych centrów oddziaływań w dynamice Lagrange’a w symulacjach z gruboziarnistym polem sił UNRES⁸⁷:

$$y_i(t + \Delta t) = y_i(t) + v_i(t)\Delta t + \frac{a_i(t)\Delta t^2}{2} \quad (5.5c)$$

$$v_i(t + \Delta t) = v_i\left(t + \frac{1}{2}\Delta t\right) + \frac{a_i(t + \Delta t)\Delta t}{2} \quad (5.5d)$$

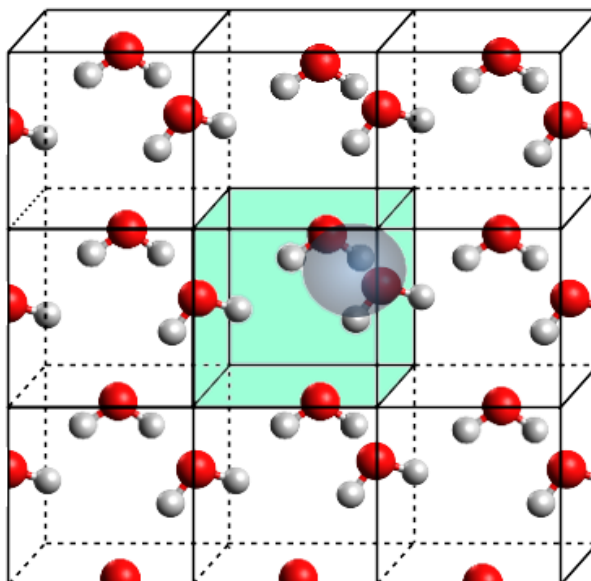
$$a_i(t) = -G^{-1}\nabla_{y_i(t)}U[y_i(t)] \quad \text{gdzie} \quad \mathbf{G} = \mathbf{A}^T\mathbf{M}\mathbf{A} + \mathbf{H} \quad (5.5e)$$

gdzie U to efektywna funkcja energii badanego układu w polu sił UNRES, \mathbf{G} to macierz bezwładności i -tych centrów oddziaływań, \mathbf{H} jest częścią macierzy \mathbf{G} opisującą ruchy rozciągające wirtualne wiązania między i -tymi centrami oddziaływań, \mathbf{A}^T jest macierzą transformacji opisującą przekształcenie uogólnionych prędkości v oraz położeń y do współrzędnych przestrzennych oraz prędkości i -tych centrów oddziaływań, a \mathbf{M} to macierz diagonalna zbudowana z mas i -tych centrów oddziaływań.

Innym istotnym czynnikiem dla przebiegu symulacji MD jest długość kroku czasowego, którego wartość zależy od rodzaju użytego do obliczeń pola sił. W pełnoatomowych polach sił długość kroku czasowego nie może przekraczać częstości oscylacyjnej obserwowanej dla najszybszego drgania w badanym związku i jest ona zwykle ograniczona do 2 femtosekund⁸⁸. Uniemożliwia to dokładne przeszukiwanie przestrzeni konformacyjnej białek i generuje spore koszty obliczeniowe takich symulacji. W symulacjach białek z użyciem gruboziarnistych pól sił ustalenie optymalnej wartości kroku czasowego nie jest tak jednoznaczne z uwagi na duży stopień uproszczenia reprezentacji tych związków. Długości kroków czasowych takich symulacji mogą sięgać nawet do 15 femtosekund⁸⁹ i są one ograniczane jedynie przez odpowiednie algorytmy numeryczne np. VTS⁸⁹ (ang.: *Variable Time Step*), który skraca krok czasowy symulacji, wówczas gdy energia kinetyczna wszystkich centrów oddziaływań przekroczy pewną dozwoloną wartość graniczną.

Przy wyborze parametrów symulacji MD konieczne jest również ustalenie typu warunków brzegowych – periodycznych lub nieperiodycznych – dla modelowanego układu. Technika periodycznych warunków brzegowych (ang.: *Periodic Boundary Conditions*, PBC) (Rys. 11) polega na dokładnym powielaniu pudła symulacyjnego układu. Wymaga to ustalenia tzw. promienia odcięcia (ang.: *cut-off*), czyli maksymalnej odległości na jaką mogą oddziaływać ze sobą cząstki (centra oddziaływań) w pudle symulacyjnym i/lub jego dokładnej replice. Oddziaływania te są możliwe jedynie wówczas, gdy każda z nich znajdzie się w obrębie zadanego promienia odcięcia. Cząstka, która opuszcza pudło symulacyjne pojawi

się po przeciwnej jego stronie w jednej z jego replik, lecz ze zmienionymi parametrami swoich współrzędnych przestrzennych.



Rys. 11. Schematyczne przedstawienie zachowania cząsteczki wody w symulowanym układzie z zastosowaniem periodycznych warunków brzegowych. Kolorem seledynowym oznaczono pudło symulacyjne, a pozostałe obszary przedstawiają jego dokładne repliki. Promień odcięcia między dwiema cząstkami został oznaczony kolorem szarym.

Symulacje dynamiki molekularnej mogą być prowadzone w różnych zespołach statystycznych odzwierciedlających warunki panujące w badanym układzie⁹⁰. Większość symulacji układów biologicznych jest prowadzona w zespole kanonicznym umożliwiającym dokładne odtworzenie warunków panujących w ich środowisku naturalnym. Do kontroli temperatury w symulacjach MD opracowano różnego typu algorytmy termostatowania. Zadaniem każdego z nich jest utrzymanie wewnątrz badanego układu względnie stałej wartości temperatury podczas trwania całej symulacji. Jedną z najczęściej stosowanych technik termostatowania jest termostat Berendsena⁹¹. Jego zasada działania polega na połączeniu symulowanego układu z zewnętrznym źródłem ciepła, który dokonuje skalowania prędkości cząstek (lub centrów oddziaływań) w każdym kroku czasowym symulacji. Prędkości te skalowane są w taki sposób, by aktualna temperatura symulowanego układu (T_k) odpowiadała zadanej temperaturze panującej w termostacie (T_0):

$$v_i(t + \Delta t) = \sqrt{\left[1 + \frac{\Delta t}{\tau} \left(\frac{T_0}{T_k} - 1\right)\right]} v_i(t) \quad (5.5f)$$

gdzie τ_T to parametr empiryczny opisujący stałą sprzężenia symulowanego układu z termostatem, a Δt to długość kroku czasowego symulacji. Jest to dość prosty i numerycznie

stabilny algorytm umożliwiający szybkie i efektywne doprowadzenie symulowanego układu do stanu równowagi termodynamicznej. Towarzyszą temu nieakceptowalne fluktuacje energii całkowitej układu, co skutkuje tym, że opisywany sposób termostatowania nie generuje poprawnego rozkładu kanonicznego dla badanego układu.

Inną koncepcję do regulacji temperatury w symulowanym układzie reprezentuje termostat Langevina⁹². Jest on często stosowany w symulacjach z gruboziarnistym polem sił UNRES i polega na połączeniu każdego j -tego centrum oddziaływań z zewnętrznym zbiornikiem ciepła. W każdym kroku czasowym symulacji na wszystkie te centra działa pewna losowa siła oraz tarcie, które powodują zmniejszenie prędkości ich poruszania się w badanym układzie:

$$m_j \vec{a}_j = -\vec{F}_j - m_j \lambda \vec{v}_j + \vec{W}_j \quad (5.5g)$$

gdzie λ jest współczynnikiem siły tarcia, F_j to składowe wszystkich sił działających na j -te centrum oddziaływań, natomiast W_j to losowa siła dana jako:

$$\langle W_j(0)W_j(t) \rangle = 2\lambda m_i k_B T \delta(t) \quad (5.5h)$$

Termostat Langevina generuje poprawny rozkład kanoniczny dla symulowanych układów, a także pozwala na stosowanie dłuższych kroków czasowych w porównaniu do niestochastycznych technik termostatowania. Zadaniem tego termostatu w symulacjach z użyciem pola sił UNRES jest doprowadzenie badanego układu do stanu równowagi termodynamicznej. W celu przyspieszenia tych symulacji współczynnik λ jest skalowany przez wartość z zakresu od 0.001 do 0.1⁹³.

5. 5. 1. Potencjał oddziaływań Lennarda-Jonesa vs pseudotwardych kul Mie

Rodzaj użytego do obliczeń potencjału oddziaływań ma bardzo duży wpływ na dokładność wyników symulacji MD. Potencjał oddziaływań jest funkcją opisującą zależność energii potencjalnej symulowanego układu od odległości oddziałujących w nim cząstek (centrów oddziaływań). Najpopularniejszym spośród nich jest potencjał Lennarda-Jonesa (Rys. 12)^{81,82} dany jako:

$$U_{LJ}(r_{ij}) = 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (5.5.1a)$$

gdzie $U(r_{ij})$ to energia oddziaływań pomiędzy parą cząstek i oraz j znajdujących się w odległości r względem siebie. Parametrem opisującym głębokość studni tego potencjału

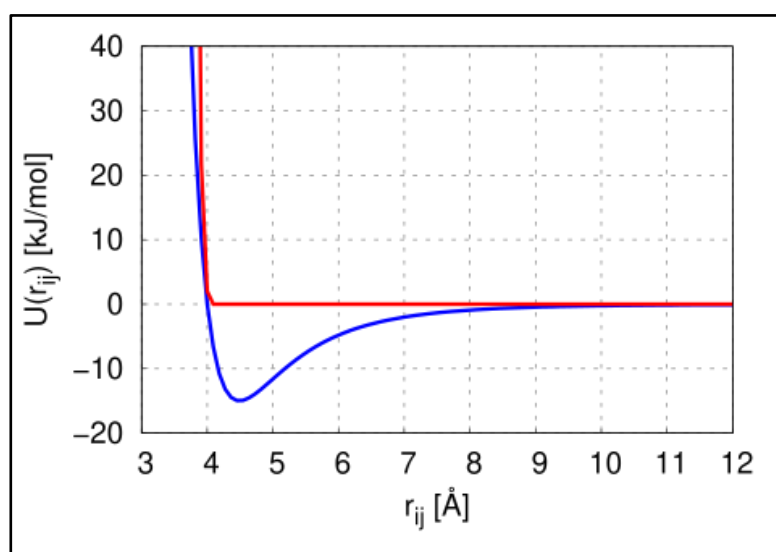
jest ϵ_{ij} , natomiast σ_{ij} opisuje średnicę oddziałujących ze sobą cząstek. Obydwie te wielkości są wyznaczone przy pomocy równań Barthelota-Lorentza⁹⁴ w następujący sposób:

$$\epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}} \quad \sigma_{ij} = \frac{\sigma_{ii} + \sigma_{jj}}{2} \quad (5.5.1b)$$

Pierwszy człon równania (5.5.1a) odnosi się do oddziaływań odpychających (r^{-12}), a drugi (r^{-6}) ma związek z oddziaływaniami przyciągającymi. Potencjał Lennarda-Jonesa jest uogólnioną formą dużo od niego starszego, lecz mniej spopularyzowanego potencjału twardych kul Mie⁹⁵ (Rys. 12) mającego następującą postać:

$$U_{\text{Mie}}(r_{ij}) = \begin{cases} 50 \left(\frac{50}{49}\right)^{49} \epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{50} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{49} \right] + \mathbf{C}, & r_{ij} < \frac{50}{49}\sigma_{ij} \\ 0, & r_{ij} \geq \frac{50}{49}\sigma_{ij} \end{cases} \quad (5.5.1c)$$

gdzie zmienna \mathbf{C} opisuje energię oddziałujących ze sobą cząstek w ustalonym zakresie wartości promienia odcięcia ($r_{\text{cut-off}}$). Zastosowanie tej zmiennej sprawia, że dla minimalnej odległości oddziałujących ze sobą cząstek ich energia potencjalna $U(r_{ij})$ przyjmuje wartości różne od zera.



Rys. 12. Wykresy przedstawiające zależności zmian energii potencjalnej $U(r)$ względem odległości oddziałujących ze sobą cząstek. Kolorem niebieskim oznaczono potencjał Lennarda-Jonesa, a czerwonym potencjał twardych kul Mie.

Wzrostowi odległości dwóch oddziałujących ze sobą cząstek towarzyszy spadek wartości energii potencjalnej układu, który jest szybszy i bardziej gwałtowny w przypadku potencjału twardych kul Mie. Potencjał ten nie posiada żadnego minimum, a najmniejsza wartość jaką może osiągnąć jest równa zero. Natomiast potencjał Lennarda-Jonesa osiąga minimum w punkcie $U(r) = -\epsilon_{ij}$, w którym siły przyciągania są równoważone przez siły odpychania dla

oddziałujących ze sobą cząstek. Dopiero po przekroczeniu tej odległości wartość energii potencjalnej układu dąży do zera.

5. 6. Metody poprawy próbkowania w symulacjach dynamiki molekularnej

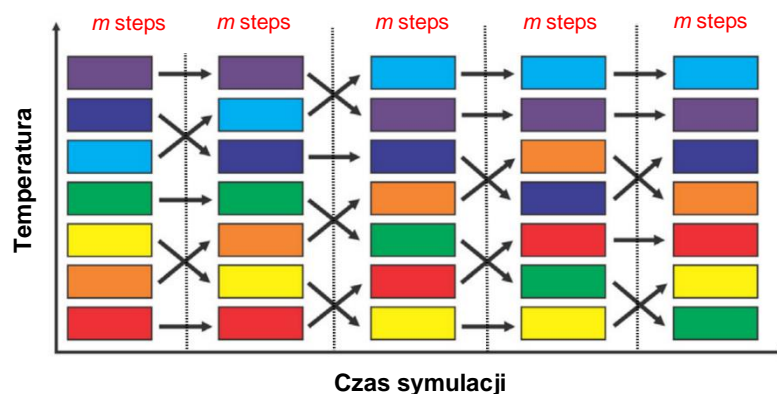
Konwencjonalne symulacje MD prowadzone dla układów biologicznych mają bardzo duże trudności z pokonywaniem wysokich barier energetycznych występujących na hiperpowierzchni energii potencjalnej takich układów. W celu dokładnego przeszukiwania przestrzeni konformacyjnej układów biologicznych opracowano kilka rozszerzonych wariantów metody MD, które znacznie redukują czas trwania takich symulacji.

Jednym z takich wariantów jest metoda dynamiki molekularnej z wymianą replik (ang.: *Replica Exchange Molecular Dynamics*, REMD)^{96,97} (Rys. 13). Metoda ta opiera się na stochastycznym algorytmie Metropolis-Hastingsa⁹⁸ i polega na równoczesnym symulowaniu wielu replik badanego układu różniących się między sobą temperaturą. Po zadanej liczbie kroków MD dokonywane są próby wymiany temperatur między sąsiednimi replikami z prawdopodobieństwem ρ obliczonym na podstawie kryterium Metropolis⁹⁸. W symulacjach REMD z użyciem gruboziarnistego pola sił UNRES prawdopodobieństwo ρ dla wymiany temperatur między sąsiednimi replikami a i b jest określone w następujący sposób⁹⁹:

$$\rho(Y_a \rightarrow Y_b) = \min [1, \exp(-\Delta)] \quad (5.6a)$$

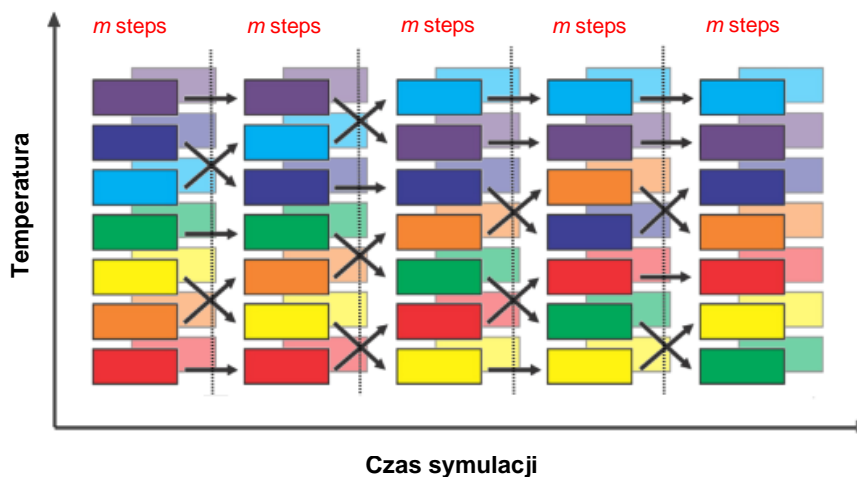
$$\Delta = [\beta_b V_b(\mathbf{Y}_b; \beta_b) - \beta_a V_a(\mathbf{Y}_b; \beta_a)] - [\beta_b V_b(\mathbf{Y}_a; \beta_b) - \beta_a V_a(\mathbf{Y}_a; \beta_a)] \quad (5.6b)$$

gdzie $\beta_a = (RT_a)^{-1}$, R to stała gazowa, T jest temperaturą a -tej repliki, V to funkcja energii potencjalnej dla a -tej repliki (zawierająca potencjały ograniczające nakładane na symulowaną strukturę białka), \mathbf{Y}_a to wektor współrzędnych przestrzennych konformacji z a -tej repliki w punkcie próbnym wymiany temperatur. Kryterium Metropolis⁹⁸ jest spełnione dla przypadku, gdy wartość Δ będzie mniejsza lub równa zero. W przeciwnej sytuacji generowana jest liczba pseudolosowa x z zakresu od 0 do 1, która jest porównywana z wielkością $\exp(-\Delta)$. Próba wymiany temperatur między dwiema replikami jest akceptowana, gdy liczba pseudolosowa x będzie mniejsza lub równa $\exp(-\Delta)$. W pozostałych przypadkach próba wymiany temperatur zostaje odrzucona.



Rys. 13. Schematyczne przedstawienie przebiegu symulacji dynamiki molekularnej z wymianą replik (REMD)¹⁰⁰.

Rozszerzeniem metody REMD jest dynamika molekularna ze zwielokrotnioną wymianą replik (ang.: *Multiplexed-Replica Exchange Molecular Dynamics*, MREMD)¹⁰¹. Metoda ta polega na jednoczesnym i wielokrotnym powieleniu liczby replik dla każdej temperatury, co prowadzi do utworzenia serii wielu równoległych względem siebie płaszczyzn (Rys. 14). Metoda MREMD oferuje możliwość wymiany temperatur między sąsiednimi replikami leżącymi w tej samej oraz sąsiadujących ze sobą płaszczyznach. Umożliwia to bardziej dokładne próbkowanie przestrzeni konformacyjnej badanego układu w porównaniu do metody REMD.



Rys. 14. Schematyczne przedstawienie przebiegu symulacji dynamiki molekularnej ze zwielokrotnioną wymianą replik (MREMD)¹⁰⁰.

Kolejną metodą oferującą zauważalną poprawę sposobu próbkowania przestrzeni konformacyjnej badanego układu jest dynamika molekularna z hamiltonowską wymianą replik (ang.: *Hamiltonian-Replica Exchange Molecular Dynamics*, H-REMD)⁷⁷ (Rys. 15). Metoda ta polega na dwuwymiarowym powieleniu liczby replik różniących się temperaturą lub potencjałem ograniczającym o określonej wadze siły więzów, który wnosi stosowny wkład do

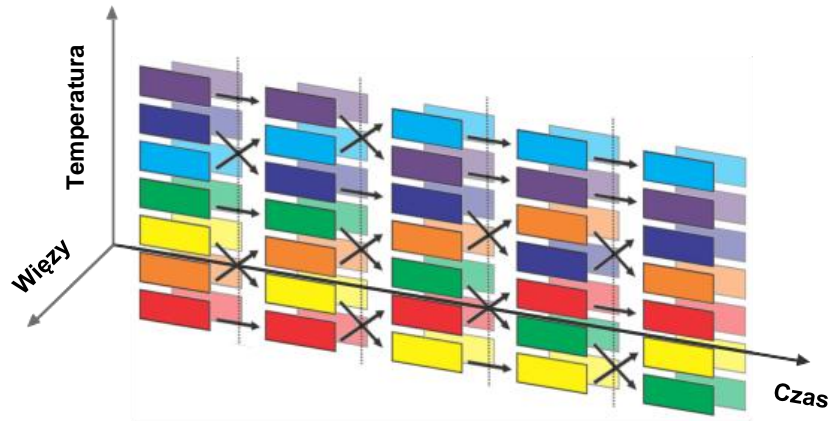
wartości energii potencjalnej badanego układu (lub wartości RFE w gruboziarnistym polu sił UNRES). Temperatura oraz parametry energii są wymieniane losowo lub w określonym kierunku w regularnych odstępach czasu między sąsiednimi replikami leżącymi na tej samej lub sąsiedniej płaszczyźnie. W symulacjach H-REMD z gruboziarnistym polem sił UNRES prawdopodobieństwo ρ tych wymian jest określone w następujący sposób⁹⁹:

$$\rho(Y_{ab} \rightarrow Y_{cd}) = \min [1, \exp(-\Delta)] \quad (5.6c)$$

$$(c, d) \in (a + 1, b), (a, b + 1), (a + 1, b + 1) \quad (5.6d)$$

$$\Delta = [\beta_c V_d(\mathbf{Y}_{cd}; \beta_c) - \beta_a V_b(\mathbf{Y}_{cd}; \beta_a) - [\beta_c V_d(\mathbf{Y}_{ab}; \beta_c) - \beta_a V_a(\mathbf{Y}_{ab}; \beta_a)]] \quad (5.6e)$$

gdzie zmienne a, b, c, d to indeksy określonych replik, $\beta_a = (RT_a)^{-1}$, $\beta_c = (RT_c)^{-1}$, R jest stałą gazową, T_a oraz T_c to wartości temperatury dla repliki a oraz c , V_b oraz V_d to wartości energii potencjalnej uwzględniające potencjał ograniczający o wadze siły więzów dla repliki b oraz d , natomiast \mathbf{Y}_{ab} oraz \mathbf{Y}_{cd} to zestawy zmiennych, które opisują stan trajektorii w temperaturze T_a i sile więzów z repliki b oraz w temperaturze T_c i sile więzów z repliki d .



Rys. 15. Schematyczne przedstawienie przebiegu symulacji dynamiki molekularnej z hamiltonowską wymianą replik (H-REMD)¹⁰⁰.

5. 7. Metody analizy danych

Podstawowym celem prowadzenia każdej symulacji jest wygenerowanie za jej pomocą jak najbardziej dokładnych wyników obliczeń dających wgląd w termodynamikę oraz strukturę badanego układu w warunkach zbliżonych do naturalnych. Bezpośrednie uzyskanie tych informacji na podstawie samej trajektorii (jednej lub wielu) jest niemożliwe do zrealizowania. Dlatego do interpretacji wyników symulacji opracowano odpowiednie metody obliczeniowe, których zadaniem jest przetworzenie danych z trajektorii w taki sposób, by wyodrębnić z nich możliwie jak najwięcej informacji nt. struktury, stabilności i termodynamiki badanego układu.

5. 7. 1. Metoda ważonych histogramów

Metoda ważonych histogramów (ang.: *Weighted Histogram Analysis Method*, WHAM)^{78,79} jest jedną z metod służących do analizy i przetwarzania wyników symulacji komputerowych. Dzięki odpowiedniemu przetworzeniu danych z trajektorii badanego układu możliwe jest m. in. określenie potencjału średniej siły dla oddziałujących ze sobą cząstek (lub centrów oddziaływań). Podstawy tej metody zostały opracowane w 1989 r. przez Swendsena i Ferrenberga¹⁰², a kilka lat później dwójka innych badaczy Kumar i Rosenberg po raz pierwszy zastosowali ją do interpretacji wyników symulacji wykonanej dla deoksyadenozyny¹⁰³.

Metoda WHAM polega na pogrupowaniu wszystkich konformacji badanego związku do osobnych histogramów względem temperatury symulacji. Każdy z histogramów opisuje rozkład prawdopodobieństwa energii całkowitej E tego związku w określonej temperaturze. W dalszej kolejności histogramy te są łączone ze sobą w taki sposób by ich obszary mniej lub bardziej pokrywały się ze sobą. Dzięki temu w każdej z temperatur wszystkie stany energetyczne badanego związku będą próbkowane z identycznym prawdopodobieństwem. W wyniku tego otrzymujemy pojedynczy płaski histogram $H(E)$ na podstawie którego możemy określić prawdopodobieństwo znalezienia badanego związku w stanie o energii całkowitej E w sposób następujący:

$$\rho(\beta, E) = \frac{\sum_{i=1}^M H(E_i) \exp(-\beta E_i)}{\sum_{j=1}^M N_j \exp(f_j - \beta d_j)} \quad (5.7.1a)$$

gdzie $\beta = (k_B T)^{-1}$, E_i to energia całkowita / współrzędne i -tej konformacji, N_j to liczba konformacji badanego związku z j -tej trajektorii, f_j to bezwymiarowa energia swobodna j -tej trajektorii, M to ogólna liczba konformacji zebranych z wszystkich trajektorii w toku danej symulacji, natomiast $H(E_i)$ to histogram dla i -tej struktury.

Metoda WHAM umożliwia analizę trajektorii uzyskanych z symulacji dynamiki molekularnej^{81,82} oraz jej rozszerzeń z wymianą replik (REMD)^{96,97}, zwielokrotnioną wymianą replik (MREMD)¹⁰¹ oraz hamiltonowską wymianą replik (H-REMD)⁷⁷. Jej atutem jest możliwość oszacowania błędów powstałych podczas próbkowania przestrzeni konformacyjnej badanych układów, a także analiza krzywych pojemności cieplnej dla losowo wybranych fragmentów trajektorii pozwalająca określić poziom zrównoważenia symulowanego układu¹⁰⁴.

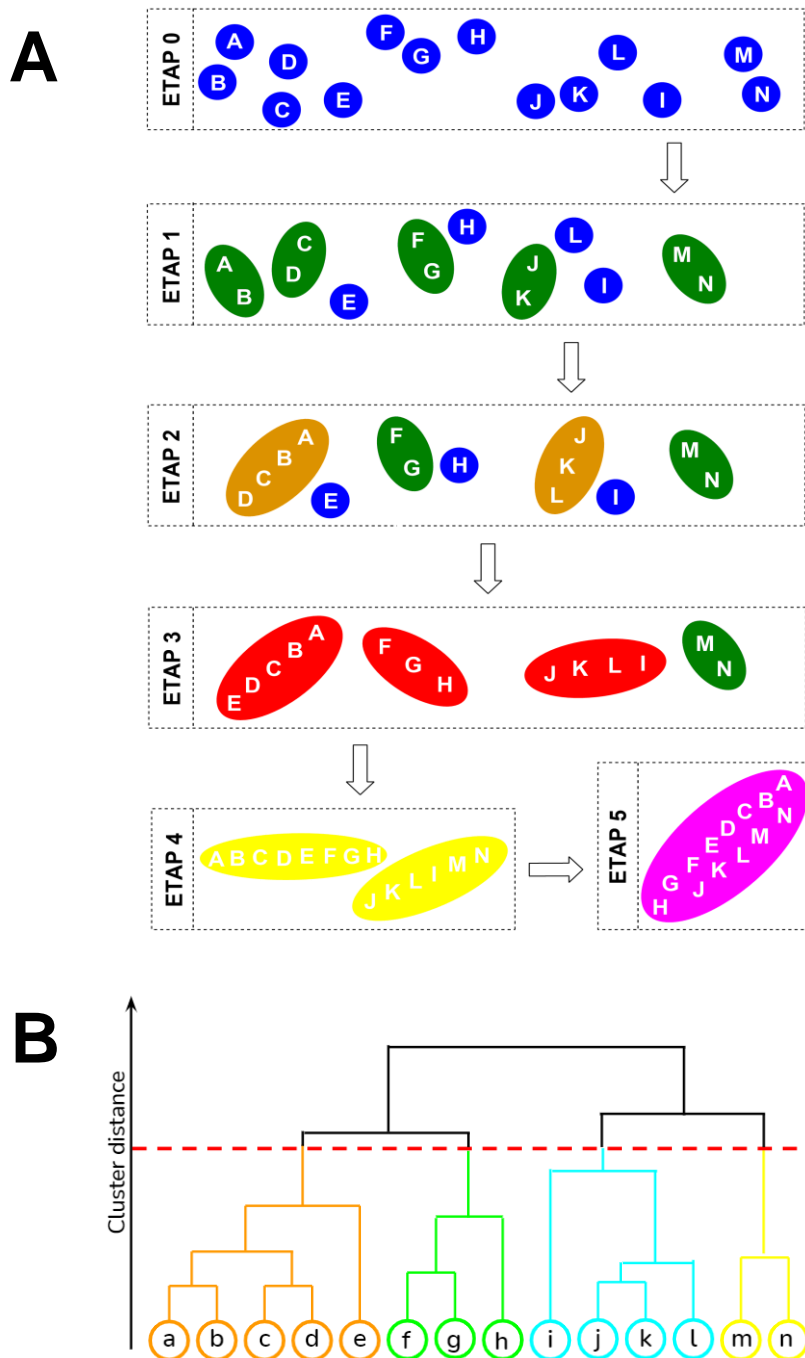
5. 7. 2. Metody analizy skupień

Z analizy wyników symulacji wykonanych przy pomocy metody WHAM^{78,79} uzyskujemy zwykle bardzo duże zbiory danych liczące nawet setki tysięcy korzystnych energetycznie konformacji badanego związku. Do analizy tak dużego zbioru danych służą metody analizy skupień, które polegają na wyselekcjonowaniu najbardziej reprezentatywnych z tego zbioru konformacji. Dzięki temu początkowy zbiór danych zostaje zredukowany do kilku lub kilkunastu konformacji¹⁰⁵. Ze względu na odmienne algorytmy grupowania konformacji do określonych skupień metody te dzielą się na hierarchiczne i niehierarchiczne^{105,106}.

Metody hierarchiczne są bardziej czasochłonne obliczeniowo. Uzyskane za ich pomocą wyniki obliczeń są łatwiejsze do interpretacji, ponieważ są przedstawiane za pomocą tzw. dendrogramu (Rys. 16B), który dostarcza wgląd w najbardziej optymalną wartość wielkości odcięcia, jakiej należy użyć w celu wygenerowania określonej liczby skupień. W metodach hierarchicznych proces tworzenia nowych skupień przebiega przez wiele etapów pośrednich. Początkowo każda konformacja stanowi odrębne skupienie i dopiero w kolejnych etapach dochodzi do tworzenia mniejszej liczby nowych skupień. Trwa to do momentu uzyskania jednego skupienia zawierającego najbardziej podobne do siebie konformacje. Stopień podobieństwa tych konformacji jest obliczany przy pomocy takich miar jak np. RMSD^{16,17}, GDT-TS^{18,19} czy TM-score^{20,21}. Jest to dość wymagająca obliczeniowo procedura, która wynika z konieczności każdorazowego porównywania ze sobą konformacji w obrębie każdego ze skupień w etapie poprzedzającym utworzenie nowego. Do najbardziej popularnych metod tego typu należy metoda najbliższego sąsiada (ang.: *single-link*, *SLINK*)¹⁰⁷ (Rys. 16A) oraz minimum wariancji Ward'a (ang.: *Ward's Minimum-Variance Method*)¹⁰⁸. W pierwszej z tych metod dwa najbardziej podobne do siebie i najbliższe sąsiadujące ze sobą skupienia są przydzielane do tego samego, nowo tworzonego skupienia. Druga z metod umożliwia wyodrębnienie bardziej podobnych do siebie konformacji badanego związku i jest ona często stosowana do analizy wyników symulacji REMD^{96,97}, MREMD¹⁰¹ oraz H-REMD⁷⁷ w gruboziarnistym polu sił UNRES. Idea tej metody polega na grupowaniu konformacji w taki sposób, by wariancje między nimi wewnątrz nowo tworzonego skupienia były możliwie jak najmniejsze.

Metody niehierarchiczne są dużo szybsze i bardziej dokładne obliczeniowo od metod hierarchicznych. Są jednak dużo trudniejsze do zaimplementowania, ponieważ wymagają wstępnego określenia liczby skupień do których będą grupowane analizowane konformacje badanego związku. Jedną z najstarszych i najbardziej znanych metod niehierarchicznych jest metoda *K*-średnich (ang.: *K-means*)¹⁰⁹, która polega na wstępnym określeniu liczby skupień

i środków każdego z nich. Dla każdej konformacji obliczana jest jej odległość względem środka każdego ze skupień. Do każdego ze skupień przydzielane są jedynie te konformacje, które znajdują się najbliżej ich środka. W dalszej kolejności określone są nowe środki skupień – są nimi średnie arytmetyczne ze współrzędnych przestrzennych wszystkich konformacji wchodzących w skład danego skupienia. Iteracje te trwają aż do momentu uzyskania zadanej liczby skupień, których konformacje będą odznaczać się największym stopniem względnego podobieństwa do siebie.



Rys. 16. Schematyczne przedstawienie zasady działania metody najbliższego sąsiada (A) oraz wygenerowany dla niego dendrogram (B), w którym czerwoną linią oznaczono optymalną wartość odcięcia potrzebną do wygenerowania zadanej liczby skupień¹¹⁰.

5. 8. Metody obliczania energii swobodnej układu

Do właściwego zrozumienia oraz opisu większości modelowanych procesów chemicznych wymagana jest znajomość wartości zmian energii swobodnych (lub potencjału średniej siły) układu symulowanego w różnych warunkach środowiskowych. W chemii obliczeniowej opracowano do tej pory wiele metod służących do wykonywania tego typu obliczeń. Każda z nich opiera się na podstawowych prawach mechaniki statystycznej i polega na określeniu zmian wartości energii swobodnej (lub potencjału średniej siły) we wszystkich próbkowanych fragmentach przestrzeni konformacyjnej badanego układu.

Znajomość zmian wartości energii swobodnej jest bardzo przydatna w modelowaniu procesów mających związek z solwatacją wielu cząstek (jonów, atomów, cząsteczek) w roztworach wodnych. Obliczenia te mają duże zastosowanie praktyczne zarówno w nauce jak i przemyśle, ponieważ dostarczają kluczowych informacji dotyczących sposobu funkcjonowania wielu układów biologicznych.

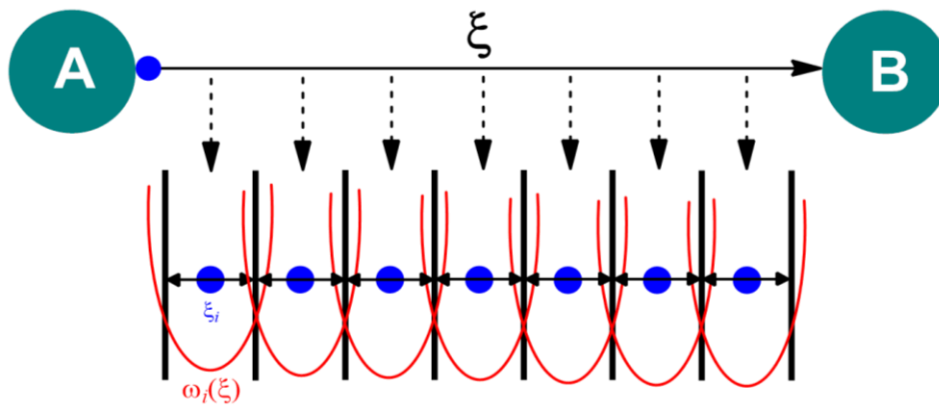
W kolejnych podrozdziałach została przedstawiona szczegółowa charakterystyka metod, które stosowałam podczas prowadzonych przez siebie badań.

5. 8. 1. Metoda próbkowania parasolowego

Metoda próbkowania parasolowego (ang.: *Umbrella Sampling*, US)^{111,112} służy do obliczenia wartości potencjału średniej siły (ang.: *Potential Mean Force*, PMF) układu symulowanego wzdłuż pewnego parametru ξ zwanego współrzędną reakcji. Zmiany konformacyjne takiego układu są dokonywane wzdłuż współrzędnej ξ i opisywane za pomocą tzw. ścieżki reakcji. Ścieżka ta jest podzielona na tzw. okna, w których próbkowane są odrębne fragmenty przestrzeni konformacyjnej badanego układu. W każdym z okien na badany układ nakładany jest potencjał harmoniczny $\omega_i(\xi)$ dany jako:

$$\omega_i(\xi) = k_i(\xi - \hat{\xi})^2 \quad (5.8.1a)$$

gdzie k_i to stała siłowa, ξ to współrzędna reakcji, $\hat{\xi}$ jest częścią współrzędnej reakcji zależną od współrzędnych przestrzennych układu w próbkowanym fragmencie jego przestrzeni konformacyjnej, natomiast $\omega_i(\xi)$ to potencjał harmoniczny nakładany na próbkowany region przestrzeni konformacyjnej układu w i -tym oknie symulacji. Rola tego potencjału polega na utrzymaniu badanego układu w pobliżu zdefiniowanej wartości współrzędnej reakcji. Umożliwia on poprawienie efektywności próbkowania, ponieważ wymusza próbkowanie trudnodostępnych, wysokoenergetycznych regionów przestrzeni konformacyjnej badanego układu.



Rys. 17. Schematyczne przedstawienie sposobu próbkowania przestrzeni konfiguracyjnej układu przy przejściu między stanem początkowym A i końcowym B na hiperpowierzchni jego energii potencjalnej w metodzie *umbrella sampling*.

Rozkład prawdopodobieństwa $\rho_i(\xi)$ próbkowania przestrzeni konfiguracyjnej układu w i -tym oknie symulacji jest wówczas dany jako:

$$\rho_i(\xi) = \omega_i(\xi) \frac{\int \exp[-\beta U(\Gamma)] \delta(\xi - \hat{\xi}) d\Gamma}{\int \omega_i(\hat{\xi}) \exp[-\beta U(\Gamma)] d\Gamma} \quad (5.8.1b)$$

gdzie $\beta = (k_B T)^{-1}$, $d\Gamma$ oznacza całkowanie po wszystkich stopniach swobody układu, $U(\Gamma)$ to energia potencjalna układu zależna od wszystkich stopni swobody Γ tego układu, a zmienna δ to różnica średnich odchyżeń między dwoma stanami symulowanego układu. Na podstawie rozkładu $\rho_i(\xi)$ obliczana jest wartość potencjału średniej siły $G_i(\xi)$ w i -tym oknie symulacji:

$$G_i(\xi) = -\frac{1}{\beta} \ln \rho_i(\xi) - \omega(\xi) + F_i \quad (5.8.1c)$$

W powyższym równaniu zmienna F_i jest siłą działającą na środek ciężkości układu w i -tym oknie symulacji:

$$F_i = -\frac{1}{\beta} \ln \langle \exp(-\beta \omega_i(\xi)) \rangle \quad (5.8.1d)$$

Wartości potencjału średniej siły $G_i(\xi)$ z wszystkich okien symulacji są następnie przetwarzane przy pomocy metody WHAM^{78,79}. Z uzyskanego histogramu wyznaczane jest ogólne prawdopodobieństwo $\rho(\xi)$ znalezienia badanego układu w najniższym energetycznie obszarze jego przestrzeni konformacyjnej. Na podstawie wartości $\rho(\xi)$ obliczany jest potencjał średniej siły $G(\xi)$ w najniższym energetycznie stanie badanego układu:

$$G(\xi) = -k_B T \ln \frac{\rho(\xi)}{\Delta V(\xi)} \quad (5.8.1e)$$

gdzie k_B to stała Boltzmanna, a $\Delta V(\xi)$ to objętość próbkowanej przestrzeni konfiguracyjnej układu wzdłuż współrzędnej reakcji ξ .

Metoda *umbrella sampling* jest często wykorzystywana do próbkowania przestrzeni konfiguracyjnej białek w symulacjach z wykorzystaniem gruboziarnistego pola sił UNRES. Na dokładność próbkowania przestrzeni konfiguracyjnej badanego układu ma wpływ ustalona dla danej symulacji wartość współrzędnej reakcji ξ oraz stałej siłowej k . Jeśli współrzędna reakcji jest dobrana w sposób nieprawidłowy to podczas próbkowania dochodzi do pominięcia wielu istotnych regionów przestrzeni konfiguracyjnej, co skutkuje obraniem nieprawidłowej ścieżki reakcji opisującej niepoprawne zmiany strukturalne i dynamikę białek. Wartość stałej siłowej k powinna być rozsądnie duża, tak by możliwe było pokonywanie wysokich barier energetycznych na hiperpowierzchni energii potencjalnej białek. Jednak obranie zbyt dużych wartości k może być źródłem wielu błędów powstałych podczas numerycznego całkowania równań ruchu wszystkich centrów oddziaływań w białkach.

5. 8. 2. Metoda zaburzenia energii swobodnej

Metoda zaburzenia energii swobodnej układu (ang.: *Free Energy Perturbation*, FEP)¹¹² jest jedną z najstarszych i najczęściej stosowanych metod do obliczania zmian wartości energii swobodnej ΔG symulowanego układu. Po raz pierwszy użyto jej do obliczenia zmian wartości ΔG dla procesu hydratacji metanolu i etanu w wodzie¹¹³, natomiast jej dzisiejsze zastosowania skupiają się głównie na komputerowym projektowaniu leków poprzez obliczanie zmian wartości energii swobodnej dla procesów wiązania ligandu z białkiem w symulowanym roztworze¹¹⁴.

W metodzie FEP zmiana wartości energii swobodnej ΔG między dwoma nieznacznie różniącymi się między sobą stanami energetycznymi jest postrzegana w charakterze perturbacji i liczona w następujący sposób¹¹²:

$$\Delta G(0 \rightarrow 1) = -\frac{1}{\beta} \ln \langle \exp(-\beta(U_1 - U_0)) \rangle_0 \quad (5.8.2a)$$

$$\Delta G(1 \rightarrow 0) = -\frac{1}{\beta} \ln \langle \exp(-\beta(U_0 - U_1)) \rangle_1 \quad (5.8.2b)$$

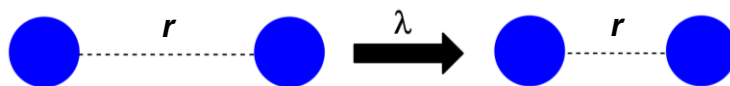
gdzie 0 to stan niezaburzony (początkowy), a 1 jest stanem zaburzonym (końcowym), U_0 i U_1 to energie potencjalne reprezentujące każdy z tych stanów, $\beta = 1/k_B T$ (gdzie k_B to stała Boltzmana), ΔG to zmiana wartości energii swobodnej przy przejściu między tymi dwoma stanami, a człony $\langle \exp(-\beta(U_1 - U_0)) \rangle_0$ oraz $\langle \exp(-\beta(U_0 - U_1)) \rangle_1$ to średnie arytmetyczne konfiguracji równowagowych stanu niezaburzonego (0) oraz stanu zaburzonego (1). Obydwa powyższe wzory są zasadne, jedynie dla przypadku, gdy masa cząstek ulegających transformacjom chemicznym jest stała w trakcie całej symulacji, a zaburzenie układu jest niewielkie. W każdej z tych przemian generowane są odmienne wartości zmian energii

swobodnej, które świadczą o występowaniu jednego preferowanego kierunku transformacji przemian chemicznych.

Dokładność obliczeniowa metody FEP jest zależna od rozkładów prawdopodobieństwa próbkowania stanu 1 i 0 w niskoenergetycznych regionach przestrzeni konfiguracyjnej badanego układu. Jeśli rozkłady te są jednakowe lub w dużej mierze pokrywają się ze sobą to możliwe jest dokładne oszacowanie zmian wartości ΔG przy przejściu między stanem $0 \rightarrow 1$ lub odwrotnie. Duże trudności obliczeniowe pojawiają przy stosunkowo dużych parametrach zaburzenia symulowanego układu. W tego typu przypadkach zmiany wartości ΔG są liczone z uwzględnieniem wielu stanów pośrednich w następujący sposób:

$$\Delta G = -\frac{1}{\beta} \sum_{i=1}^{N-1} \ln \left\langle \exp(-\beta \Delta \lambda_i \Delta U) \right\rangle_{\lambda_i} \quad (5.8.2c)$$

gdzie $\beta = 1/k_B T$ (k_B – stała Boltzmana, T – temperatura), N to liczba próbkowanych stanów pośrednich, ΔU to różnica energii potencjalnej dla dwóch kolejnych stanów próbkowanych w stanie i , natomiast λ jest parametrem sprzężenia dla pary próbkowanych stanów. Parametr λ przyjmuje wartości z zakresu od 0 do 1. Opisuje on zmiany zachodzące w symulowanym układzie powstałe na skutek przemiany jednego stanu w drugi, a także ma kluczowy wpływ na uzyskiwaną dokładność wyników obliczeń. Parametr ten jest stosowany m. in. do obliczania wartości PMF dla różnych odległości dwóch oddziałujących ze sobą cząstek (Rys. 18).



Rys. 18. Schematyczne przedstawienie przykładowego zastosowania parametru λ w obliczeniach wartości PMF względem różnych odległości dwóch oddziałujących ze sobą cząstek.

5. 8. 3. Metoda Widom'a wprowadzania cząstki

Kolejną popularną metodą służącą do obliczania zmian wartości energii swobodnej ΔG wynikającej z wprowadzenia niepolarniej cząstki do wody lub innego polarnego rozpuszczalnika w symulowanym układzie jest metoda Widom'a zwana także metodą wprowadzania cząstki (ang.: *Particle Insertion*, PI)^{115,116}. W metodzie tej proces rozpuszczania substancji niepolarniej wprowadzanej do polarnego rozpuszczalnika składa się z dwóch etapów. Pierwszy z nich polega na samoistnym wytworzeniu przez strukturę rozpuszczalnika hydrofobowych wnęk o różnej wielkości poprzez które substancja niepolarna będzie mogła wnikać do wnętrza symulowanego roztworu. Analiza procesu tworzenia tych wnęk stanowi cenne źródło informacji nt. strukturalnych i fizykochemicznych zmian cząsteczek rozpuszczalnika wewnątrz pierwszej warstwy solwacyjnej substancji niepolarniej. Drugi etap

polega na obliczeniu energii wzajemnych oddziaływań między niepolarną cząstką i sąsiadującymi z nią cząsteczkami rozpuszczalnika.

Idea tej metody polega na wielokrotnym wprowadzaniu niepolarnej cząstki w różne losowe miejsca termodynamicznie zrównoważonej struktury czystego rozpuszczalnika. Energia swobodna tego procesu jest wyrażana przy pomocy potencjału chemicznego μ_{exc} opisywanego jako:

$$\mu_{exc} = -k_B T \ln \left\langle \exp \left(\frac{-\Delta U}{k_B T} \right) \right\rangle \quad (5.8.3a)$$

gdzie ΔU to różnica energii potencjalnej między układem składającym się z N -cząsteczek rozpuszczalnika i jednej cząstki substancji niepolarnej oraz układem zbudowanym z N -cząsteczek rozpuszczalnika, k_B to stała Boltzmanna, T to temperatura układu, a człon $\langle \dots \rangle_N$ to średnia z wszystkich konformacji struktury czystego rozpuszczalnika po każdorazowym wprowadzeniu do niego pojedynczej cząstki substancji niepolarnej.

5. 9. Teoretyczne metody przewidywania struktur białek

Znajomość struktur przestrzennych białek umożliwia zrozumienie mechanizmu funkcjonowania żywych komórek. Wiedza ta ma kluczowe znaczenie w wielu zastosowaniach medycznych i stanowi cenne źródło informacji dla komputerowego projektowania leków. Niestety liczba struktur białek oznaczonych metodami eksperymentalnymi i zdeponowanych w bazie PDB (ang.: *Protein Data Bank*)¹¹⁷ jest niewielka (liczba struktur: 185728, stan na 22.04.2022 r.) w porównaniu do liczby sekwencji aminokwasowych w bazach danych takich jak UniProt (ang.: *The Universal Protein Resource*)¹ oraz Swiss-Prot¹¹⁸ (230 895 644, stan na 22.04.2022 r.). Istnieje zatem konieczność opracowania metod służących do teoretycznego przewidywania struktur białek, które mogłyby stanowić alternatywę dla powszechnie używanych do tego celu metod eksperymentalnych takich jak np. krystalografia rentgenowska (ang.: *X-Ray crystallography*)¹¹⁹, spektroskopia jądrowego rezonansu magnetycznego (ang.: *Nuclear Magnetic Resonance, NMR*)¹²⁰ lub mikroskopia krioelektronowa (ang.: *cryogenic Electron Microscopy, cryo-EM*)¹²¹.

Tradycyjne metody przewidywania struktur białek dzielą się na dwie grupy: metody oparte na modelowaniu homologicznym oraz metody modelowania *de novo*. Metody te opierają się na hipotezie Anfinsena¹²² mówiącej o tym, że sekwencja aminokwasowa białka warunkuje jego strukturę przestrzenną, a stan natywny białka jest osiągnany jedynie w punkcie odpowiadającym globalnemu minimum energii swobodnej w jego przestrzeni konformacyjnej. Pierwsza grupa metod do przewidywania struktur białek wykorzystuje podobieństwo

sekwencyjne między modelowanym białkiem oraz jego homologami o znanych strukturach przestrzennych. Natomiast metody *de novo* opierają na podstawowych prawach fizyki oraz na hipotezie Anfinsena¹²². Założeniem metod *de novo* jest znalezienie struktury białka o jak najmniejszej wartości energii swobodnej przy użyciu różnego typu wyrażeń służących do opisu tej energii¹²³. Większość z obecnie stosowanych metod teoretycznych ma charakter hybrydowy np. opracowane w naszym Zespole podejście obliczeniowe łączące modelowanie homologiczne z opartym na fizyce oddziaływań gruboziarnistym polem sił UNRES^{124,125}. Innym przykładem metody hybrydowej jest AlphaFold¹²⁶, który stanowi połączenie metod głębokiego uczenia (ang.: *deep learning*) z sieciami neuronowymi oraz informacjami pobranymi z sekwencyjnych oraz strukturalnych baz danych.

5. 9. 1. Modelowanie homologiczne

Metody modelowania homologicznego (ang.: *Homology Modeling*)¹²⁷ polegają na przewidywaniu struktury przestrzennej białka na podstawie podobieństw sekwencyjnych między modelowanym białkiem oraz jego homologami o znanych strukturach oznaczonych metodami eksperymentalnymi. Przydatnymi narzędziami umożliwiającymi łatwe i szybkie modelowanie homologiczne białek są serwery internetowe takie jak np. PSIPRED (ang.: *PSI-blast based secondary structure PREDiction*)¹²⁸, które w oparciu o informacje zakodowane w sekwencji aminokwasowej modelowanego białka z ok. 80% prawdopodobieństwem przewidują jego strukturę II-rzędową. W dalszej kolejności przy pomocy specjalistycznego oprogramowania takiego jak np. PSI-BLAST (ang.: *PSI-Basic Local Alignment Search Tools*)¹²⁹ przeszukiwane są bazy danych eksperymentalnych w celu znalezienia białek o możliwie jak największym podobieństwie sekwencyjnym do modelowanego wzorca. Sekwencje tych białek są w pierwszej kolejności nakładane na siebie, a następnie dopasowywane do sekwencji wzorca. Białka o największym podobieństwie sekwencyjnym do modelowanego wzorca są wykorzystywane do tworzenia modeli jego III-rzędowych struktur. Modele te są zazwyczaj tworzone przy pomocy algorytmów wykorzystujących jako szkielet najbardziej podobny i kompletny homolog do modelowanego białka. Pozostałe obszary tych modeli są tworzone na podstawie fragmentów innych homologów o dużym podobieństwie do lokalnej sekwencji i struktury II-rzędowej modelowanego białka. Wszystkie brakujące elementy takiego modelu, np. pętle oraz nieustrukturyzowane fragmenty są tworzone od podstaw – najpierw odtwarzane są łańcuchy boczne białka, później jego łańcuch główny, a następnie każdy z nich jest poddawany krótkiej optymalizacji. Modele te są następnie grupowane do osobnych skupień na podstawie miar określających podobieństwo ich struktur do białek homologicznych dla badanego wzorca.

Strukturę reprezentacyjną modelowanego białka stanowi model odznaczający się największym podobieństwem strukturalnym do jego homologów. Poprawność stereochemiczna takiego modelu jest sprawdzana np. w programie PROCHECK¹³⁰, który generuje unikalny dla niego wykres Ramachandran.

Dokładność opisywanych metod zależy od procentowego podobieństwa sekwencji białek homologicznych do modelowanego wzorca, a także od sposobu jego modelowania. Podobieństwo to powinno być możliwie jak największe – jeśli wynosi ono mniej niż 30% to uzyskanie poprawnego dopasowania sekwencyjnego białek homologicznych do modelowanego wzorca jest bardzo trudne i wpływa niekorzystnie na wynik dalszych przewidywań. Aby ograniczyć tę możliwość procedura modelowania jest zazwyczaj wykonywana na zestawie nie jednego, lecz wielu białek homologicznych.

Zaletą metod modelowania homologicznego jest przede wszystkim niski koszt obliczeniowy, a także możliwość zdalnego przewidywania struktur białek przy pomocy serwerów internetowych takich jak np. I-TASSER (ang.: *Iterative Threading ASSEmbly Refinement*)¹³¹, RaptorX¹³² lub HHpred (ang.: *Homology detection and structure prediction server using Hidden Markov Models*)¹³³. Wadą opisywanych metod jest duży spadek dokładności przewidywań, który następuje wraz ze spadkiem podobieństwa sekwencji między modelowanym białkiem i jego homologami o znanej strukturze eksperymentalnej.

5. 9. 2. Metody *de novo* przewidywania struktur białek

Metody *de novo*^{134,135} polegają na przewidywaniu struktur przestrzennych białek przy pomocy podstawowych praw fizyki oraz hipotezy Anfinsena¹²². W metodach *de novo* znalezienie struktury natywnej dla zadanej sekwencji aminokwasowej białka wymaga dokładnego przeszukania jego przestrzeni konformacyjnej w obszarach o niskich wartościach energii swobodnej. Do eksploracji przestrzeni konformacyjnej badanego białka służą metody symulacji komputerowych takie jak Monte Carlo¹³⁶ lub dynamika molekularna^{81,82} albo jej rozszerzenia uwzględniające wymianę replik^{96,97,101,77}. Preferowane jest przy tym użycie gruboziarnistych pól sił jak np. UNRES, które z uwagi na uproszczony sposób reprezentacji badanego białka umożliwiają znaczne skrócenie czasu trwania takiej symulacji. Wyniki symulacji z polem sił UNRES są następnie przetwarzane przy pomocy metody WHAM^{78,79}, a konformacje o najniższej energii swobodnej są grupowane do osobnych skupień przy pomocy metody Ward'a¹⁰⁸. Z wszystkich skupień wybierana jest jedna reprezentatywna konformacja badanego białka, której energia swobodna jest możliwie jak najniższa.

Metody *de novo* służą do modelowania III- i IV-rzędowych struktur białek oraz dostarczają kluczowych informacji nt. dynamiki procesu ich związania¹³⁷. Główną wadą tych metod są duże trudności związane z uzyskaniem dokładnego, globalnego minimum wartości energii swobodnej dla modelowanych białek, a także niedokładna (tj. przybliżona) postać funkcji energii w stosowanych do obliczeń polach sił.

5. 9. 3. Połączenie modelowania homologicznego z polem sił UNRES

Innym sposobem służącym do przewidywania struktur przestrzennych białek jest podejście hybrydowe opracowane w zespole Prof. Adama Liwo i Prof. Cezarego Czaplewskiego, które polega na połączeniu metod modelowania homologicznego z gruboziarnistym polem sił UNRES^{124,125}. Podejście to odznacza się dużą dokładnością przewidywań struktur białek z poprawnym sposobem upakowania występujących wśród nich domen.

Idea tego hybrydowego podejścia polega na wyborze najlepszych jakościowo modeli białek o tej samej sekwencji do modelowanego wzorca. Modele te są pozyskiwane z wielu serwerów zewnętrznych, a ich jakość oceniana jest przy pomocy zewnętrznego serwera DeepQA¹³⁸. Modele te są porównywane ze sobą w celu znalezienia wspólnych dla nich fragmentów strukturalnych zwanych konsensusami. Na podstawie tych fragmentów wyprowadzane są potencjały więzów V_x , tj. *quasi*-harmoniczne funkcje logarytmiczne Gaussa, które są nakładane na modelowane białko i mają następującą postać:

$$V_X = V_d + V_\theta + V_\gamma + V_{SC} \quad (5.9.3a)$$

gdzie V_d to więzy nakładane na odległości pomiędzy dwoma kolejnymi atomami C^α , V_θ to więzy nakładane na wirtualne kąty walencyjne θ , V_γ to więzy nakładane na wirtualne kąty torsyjne γ , a V_{SC} to więzy nakładane na wektory położenia środków zjednoczonych łańcuchów bocznych. Reprezentację matematyczną każdego z tych więzów przedstawiają poniższe równania:

$$V_d = - \sum_{j < k} \ln \left\{ \sum_{n=1}^N \exp \left[- \frac{1}{2} \left(\frac{d_{jk} - d_{jk}^{(n)}}{\sigma_{d_{xj}}^{(n)}} \right)^2 \right] \right\} \quad (5.9.3b)$$

$$V_\theta = - \sum_j \ln \left\{ \sum_{n=1}^N \exp \left[- \frac{1}{2} \left(\frac{\theta - \theta_j^{(n)}}{\sigma_{\theta_j}^{(n)}} \right)^2 \right] \right\} \quad (5.9.3c)$$

$$V_\gamma = - \sum_j \ln \left\{ \sum_{n=1}^N \exp \left[- \frac{1}{2} \left(\frac{\gamma - \gamma_j^{(n)}}{\sigma_{\gamma_j}^{(n)}} \right)^2 \right] \right\} \quad (5.9.3d)$$

$$V_{SC} = - \sum_j \ln \left\{ \sum_{n=1}^N \exp \left[- \frac{1}{2} \left(\frac{\|\hat{\mathbf{r}}_{SC_j} - \hat{\mathbf{r}}_{SC_j}^{(n)}\|^2}{\sigma_{\gamma_j}^{(n)}} \right)^2 \right] \right\} \quad (5.9.3e)$$

gdzie d_{jk} oraz $d_{jk}(n)$ to odległości między dwoma najbliższymi atomami C^α w modelowanej i n -tej strukturze odniesienia, θ_j oraz $\theta_j(n)$ to wirtualne kąty walencyjne pomiędzy trzema kolejnymi atomami C^α w modelowanej oraz n -tej strukturze odniesienia, γ_j oraz $\gamma_j(n)$ to wirtualne kąty torsyjne zawarte między czterema kolejnymi atomami C^α w modelowanej oraz n -tej strukturze odniesienia, natomiast r_{SC_j} oraz $r_{SC_j}(n)$ to wektory kartezjańskich współrzędnych położenia dla środka j -tego zjednoczonego łańcucha bocznego w symulowanej i n -tej strukturze odniesienia, a parametr σ to odchylenie standardowe każdego z tych więzów.

Informacje strukturalne dla zadanej sekwencji modelowanego białka są wyodrębniane z bazy dostępnych fragmentów bibliotek przy pomocy metody DFA (ang.: *Dynamic Fragment Assembly*)^{139,140}. Pozyskane w ten sposób informacje są następnie przekształcane do odpowiednich pseudopotencjałów energii:

$$e_{DFA} = w_{DFA,dist} e_{DFA,dist} + w_{DFA,angle} e_{DFA,angle} + w_{DFA,nn} e_{DFA,nn} \quad (5.9.3f)$$

gdzie e_{DFA} to sumaryczna energia wszystkich pseudopotencjałów, człony $e_{DFA,dist}$ oraz $e_{DFA,angle}$ służą do upodobnienia lokalnej struktury modelowanego białka do fragmentów bibliotek, natomiast $e_{DFA,nn}$ opisuje liczbę reszt aminokwasowych wokół sąsiednich atomów C^α w modelowanej strukturze białka względem użytych do obliczeń fragmentów bibliotek. Funkcja energii całkowitej badanego białka jest wówczas wyrażana następująco:

$$U = U_{UNRES} + E_{DFA} + w_d V_d + w_\theta V_\theta + w_\gamma V_\gamma + w_{SC} V_{SC} \quad (5.9.2g)$$

gdzie U_{UNRES} to efektywna funkcja energii w polu sił UNRES, e_{DFA} to sumaryczny pseudopotencjał energii w metodzie DFA, V_x to potencjały więzów nakładane na modelowane białko, a w_x to wagi tych więzów.

Po nałożeniu więzów na modelowane białko jego przestrzeń konformacyjna jest przeszukiwana w celu znalezienia stanu o możliwie jak najniższej wartości energii swobodnej. Aby tego dokonać wykonuje się symulacje MREMD¹⁰¹ w gruboziarnistym polu sił UNRES. Wyniki tych symulacji są następnie przetwarzane i analizowane przy pomocy metody WHAM^{78,79} oraz metody Ward'a¹⁰⁸ w celu wyłonienia najbardziej reprezentatywnej konformacji dla stanu natywnego modelowanego białka. Konformacja ta jest następnie konwertowana do reprezentacji pełnoatomowej pola sił AMBER w wersji ff14SB¹⁴¹ przy pomocy algorytmów PULCHRA¹⁴² i SCWRL¹⁴³. W dalszym etapie energia potencjalna struktury modelowanego białka jest minimalizowana, a następnie białko to jest poddawane

krótkiej symulacji MD w celu wyeliminowania potencjalnych ograniczeń sterycznych oraz poprawienia przestrzennego ułożenia jego łańcuchów bocznych.

5. 9. 4. Metoda AlphaFold

Metoda AlphaFold2¹²⁶ jest jedną z metod służących do teoretycznego przewidywania struktur białek i kompleksów, która została najwyżej oceniona w 14-tej edycji eksperymentu naukowego CASP. Jest to metoda opierająca się na sztucznej inteligencji, która do przewidywań wykorzystuje metody głębokiego uczenia (ang.: *deep learning*) oraz duże zbiory danych składające się z sekwencji oraz struktur oznaczonych eksperymentalnie białek zdeponowanych w bazie PDB¹¹⁷. Mechanizm modelowania białek przy pomocy tej metody przypomina sposób w jaki składane są puzzle. W pierwszej kolejności przeszukiwana jest baza danych PDB w celu znalezienia białek o sekwencjach homologicznych do modelowanego wzorca. Następnie spośród znalezionych białek wyszukiwane są wspólne dla nich fragmenty sekwencji, które są grupowane do osobnych skupisk. W ostatnim etapie modelowania poszukiwane są sposoby połączenia tych skupisk w jedną całość pasującą do modelowanego białka.

Opisywana metoda odznacza się bardzo wysoką dokładnością przewidywań (na poziomie atomowym) nawet dla białek zawierających ponad 2000 reszt aminokwasowych. W 14-tej edycji CASP blisko 2/3 struktur białek przewidzianych przy pomocy AlphaFold2 nie różniło się znacznie od struktur uzyskanych przy pomocy czasochłonnych metod doświadczalnych takich jak krystalografia rentgenowska (X-Ray)¹¹⁹, spektroskopia jądrowego rezonansu magnetycznego (NMR)¹²⁰ czy mikroskopia krioelektronowa (cryo-EM)¹²¹. Główna wada tej metody ma związek z odkształcaniem się struktury reszt aminokwasowych podczas ich składania do trójwymiarowej struktury modelowanego białka.

5. 9. 4. Eksperyment CASP

CASP (ang.: *Critical Assessment of Protein Structure Prediction*)¹⁴⁴ jest cyklicznym eksperymentem odbywającym się regularnie co dwa lata od 1994 r., w którym każdorazowo uczestniczy wiele grup badawczych z całego świata. Głównym celem tego eksperymentu jest testowanie przez jego uczestników nowych i bardziej dokładnych metod służących do komputerowego modelowania III- i IV-rzędowych struktur białek.

Zadaniem każdej z grup uczestniczących w eksperymencie CASP jest jak najdokładniejsze określenie III- i IV-rzędowej struktury testowanych białek na podstawie ich sekwencji aminokwasowych oraz różnej jakości ich modeli serwerowych. Eksperymentalna

struktura badanych białek została wcześniej oznaczona, lecz do momentu zakończenia eksperymentu CASP nie jest ona udostępniana w bazie PDB¹¹⁷.

Do przewidywania struktur białek wykorzystywane są metody modelowania homologicznego¹²⁷, metody *de novo*^{134,135} oraz metody hybrydowe. Każda z grup ma tyle samo czasu na określenie struktury badanego białka i wysłanie pięciu ich zdaniem najlepszych modeli badanego białka do organizatorów CASP. Czas ten jest uzależniony od stosowanego protokołu przewidywań – dla białek określanych wyłącznie na podstawie ich sekwencji aminokwasowej termin ten wynosi 3 dni, natomiast dla białek modelowanych na podstawie ich sekwencji oraz modeli serwerowych termin ten jest dłuższy i wynosi 21 dni.

Ogólny czas trwania eksperymentu CASP wynosi blisko 9 miesięcy – w przeciągu pierwszych 5 miesięcy wszystkie grupy badawcze, które biorą w nim udział zajmują się przewidywaniem struktury białek na podstawie opracowanych przez siebie protokołów przewidywań. Przez kolejne 4 miesiące wyniki tych przewidywań są oceniane przez niezależny zespół ekspertów oraz zautomatyzowane metody bazujące na popularnych miarach porównawczych takich jak np. RMSD^{16,17}, GDT-TS^{18,19} czy TM-score^{20,21}. Ocenie podlega również przyjęta przez każdą z grup badawczych metodologia obliczeń. Każda edycja eksperymentu CASP kończy się konferencją na której zwycięzcy tego konkursu oraz inne najlepiej ocenione grupy badawcze prezentują wyniki swoich przewidywań oraz stosowaną metodologię obliczeń.

Od 2003 r. do eksperymentu CASP został włączony projekt o nazwie CAPRI¹⁴⁵, który polega na przewidywaniu oddziaływań białko-ligand oraz białko-białko. Wyniki badań uzyskane w ramach projektu CAPRI stanowią cenny wgląd w nowe sposoby modelowania białek i ich kompleksów.

Podczas swoich studiów doktoranckich uczestniczyłam w dwóch ostatnich edycjach eksperymentu CASP – po raz pierwszy w 2018 r. w jego 13-tej edycji^{146,125} oraz dwa lata później w 14-tej edycji¹⁴⁷. Podczas każdej z nich brałam udział w pracach dwóch grup badawczych o nazwie UNRES oraz KIAS-Gdansk (w przypadku 14-tej edycji UNRES-template). W obydwu tych grupach obliczenia były prowadzone przy użyciu metody MREMD¹⁰¹ oraz gruboziarnistego pola sił UNRES w wersji NEWCT-9P³⁹. Protokół predykcyjny grupy UNRES bazował na metodach opartych wyłącznie na fizyce oddziaływań, natomiast w przypadku grupy KIAS-Gdansk (później UNRES-template) wykorzystywane były także modele serwerowe dostępne dla każdego z badanych białek. W trakcie swojego uczestnictwa w dwóch ostatnich edycjach eksperymentu CASP zajmowałam się

przewidywaniem 20 struktur białek monomerycznych^{125,146,147} oraz 1 kompleksu białkowego składającego się z trzech jednostek heteromerycznych¹⁴⁸.

6. BADANIA WŁASNE

W niniejszej części pracy zostaną zaprezentowane projekty badawcze, w których brałam główny lub znaczący udział. Projekty te stanowią część mojego dotychczasowego dorobku naukowego, a uzyskane za ich pomocą wyniki obliczeń pomogły mi zrealizować przedstawiony wcześniej cel pracy.

6. 1. Hydratacja hydrofobowa i oddziaływania hydrofobowe cząstek Lennarda-Jonesa oraz pseudotwardych kul Mie w różnych modelach wody

Celem tego projektu było scharakteryzowanie termodynamiki hydratacji hydrofobowej jednego atomu Xe oraz asocjacji dwóch atomów Xe w wodzie w zakresie temperatur od 273 do 373 K. Dla każdego z tych procesów przeanalizowano również zmiany strukturalne wody w obrębie pierwszej warstwy solwatacyjnej atomów Xe w temperaturze 273 i 373 K.

Do moich zadań w tym projekcie należało wykonanie symulacji MD dla procesu hydratacji pojedynczego atomu Xe oraz asocjacji dwóch atomów Xe. Moja rola sprowadzała się również do obliczenia temperaturowych zależności termodynamicznych charakteryzujących obydwie te procesy. Ponadto dokonałam analizy struktury wody w pierwszej warstwie solwatacyjnej jednego oraz dwóch atomów Xe.

Opis procedury obliczeniowej:

Procedura obliczeniowa tego projektu badawczego składała się z kilku etapów. W każdym z nich do opisu oddziaływań Xe-Xe oraz Xe-woda wykorzystano dwa różne potencjały oddziaływań: Lennarda-Jonesa^{81,82} oraz pseudotwardych kul Mie⁹⁵:

$$U_{LJ}(r_{ij}) = 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (6.1a)$$

$$U_{Mie}(r_{ij}) = \begin{cases} 50 \left(\frac{50}{49} \right)^{49} \epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{50} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{49} \right] + \mathbf{C}, & r_{ij} < \frac{50}{49} \sigma_{ij} \\ 0, & r_{ij} \geq \frac{50}{49} \sigma_{ij} \end{cases} \quad (6.1b)$$

Zgodnie z każdym z nich atom Xe był opisywany jako niepolarna cząstka Lennarda-Jonesa (LJ-Xe) lub pseudotwarda kula Mie o średnicy atomu Xe (Mie-Xe). Badania te były prowadzone dla zakresu temperatur od 273 do 373 K z wykorzystaniem trzech modeli wody: SPC/E⁶⁵, TIP3P-Ew⁶⁶ oraz TIP4P-2005⁶⁷.

Tab. 1. Parametry pola sił dla atomu Xe¹⁴⁹ oraz modeli wody SPC/E⁶⁵, TIP3P-Ew⁶⁶ i TIP4P-2005⁶⁷.

Typ atomu	σ [Å]	ϵ [kcal/mol]	q [eu]
Xe	3.9750	0.42665	0.0000
SPC/E			
O	3.1660	0.15535	-0.8476
H	0.0000	0.00000	0.4238
TIP3P-Ew			
O	3.1880	0.10200	-0.8300
H	0.0000	0.00000	0.4150
TIP4P-2005			
O	3.1589	0.18520	0.0000
H	0.0000	0.00000	0.5564
X	0.0000	0.00000	-1.1128

W potencjale Lennarda-Jonesa parametry ϵ oraz σ dla oddziaływań przyciągających woda-Xe zostały wyznaczone na podstawie równań Barthelota-Lorentza⁹⁴:

$$\epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}} \quad \sigma_{ij} = \frac{\sigma_{ii} + \sigma_{jj}}{2} \quad (6.1c)$$

Pierwszy etap obliczeń polegał na zbadaniu procesu hydratacji hydrofobowej pojedynczego atomu Xe w wodzie. Obliczenia te były prowadzone przy użyciu metody MD^{81,82} dla pięciu temperatur: 273, 298, 323, 348 i 373 K w stałych warunkach ciśnienia $p = 1$ atm dla pudła symulacyjnego składającego się z 1000 cząsteczek trzech podanych wyżej modeli wody. Wartości temperatury i ciśnienia były kontrolowane za pomocą termostatu oraz barostatu Nose-Hoovera¹⁵⁰. Symulacje te prowadzono z zastosowaniem periodycznych warunków brzegowych. Promień odcięcia dla oddziałujących ze sobą cząstek wynosił 10 Å. Geometria cząsteczek wody podczas symulacji była utrzymywana przy pomocy algorytmu SHAKE'a¹⁵¹. Wartość kroku czasowego symulacji MD wynosiła 2 femtosekundy. Na podstawie wyników symulacji MD obliczono zmiany gęstości modeli wody, które porównano z danymi eksperymentalnymi dla prawdziwej wody (Rys. 19). Wyniki symulacji MD zostały również użyte do obliczenia zmian wartości energii swobodnej Gibbsa ΔG dla procesu hydratacji pojedynczego atomu Xe w wodzie. Obliczenia te prowadzono przy użyciu metody PI^{115,116} w ramach której atom Xe był wielokrotnie wprowadzany w różne losowe miejsca wszystkich analizowanych konformacji wody. Pozostałe wielkości termodynamiczne opisujące proces

hydratacji, takie jak entalpia ΔH oraz człon entropowy $T\Delta S$ (pomnożony przez wartość -1) zostały wyznaczone na drodze dopasowania funkcji parabolicznej do zależnej od temperatury wartości energii swobodnej ΔG .

Drugi etap obliczeń polegał na określeniu termodynamiki oddziaływań hydrofobowych dla dwóch atomów Xe w wodzie. Badania te sprowadzały się do wyznaczenia wartości potencjału średniej siły $\Delta G [r]$ względem odległości dwóch oddziałujących ze sobą atomów Xe. Do obliczeń wykorzystano metodę MD^{81,82} z próbkowaniem parasolowym^{111,112}. Symulacje te były prowadzone dla 21 temperatur z zakresu od 273 do 373 K (co 5 K każda) w stałych warunkach ciśnienia $p = 1\text{atm}$. Pudło symulacyjne było zbudowane z 1728 cząsteczek trzech modeli wody: SPC/E, TIP3P-Ew i TIP4P-2005 oraz dwóch atomów Xe. Ścieżka reakcji składała się z 25 okien, w której współrzędną reakcji była odległość Xe-Xe. Odległość równowagowa dla pary atomów Xe zmieniała się w zakresie od 3.5 Å dla pierwszego okna do 15.5 Å dla ostatniego okna symulacji. Czas trwania symulacji w każdym z tych okien wynosił 10 nanosekund. Pozostałe parametry symulacji były takie same jak dla procesu hydratacji pojedynczego atomu Xe w wodzie. Wyniki tych symulacji zostały przetworzone przy pomocy metody WHAM^{78,79} w celu określenia ogólnego profilu zmian wartości $\Delta G [r]$ dla różnych odległości Xe-Xe w każdej z badanych temperatur. Wartości $\Delta H [cm]$ oraz $T\Delta S [cm]$ dla minimum kontaktowego (ang.: *contact minimum, cm*) konfiguracji Xe-Xe zostały określone na drodze dopasowania funkcji parabolicznej do wartości $\Delta G [cm]$.

Trzeci etap badań polegał na określeniu i przeanalizowaniu struktury trzech użytych do obliczeń modeli wody w pierwszej warstwie solwatacyjnej jednego oraz dwóch atomów Xe w minimum kontaktowym ich konfiguracji. W tym celu wykonano 10 nanosekundowe symulacje MD^{81,82} dla dwóch temperatur równych 273 i 373 K. Wyniki tych symulacji zostały użyte do wygenerowania dwuwymiarowych rozkładów obejmujących:

- a) gęstość upakowania cząsteczek wody,
- b) średnią orientację przestrzenną cząsteczek wody, która była liczona jako cosinus kąta zawartego między wektorem momentu dipolowego wody i prostą prostopadłą do osi przechodzącej przez środek jednego oraz dwóch atomów Xe,
- c) średnią liczbę wiązań wodorowych między cząsteczkami wody względem promienia odcięcia $r_{\text{cut-off}} = 3.5 \text{ \AA}$ dla odległości O – O oraz $r_{\text{cut-off}} = 30^\circ$ dla kąta H – O ... O,

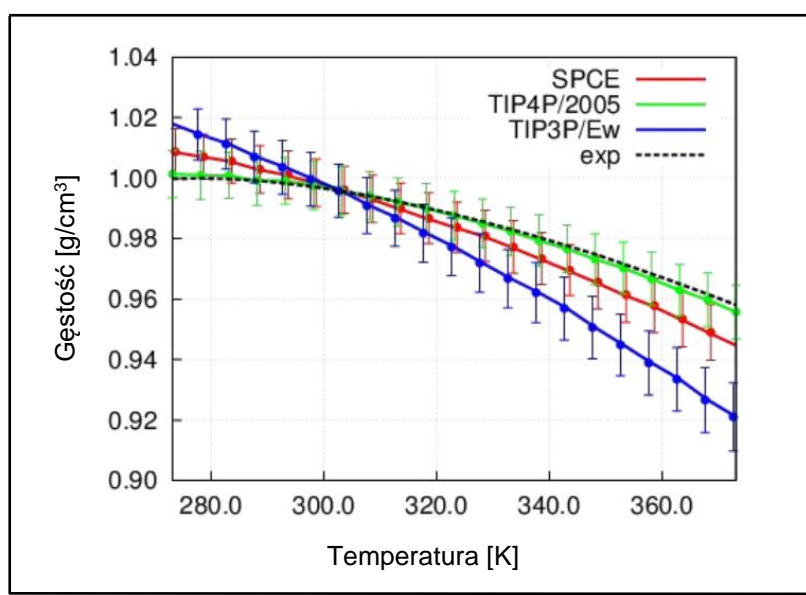
d) sposób uporządkowania tetraedrycznego cząsteczek wody, który był opisywany przez parametr q obliczony na podstawie wzoru Erringtona-Debenettiego¹⁵²:

$$q_i = 1 - (3/8) \sum_j \sum_k \left[\cos\theta_{ijk} + (1/3) \right]^2 \quad (6.1d)$$

gdzie sumowanie dotyczyło czterech najbliższych cząsteczek sąsiadujących z badaną cząsteczką i , natomiast zmienna θ_{ijk} opisywała kąt zawarty między trzema cząsteczkami wody z cząsteczką i znajdującą się pośrodku nich. Parametr q przyjmuje wartości z zakresu od 0 dla gazu idealnego do 1 dla idealnej, czworościennej sieci wiązań wodorowych w wodzie.

Wyniki obliczeń dla procesu hydratacji hydrofobowej:

Model TIP4P-2005 charakteryzował się najlepszym odtworzeniem właściwości i struktury prawdziwej wody. Pozostałe modele takie jak SPC/E oraz TIP3P-Ew wykazywały dużo większe odchylenia od gęstości eksperymentalnych wyznaczonych dla prawdziwej wody w zakresie temperatur od 273 do 373 K przy stałym ciśnieniu $p = 1$ atm (Rys. 19).



Rys. 19. Porównanie gęstości obliczonych dla trzech modeli wody: TIP3P-Ew (niebieska linia), SPC/E (czerwona linia) i TIP4P-2005 (zielona linia) z danymi eksperymentalnymi wyznaczonymi dla prawdziwej wody (czarna linia przerywana) w zakresie temperatur od 273 do 373 K przy ciśnieniu $p = 1$ atm.

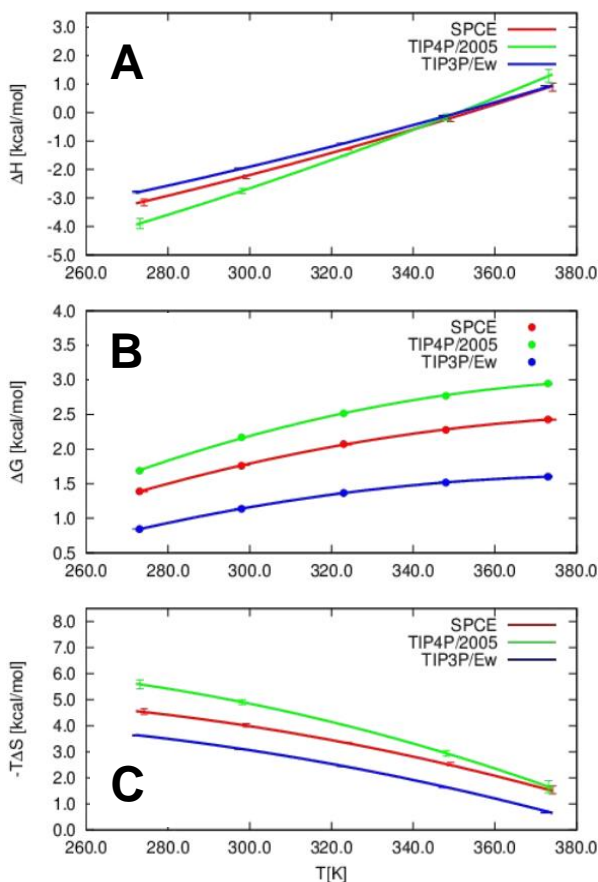
Na Rys. 20 i 21 w panelach A-C przedstawione zostały wykresy opisujące zależności zmian wartości energii swobodnej ΔG , entalpii ΔH oraz członu entropowego $T\Delta S$ (pomnożonego przez wartość -1) od temperatury dla procesu hydratacji pojedynczego atomu Xe w modelu SPC/E, TIP3P-Ew oraz TIP4P-2005.

Entalpia hydratacji ΔH cząstki LJ-Xe była początkowo ujemna i rosła wraz z temperaturą osiągając wartość 0 kcal/mol w temperaturze ok. 350 K. Powyżej tej temperatury wartości ΔH dla każdego modelu wody były dodatnie (Rys. 20, panel A). Wartości ΔG dla cząstki LJ-Xe były zawsze dodatnie – dla modelu TIP3P-Ew zmieniały się od ok. 0.8 do ok. 1.5 kcal/mol, w modelu SPC/E od ok. 1.4 do ok. 2.4 kcal/mol, a w modelu TIP4P-2005 od ok. 1.6 do ok. 2.9 kcal/mol (Rys. 20, panel B). Człon entropowy $T\Delta S$ dla procesu hydratacji cząstki LJ-Xe był początkowo ujemny i zwiększał się wraz z temperaturą. Wartości $T\Delta S$ dla modelu TIP3P-Ew zmieniały się od ok. -3.9 do -1.0 kcal/mol, dla SPC/E od ok. -4.5 do -2.0 kcal/mol, natomiast dla TIP4P-2005 od ok. -5.5 do -2.0 kcal/mol (Rys. 20, panel C). Wartości ΔG , ΔH oraz $T\Delta S$, które zostały uzyskane dla cząstki LJ-Xe w temperaturze do ok. 360 K są zgodne z danymi eksperymentalnymi dla procesu hydratacji pojedynczego atomu Xe¹⁵³.

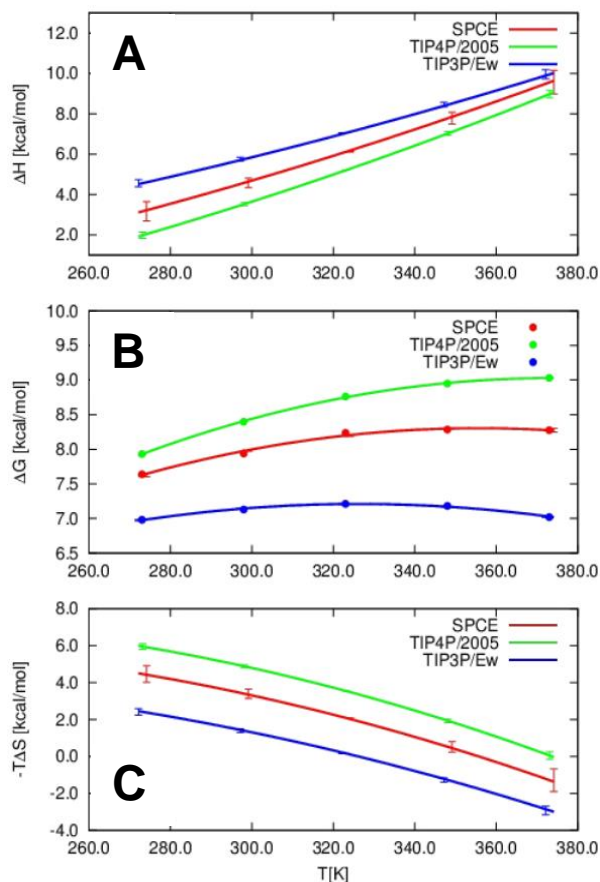
W przypadku pseudotwardych kul Mie-Xe wartości ΔH były zawsze dodatnie i rosły wraz z temperaturą. Dla modelu TIP3P-Ew wartości te zmieniały się od ok. 4.5 do 10.0 kcal/mol, dla modelu SPC/E od 3.0 do 9.5 kcal/mol, a dla modelu TIP4P-2005 od 2.0 do 9.0 kcal/mol (Rys. 21, panel A). Procesowi hydratacji pseudotwardej kuli Mie-Xe towarzyszyły zawsze dodatnie wartości ΔG , które rosły wraz z temperaturą. Dla modelu TIP3P-Ew wartości te oscylowały wokół 7 kcal/mol, w modelu SPC/E zmieniały się od ok. 7.6 do ok. 8.3 kcal/mol, natomiast w modelu TIP4P-2005 zmiany te plasowały się w zakresie od ok. 7.9 do ok. 9.1 kcal/mol (Rys. 21, panel B). Wartości $T\Delta S$ były początkowo ujemne i rosły wraz z temperaturą. W modelu TIP3P-Ew wartości te zmieniały się od ok. -2.5 do ok. 2.0 kcal/mol, natomiast w modelu SPC/E od ok. -4.5 do 1.0 kcal/mol. Najwyższe wartości $T\Delta S$ zostały osiągnięte w modelu TIP4P-2005, gdzie zmieniały się od -6.0 kcal/mol w temperaturze 273 K do 0 kcal/mol w temperaturze 373 K (Rys. 21, panel C). W modelu TIP4P-2005 w zakresie temperatur od 273 do ok. 350 K osiągnięto podobne wartości $T\Delta S$ dla procesu hydratacji cząstki LJ-Xe oraz pseudotwardej kuli Mie-Xe. Było to wynikiem tego, że model ten jako jedyny spośród użytych do obliczeń najlepiej odtwarzał strukturę i właściwości prawdziwej wody (Rys. 19).

Termodynamika procesu hydratacji dla pseudotwardej kuli Mie-Xe bardzo różniła się od uzyskanej dla cząstki LJ-Xe. Można to wytłumaczyć tym, że pseudotwarde kule Mie o rozmiarach atomu Xe są bardziej hydrofobowe od cząstek LJ-Xe, ponieważ nie są zdolne do oddziaływań przyciągających z sąsiednimi cząsteczkami wody. Dzięki temu lepiej nadają się

do badania efektu hydrofobowego substancji niepolarnych rozpuszczanych w wodzie, ponieważ termodynamika hydratacji substancji zależy jedynie od właściwości i struktury wody.



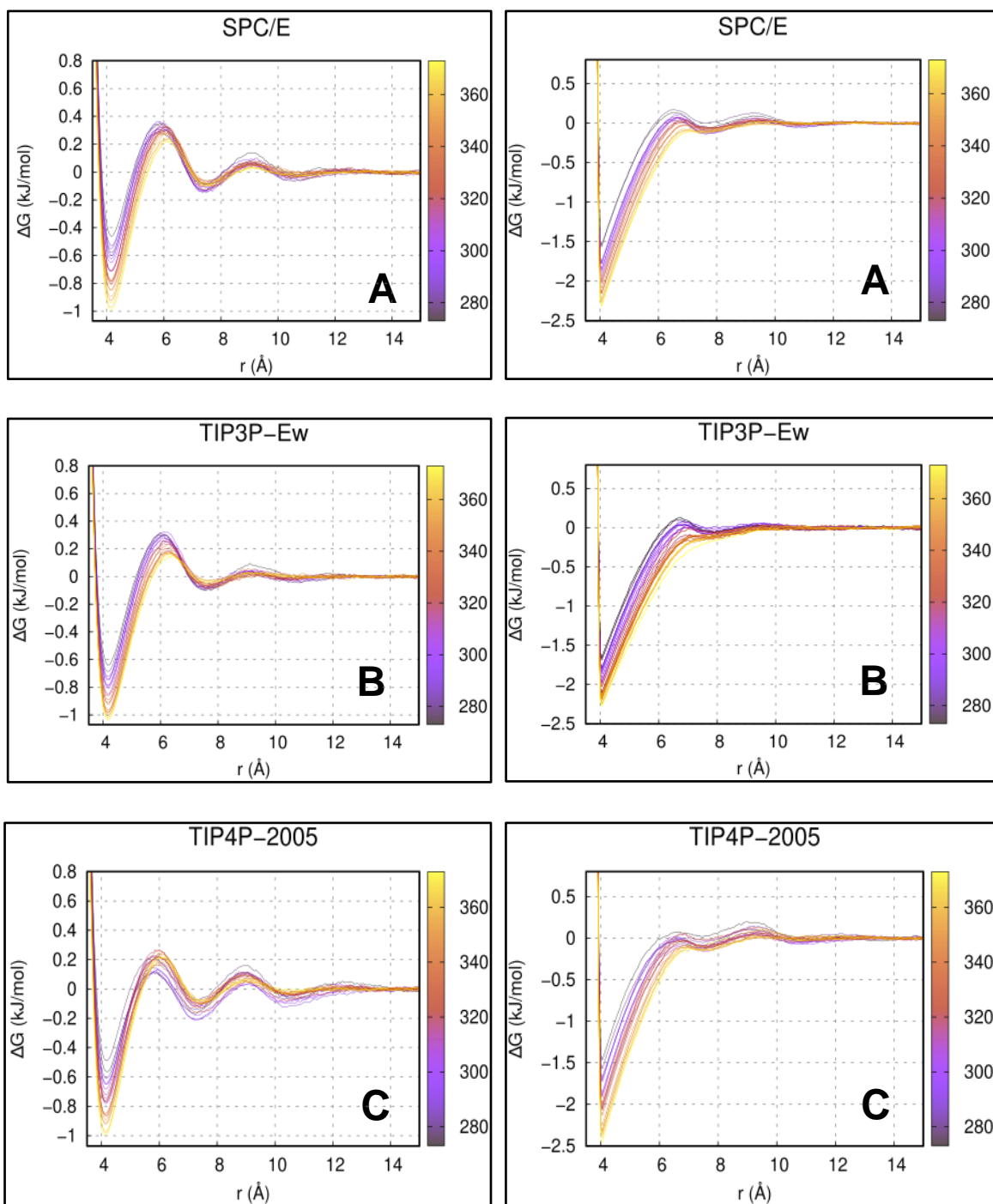
Rys. 20. Wykresy przedstawiające zakres zmian wartości ΔH (panel A), ΔG (panel B) oraz $T\Delta S$ (panel C) względem temperatury dla procesu hydratacji cząstki LJ-Xe symulowanej w modelu TIP3P-Ew, SPC/E oraz TIP4P-2005.



Rys. 21. Wykresy przedstawiające zakres zmian wartości ΔH (panel A), ΔG (panel B) oraz $T\Delta S$ (panel C) względem temperatury dla procesu hydratacji pseudotwardej kuli Mie-Xe symulowanej w modelu TIP3P-Ew, SPC/E oraz TIP4P-2005.

Wyniki obliczeń dla oddziaływań hydrofobowych

Na Rys. 22 i 23 zostały przedstawione wykresy opisujące zmiany wartości potencjału średniej siły $\Delta G [r]$ dla oddziaływań hydrofobowych dwóch cząstek LJ-Xe oraz pseudotwardych kul Mie-Xe symulowanych w modelu SPC/E (panel A), TIP3P-Ew (panel B) oraz TIP4P-2005 (panel C) w zakresie temperatur od 273 do 373 K (co 5 K każda) oraz w stałych warunkach ciśnienia $p = 1$ atm.



Rys. 22. Wykresy przedstawiające zależność zmian wartości potencjału średniej siły $\Delta G[r]$ względem odległości dwóch oddziałujących ze sobą cząstek LJ-Xe w zakresie temperatur od 273 do 373 K.

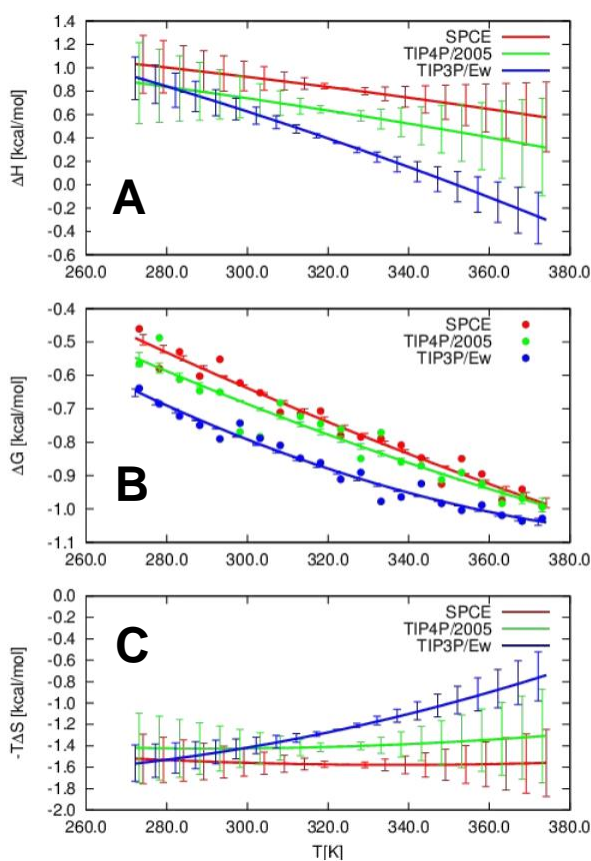
Rys. 23. Wykresy przedstawiające zależność zmian wartości potencjału średniej siły $\Delta G[r]$ względem odległości dwóch oddziałujących ze sobą pseudotwardych kul Mie-Xe w zakresie temperatur od 273 do 373 K.

Pierwsze minimum widoczne na Rys. 22 i 23 opisuje tzw. minimum kontaktowe konfiguracji, czyli najmniejszą odległość w jakiej mogą ze sobą oddziaływać dwa atomy Xe w wodzie. Położenie tego minimum jest niezależne od temperatury i rodzaju użytego do obliczeń modelu wody – wzrost temperatury przyczynia się jedynie do obniżenia wartości $\Delta G[r]$ dla dwóch atomów Xe w minimum kontaktowym ich konfiguracji. Pierwsze maksimum

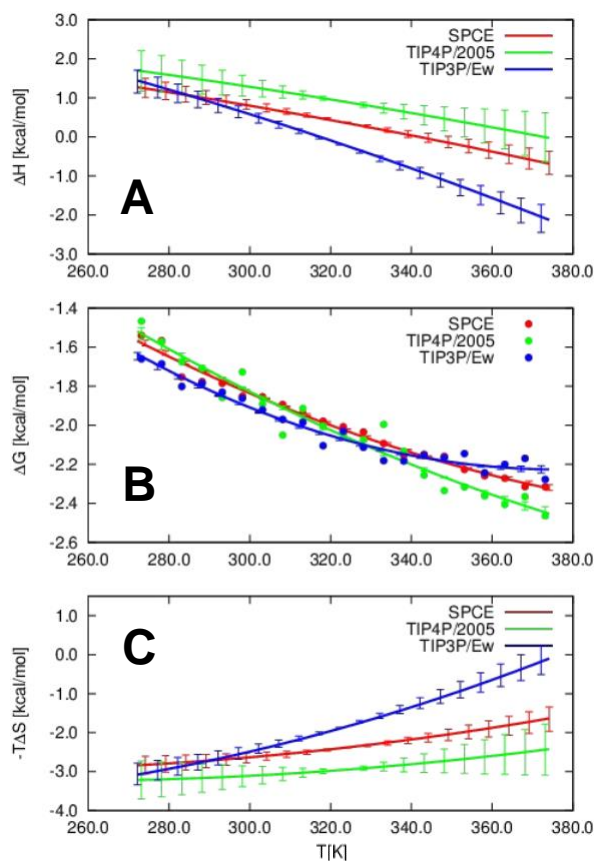
widoczne na każdym z tych wykresów opisuje tzw. barierę desolwatacyjną, czyli taką konfigurację dwóch atomów Xe do której nie może wniknąć żadna sąsiadująca z nimi cząsteczka wody. Drugie minimum określane jako minimum separowanego rozpuszczalnika charakteryzuje dalekozasięgową konfigurację atomów Xe między którymi znajduje się jedna cząsteczka wody.

W każdym z trzech użytych do obliczeń modeli wody minimum kontaktowe dla konfiguracji dwóch cząstek LJ-Xe zostało osiągnięte dla odległości ok. 4.2 Å, bariera desolwatacyjna dla odległości ok. 6.0 Å, a minimum separowanego rozpuszczalnika w okolicy 7.2 Å (Rys. 22, panel A-C). W przypadku cząstek Mie-Xe pierwsze minimum zostało osiągnięte dla odległości ok. 4.0 Å, bariera desolwatacyjna dla odległości ok. 7.0 Å, a minimum separowanego rozpuszczalnika w okolicy 8.0 Å (Rys. 23, panel A-C).

Na Rys. 24 i 25 znajdują się wykresy przedstawiające zależności zmian wartości ΔG [cm] (panel A), ΔH [cm] (panel B) oraz $T\Delta S$ [cm] (panel C) względem temperatury w minimum kontaktowym konfiguracji dwóch cząstek LJ-Xe oraz pseudotwardych kul Mie-Xe.



Rys. 24. Wykresy przedstawiające zakres zmian wartości ΔH [cm] (panel A), ΔG [cm] (panel B) oraz $T\Delta S$ [cm] (panel C) względem temperatury w minimum kontaktowym dwóch cząstek LJ-Xe w modelu TIP3P-Ew, SPC/E oraz TIP4P-2005.



Rys. 25. Wykresy przedstawiające zakres zmian wartości ΔH [cm] (panel A), ΔG [cm] (panel B) oraz $T\Delta S$ [cm] (panel C) względem temperatury w minimum kontaktowym dwóch pseudotwardych kul Mie-Xe w modelu TIP3P-Ew, SPC/E oraz TIP4P-2005.

Wartości ΔG [cm] dla dwóch cząstek LJ-Xe (Rys. 24, panel B) i pseudotwardych kul Mie-Xe (Rys. 25, panel B) w minimum kontaktowym ich konfiguracji zmniejszały się wraz ze wzrostem temperatury. W każdym z użytych do obliczeń modeli wody mniejsze wartości ΔG [cm] zostały osiągnięte dla dwóch pseudotwardych kul Mie-Xe. Wynik ten świadczy o większej sile oddziaływań przyciągających między pseudotwardymi kulami Mie-Xe w porównaniu do cząstek LJ-Xe. W modelu TIP4P-2005, który najlepiej odtwarzał strukturę i właściwości prawdziwej wody wartości ΔG [cm] mieściły się w zakresie od -0.55 do -0.95 kcal/mol dla cząstek LJ-Xe oraz od -1.45 do -2.45 kcal/mol dla pseudotwardych kul Mie-Xe.

Wartości ΔH [cm] dla cząstek LJ-Xe oraz pseudotwardych kul Mie-Xe zmniejszały się wraz z temperaturą (Rys. 24 i 25, panel A). Dla cząstek LJ-Xe w modelu SPC/E i TIP4P-2005 wartości te były zawsze dodatnie, natomiast w modelu TIP3P-Ew i temperaturze powyżej 350 K entalpia oddziaływań hydrofobowych zaczęła przyjmować wartości ujemne. W przypadku pseudotwardych kul Mie-Xe jedynie w modelu TIP4P-2005 wartości ΔH [cm] były zawsze dodatnie i większe niż dla cząstek LJ-Xe.

Człon entropowy $T\Delta S$ [cm] dla cząstek LJ-Xe oraz pseudotwardych kul Mie-Xe był zawsze dodatni i zwiększał się wraz z temperaturą (Rys. 24 i 25, panel C). We wszystkich trzech modelach wody większe wartości $T\Delta S$ [cm] zostały osiągnięte dla pseudotwardych kul Mie-Xe – największe wartości oscylujące w ok. 3.0 kcal/mol zostały uzyskane w modelu TIP4P-2005 (Rys. 25, panel C). Dodatnie wartości $T\Delta S$ [cm] dla oddziaływań hydrofobowych dwóch cząstek LJ-Xe i pseudotwardych kul Mie-Xe dowodzą temu, że proces tworzenia minimum kontaktowego ich konfiguracji był całkowicie lub częściowo sterowany przez entropię.

Dla cząstek LJ-Xe oraz pseudotwardych kul Mie-Xe największe odchylenia od wartości ΔH [cm] oraz $T\Delta S$ [cm] zostały uzyskane dla temperatury 273 oraz 373 K i wynikały one ze sposobu dopasowywania funkcji parabolicznej do energii swobodnej ΔG [cm] w minimum kontaktowym konfiguracji Xe-Xe (Rys. 24 i 25, panel A i C).

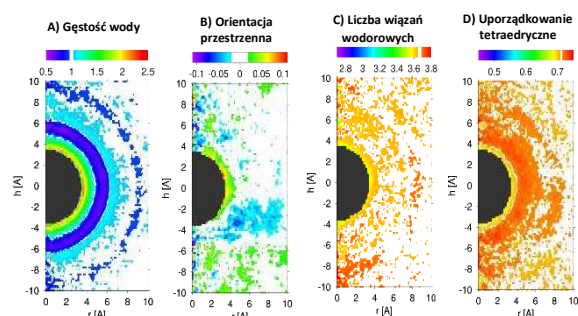
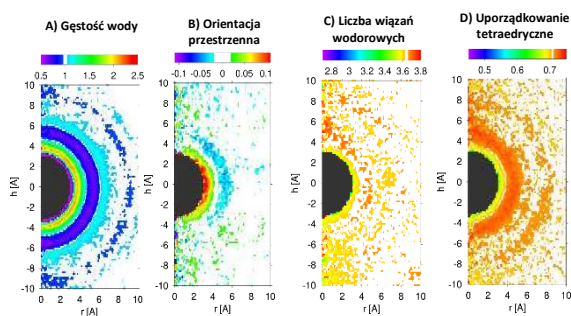
Analiza struktury wody

Na Rys. 26-29 zostały przedstawione dwuwymiarowe wykresy opisujące strukturę wody w pierwszej warstwie solwatacyjnej jednej oraz dwóch cząstek LJ-Xe i pseudotwardych kul Mie-Xe w modelu TIP4P-2005 i temperaturze 273 i 373 K.

Każdy z rozkładów został uśredniony względem kąta azymutalnego i wyrażony we współrzędnych cylindrycznych za pomocą osi \mathbf{h} przechodzącej przez środek dwóch atomów Xe i prostopadłej do niej osi \mathbf{r} ; w przypadku pojedynczego atomu Xe współrzędne te zostały zdefiniowane wokół osi o współrzędnych (1,1,1). Dla uproszczenia analizy struktury wody w każdym z rozkładów zastosowano tę samą paletę barw, w której kolor biały opisuje średnią wartość badanej wielkości dla cząsteczek wody znajdujących się w dużych odległościach względem atomów Xe.

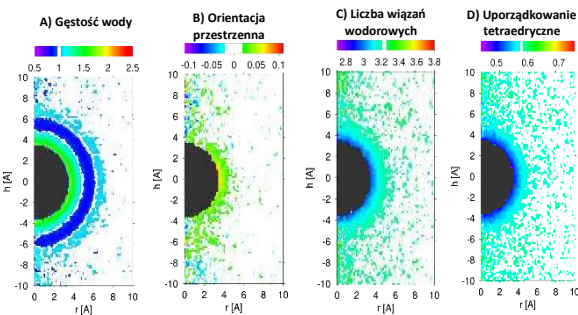
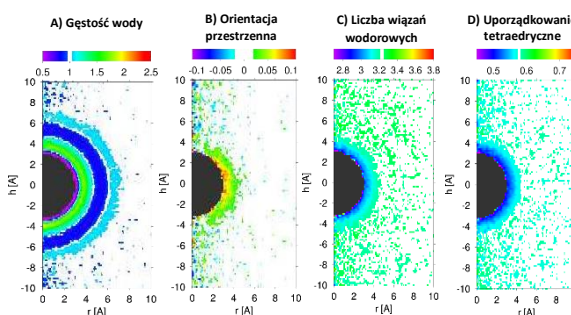
LJ-Xe ($T = 273$ K)

Mie-Xe ($T = 273$ K)



LJ-Xe ($T = 373$ K)

Mie-Xe ($T = 373$ K)

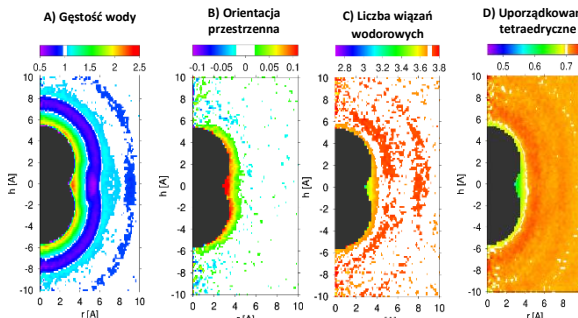
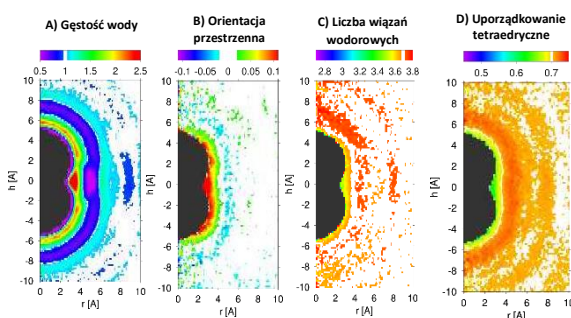


Rys. 26. Rozkłady gęstości, średniej orientacji przestrzennej, średniej liczby wiązań wodorowych oraz uporządkowania tetraedycznego cząstek wody wokół pojedynczej cząstki LJ-Xe w $T = 273$ K i $T = 373$ K.

Rys. 27. Rozkłady gęstości, średniej orientacji przestrzennej, średniej liczby wiązań wodorowych oraz uporządkowania tetraedycznego wody wokół pseudotwardej kuli Mie-Xe w $T = 273$ K i $T = 373$ K.

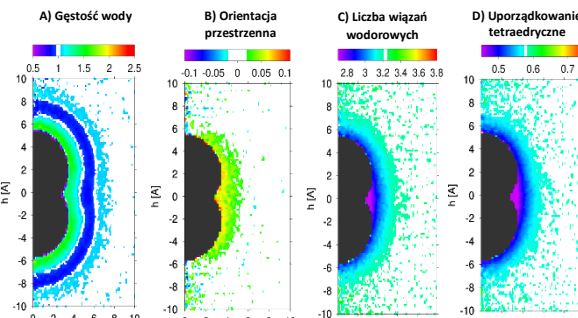
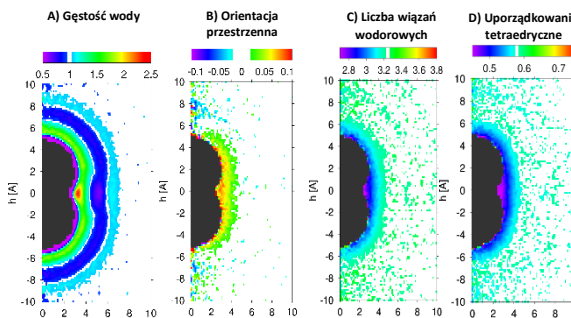
LJ-Xe ($T = 273$ K)

Mie-Xe ($T = 273$ K)



LJ-Xe ($T = 373$ K)

Mie-Xe ($T = 373$ K)



Rys. 28. Rozkłady gęstości, średniej orientacji przestrzennej, średniej liczby wiązań wodorowych oraz uporządkowania tetraedycznego cząstek wody wokół dwóch cząstek LJ-Xe w $T = 273$ K i $T = 373$ K.

Rys. 29. Rozkłady gęstości, średniej orientacji przestrzennej, średniej liczby wiązań wodorowych oraz uporządkowania tetraedycznego cząstek wody wokół dwóch pseudotwardych kul Mie-Xe w $T = 273$ K i $T = 373$ K.

Gęstość upakowania cząsteczek wody zależała od temperatury i sposobu reprezentacji atomów Xe (Rys. 26-29, panel A). Większe gęstości wody zostały zaobserwowane w temperaturze 273 K dla jednej (Rys. 26A) oraz dwóch cząstek LJ-Xe (Rys. 28A). Sposób upakowania cząsteczek wody wokół dwóch cząstek LJ-Xe (Rys. 28A) i pseudotwardych kul Mie-Xe (Rys. 29A) w minimum kontaktowym ich konfiguracji był dość nierównomierny – można to było zaobserwować na przykładzie wody gromadzącej się w środkowym obszarze nakładania się ich powłok solwatacyjnych. W przypadku cząstek LJ-Xe akumulacja ta była widoczna w obydwu temperaturach, a dla pseudotwardych kul Mie-Xe udało się ją zaobserwować jedynie w temperaturze 273 K.

Większy stopień średniego uporządkowania przestrzennego wody został osiągnięty w temperaturze 273 K dla jednej oraz dwóch pseudotwardych kul Mie-Xe (Rys. 29, panel B).

W pierwszej warstwie solwatacyjnej jednej oraz dwóch cząstek LJ-Xe i pseudotwardych kul Mie-Xe zaobserwowano mniejszą liczbę wiązań wodorowych niż w strukturze czystej wody (Rys. 26-29, panel C). Było to konsekwencją tego, że obecność tych cząstek w roztworze ograniczała ilość dostępnej przestrzeni do tworzenia wiązań wodorowych między cząsteczkami wody. Większa liczba wiązań wodorowych została osiągnięta w temperaturze 273 K wokół jednej oraz dwóch pseudotwardych kul Mie-Xe (Rys. 26-29, panel C). Podczas przejścia do temperatury 373 K nastąpił spadek średniej liczby tych wiązań między cząsteczkami wody.

Uporządkowanie tetraedryczne wody było większe w pierwszej warstwie solwatacyjnej jednej oraz dwóch cząstek LJ-Xe oraz Mie-Xe niż w strukturze czystej wody. Wyższe wartości parametru q opisującego uporządkowanie tetraedryczne wody zostały osiągnięte dla niższej z analizowanych temperatur.

Uzyskane wyniki są zgodne z teoretycznymi założeniami modelu góry lodowej Franka i Evansa³². Energia pojedynczego wiązania wodorowego między cząsteczkami wody była większa w pierwszej warstwie solwatacyjnej niepolarnych cząstek niż w strukturze czystej wody. Towarzyszyły temu dodatnie wartości entalpii $\Delta H [cm]$ i członu entropowego $T\Delta S [cm]$ w minimum kontaktowym konfiguracji dwóch cząstek LJ-Xe oraz pseudotwardych Mie-Xe. Zależności te są jednak najlepiej zobrazowane w modelu TIP4P-2005 dla pseudotwardych kul Mie-Xe.

Wyniki moich badań dotyczących termodynamiki hydratacji i oddziaływań hydrofobowych dla cząstek LJ-Xe oraz pseudotwardych kul Mie-Xe w modelach SPC/E, TIP3P-Ew oraz TIP4P-2005 znajdują się w załączniku 1 stanowiącym publikację „Hydrophobic hydration and pairwise hydrophobic interaction of Lennard-Jones and Mie particles in different water models”, Zięba K., Czaplewski C., Liwo A., Graziano G., *Phys. Chem. Chem. Phys.*, 2020, **22**, 4758-4771.

6. 2. Modelowanie molekularne zależności energii swobodnej hydratacji od rozmiaru niepolarnej cząstki

Celem tego projektu było określenie zależności między temperaturą, a energią swobodną Gibbsa ΔG , entalpią ΔH , członem entropowym $T\Delta S$ oraz pojemnością cieplną ΔC_v dla procesu hydratacji niepolarnej cząstki w wodzie, której średnica σ_X w trakcie symulacji zmieniała się w zakresie od 3 do 18 Å. W ramach projektu przeanalizowano również zmiany strukturalne wody w pierwszej warstwie solwatacyjnej tej cząstki o średnicy σ_X równej 3 oraz 18 Å w temperaturze 273 i 373 K.

Moja rola w tym projekcie sprowadzała się do wykonania symulacji MD^{81,82} dla procesu hydratacji niepolarnej rosnącej cząstki, obliczenia wartości energii swobodnej $\Delta G(T)$ hydratacji tej cząstki przy pomocy metody FEP¹¹² oraz PI^{115,116}, określenia pozostałych wielkości termodynamicznych charakteryzujących ten proces takich jak ΔH , $T\Delta S$ oraz ΔC_v i ostatecznie do przeanalizowania zmian strukturalnych wody w pierwszej warstwie solwatacyjnej badanej cząstki.

Opis procedury obliczeniowej:

Do obliczeń wykorzystano dwa odmienne typy potencjałów oddziaływań – pierwszym z nich był potencjał Lennarda-Jonesa^{81,82}, a drugim potencjał pseudotwardych kul Mie⁹⁵. W zależności od użytego do obliczeń typu potencjału niepolarna cząstka była opisywana jako realna cząstka Lennarda-Jonesa lub pseudotwarda kula Mie niezdolna do oddziaływań przyciągających z sąsiednimi cząsteczkami wody. Energia swobodna ΔG dla procesu hydratacji niepolarnej cząstki w wodzie została wyznaczona przy pomocy metody MD^{81,82} oraz FEP¹¹² lub PI^{115,116}. Obliczenia te zostały wykonane dla pięciu temperatur: 273, 298, 323, 348 i 373 K w stałych warunkach ciśnienia $p = 1 \text{ atm}$.

Pierwszy etap badań polegał na obliczeniu średniej względnej wartości ΔG dla procesu hydratacji niepolarnej cząstki w wodzie przy pomocy metody MD^{81,82} oraz FEP¹¹². W tym przypadku niepolarna cząstka została umieszczona w pudle symulacyjnym składającym się z 1728 cząsteczek dwóch różnych modeli wody takich jak SPC/E⁶⁵ i TIP4P-2005⁶⁷, których parametry pól sił zostały przedstawione w Tabeli 1 w rozdziale 6.1. W metodzie FEP średnica niepolarnej cząstki σ_{X-O_w} była liczona względem atomu tlenu O_w sąsiedniej cząsteczki wody w następujący sposób:

$$\sigma_{X-O_w} = \frac{\sigma_X + \sigma_{O_w}}{2} \quad (6.2a)$$

gdzie σ_X to bezwzględna średnica niepolarniej cząstki przyjmująca wartości z zakresu od 3 do 18 Å, natomiast σ_{Ow} to średnica atomu tlenu sąsiedniej cząsteczki wody, której wartość była zależna od rodzaju użytego do obliczeń modelu wody. Wartość średnicy σ_{X-Ow} w trakcie symulacji zmieniała się od 3 do 11 Å ze stałą sprężenia $\lambda = 0.01$ oraz krokiem $\Delta\sigma_{X-Ow}$ równym 0.08 Å i była liczona w następujący sposób:

$$\Delta\sigma_{X-Ow} = (11.0\text{Å} - 3.0\text{Å})\lambda + 3.0\text{Å} \quad (6.2b)$$

Głębokość studni ϵ_{X-Ow} dla każdego potencjału oddziaływań była zawsze stała i przyjmowała wartość równą 0.1 kcal/mol. Symulacje MD składały się z 20 000 000 kroków z krokiem czasowym wynoszącym 2 femtosekundy. Do obliczeń wykorzystano technikę periodycznych warunków brzegowych z promieniem odcięcia $r_{cut-off}$ wynoszącym 10 Å. Temperatura i ciśnienie panujące w symulowanym układzie były kontrolowane przy pomocy termostatu oraz barostatu Nose-Hoovera¹⁵⁰, a geometria cząsteczek wody była kontrolowana przy pomocy algorytmu SHAKE'a¹⁵¹. Ścieżka reakcji została podzielona na 100 okien liczących po 200 000 kroków. W każdym z nich pierwsze 40 000 kroków zostało poświęcone na zrównoważenie energii całkowitej układu. Dane zebrane z ostatnich 160 000 kroków symulacji zostały uśrednione, a następnie użyte do obliczenia średniej względnej wartości ΔG dla procesu hydratacji rosnącej niepolarniej cząstki. Parametrem zaburzenia energii swobodnej ΔG była średnica σ_{X-Ow} niepolarniej cząstki. Pozostałe wielkości termodynamiczne takie jak ΔH , $T\Delta S$ oraz ΔC_v zostały wyznaczone na drodze dopasowania funkcji parabolicznej do wartości ΔG odpowiadającej kilku wybranym średnicom σ_{X-Ow} niepolarniej cząstki równym: 3.08, 3.56, 4.04, 5.96, 9.00 i 11.00 Å.

Drugi etap badań polegał na obliczeniu średniej bezwzględnej wartości ΔG dla procesu hydratacji rosnącej niepolarniej cząstki o średnicy bezwzględnej σ_X w zakresie od 3 do 6 Å (z krokiem 0.5 Å). W przypadku większych rozmiarów analizowanej cząstki struktura wody nie była w stanie samodzielnie wytworzyć adekwatnej do jej średnicy wielkości hydrofobowej wnęki, co przekładało się na dążące do nieskończoności wartości ΔG jej hydratacji. Do obliczeń wykorzystano metodę MD^{81,82} oraz PI^{115,116}, gdzie niepolarna cząstka została umieszczona w pudle symulacyjnym składającym się z 1000 cząsteczek dwóch różnych modeli wody – SPC/E⁶⁵ oraz TIP4P-2005⁶⁷. Symulacje MD składały się z 10 000 000 kroków z krokiem czasowym równym 2 femtosekundy. Pozostałe parametry symulacji były takie same jak w poprzednio opisywanym etapie obliczeniowym. Dla każdej z analizowanych temperatur wygenerowano po 10 000 konformacji struktury wody do których w losowe miejsca wprowadzano łącznie 1 mln razy niepolarną cząstkę. Wartości ΔG dla kilku wybranych średnic σ_X równych 3.00, 3.50 i 4.00 Å zostały porównane z wartościami ΔG z metody FEP dla cząstki

o średnicy σ_{X-Ow} równej 3.08, 3.56 oraz 4.04 Å. Porównanie to umożliwiło określenie bezwzględnej wartości ΔG dla procesu hydratacji niepolarniej cząstki w wodzie, która była liczona w następujący sposób:

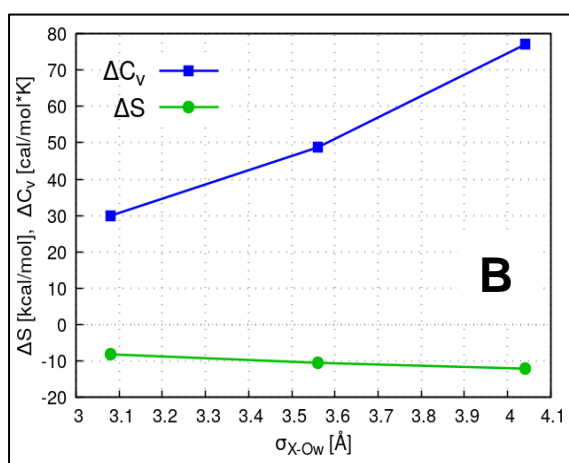
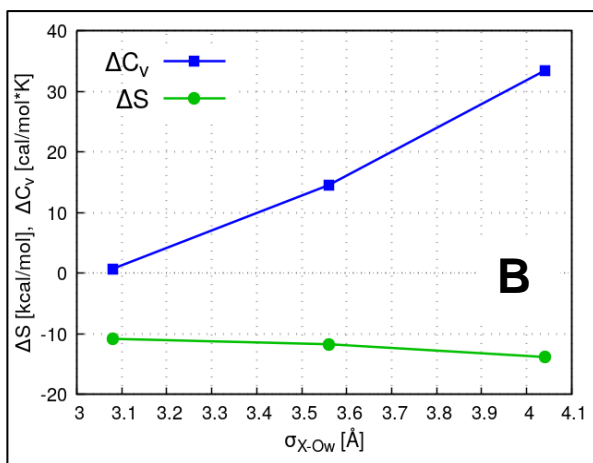
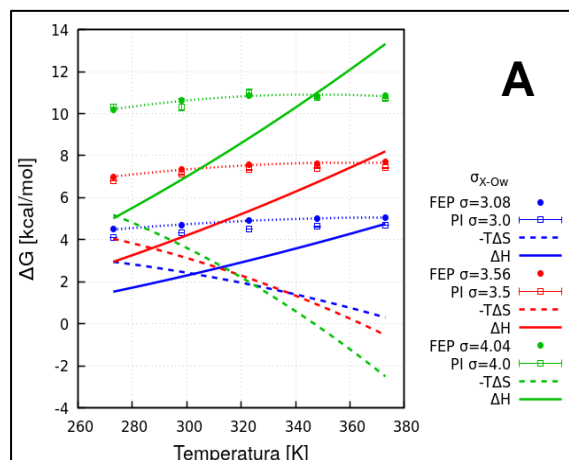
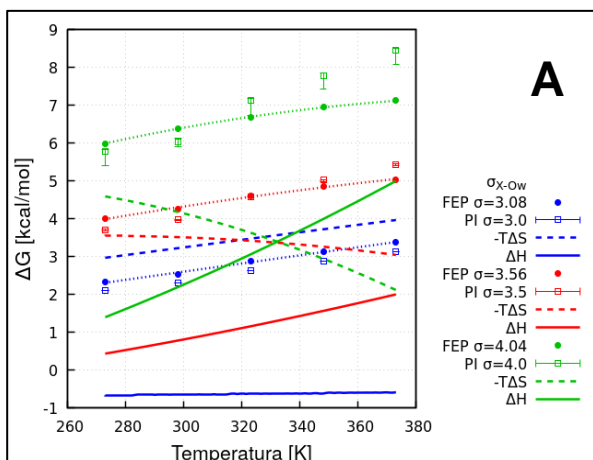
$$\Delta G = \Delta G^{FEP} + \Delta G_{\sigma_x=3.0\text{\AA}}^{PI} \quad (6.2c)$$

gdzie ΔG^{FEP} to średnia względna energia swobodna hydratacji, a ΔG^{PI} to średnia bezwzględna energia swobodna hydratacji liczona przy pomocy metody PI dla małych cząstek od średnicy σ_x równej 3 Å.

Ostatni etap badań polegał na przeanalizowaniu struktury wody w pierwszej warstwie solwatacyjnej dużej ($\sigma_{X-Ow} = 11.00$ Å) oraz małej ($\sigma_{X-Ow} = 3.08$ Å) cząstki Lennarda-Jonesa lub pseudotwardej kuli Mie umieszczonych w środowisku 1728 cząsteczek dwóch modeli wody: SPC/E⁶⁵ i TIP4P-2005⁶⁷ w temperaturze 273 i 373 K. Trajektorie uzyskane z symulacji MD dla obydwu tych cząstek zostały użyte do przeprowadzenia analizy porównawczej w pierwszej warstwie solwatacyjnej każdej z nich pod kątem: a) gęstości upakowania cząsteczek wody, b) kątovej orientacji cząsteczek wody, c) liczby wiązań wodorowych między cząsteczkami wody oraz d) sposobu uporządkowania tetraedrycznego cząsteczek wody. Każda z wymienionych właściwości wody została wyznaczona w ten sam sposób jak w przypadku procesu hydratacji pojedynczego atomu Xe opisywanego w rozdziale 6.1.

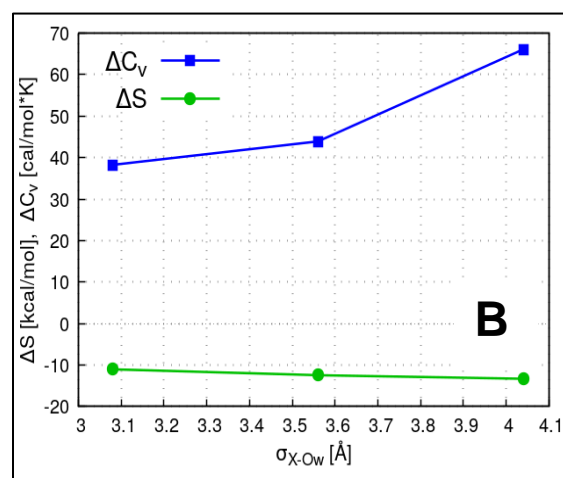
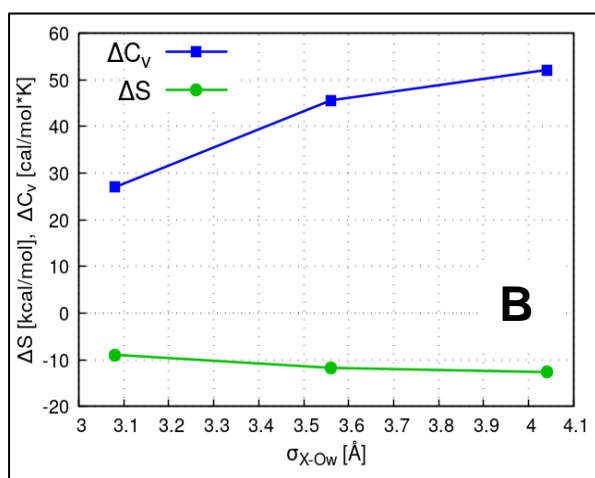
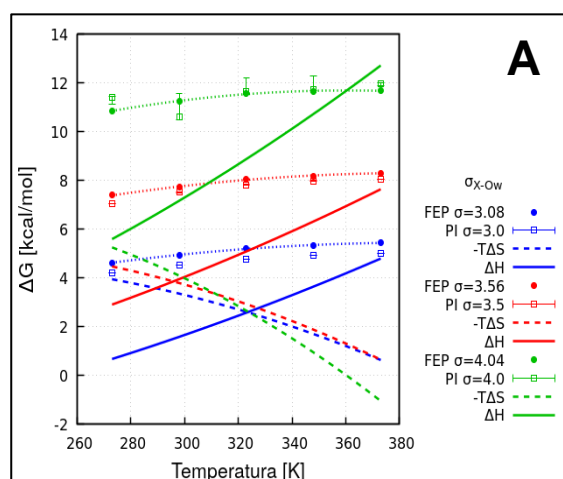
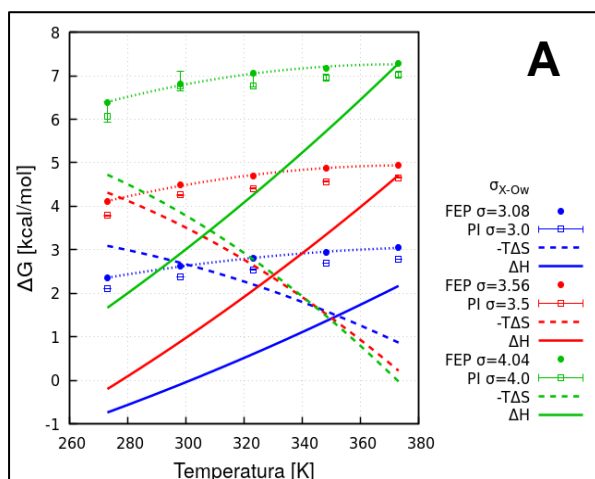
Wyniki obliczeń:

Na Rys. 30-33 w panelach A oraz B zostały zamieszczone wykresy przedstawiające zależności zmian wartości ΔG , ΔH , $T\Delta S$ oraz ΔC_v obliczone przy pomocy metody FEP względem temperatury dla procesu hydratacji niepolarniej cząstki Lennarda-Jonesa oraz pseudotwardej kuli Mie o średnicy względnej σ_{X-Ow} równej 3.08, 3.56 i 4.04 Å dla modelu SPC/E (Rys. 30 i 31) oraz TIP4P-2005 (Rys. 32 i 33). Ponadto na każdym z tych wykresów zostały zamieszczone wartości $\Delta G(T)$ obliczone przy pomocy metody PI dla obydwu typów cząstek o średnicy bezwzględnej σ_x równej 3.0, 3.5 oraz 4.0 Å.



Rys. 30. Panel A przedstawia temperaturowe zależności zmian wartości ΔG , ΔH oraz $T\Delta S$ dla procesu hydratacji rosnącej cząstki Lennarda-Jonesa w modelu SPC/E. Na panelu B przedstawiono zależności ΔC_v oraz ΔS względem rozmiaru rosnącej cząstki Lennarda-Jonesa w modelu SPC/E dla zakresu temperatur od 273 do 373 K.

Rys. 31. Panel A przedstawia temperaturowe zależności zmian wartości ΔG , ΔH oraz $T\Delta S$ dla procesu hydratacji rosnącej pseudo twardej kuli Mie w modelu SPC/E. Na panelu B przedstawiono zależności ΔC_v oraz ΔS względem rozmiaru rosnącej pseudo twardej kuli Mie w modelu SPC/E dla zakresu temperatur od 273 do 373 K.



Rys. 32. Panel A przedstawia temperaturowe zależności zmian wartości ΔG , ΔH oraz $T\Delta S$ dla procesu hydratacji rosnącej cząstki Lennarda-Jonesa w modelu TIP4P-2005. Na panelu B przedstawiono zależności ΔC_v oraz ΔS względem rozmiaru rosnącej cząstki Lennarda-Jonesa w modelu TIP4P-2005 dla zakresu temperatur od 273 do 373 K.

Rys. 33. Panel A przedstawia temperaturowe zależności zmian wartości ΔG , ΔH oraz $T\Delta S$ dla procesu hydratacji rosnącej pseudotwardej kuli Mie w modelu TIP4P-2005. Na panelu B przedstawiono zależności ΔC_v oraz ΔS względem rozmiaru rosnącej pseudotwardej kuli Mie w modelu TIP4P-2005 dla zakresu temperatur od 273 do 373 K.

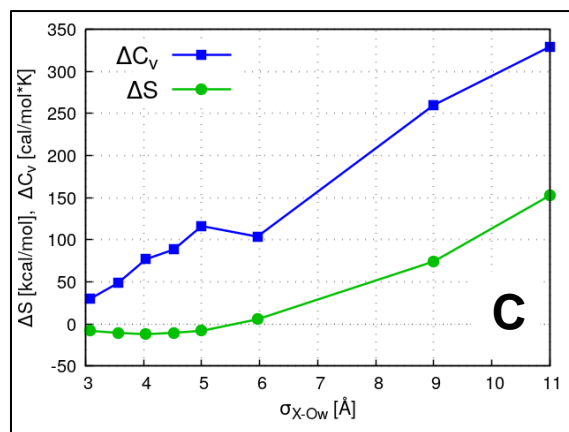
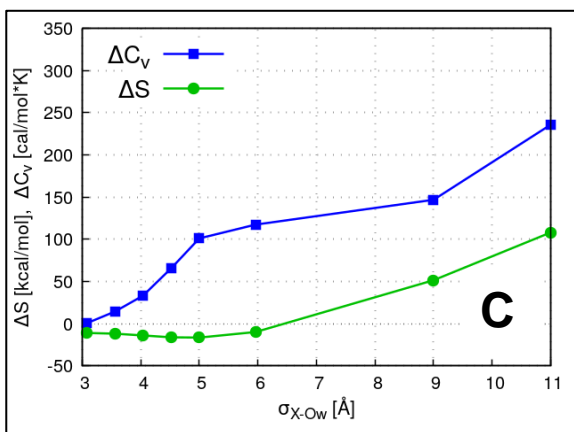
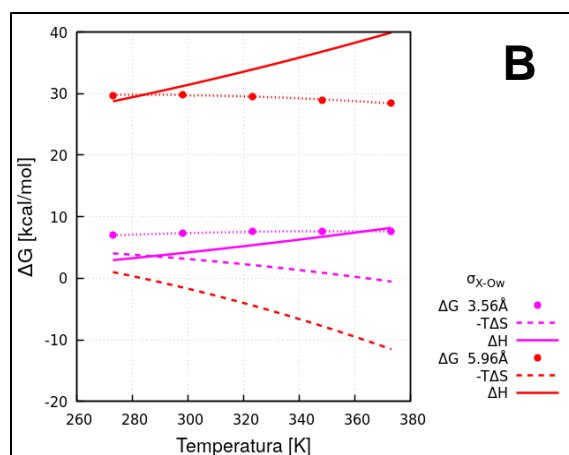
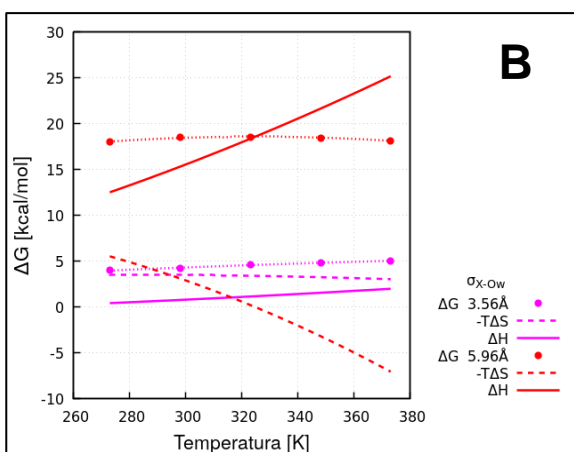
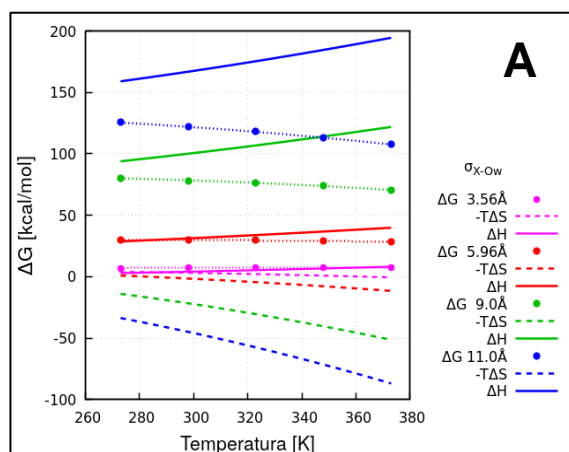
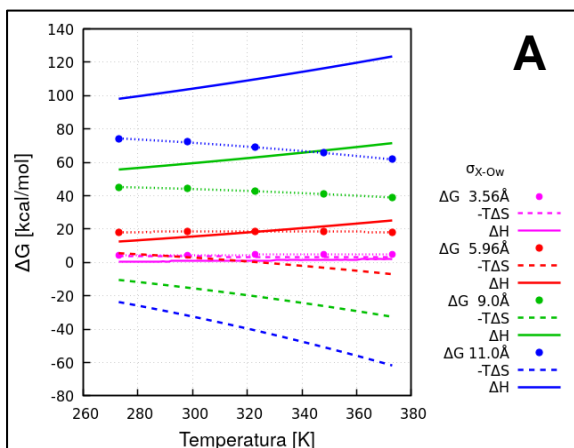
W procesie hydratacji małej cząstki o średnicy σ_{X-Ow} od 3.08 do 4.04 Å wzrostowi temperatury układu w obydwu modelach wody towarzyszył wzrost wartości ΔH , $T\Delta S$ oraz ΔG (Rys. 30-33, panel A). Wyższe wartości każdej z tych wielkości zostały osiągnięte dla pseudotwardej kuli Mie (Rys. 31 i Rys. 33, panel A). Entalpia hydratacji pseudotwardej kuli Mie była zawsze dodatnia i podobna w obydwu modelach wody. Ujemne wartości ΔH zostały osiągnięte dla cząstki Lennarda-Jonesa o średnicy $\sigma_{X-Ow} = 3.08$ Å w zakresie temperatur od 273 do 373 K w modelu SPC/E (Rys. 30, panel A) oraz w modelu TIP4P-2005 i temperaturze poniżej 300 K (Rys. 32, panel A). Ujemne wartości ΔH zostały również zaobserwowane dla cząstki Lennarda-Jonesa o średnicy σ_{X-Ow} równej 3.56 Å w modelu TIP4P-2005 i temperaturze poniżej 280 K.

Dla pseudotwardej kuli Mie w modelu SPC/E i TIP4P-2005 osiągnięto bardzo podobne wartości członu entropowego $T\Delta S$ (Rys. 31 i Rys. 33, panel A). Najwyższe wartości $T\Delta S$ zostały uzyskane w modelu TIP4P-2005 i temperaturze 373 K dla pseudotwardej kuli Mie o średnicy σ_{X-Ow} równej 4.04 Å (Rys. 33, panel A). Proces hydratacji cząstki Lennarda-Jonesa o średnicy $\sigma_{X-Ow} = 3.08$ Å w modelu SPC/E odznaczał się brakiem wzajemnej kompensacji entalpowo-entropowej. Skutkowało to odmiennymi zależnościami wartości ΔH oraz $T\Delta S$ od temperatury w porównaniu do pozostałych rozmiarów rosnącej cząstki (Rys. 30, panel A).

Wraz ze wzrostem temperatury układu rosły wartości pojemności cieplnej ΔC_v , które dla każdego modelu wody i rodzaju użytego do obliczeń potencjału oddziaływań były zawsze dodatnie (Rys. 30-33, panel B). Największe wartości ΔC_v zostały osiągnięte dla pseudotwardej kuli Mie w modelu SPC/E.

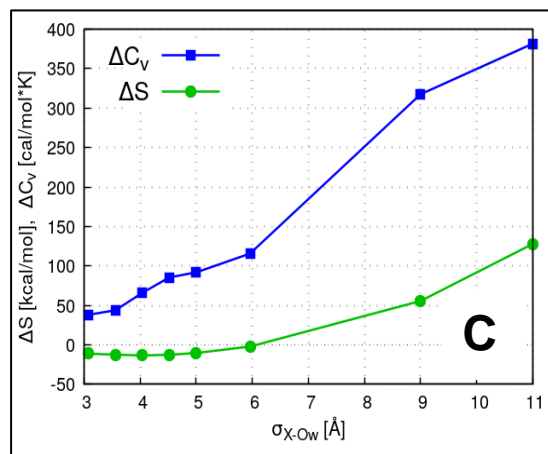
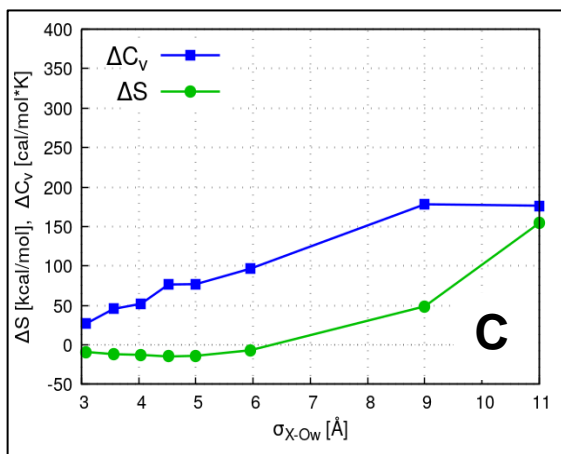
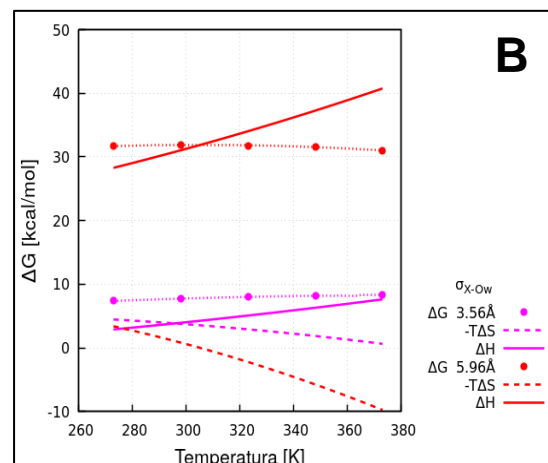
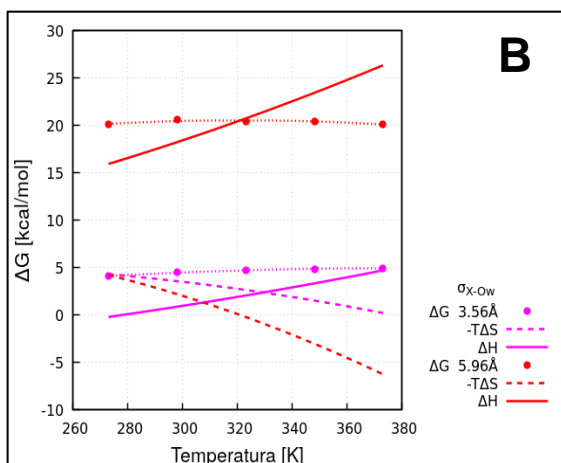
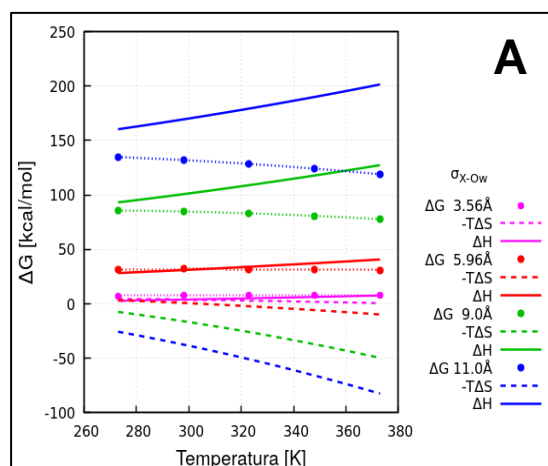
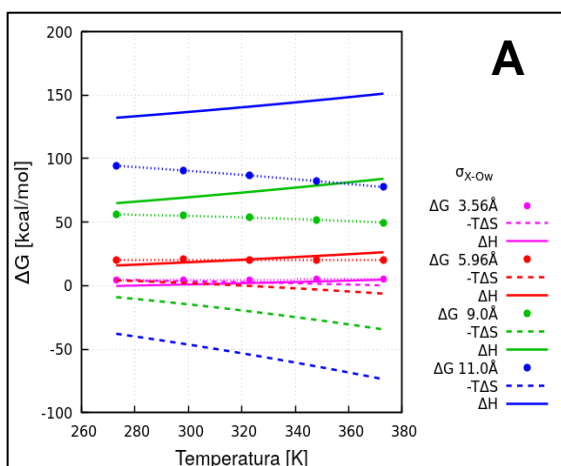
W obydwu modelach wody wartości energii swobodnych ΔG obliczone przy pomocy metody FEP oraz PI dla cząstki Lennarda-Jonesa oraz pseudotwardej kuli Mie były bardzo do siebie podobne i rosły wraz z temperaturą układu (rys. 30-33, panel A). Największe wartości ΔG zostały osiągnięte dla pseudotwardej kuli Mie o średnicy σ_{X-Ow} równej 4.04 Å w modelu TIP4P-2005 (Rys. 33, panel A).

Na Rys. 34-37 w panelach A-C zostały zamieszczone wykresy przedstawiające zależności zmian wartości ΔG , ΔH , $T\Delta S$ oraz ΔC_v od temperatury dla procesu hydratacji niepolarniej cząstki Lennarda-Jonesa oraz pseudotwardej kuli Mie o średnicy σ_{X-Ow} równej 3.56, 5.96, 9.00 i 11.00 Å w modelu SPC/E (Rys. 34 i 35) oraz TIP4P-2005 (Rys. 36 i 37).



Rys. 34. Wykres temperaturowych zależności zmian wartości ΔG , ΔH oraz $T\Delta S$ dla cząstki Lennarda-Jonesa o średnicy σ_{x-ow} równej 3.56, 5.96, 9.00 i 11.00 Å w modelu SPC/E (panel A). Panel B przedstawia powiększenie tych zależności dla $\sigma_{x-ow} = 3.56$ i 5.96 Å. Panel C przedstawia zależność zmian wartości ΔS oraz ΔC_v względem średnicy rosnącej cząstki Lennarda-Jonesa w zakresie temperatur od 273 do 373 K.

Rys. 35. Wykres temperaturowych zależności zmian wartości ΔG , ΔH oraz $T\Delta S$ dla pseudo twardej kuli Mie o średnicy σ_{x-ow} równej 3.56, 5.96, 9.00 i 11.00 Å w modelu SPC/E (panel A). Panel B przedstawia powiększenie tych zależności dla $\sigma_{x-ow} = 3.56$ i 5.96 Å. Panel C przedstawia zależność zmian wartości ΔS oraz ΔC_v względem średnicy rosnącej pseudotwardej kuli Mie w zakresie temperatur od 273 do 373 K.



Rys. 36. Wykres temperaturowych zależności zmian wartości ΔG , ΔH oraz $T\Delta S$ dla cząstki Lennarda-Jonesa o średnicy σ_{X-Ow} równej 3.56, 5.96, 9.00 i 11.00 Å w modelu TIP4P-2005 (panel A). W panelu B przedstawiono powiększenie tych zależności dla $\sigma_{X-Ow} = 3.56$ i 5.96 Å. Panel C przedstawia zależność zmian wartości ΔS oraz ΔC_v względem średnicy rosnącej cząstki Lennarda-Jonesa w zakresie temperatur od 273 do 373 K.

Rys. 37. Wykres temperaturowych zależności zmian wartości ΔG , ΔH oraz $T\Delta S$ dla pseudo twardej kuli Mie o średnicy σ_{X-Ow} równej 3.56, 5.96, 9.0 i 11 Å w modelu TIP4P-2005 (panel A). W panelu B przedstawiono powiększenie tych zależności dla $\sigma_{X-Ow} = 3.56$ i 5.96 Å. Panel C przedstawia zależność zmian wartości ΔS oraz ΔC_v względem średnicy rosnącej pseudotwardej kuli Mie w zakresie temperatur od 273 do 373 K.

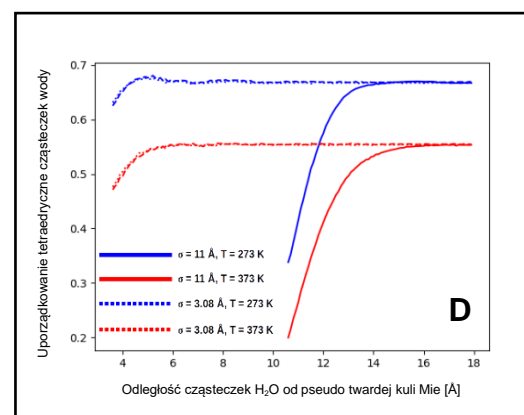
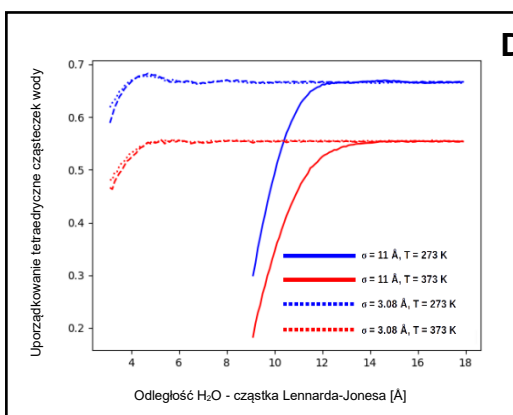
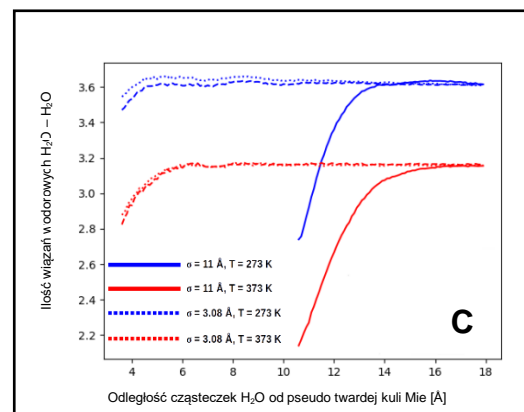
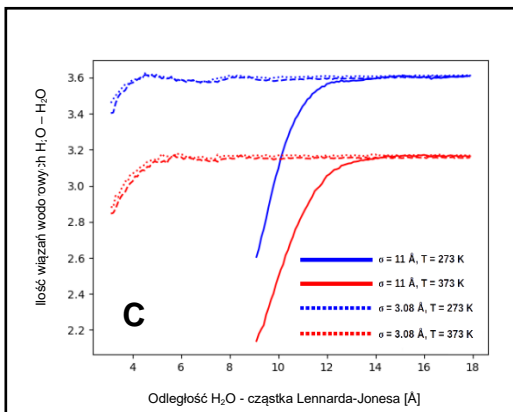
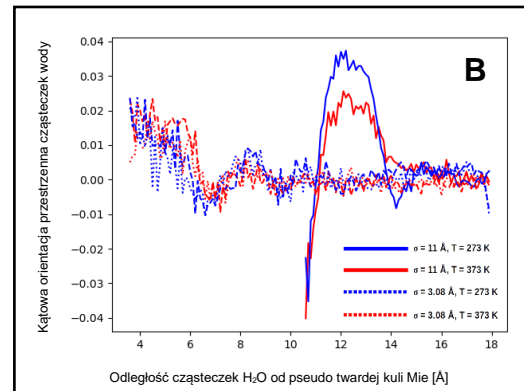
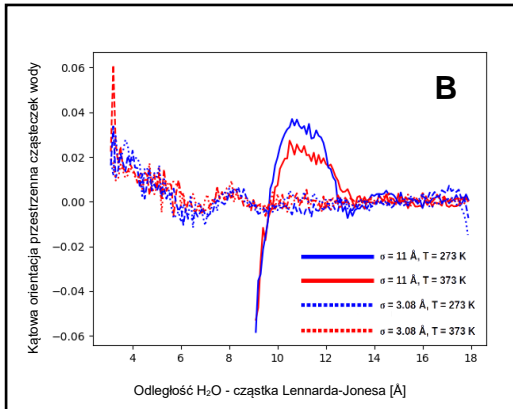
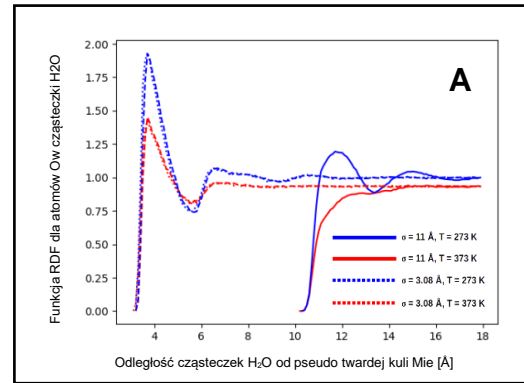
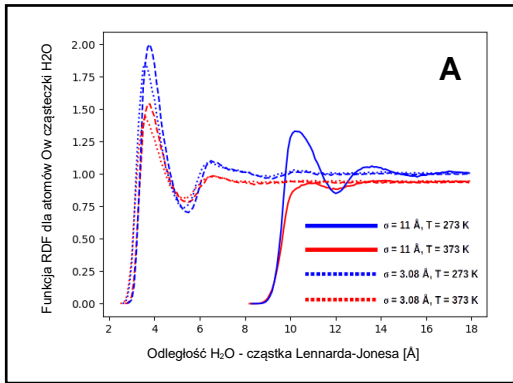
W procesie hydratacji niepolarnej cząstki Lennarda-Jonesa oraz pseudotwardej kuli Mie o średnicy σ_{X-Ow} 3.56, 5.96, 9.00 i 11.00 Å wzrost temperatury układu pociągał za sobą wzrost wartości ΔH oraz $T\Delta S$ (Rys. 34-37, panel A i B). W obydwu modelach wody wartości ΔH były zawsze dodatnie. Z kolei wartości $T\Delta S$ były zawsze dodatnie jedynie dla dużych cząstek o średnicach σ_{X-Ow} równych 9.00 i 11.00 Å. Większe wartości tych wielkości zostały osiągnięte dla pseudotwardej kuli Mie. W modelach SPC/E i TIP4P-2005 dla pseudotwardej kuli Mie o średnicy σ_{X-Ow} od 3.5 do 9.0 Å uzyskano podobne wartości ΔH oraz $T\Delta S$ (Rys. 35 i 37, panel A i B). Dla większych cząstek Mie o średnicy $\sigma_{X-Ow} = 11.00$ Å w modelu TIP4P-2005 oraz $T > 360$ K uzyskano wyższe wartości ΔH , a w $T < 300$ K niższe wartości $T\Delta S$ w porównaniu do obydwu tych wielkości uzyskanych w modelu SPC/E (Rys. 35 i 37, panel A i B).

Wartości ΔG w obydwu modelach wody zmniejszały się wraz ze wzrostem temperatury i rosły wraz ze wzrostem średnicy σ_{X-Ow} cząstki Lennarda-Jonesa oraz pseudotwardej kuli Mie (Rys. 34-37, panel A i B). Wyższe wartości ΔG zostały uzyskane dla pseudotwardej kuli Mie w modelu TIP4P-2005 – największe dla średnicy σ_{X-Ow} równej 11.00 Å (Rys. 37, panel A).

Uzyskane wyniki są zgodne z teorią Pratta i Chandlera³⁵ mówiącą o ścisłej zależności entropii układu od wielkości hydrofobowej wnęki tworzonej wewnątrz struktury wody. Zgodnie z tą teorią entropia hydratacji małych niepolarnych cząstek powinna być dodatnia i przyjmować jak największe wartości. Wyniki obliczeń uzyskane dla cząstki Lennarda-Jonesa oraz pseudotwardej kuli Mie o średnicy $\sigma_{X-Ow} < 5.96$ Å są zgodne z danymi eksperymentalnymi dla etanu¹⁵⁴. W przypadku większych rozmiarów obydwu typów cząstek uzyskano dużą zgodność z danymi eksperymentalnymi dla 2,2,4-trimetylopentanu oraz cykloheksanu¹⁵⁴.

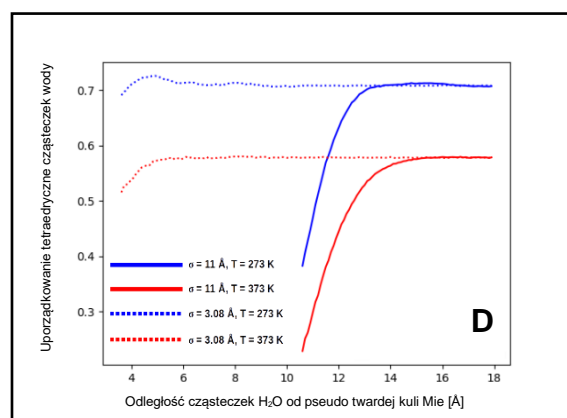
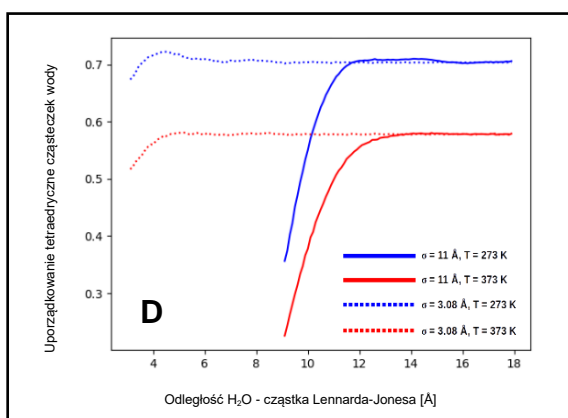
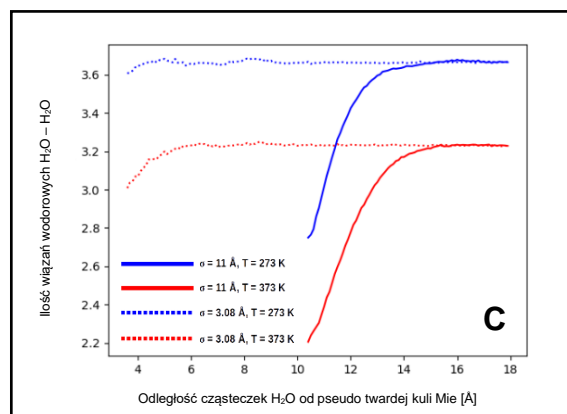
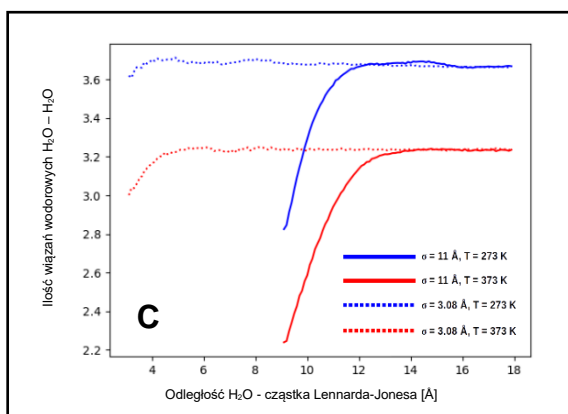
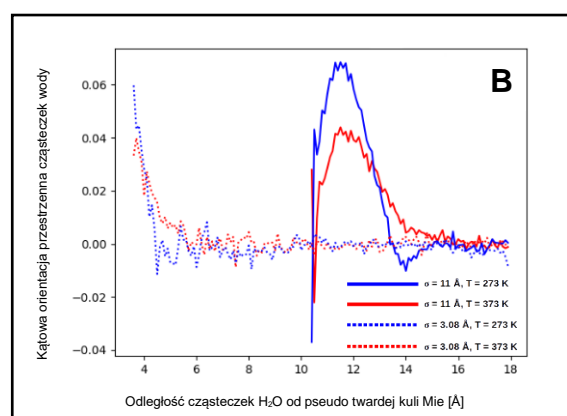
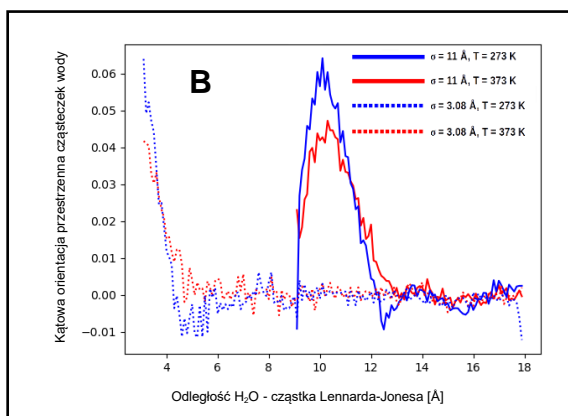
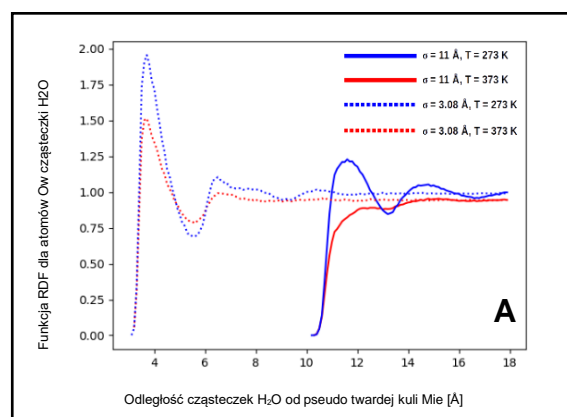
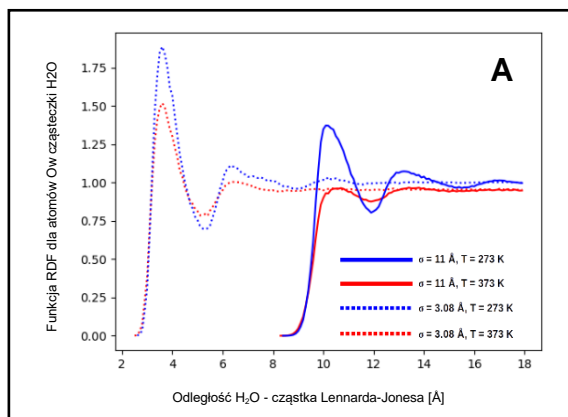
Wraz ze wzrostem temperatury oraz średnicy cząstki Lennarda-Jonesa i pseudotwardej kuli Mie dochodziło do wzrostu wartości pojemności cieplnej układu ΔC_v (Rys. 34-37, panel C). Większe wartości tej wielkości zostały osiągnięte dla pseudotwardej kuli Mie w modelu TIP4P-2005 – największe dla średnicy σ_{X-Ow} równej 11.00 Å (Rys. 37, panel C).

Na Rys. 38-41 zostały przedstawione wykresy opisujące zmiany: a) gęstości cząsteczek wody, b) ich kątowej orientacji przestrzennej, c) liczby wiązań wodorowych H₂O-H₂O oraz d) uporządkowania tetraedrycznego wody w pierwszej warstwie solwatacyjnej cząstki Lennarda-Jonesa oraz pseudotwardej kuli Mie o średnicy σ_{X-Ow} równej 3.08 oraz 11.00 Å. Każda z tych wielkości została obliczona dla modelu SPC/E (Rys. 38 i 39) oraz TIP4P-2005 (Rys. 40 i 41) w temperaturze 273 i 373 K.



Rys. 38. Wykresy gęstości cząsteczek wody (panel A), ich kątowej orientacji przestrzennej (panel B), liczby tworzonych przez nie wiązań wodorowych (panel C) oraz ich uporządkowania tetraedrycznego (panel D) w pierwszej warstwie solwatacyjnej cząstki Lennarda-Jonesa o średnicy σ_{Ow} równej 3.08 oraz 11.00 Å dla modelu SPC/E w $T = 273$ K (niebieski) oraz $T = 373$ K (czerwony).

Rys. 39. Wykresy gęstości cząsteczek wody (panel A), ich kątowej orientacji przestrzennej (panel B), liczby tworzonych przez nie wiązań wodorowych (panel C) oraz ich uporządkowania tetraedrycznego (panel D) w pierwszej warstwie solwatacyjnej pseudotwardej kuli Mie o średnicy σ_{Ow} równej 3.08 oraz 11 Å dla modelu SPC/E w $T = 273$ K (niebieski) oraz $T = 373$ K (czerwony).



Rys. 40. Wykresy gęstości cząsteczek wody (panel A), ich kątowej orientacji przestrzennej (panel B), liczby tworzonych przez nie wiązań wodorowych (panel C) oraz ich uporządkowania tetraedycznego (panel D) w pierwszej warstwie solwacyjnej cząstki Lennarda-Jonesa o średnicy σ_{X-Ow} równej 3.08 oraz 11 Å dla modelu TIP4P-2005 w $T = 273$ K (niebieski) oraz $T = 373$ K (czerwony).

Rys. 41. Wykresy gęstości cząsteczek wody (panel A), ich kątowej orientacji przestrzennej (panel B), liczby tworzonych przez nie wiązań wodorowych (panel C) oraz ich uporządkowania tetraedycznego (panel D) w pierwszej warstwie solwacyjnej pseudotwardej kuli Mie o średnicy σ_{X-Ow} równej 3.08 oraz 11 Å dla modelu TIP4P-2005 w $T = 273$ K (niebieski) oraz $T = 373$ K (czerwony).

W obydwu modelach wody większe gęstości upakowania cząsteczek wody zostały osiągnięte w temperaturze 273 K dla mniejszej cząstki Lennarda-Jonesa oraz pseudotwardej kuli Mie. W tej temperaturze wyższe gęstości wody zostały osiągnięte w modelu TIP4P-2005 w pierwszej warstwie solwatacyjnej pseudotwardej kuli Mie o średnicy σ_{X-Ow} równej 3.08 Å. (Rys. 39 i 40, panel A). W wyższej z temperatur największą gęstość wody osiągnięto w modelu TIP4P-2005 w pierwszej warstwie solwatacyjnej cząstki Lennarda-Jonesa.

Analiza rozkładu cosinusa kąta zawartego między prostą prostopadłą do środka cząstki Lennarda-Jonesa oraz pseudotwardej kuli Mie, a wektorem momentu dipolowego wody pokazuje spore różnice między rozmiarem każdej z tych cząstek, temperaturą układu oraz modelem wody użytym do obliczeń (Rys. 38-42, panel B). W temperaturze 273 K większe wartości tego kąta zostały osiągnięte dla niepolarnych cząstek o średnicy σ_{X-Ow} równej 11.00 Å. Największy stopień uporządkowania przestrzennego cząsteczek wody został osiągnięty w pierwszej warstwie solwatacyjnej pseudotwardej kuli Mie w modelu TIP4P-2005 i temperaturze 273 K (Rys. 41, panel B).

Większa liczba wiązań wodorowych została osiągnięta w modelu TIP4P-2005 i temperaturze 273 K w pierwszej warstwie solwatacyjnej cząstki Lennarda-Jonesa oraz pseudotwardej kuli Mie o średnicy σ_{X-Ow} równej 3.08 Å (Rys. 38 i 40, panel C). Miało to związek z większą powierzchnią przestrzeni dostępnej dla wody po wytworzeniu w jej strukturze niewielkiej hydrofobowej wnęki. W temperaturze 273 K liczba wiązań wodorowych w otoczeniu każdej z tych cząstek była podobna.

Największe uporządkowanie tetraedryczne wody zostało osiągnięte w modelu TIP4P-2005 i temperaturze 273 K dla odległości ok. 5 Å cząsteczek wody względem pierwszej warstwy solwatacyjnej cząstki Lennarda-Jonesa oraz pseudotwardej kuli Mie o średnicy σ_{X-Ow} równej 3.08 Å (Rys. 41, panel D).

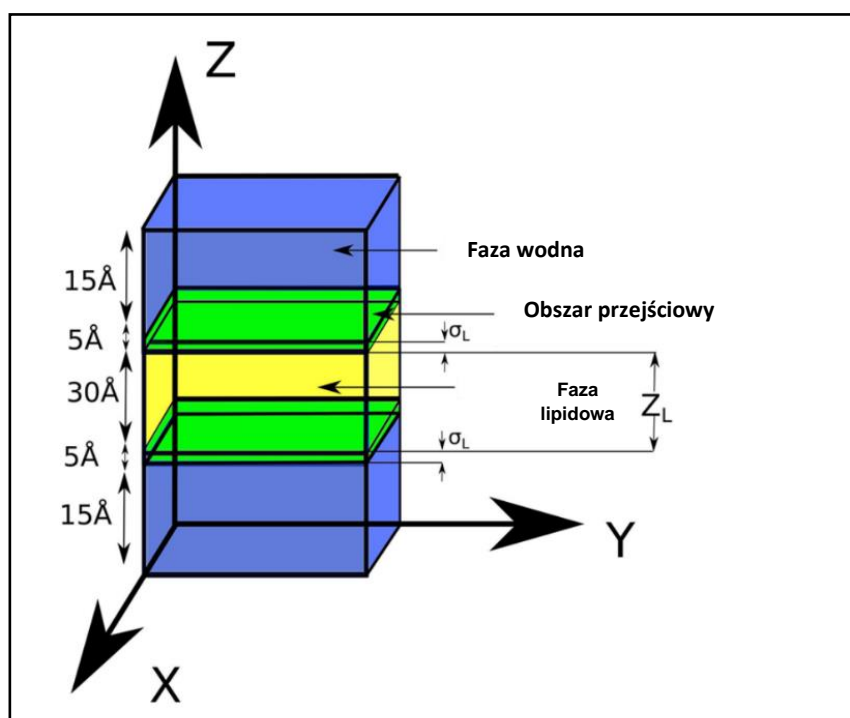
6. 3. Rozszerzanie gruboziarnistego pola sił UNRES do białek membranowych w dwuwarstwie lipidowej

Celem tego projektu było rozszerzenie gruboziarnistego pola sił UNRES do symulacji białek membranowych w środowisku lipidowym.

Moja rola w tym projekcie sprowadzała się do analizy i interpretacji wyników symulacji MREMD wykonanych dla α -helikalnych białek membranowych w fazie wodnej, lipidowej oraz dwuwarstwie lipidowej w testowanym polu sił UNRES. Do moich zadań należała również ocena zdolności predykcyjnych testowanej wersji pola sił UNRES w obydwu typach środowisk oraz analiza zmian konformacyjnych białek podczas procesu ich zwijania w badanym obszarze symulacyjnym.

Opis procedury obliczeniowej:

Do badań wykorzystano pionowe pudło symulacyjne składające się z fazy wodnej oraz dwuwarstwy lipidowej – o ściśle określonych wymiarach (Rys. 42). Rozmiar fazy lipidowej rozciągał się od $(Z_{\text{box}} - Z_L) / 2$ do $(Z_{\text{box}} + Z_L) / 2$, gdzie Z_{box} był rozmiarem pudła symulacyjnego wzdłuż osi Z, a parametr Z_L określał grubość fazy lipidowej. Grubość obszaru przejściowego σ_L została dobrana w taki sposób, by zapewnić swobodne przejście białka membranowego z fazy wodnej do lipidowej i odwrotnie.



Rys. 42. Schematyczne przedstawienie pudła symulacyjnego użytego do obliczeń. Kolorem niebieskim oznaczono fazę wodną, żółtym fazę lipidową, a zielonym obszar przejściowy.

Obliczenia te były wykonywane z użyciem gruboziarnistego pola sił UNRES w wersji OPT-WTFSA-1⁶⁹, w której domyślne wyrażenie na efektywną funkcję energii zostało rozszerzone do opisu centrów oddziaływań łańcucha polipeptydowego białek w fazie lipidowej:

$$\begin{aligned}
U = & w_{SC} \sum_{i < j} U_{SC_i SC_j} + w_{SCP} \sum_{i \neq j} U_{SC_i P_j} + w_{pp}^{VDW} \sum_{i < j-1} U_{P_i P_j}^{VDW} \\
& + w_{pp}^{el} f_2(T) \sum_{i < j-1} U_{P_i P_j}^{el} + w_{bond} \sum_i U_{bond}(d_i) + w_b \sum_i U_b(\theta_i) \\
& + w_{corr}^{(3)} f_3(T) \sum_{i < j} U_{corr;ij}^{(3)} + w_{corr}^{(4)} f_4(T) \sum_{i < j} U_{corr;ij}^{(4)} \\
& + w_{turn}^{(3)} f_3(T) \sum_i U_{turn;i}^{(3)} + w_{turn}^{(4)} f_4(T) \sum_i U_{turn;i}^{(4)} \\
& + w_{rot} \sum_i U_{rot}(\alpha_{SC_i}, \beta_{SC_i}) + w_{ssbond} \sum_i U_{ssbond;i} \\
& + w_{tor} f_2(T) \sum_i U_{tor}(\gamma_i) + w_{tord} f_3(T) \sum_i U_{tord}(\gamma_i, \gamma_{i+1}) \\
& + w_{SC-corr} f_2(T) \sum_{m=1}^3 \sum_i U_{SC-corr}(\tau_i^{(m)}) \\
& + w_{liptrans} \left(\sum_i U_{P_i}^{liptrans} + \sum_i U_{SC_i}^{liptrans} \right)
\end{aligned} \tag{6.3a}$$

Rozszerzenie gruboziarnistego pola sił UNRES do środowiska lipidowego wiązało się z koniecznością zmodyfikowania kilku istniejących w powyższym równaniu członów energetycznych w taki sposób, by każdy z nich umożliwiał jak najdokładniejszy opis oddziaływań zjednoczonych łańcuchów bocznych (**SC**) oraz zjednoczonych grup peptydowych (**P**) białek w fazie wodnej oraz lipidowej.

Jednym z członów, który wymagał stosowanej modyfikacji był potencjał $U_{SC_i SC_j}$ opisujący średnią energię swobodną oddziaływań hydrofobowych (hydrofilowych) między zjednoczonymi łańcuchami bocznymi białek. Energia tych oddziaływań w polu sił UNRES jest opisywana za pomocą potencjału Gaya-Berne'a¹⁵⁵ w następujący sposób:

$$U_{GB} = 4 \left[\epsilon_{ij} x_{ij}^{12} - \epsilon_{ij} x_{ij}^6 \right] \tag{6.3b}$$

$$\epsilon_{ij} = \epsilon^\circ \epsilon [\hat{\mathbf{u}}_i \circ \hat{\mathbf{r}}_{ij}, \hat{\mathbf{u}}_j \circ \hat{\mathbf{r}}_{ij}, \hat{\mathbf{u}}_i \circ \hat{\mathbf{u}}_j] \tag{6.3c}$$

$$x_{ij} = \frac{\sigma_{ij}^\circ}{r_{ij} - \sigma_{ij}, [\hat{\mathbf{u}}_i \circ \hat{\mathbf{r}}_{ij}, \hat{\mathbf{u}}_j \circ \hat{\mathbf{r}}_{ij}, \hat{\mathbf{u}}_i \circ \hat{\mathbf{u}}_j] + \sigma_{ij}^\circ} \tag{6.3d}$$

gdzie parametry ε° oraz σ_{ij}° określają odpowiednio głębokość studni potencjału Gaya-Berne'a i odległość dwóch równoległe zorientowanych względem siebie osi głównych zjednoczonych łańcuchów bocznych białek, których energia wzajemnych oddziaływań przyjmuje wartość zero, $\hat{\mathbf{u}}_i$ oraz $\hat{\mathbf{u}}_j$ to jednostkowe wektory osi SC \dots C $^\alpha$, r_{ij} to jednostkowy wektor reprezentujący odległość między osiami głównymi SC $_i$ oraz SC $_j$, natomiast x_{ij} jest zmniejszoną odwrotnością odległości oddziałujących ze sobą zjednoczonych grup bocznych białek.

Modyfikacja potencjału $U_{SC_iSC_j}$ wiązała się z koniecznością określenia parametrów ε_L° w fazie lipidowej (parametry ε_w° dla fazy wodnej były takie same jak w pierwotnej wersji OPT-WTFSA-1 pola sił UNRES). W tej wersji pola sił wymiary łańcucha bocznego białek oraz ich anizotropie były takie same dla fazy wodnej i lipidowej. Do wyznaczenia parametrów ε_L° zastosowano procedurę obliczeniową opracowaną przez Levitt'a¹⁵⁶. W pierwszej kolejności zsumowano parametry ε opisujące głębokość studni potencjału Lennarda-Jonesa dla wszystkich atomów każdego łańcucha bocznego białek w polu sił AMBER ff14SB¹⁵⁷. Następnie przy pomocy reguły Lorentza-Barthelota⁹⁴ obliczono wartości $\varepsilon_{L;XY}^\circ$ dla dwóch oddziałujących ze sobą typów łańcuchów bocznych (tu: X oraz Y) w fazie lipidowej:

$$\varepsilon_{L;X,Y}^\circ = \sqrt{\varepsilon_{L;X}^\circ \varepsilon_{L;Y}^\circ} \quad (6.3e)$$

Parametry ε_L° oraz ε_w° dla wszystkich możliwych kombinacji oddziałujących ze sobą reszt aminokwasowych w zjednoczonych łańcuchach bocznych białek zostały przedstawione w Tabeli S1 w suplemencie do załącznika 2 stanowiącego publikację „*Extension of the UNRES Coarse-Grained Force Field to Membrane Proteins in the Lipid Bilayer*” K. Zięba, M. Ślusarz, R. Ślusarz, A. Liwo, C. Czaplowski, A. K. Sieradzan, J. Phys. Chem. B, 2019, **123**, 7829-7839. Wartości ε_L° oraz ε_w° zostały następnie wykorzystane do obliczenia ogólnego parametru ε° w fazie wodnej oraz lipidowej:

$$\varepsilon^\circ = \left(1 - \frac{s_i + s_j}{2}\right) \varepsilon_w^\circ + \frac{s_i + s_j}{2} \varepsilon_L^\circ \quad (6.3f)$$

gdzie zmienna s_i opisuje położenie i -tego zjednoczonego łańcucha bocznego białek w fazie wodnej ($s_i = 0$), lipidowej ($s_i = 1$) oraz w obszarze przejściowym pomiędzy lipidem i wodą ($0 < s_i < 1$) i jest ona dana jako:

$$s_i = \begin{cases} 0 & z_i > (Z_{\text{box}} + Z_L)/2 + \sigma_L \\ 3\lambda(z_i)^2 - 2\lambda(z_i)^3 & (Z_{\text{box}} + Z_L)/2 < z_i < (Z_{\text{box}} + Z_L)/2 + \sigma_L \\ 1 & (Z_{\text{box}} - Z_L)/2 < z_i < (Z_{\text{box}} + Z_L)/2 \end{cases} \quad (6.3g)$$

$$\lambda(z) = \begin{cases} \frac{(Z_{\text{box}} + Z_L)/2 - z + \sigma_L}{\sigma_L} & (Z_{\text{box}} + Z_L)/2 < z < (Z_{\text{box}} + Z_L)/2 + \sigma_L \\ \frac{(Z_{\text{box}} - Z_L)/2 + z - \sigma_L}{\sigma_L} & (Z_{\text{box}} - Z_L)/2 - \sigma_L < z < Z_{\text{box}} - Z_L \end{cases} \quad (6.3h)$$

Zmienna z_i opisuje współrzędną kartezjańską i -tego zjednoczonego łańcucha bocznego białek. W analogiczny sposób definiowana była również zmienna s_j odnosząca się do j -tego zjednoczonego łańcucha bocznego białek.

Dalszej modyfikacji uległy również potencjały korelacyjne $U_{\text{corr}}^{(3)}$ i $U_{\text{corr}}^{(4)}$ oraz skrętne $U_{\text{turn}}^{(3)}$ i $U_{\text{turn}}^{(4)}$, które są niezbędne do prawidłowego określenia II-rzędowych struktur białek, a także potencjał $U_{\text{pp}}^{\text{el}}$ opisujący oddziaływania elektrostatyczne między zjednoczonymi grupami peptydowymi. Modyfikacja tych potencjałów miała na celu odzwierciedlenie większej siły oddziaływań elektrostatycznych między zjednoczonymi grupami peptydowymi w fazie lipidowej. Aby tego dokonać potencjały $U_{\text{corr}}^{(3)}$, $U_{\text{turn}}^{(3)}$ i $U_{\text{turn}}^{(4)}$ zostały przeskalowane przez funkcję f_L opisaną równaniem 6.3i, natomiast potencjały $U_{\text{pp}}^{\text{el}}$ oraz $U_{\text{corr}}^{(4)}$ które są proporcjonalne do drugiej potęgi wirtualnych momentów dipolowych zjednoczonych grup peptydowych zostały przeskalowane przez funkcję f_L^2 :

$$f_L = \frac{s_i + s_j}{2} s_L + 1 \quad (6.3i)$$

gdzie zmienne s_i oraz s_j opisują wzory 6.3g i 6.3h, a s_L jest współczynnikiem skalującym fazę lipidową, który przyjmuje wartości z zakresu od 0.4 do 2.5 (z krokiem 0.3).

Do efektywnej funkcji energii UNRES wprowadzono także dwa nowe potencjały oddziaływań takie jak $U_{\text{pi}}^{\text{liptrans}}$ oraz $U_{\text{SCi}}^{\text{liptrans}}$, które opisują transfer zjednoczonych grup peptydowych i zjednoczonych łańcuchów bocznych z fazy wodnej do lipidowej. Potencjały te uwzględniają również protonowanie i deprotonowanie naładowanych reszt aminokwasowych obydwu centrów oddziaływań podczas przejścia między fazą wodną i lipidową. Zostały one sparametryzowane na podstawie współczynnika podziału dla amidów kwasu N-aminoacetylowego w układzie woda / n -oktanol¹⁵⁸ i są przedstawiane jako:

$$U_i^{\text{liptrans}} = \Delta G_i \times s_i \quad (6.3j)$$

gdzie ΔG_i to energia swobodna transferu i -tego centrum oddziaływań z fazy wodnej do lipidowej. Dla zjednoczonych łańcuchów bocznych przyjęto, że wartość ΔG_{SCi} była równa eksperymentalnie wyznaczonej wartości energii swobodnej transferu każdego aminokwasu z fazy wodnej do n -oktanolowej¹⁵⁸. Rozszerzone pole sił UNRES zostało przetestowane dla dwóch wartości energii swobodnych ΔG_{pi} transferu zjednoczonych grup peptydowych – pierwszej równej 0 kcal/mol oraz drugiej (eksperymentalnej) równej

6.11 kcal/mol¹⁵⁸. Pierwsza z tych wartości wynikała ze sposobu kalibracji pola sił UNRES w wersji OPT-WTFSA-1 podczas której przyjęto założenie, że grupy peptydowe są umieszczone w niepolarnym wnętrzu peptydu.

Pozostałe potencjały takie jak U_b , U_{bond} , U_{tor} , U_{tord} oraz U_{rot} były uwzględniane w ten sam sposób w fazie wodnej oraz lipidowej. Człony wagowe tych oraz wyżej przedstawionych potencjałów $U_{corr}^{(3)}$, $U_{corr}^{(4)}$, $U_{turn}^{(3)}$, $U_{turn}^{(4)}$, U_{pp}^{el} i U_{SCiSCj} nie uległy zmianie względem podstawowej wersji OPT-WTFSA-1 pola sił UNRES.

Tak rozszerzone pole sił UNRES było następnie testowane pod kątem zdolności do zachowania energii całkowitej układu. W tym celu wykonano trzy serie mikrokanonicznych symulacji MD^{81,82} dla struktury eksperymentalnej białka o kodzie PDB: 1GAB, która na potrzeby testów została skonwertowana do reprezentacji gruboziarnistej pola sił UNRES. Obszar symulacyjny miał wymiary 100 x 100 x 70 Å – grubość fazy wodnej była równa 15 Å, fazy lipidowej 30 Å, a obszaru przejściowego 5 Å (po każdej ze stron fazy lipidowej). Do całkowania równań ruchu zastosowano algorytm VTS⁸⁹. Każda symulacja składała się z 1 000 000 kroków z krokiem czasowym wynoszącym 0.498 femtosekundy. W fazie lipidowej oddziaływania elektrostatyczne były skalowane przez współczynnik s_L równy 0.4, 1.0 oraz 2.0.

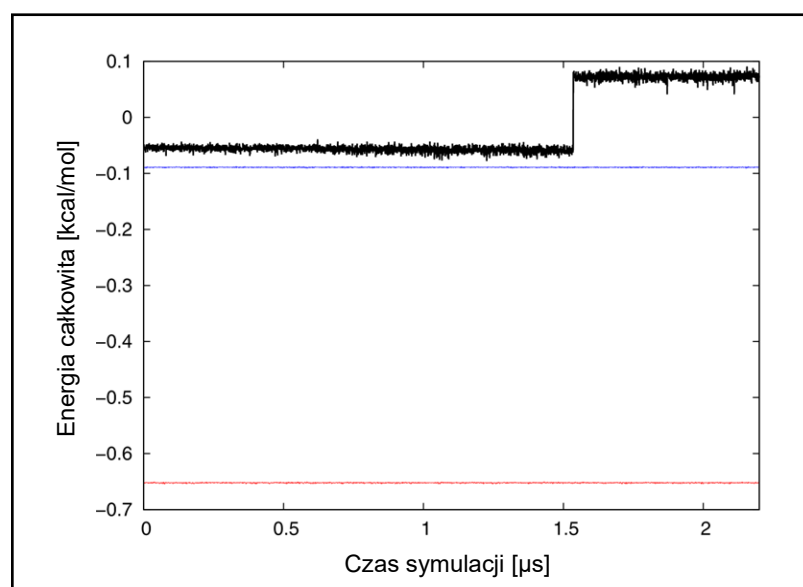
W dalszej kolejności rozszerzone pole sił UNRES było testowane pod kątem zdolności do zachowania temperatury panującej w symulowanym układzie. W tym celu wykonano cztery serie kanonicznych symulacji MD^{81,82} dla struktury eksperymentalnej białka o kodzie PDB: 1GAB, która podobnie jak poprzednio została skonwertowana do reprezentacji gruboziarnistej pola sił UNRES. Do testów użyto termostatu Langevina⁹², Berendsena⁹¹ oraz Nose-Hoovera¹⁵⁰. Wymiary obszaru symulacyjnego, w tym także grubość fazy wodnej, lipidowej oraz obszaru przejściowego były takie same jak w symulacjach mikrokanonicznych. Każda z symulacji składała się z 20 000 000 kroków z krokiem czasowym równym 4.89 femtosekund. Oddziaływania elektrostatyczne w fazie lipidowej były skalowane przez współczynnik s_L równy 0.4. Stała sprzężenia τ symulowanego układu z termostatem Berendsena wynosiła 4.89 femtosekund, natomiast masa pudła symulacyjnego związana z dodatkowym stopniem swobody układu w technice termostatowania Nose-Hoovera była równa 5.0.

Ostatni etap badań polegał na przetestowaniu zdolności rozszerzonego pola sił UNRES do przewidywania struktur białek w fazie wodnej oraz lipidowej. Do testów wykorzystano dziesięć α -helikalnych białek membranowych o kodach PDB: 1IIJ (35 reszt), 1N7L (53 reszty), 1WAZ (46 reszt), 1WU0 (72 reszty), 2KSD (75 reszt), 2LOP (75 reszt), 1VRY (76 reszt), 1A91 (79 reszt), 2K9P (80 reszt) oraz 2MOZ (81 reszt), których struktury eksperymentalnie

zostały zbadane przy pomocy metody NMR¹²⁰ w błonie lipidowej, a ich podobieństwo sekwencyjne było bardzo małe. Obszar symulacyjny miał kształt prostopadłościanu o wymiarach 470 x 70 Å. Grubość fazy lipidowej była równa 30 Å z obszarami przejściowymi o wielkości 5 Å (po każdej ze stron fazy lipidowej). Dla każdego z tych białek wykonano symulacje MREMD¹⁰¹ dla dwóch wartości energii swobodnej $\Delta G_{pi} = 0$ kcal/mol oraz $\Delta G_{pi} = 6.11$ kcal/mol. Symulacje te składały się łącznie z 48 replik i były prowadzone w 24 różnych temperaturach z zakresu od 250 do 480 K (co 10 K każda), w których na jedną temperaturę przypadały dwie repliki. Każda symulacja składała się z 70 000 000 kroków z długością kroku czasowego wynoszącą 4.89 femtosekund. Faza lipidowa była skalowana przez współczynnik s_L równy 1.3, a dwuwarstwa lipidowa przez wartości s_L z zakresu od 0.4 do 2.5 (z krokiem 0.3). Symulacje MREMD polegały na próbkowaniu przestrzeni konformacyjnej badanych białek w celu znalezienia struktury o możliwie jak najniższej wartości energii swobodnej. Wyniki symulacji zostały przetworzone przy pomocy metody WHAM^{78,79} w celu określenia zmian wartości pojemności cieplnej ΔC_v od temperatury, wartości C^αRMSD dla symulowanych i eksperymentalnych struktur białek oraz prawdopodobieństwa występowania symulowanych konformacji białek w niskoenergetycznych obszarach ich przestrzeni fazowej. Z dużym prawdopodobieństwem stwierdzono, że konformacje uzyskane w temperaturach 260, 280 i 300 K odznaczały się najniższymi wartościami energii swobodnych. Na podstawie miary C^αRMSD konformacje te zostały pogrupowane do pięciu osobnych skupień przy pomocy metody Ward'a¹⁰⁸. Spośród wszystkich skupień wybrano jedno o najniższej średniej wartości C^αRMSd względem struktury eksperymentalnej badanego białka. W dalszej kolejności z tego skupienia wybrano najbardziej reprezentatywną konformację symulowanego białka o najbliższej wartości C^αRMSd względem jego struktury eksperymentalnej. Dla wybranej konformacji białka obliczone zostały wartości miar GDT-TS^{18,19} oraz TM-score^{20,21} względem struktury eksperymentalnej. W dalszej kolejności dla obydwu wartości ΔG_{pi} wyznaczono najbardziej optymalną wartość współczynnika s_L skalującego dwuwarstwę lipidową. W tym celu dokonano uśrednienia wszystkich wartości GDT-TS oraz TM-score dla zbioru reprezentatywnych struktur badanych białek względem wartości współczynnika skalowania s_L . Konformacje reprezentatywne białek, które zostały uzyskane w fazie wodnej, lipidowej oraz w dwuwarstwie lipidowej z najbardziej optymalną wartością współczynnika s_L zostały skonwertowane do struktur pełnoatomowych przy pomocy algorytmu PULCHRA¹⁴².

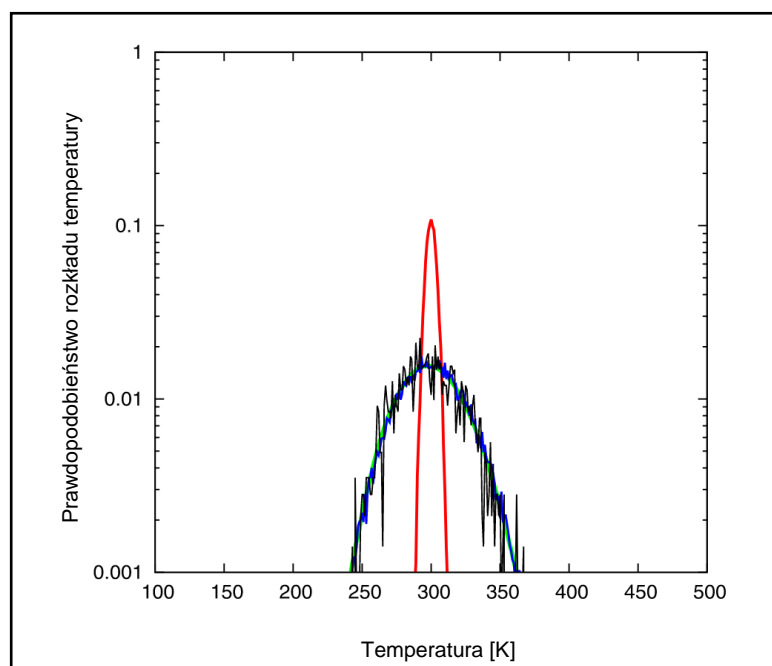
Wyniki obliczeń:

Rozszerzone pole sił UNRES odznaczało się dużą zdolnością do zachowania energii całkowitej układu podczas trwania symulacji (Rys. 43). Dla współczynników skalowania s_L równych 0.4 oraz 1.0 zaobserwowano niewielkie fluktuacje energii całkowitej wokół jej wartości średniej. Większe fluktuacje ze skokiem energii całkowitej o ok. 0.2 kcal/mol zostały zaobserwowane dla współczynnika s_L równego 2.0. Tak duży skok energii całkowitej był prawdopodobnie wynikiem większych zmian konformacyjnych białka o kodzie PDB: 1GAB przy dużych wartościach s_L skalujących fazę lipidową. Uzyskane wyniki świadczą o tym, że symulacje z rozszerzonym do środowiska lipidowego polem sił UNRES wykazywały zachowanie symplektyczne, a energia całkowita badanego układu (poza sporadycznym skokiem dla $s_L = 2.0$) nie wykazywała dużych odstępstw od swojej wartości średniej.



Rys. 43. Wykres zmian wartości energii całkowitej układu dla współczynnika skalowania fazy lipidowej $s_L = 0.4$ (czerwony), $s_L = 1.0$ (niebieski) oraz $s_L = 2.0$ (czarny) względem czasu trwania symulacji w rozszerzonych do środowiska lipidowego polu sił UNRES.

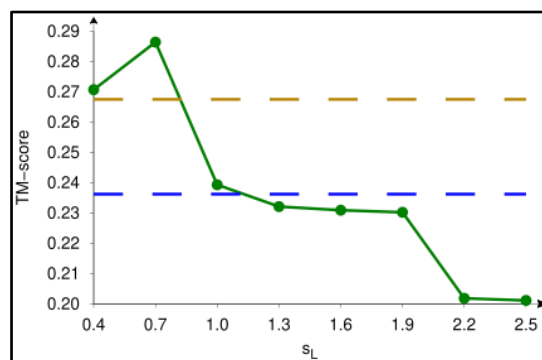
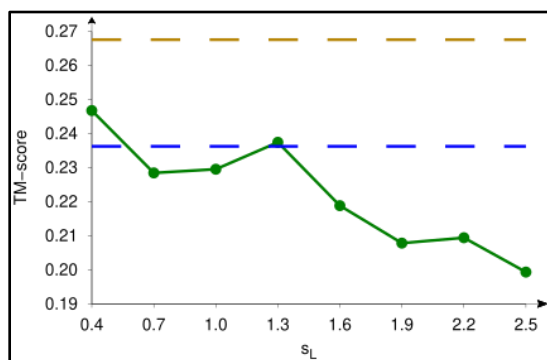
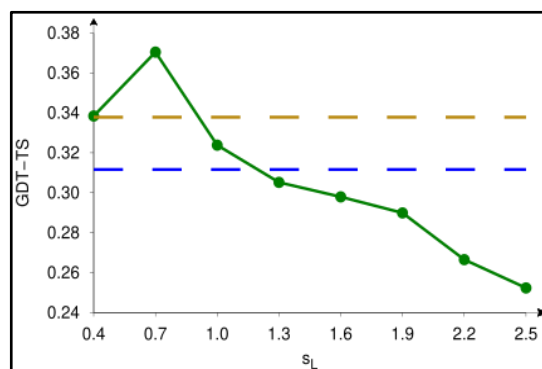
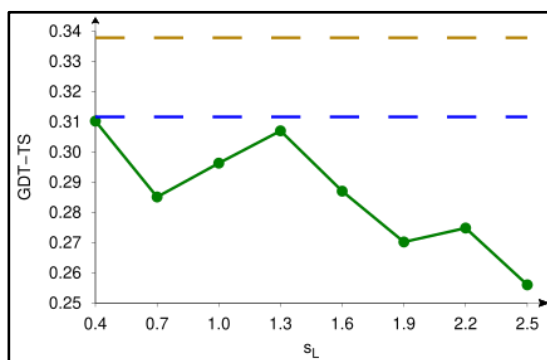
Najbardziej korzystny rozkład temperatur został uzyskany dla symulacji z zastosowaniem termostatu Langevina, ponieważ on jako jedyny najlepiej pokrywał się z teoretycznym rozkładem Maxwella-Boltzmana. Dla termostatu Nose-Hoovera uzyskano mniej korzystny rozkład temperatur, który był wynikiem większych fluktuacji temperatur względem zadanego rozkładu teoretycznego. Najbardziej niekorzystny rozkład temperatur został osiągnięty dla termostatu Berendsena – ten rozkład był zbyt wąski i bardzo odbiegał od temperaturowego rozkładu Maxwella-Boltzmana (Rys. 44).



Rys. 44. Wykres przedstawiający rozkłady temperatur uzyskane w symulacjach MD z testowanym polem sił UNRES oraz termostatem Langevina (niebieska linia), Berendsena (czerwona linia) i Nose-Hoovera (czarna linia). Kolorem zielonym oznaczono teoretyczny rozkład temperatur Maxwella-Boltzmana.

Białka membranowe, które były symulowane z energią swobodną transferu zjednoczonych grup peptydowych z fazy wodnej do lipidowej $\Delta G_{pi} = 0$ kcal/mol (Rys. 46) posiadały większe wartości miar GDT-TS i TM-score względem swoich struktur eksperymentalnych w porównaniu do symulacji z eksperymentalną wartością ΔG_{pi} równą 6.11 kcal/mol (Rys. 45).

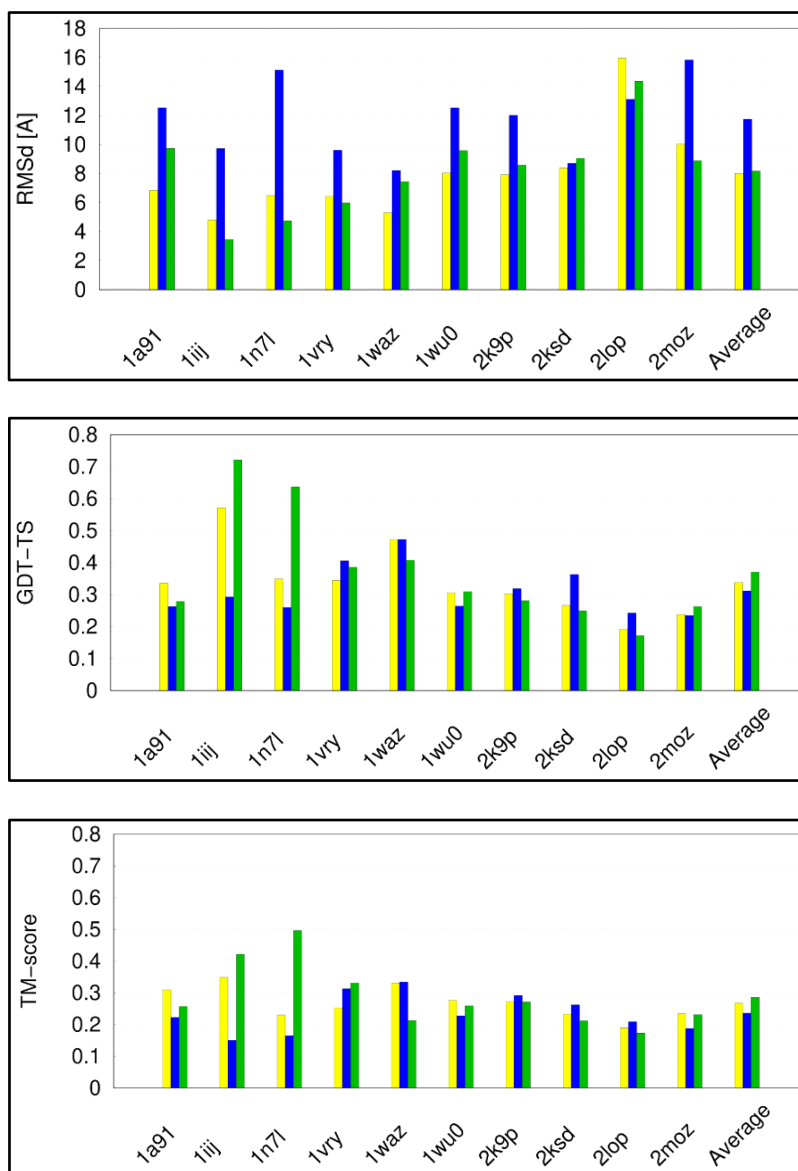
Największe wartości GDT-TS oraz TM-score zostały osiągnięte dla $\Delta G_{pi} = 0$ kcal/mol w dwuwarstwie lipidowej skalowanej przez współczynnik s_L równy 0.7 (Rys. 46). Najniższe wartości każdej z tych miar zostały osiągnięte w fazie wodnej (Rys. 46). Uzyskane wyniki świadczą o tym, że testowane pole sił UNRES wykazywało największe zdolności predykcyjne w dwuwarstwie lipidowej o współczynniku skalowania $s_L = 0.7$ i względnie dobre, lecz nieco gorsze w środowisku całkowicie lipidowym. W fazie wodnej symulowane struktury białek membranowych były bardzo dalekie od struktur eksperymentalnych wyznaczonych w błonie lipidowej.



Rys. 45. Zależność zmian średnich wartości GDT-TS oraz TM-score względem współczynnika skalowania s_L dla reprezentatywnych konformacji białek membranowych uzyskanych z symulacji MREMD z energią swobodną ΔG_{pi} równą 6.11 kcal/mol. Kolorem niebieskim oznaczono fazę wodną, żółtym lipidową, a zielonym dwuwarstwę lipidową.

Rys. 46. Zależność zmian średnich wartości GDT-TS oraz TM-score względem współczynnika skalowania s_L dla reprezentatywnych konformacji białek membranowych uzyskanych z symulacji MREMD z energią swobodną ΔG_{pi} równą 0 kcal/mol. Kolorem niebieskim oznaczono fazę wodną, żółtym lipidową, a zielonym dwuwarstwę lipidową.

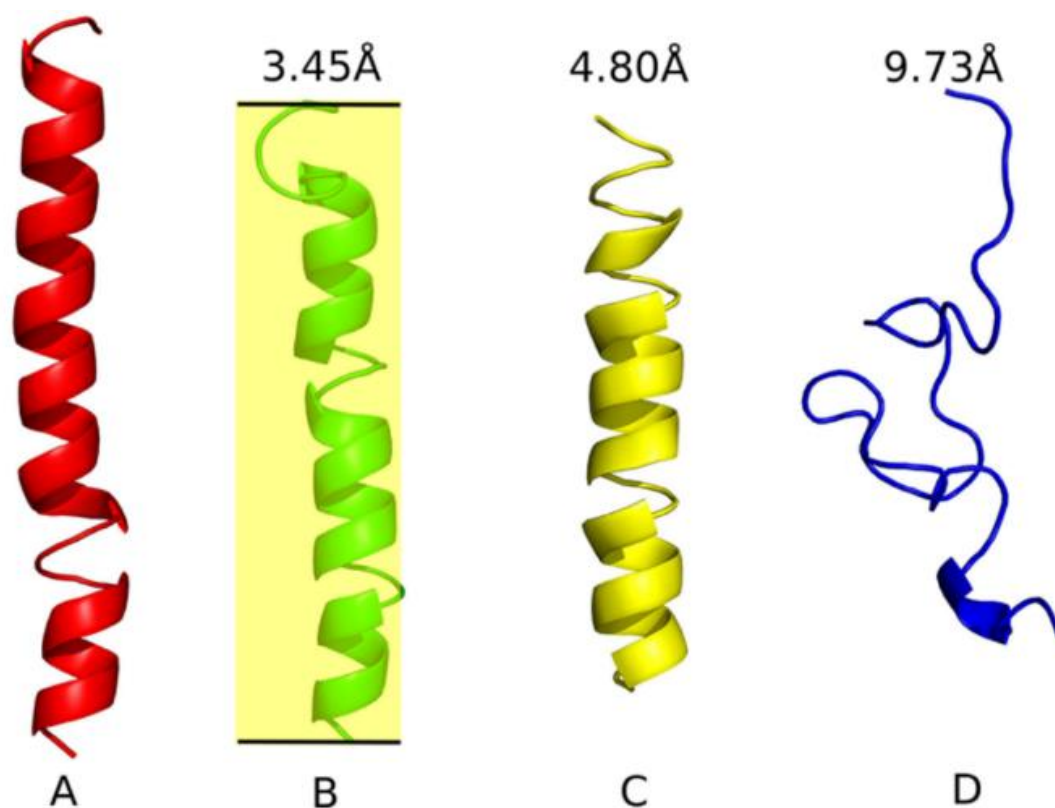
W dwuwarstwie lipidowej skalowanej przez współczynnik $s_L = 0.7$ struktury najbliższe eksperymentalnym zostały osiągnięte dla dwóch białek o kodach PDB: 1IIJ oraz 1N7L (Rys. 47). Dla symulowanej i eksperymentalnej struktury białka 1IIJ osiągnięto następujące wartości miar porównawczych: RMSD ≈ 4 Å, GDT-TS ≈ 0.75 oraz TM-score ≈ 0.45 . W przypadku białka 1N7L wartości każdej z tych miar były następujące: RMSD ≈ 5 Å, GDT-TS ≈ 0.65 oraz TM-score ≈ 0.5 . W fazie lipidowej struktura najbliższa eksperymentalnej została uzyskana dla białek o kodzie PDB: 1IIJ oraz 1WAZ (Rys. 47). Dla pierwszego z nich osiągnięto następujące wartości miar porównawczych: RMSD ≈ 5 Å, GDT-TS ≈ 0.6 oraz TM-score ≈ 0.35 . Natomiast dla białka 1WAZ wartości RMSD były równe ok. 6 Å, GDT-TS ok. 0.5, a TM-score ok. 0.35.



Rys.47. Wykresy słupkowe wartości RMSD, GDT-TS oraz TM-score dla struktur eksperymentalnych badanych białek oraz ich reprezentatywnych konformacji uzyskanych z symulacji MREMD dla $\Delta G_{pi} = 0$ kcal/mol w fazie wodnej (niebieski), lipidowej (żółty) oraz dwuwarstwie lipidowej skalowanej przez współczynnik $s_L = 0.7$ (zielony).

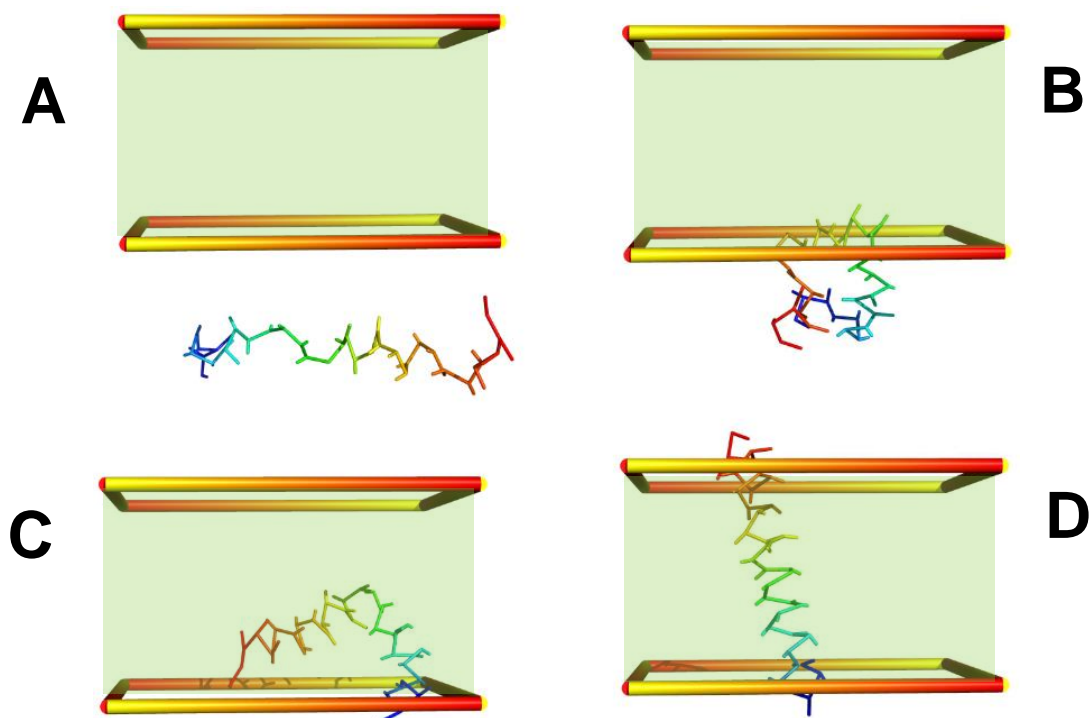
Struktura badanych białek membranowych była zależna od środowiska w jakim były one symulowane. Można to dokładniej pokazać na przykładzie białka o kodzie PDB: 1IIJ, które było najmniejszym i najprostszym spośród wszystkich badanych w tym projekcie białek membranowych. Struktura eksperymentalna tego białka ma kształt pojedynczej α -helisy (Rys. 48A). Białko to jest zbudowane głównie z hydrofobowych reszt aminokwasowych i glicyny, a polarne reszty tego białka znajdują się wyłącznie na N- i C-końcach jego łańcucha polipeptydowego. W dwuwarstwie lipidowej skalowanej przez współczynnik s_L równy 0.7 konformacja reprezentatywna tego białka była zbliżona do struktury eksperymentalnej; była jednak od niej krótsza i posiadała wyraźnie zniekształcony N-koniec (Rys. 48B). Środkowy fragment tego białka znajdował się w niepolarnym wnętrzu

membrany, a polarne końce tego peptydu rozciągały się po obu stronach granic membrany. W fazie lipidowej białko 1IJ miało kształt skompresowanej α -helisy (Rys. 48C). W fazie wodnej nie udało się odtworzyć struktury II-rzędowej tego białka (Rys. 48D). W tym środowisku polarne końce białka 1IJ uległy rozciągnięciu, a jego hydrofobowe reszty tworzyły ciasno skompresowane skupienie blokujące możliwość odtworzenia jego II-rzędowej struktury. W przypadku pozostałych białek membranowych zaobserwowano podobne zależności zmian strukturalnych od środowiska reakcji.



Rys. 48. Struktura eksperymentalna białka 1IJ (A), konformacja reprezentatywna białka 1IJ w dwuwarstwie lipidowej skalowanego przez współczynnik $s_L = 0.7$ (B), fazie lipidowej (C) oraz środowisku wodnym (D). Nad każdą z symulowanych struktur białka 1IJ umieszczono wartość RMSD liczoną względem struktury eksperymentalnej.

Ostatni etap badań sprowadzał się do określenia procesu zwiwania każdego z badanych białek membranowych w testowanym polu sił UNRES. Na przykładzie białka o kodzie PDB: 1IIJ można zauważyć, że proces ten rozpoczynał się w momencie całkowitego zanurzenia tego białka w fazie lipidowej (Rys. 49B) i kończył wówczas, gdy jego polarne końce stykały się z obydwoma granicami dwuwarstwy lipidowej (Rys. 49D).



Rys. 49 Wybrane fragmenty trajektorii białka o kodzie PDB: 1IIJ przedstawiające proces jego zwiwania się w dwuwarstwie lipidowej skalowanej przez współczynnik $s_L = 0.7$. Trajektoria tego białka została uzyskana na drodze symulacji MREMD w testowanym polu sił UNRES z energią swobodną ΔG_{pi} równą 0 kcal/mol.

Wyniki moich badań dotyczących rozszerzenia pola sił UNRES do środowiska lipidowego znajdują się w załączniku 2 stanowiącym publikację „*Extension of the UNRES Coarse-Grained Force Field to Membrane Proteins in the Lipid Bilayer*” Zięba K., Ślusarz M., Ślusarz R., Liwo A., Czaplowski C., Sieradzan A. K., J. Phys. Chem. B, 2019, **123**, 7829-7839 oraz w suplemencie 2A do tego załącznika.

6. 4. Rozszerzanie parametryzacji pola sił UNRES na 5 lokalnych typów reszt aminokwasowych białek

Założeniem tego projektu badawczego było rozszerzenie obecnej wersji NEWCT-9P pola sił UNRES z trzech³⁹ do pięciu lokalnych typów reszt aminokwasowych białek w celu zwiększenia zdolności tego pola do przewidywania struktur białek w środowisku wodnym.

Moja rola w tym projekcie sprowadzała się do kalibracji i optymalizacji nowo sparametryzowanego pola sił UNRES, a następnie do walidacji tego pola pod kątem testowania jego zdolności predykcyjnych przy wykorzystaniu do tego celu odpowiednich białek testowych.

Kalibracja i walidacja pola sił UNRES:

Osiągnięcie wyżej postawionego celu badawczego wiązało się z koniecznością poprawy opisu oddziaływań lokalnych i korelacyjnych w aktualnej wersji pola sił UNRES o nazwie NEWCT-9P³⁹, które zostało szczegółowo omówione w rozdziale 5.4.2.

Pierwszy etap badań polegał na zmodyfikowaniu potencjałów U_{tor} , U_{turn} oraz U_{corr} i zastąpieniu ich nowo sparametryzowanymi potencjałami rozszerzonymi do pięciu lokalnych typów reszt aminokwasowych białek. Podział ten jest bardziej użyteczny od dotychczas stosowanego, ponieważ kładzie większy nacisk na wzajemne podobieństwa strukturalne każdej z reszt aminokwasowych, ich właściwości oraz lokalizację wewnątrz struktury białek.

Tab. 1. Przypisanie reszt aminokwasowych białek do uproszczonego alfabetu złożonego z trzech grup¹⁵⁹.

Numer grupy	Reszty aminokwasowe
1	Gly
2	Ala, Cys, Asp, His, Glu, Phe, Leu, Ile, Lys, Gln, Met, Asn, Thr, Arg, Ser, Tyr, Val, Trp
3	Pro

Tab. 2. Przypisanie reszt aminokwasowych białek do uproszczonego alfabetu złożonego z pięciu grup¹⁵⁹.

Numer grupy	Reszty aminokwasowe
1	Gly
2	Asp, Asn
3	His, Ala, Glu, Arg, Lys, Gln, Thr, Ser
4	Ile, Cys, Phe, Val, Leu, Met, Tyr, Trp
5	Pro

Analityczne wyrażenia na potencjały korelacyjne i lokalne zostały wyprowadzone w sposób analogiczny jak w przypadku parametryzacji wersji NEWCT-9P pola sił UNRES z podziałem reszt aminokwasowych na trzy grupy³⁹ (Tab. 1) z tą różnicą, że do obecnych badań wykorzystano pięć prostych modelowych układów składających się z terminalnie blokowanej glicyny, L-alaniny, L-proliny, L-leucyny oraz L-asparaginy. Następnie dla każdego z tych układów wykonano obliczenia na poziomie MP2⁴⁸ z bazą 6-31G** w celu wyznaczenia powierzchni ich energii potencjalnej. Wyniki tych obliczeń posłużyły do określenia wartości PMF dla potencjałów korelacyjnych i lokalnych. Do każdej z obliczonych wartości PMF dopasowano wyprowadzone wcześniej wyrażenia analityczne i tym sposobem uzyskano zmodyfikowane formy funkcyjne potencjałów U_{tor} , U_{turn} oraz U_{corr} , które zostały zaimplementowane do pola sił UNRES przy pomocy jednowymiarowego szeregu Fouriera.

Kolejny etap badań polegał na iteracyjnej kalibracji i optymalizacji pola sił UNRES rozszerzonego do pięciu lokalnych typów reszt aminokwasowych białek. Kalibracja była wykonywana przy pomocy metody H-REMD⁷⁷, a jej wyniki zostały przetworzone przy pomocy metody WHAM^{78,79}. Do optymalizacji parametrów pola sił wykorzystano metodę największego podobieństwa (ang.: *Maximum Likelihood*, MAXLIKE)⁸⁰, której idea działania została szczegółowo opisana w rozdziale 5.4.2.1.

Ostatni etap badań polegał na walidacji nowo sparametryzowanego, skalibrowanego i zoptymalizowanego pola sił UNRES przy pomocy metody MREMD¹⁰¹ na zestawie wielu białek testowych zawierających różne typy struktur II-rzędowych.

Charakterystyka białek treningowych:

Kalibracja i optymalizacja pola sił UNRES składała się z 8 iteracji. Do obliczeń wykorzystano zestaw 9 małych białek treningowych pozbawionych N- i C-końców. Białka te posiadały różne typy struktur II-rzędowych, a ich eksperymentalne zespoły konformacyjne zostały wygenerowane w różnych warunkach temperaturowych (Rys. 50). Szczegółowe informacje nt. każdego z nich zostały przedstawione w poniższej tabeli.

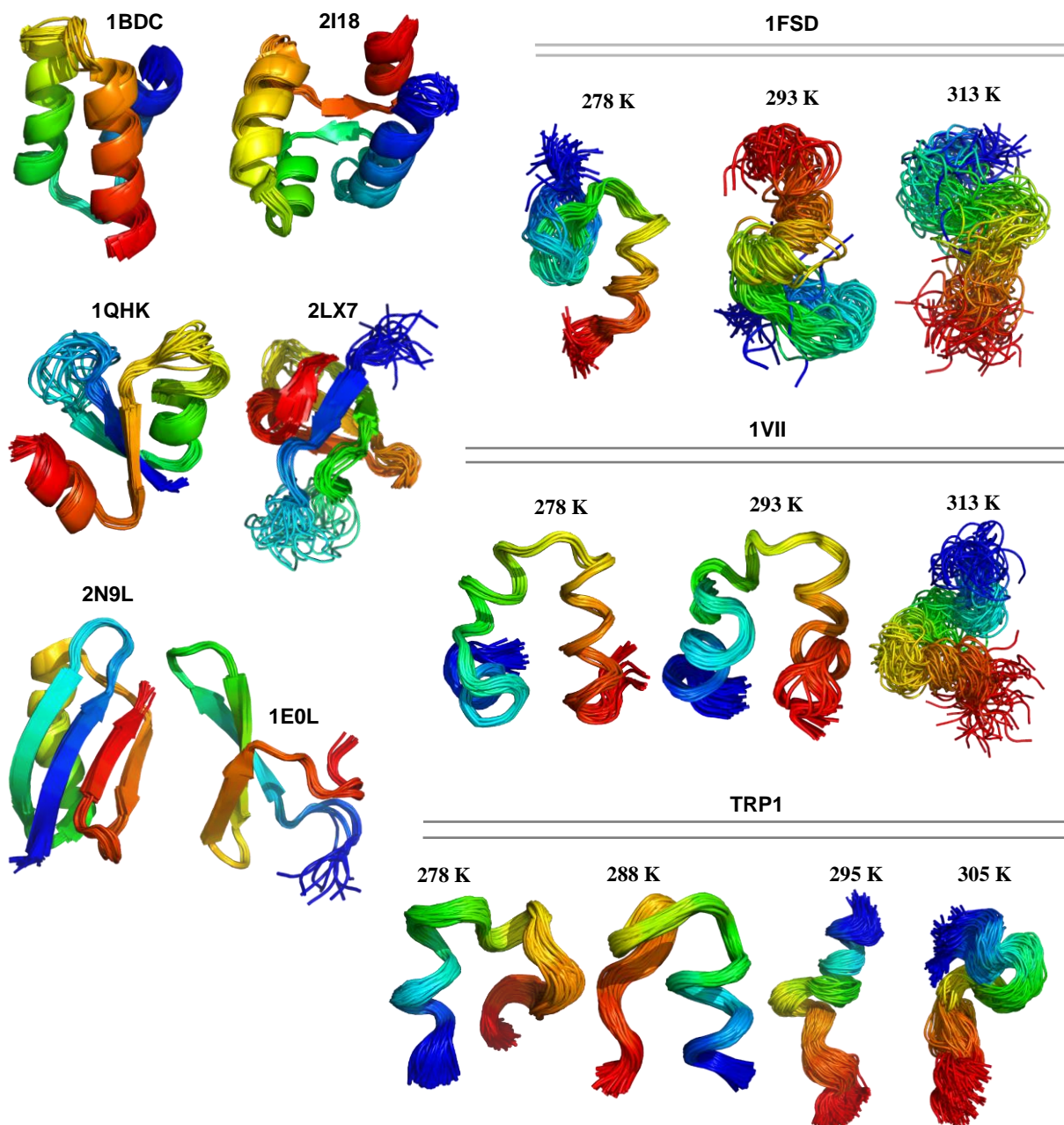
Tab.3. Szczegółowe informacje nt. białek treningowych użytych do parametryzacji pola sił UNRES w wersji NEWCT-9P.

Nazwa	Kod PDB	Liczba reszt	Typ struktur II-rzędowych	Temperatura** [K]
Domena SH3 specyficznie hamowanego wzrostu białka GAS7 człowieka	2LX7	60	$\alpha + \beta$	298
Syntetyczna konstrukcja pełnej sekwencji motywu $\beta\beta\alpha$	1FSD	28	$\alpha + \beta$	278, 293, 313
Immunoglobulina G wiążąca białko G	2N9L	57	$\alpha + \beta$	295
N-terminalna domena rybonukleazy H	1QHK	47	$\alpha + \beta$	298
Domena FBP28WW białka wiążącego w organizmie myszy	1E0L	28	β	285
C-terminalna domena motywu α -pętla- α białka wiążącego wapń w pelzaku czerwonej (<i>Entamoeba histolytica</i>)	2I18	64	α	308
Domena B wiążąca immunoglobulinę w białku A gronkowca złocistego	1BDD	46	α	298
Wilina	1VII	35	α	278, 293, 313
Klatka tryptofanowa (TRP1)*	-	20	α	278, 288, 295, 305

*Struktura niezdeponowana jeszcze w bazie PDB

**Temperatura, w której wygenerowane zostały eksperymentalne zespoły konformacyjne danego białka

Wybór tych konkretnych białek był podyktowany ich dużym stopniem zróżnicowania strukturalnego oraz małą liczbą reszt aminokwasowych. Struktury wszystkich tych białek zostały oznaczone metodą NMR¹²⁰, przy czym klatka tryptofanu (oznaczona jako TRP1) o sekwencji DAYAQLKDGGPSAGRPPPS, a także białka o kodach PDB: 1VII oraz 1FSD zostały wyznaczone w kilku różnych temperaturach, tj. poniżej, w pobliżu oraz powyżej temperatury topnienia⁶⁸.



Rys. 50. Zespoły konformacyjne struktur eksperymentalnych białek treningowych stosowanych na etapie kalibracji i optymalizacji pola sił UNRES.

W każdej iteracji pole sił UNRES było kalibrowane na zestawie wielu zespołów konformacyjnych białek treningowych (Rys. 50) przy pomocy metody H-REMD⁷⁷. Do symulacji w iteracji „0” użyto zestawu parametrów z wersji NEWCT-9P pola sił UNRES sparametryzowanej na 3 typy lokalnych reszt aminokwasowych białek. W każdej kolejnej iteracji parametry te były optymalizowane przy pomocy metody MAXLIKE⁸⁰.

Na odległości między atomami C^α oraz między zjednoczonymi łańcuchami bocznymi w symulowanych i eksperymentalnych zespołach konformacyjnych białek treningowych nakładane były więzy pozycyjne q dane jako¹⁶⁰:

$$q = 1 - \frac{2}{(n_{res} - 2)(n_{res} - 3)} \sum_{i=4}^n \sum_{j=1}^{i-3} \left(\exp \left[-\frac{(d_{C_i^{\alpha}C_j^{\alpha}} - d_{C_i^{\alpha}C_j^{\alpha}}^{ref})^2}{0.25(d_{C_i^{\alpha}C_j^{\alpha}}^{ref})^2} \right] + \exp \left[-\frac{(d_{SC_iSC_j} - d_{SC_iSC_j}^{ref})^2}{0.25(d_{SC_iSC_j}^{ref})^2} \right] \right) \quad (6.4b)$$

gdzie d_{SC_i} oraz $d_{SC_i}^{(ref)}$ to odległości między środkami dwóch sąsiednich łańcuchów bocznych w symulowanych i referencyjnych zespołach konformacyjnych badanego białka treningowego, a $d_{C_i^{\alpha}C_j^{\alpha}}$ oraz $d_{C_i^{\alpha}C_j^{\alpha}}^{(ref)}$ to odległości między dwoma sąsiednimi atomami C^α w obydwu typach porównywanych ze sobą zespołów konformacyjnych badanego białka. Symulacje H-REMD składały się z 30 000 000 kroków z krokiem czasowym równym 4.89 femtosekundy. Zostały one wykonane dla 12 temperatur: 260, 262, 266, 271, 276, 282, 288, 296, 304, 315, 333, 370 K oraz 12 replik (gdzie na jedną replikę przypadały 4 temperatury) ze stałą siłową k wynoszącą 0, 2, 4, 8, 16, 32, 64 i 128 kcal/mol. Repliki te były wymieniane co 10 000 kroków. Do całkowania równań ruchu zastosowano algorytm A-MTS¹⁶¹. Wyniki symulacji zostały przetworzone przy pomocy metody WHAM^{78,79} w celu określenia parametrów pola sił po danej iteracji, które były następnie optymalizowane przy pomocy metody MAXLIKE⁸⁰. Do obliczenia funkcji optymalnego podobieństwa l^* wykorzystano zespoły konformacyjne wszystkich białek treningowych, które zostały wygenerowane w ostatnich 1000 kroków każdej trajektorii. Po wyznaczeniu funkcji l^* następowało zakończenie bieżącej iteracji. Przed rozpoczęciem nowej iteracji funkcje l^* z aktualnej i poprzedniej iteracji były ze sobą porównywane.

Dla zadanych parametrów pola sił UNRES dokonano jego wstępnej walidacji na zbiorze składającym się z zespołów konformacyjnych wszystkich białek treningowych. Były to symulacje MREMD¹⁰¹ składające się z 20 000 000 kroków z krokiem czasowym równym 4.89 femtosekundy. Symulacje te były prowadzone dla 12 temperatur: 260, 262, 266, 276, 282, 288, 296, 304, 315, 333 i 370 K i 12 replik, gdzie na jedną replikę przypadały 4 temperatury. Wymiana replik odbywała się po 10 000 kroków. Wyniki symulacji zostały przetworzone przy pomocy metody WHAM^{78,79} w celu określenia zmian pojemności cieplnej układu, prawdopodobieństwa występowania konformacji badanych białek w niskoenergetycznych regionach ich przestrzeni fazowej oraz wartości C^αRMSD dla symulowanych i eksperymentalnych struktur białek treningowych. Konformacje uzyskane w temperaturach 260, 280 i 300 K zostały następnie pogrupowane do pięciu osobnych skupień przy pomocy metody Ward'a¹⁰⁸. Do dalszej analizy wybrano skupienie o najniższej średniej wartości C^αRMSD względem struktur eksperymentalnych. Dla pierwszej oraz najbardziej

reprezentatywnej konformacji z tego skupienia obliczone zostały wartości GDT-TS^{18,19} względem odpowiadających im struktur eksperymentalnych.

Kolejny etap badań sprowadzał się do pełnowymiarowej walidacji pola sił UNRES przy pomocy metody MREMD¹⁰¹ oraz zestawu wielu białek testowych.

Walidacja pola sił UNRES:

Do walidacji pola sił UNRES wykorzystano zestaw 77 białek testowych o różnej liczbie reszt aminokwasowych oraz typach struktury II-rzędowej, spośród których 22 było białkami α -helikalnymi, 24 zawierały wyłącznie β -kartki, a 25 stanowiły białka zbudowane zarówno z α -helis jak i β -kartek. Kody PDB oraz liczba reszt aminokwasowych dla każdego z tych białek zostały przedstawione w poniższej tabeli.

Tab. 4. Zestawienie białek używanych w testach pola sił UNRES rozszerzonego do 5 typów lokalnych reszt aminokwasowych.

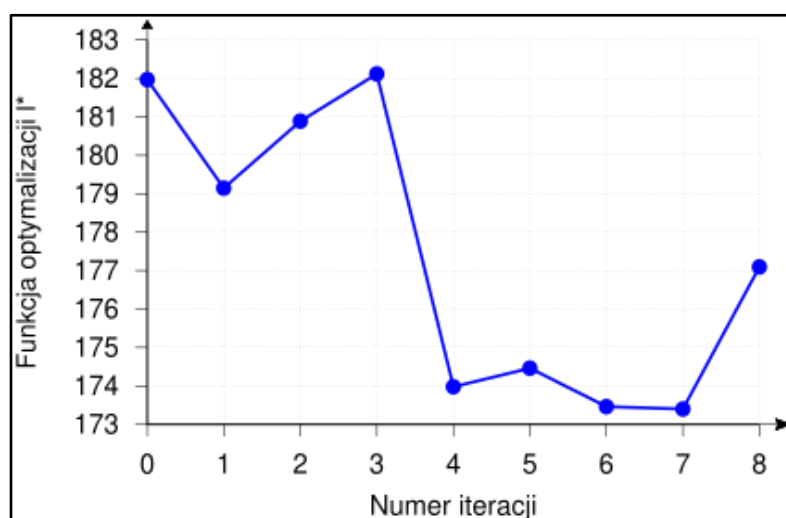
α (28 białek)		β (24 białka)		$\alpha + \beta$ (25 białek)	
Kod PDB	Liczba reszt	Kod PDB	Liczba reszt	Kod PDB	Liczba reszt
1L2Y	20	1ED7	45	1HNS	47
1RIJ	23	1TPM	50	1E0G	48
1YRF	35	1BK2	57	1EM7	56
1PPT	36	4F98-A	62	1PGA	56
1RES	43	1FYN	62	2FMR	65
1BBL	51	2LGN	66	1SAP	66
1GAB	53	1CSP	67	3CI2	66
1ENH	54	1HRE	67	1HYW	68
1PRU	56	1MJC	69	1STU	68
1FEX	59	1AH9	71	1BBY	69
1KOY	62	4AIT	74	1VIG	71
2L09	62	1IYU	79	1UBQ	76
1LQ7	67	1PHT	85	2N2U	77
1E68	70	1GH5	87	2PTL	78
1POU	71	2KYW	87	4N6T	79
1BG8-A	76	1WKT	88	5G3Q-B	79
1CLB	76	1TEN	90	1GHH	80
1J7O	76	2LVC	91	1LEA	84
1ACP	77	3NCM	92	1OPD	85
1EO0	77	1WIU	93	1IQO	88
1HYP	80	3PUC	98	1PTF	88
2HEP	85	1ACX	108	1TIG	94
1A1W	91	4QRL	110	2ACY	98
2YGS	92	3POJ	113	2MQ8	112
2CRB	97			1THX	115
1P68	102				
1K40	126				
1A63	130				

Dla każdego z białek testowych przeprowadzono symulacje MREMD¹⁰¹, której parametry i sposób analizy uzyskanych z niej wyników były takie same jak w przypadku

wstępnej walidacji pola sił UNRES dokonanej na zbiorze wielu zespołów konformacyjnych białek treningowych.

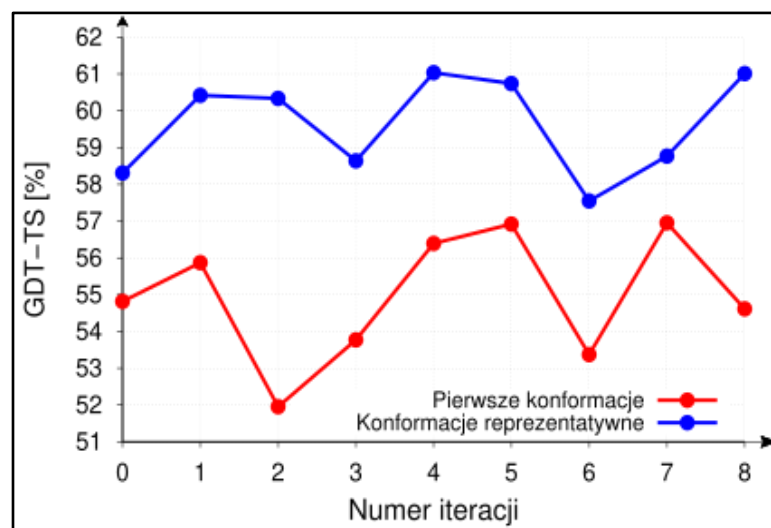
Interpretacja wyników obliczeń:

Funkcja optymalizacji l^* dla symulowanych i eksperymentalnych zespołów konformacyjnych białek treningowych zmniejszała swoją wartość aż do siódmej iteracji włącznie. Na tej podstawie można stwierdzić, że rozszerzenie parametryzacji pola sił UNRES do pięciu lokalnych typów reszt aminokwasowych wpłynęło na poprawę zdolności predykcyjnych tego pola dla 9 badanych białek treningowych (Rys. 51).



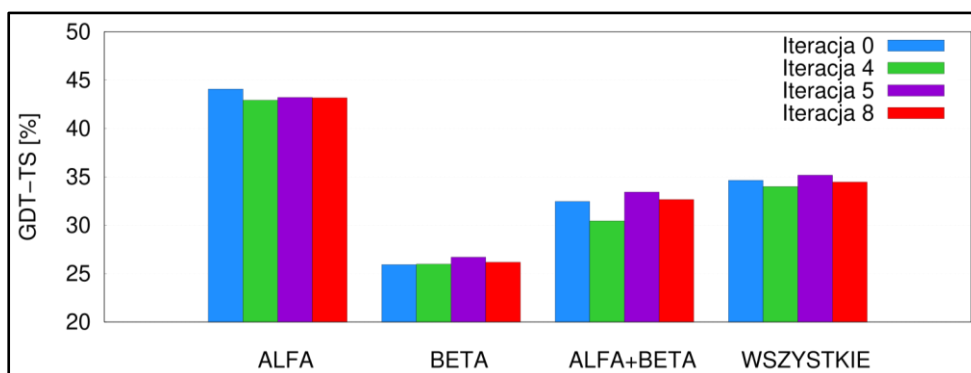
Rys. 51. Wykres zależności funkcji optymalizacji l^* od liczby iteracji w cyklu optymalizacyjnym pola sił UNRES rozszerzonego na 5 lokalnych typów reszt aminokwasowych.

Wartości miar GDT-TS uzyskane dla pierwszych i reprezentatywnych konformacji białek treningowych względem ich struktur eksperymentalnych świadczą o poprawie efektywności przewidywań testowanego pola sił UNRES (Rys. 52). Między iteracją „0” oraz iteracją 7 dla pierwszych konformacji z reprezentatywnego skupienia zaobserwowano ok. 2% wzrost wartości GDT-TS z 55 do 57%, a dla reprezentatywnych konformacji z tego skupienia wartości tej miary wzrosły o ok. 1% z 58 do 59%. Uzyskane wartości GDT-TS świadczą o dobrym podobieństwie między symulowanymi i eksperymentalnymi strukturami białek treningowych.

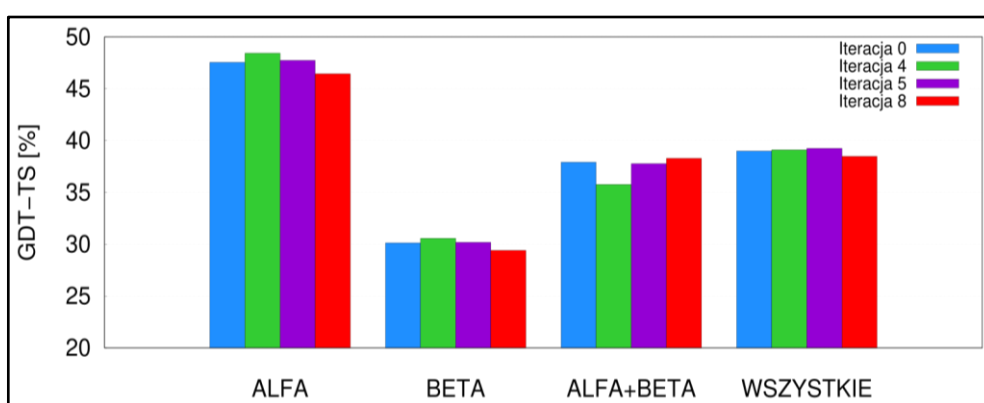


Rys. 52. Wykres zmian średnich wartości GDT-TS dla pierwszych i reprezentatywnych konformacji białek treningowych względem liczby iteracji w cyklu optymalizacyjnym pola sił UNRES.

W przypadku wszystkich białek testowych niewielką poprawę zdolności predykcyjnych pola sił UNRES uzyskano dopiero w piątej iteracji. Pierwsze konformacje białek pochodzące ze skupienia o najniższej średniej wartości C^{α} RMSD posiadały mniejsze średnie wartości GDT-TS w stosunku do konformacji reprezentatywnych (Rys. 53 i 54). Testowane pole sił UNRES najlepiej odtworzyło strukturę przestrzenną α -helikalnych białek testowych. Średnie z wartości GDT-TS dla konformacji reprezentatywnych tych białek względem ich struktur eksperymentalnych uplasowały się na poziomie 48.43% (Rys. 54). Pole sił UNRES miało największe problemy z odtworzeniem białek o strukturze β -kardki. Największe średnie wartości GDT-TS dla reprezentatywnych konformacji tych białek zostały osiągnięte w czwartej iteracji, gdzie wynosiły 30.57% (Rys. 54). W przypadku białek zawierających α -helisy i β -kardki największe średnie wartości GDT-TS z ich konformacji reprezentatywnych względem struktur eksperymentalnych były równe 39.25% i zostały one osiągnięte w iteracji ósmej (Rys. 54).



Rys. 53. Zestawienie średnich wartości GDT-TS dla pierwszych konformacji wszystkich typów białek testowych pochodzących ze skupień o najniższych średnich wartościach C^αRMSD względem ich struktur eksperymentalnych.



Rys. 54. Zestawienie średnich wartości GDT-TS dla reprezentatywnych konformacji wszystkich typów białek testowych pochodzących ze skupień o najniższych średnich wartościach C^αRMSD względem ich struktur eksperymentalnych.

6. 5. Przewidywanie struktur białek przy użyciu gruboziarnistego pola sił UNRES oraz informacji z baz danych w 13-tej edycji eksperymentu naukowego CASP

Cel tego projektu sprowadzał się do przewidywania III- i IV-rzędowych struktur białek w ramach udziału wszystkich członków grupy badawczej Prof. Adama Liwo i Prof. Cezarego Czaplewskiego w 13-tej edycji eksperymentu naukowego CASP jako grupa KIAS-Gdansk. Do przewidywań wykorzystano hybrydowe podejście łączące gruboziarniste pole sił UNRES z modelowaniem homologicznym oraz wieloma modelami serwerowymi dla zadanej sekwencji aminokwasowej badanego białka.

Moja rola w tym projekcie polegała na określeniu struktury III-rzędowej dla 9 białek o liczbie reszt aminokwasowych od 199 do 529.

Opis procedury przewidywań:

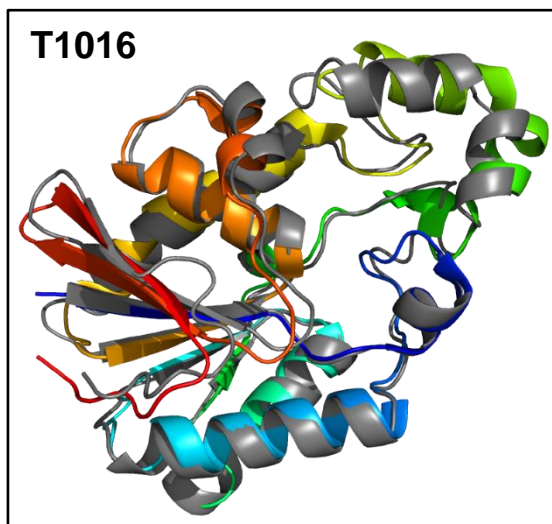
Protokół przewidywań grupy KIAS-Gdansk został szczegółowo opisany w rozdziale 5.9.3. Bazował on na połączeniu gruboziarnistego pola sił UNRES w wersji OPT-WTFSA-1⁶⁹ z modelowaniem homologicznym i więzami pozycyjnymi wyprowadzonymi na podstawie fragmentów konsensusowych z najlepszych jakościowo modeli serwerowych dla zadanej sekwencji aminokwasowej białek. W tym projekcie wagi więzów nakładanych na długości wirtualnych wiązań były równe 0.5, a dla wirtualnych kątów walencyjnych, torsyjnych oraz środków zjednoczonych łańcuchów bocznych były one równe 1.0.

Cykl badań składał się z wielu etapów. Pierwszy z nich sprowadzał się do wyboru najlepszych jakościowo modeli serwerowych dla zadanej sekwencji aminokwasowej, których jakość była oceniana przy pomocy zewnętrznego serwera DeepQA¹³⁸. Był to etap mający kluczowe znaczenie dla całej procedury obliczeniowej, ponieważ wybór większej liczby, gorszych jakościowo modeli mógłby niekorzystnie wpłynąć na oczekiwaną dokładność przewidywań badanych białek. W grupie KIAS-Gdansk do przewidywań najczęściej używano modeli pochodzących z serwerów QUARK, BAKER-ROSETTASERVER, Zhang-Server, RaptorX-DeepModeller oraz RaptorX-TBM. Kolejny etap badań polegał na wyodrębnieniu z najlepszych jakościowo modeli tzw. fragmentów konsensusowych, czyli wspólnych dla każdego z nich fragmentów strukturalnych składających się co najmniej z 20 reszt aminokwasowych, których wartości C^αRMSD byłyby mniejsze niż 7 Å. Na podstawie tych fragmentów zostały wyprowadzone więzy pozycyjne, które następnie były nakładane na modelowane białko. Etap czwarty tych badań polegał na próbkowaniu przestrzeni konformacyjnej badanego białka przy wykorzystaniu do tego celu metody MREMD¹⁰¹. Symulacje te były prowadzone dla 12 temperatur – 260, 262, 266, 271, 276, 282, 288, 296, 304,

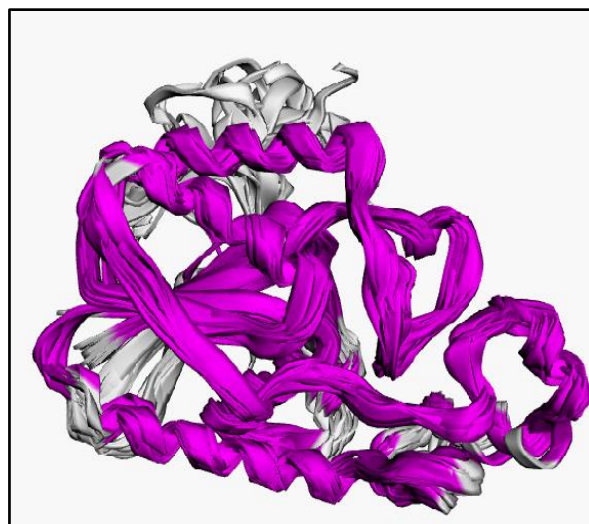
315, 333 i 370 K oraz 12 replik, w których na jedną replikę przypadały 4 temperatury. Do kontroli temperatury symulowanego układu zastosowano termostat Langevina⁹². W celu przyspieszenia tych obliczeń współczynnik tarcia λ był skalowany przez wartość równą 0.1. Każda symulacja składała się z 20 000 000 kroków o długości kroku czasowego równego 4.86 femtosekund. Do całkowania równań ruchu wykorzystano algorytm A-MTS¹⁶¹. W etapie piątym dane z ostatnich 200 kroków symulacji zostały przetworzone przy pomocy metody WHAM^{78,79} w celu wyznaczenia zależności zmian pojemności cieplnej układu od temperatury, prawdopodobieństwa wystąpienia konformacji symulowanego białka w niskoenergetycznych regionach jego przestrzeni fazowej oraz wartości C^αRMSD dla niskoenergetycznych konformacji tego białka. Niskoenergetyczne konformacje badanych przeze mnie białek były zazwyczaj uzyskiwane w trzech temperaturach: 260, 280 i 300 K. Każda z nich była następnie grupowana przy pomocy metody Ward'a¹⁰⁸ do pięciu osobnych skupień. Konformacją reprezentacyjną badanego białka była struktura posiadająca najniższe wartości C^αRMSD względem średniej tej miary dla danego skupienia. Kolejny etap badań polegał na przekształceniu konformacji reprezentacyjnej badanego białka do struktury pełnoatomowej w wersji ff14SB¹⁵⁷ pola sił AMBER przy pomocy algorytmów PULCHRA¹⁴² oraz SCWRL¹⁴³. W ostatnim etapie badań należało zminimalizować energię potencjalną struktury pełnoatomowej oraz wykonać dla niej krótką symulację MD^{81,82} mającą na celu usunięcie potencjalnych zawał sterycznych oraz zoptymalizowanie ułożenia przestrzennego łańcuchów bocznych w cząsteczce badanego białka.

Przykładowe wyniki przewidywań:

Poniższe rysunki przedstawiają nałożenie trzech spośród dziewięciu badanych przeze mnie III-rzędowych struktur białek o kodach T1016 (Rys. 55), T1018 (Rys. 57) oraz T1021s2 (Rys. 59) z odpowiadającymi im strukturami eksperymentalnymi pobranymi z bazy PDB¹¹⁷.

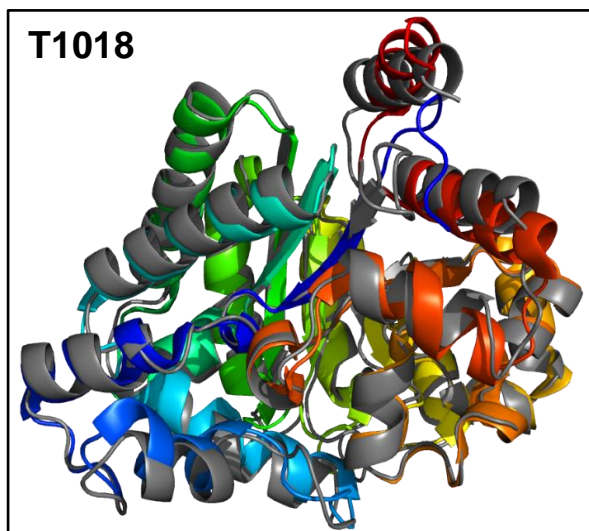


Rys. 55. Porównanie najlepszej spośród pięciu przewidzianych w $T = 280$ K struktur białka T1016 (kolor szary) względem domeny A struktury eksperymentalnej białka o kodzie PDB 6E4B.

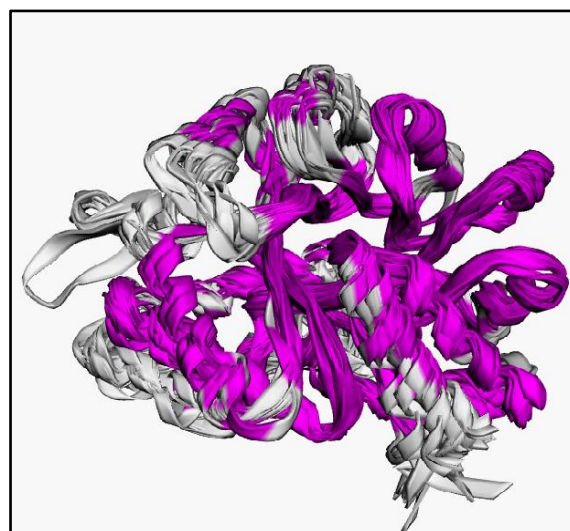


Rys. 56. Rysunek przedstawiający nałożenie najlepszych jakościowo modeli serwerowych dla białka T1016. Kolorem fioletowym oznaczono odpowiadające im fragmenty konsensusowe.

Białko T1016 jest zbudowane z α -helis oraz β -kartek i składa się z 203 reszt aminokwasowych. Do wyprowadzenia więzów dla tego białka użyłam łącznie 22 modeli pochodzących z serwerów BAKER-ROSETTASERVER, Delta-Gelly-Server, Seok-server, RaptorX-DeepModeller, Seok-assembly, MULTICOM_CLUSTER, MULTICOM_NOVEL oraz QUARK. Fragmenty konsensusowe tych modeli zostały oznaczone kolorem fioletowym na Rys. 56. Przewidywana przeze mnie struktura została sklasyfikowana na 11 miejscu wśród najlepszych przewidywań z wartością GDT-TS względem struktury eksperymentalnej równą 81.68%. Największe różnice między zmodelowaną i eksperymentalną strukturą tego białka są widoczne w obszarach pętli, długości niektórych β -kartek oraz lokalnych przemieszczeniach pojedynczych α -helis (Rys. 55).

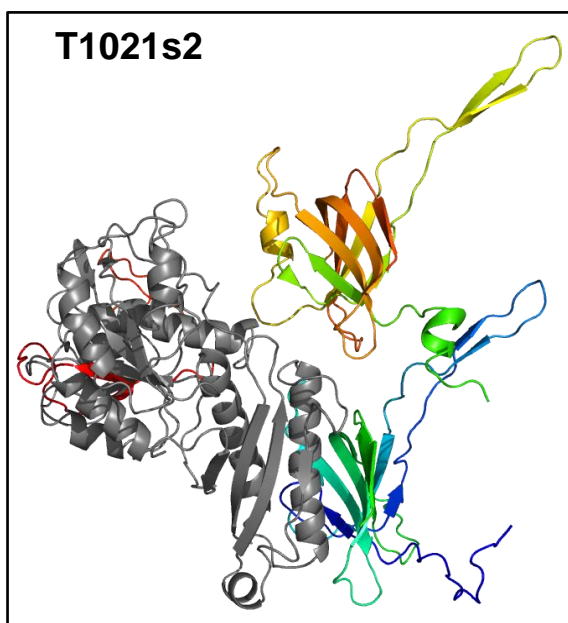


Rys. 57. Porównanie najlepszej spośród pięciu przewidzianych w $T = 280$ K struktur białka T1018 (kolor szary) względem domeny A struktury eksperymentalnej białka o kodzie PDB 6N91.

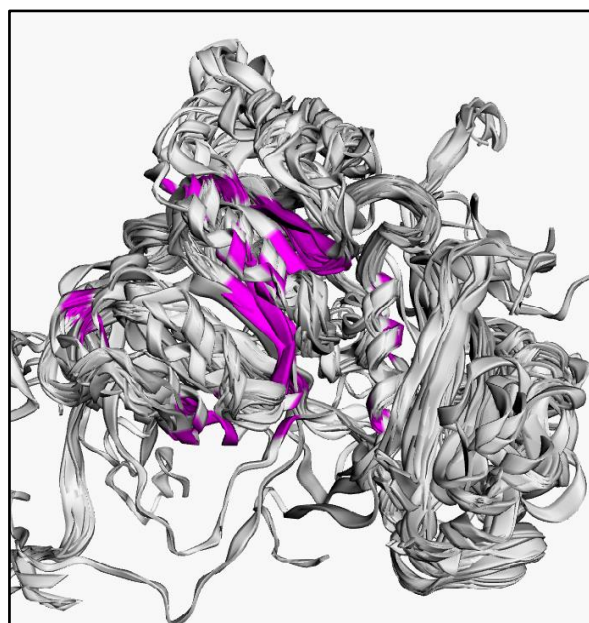


Rys. 58. Rysunek przedstawiający nałożenie najlepszych jakościowo modeli serwerowych dla białka T1018. Kolorem fioletowym oznaczono odpowiadające im fragmenty konsensusowe.

Białko T1018 jest zbudowane z α -helis oraz β -kardtek i składa się z 334 reszt aminokwasowych. Do jego zmodelowania wykorzystałam łącznie 20 modeli pochodzących z serwerów, Zhang-Server, QUARK, RaptorX-TBM, Seok-server, Yang-server oraz slbio_server oraz BAKER-ROSETTASERVER. Fragmenty konsensusowe tych modeli zostały oznaczone na Rys. 58 kolorem fioletowym. Wartość GDT-TS dla modelowanego białka względem jego struktury eksperymentalnej była równa 85.70%. Największe dysproporcje między symulowaną i eksperymentalną strukturą badanego białka były widoczne w obszarze pętli, długości β -kardtek, a także lokalnych przemieszczeniach pojedynczych α -helis (Rys. 57).



Rys. 59. Porównanie najlepszej spośród pięciu przewidzianych w $T = 300\text{ K}$ struktur białka T1021s2 (kolor szary) względem struktury eksperymentalnej białka o kodzie PDB 6RAP.



Rys. 60. Rysunek przedstawiający nałożenie najlepszych jakościowo modeli serwerowych dla białka T1021s2. Kolorem fioletowym oznaczono odpowiadające im fragmenty konsensusowe.

Białko o kodzie T1021s2 jest zbudowane z α -helis oraz β -katek i zawiera 354 reszty aminokwasowe. Do wyprowadzenia więzów dla tego białka wykorzystałam 20 modeli otrzymanych z serwerów takich jak Zhang-Server, RaptorX-TBM, RaptorX-DeepModeller, QUARK oraz BAKER_ROSETTASERVER. Fragmenty konsensusowe tych modeli zostały oznaczone na Rys. 60 kolorem fioletowym. Wartość GDT-TS dla modelowanego białka względem jego struktury eksperymentalnej wynosiła 65.62 %. Mniejsze podobieństwo obydwu tych struktur w porównaniu do podobieństwa uzyskanego dla dwóch wyżej opisywanych białek było wynikiem wyboru gorszych jakościowo modeli serwerowych i małej liczby wspólnych dla nich fragmentów konsensusowych w stosunku do liczby reszt aminokwasowych w każdym z nich. Skutkowało to nieprawidłowym sposobem upakowania poprawnie przewidzianych typów struktur II-rzędowych i nieprawidłową ich liczbą, ponieważ w zmodelowanej strukturze wiele spośród β -katek zostało zastąpione α -helisami. (Rys. 59).

Wyniki badań związanych z przewidywaniem III- i IV-rzędowych struktur białek przy pomocy opisywanego protokołu przewidywań dla grupy KIAS-Gdansk znajdują się w załączniku 3 stanowiącym publikację „Improved Consensus-Fragment Selection in Template-Assisted Prediction of Protein Structures with the UNRES Force Field in CASP13”, Karczyńska A. S., Zięba K., Uciechowska U., Mozolewska M. A, Krupa P., Lubecka E. A., Lipska A. G., Sikorska C., Samsonov S. A., Sieradzan A. K., Giełdoń A., Liwo A., Ślusarz R., Ślusarz M., Lee J., Joo K., Czaplewski C., *J. Chem. Inf. Model.*, 2020, **60**, 1844-1864.

7. PODSUMOWANIE I WNIOSKI

1. W modelach SPC/E, TIP3P-Ew oraz TIP4P-2005 wzrost temperatury procesu hydratacji niepolarniej cząstki LJ-Xe oraz pseudotwardej kuli Mie-Xe o średnicy $\sigma_X = 3 \text{ \AA}$ pociągał za sobą wzrost wartości energii swobodnej ΔG , entalpii ΔH oraz członu entropowego $T\Delta S$. Wyniki uzyskane w każdym z tych modeli dla cząstki LJ-Xe i temperatury poniżej 360 K są zgodne z danymi eksperymentalnymi dla atomu Xe.
2. Wzrost temperatury procesu hydratacji rosnącej cząstki Lennarda-Jonesa oraz pseudotwardej kuli Mie o średnicy σ_X od 3 do 6 \AA pociągał za sobą wzrost wartości energii swobodnej ΔG , entalpii ΔH oraz członu entropowego $T\Delta S$. Dla większych cząstek o średnicy $\sigma_X > 6 \text{ \AA}$ wartości ΔH oraz $T\Delta S$ zwiększały się wraz z temperaturą, lecz towarzyszył temu spadek wartości ΔG . Zmiany wartości tych wielkości termodynamicznych od temperatury są zgodne z teorią hydrofobowości Pratta i Chandlera oraz danymi eksperymentalnymi dla hydratacji etanu, cykloheksanu i 2,2,4-trimetylopentanu.
3. Wzrostowi temperatury procesu asocjacji dwóch cząstek LJ-Xe oraz pseudotwardych kul Mie-Xe w minimum kontaktowym ich konfiguracji towarzyszył wzrost wartości $T\Delta S [cm]$ oraz spadek wartości $\Delta G [cm]$ i $\Delta H [cm]$.
4. W procesie hydratacji i asocjacji modelowanych przeze mnie układów większe uporządkowanie strukturalne wody zostało osiągnięte w pierwszej warstwie solwatacyjnej jednej oraz dwóch pseudotwardych kul Mie w minimum kontaktowym ich konfiguracji. Można to uzasadnić tym, że pseudotwarde kule Mie są bardziej hydrofobowe od cząstek Lennarda-Jonesa, ponieważ nie są zdolne do oddziaływań przyciągających z wodą. Dzięki temu mogą one oddziaływać ze sobą z większą siłą na mniejszych odległościach kontaktowych niż cząstki Lennarda-Jonesa. Temperaturowe zależności termodynamiczne towarzyszące efektowi hydrofobowemu pseudotwardych kul Mie w wodzie zależą jedynie od rodzaju użytego do obliczeń modelu wody. Wyniki obliczeń, które odznaczały się największą zgodnością z teoretycznym modelem góry lodowej Franka i Evansa zostały osiągnięte dla pseudotwardych kul Mie w modelu TIP4P-2005. Uzyskane przeze mnie wyniki są pomocne w zrozumieniu molekularnych podstaw efektu hydrofobowego nawet dla tak dużych układów jak białka.

5. Rozszerzone do środowiska lipidowego pole sił UNRES wykazywało największe zdolności predykcyjne dla energii swobodnej transferu zjednoczonych grup peptydowych z fazy wodnej do lipidowej ΔG_{pi} równej 0 kcal/mol w dwuwarstwie lipidowej skalowanej przez współczynnik s_L równy 0.7. W tych warunkach struktury symulowanych białek były bardzo podobne do eksperymentalnych. Struktury białek, które były całkowicie zanurzone w fazie lipidowej miały skompresowany kształt. Białka symulowane w fazie wodnej nie posiadały struktury II-rzędowej.
6. Rozszerzenie parametryzacji pola sił UNRES z trzech do pięciu lokalnych typów reszt aminokwasowych białek wpłynęło na poprawę jego zdolności predykcyjnych w środowisku wodnym. Największa poprawa była widoczna dla α -helikalnych białek testowych. Największe trudności wystąpiły podczas próby odtworzenia i upakowania struktur białek zawierających wyłącznie same β -kartki.
7. Protokół obliczeniowy używany przez grupę KIAS-Gdansk w 13-edycji eksperymentu CASP odznaczał się dużą dokładnością przewidywań, która w głównej mierze była zależna od jakości modeli dostępnych dla badanego białka, które zostały pobrane z zewnętrznych serwerów, a także od liczby wspólnych dla nich fragmentów konsensusowych.

8. LISTA PUBLIKACJI STANOWIĄCYCH ZAŁĄCZNIKI DO PRACY

1. **Zięba K.**, Czaplewski C., Liwo A., Graziano G., Hydrophobic hydration and pairwise hydrophobic interaction of Lennard-Jones and Mie particles in different water models, *Phys. Chem. Chem. Phys.*, 2020, **22**, 4758-4771. IF = 3,676.
2. **Zięba K.**, Ślusarz M., Ślusarz R., Liwo A., Czaplewski C., Sieradzan A. K., Extension of the UNRES Coarse-Grained Force Field to Membrane Proteins in the Lipid Bilayer, *J. Phys. Chem. B*, 2019, **123**, 7829-7839. IF = 2,857.
- 2A. **Zięba K.**, Ślusarz M., Ślusarz R., Liwo A., Czaplewski C., Sieradzan A. K., Supporting Information for Extension of the UNRES Coarse-Grained Force Field to Membrane Proteins in the Lipid Bilayer.
3. Karczyńska A. S., **Zięba K.**, Uciechowska U., Mozolewska M. A., Krupa P., Lubecka E. A., Lipska A. G., Sikorska C., Samsonov S. A., Sieradzan A. K., Gieldoń A., Liwo A., Ślusarz R., Ślusarz M., Lee J., Joo K., Czaplewski C., Improved Consensus-Fragment Selection in Template-Assisted Prediction of Protein Structures with the UNRES Force Field in CASP13, *J. Chem. Inf. Model.*, 2020, **60**, 1844-1864. IF = 4,956.

9. LISTA POZOSTAŁYCH PUBLIKACJI

Publikacje z listy filadelfijskiej:

1. Lubecka E. A., Karczyńska A. S., Lipska A. G., Sieradzan A. K., **Zięba K.**, Sikorska C., Uciechowska U., Samsonov S. A., Krupa P., Mozolewska M. A., Golon Ł., Giełdoń A., Czaplewski C., Ślusarz R., Ślusarz M., Crivelli S. N., Liwo A., Evaluation of the scale-consistent UNRES force field in template-free prediction of protein structures in the CASP13 experiment, *Journal of Molecular Graphics and Modeling*, 2019, **92**, 154-166. IF = 2,079.
2. Antoniak A., Biskupek I., Bojarski K. K., Czaplewski C., Giełdoń A., Kogut M. M., Kogut M., Krupa P., Lipska A. G., Liwo A., Lubecka E. A., Marcisz M., Maszota-Zieleniak M., Samsonov S. A., Sieradzan A. K., Ślusarz M. J., Ślusarz R., Wesołowski P. A., **Zięba K.**, Modeling protein structures with the coarse-grained UNRES force field in the CASP14 experiment, *Journal of Molecular Graphics and Modeling*, 2021, **108**, 1-11. IF = 2,518.
3. Lensink M. F., Brysbaert G., Mauri T., Nadzirin N., Velankar S., Chaleil R. A. G., Clarence T., Bates P. A., Kong R., Liu B., Yang G., Liu M., Shi H., Lu X., Chang S., Roy R. S., Quadir F., Liu J., Cheng J., Antoniak A., Czaplewski C., Giełdoń A., Kogut M., Lipska A. G., Liwo A., Lubecka E. A., Maszota-Zieleniak M., Sieradzan A. K., Ślusarz R., Wesołowski P. A., **Zięba K.**, del Carpio Muñoz C. A., Ichiishi E., Harmalkar A., Gray J. J., Bonvin A. M. J. J., Ambrosetti F., Honorato R. V., Jandova Z., Jiménez-García B., Koukos P. I., Van Keulen S., Van Noort C. W., Réau M., Roel-Touris J., Kotelnikov S., Padhorny D., Porter K. A., Alekseenko A., Ignatov M., Desta I., Ashizawa R., Sun Z., Ghani U., Hashemi N., Vajda S., Kozakov D., Rosell M., Rodríguez-Lumbreras L. A., Fernandez-Recio J., Karczynska A., Grudin S., Yan Y., Li H., Lin P., Huang S. Y., Christoffer C., Terashi G., Verburt J., Sarkar D., Aderinwale T., Wang X., Kihara D., Nakamura T., Hanazono Y., Gowthaman R., Guest J. D., Yin R., Taherzadeh G., Pierce B. G., Barradas-Bautista D., Cao Z., Cavallo L., Oliva R., Sun Y., Zhu S., Shen Y., Park T., Woo H., Yang J., Kwon S., Won J., Seok C., Kiyota Y., Kobayashi S., Harada Y., Takeda-Shitaka M., Kundrotas P. J., Singh A., Vakser I. A., Dapkūnas J., Olechnovič K., Venclovas Č., Duan R., Qiu L., Xu X., Zhang S., Zou X., Wodak S. J., Prediction of protein assemblies, the next frontier: The CASP14-CAPRI experiment, *Proteins*, 2021; 89:1800 -1823. IF = 3,756.

10. WYKAZ RYSUNKÓW

- Rys. 1. Schematyczne przedstawienie łańcucha polipeptydowego białek z zaznaczonymi kątami dwuściennymi φ oraz ψ . Kolorem pomarańczowym oznaczono wiązanie peptydowe, niebieskim N-koniec, a czerwonym C-koniec łańcucha polipeptydowego.
- Rys. 2. Przykładowe reprezentacje przestrzenne białka fibrylarnego – kolagenu (A), globularnego - mioglobiny (B) oraz membranowego – bakteriorodopsyny (C) (membrana została oznaczona kolorem czerwonym).
- Rys. 3. Przykładowe reprezentacje przestrzenne struktur białek o różnym stopniu rzędowości.
- Rys. 4. Schematyczny wykres Ramachandrana przedstawiający dozwolone i niedozwolone sterycznie typy zgięć w łańcuchu głównym białka¹⁵.
- Rys. 5. Schematyczne przedstawienie efektu hydrofobowego dla dwóch niepolarnych cząstek wprowadzonych do wody. Kolor zielony przedstawia solwatowane cząstki, a niebieski sąsiadujące z nimi cząsteczki wody.
- Rys. 6. Schematyczne przedstawienie etapu inicjacji fałdowania struktur białek. Oddziaływania hydrofobowe między parą niepolarnych łańcuchów bocznych zostały oznaczone linią przerywaną. Struktury niezwinięta (po lewej) i zwinięta (po prawej) istnieją w stanie równowagi chemicznej³⁷.
- Rys. 7. Schemat obrazujący występowanie oddziaływań lokalnych w łańcuchu głównym białek w gruboziarnistym polu sił UNRES³⁹.
- Rys. 8. Projekcja przestrzenna białka o kodzie PDB: 1III w reprezentacji pełnoatomowej (A), zjednoczonych atomów (B) oraz gruboziarnistej na przykładzie pola UNRES (C).
- Rys. 9. Schematyczne przedstawienie A) trójpunktowego oraz B) czteropunktowego modelu wody.
- Rys. 10. Schematyczne przedstawienie łańcucha polipeptydowego białek w gruboziarnistym polu sił UNRES. Kolorem niebieskim oznaczono zjednoczone grupy peptydowe (p), białym atomy C^α, a elipsy oznaczone kolorem różowym, fioletowym i zielonym reprezentują zjednoczone łańcuchy boczne (SC). Zmienna θ_i to kąt walencyjny zawarty między trzema kolejnymi atomami C^α, zmienna γ_i to kąt torsyjny zawarty między czterema kolejnymi atomami C^α, natomiast kąty oznaczone jako α_i oraz β_i opisują położenie zjednoczonych łańcuchów bocznych względem łańcucha głównego białek⁶⁸.
- Rys. 11. Schematyczne przedstawienie zachowania cząsteczki wody w symulowanym układzie z zastosowaniem periodycznych warunków brzegowych. Kolorem seledynowym oznaczono pudło symulacyjne, pozostałe obszary przedstawiają jego dokładne repliki. Promień odcięcia między dwiema cząstkami został oznaczony kolorem szarym.
- Rys. 12. Wykresy przedstawiające zależności zmian energii potencjalnej $U(r)$ względem odległości oddziałujących ze sobą cząstek. Kolorem niebieskim oznaczono potencjał Lennarda-Jonesa, a czerwonym potencjał twardych kul Mie.
- Rys. 13. Schematyczne przedstawienie przebiegu symulacji dynamiki molekularnej z wymianą replik (REMD)¹⁰⁰.
- Rys. 14. Schematyczne przedstawienie przebiegu symulacji dynamiki molekularnej ze zwielokrotnioną wymianą replik (MREMD)¹⁰⁰.

- Rys. 15. Schematyczne przedstawienie przebiegu symulacji dynamiki molekularnej z hamiltonowską wymianą replik (H-REMD)¹⁰⁰.
- Rys. 16. Schematyczne przedstawienie zasady działania metody najbliższego sąsiada (A) oraz wygenerowany dla niego dendrogram (B), w którym czerwoną linią oznaczono optymalną wartość odcięcia potrzebną do wygenerowania zadanej liczby skupień¹¹⁰.
- Rys. 17. Schematyczne przedstawienie sposobu próbkowania przestrzeni konfiguracyjnej układu przy przejściu między stanem początkowym A i końcowym B na hiperpowierzchni jego energii potencjalnej w metodzie *umbrella sampling*.
- Rys. 18. Schematyczne przedstawienie przykładowego zastosowania parametru λ w obliczeniach wartości PMF względem różnych odległości dwóch oddziałujących ze sobą cząstek.
- Rys. 19. Porównanie gęstości obliczonych dla trzech modeli wody: TIP3P-Ew (niebieska linia), SPC/E (czerwona linia) i TIP4P-2005 (zielona linia) z danymi eksperymentalnymi wyznaczonymi dla prawdziwej wody (czarna linia przerywana) w zakresie temperatur od 273 do 373 K przy ciśnieniu $p = 1$ atm.
- Rys. 20. Wykresy przedstawiające zakres zmian wartości ΔH (panel A), ΔG (panel B) oraz $T\Delta S$ (panel C) względem temperatury dla procesu hydratacji cząstki LJ-Xe symulowanej w modelu TIP3P-Ew, SPC/E oraz TIP4P-2005.
- Rys. 21. Wykresy przedstawiające zakres zmian wartości ΔH (panel A), ΔG (panel B) oraz $T\Delta S$ (panel C) względem temperatury dla procesu hydratacji pseudotwardej kuli Mie-Xe symulowanej w modelu TIP3P-Ew, SPC/E oraz TIP4P-2005.
- Rys. 22. Wykresy przedstawiające zależność zmian wartości potencjału średniej siły $\Delta G[r]$ względem odległości dwóch oddziałujących ze sobą cząstek LJ-Xe w zakresie temperatur od 273 do 373 K.
- Rys. 23. Wykresy przedstawiające zależność zmian wartości potencjału średniej siły $\Delta G[r]$ względem odległości dwóch oddziałujących ze sobą pseudotwardych kul Mie-Xe w zakresie temperatur od 273 do 373 K.
- Rys. 24. Wykresy przedstawiające zakres zmian wartości $\Delta H [cm]$ (panel A), $\Delta G [cm]$ (panel B) oraz $T\Delta S [cm]$ (panel C) względem temperatury w minimum kontaktowym dwóch cząstek LJ-Xe w modelu TIP3P-Ew, SPC/E oraz TIP4P-2005.
- Rys. 25. Wykresy przedstawiające zakres zmian wartości $\Delta H [cm]$ (panel A), $\Delta G [cm]$ (panel B) oraz $T\Delta S [cm]$ (panel C) względem temperatury w minimum kontaktowym dwóch pseudotwardych kul Mie-Xe w modelu TIP3P-Ew, SPC/E oraz TIP4P-2005.
- Rys. 26. Rozkłady gęstości, średniej orientacji przestrzennej, średniej liczby wiązań wodorowych oraz uporządkowania tetraedrycznego cząstek wody wokół pojedynczej cząstki LJ-Xe w $T = 273$ K i $T = 373$ K.
- Rys. 27. Rozkłady gęstości, średniej orientacji przestrzennej, średniej liczby wiązań wodorowych oraz uporządkowania tetraedrycznego wody wokół pseudotwardej kuli Mie-Xe w $T = 273$ K i $T = 373$ K.
- Rys. 28. Rozkłady gęstości, średniej orientacji przestrzennej, średniej liczby wiązań wodorowych oraz uporządkowania tetraedrycznego cząstek wody wokół dwóch cząstek LJ-Xe w $T = 273$ K i $T = 373$ K.

- Rys. 29. Rozkłady gęstości, średniej orientacji przestrzennej, średniej liczby wiązań wodorowych oraz uporządkowania tetraedrycznego cząstek wody wokół dwóch pseudotwardych kul Mie-Xe w $T = 273 \text{ K}$ i $T = 373 \text{ K}$.
- Rys. 30. Panel A przedstawia temperaturowe zależności zmian wartości ΔG , ΔH oraz $T\Delta S$ dla procesu hydratacji rosnącej cząstki Lennarda-Jonesa w modelu SPC/E. Na panelu B przedstawiono zależności ΔC_v oraz ΔS względem rozmiaru rosnącej cząstki Lennarda-Jonesa w modelu SPC/E dla zakresu temperatur od 273 do 373 K.
- Rys. 31. Panel A przedstawia temperaturowe zależności zmian wartości ΔG , ΔH oraz $T\Delta S$ dla procesu hydratacji rosnącej pseudo twardej kuli Mie w modelu SPC/E. Na panelu B przedstawiono zależności ΔC_v oraz ΔS względem rozmiaru rosnącej pseudo twardej kuli Mie w modelu SPC/E dla zakresu temperatur od 273 do 373 K.
- Rys. 32. Panel A przedstawia temperaturowe zależności zmian wartości ΔG , ΔH oraz $T\Delta S$ dla procesu hydratacji rosnącej cząstki Lennarda-Jonesa w modelu TIP4P-2005. Na panelu B przedstawiono zależności ΔC_v oraz ΔS względem rozmiaru rosnącej cząstki Lennarda-Jonesa w modelu TIP4P-2005 dla zakresu temperatur od 273 do 373 K.
- Rys. 33. Panel A przedstawia temperaturowe zależności zmian wartości ΔG , ΔH oraz $T\Delta S$ dla procesu hydratacji rosnącej pseudotwardej kuli Mie w modelu TIP4P-2005. Na panelu B przedstawiono zależności ΔC_v oraz ΔS względem rozmiaru rosnącej pseudotwardej kuli Mie w modelu TIP4P-2005 dla zakresu temperatur od 273 do 373 K.
- Rys. 34. Wykres temperaturowych zależności zmian wartości ΔG , ΔH oraz $T\Delta S$ dla cząstki Lennarda-Jonesa o średnicy σ_{X-Ow} równej 3.56, 5.96, 9.00 i 11.00 Å w modelu SPC/E (panel A). Panel B przedstawia powiększenie tych zależności dla $\sigma_{X-Ow} = 3.56$ i 5.96 Å. Panel C przedstawia zależność zmian wartości ΔS oraz ΔC_v względem średnicy rosnącej cząstki Lennarda-Jonesa w zakresie temperatur od 273 do 373 K.
- Rys. 35. Wykres temperaturowych zależności zmian wartości ΔG , ΔH oraz $T\Delta S$ dla pseudo twardej kuli Mie o średnicy σ_{X-Ow} równej 3.56, 5.96, 9.00 i 11.00 Å w modelu SPC/E (panel A). Panel B przedstawia powiększenie tych zależności dla $\sigma_{X-Ow} = 3.56$ i 5.96 Å. Panel C przedstawia zależność zmian wartości ΔS oraz ΔC_v względem średnicy rosnącej pseudotwardej kuli Mie w zakresie temperatur od 273 do 373 K.
- Rys. 36. Wykres temperaturowych zależności zmian wartości ΔG , ΔH oraz $T\Delta S$ dla cząstki Lennarda-Jonesa o średnicy σ_{X-Ow} równej 3.56, 5.96, 9.00 i 11.00 Å w modelu TIP4P-2005 (panel A). W panelu B przedstawiono powiększenie tych zależności dla $\sigma_{X-Ow} = 3.56$ i 5.96 Å. Panel C przedstawia zależność zmian wartości ΔS oraz ΔC_v względem średnicy rosnącej cząstki Lennarda-Jonesa w zakresie temperatur od 273 do 373 K.
- Rys. 37. Wykres temperaturowych zależności zmian wartości ΔG , ΔH oraz $T\Delta S$ dla pseudo twardej kuli Mie o średnicy σ_{X-Ow} równej 3.56, 5.96, 9.0 i 11 Å w modelu TIP4P-2005 (panel A). W panelu B przedstawiono powiększenie tych zależności dla $\sigma_{X-Ow} = 3.56$ i 5.96 Å. Panel C przedstawia zależność zmian wartości ΔS oraz ΔC_v względem średnicy rosnącej pseudotwardej kuli Mie w zakresie temperatur od 273 do 373 K.
- Rys. 38. Wykresy gęstości cząsteczek wody (panel A), ich kątowej orientacji przestrzennej (panel B), liczby tworzonych przez nie wiązań wodorowych (panel C) oraz ich uporządkowania tetraedrycznego (panel D) w pierwszej warstwie solwatacyjnej cząstki Lennarda-Jonesa o średnicy σ_{X-Ow} równej 3.08 oraz 11.00 Å dla modelu SPC/E w $T = 273 \text{ K}$ (niebieski) oraz $T = 373 \text{ K}$ (czerwony).

- Rys. 39. Wykresy gęstości cząsteczek wody (panel A), ich kątowej orientacji przestrzennej (panel B), liczby tworzonych przez nie wiązań wodorowych (panel C) oraz ich uporządkowania tetraedrycznego (panel D) w pierwszej warstwie solwatacyjnej pseudotwardej kuli Mie o średnicy σ_{X-Ow} równej 3.08 oraz 11.00 Å dla modelu SPC/E w T = 273 K (niebieski) oraz T = 373 K (czerwony).
- Rys. 40. Wykresy gęstości cząsteczek wody (panel A), ich kątowej orientacji przestrzennej (panel B), liczby tworzonych przez nie wiązań wodorowych (panel C) oraz ich uporządkowania tetraedrycznego (panel D) w pierwszej warstwie solwatacyjnej cząstki Lennarda-Jonesa o średnicy σ_{X-Ow} równej 3.08 oraz 11 Å dla modelu TIP4P-2005 w T = 273 K (niebieski) oraz T = 373 K (czerwony).
- Rys. 41. Wykresy gęstości cząsteczek wody (panel A), ich kątowej orientacji przestrzennej (panel B), liczby tworzonych przez nie wiązań wodorowych (panel C) oraz ich uporządkowania tetraedrycznego (panel D) w pierwszej warstwie solwatacyjnej pseudotwardej kuli Mie o średnicy σ_{X-Ow} równej 3.08 oraz 11 Å dla modelu TIP4P-2005 w T = 273 K (niebieski) oraz T = 373 K (czerwony).
- Rys. 42. Schematyczne przedstawienie pudła symulacyjnego użytego do obliczeń. Kolorem niebieskim oznaczono fazę wodną, żółtym fazę lipidową, a zielonym obszar przejściowy.
- Rys. 43. Wykres zmian wartości energii całkowitej układu dla współczynnika skalowania fazy lipidowej $s_L = 0.4$ (czerwony), $s_L = 1.0$ (niebieski) oraz $s_L = 2.0$ (czarny) względem czasu trwania symulacji w rozszerzonych do środowiska lipidowego polu sił UNRES.
- Rys. 44. Wykres przedstawiający rozkłady temperatur uzyskane w symulacjach MD z testowanym polem sił UNRES oraz termostatem Langevina (niebieska linia), Berendsena (czerwona linia) i Nose-Hoovera (czarna linia). Kolorem zielonym oznaczono teoretyczny rozkład temperatur Maxwella-Boltzmann
- Rys. 45. Zależność zmian średnich wartości GDT-TS oraz TM-score względem współczynnika skalowania s_L dla reprezentatywnych konformacji białek membranowych uzyskanych z symulacji MREMD z energią swobodną ΔG_{pi} równą 6.11 kcal/mol. Kolorem niebieskim oznaczono fazę wodną, żółtym lipidową, a zielonym dwuwarstwę lipidową.
- Rys. 46. Zależność zmian średnich wartości GDT-TS oraz TM-score względem współczynnika skalowania s_L dla reprezentatywnych konformacji białek membranowych uzyskanych z symulacji MREMD z energią swobodną ΔG_{pi} równą 0 kcal/mol. Kolorem niebieskim oznaczono fazę wodną, żółtym lipidową, a zielonym dwuwarstwę lipidową.
- Rys. 47. Wykresy słupkowe wartości RMSD, GDT-TS oraz TM-score dla struktur eksperymentalnych badanych białek oraz ich reprezentatywnych konformacji uzyskanych z symulacji MREMD dla $\Delta G_{pi} = 0$ kcal/mol w fazie wodnej (niebieski), lipidowej (żółty) oraz dwuwarstwie lipidowej skalowanej przez współczynnik $s_L = 0.7$ (zielony).
- Rys. 48. Struktura eksperymentalna białka 1IIJ (A), konformacja reprezentatywna białka 1IIJ w dwuwarstwie lipidowej skalowanego przez współczynnik $s_L = 0.7$ (B), fazie lipidowej (C) oraz środowisku wodnym (D). Nad każdą z symulowanych struktur białka 1IIJ umieszczono wartość RMSD liczoną względem struktury eksperymentalnej.
- Rys. 49. Wybrane fragmenty trajektorii białka o kodzie PDB: 1IIJ przedstawiające proces jego zwijania się w dwuwarstwie lipidowej skalowanej przez współczynnik $s_L = 0.7$. Trajektorja tego białka została uzyskana na drodze symulacji MREMD w testowanym polu sił UNRES z energią swobodną ΔG_{pi} równą 0 kcal/mol.

- Rys. 50. Zespoły konformacyjne struktur eksperymentalnych białek treningowych stosowanych na etapie kalibracji i optymalizacji pola sił UNRES.
- Rys. 51. Wykres zależności funkcji optymalizacji l^* od liczby iteracji w cyklu optymalizacyjnym pola sił UNRES rozszerzonego na 5 lokalnych typów reszt aminokwasowych.
- Rys. 52. Wykres zmian średnich wartości GDT-TS dla pierwszych i reprezentatywnych konformacji białek treningowych względem liczby iteracji w cyklu optymalizacyjnym pola sił UNRES.
- Rys. 53. Zestawienie średnich wartości GDT-TS dla pierwszych konformacji wszystkich typów białek testowych pochodzących ze skupień o najniższych średnich wartościach C^α RMSD względem ich struktur eksperymentalnych.
- Rys. 54. Zestawienie średnich wartości GDT-TS dla reprezentatywnych konformacji wszystkich typów białek testowych pochodzących ze skupień o najniższych średnich wartościach C^α RMSD względem ich struktur eksperymentalnych.
- Rys. 55. Porównanie najlepszej spośród pięciu przewidzianych w $T = 280$ K struktur białka T1016 (kolor szary) względem domeny A struktury eksperymentalnej białka o kodzie PDB 6E4B.
- Rys. 56. Rysunek przedstawiający nałożenie najlepszych jakościowo modeli serwerowych dla białka T1016. Kolorem fioletowym oznaczono odpowiadające im fragmenty konsensusowe.
- Rys. 57. Porównanie najlepszej spośród pięciu przewidzianych w $T = 280$ K struktur białka T1018 (kolor szary) względem domeny A struktury eksperymentalnej białka o kodzie PDB 6N91.
- Rys. 58. Rysunek przedstawiający nałożenie najlepszych jakościowo modeli serwerowych dla białka T1018. Kolorem fioletowym oznaczono odpowiadające im fragmenty konsensusowe.
- Rys. 59. Porównanie najlepszej spośród pięciu przewidzianych w $T = 300$ K struktur białka T1021s2 (kolor szary) względem struktury eksperymentalnej białka o kodzie PDB 6RAP.
- Rys. 60. Rysunek przedstawiający nałożenie najlepszych jakościowo modeli serwerowych dla białka T1021s2. Kolorem fioletowym oznaczono odpowiadające im fragmenty konsensusowe.

11. LITERATURA

- [1] Bairoch, A. The Universal Protein Resource (UniProt). *Nucleic Acids Res.* **2004**, *33* (Database issue), D154–D159.
- [2] Shaw, D. E.; Deneroff, M. M.; Dror, R. O.; Kuskin, J. S.; Larson, R. H.; Salmon, J. K.; Young, C.; Batson, B.; Bowers, K. J.; Chao, J. C.; Eastwood, M. P.; Gagliardo, J.; Grossman, J. P.; Ho, C. R.; Ierardi, D. J.; Kolossváry, I.; Klepeis, J. L.; Layman, T.; McLeavey, C.; Moraes, M. A.; Mueller, R.; Priest, E. C.; Shan, Y.; Spengler, J.; Theobald, M.; Towles, B.; Wang, S. C. Anton, a Special-Purpose Machine for Molecular Dynamics Simulation. *Commun. ACM* **2008**, *51* (7), 91–97.
- [3] Kessel, A.; Ben-Tal, N. *Introduction to Proteins: Structure, Function, and Motion*, Second edition.; Chapman & Hall/CRC mathematical and computational biology series; CRC Press, Taylor & Francis Group: Boca Raton, 2018.
- [4] Stoker, H. S. *General, Organic, and Biological Chemistry*, Sixth edition.; Brooks/Cole Cengage Learning: Belmont, CA, 2013.
- [5] Berg, J. M.; Tymoczko, J. L.; Stryer, L.; Clarke, N. D. *Biochemistry*, 5. ed., [Nachdr.], international ed.; Freeman: New York, 2003.
- [6] Voet, D.; Voet, J. G. *Biochemistry*, 4th ed.; John Wiley & Sons: Hoboken, NJ, 2011.
- [7] Garrett, R.; Grisham, C. M. *Biochemistry*, 4th ed.; Brooks/Cole, Cengage Learning: Belmont, CA, 2010.
- [8] Polypeptide Chain Configurations in Crystalline Proteins. *Proc. R. Soc. Lond. Ser. Math. Phys. Sci.* **1950**, *203* (1074), 321–357.
- [9] McMurry, J. *Organic Chemistry*, 8e ed.; Brooks/Cole, Cengage Learning: Belmont, CA, 2012.
- [10] Kabsch, W.; Sander, C. Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* **1983**, *22* (12), 2577–2637.
- [11] Shoulders, M. D.; Raines, R. T. Collagen Structure and Stability. *Annu. Rev. Biochem.* **2009**, *78* (1), 929–958.
- [12] Sewald, N.; Jakubke, H.-D. *Peptides: Chemistry and Biology*, 2., rev.updated ed.; Wiley-VCH: Weinheim, 2009.
- [13] Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. Stereochemistry of Polypeptide Chain Configurations. *J. Mol. Biol.* **1963**, *7* (1), 95–99.
- [14] Wang, Y.; Zhang, H.; Zhong, H.; Xue, Z. Protein Domain Identification Methods and Online Resources. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 1145–1153.
- [15] <https://www.Peptideweb.Com/Ramachandran-Plot-as-a-Tool-for-Peptide-and-Protein-Structures-Quality-Determination>.
- [16] Knapp, B.; Frantal, S.; Cibena, M.; Schreiner, W.; Bauer, P. Is an Intuitive Convergence Definition of Molecular Dynamics Simulations Solely Based on the Root Mean Square Deviation Possible? *J. Comput. Biol.* **2011**, *18* (8), 997–1005.
- [17] Alapati, R.; Shuvo, Md. H.; Bhattacharya, D. SPECS: Integration of Side-Chain Orientation and Global Distance-Based Measures for Improved Evaluation of Protein Structural Models. *PLOS ONE* **2020**, *15* (2), e0228245.
- [18] Mariani, V.; Biasini, M.; Barbato, A.; Schwede, T. LDDT: A Local Superposition-Free Score for Comparing Protein Structures and Models Using Distance Difference Tests. *Bioinformatics* **2013**, *29* (21), 2722–2728.

- [19] Modi, V.; Xu, Q.; Adhikari, S.; Dunbrack, R. L. Assessment of Template-Based Modeling of Protein Structure in CASP11: Template-Based Modeling Assessment in CASP11. *Proteins Struct. Funct. Bioinforma.* **2016**, *84*, 200–220.
- [20] Xu, J.; Zhang, Y. How Significant Is a Protein Structure Similarity with TM-Score = 0.5? *Bioinformatics* **2010**, *26* (7), 889–895.
- [21] Kufareva, I.; Abagyan, R. Methods of Protein Structure Comparison. In *Homology Modeling*; Orry, A. J. W., Abagyan, R., Eds.; Methods in Molecular Biology; Humana Press: Totowa, NJ, 2011; Vol. 857, pp 231–257.
- [22] Siew, N.; Elofsson, A.; Rychlewski, L.; Fischer, D. MaxSub: An Automated Measure for the Assessment of Protein Structure Prediction Quality. *Bioinformatics* **2000**, *16* (9), 776–785.
- [23] Levitt, M.; Gerstein, M. A Unified Statistical Framework for Sequence Comparison and Structure Comparison. *Proc. Natl. Acad. Sci.* **1998**, *95* (11), 5913–5920.
- [24] *Encyclopedia of Physical Science and Technology*, 3rd ed.; Meyers, R. A., Ed.; Academic Press: San Diego, 2002.
- [25] Pal, S. *Fundamentals of Molecular Structural Biology*; Academic Press, an imprint of Elsevier: London, United Kingdom ; San Diego, California, 2020.
- [26] Steiner, T. The Hydrogen Bond in the Solid State. *Angew. Chem. Int. Ed.* **2002**, *41* (1), 48–76.
- [27] Creighton, T. E. Disulfide Bond Formation in Proteins. In *Methods in Enzymology*; Elsevier, 1984; Vol. 107, pp 305–329.
- [28] *Novel Therapeutic Proteins: Selected Case Studies*; Dembowsky, K., Stadler, P., Eds.; Wiley-VCH: Weinheim ; New York, 2001.
- [29] Kunz, W.; Tsurko, J. *Thermodynamics of Amino Acid and Protein Solutions*; Transworld Research Network: Kerala, 2010.
- [30] Schauerperl, M.; Podewitz, M.; Waldner, B. J.; Liedl, K. R. Enthalpic and Entropic Contributions to Hydrophobicity. *J. Chem. Theory Comput.* **2016**, *12* (9), 4600–4610.
- [31] Archer, D. G.; Wang, P. The Dielectric Constant of Water and Debye-Hückel Limiting Law Slopes. *J. Phys. Chem. Ref. Data* **1990**, *19* (2), 371–411.
- [32] Frank, H. S.; Evans, M. W. Free Volume and Entropy in Condensed Systems III. Entropy in Binary Liquid Mixtures; Partial Molal Entropy in Dilute Solutions; Structure and Thermodynamics in Aqueous Electrolytes. *J. Chem. Phys.* **1945**, *13* (11), 507–532.
- [33] Chandler, D. Interfaces and the Driving Force of Hydrophobic Assembly. *Nature* **2005**, *437* (7059), 640–647.
- [34] Némethy, G.; Scheraga, H. A. Structure of Water and Hydrophobic Bonding in Proteins. I. A Model for the Thermodynamic Properties of Liquid Water. *J. Chem. Phys.* **1962**, *36* (12), 3382–3400.
- [35] Pratt, L. R.; Chandler, D. Theory of the Hydrophobic Effect. *J. Chem. Phys.* **1977**, *67* (8), 3683–3704.
- [36] Némethy, G.; Scheraga, H. A. THE STRUCTURE OF WATER AND HYDROPHOBIC BONDING IN PROTEINS. III. THE THERMODYNAMIC PROPERTIES OF HYDROPHOBIC BONDS IN PROTEINS ^{1,2}. *J. Phys. Chem.* **1962**, *66* (10), 1773–1789.

- [37] Dyson, H. J.; Wright, P. E.; Scheraga, H. A. The Role of Hydrophobic Interactions in Initiation and Propagation of Protein Folding. *Proc. Natl. Acad. Sci.* **2006**, *103* (35), 13057–13061.
- [38] Doyle, R.; Simons, K.; Qian, H.; Baker, D. Local Interactions and the Optimization of Protein Folding. *Proteins Struct. Funct. Genet.* **1997**, *29* (3), 282–291.
- [39] Liwo, A.; Sieradzan, A. K.; Lipska, A. G.; Czaplewski, C.; Joung, I.; Żmudzińska, W.; Hałabis, A.; Ołdziej, S. A General Method for the Derivation of the Functional Forms of the Effective Energy Terms in Coarse-Grained Energy Functions of Polymers. III. Determination of Scale-Consistent Backbone-Local and Correlation Potentials in the UNRES Force Field and Force-Field Calibration and Validation. *J. Chem. Phys.* **2019**, *150* (15), 155104.
- [40] Kumar, A.; Baruah, A.; Biswas, P. Role of Local and Nonlocal Interactions in Folding and Misfolding of Globular Proteins. *J. Chem. Phys.* **2017**, *146* (6), 065102.
- [41] Muñoz, V.; Serrano, L. Local versus Nonlocal Interactions in Protein Folding and Stability – an Experimentalist’s Point of View. *Fold. Des.* **1996**, *1* (4), R71–R77.
- [42] Kolos, W. *Chemia kwantowa*; Państwowe Wydaw. Naukowe: Warszawa, 1986.
- [43] *Encyclopedia of Inorganic Chemistry*, 2nd ed.; King, R. B., Ed.; Wiley: Chichester, West Sussex, England; Hoboken, NJ, 2005.
- [44] Sadlej, J. *Obliczeniowe metody chemii kwantowej CNDO, INDO, NDDO, ab initio*; Państw. Wydaw. Naukowe: Warszawa, 1988.
- [45] Piela, L.; Wydawnictwo Naukowe PWN. *Idee chemii kwantowej*; Wydawnictwo Naukowe PWN: Warszawa, 2021.
- [46] Born, M.; Oppenheimer, R. Zur Quantentheorie der Molekeln. *Ann. Phys.* **1927**, *389* (20), 457–484.
- [47] Roothaan, C. C. J. Self-Consistent Field Theory for Open Shells of Electronic Systems. *Rev. Mod. Phys.* **1960**, *32* (2), 179–185.
- [48] Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; Dover Publications: Mineola, N.Y., 1996.
- [49] Christensen, A. S.; Kubař, T.; Cui, Q.; Elstner, M. Semiempirical Quantum Mechanical Methods for Noncovalent Interactions for Chemical and Biochemical Applications. *Chem. Rev.* **2016**, *116* (9), 5301–5337.
- [50] Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods VI: More Modifications to the NDDO Approximations and Re-Optimization of Parameters. *J. Mol. Model.* **2013**, *19* (1), 1–32.
- [51] Bikadi, Z.; Hazai, E. Application of the PM6 Semi-Empirical Method to Modeling Proteins Enhances Docking Accuracy of AutoDock. *J. Cheminformatics* **2009**, *1* (1), 15.
- [52] Vanommeslaeghe, K.; Guvench, O.; MacKerell, A. D. Molecular Mechanics. *Curr. Pharm. Des.* **2014**, *20* (20), 3281–3292.
- [53] González, M. A. Force Fields and Molecular Dynamics Simulations. *Éc. Thématique Société Fr. Neutron.* **2011**, *12*, 169–200.
- [54] Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117* (19), 5179–5197.

- [55] Zhu, X.; Lopes, P. E. M.; MacKerell, A. D. Recent Developments and Applications of the CHARMM Force Fields. *WIREs Comput. Mol. Sci.* **2012**, *2* (1), 167–185.
- [56] Liu, Z.; Wu, X.; Wang, W. A Novel United-Atom Force Field for Imidazolium-Based Ionic Liquids. *Phys. Chem. Chem. Phys.* **2006**, *8* (9), 1096.
- [57] Sieradzan, A. K.; Czaplewski, C.; Krupa, P.; Mozolewska, M. A.; Karczyńska, A. S.; Lipska, A. G.; Lubecka, E. A.; Gołaś, E.; Wirecki, T.; Makowski, M.; Ołdziej, S.; Liwo, A. Modeling the Structure, Dynamics, and Transformations of Proteins with the UNRES Force Field. In *Protein Folding*; Muñoz, V., Ed.; Methods in Molecular Biology; Springer US: New York, NY, 2022; Vol. 2376, pp 399–416.
- [58] Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **2007**, *111* (27), 7812–7824.
- [59] Kolinski, A. Protein Modeling and Structure Prediction with a Reduced Representation. *Acta Biochim. Polym.* **2004**, 349–371.
- [60] *Molecular Modeling: Basic Principles and Applications*, 2nd ed.; Höltje, H.-D., Ed.; Methods and principles in medicinal chemistry; Wiley-VCH: Weinheim, 2003.
- [61] *Computational Pharmaceutics: Application of Molecular Modeling in Drug Delivery*; Ouyang, D., Smith, S. C., Eds.; Advances in pharmaceutical technology; John Wiley & Sons Ltd: Chichester, West Sussex, United Kingdom; Hoboken, 2015.
- [62] On the Determination of Molecular Fields. —II. From the Equation of State of a Gas. *Proc. R. Soc. Lond. Ser. Contain. Pap. Math. Phys. Character* **1924**, *106* (738), 463–477.
- [63] Izadi, S.; Anandakrishnan, R.; Onufriev, A. V. Building Water Models: A Different Approach. *J. Phys. Chem. Lett.* **2014**, *5* (21), 3863–3871.
- [64] Vega, C.; de Miguel, E. Surface Tension of the Most Popular Models of Water by Using the Test-Area Simulation Method. *J. Chem. Phys.* **2007**, *126* (15), 154707.
- [65] Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91* (24), 6269–6271.
- [66] Price, D. J.; Brooks, C. L. A Modified TIP3P Water Potential for Simulation with Ewald Summation. *J. Chem. Phys.* **2004**, *121* (20), 10096–10103.
- [67] Abascal, J. L. F.; Vega, C. A General Purpose Model for the Condensed Phases of Water: TIP4P/2005. *J. Chem. Phys.* **2005**, *123* (23), 234505.
- [68] Liwo, A.; Czaplewski, C. Extension of the Force-Matching Method to Coarse-Grained Models with Axially Symmetric Sites to Produce Transferable Force Fields: Application to the UNRES Model of Proteins. *J. Chem. Phys.* **2020**, *152* (5), 054902.
- [69] Krupa, P.; Hałabis, A.; Żmudzińska, W.; Ołdziej, S.; Scheraga, H. A.; Liwo, A. Maximum Likelihood Calibration of the UNRES Force Field for Simulation of Protein Structure and Dynamics. *J. Chem. Inf. Model.* **2017**, *57* (9), 2364–2377.
- [70] Krupa, P.; Sieradzan, A. K.; Rackovsky, S.; Baranowski, M.; Ołdziej, S.; Scheraga, H. A.; Liwo, A.; Czaplewski, C. Improvement of the Treatment of Loop Structures in the UNRES Force Field by Inclusion of Coupling between Backbone- and Side-Chain-Local Conformational States. *J. Chem. Theory Comput.* **2013**, *9* (10), 4620–4632.

- [71] Rojas, A.; Liwo, A.; Browne, D.; Scheraga, H. A. Mechanism of Fiber Assembly: Treatment of A β Peptide Aggregation with a Coarse-Grained United-Residue Force Field. *J. Mol. Biol.* **2010**, *404* (3), 537–552.
- [72] Maisuradze, G. G.; Liwo, A.; Scheraga, H. A. Relation between Free Energy Landscapes of Proteins and Dynamics. *J. Chem. Theory Comput.* **2010**, *6* (2), 583–595.
- [73] Karczyńska, A.; Mozolewska, M. A.; Krupa, P.; Gieldoń, A.; Bojarski, K. K.; Zaborowski, B.; Liwo, A.; Ślusarz, R.; Ślusarz, M.; Lee, J.; Joo, K.; Czaplewski, C. Use of the UNRES Force Field in Template-Assisted Prediction of Protein Structures and the Refinement of Server Models: Test with CASP12 Targets. *J. Mol. Graph. Model.* **2018**, *83*, 92–99.
- [74] He, Y.; Liwo, A.; Weinstein, H.; Scheraga, H. A. PDZ Binding to the BAR Domain of PICK1 Is Elucidated by Coarse-Grained Molecular Dynamics. *J. Mol. Biol.* **2011**, *405* (1), 298–314.
- [75] Gołaś, E.; Maisuradze, G. G.; Senet, P.; Ołdziej, S.; Czaplewski, C.; Scheraga, H. A.; Liwo, A. Simulation of the Opening and Closing of Hsp70 Chaperones by Coarse-Grained Molecular Dynamics. *J. Chem. Theory Comput.* **2012**, *8* (5), 1750–1764.
- [76] Zięba, K.; Ślusarz, M.; Ślusarz, R.; Liwo, A.; Czaplewski, C.; Sieradzan, A. K. Extension of the UNRES Coarse-Grained Force Field to Membrane Proteins in the Lipid Bilayer. *J. Phys. Chem. B* **2019**, *123* (37), 7829–7839.
- [77] Fukunishi, H.; Watanabe, O.; Takada, S. On the Hamiltonian Replica Exchange Method for Efficient Sampling of Biomolecular Systems: Application to Protein Structure Prediction. *J. Chem. Phys.* **2002**, *116* (20), 9058–9067.
- [78] Chodera, J. D.; Swope, W. C.; Pitera, J. W.; Seok, C.; Dill, K. A. Use of the Weighted Histogram Analysis Method for the Analysis of Simulated and Parallel Tempering Simulations. *J. Chem. Theory Comput.* **2007**, *3* (1), 26–41.
- [79] Zhang, C.; Lai, C.-L.; Pettitt, B. M. Accelerating the Weighted Histogram Analysis Method by Direct Inversion in the Iterative Subspace. *Mol. Simul.* **2016**, *42* (13), 1079–1089.
- [80] Zaborowski, B.; Jagieła, D.; Czaplewski, C.; Hałabis, A.; Lewandowska, A.; Żmudzińska, W.; Ołdziej, S.; Karczyńska, A.; Omieczynski, C.; Wirecki, T.; Liwo, A. A Maximum-Likelihood Approach to Force-Field Calibration. *J. Chem. Inf. Model.* **2015**, *55* (9), 2050–2070.
- [81] Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*, Second edition.; Oxford University Press: Oxford, United Kingdom, 2017.
- [82] Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed.; Computational science series; Academic Press: San Diego, 2002.
- [83] Chodera, J. D.; Swope, W. C.; Pitera, J. W.; Dill, K. A. Long-Time Protein Folding Dynamics from Short-Time Molecular Dynamics Simulations. *Multiscale Model. Simul.* **2006**, *5* (4), 1214–1226.
- [84] Śledź, P.; Caflisch, A. Protein Structure-Based Drug Design: From Docking to Molecular Dynamics. *Curr. Opin. Struct. Biol.* **2018**, *48*, 93–102.
- [85] Fatemi, S. M.; Foroutan, M. Recent Developments Concerning the Dispersion of Carbon Nanotubes in Surfactant/Polymer Systems by MD Simulation. *J. Nanostructure Chem.* **2016**, *6* (1), 29–40.
- [86] Nieszporek, K.; Nieszporek, J.; Trojak, M. Calculations of Shear Viscosity, Electric Conductivity and Diffusion Coefficients of Aqueous Sodium Perchlorate Solutions from Molecular Dynamics Simulations. *Comput. Theor. Chem.* **2016**, *1090*, 52–57.

- [87] Liwo, A.; Khalili, M.; Scheraga, H. A. *Ab Initio* Simulations of Protein-Folding Pathways by Molecular Dynamics with the United-Residue Model of Polypeptide Chains. *Proc. Natl. Acad. Sci.* **2005**, *102* (7), 2362–2367.
- [88] Feenstra, K. A.; Hess, B.; Berendsen, H. J. C. Improving Efficiency of Large Time-Scale Molecular Dynamics Simulations of Hydrogen-Rich Systems. *J. Comput. Chem.* **1999**, *20* (8), 786–798.
- [89] Khalili, M.; Liwo, A.; Rakowski, F.; Grochowski, P.; Scheraga, H. A. Molecular Dynamics with the United-Residue Model of Polypeptide Chains. I. Lagrange Equations of Motion and Tests of Numerical Stability in the Microcanonical Mode. *J. Phys. Chem. B* **2005**, *109* (28), 13785–13797.
- [90] Hünenberger, P. H. Thermostat Algorithms for Molecular Dynamics Simulations. In *Advanced Computer Simulation*; Dr. Holm, C., Prof. Dr. Kremer, K., Eds.; Abe, A., Joanny, J.-F., Albertsson, A.-C., Duncan, R., Kausch, H.-H., Kobayashi, S., Dušek, K., Lee, K.-S., de Jeu, W. H., Leibler, L., Nuyken, O., Long, T. E., Terentjev, E. M., Voit, B., Manners, I., Wegner, G., Möller, M., Series Eds.; Advances in Polymer Science; Springer Berlin Heidelberg: Berlin, Heidelberg, 2005; Vol. 173, pp 105–149.
- [91] Morishita, T. Fluctuation Formulas in Molecular-Dynamics Simulations with the Weak Coupling Heat Bath. *J. Chem. Phys.* **2000**, *113* (8), 2976–2982.
- [92] Grest, G. S.; Kremer, K. Molecular Dynamics Simulation for Polymers in the Presence of a Heat Bath. *Phys. Rev. A* **1986**, *33* (5), 3628–3631.
- [93] Khalili, M.; Liwo, A.; Jagielska, A.; Scheraga, H. A. Molecular Dynamics with the United-Residue Model of Polypeptide Chains. II. Langevin and Berendsen-Bath Dynamics and Tests on Model α -Helical Systems. *J. Phys. Chem. B* **2005**, *109* (28), 13798–13810.
- [94] Duh, D.-M.; Henderson, D.; Rowley, R. L. Some Effects of Deviations from the Lorentz-Berthelot Combining Rules for Mixtures of Lennard-Jones Fluids. *Mol. Phys.* **1997**, *91* (6), 1143–1147.
- [95] Jover, J.; Haslam, A. J.; Galindo, A.; Jackson, G.; Müller, E. A. Pseudo Hard-Sphere Potential for Use in Continuous Molecular-Dynamics Simulation of Spherical and Chain Molecules. *J. Chem. Phys.* **2012**, *137* (14), 144505.
- [96] Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314* (1–2), 141–151.
- [97] Zhang, W.; Wu, C.; Duan, Y. Convergence of Replica Exchange Molecular Dynamics. *J. Chem. Phys.* **2005**, *123* (15), 154105.
- [98] Hastings, W. K. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika* **1970**, *57* (1), 97–109.
- [99] Karczyńska, A. S.; Czaplewski, C.; Krupa, P.; Mozolewska, M. A.; Joo, K.; Lee, J.; Liwo, A. Ergodicity and Model Quality in Template-Restrained Canonical and Temperature/Hamiltonian Replica Exchange Coarse-Grained Molecular Dynamics Simulations of Proteins. *J. Comput. Chem.* **2017**, *38* (31), 2730–2746.
- [100] Bernardi, R. C.; Melo, M. C. R.; Schulten, K. Enhanced Sampling Techniques in Molecular Dynamics Simulations of Biological Systems. *Biochim. Biophys. Acta BBA - Gen. Subj.* **2015**, *1850* (5), 872–877.
- [101] Rhee, Y. M.; Pande, V. S. Multiplexed-Replica Exchange Molecular Dynamics Method for Protein Folding Simulation. *Biophys. J.* **2003**, *84* (2), 775–786.

- [102] Ferrenberg, A. M.; Swendsen, R. H. Optimized Monte Carlo Data Analysis. *Comput. Phys.* **1989**, *3* (5), 101.
- [103] Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. THE Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13* (8), 1011–1021.
- [104] Liwo, A.; Khalili, M.; Czaplewski, C.; Kalinowski, S.; Ołdziej, S.; Wachucik, K.; Scheraga, H. A. Modification and Optimization of the United-Residue (UNRES) Potential Energy Function for Canonical Simulations. I. Temperature Dependence of the Effective Energy Function and Tests of the Optimization Method with Single Training Proteins. *J. Phys. Chem. B* **2007**, *111* (1), 260–285.
- [105] *Cluster Analysis*, 5th ed.; Everitt, B., Ed.; Wiley series in probability and statistics; Wiley: Chichester, West Sussex, U.K, 2011.
- [106] Peng, J.; Wang, W.; Yu, Y.; Gu, H.; Huang, X. Clustering Algorithms to Analyze Molecular Dynamics Simulation Trajectories for Complex Chemical and Biological Systems. *Chin. J. Chem. Phys.* **2018**, *31* (4), 404–420.
- [107] Sibson, R. SLINK: An Optimally Efficient Algorithm for the Single-Link Cluster Method. *Comput. J.* **1973**, *16* (1), 30–34.
- [108] Ward, J. H. Hierarchical Grouping to Optimize an Objective Function. *J. Am. Stat. Assoc.* **1963**, *58* (301), 236–244.
- [109] Peña, J. M.; Lozano, J. A.; Larrañaga, P. An Empirical Comparison of Four Initialization Methods for the K-Means Algorithm. *Pattern Recognit. Lett.* **1999**, *20* (10), 1027–1040.
- [110] [https://Towardsdatascience.Com/Hierarchical-Clustering-Explained-E59b13846da8](https://Towardsdatascience.com/Hierarchical-Clustering-Explained-E59b13846da8).
- [111] Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23* (2), 187–199.
- [112] *Free Energy Calculations: Theory and Applications in Chemistry and Biology*, Study ed.; Chipot, C., Pohorille, A., Eds.; Springer series in chemical physics; Springer: New York, 2007.
- [113] Jorgensen, W. L.; Ravimohan, C. Monte Carlo Simulation of Differences in Free Energies of Hydration. *J. Chem. Phys.* **1985**, *83* (6), 3050–3054.
- [114] Abel, R.; Wang, L.; Harder, E. D.; Berne, B. J.; Friesner, R. A. Advancing Drug Discovery through Enhanced Free Energy Calculations. *Acc. Chem. Res.* **2017**, *50* (7), 1625–1632.
- [115] Widom, B. Potential-Distribution Theory and the Statistical Mechanics of Fluids. *J. Phys. Chem.* **1982**, *86* (6), 869–872.
- [116] Tomás-Oliveira, I.; Wodak, S. J. Thermodynamics of Cavity Formation in Water and *n*-Hexane Using the Widom Particle Insertion Method. *J. Chem. Phys.* **1999**, *111* (18), 8576–8587.
- [117] Berman, H. M. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235–242.
- [118] Boeckmann, B. The SWISS-PROT Protein Knowledgebase and Its Supplement TrEMBL in 2003. *Nucleic Acids Res.* **2003**, *31* (1), 365–370.
- [119] Ladd, M. F. c. *Structure Determination by X-Ray Crystallography: Analysis by x-Rays and Neutrons*; Springer: New York, 2012.
- [120] Wüthrich, K. Protein Structure Determination in Solution by Nuclear Magnetic Resonance Spectroscopy. *Science* **1989**, *243* (4887), 45–50.

- [121] Nogales, E. The Development of Cryo-EM into a Mainstream Structural Biology Technique. *Nat. Methods* **2016**, *13* (1), 24–27.
- [122] Anfinsen, C. B. Principles That Govern the Folding of Protein Chains. *Science* **1973**, *181* (4096), 223–230.
- [123] Anfinsen, C. B.; Scheraga, H. A. Experimental and Theoretical Aspects of Protein Folding. In *Advances in Protein Chemistry*; Elsevier, 1975; Vol. 29, pp 205–300.
- [124] Mozolewska, M. A.; Krupa, P.; Zaborowski, B.; Liwo, A.; Lee, J.; Joo, K.; Czaplewski, C. Use of Restraints from Consensus Fragments of Multiple Server Models To Enhance Protein-Structure Prediction Capability of the UNRES Force Field. *J. Chem. Inf. Model.* **2016**, *56* (11), 2263–2279.
- [125] Karczyńska, A. S.; Zięba, K.; Uciechowska, U.; Mozolewska, M. A.; Krupa, P.; Lubecka, E. A.; Lipska, A. G.; Sikorska, C.; Samsonov, S. A.; Sieradzan, A. K.; Giełdoń, A.; Liwo, A.; Ślusarz, R.; Ślusarz, M.; Lee, J.; Joo, K.; Czaplewski, C. Improved Consensus-Fragment Selection in Template-Assisted Prediction of Protein Structures with the UNRES Force Field in CASP13. *J. Chem. Inf. Model.* **2020**, *60* (3), 1844–1864.
- [126] Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589.
- [127] *Computational Biology*; Fenyó, D., Ed.; Springer protocols; Humana Press: New York, NY, 2010.
- [128] McGuffin, L. J.; Bryson, K.; Jones, D. T. The PSIPRED Protein Structure Prediction Server. *Bioinformatics* **2000**, *16* (4), 404–405.
- [129] Altschul, S. Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs. *Nucleic Acids Res.* **1997**, *25* (17), 3389–3402.
- [130] Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. PROCHECK: A Program to Check the Stereochemical Quality of Protein Structures. *J. Appl. Crystallogr.* **1993**, *26* (2), 283–291.
- [131] Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. The I-TASSER Suite: Protein Structure and Function Prediction. *Nat. Methods* **2015**, *12* (1), 7–8.
- [132] Källberg, M.; Wang, H.; Wang, S.; Peng, J.; Wang, Z.; Lu, H.; Xu, J. Template-Based Protein Structure Modeling Using the RaptorX Web Server. *Nat. Protoc.* **2012**, *7* (8), 1511–1522.
- [133] Soding, J.; Biegert, A.; Lupas, A. N. The HHpred Interactive Server for Protein Homology Detection and Structure Prediction. *Nucleic Acids Res.* **2005**, *33* (Web Server), W244–W248.
- [134] Pan, X.; Kortemme, T. Recent Advances in de Novo Protein Design: Principles, Methods, and Applications. *J. Biol. Chem.* **2021**, *296*, 100558.
- [135] Huang, P.-S.; Boyken, S. E.; Baker, D. The Coming of Age of de Novo Protein Design. *Nature* **2016**, *537* (7620), 320–327.
- [136] Barbu, A. G.; Zhu, S.-C. *Monte Carlo Methods*; Springer: Singapore, 2020.

- [137] Maisuradze, G. G.; Senet, P.; Czaplewski, C.; Liwo, A.; Scheraga, H. A. Investigation of Protein Folding by Coarse-Grained Molecular Dynamics with the UNRES Force Field. *J. Phys. Chem. A* **2010**, *114* (13), 4471–4485.
- [138] Cao, R.; Bhattacharya, D.; Hou, J.; Cheng, J. DeepQA: Improving the Estimation of Single Protein Model Quality with Deep Belief Networks. *BMC Bioinformatics* **2016**, *17* (1), 495.
- [139] Sasaki, T. N.; Sasai, M. A Coarse-Grained Langevin Molecular Dynamics Approach to Protein Structure Reproduction. *Chem. Phys. Lett.* **2005**, *402* (1–3), 102–106.
- [140] Sasaki, T. N.; Cetin, H.; Sasai, M. A Coarse-Grained Langevin Molecular Dynamics Approach to de Novo Protein Structure Prediction. *Biochem. Biophys. Res. Commun.* **2008**, *369* (2), 500–506.
- [141] Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696–3713.
- [142] Rotkiewicz, P.; Skolnick, J. Fast Procedure for Reconstruction of Full-Atom Protein Models from Reduced Representations. *J. Comput. Chem.* **2008**, *29* (9), 1460–1465.
- [143] Wang, Q.; Canutescu, A. A.; Dunbrack, R. L. SCWRL and MolIDE: Computer Programs for Side-Chain Conformation Prediction and Homology Modeling. *Nat. Protoc.* **2008**, *3* (12), 1832–1847.
- [144] Kryshtafovych, A.; Schwede, T.; Topf, M.; Fidelis, K.; Moult, J. Critical Assessment of Methods of Protein Structure Prediction (CASP)—Round XIII. *Proteins Struct. Funct. Bioinforma.* **2019**, *87* (12), 1011–1020.
- [145] Janin, J.; Henrick, K.; Moult, J.; Eyck, L. T.; Sternberg, M. J. E.; Vajda, S.; Vakser, I.; Wodak, S. J. CAPRI: A Critical Assessment of PRedicted Interactions. *Proteins Struct. Funct. Genet.* **2003**, *52* (1), 2–9.
- [146] Lubecka, E. A.; Karczyńska, A. S.; Lipska, A. G.; Sieradzan, A. K.; Zięba, K.; Sikorska, C.; Uciechowska, U.; Samsonov, S. A.; Krupa, P.; Mozolewska, M. A.; Golon, Ł.; Giełdoń, A.; Czaplewski, C.; Ślusarz, R.; Ślusarz, M.; Crivelli, S. N.; Liwo, A. Evaluation of the Scale-Consistent UNRES Force Field in Template-Free Prediction of Protein Structures in the CASP13 Experiment. *J. Mol. Graph. Model.* **2019**, *92*, 154–166.
- [147] Antoniak, A.; Biskupek, I.; Bojarski, K. K.; Czaplewski, C.; Giełdoń, A.; Kogut, M.; Kogut, M. M.; Krupa, P.; Lipska, A. G.; Liwo, A.; Lubecka, E. A.; Marcisz, M.; Maszota-Zieleniak, M.; Samsonov, S. A.; Sieradzan, A. K.; Ślusarz, M. J.; Ślusarz, R.; Wesołowski, P. A.; Zięba, K. Modeling Protein Structures with the Coarse-Grained UNRES Force Field in the CASP14 Experiment. *J. Mol. Graph. Model.* **2021**, *108*, 108008.

- [148] Lensink, M. F.; Brysbaert, G.; Mauri, T.; Nadzirin, N.; Velankar, S.; Chaleil, R. A. G.; Clarence, T.; Bates, P. A.; Kong, R.; Liu, B.; Yang, G.; Liu, M.; Shi, H.; Lu, X.; Chang, S.; Roy, R. S.; Quadir, F.; Liu, J.; Cheng, J.; Antoniak, A.; Czaplewski, C.; Gieldoń, A.; Kogut, M.; Lipska, A. G.; Liwo, A.; Lubecka, E. A.; Maszota-Zieleniak, M.; Sieradzan, A. K.; Ślusarz, R.; Wesółowski, P. A.; Zięba, K.; Del Carpio Muñoz, C. A.; Ichiishi, E.; Harmalkar, A.; Gray, J. J.; Bonvin, A. M. J. J.; Ambrosetti, F.; Vargas Honorato, R.; Jandova, Z.; Jiménez-García, B.; Koukos, P. I.; Van Keulen, S.; Van Noort, C. W.; Réau, M.; Roel-Touris, J.; Kotelnikov, S.; Padhorny, D.; Porter, K. A.; Alekseenko, A.; Ignatov, M.; Desta, I.; Ashizawa, R.; Sun, Z.; Ghani, U.; Hashemi, N.; Vajda, S.; Kozakov, D.; Rosell, M.; Rodríguez-Lumbreras, L. A.; Fernandez-Recio, J.; Karczynska, A.; Grudin, S.; Yan, Y.; Li, H.; Lin, P.; Huang, S.; Christoffer, C.; Terashi, G.; Verburgt, J.; Sarkar, D.; Aderinwale, T.; Wang, X.; Kihara, D.; Nakamura, T.; Hanazono, Y.; Gowthaman, R.; Guest, J. D.; Yin, R.; Taherzadeh, G.; Pierce, B. G.; Barradas-Bautista, D.; Cao, Z.; Cavallo, L.; Oliva, R.; Sun, Y.; Zhu, S.; Shen, Y.; Park, T.; Woo, H.; Yang, J.; Kwon, S.; Won, J.; Seok, C.; Kiyota, Y.; Kobayashi, S.; Harada, Y.; Takeda-Shitaka, M.; Kundrotas, P. J.; Singh, A.; Vakser, I. A.; Dapkūnas, J.; Olechnovič, K.; Venclovas, Č.; Duan, R.; Qiu, L.; Xu, X.; Zhang, S.; Zou, X.; Wodak, S. J. Prediction of Protein Assemblies, the next Frontier: The CASP14-CAPRI Experiment. *Proteins Struct. Funct. Bioinforma.* **2021**, 89 (12), 1800–1823.
- [149] Czaplewski, C.; Kalinowski, S.; Liwo, A.; Scheraga *, H. A. Comparison of Two Approaches to Potential of Mean Force Calculations of Hydrophobic Association: Particle Insertion and Weighted Histogram Analysis Methods. *Mol. Phys.* **2005**, 103 (21–23), 3153–3167.
- [150] Brańka, A. C.; Kowalik, M.; Wojciechowski, K. W. Generalization of the Nosé–Hoover Approach. *J. Chem. Phys.* **2003**, 119 (4), 1929–1936.
- [151] Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, 23 (3), 327–341.
- [152] Errington, J. R.; Debenedetti, P. G. Relationship between Structural Order and the Anomalies of Liquid Water. *Nature* **2001**, 409 (6818), 318–321.
- [153] Graziano, G. Shedding Light on the Hydrophobicity Puzzle. *Pure Appl. Chem.* **2016**, 88 (3), 177–188.
- [154] Pereda, S.; Awan, J. A.; Mohammadi, A. H.; Valtz, A.; Coquelet, C.; Brignole, E. A.; Richon, D. Solubility of Hydrocarbons in Water: Experimental Measurements and Modeling Using a Group Contribution with Association Equation of State (GCA-EoS). *Fluid Phase Equilibria* **2009**, 275 (1), 52–59.
- [155] Liwo, A.; O’dziej, S.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scheraga, H. A. A United-Residue Force Field for off-Lattice Protein-Structure Simulations. I. Functional Forms and Parameters of Long-Range Side-Chain Interaction Potentials from Protein Crystal Data. *J. Comput. Chem.* **1997**, 18 (7), 849–873.
- [156] Levitt, M. A Simplified Representation of Protein Conformations for Rapid Simulation of Protein Folding. *J. Mol. Biol.* **1976**, 104 (1), 59–107.
- [157] Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E.; Wang, J.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Kollman, P. A. *AMBER14*; University of California: San Francisco, 2014.
- [158] Fauchere, J. L.; Pliska, V. Hydrophobic Parameters II of Amino-Acid Side Chains from the Partitioning of N-Acetyl-Amino-Acid Amides. *Eur. J. Med. Chem.* **1983**, 18 (4), 369–375.

- [159] Solis, A. D.; Rackovsky, S. Optimized Representations and Maximal Information in Proteins. *Proteins* **2000**, *38* (2), 149–164.
- [160] Eastwood, M. P.; Hardin, C.; Luthey-Schulten, Z.; Wolynes, P. G. Statistical Mechanical Refinement of Protein Structure Prediction Schemes: Cumulant Expansion Approach. *J. Chem. Phys.* **2002**, *117* (9), 4602–4615.
- [161] Rakowski, F.; Grochowski, P.; Lesyng, B.; Liwo, A.; Scheraga, H. A. Implementation of a Symplectic Multiple-Time-Step Molecular Dynamics Algorithm, Based on the United-Residue Mesoscopic Potential Energy Function. *J. Chem. Phys.* **2006**, *125* (20), 204107.