



AKADEMIA GÓRNICZO-HUTNICZA IM. STANISŁAWA STASZICA W KRAKOWIE

DZIEDZINA biofizyka

DYSCYPLINA nauki fizyczne

ROZPRAWA DOKTORSKA

Wielowymiarowa analiza modulowanych struktur makromolekuł z wykorzystaniem nowatorskich korekt fononowych na przykładzie kompleksu Hyp-1/ANS

Autor: mgr Joanna Śmietańska

Promotor rozprawy: prof. dr hab. Janusz Wolny
Drugi promotor: prof. dr hab. Mariusz Jaskólski

Praca wykonana w: AGH, Wydział Fizyki i Informatyki Stosowanej, Katedra Fizyki
Materii Skondensowanej

Oświadczenie autora rozprawy:

Oświadczam, świadoma odpowiedzialności karnej za poświadczenie nieprawdy, że niniejszą pracę doktorską wykonałam osobiście i samodzielnie i że nie korzystałam ze źródeł innych niż wymienione w pracy.

data, podpis autora

Oświadczenie promotorów rozprawy:

Niniejsza rozprawa jest gotowa do oceny przez recenzentów.

data, podpis promotora rozprawy

data, podpis promotora rozprawy

Podziękowania

Składam ogromne podziękowania dla moich Promotorów, prof. dr hab. Janusza Wolnego oraz prof. dr hab. Mariusza Jaskólskiego, za niezwykle cenną wiedzę, wsparcie, a także motywację, którą przekazywali mi podczas całego procesu pisania pracy doktorskiej.

Chciałabym również podziękować dr Radosławowi Strzałce i dr Ireneuszowi Bugańskiemu za miłą współpracę, wspierającą atmosferę i ciekawe dyskusje na temat moich badań.

Pragnę wyrazić moje podziękowania dr Joannie Śliwiak i dr. hab. Mirosławowi Gilskiemu za udostępnienie danych eksperymentalnych, bezcenne wskazówki oraz wsparcie i pomoc na każdym etapie badań.

Nie mogę też zapomnieć o doktorantach z Katedry Fizyki Materii Skondensowanej, którzy byli moimi kolegami i pomogli mi w wielu aspektach mojej pracy doktorskiej. Ich wkład i współpraca były nieocenione.

Ostatecznie, składam szczególne podziękowania dla moich bliskich: Rodziców, Brata i Przyjaciół, którzy zawsze byli przy mnie i dawali mi wsparcie w trudnych momentach. Bez Waszej pomocy, rady i wsparcia nie byłoby mnie tam, gdzie jestem dzisiaj.

Pracę dedykuję Mateuszowi

Niniejsza rozprawa doktorska została wykonana w ramach grantu NCN PRELUDIUM nr 2020/37/N/ST3/01501 oraz Programu Operacyjnego Wiedza Edukacja Rozwój, nr projektu POWR.03.02.00-00-I004/16, współfinansowanego ze środków Unii Europejskiej. Część obliczeń powstała z wykorzystaniem infrastruktury PL-Grid.



Spis treści

Podziękowania.....	3
1. Streszczenie.....	6
2. Wstęp	8
2.1. Krystalografia białek.....	9
2.2. Techniki krystalizacji białek	10
2.3. Metody rozwiązywania problemu fazowego	11
2.3.1. Podstawienie molekularne.....	11
2.3.2. Wielokrotne podstawienie izomorficzne (MIR).....	11
2.3.3. Dostrojona dyfrakcja anomalna	12
2.4. Mapy gęstości elektronowej białek.....	13
2.5. Tworzenie modelu struktury i udokładnianie	14
2.6. Walidacja struktury	15
2.7. Struktury modulowane.....	16
2.7.1. Struktury modulowane białek	21
2.7.2. Czynniki przemieszczenia atomowego (ADPs) w modelach białek	21
2.8. Zbliżniaczenie	22
2.9. Kwazikryształy	26
2.9.1. Kwazikryształ 1D – ciąg Fibonacciego.....	27
2.9.2. Parkietaż Penrose’a (Penrose tilings)	28
2.9.3. Metoda średniej komórki elementarnej (AUC).....	29
2.9.4. Struktury modulowane a kwazikryształy.....	31
2.10. Symulacje dynamiki molekularnej struktur białek	31
3. Cel rozprawy.....	35
4. Materiały i metody.....	37
4.1. Kompleks białkowy Hyp-1/ANS.....	37
4.2. Dane eksperymentalne	38
4.3. Rozwiązanie struktury	38
4.3.1. Phaser	38
4.4. Udokładnienie modelu	39
4.4.1. REFMAC5	39

4.5.	Wizualizacja danych	39
4.5.1.	Coot	39
4.5.2.	PyMOL	40
4.5.3.	UCSF Chimera	40
4.6.	Testy zbliżniaczenia	40
4.6.1.	TRUNCATE	40
4.7.	Symulacje białka Hyp-1 metodami dynamiki molekularnej	41
4.7.1.	GROMACS	41
4.8.	Program do udokładniania modelu białka	41
4.8.1.	Optymalizacja problemu przy użyciu algorytmu metaheurystycznego MEIGO	41
4.8.2.	Skrypty MATLAB do przygotowania pliku wejściowego	43
4.8.3.	Przekształcanie danych do nadprzestrzeni (3+1)D	44
4.8.4.	Generowanie wiązań stereochemicznych	45
4.8.5.	Udokładnianie modelu	46
4.8.6.	Skrypty do generowania graficznej reprezentacji modelu i jego weryfikacji	48
5.	Dane doświadczalne wykorzystane w rozprawie	49
5.1.	Otrzymanie kompleksu białkowego Hyp-1/ANS	49
5.2.	Rozwiązanie struktury Hyp-1/ANS	50
6.	Wyniki	53
6.1.	Kompleksy białkowe Hyp-1/ANS jako struktury modulowane	53
6.2.	Testy zbliżniaczenia	56
6.3.	Upakowanie molekuł w komórce elementarnej	57
6.4.	Podobieństwo struktury 9Hyp/ANS do modulowanej struktury 7Hyp/ANS	60
6.5.	Udokładnienie modelu 9Hyp/ANS	62
6.6.	Miejsca wiązania ANS	64
6.7.	Symulacje białka Hyp-1 metodami dynamiki molekularnej	65
6.7.1.	Model Hyp-1 użyty do symulacji	65
6.7.2.	Algorytm symulacji MD białka Hyp-1	66
6.8.	Proces przekształcania superkomórki 7Hyp/ANS w jej reprezentację w przestrzeni wielowymiarowej (3+1)D	68
6.9.	Udokładnienie modelu 7Hyp-ANS w komórce (3+1)D	72
6.10.	Weryfikacja poprawności stereochemicznej struktury	73
6.11.	Wprowadzenie poprawek związanych z obecnością fononów	79
7.	Dyskusja	80
7.1.	Wpływ warunków krystalizacji na modulację struktury	80
7.2.	Artefakty pozostałe po krystalizacji	81

7.3.	Analiza zbliżnienia	81
7.4.	Obsadzenie modelu struktury cząsteczkami ANS	82
7.5.	Konformacja cząsteczek ANS i ich wpływ na formowanie nadstruktury	85
7.6.	Identyfikacja obecności liganda melatoniny w mapach gęstości elektronowej	86
7.7.	Porównanie eksperymentalnego i symulowanego metodami MD modelu struktury Hyp-1	88
7.8.	Porównanie otrzymanych i doświadczalnych ADPs	89
7.9.	Symulowane rozkłady prawdopodobieństwa kątów łańcucha bocznego $P(\chi_1)$ dla Pro, Ser, Cys, Val i Thr	92
7.10.	Symulowane bimodalne rozkłady prawdopodobieństwa kątów łańcucha bocznego $P(\chi_1, \chi_2)$ dla Leu, Phe, Tyr, His, Asp i Asn	96
7.11.	Symulowane rozkłady prawdopodobieństwa kątów łańcucha bocznego $P(\chi_1, \chi_2, \chi_3)$ dla Glu, Gln i Met	99
7.12.	Symulowane rozkłady prawdopodobieństwa kątów łańcucha bocznego $P(\chi_1, \chi_2, \chi_3, \chi_4)$ dla Lys i Arg	102
7.13.	Wyniki udokładnienia w nadprzestrzeni (3+1)D oraz korekt fononowych	104
8.	Podsumowanie	105
9.	Bibliografia.....	108
10.	Dorobek naukowy	114

1. Streszczenie

Zjawisko modulacji struktury jest stosunkowo dobrze poznane w krystalografii małowymiarowej, ale jego występowanie w makromolekularnych kryształach białkowych było zaskakujące. Fizyczne przejawy tego zjawiska obejmują obserwację dodatkowych refleksów między głównymi pikami braggowskimi na dyfraktogramach. W rezultacie modulacji, symetria translacyjna kryształu zostaje zaburzona w przestrzeni trójwymiarowej, a periodyczność struktury zostaje przywrócona tylko w wyższych wymiarach. Wymaga to stosowania specjalistycznych metod analizy wielowymiarowej do prawidłowego wskaźnikowania obrazów dyfrakcyjnych i opisu struktury. Modulacja struktury może być spowodowana zarówno periodycznymi zmianami pozycji atomów w miejscach określonych przez symetrię przestrzenną komórki, jak i okresowymi zmianami obsadzenia danej pozycji krystalograficznej.

Dotychczasowe metody rozwiązywania i udokładniania struktur rutynowo stosowane w krystalografii białek nie są odpowiednie do przeprowadzenia kompleksowej analizy modulowanych struktur. Ograniczenie do opisu w przestrzeni trójwymiarowej oznacza przyjęcie wymierności modulacji, gdzie porządek translacyjny zostaje przywrócony po określonej, całkowitej liczbie komórek elementarnych. Wówczas strukturę należy analizować w powiększonej superkomórce, co prowadzi do znacznego wzrostu liczby

parametrów w przypadku złożonych struktur białkowych oraz umożliwia uzyskanie tylko przybliżonego modelu.

Brak odpowiednich narzędzi do analizy makromolekularnych struktur modulowanych prowadzi do poważnych problemów z właściwym wskaźnikowaniem i przetwarzaniem danych dyfrakcyjnych, a następnie konstruowaniem kompletnego modelu struktury o zadowalających wskaźnikach rozbieżności. Dotychczas udało się przeprowadzić pełną analizę strukturalną tylko dla kilku modulowanych kryształów białek. Należą do nich kompleksy białka Hyp-1 z dziurawca zwyczajnego (*Hypericum perforatum*) z fluorescencyjnym ligandem ANS (8-anilinonaftaleno-1-sulfonianem). W zależności od warunków krystalizacji, kompleksy białkowe Hyp-1/ANS mogą tworzyć kryształy o siedmio- (7Hyp/ANS) lub dziewięciokrotnej modulacji struktury (9Hyp/ANS) wzdłuż kierunku *c* grupy przestrzennej C2.

W ramach rozprawy doktorskiej porównywałam dwie modulowane struktury krystaliczne kompleksu białkowego Hyp-1/ANS. Pierwsza z nich, oznaczona jako 7Hyp/ANS, została otrzymana, rozwiązana i opisana przez dr Joannę Śliwiak oraz grupę prof. Mariusza Jaskólskiego w 2015 roku i posiadała siedmiokrotną modulację wzdłuż osi *c* oraz zawierała 28 niezależnych molekuł białka w rozbudowanej superkomórce. Obecność modulacji i elementów tNCS (Translational Non-Crystallographic Symmetry, niekrystalograficzna symetria translacyjna) była połączona ze zbliżeniem kryształu, co stanowiło dodatkową trudność podczas rozwiązania i udokładniania struktury. Ostatecznie struktura została rozwiązana i udokładniona w podejściu superkomórki, a następnie wyniki badań zdeponowano w bazie PDB (Protein Data Bank) z kodem 4N3E. Wkrótce potem podczas zmiany warunków krystalizacji i kokrystalizacji w obecności hormonu roślinnego melatoniny, otrzymano kolejną modulowaną strukturę kompleksu Hyp-1/ANS. Omawiany model kompleksu 9Hyp/ANS z dziewięciokrotną modulacją wzdłuż osi *c* składał się finalnie z 36 molekuł białka Hyp-1 rozmieszczonych zgodnie z motywem zawierającym 4 molekuły (2 dimery) Hyp-1 powtórzonym 9 razy wzdłuż *c* ($4 \times 9 = 36$). Struktura 9Hyp/ANS została rozwiązana przez dr Joannę Śliwiak analogicznie do poprzedniego przypadku 7Hyp/ANS, a następnie udokładniona przeze mnie w ujęciu superkomórki.

Przedstawiona rozprawa doktorska składa się z trzech zasadniczych części i zadań związanych z analizą struktury krystalicznej białka Hyp-1. Pierwsza część obejmuje udokładnienie modelu struktury kryształu 9Hyp/ANS w ujęciu superkomórki przy użyciu konwencjonalnego oprogramowania oraz związanej z nim analizy elementów strukturalnych kryształu, szczególnie kwestii związanych z upakowaniem atomów w obrębie komórki elementarnej, perturbacjami periodyczności struktury, porównaniem ze strukturą 7Hyp/ANS oraz rozmieszczeniem liganda.

W drugiej części dokonałam szczegółowej analizy danych dyfrakcyjnych obu modulowanych kryształów Hyp-1/ANS, udokładnienia oraz metodycznego opisu

i udokładnienia struktury 7Hyp/ANS w przestrzeni wielowymiarowej. W procesie udokładniania struktury 7Hyp/ANS zostało użyte autorskie, specjalistyczne oprogramowanie stworzone w środowisku Matlab. Zrezygnowanie z uproszczonego założenia wymierności modulacji oraz wprowadzenie dodatkowych poprawek uwzględniających nieporządek w strukturze pozwoliło na uzyskanie nowych modeli oraz poprawę ich wskaźników rozbieżności. Opracowany pakiet został następnie rozbudowany o kolejne moduły umożliwiające wizualizację danych oraz wprowadzanie poprawek związanych z drganiami termicznymi sieci krystalicznej (fononami).

Trzecia część pracy dotyczy zastosowania metod dynamiki molekularnej do badania ruchów termicznych i konformacji łańcuchów bocznych białka Hyp-1 w oparciu o czysto fizyczne podejście, w oderwaniu od więzów narzucanych podczas udokładniania struktury krystalicznej. Uzyskane informacje pozwoliły określić preferowane energetycznie konformacje w obrębie białka oraz ich porównanie z wariantami dobranymi w procesie udokładniania struktury krystalicznej.

2. Wstęp

Obserwacja dyfrakcji promieniowania rentgenowskiego (promieniowania X) na kryształach przez Friedricha, Knippinga oraz Lauego w 1912 roku jest jednym z najważniejszych odkryć w historii nauki, które popchnęło badaczy z całego świata do przeprowadzania własnych eksperymentów. W listopadzie 1912 roku Towarzystwu Filozoficznemu w Cambridge zostało przedstawione równanie Bragga determinujące warunki konstruktywnej interferencji wiązek promieniowania rozproszonego na atomach sieci krystalicznej, a w lipcu 1913 roku zostały opublikowane pierwsze kompletne struktury krystaliczne diamentu, NaCl, KCl, KBr i KI [1]. W 1913 roku Ewald zaprezentował geometryczne przedstawienie równań interferencyjnych Laue'go z siecią odwrotną i sferą odbicia [2]. Bardzo istotny problem dyfrakcji na kryształach o zaburzonej temperaturze został rozwiązany w latach 1913-1914 przez Debye'a [3], który wyprowadził izotropowy czynnik temperaturowy B (stosowany do dziś z poprawką wprowadzoną przez Wallera w 1923 r.) oraz temperaturę Debye'a charakteryzującą właściwości sprężyste kryształów. Niezależnie od siebie Debye & Scherrer w 1916 roku [4] oraz Hull w 1917 roku [5], wynaleźli metodę proszkową, która stała się niezmiernie użyteczna w metalurgii i wielu innych dziedzinach badań strukturalnych.

Podstawowym założeniem klasycznej krytalografii jest istnienie dyskretnej symetrii translacyjnej, która determinuje możliwe elementy symetrii kryształu. W 1919 roku Niggli opublikował w swojej książce *Geometrische Kristallographie des Diskontinuums* 230 możliwych grup przestrzennych kryształów, w tym wszystkie pozycje równoważne i wiele diagramów, a także omówił metody wyznaczania grup przestrzennych poprzez wygaszenia systematyczne [6]. Wczesne określanie struktury krystalicznej odbywało się metodą prób i błędów, tj. metodą zgadywania na podstawie wiedzy o dyfrakcji, symetrii i geometrii kryształu oraz poprzez analogię do poznanych struktur. Niemniej, do 1930 roku udało się

wyznaczyć i opisać wiele struktur nieorganicznych oraz stopów. W latach trzydziestych XX wieku do powszechnego użytku weszły metody Fouriera, chociaż już znacznie wcześniej zdano sobie sprawę, że czynniki struktury F są w istocie współczynnikami szeregu Fouriera.

W 1934 roku zaproponowano metodę Pattersona [7], która pozwala tworzyć mapy wektorowe, których maksima znajdują się w położeniach odpowiadających względnym odległościom między atomami w rzeczywistej strukturze. Jeśli pary atomów należą do tej samej cząsteczki, to odpowiadające im wektory są stosunkowo krótkie, a ich punkty końcowe znajdują się niedaleko od początku mapy Pattersona; nazywa się je wektorami własnymi Pattersona. Gdyby nie istniały wektory międzycząsteczkowe (wektory krzyżowe Pattersona), wewnętrzny region mapy Pattersona byłby jednakowy dla tej samej cząsteczki w różnych strukturach krystalicznych, a jedyna różnica dotyczyłaby operacji rotacji. Dla cząsteczek homologicznych mapy Pattersona nie są dokładnie takie same, choć bardzo podobne. Dlatego wektory własne Pattersona mogą dostarczyć relacji rotacyjnych pomiędzy znanymi i nieznanymi strukturami molekularnymi.

Dość wcześnie zdano sobie sprawę, że powinno być możliwe wyznaczenie fazy bezpośrednio z obserwowanych wartości bezwzględnych czynników struktury, ponieważ standardowy model kulisto-atomowy zawiera o wiele mniej parametrów (współrzędnych atomowych) do określenia, niż liczba przeprowadzonych obserwacji. Podstawy nowoczesnych metod bezpośrednich zostały wprowadzone wraz z równaniem Sayre'a, pracami Hauptmana i Karle (1953) oraz zaprzęgnięciem do pracy komputerów. Najnowszą metodą wyznaczania struktury jest technika *charge flipping*, zaproponowana po raz pierwszy w 2004 roku [8]. Jest ona koncepcyjnie nowa, ponieważ nie opiera się na założeniu, że struktura składa się z atomów o znanej gęstości elektronowej, a jedynie zakłada, że gęstość rozpraszania (gęstość elektronów) jest skoncentrowana w kilku regionach, podczas gdy większość komórki elementarnej jest zasadniczo pusta. Metoda ta nie wymaga początkowych informacji o symetrii, gdyż struktura jest rozwiązana domyślnie w prymitywnej trójskośnej grupie przestrzennej $P1$, znajomość składu chemicznego nie jest wymagana, a wszelkie ograniczenia symetrii wydają się jedynie utrudniać rozwiązanie struktury. Dane dyfrakcyjne nie muszą być znormalizowane, a współczynnik skali nie musi być znany.

2.1. Krystalografia białek

Rozwój krystalografii makromolekularnej jest historią dość analogiczną do opisanej powyżej, ale o 20-30 lat późniejszą niż rozwój krystalografii małowcząsteczkowej. W 1934 roku J. D. Bernal uzyskał pierwsze udane zdjęcia dyfrakcyjne uwodnionego białka. Niespełna 25 lat później pierwsze struktury białkowe, mioglobiny i hemoglobiny, zostały z powodzeniem określone przez J. C. Kendrew i M. F. Perutza. Gdy w 1971 roku został utworzony Protein Data Bank (PDB), początkowo zdeponowano tam jedynie siedem struktur białkowych. Po prawie trzech dekadach powolnego napływu struktur, w połowie lat 90. PDB otrzymał ogromny impuls wchodząc w fazę wykładniczego wzrostu. Głównymi

czynnikami były: postępy w technologii komputerowej i informacyjnej zapewniające potrzebną moc obliczeniową dla skomplikowanych obliczeń, jak również bardziej wydajne algorytmy i środki automatyzacji eksperymentów, wprowadzenie inżynierii genetycznej dla łatwego wytwarzania praktycznie dowolnego białka w "fabrykach" komórek bakteryjnych oraz powszechne użycie potężnych synchrotronowych źródeł promieniowania X [9].

Wprowadzenie źródeł synchrotronowych trzeciej generacji oraz ultraszybkich pikselowych detektorów powierzchniowych (APD) zredukowało czas pomiaru danych do sekund, przy jednoczesnym zmniejszeniu wielkości kryształów do mikrometrów oraz ogólnej poprawie jakości danych. Dodatkowe korzyści są powiązane z opracowaniem metody MAD (Multiwavelength Anomalous Diffraction, dostrojonej dyfrakcji anomalnej) oraz jej uproszczonej pochodnej, SAD (Single-wavelength Anomalous Diffraction), jako technik rozwiązywania problemu fazowego [10]. Bazują one na rozpraszaniu przestrajalnego promieniowania synchrotronowego przez atomy wykazujące dyfrakcję anomalną, takie jak selen, które mogą być wprowadzone do rekombinowanych białek w postaci selenometioniny Se-Met.

W idealnym przypadku kryształy mogą być opisane jako nieskończone matryce, których bloki konstrukcyjne (jednostki asymetryczne) są ułożone zgodnie z dobrze zdefiniowanymi regułami symetrii tworząc 230 grup przestrzennych. Powstałe komórki elementarne są powtarzane w trzech wymiarach poprzez translacje. Istotnym jest, że białka i kwasy nukleinowe nie krystalizują w grupach przestrzennych z symetrią inwersyjną, ponieważ składają się z enancjomerów (odpowiednio L-aminokwasów i D-cukrów), co zmniejsza liczbę możliwych grup przestrzennych z 230 do 65 [9-11].

2.2. Techniki krystalizacji białek

Podstawą do określenia struktury metodami rentgenowskimi jest uzyskanie wysokiej jakości, czystych i homogenicznych monokryształów białka. Krystalizacja białek, kwasów nukleinowych i dużych kompleksów biologicznych, takich jak wirusy, rozpoczyna się od uzyskania kontrolowanego stanu przesylenia. Istnieje wiele metod krystalizacji białek, z których wszystkie mają na celu doprowadzenie roztworu białka do stanu przesylenia, co wymusza przejście makrocząsteczek w stan stały, czyli krystaliczny [9,10]. Krystalizacja makrocząsteczek jest procesem wieloparametrycznym obejmującym trzy główne etapy: nukleację, wzrost i terminację wzrostu. Stadium przesylenia jest osiągnięte poprzez dodanie łagodnych czynników strącających, takich jak neutralne sole lub polimery, lub w wyniku manipulacji różnymi parametrami, takimi jak temperatura, siła jonowa i pH. W procesie krystalizacji ważne są również czynniki, które mogą wpływać na stan strukturalny makrocząsteczki, takie jak jony metali, inhibitory, kofaktory lub inne konwencjonalne małe cząsteczki. Opracowano wiele technik krystalizacji, a do najczęściej stosowanych należą: dializa, dyfuzja par metodą wiszącej lub siedzącej kropli i dyfuzja cieczy [9]. W ostatnich latach osiągnięto szybki wzrost liczby sukcesów w krystalizacji

makrocząsteczek dzięki pojawieniu się praktycznych, łatwych w użyciu zestawów przesiewowych oraz zastosowaniu robotyki laboratoryjnej.

2.3. Metody rozwiązywania problemu fazowego

Wyznaczenie gęstości elektronowej wymaga jednoczesnej znajomości amplitudy oraz fazy fal rozproszonych na tej gęstości elektronowej. Jedyną wielkością dostępną eksperymentalnie jest natężenie danego refleksu proporcjonalne do kwadratu amplitudy fali rozproszonej, natomiast nie istnieje żadna bezpośrednia metoda do wyznaczenia fazy fal rozproszonych. Stanowi to istotę problemu fazowego [9]. Obok konieczności uzyskania dobrze rozprasających monokryształów, problem fazowy jest drugim istotnym ograniczeniem w krystalografii białek. Eksperymentalnie problem fazowy w krystalografii białek może być rozwiązany za pomocą kilka metod:

- Podstawienie molekularne (Molecular Replacement, MR)
- Wielokrotne podstawienie izomorficzne (Multiple Isomorphous Replacement, MIR)
- Dostrojona dyfrakcja anomalna o wielu długościach fali (MAD) lub o jednej długości fali (SAD)

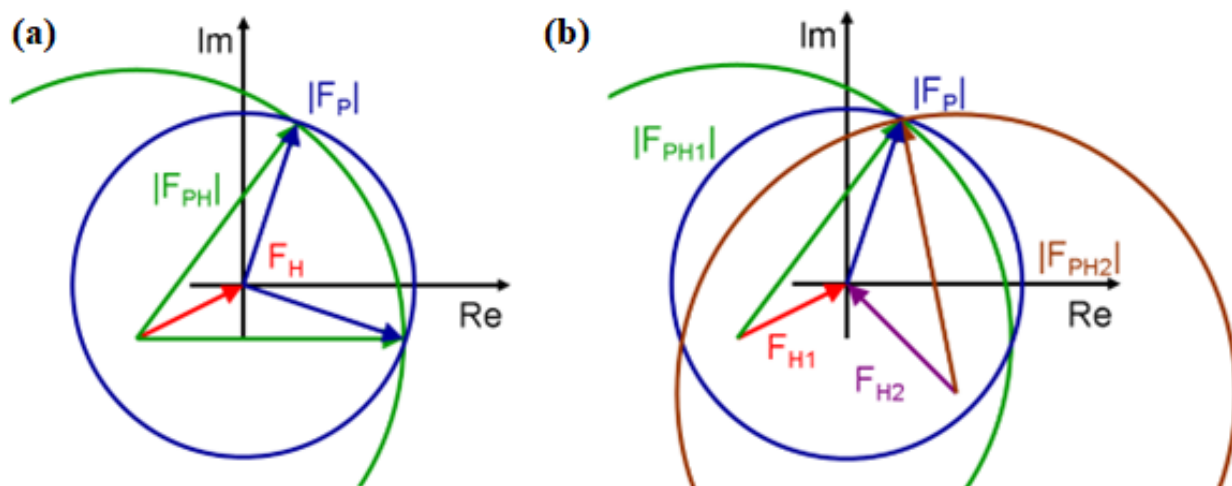
2.3.1. Podstawienie molekularne

Podstawienie molekularne wykorzystuje model białka o znanej strukturze. Może to być dobrze opisany mutant badanego białka, struktura białka z tej samej rodziny w przypadku białka dzikiego krystalizowanego po raz pierwszy bądź białko homologiczne pochodzące z innego organizmu. Wychodząc z modelu o znanej strukturze należy obliczyć jego czynniki struktury. Następnie wyznacza się położenie i orientację cząsteczki modelu względem komórki elementarnej badanego kryształu. Aby możliwie jak najwierniej odtworzyć zmierzony obraz dyfrakcyjny, testom podlegają wszystkie możliwe położenia i orientacje modelu w komórce elementarnej. W przypadku komórki elementarnej zawierającej N cząsteczek białka oznacza to wyznaczenie $6N$ parametrów: trzech kątów towarzyszących obrotowi cząsteczki oraz trzech wektorów translacji. Dla uproszczenia próbkuje się najpierw przestrzeń rotacji, po czym wybierane jest najlepsze rozwiązanie (jedno lub kilka) i tylko dla tych wybranych rozwiązań próbkowana jest przestrzeń translacji. Wstępne fazy odpowiednich zmierzonych czynników pochodzą z faz policzonych dla modelu. Podobnie wstępna mapa gęstości elektronowej jest liczona na podstawie faz przeniesionych z modelu.

2.3.2. Wielokrotne podstawienie izomorficzne (MIR)

Ta eksperymentalna metoda wyznaczania faz polega na podstawieniu w kilku miejscach struktury atomów metali ciężkich, takich jak Hg, Au i Pt, co prowadzi do uzyskania pochodnej wzbogaconej o atomy z dużą liczbą elektronów. Pozycje atomów ciężkich określa się na podstawie różnic w amplitudzie czynnika struktury pomiędzy natywnymi (niemodyfikowanymi) i pochodnymi zestawami danych, posilując się mapą Pattersona lub metodami bezpośrednimi. Po wyznaczeniu pozycji atomów ciężkich można obliczyć odpowiadające im położeniom amplitudy i fazy, a następnie porównać

obserwowane amplitudy struktury natywnej (F_P) i pochodnej (F_{PH}) do oszacowania faz całej struktury (Rys. 1). W przypadku jednej pochodnej możliwe są dwa rozwiązania równań, co można wyeliminować poprzez dodanie dalszych pochodnych.

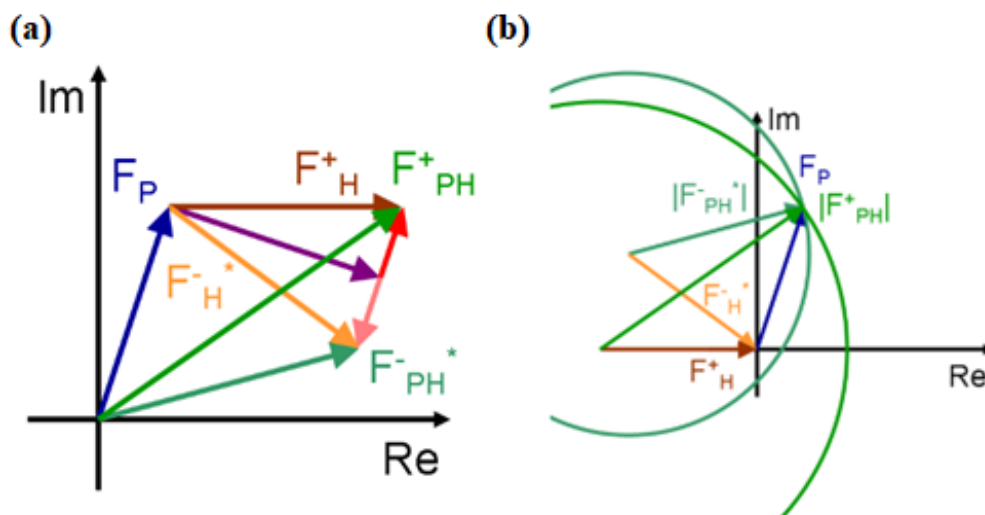


Rysunek 1. Konstrukcje Harkera przedstawiające rozwiązanie czynników struktury ciężkiego atomu (F_H) z wykorzystaniem amplitudy kryształu natywnego (F_P) i pochodnej (F_{PH}). Metoda pojedynczego podstawienia izomorficznego (SIR) dostarcza dwie możliwe fazy dla (F_P). W przypadku MIR niejednoznaczność fazowa jest rozwiązywana poprzez uwzględnienie informacji z drugiego lub dalszych kryształów pochodnych (b) [3].

2.3.3. Dostrojona dyfrakcja anomalna

Metoda dostrojonej dyfrakcji anomalnej opiera się na zjawisku anomального rozpraszania promieniowania rentgenowskiego ze zmianą amplitudy i fazy przez ciężkie atomy przy długościach fal bliskich ich krawędziom absorpcji. Stąd wynikają zmiany natężeń dyfrakcji, które są wykorzystywane do wyznaczenia podstruktury atomów anomalnych. Podstawowa procedura wykorzystuje zmodyfikowane białka zawierające selenometioninę (Se-Met) wprowadzoną zamiast reszt metioniny jako źródło rozpraszania anomального, jednakże można zastosować tę metodę dla innych metaloprotein zawierających jeden lub więcej endogennych metali (np. Fe, Cu, Mo, Zn, Ni). Podczas rozpraszania anomального współczynnik rozpraszania atomowego zawiera trzy składniki: normalne rozpraszanie zależne od kąta Bragga, poprawkę rzeczywistą (f') i czynnik anomalny (f'') związane z długością fali i przesunięte w fazie o 90 stopni względem siebie, jak przedstawiono na Rys. 2. W wyniku złamania prawa Friedela (mówiącego, że intensywności refleksów o wskaźnikach h, k, l i $-h, -k, -l$ są sobie równe) powstają różnice anomalne, które można wykorzystać do lokalizacji atomów rozpraszających anomalnie. W typowym eksperymencie MAD dokonuje się pomiaru zestawów danych dyfrakcyjnych przy trzech długościach fali: w maksimum krzywej absorpcji (najwyższe f''), w punkcie przegięcia krzywej absorpcji (minimalne f') oraz dla długości fali powyżej krawędzi absorpcji, gdzie f' jest bliskie zeru, a f'' może być jeszcze stosunkowo duże. Wprowadzenie techniki SAD

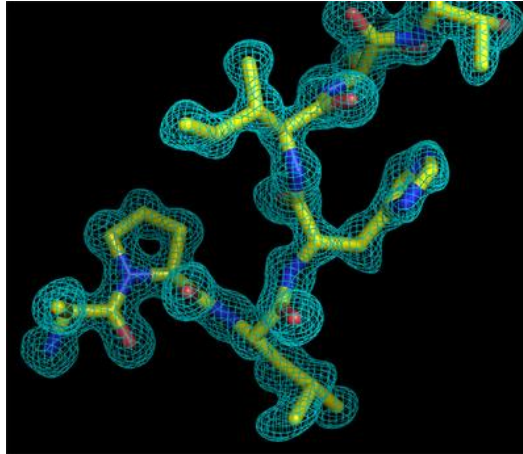
umożliwia pomiar zestawu danych tylko dla jednej długości fali, zazwyczaj w maksimum absorpcji. W ten sposób minimalizuje się uszkodzenia radiacyjne kryształu oraz problem nieizomorfizmu, ale zastosowanie metody wymaga bardzo dokładnych pomiarów.



Rysunek 2. Konstrukcje Harkera obrazujące wyznaczenie wkładów do czynników struktury od anomalnych wkładów ciężkich atomów (F_H^+ i F_H^-) (a). Rysując okręgi wokół końców obu czynników struktury i biorąc pod uwagę amplitudy anomalnej pary Friedla (F_{PH}^+ , F_{PH}^-) można wyznaczyć fazę (b) [3].

2.4. Mapy gęstości elektronowej białek

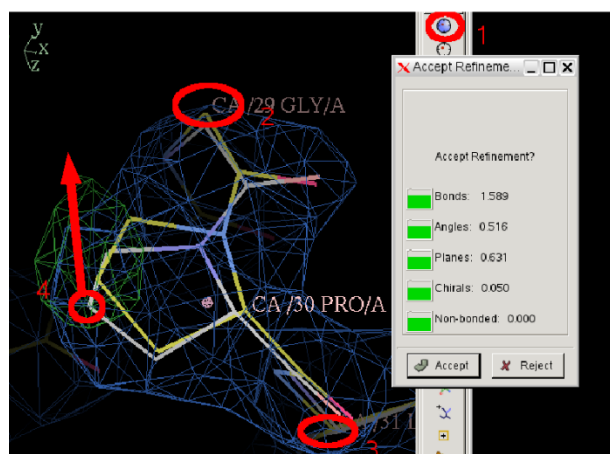
Gęstość elektronowa stanowi odwrotną transformatę Fouriera obrazów dyfrakcyjnych i można ją obliczyć znając amplitudy oraz fazy fal rozproszonych. Rozdzielczość obecnie uzyskiwanych struktur białek sięga 1 Å. Ruchliwość białek oraz defekty sieci krystalicznej narzucają istotne ograniczenia rozdzielczości uzyskiwanych map gęstości elektronowej, na których nie można odróżnić pojedynczych atomów, a jedynie obszar gęstości elektronowej, w którym atomy nie są wyraźnie oddzielone od siebie (Rys. 3). Wprowadzenie większych kątów rozpraszania powoduje zwiększenie liczby słabych refleksów o niewielkiej intensywności, które mogą zawierać informację o istotnych detalach strukturalnych. Rozmycie map gęstości elektronowej kryształów białek jest również powiązane z niezbyt wysokim stosunkiem sygnału do szumu.



Rysunek 3. Mapa gęstości elektronowej białka z wbudowanym fragmentem łańcucha peptydowego.

2.5. Tworzenie modelu struktury i udokładnianie

Rozwiązanie struktury białka odbywa się na podstawie dostępnych informacji o badanym białku, zwłaszcza sekwencji aminokwasowej, gdyż mapy gęstości elektronowej rzadko mają rozdzielczość atomową. Jeżeli rozdzielczość jest lepsza niż $\sim 3.2 \text{ \AA}$, można wykorzystać uzyskane mapy do automatycznego budowania modeli. Oprogramowania PHENIX [12], SHELX [13], ARP/WARP [14] i BUCCANEER [15] są rutynowo stosowane do wstępnego budowania i udoskonalania modeli poprzez iteracyjne łączenie faz oraz procedurę ulepszania modelu w kolejnych cyklach budowy. Po wygenerowaniu dobrego modelu wyjściowego wygodne narzędzie do dalszego ręcznego budowania modelu i jego korekty stanowi program graficzny COOT [16]. Udokładnianie struktury to iteracyjny proces cyklicznego poprawiania, obliczania rozbieżności, wprowadzania dalszych poprawek, aż do osiągnięcia założonego poziomu zgodności modelu (Rys. 4). Podczas budowy i udokładniania modelu nakłada się ograniczenia na długości i kąty wiązań określone dzięki poprzedniemu rozwiązaniu tysięcy innych struktur białkowych. Na podstawie tych ograniczeń można racjonalnie budować główny łańcuch polipeptydowy białka, podczas gdy łańcuchy boczne aminokwasów modeluje się na podstawie baz danych konformacji reszt aminokwasowych i częstotliwości ich występowania w białkach.



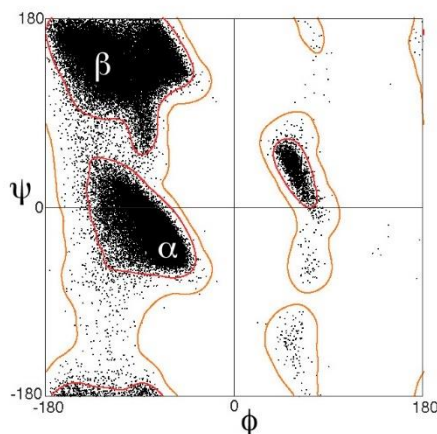
Rysunek 4. Proces ręcznego poprawiania modelu białka i dopasowywania wiązań stereochemicznych w programie Coot. W tym stanie można ręcznie manipulować atomami i przeciągać je w optymalne dla struktury położenia w mapie gęstości elektronowej.

2.6. Walidacja struktury

Decyzja o rezygnacji z dalszego udokładniania modelu struktury jest zawsze arbitralna i następuje po osiągnięciu zakładanego poziomu zgodności. Na ostatnim etapie rozwiązywania struktury dokonuje się sprawdzenia jakości uzyskanej struktury. Umożliwia to weryfikację, na ile wiarygodne są elementy dopasowane przez eksperymentatora do rozmytych obszarów gęstości elektronowej. Miarą jakości dopasowania uzyskanej struktury do danych dyfrakcyjnych są wskaźniki rozbieżności R zdefiniowane jak we wzorze (1) jako względna różnica między amplitudami czynników struktury obliczonych dla modelu (F_{calc}) a wyznaczonymi doświadczalnie na podstawie pomiaru intensywności refleksów (F_{exp}):

$$R = \frac{\sum_{hkl} |F_{hkl}^{exp}| - |F_{hkl}^{calc}|}{\sum_{hkl} |F_{hkl}^{exp}|} \quad (1)$$

Mniejsza wartość R świadczy o lepszym dopasowaniu modelu do danych dyfrakcyjnych. Za dobrze udokładnione zwykle uznaje się struktury o $R < 0,2$. Dodatkowy wskaźnik rozbieżności R_{free} dla puli refleksów wyłączonych z udokładniania służy do oceny, czy udokładniane parametry mają pełne oparcie statystyczne w informacji zmierzonej podczas eksperymentu. Wartość R_{free} nie powinna znacząco przekraczać wartości R . Istotne jest również sprawdzenie geometrycznej poprawności uzyskanej struktury i nałożonych na nią wiązań stereochemicznych. Białka mają typowe przedziały wartości kątów i wiązań, toteż należy wyeliminować znaczące odstępstwa od tych wielkości. Wiązania peptydowe muszą być płaskie, natomiast kąty ψ i ϕ łańcucha głównego zawierać się w określonych przedziałach przedstawionych jako obszary dozwolone na mapach Ramachandrana [17]. W bazach danych porównuje się konformacje łańcuchów bocznych aminokwasów [18,19]. Sprawdzeniu podlegają również kontakty van der Waalsa, co umożliwia ocenę odległości między atomami. Weryfikowane są również wzajemne oddziaływania aminokwasów hydrofobowych, a także aminokwasów polarnych z innymi polarnymi lub wodą.



Rysunek 5. Wykres Ramachandrana, na którym każdy punkt odpowiada parze kątów torsyjnych (dwuściennych) ψ i ϕ występujących w pojedynczym aminokwasie łańcucha białkowego. Obszary dozwolone są zaznaczone czerwonymi obwiedniami.

2.7. Struktury modulowane

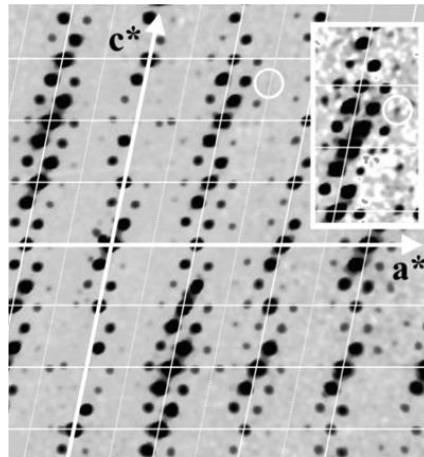
Podstawową własnością klasycznego kryształu jest translacyjna periodyczność, której bezpośrednią konsekwencją stanowi istnienie wzoru dyfrakcyjnego dla promieniowania X, neutronów lub elektronów z ostrymi pikami dyfrakcyjnymi zlokalizowanymi w punktach siatki zdefiniowanych przez podstawowe wektory a_1^* , a_2^* oraz a_3^* :

$$h = h_1 a_1^* + h_2 a_2^* + h_3 a_3^* \quad (2)$$

Pierwszymi wzmiankami o możliwości istnienia struktur przełamujących tradycyjną periodyczność sieci były prowadzone pod koniec XIX wieku badania kryształów minerału kalawerytu [20]. Stwierdzono, że płaszczyzny tego kryształu nie spełniają prawa wskaźników wymiernych. W 1935 roku Donnay [21] pokazał, że powierzchnie, które nie mogą być wywskaźnikowane, odpowiadają dodatkowym refleksom na obrazach dyfrakcji rentgenowskiej. W 1938 roku Preston zaobserwował na obrazach dyfrakcyjnych stopu Al-Cu obecność dodatkowych refleksów-satelit regularnie rozmieszczonych wokół normalnych, silnych refleksów głównych [22]. Wy tłumaczeniem jest kombinacja sinusoidalnej fali gęstości i sinusoidalnej fali odpowiadającej za przesunięcie atomów z ich pozycji równowagowych.

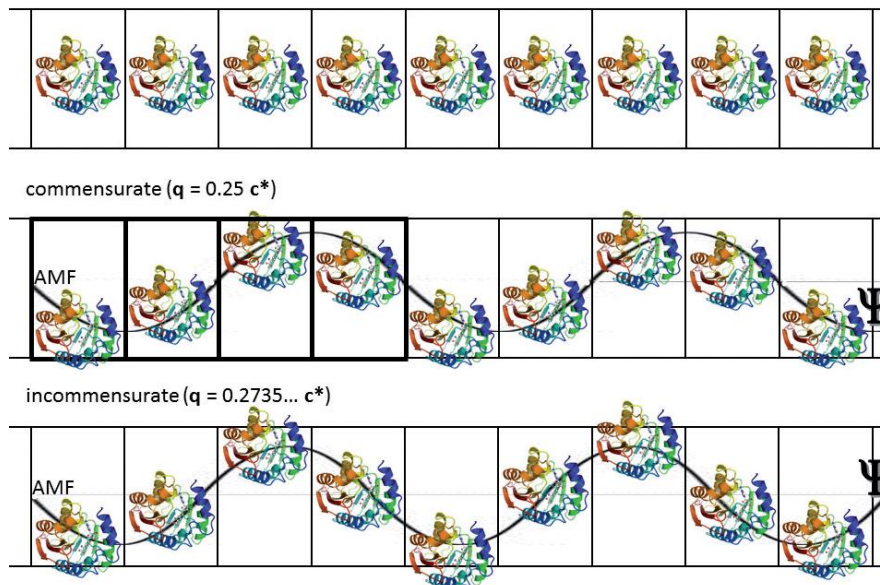
W przypadku niektórych związków pojawiają się dodatkowe ostre piki dyfrakcyjne, tzw. satelity, zlokalizowane poza węzłami sieci [10,23]. Piki te można wskaźnikować przy zastosowaniu jednego lub więcej dodatkowych wektorów modulacji oznaczanych jako q_j , gdzie ($j = 0,1,2,\dots,d$):

$$h = h_1 a_1^* + h_2 a_2^* + h_3 a_3^* + m_1 q_1 + m_2 q_2 + \dots + m_d q_d \quad (3)$$



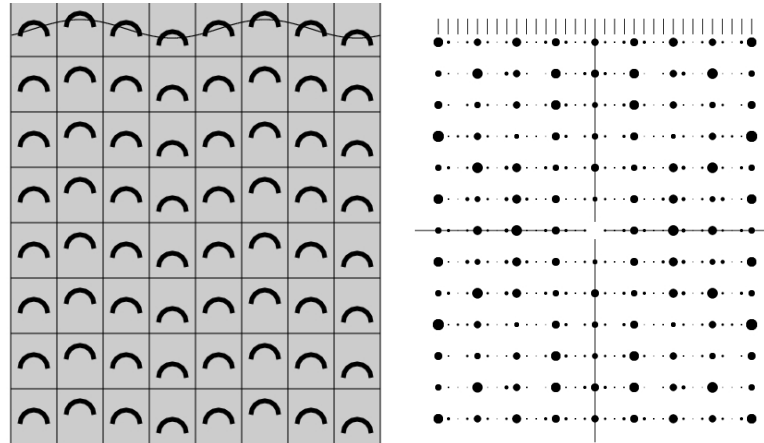
Rysunek 6. Obraz dyfrakcyjny warstwy $(h,0,l)$ dla kryształu Na_2CO_3 . Biała siatka przedstawia siatkę głównych refleksów. Refleksy satelitarne są regularnie przesunięte względem głównych refleksów przez wektor modulacji q [14].

Za modulację uważa się tu okresową deformację struktury podstawowej posiadającej symetrię danej grupy przestrzennej. O modulowanej strukturze kryształu można mówić, gdy ścisły, krótkozasięgowy porządek translacyjny od jednej komórki jednostkowej do następnej zostaje utracony, ale uporządkowanie długiego zasięgu ulega przywróceniu przez periodyczną atomową funkcję modulacyjną Ψ (Atomic Modulation Function, AMF) [24]. Jeżeli periodyczność modulacji nie jest zgodna z periodycznością struktury podstawowej, to zmodulowaną strukturę krystaliczną nazywamy niewspółmierną. W przypadku struktury niewspółmiernej przynajmniej jedna składowa wektora modulacji podstawowej q_j względem struktury bazowej jest liczbą niewymierną [25].

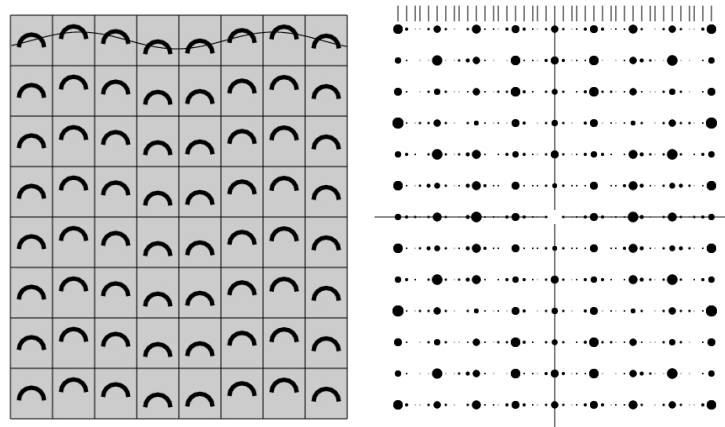


Rysunek 7. Ilustracja struktury idealnie periodycznej (na górze), współmiernie modulowanej (w środku) i niewspółmiernie modulowanej (na dole) wraz z odpowiadającymi im atomowymi funkcjami modulacji Ψ oraz wektorami modulacji q [5].

W obrazie dyfrakcyjnym takiej fazy modulowanej obserwuje się refleksy główne (dla których $m = 0$) oraz satelitarne (dla których $m \neq 0$). Natężenia satelitów szybko spadają dla dużych m , tak że obserwowane piki dyfrakcyjne pozostają rozdzielone, choć wektory w postaci $[m_d q_d]$ mogą się dowolnie zbliżać do siebie. Gdy wszystkie składowe wektorów modulacji są liczbami wymiernymi, mamy do czynienia ze strukturą współmiernie modulowaną lub tzw. nadstrukturą.



Rysunek 8. Schemat obrazu dyfrakcyjnego struktury współmiernie modulowanej z wektorem $q = 0,25c^*$ (periodyczność przywrócona co 4 komórki elementarne). Widoczne jest równomierne rozmieszczenie trzech słabych refleksów satelitarnych między refleksami głównymi [11].



Rysunek 9. Schemat obrazu dyfrakcyjnego struktury niewspółmiernie modulowanej w której okres AMF nie odpowiada periodyczności struktury podstawowej. Refleksy satelitarne są rozmieszczone nieregularnie względem głównych [11].

W większości przypadków do opisu pozycji pików braggowskich wystarczy pojedynczy wektor modulacji, co sprowadza wzór (3) do postaci:

$$h = h_1 a_1^* + h_2 a_2^* + h_3 a_3^* + m q \quad (4)$$

Struktury niewspółmiernie modulowane nie spełniają wymogu zachowania symetrii translacyjnej stawianego kryształom, choć ich uporządkowanie dalekiego zasięgu jest równie doskonałe jak w przypadku zwykłych kryształów.

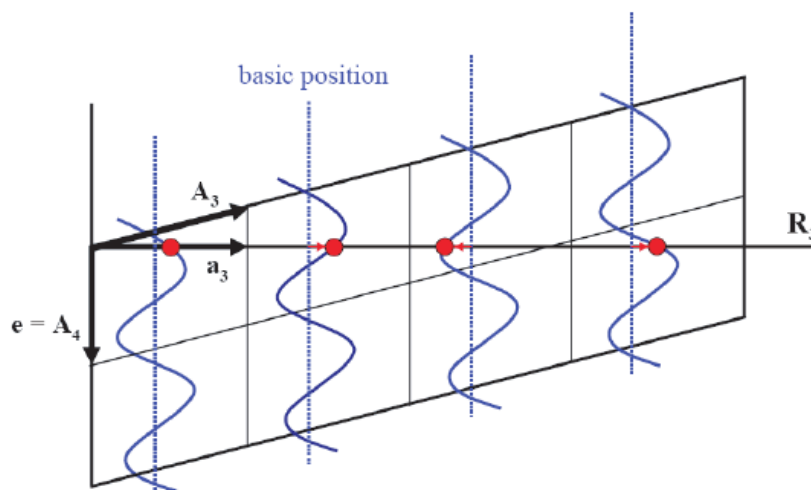
Ze względu na podłoże zjawiska można wyodrębnić kilka typów modulacji struktury:

- (i) modulację przesunięciową - polegającą na okresowym przesunięciu od pozycji atomowych struktury podstawowej;
- (ii) modulację obsadzeniową - pozycje atomowe struktury podstawowej są zajmowane zgodnie z okresową funkcją prawdopodobieństwa;
- (iii) modulację magnetyczną - następuje periodyczna zmiana orientacji atomowego momentu magnetycznego;
- (iv) strukturę poprzeraстанą - struktura podstawowa składa się z dwóch lub więcej podsystemów, z których każdy ma własną symetrię grupy przestrzennej (przy zaniedbaniu wzajemnych oddziaływań), taką, że w ogólności odpowiadające im sieci są niewspółmierne.

Podstawową konsekwencją złamania symetrii translacyjnej jest brak możliwości przypisania kryształu do jednej z 230 grup przestrzennych, co narzuca konieczność znalezienia innej metody opisu. Do opisu struktur modulowanych opracowano specjalną teorię nadprzestrzeni, której rozwój jest ściśle związany z nazwiskami Aloysio Jannera, Teda Janssena (1987) [26-28] oraz Petera de Wolffa (1974) [29]. Janner & Janssen sformułowali formalizm $(n + d)$ - wymiarowych grup przestrzennych, gdzie n określa wymiar struktury podstawowej, a d wymiar modulacji. W przypadku struktury z jednym wektorem modulacji, opis kryształu w nadprzestrzeni $(3 + 1)D$ opiera się na konstrukcji, która sztucznie przenosi wszystkie satelity w czwarty wymiar z przesunięciem proporcjonalnym do wskaźnika satelity (Rys. 10). Przyjmując, że modulacja jest sinusoidalną falą płaską, można opisać strukturę jako położenia atomów struktury podstawowej o trójwymiarowej symetrii przestrzenno-grupowej, okresowo przemieszczanych zgodnie z falą modulacyjną [30]. Położenie j -tego atomu w p -tej komórce elementarnej struktury podstawowej względem wybranego początku układu współrzędnych można zapisać tak jak we wzorze (5) jako:

$$\mathbf{r}(\mathbf{p}, j) = \mathbf{r}_j + \mathbf{p} + A \sin[(\mathbf{r}_j + \mathbf{p}) \cdot \mathbf{q} + \varphi_0] \quad (5)$$

gdzie \mathbf{p} - określa wektor sieciowy, \mathbf{r}_j - wektor położenia atomu w danej komórce elementarnej, A - amplitudę modulacji, \mathbf{q} - wektor modulacji, zaś φ_0 - fazą początkową modulacji.



Rysunek 10. Reprezentacja położenia atomu struktury modulowanej przesunięciowo w nadprzestrzeni $(3 + 1)d$: Pozycje atomów są opisane przez harmoniczną funkcją modulacji. W czerwonych punktach następuje przecięcie funkcji modulacji z przestrzenią rzeczywistą R_3 . Te przecięcia definiują strukturę kryształu [30].

W wielu przypadkach rozwiązanie struktury modulowanej sprowadza się do dwóch zasadniczych kroków. W pierwszym z nich do rozwiązania są używane tylko główne refleksy dając strukturę średnią [31]. Standardowe metody rozwiązywania uogólnione na struktury modulowane zostały zaimplementowane w niektórych programach.. Jednym z nich jest program DIMS [32] napisanego przez Fan Hai-fu wykorzystujący metody bezpośrednio oparte na równaniu Sayre'a. Najbardziej obiecującą metodą do rozwiązywania *ab initio* silnie modulowanych struktur jest *charge flipping* zaimplementowany do programu Superflip i rozpowszechniany razem z programem Jana2006 [33]. W tej metodzie wstępne funkcje modulacji otrzymuje się bezpośrednio na wyjściu programu, którym jest mapa gęstości elektronowej w nadprzestrzeni. Niestety, oba te programy nie są przystosowane do rozwiązywania struktur krystalicznych białek.

Otrzymany model strukturalny zwykle jest niedoskonały i wykazuje takie cechy jak rozszczępienie pozycji atomowych lub zawyżone parametry ADP, wyrażone poprzez nierealnie duże elipsoidy. Podczas wstępnych stadiów udokładniania odszukuje się zmodulowane atomy, określa rodzaj modulacji oraz odszukuje wektor modulacji poprzez udokładnianie w małych, losowo wybranych przemieszczeniach. Choć nowoczesne oprogramowanie dyfraktometrów znacznie ułatwia etapy pomiaru danych, wskaźnikowania obrazów dyfrakcyjnych, przetwarzania danych i rozwiązywania struktury, analiza struktur modulowanych wciąż wymaga wiele uwagi. Trudności zaczynają się po etapie rozwiązywania, gdyż interpretacja danych wyjściowych z *charge flippingu* wymaga doświadczenia, manualnego przeglądania pików oraz analizy map gęstości elektronowej. Interpretacja ta jest kluczowa dla właściwego doboru parametrów modulacji i może być specyficzna dla różnych struktur modulowanych. Dobór ten jest ważny nie tylko

do udokładnienia struktury, ale przede wszystkim dla interpretacji wyników, gdzie dążymy do zrozumienia przyczyn, dla których struktura jest modulowana.

2.7.1. Struktury modulowane białek

Struktury modulowane zostały zaobserwowane i stosunkowo dobrze opisane w krystalografii małowcząsteczkowej, ale są one niezwykle rzadko spotykane w krystalografii białek. Prawdopodobnie wynika to z faktu, że takie przypadki, nawet jeśli się zdarzały, były w przeszłości odrzucane jako zbyt trudne, a dane jako nienadające się do przetwarzania. Możliwość istnienia niewspółmiernie modulowanych struktur białkowych oraz próby ich metodycznego badania zostały po raz pierwszy opisane przez Lovelace'a et al. w 2004 roku [34]. Pozostałe prace opisują dyfrakcję (3+1)-wymiarowych niewspółmiernie modulowanych kryształów kompleksu białkowego profilina-aktyna [35] oraz próby wprowadzenia opisu w przestrzeni wielowymiarowej do rozwiązania tej struktury [36-39]. Jedną z niewielu pomyślnie rozwiązanych zmodulowanych struktur krystalicznych białek jest struktura białka roślinnego Hyp-1 w kompleksie z barwnikiem fluorescencyjnym ANS, która została pomyślnie rozwiązana metodą podstawienia molekularnego z uwzględnieniem translacyjnej symetrii niekrystalograficznej (translational Non-Crystallographic Symmetry, tNCS) i opisana jako tetartoedrycznie zbliźniona faza współmiernie modulowana z siedmiokrotnie rozszerzoną komórką elementarną zawierającą 28 niezależnych cząsteczek białka [40,41]. Struktura została zdeponowana w bazie PDB z kodem 4N3E z końcowymi wskaźnikami rozbieżności $R = 0,223$ oraz $R_{free} = 0,278$ [42]. Zbliżony model współmiernie modulowanej struktury innego kryształu kompleksu Hyp-1/ANS z dziewięciokrotnym poszerzeniem komórki elementarnej opisano w 2021 roku. Praca J.J. Lovelace'a & G. Borgstahl [43] wskazuje na silne podobieństwo funkcji modulacji dwóch udokładnionych modeli kompleksu Hyp-1 i ANS z siedmio- oraz dziewięciokrotną modulacją struktury.

2.7.2. Czynniki przemieszczenia atomowego (ADPs) w modelach białek

Parametr ADP (Atomic Displacement Parameter), czasami nazywany również czynnikiem Debye'a-Wallera (DWF, Debye-Waller Factor), czynnikiem temperaturowym, parametrem drgań termicznych lub B-factorem jest stosowany w krystalografii białek do opisu tłumienia rozpraszania promieniowania X lub neutronów w wyniku ruchów termicznych molekuly [9,10]. Koncepcja ADP bazuje na obserwacjach Petera Debye'a z 1913 roku, który początkowo zakładał, że atomy w kryształach oscylują niezależnie wokół swojej pozycji równowagi i w konsekwencji zastosował do ich opisu klasyczne metody statystyczne. W odpowiedzi Ivar Waller zaproponował modyfikację metody Debye'a, polegającą na zastąpieniu współczynnika temperatury jego kwadratem. W ten sposób wprowadzono współczynniki Debye'a-Wallera opisane wzorem (6):

$$DWF = \exp(-B(\sin \theta / \lambda)^2), \quad (6)$$

Gdzie θ to kąt pod jakim pada promieniowanie, λ to długość fali, parametr $B = 8\pi^2\langle u^2 \rangle$, u określa chwilowe przesunięcie atomu z położenia równowagi i odzwierciedla amplitudę oscylacji atomu wokół jego pozycji równowagi w strukturze krystalicznej.

W krystalografii białek pojedyncza struktura krystaliczna białka odpowiada zwykle tylko jednej („zamrożonej”) konformacji [44]. Jeżeli w strukturze istnieją regiony o wysokiej elastyczności opisywane wysokimi wartościami ADP, prawdopodobnie w roztworze istnieje kilka różnych konformacji [45]. Analiza wartości takich ADPs wskazuje zatem badaczowi na konieczność przeprowadzenia dodatkowych badań strukturalnych we wskazanym obszarze struktury, np. w powiązaniu ze spektroskopią magnetycznego rezonansu jądrowego (NMR) lub metodami dynamiki molekularnej (MD). Struktury o niskiej rozdzielczości 3-5 Å charakteryzują się niekiedy wysokimi wartościami ADPs na poziomie aż 100-200 Å², co może prowadzić do błędnej interpretacji struktury. Z tego powodu struktury rentgenowskie o wysokiej rozdzielczości są niezbędne do uzyskania wiarygodnych ADPs [46]. Podczas interpretacji elastyczności i dynamiki, nie należy zapominać, że ADPs odzwierciedlają zarówno wibracje jak i zaburzenia statyczne, a zatem konieczne jest wykonanie wielokrotnych pomiarów w funkcji temperatury w celu oddzielenia tych dwóch efektów [47].

Systematyczna analiza wartości ADPs na podstawie modeli struktur zdeponowanych w Protein Data Bank (PDB) może być wykorzystana do identyfikacji fluktuacji atomów, łańcuchów bocznych, a nawet całych regionów struktury. Znormalizowane ADPs będące wypadkową wielu czynników, w tym rozdzielczości, wiązań chemicznych i doboru metody udokładniania biorą pod uwagę średnią wartość ADPs w danej strukturze ($\langle B \rangle$) oraz jego odchylenie standardowe σ (Wzór (7)):

$$B_{norm} = (B - \langle B \rangle) / \sigma \quad (7)$$

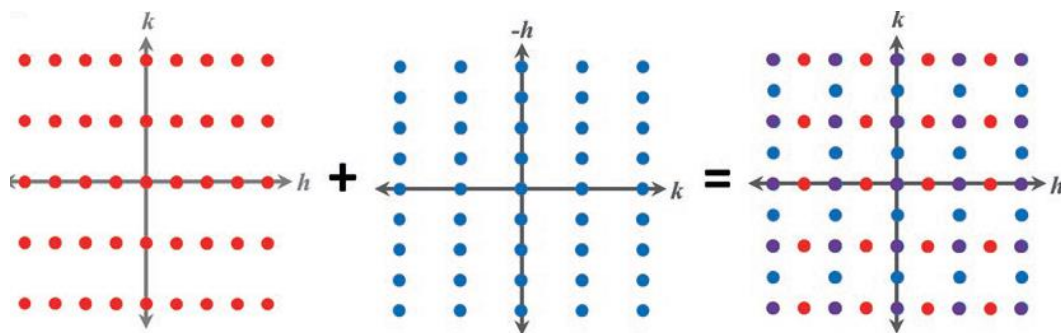
Analiza ta jest obecnie rutynowo stosowana w krystalografii białek do identyfikacji i interpretacji sztywności, elastyczności i/lub ruchów wewnętrznych w białkach [48,49].

2.8. Zbliźniaczenie

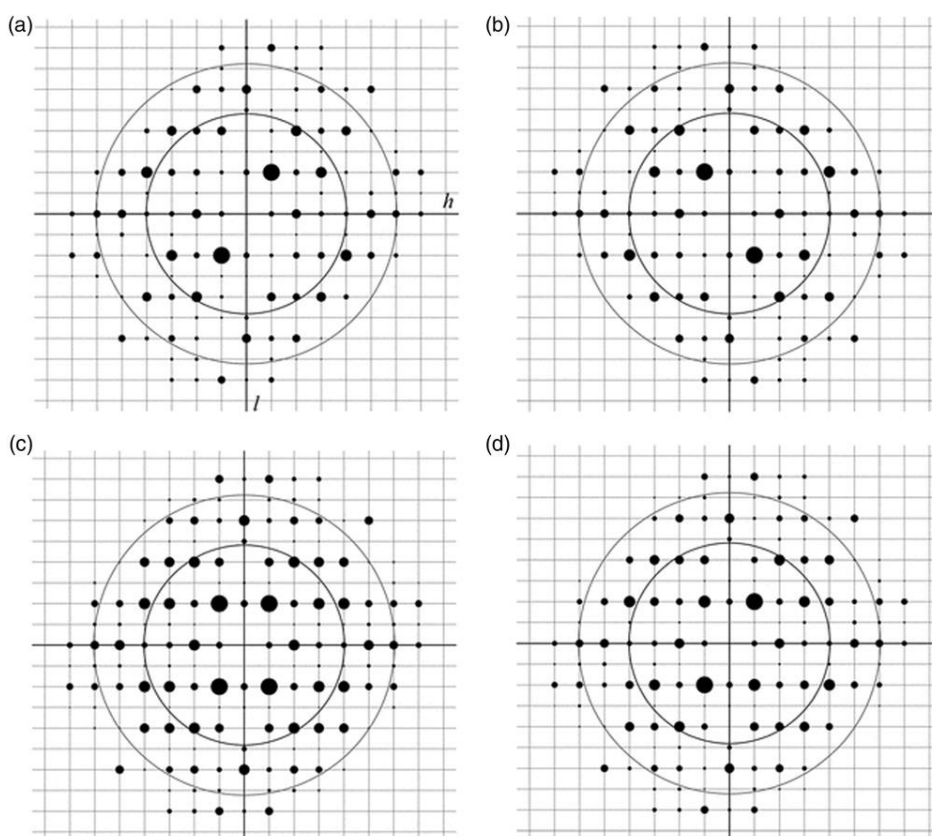
Zbliźniaczenie powstaje na etapie wzrostu kryształu, kiedy to jego powierzchnia przedstawiająca uporządkowany układ cząsteczek może działać jako dobry punkt nukleacji dla wzrostu dodatkowych, odrębnie zorientowanych domen krystalicznych powiązanych transformacją, która nie należy do symetrii jego grupy punktowej [9-11]. Zazwyczaj liczba różnych domen jest niewielka, a w większości przypadków występują tylko dwie różnie zorientowane domeny. Prowadzi to do istnienia dwóch wzajemnie obróconych obrazów dyfrakcyjnych i częściowego lub całkowitego nakładania się refleksów pochodzących od różnych domen, a w rezultacie komplikuje oszacowanie prawdziwych czynników struktury z obserwowanych natężeń (Rys. 11). Natężenie dla kryształu bliźniaczego odpowiada sumie natężeń fal odbitych od poszczególnych domen, podczas gdy w kryształ

niezblźnionym wszystkie komórki elementarne rozpraszają w fazie, a całkowity efekt odpowiada sumie amplitud [50].

Wyróżnia się kilka typów zblźnienienia: w zblźnienieniu meroedrycznym, operator zblźnienienia należy do tego samego układu krystalograficznego. Jeśli operator zblźnienienia należy do układu krystalograficznego o wyższej symetrii, co wynika z przypadkowej zależności parametrów komórki (np. w układzie romboidalnym $a = b$) mówimy o zblźnienieniu pseudomeroedrycznym. Gdy domeny bliźniacze przyjmują dwie orientacje, zblźnienienie jest hemiedryczne, gdy cztery - tetartoedryczne, osiem – ogdoedryczne, itd. (Rys. 12).



Rysunek 11. Efekt nakładania obrazów dyfrakcyjnych dwóch niezależnie rozpraszających promieniowanie rentgenowskie domen w kryształach zblźnionych. Finalny obraz dyfrakcyjny może wskazywać na wyższą symetrię struktury niż rzeczywista [9].

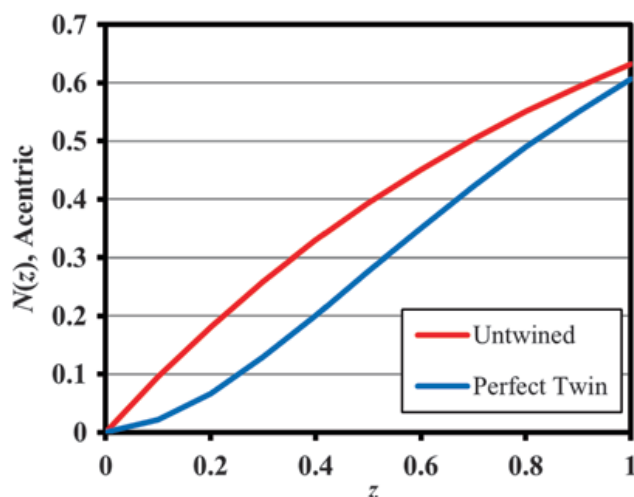


Rysunek 12. Konsekwencje działania operacją rotacji wokół osi 2-krotnej wzdłuż kierunku a na obraz dyfrakcyjny jednoskośnego kryształu o $\beta=90^\circ$. Pokazana jest tylko warstwica $h0l$: a) dyfrakcja z jednej domeny, b) dyfrakcja z innej domeny, c) superpozycja dyfrakcji z domenami o równym udziale, d) superpozycja (a) i (b) symulująca bliźniaka o 80% udziale pierwszej domeny i 20% drugiej domeny [50].

W przypadku częściowego nakładania refleksów standardowe programy do przetwarzania danych, takie jak CrysAlis (Rigaku Corporation, Tokyo, Japan.), SAINT (Bruker AXS Inc., Madison, Wisconsin, USA) czy EVAL [51], potrafią podczas przetwarzania danych rozpoznać nakładanie refleksów i kodują odpowiednie informacje (tj. które refleksy są nakładane i rozdzielane) w postaci tzw. formatu hklf5. W prostym przypadku zbliźniaczenie można łatwo rozpoznać na obrazie dyfrakcyjnym, rozwiązać strukturę standardowymi metodami biorąc pod uwagę jedną domenę lub dane niezblizniaczone, a następnie udokładnić model przy użyciu macierzy zbliźniaczenia lub danych w formacie hklf5 [50].

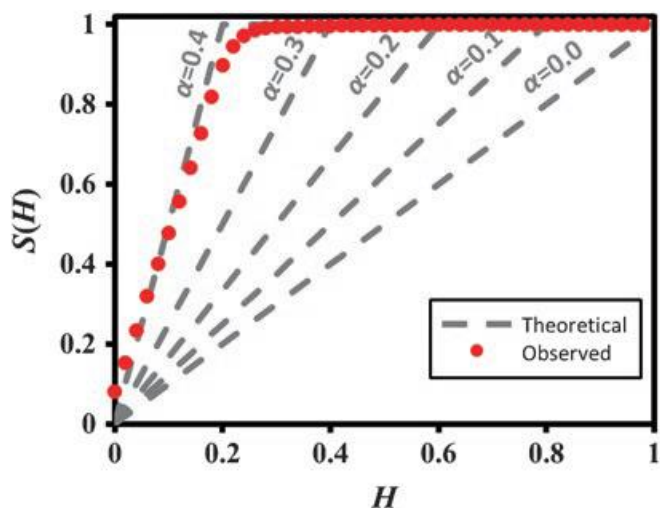
Znacznie trudniejsze do wykrycia zbliźniaczenie meroedryczne stanowi powszechny defekt kryształów makromolekuł, a identyfikacji problemu dokonuje się obserwując odstępstwa rozkładu intensywności od standardowego rozkładu Wilsona [52]. Analiza statystyczna intensywności refleksów wskazuje, że w przypadku kryształu bliźniaczego rozkład intensywności jest bardziej równomierny, z mniejszą ilością bardzo słabych i bardzo silnych odbić niż należałoby się spodziewać. Odchylenia od oczekiwanego rozkładu natężeń

mogą być zdiagnozowane przy pomocy skumulowanego rozkładu znormalizowanych intensywności, testem $N(z)$. W tym celu oblicza się znormalizowane intensywności refleksy (z) poprzez podzielenie wyniku każdego pomiarów przez średnią wartość intensywności dla danej powłoki rozdzielczości. W zależności od ułamka zblizniaczenia, skumulowany rozkład z , $N(z)$, przyjmuje różną postać. Dla kryształów niezblizniaczonych wykres ma charakter wykładniczy, natomiast dla kryształu idealnie bliźniaczego jest sigmoidalny jak na Rys. 13.



Rysunek 13. Test $N(z)$ dla refleksów od kryształów niezblizniaczonych i doskonale zblizniaczonych (równy udział dwóch domen, $\alpha = 0,5$) [9].

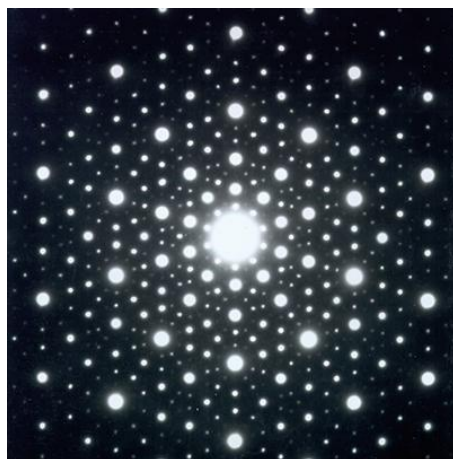
Inny test, $S(H)$ test, opisuje rozkład stosunku (H) różnicy do sumy intensywności refleksów powiązanych operacjami zblizniaczenia. Test umożliwia nie tylko potwierdzenie obecności lub braku zblizniaczenia, ale również oszacowanie ułamka zblizniaczenia, od którego zależy nachylenie krzywej rozkładu (Rys. 14). Najbardziej wiarygodny statystyczny test zblizniaczenia daje metoda zwana L-testem, polegająca na analizie skumulowanego rozkładu stosunku intensywności, $|L|$, który jest obliczany przez wybranie dwóch intensywności i podzielenie ich różnicy przez ich sumę. Różnica względem testu $S(H)$ polega na tym, że natężenia użyte do obliczenia H pochodzą od refleksów związanych z bliźniakami, podczas gdy natężenia (I_1 i I_2) użyte do obliczenia $|L|$ pochodzą od refleksów leżących blisko siebie w przestrzeni odwrotnej. Tym sposobem pomija się wpływ innych zaburzeń periodyczności, które komplikują identyfikację zblizniaczenia. Opisane testy zblizniaczenia zostały zastosowane w programie CTRUNCATE [53] dostępnym w pakiecie CCP4 lub phenix.xtriage [54] z pakietu PHENIX [55].



Rysunek 14. Skumulowany rozkład H , czyli test $S(H)$, z różnym przebiegiem wykresu zależnie od ułamka zbliźniaczenia [3].

2.9. Kwazikryształy

Klasyczna krytalografia dopuszcza istnienie osi symetrii wyłącznie o krotności 1, 2, 3, 4 i 6. Dzięki temu ograniczeniu zostaje zachowany warunek periodyczności rozumiany jako niezmienniczość całej struktury względem translacji sieciowych opisanych za pomocą wektora będącego liniową kombinacją trzech wektorów bazowych o współczynnikach całkowitych. W 1982 roku Dan Shechtman stwierdził na obrazach dyfrakcyjnych stopów Al z 10-14% Mn, Fe i Cr (Rys. 15) istnienie zabronionych przez klasyczną krytalografię dziesięciokrotnych osi symetrii [56]. Choć odkrycie to było długo kwestionowane, zaś opisująca je praca ukazała się finalnie dopiero w 1984 roku, badania Shechtmana przyczyniły się do zmiany definicji kryształu według Międzynarodowej Unii Krytalografii (IUCr). Kwazikryształy cechują się istnieniem dyskretnego obrazu dyfrakcyjnego przy jednoczesnym wykazywaniu zabronionej symetrii, dlatego nowa definicja kryształu to „Dowolny materiał charakteryzujący się wyraźnie dyskretnym obrazem dyfrakcyjnym”.



Rysunek 15. Obraz dyfrakcji elektronów dla ikozaedrycznego kwazikryształu zarejestrowany przez Dana Shechtmana [56].

W kwazikryształach symetria translacyjna w przestrzeni 3D jest całkowicie nieobecna, a dyskretny obraz dyfrakcyjny, jak to po raz pierwszy wykazał nie może być wskaźnikowany za pomocą trzech wskaźników hkl [25]. Struktury te mają jednak nierepetycyjny, długozasięgowy porządek orientacyjny, podobny w dwóch wymiarach do słynnych wzorów Penrose'a (parkietaż Penrose'a lub Penrose tillings) [57]. Oprócz braku periodyczności, kwazikryształy często charakteryzują się również obecnością symetrii rotacyjnej, np. pięcio-, dziesięcio- lub dwunastokrotnej, która jest zabroniona w klasycznych kryształach periodycznych. Skomplikowana teoria matematyczna opracowana przez Jannera i Janssen [26-28] wyjaśnia kwazikryształy poprzez przywołanie periodycznej struktury w wielowymiarowej (np. 6D) nadprzestrzeni, której rzut na trójwymiarową hiperpłaszczyznę (odpowiadającą znanej nam przestrzeni 3D) daje początek nieperiodycznej konstrukcji.

Jednowymiarowe kwazikryształy są praktycznie klasą struktur niewspółmiernie modulowanych [58]. W przestrzeni 2D kwazikryształy zawierają kwaziperiodyczne warstwy o ośmio-, dziesięcio- lub dwunastokrotnej symetrii rozmieszczone periodycznie jedna nad drugą wzdłuż wybranego kierunku. Do wskaźnikowania ich obrazów dyfrakcyjnych potrzeba pięciu liczb całkowitych. Symetria kwazikryształu (QC) może być opisana poprzez osadzenie przestrzeni fizycznej w n -wymiarowej (nD) nadprzestrzeni. W przypadku kwazikryształu dwuwymiarowego, trójwymiarowa przestrzeń fizyczna V_E jest osadzona w pięciowymiarowej przestrzeni V , która jest bezpośrednią sumą V_E i dwuwymiarowej przestrzeni wewnętrznej V_I [59,60]. Procedurę tę można uogólnić na przypadek kwazikryształów 3D, takich jak fazy ikozaedryczne, które powstają w wyniku projekcji do przestrzeni trójwymiarowej z przestrzeni 6-cio wymiarowej. Problem wyznaczania struktury polega na znalezieniu położenia i szczegółowego kształtu trójwymiarowych domen obsadzeniowych (powierzchni atomowych).

2.9.1. Kwazikryształ 1D – ciąg Fibonacciego

Prostym modelem jednowymiarowego kwazikryształu jest znany od XIII wieku ciąg Fibonacciego. Stanowi on przykład aperiodycznej struktury, w której zostało zachowane uporządkowanie dalekiego zasięgu [57]. W wersji liczbowej każdy wyraz ciągu $i > 2$ jest sumą dwóch poprzednich: 1, 1, 2, 3, 5, 8, itd. W wersji inflacyjnej ciąg jest zbudowany z dwóch odcinków, których długość pozostaje w stosunku $\tau/1$, gdzie τ jest liczbą niewymierną znaną jako „złota liczba” lub „złoty środek” ($(1 + \sqrt{5})/2 = 1,618\dots$). Łańcuch Fibonacciego można zbudować za pomocą reguły podstawiania $L \rightarrow LS$ i $S \rightarrow L$, gdzie L i S oznaczają odcinek długi i krótki:

$$S \rightarrow L$$

$$L \rightarrow L + S$$

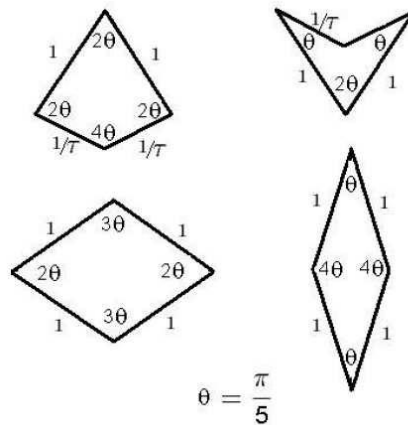
Otrzymana struktura 1D jest kwaziperiodyczna i zaczynając od S otrzymujemy postać:

$$\{S, L, LS, LSL, LSLLS, LSLLSLSL, \dots\}$$

Poprzez nieskończenie wielokrotne zastosowanie formuły otrzymuje się aperiodyczny wyraz będący modelem jednowymiarowego kwazikryształu. Zamiast atomów powtarzających się periodycznie w całej strukturze, jak w zwykłych kryształach, atomy w kwazikryształach są rozmieszczone w długich lub krótkich odległościach, które powtarzają się z niewspółmiernymi częstotliwościami [60,61].

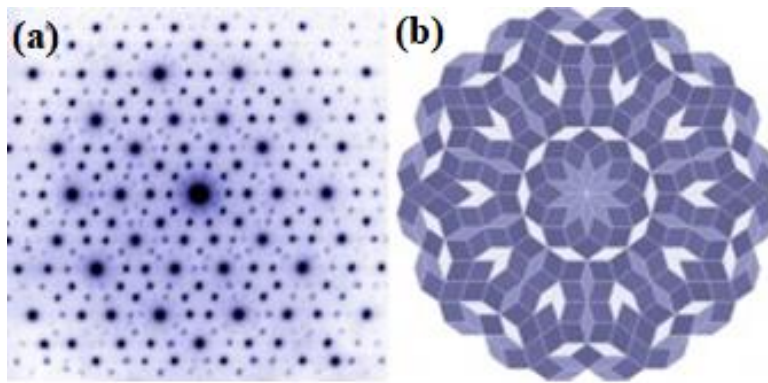
2.9.2. Parkietaż Penrose'a (Penrose tilings)

Na początku lat 70-tych XX wieku Roger Penrose pokazał, że można całkowicie pokryć płaszczyznę w sposób aperiodyczny przy użyciu skończonej liczby elementów [62]. Osiągnął to przy użyciu dwóch rodzajów figur dopasowanych regułą przylegania – najpierw były to „strzałka” i „latawiec”, a następnie dwa typy rombów, cienki i gruby (Rys. 16). Lokalna symetria parkietażu Penrose'a jest pięcio- bądź dziesięciokrotna, choć w zbiorze nieskończonym nie ma żadnej osi symetrii. Stosunek pola powierzchni rombu grubego do cienkiego wynosi τ .



Rysunek 16. Strzałka i latawiec (u góry) oraz gruby i cienki romb (na dole) używane w pokryciach Penrose'a [47].

W podejściu wielowymiarowym obserwowana w świecie rzeczywistym struktura 2D Penrose'a ma dziesięciokrotną oś obrazu dyfrakcyjnego i stanowi rzut 5D sieci regularnej na przestrzeń 2D (Rys. 17).



Rysunek 17. Obraz dyfrakcji elektronów na kwazikryształe stopu Zn-Mg-Ho (a) wraz z 2D pokryciem Penrose'a (b) [63].

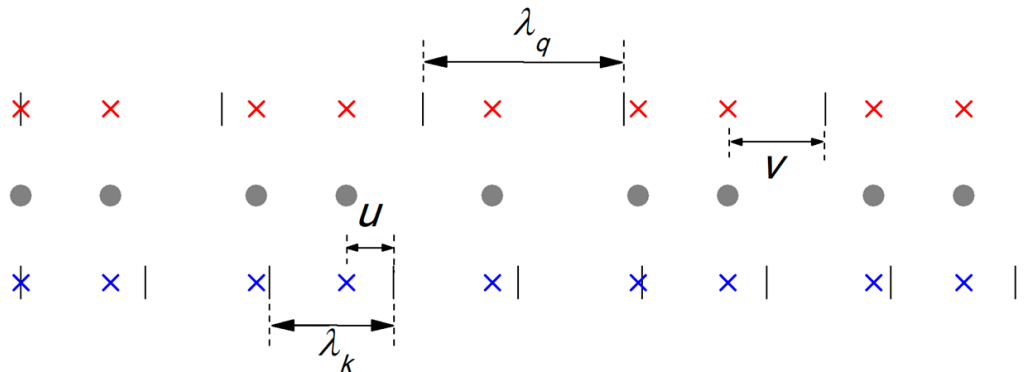
2.9.3. Metoda średniej komórki elementarnej (AUC)

Metoda analizy statystycznej stanowi jedną z nielicznych alternatyw dla analizy wielowymiarowej kwazikryształów i innych struktur cechujących się zaburzoną periodycznością. Technika ta działa tylko w przestrzeni rzeczywistej i polega na zastąpieniu pozycji atomów w przestrzeni fizycznej ich względnymi wartościami odniesionymi do pewnych periodycznych siatek referencyjnych [57-60]. Jedną z wad analizy wielowymiarowej jest to, że powierzchnie atomowe traktuje się jako niefizyczne obiekty utożsamiane z wielowymiarowymi atomami. Wobec tego zaletę metody statystycznej stanowi fakt, że utrzymuje ona wszystkie rozważania w przestrzeni fizycznej, a zatem przestrzeni, w jakiej atomy są domyślnie rozmieszczone w strukturze. Zasada konstrukcji średniej komórki elementarnej zostanie przedstawiona na przykładzie jednowymiarowym. Jak zostało stwierdzone powyżej, ciąg Fibonacciego stanowi dobry model kwazikryształu 1D o aperiodycznym rozmieszczeniu punktów o dalekozasięgowym uporządkowaniu. Układ punktów można rzutować na periodyczną sieć odniesienia o stałej sieciowej λ_k związanej z długością pewnego wektora k_0 :

$$\lambda_k = \frac{2\pi}{k_0} \quad (8)$$

Wektor k_0 dobieramy w taki sposób, aby był on podstawowym wektorem rozpraszania w przestrzeni odwrotnej. Rozkład $P(u)$ odległości punktów ciągu Fibonacciego względem sieci odniesienia tworzy *średnią komórkę elementarną*. Przedstawiony rozkład $P(u)$ nadaje się do celów statystycznych, ponieważ dla dostatecznie dużych układów jest dobrze zdefiniowany, ograniczony i w wielu przypadkach gęsty. Do opisu struktury aperiodycznej, w tym kwazikryształów, i wyznaczenia czynnika struktury potrzebna jest jeszcze jedna sieć odniesienia o stałej λ_q spełniającej zależność $\lambda_q = \frac{2\pi}{q_0}$. Wektory k_0 i q_0 są dobierane w taki sposób, aby stosunek ich długości wynosił τ . Wprowadzenie drugiej siatki odniesienia jest wymuszone przez aperiodyczność struktury. Względem drugiej sieci odniesienia można zbudować w ten sam sposób rozkład odległości punktów

$P(v)$, a wówczas pełna informacja o strukturze będzie przechowywana w funkcji pełnego rozkładu $P(u,v)$, która jest rozkładem pozycji atomowych rzutowanych na obie siatki odniesienia (Rys. 18).



Rysunek 18. Konstrukcja średniej komórki elementarnej (AUC) dla dwóch sieci referencyjnych o stałych sieciowych λ_k i λ_q . Szare atomy modelowej struktury krystalicznej przedstawione w 1D jako węzły ciągu Fibonacciego są rzutowane na sieci odniesienia (czerwone i niebieskie krzyże) [57,64].

Analityczne wyprowadzenie wzoru na czynnik struktury w podejściu statystycznym wykorzystuje własności transformacji Fouriera. Wychodząc z definicji, czynnik struktury w AUC skonstruowanej dla sieci referencyjnych po stałych λ_k i λ_q można zapisać jako:

$$\begin{aligned} F(k) = F(n, m) &= \sum_{j=1}^N \exp [(ink_0(\alpha\lambda_k + u_j + imq_0(\beta\lambda_q + v_j))] = \\ &= \sum_{j=1}^N \exp [ink_0u_j + imq_0v_j] \end{aligned} \quad (9)$$

gdzie α , β , n , m są liczbami całkowitymi, a u_j , v_j są współrzędnymi atomów w AUC względem sieci referencyjnych o stałych λ_k i λ_q . Wartości u_j oraz v_j gęsto wypełniają ograniczoną przestrzeń tworząc dobrze zdefiniowany rozkład $P(u, v)$ oraz są wzajemnie powiązane relacją skalowania $v(u)$. Tym samym można zastąpić sumowanie całkowaniem po AUC, wyrażając czynnik struktury za pomocą uogólnionej dwumodowej transformaty Fouriera:

$$F(k) = F(n, m) = \int_0^{\lambda_k} \int_0^{\lambda_q} P(u, v) \exp (ink_0u_j + imq_0v_j) dudv \quad (10)$$

Przedstawiona zasada konstrukcji rozkładu statystycznego nadaje się również do analizy struktur modulowanych, przy czym rozkład $P(u)$ charakteryzuje się znacznie bardziej złożonym kształtem. Wciąż jednak można wyprowadzić wzór na czynnik struktury dla struktur modulowanych z pojedynczą harmoniczną stosując funkcje Bessela:

$$F(k) = J_{-m} \left(k_0 A \left(n + \frac{m}{\xi} \right) \right) = J_{-m}(kA) \quad (11)$$

gdzie $k = k_0 \left(n + \frac{m}{\xi} \right)$, ξ jest czynnikiem skalującym określającym współmierność modulacji, a n i m oznaczają wskaźniki całkowite refleksów głównych i satelitarnych.

Obraz dyfrakcyjny składa się z serii pików zgrupowanych pod obwiednią zdefiniowaną funkcją Bessela kolejnych rzędów [64]. Metoda statystyczna pozwala wykazać bliskie podobieństwo między kwazikryształami a strukturami niewspółmiernie modulowanymi. W tym ujęciu kwazikryształ 1D, modelowany przez łańcuch Fibonacciego, może być rozumiany jako struktura modulowana z wieloma wektorami modulacji stanowiącymi rozwinięcie Fouriera podstawowego członu modulacji.

2.9.4. Struktury modulowane a kwazikrystały

Pojęcie kwaziperiodyczności dotyczy zarówno struktur niewspółmiernie modulowanych, jak i kwazikryształów. Obie klasy struktur dają dyskretne obrazy dyfrakcyjne przy jednoczesnym utrzymaniu dalekozasięgowego uporządkowania. W przypadku struktur niewspółmiernie modulowanych i kwazikryształów obserwuje się całkowity zanik symetrii translacyjnej. Na obrazach dyfrakcyjnych pojawiają się dodatkowe refleksy, które w przypadku struktur niewspółmiernie modulowanych wynikają z okresowej modulacji, natomiast w kwazikryształach są efektem obecności dodatkowych, zabronionych w krystalografii klasycznej elementów symetrii. Dodatkowe refleksy nie mogą zostać wywskaźnikowane przy użyciu trzech podstawowych wskaźników *hkl*.

Podstawowa różnica jest związana z faktem, że struktury niewspółmiernie modulowane są zasadniczo podstawowymi strukturami, które są okresowo zaburzane - "modulowane" - gdzie okres modulacji jest niewspółmierny z okresem struktury bazowej. Struktura kryształu jest w pełni zdeterminowana przez strukturę podstawową i falę (lub fale) modulacyjną. Gdy amplituda fali modulacji spada do zera, opis struktury niewspółmiernie modulowanej upraszcza się do opisu periodycznego. Tymczasem kwaziperiodyczne kryształy posiadające symetrie zabronione dla struktur periodycznych nie mogą być utworzone poprzez modyfikację podstawowej struktury periodycznej o tej samej symetrii.

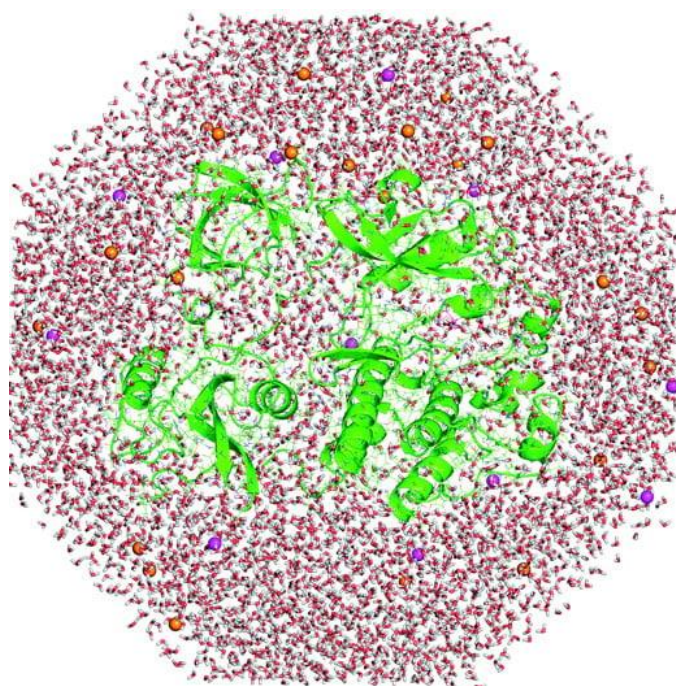
2.10. Symulacje dynamiki molekularnej struktur białek

Eksperymentalne metody badania struktury białek, a więc głównie krystalografia rentgenowska, spektroskopia NMR (Nuclear Magnetic Resonance Spectroscopy) i kriomikroskopia elektronowa (cryo-EM, Cryogenic Electron Microscopy), są obarczone kilkoma ograniczeniami. W przypadku krystalografii, pierwsze z nich wynika z konieczności otrzymania dobrze rozprasających kryształów do rejestracji obrazu dyfrakcyjnego [11]. Zamrażanie kryształów w warunkach kriogenicznych w ciekłym azocie sprawia, że konformacja badanej cząsteczki może odbiegać od konformacji w stanie natywnym i formie biologicznie czynnej białka. Spektroskopia NMR zawęży wielkość badanych struktur do białek liczących nie więcej niż 150-200 aminokwasów oraz uniemożliwia precyzyjną analizę większych układów. Dla analogicznych struktur mierzonych metodami krystalografii rentgenowskiej oraz przy użyciu spektroskopii NMR średnie odchylenie kwadratowe położenia atomów ((root-mean-square deviation, RMSD) sięga ok. 2,0 Å, co wskazuje na znaczące różnice w proponowanej konformacji.

Fundamentalne zrozumienie działania białek wymaga wskazania związku pomiędzy trójwymiarową strukturą, uzyskaną metodami krystalografii rentgenowskiej, cryo-EM lub spektroskopii NMR a dynamiką, którą znacznie trudniej jest zbadać w sposób eksperymentalny. Symulacje dynamiki molekularnej (MD) zapewniają wgląd w dynamikę struktur białkowych oraz umożliwiają eksplorację krajobrazu energii konformacyjnej dostępnej dla cząsteczek białek i przewidywanie trajektorii ruchu atomów w czasie, w oparciu o model oddziaływań międzyatomowych [65-66]. Pierwsza symulacja dynamiki molekularnej białka została opisana w 1977 roku i składała się z trajektorii 9,2 ps dla małego białka w próżni [67]. Rosnące zainteresowanie symulacjami MD wynika przede wszystkim z gwałtownej eksplozji eksperymentalnych struktur pewnych klas cząsteczek (białek błonowych, kanałów jonowych czy transporterów neurotransmiterów) krytycznych w neuronauce. Te eksperymentalne struktury stanowią punkt wyjścia dla symulacji MD, które mogą pomóc rozwiązać strukturalnych pytania dotyczące funkcjonowania kluczowych białek neuronalnych, mechanizmów ich patologicznej agregacji w pewnych warunkach, czy projektowania [68,69].

Po drugie, w ciągu ostatnich kilku lat symulacje MD stały się znacznie bardziej wydajne i dostępne dzięki rozwojowi sprzętu komputerowego, szczególnie procesorów graficznych (GPU) [70]. Nastąpił również szybki rozwój pakietów oprogramowania do prowadzenia symulacji MD.

Ogólna idea symulacji MD jest prosta - należy obliczyć siłę wywieraną na każdy atom przez wszystkie inne atomy, znając pozycje wszystkich atomów w układzie biomolekularnym (np. białko otoczone wodą jak na Rys. 19 i być może warstwą lipidów). Można więc wykorzystać prawa dynamiki Newtona do przewidywania przestrzennego położenia każdego atomu w funkcji czasu [71]. Używając odpowiednio dobranego kroku czasowego, można wielokrotnie obliczać siły działające na każdy atom, a następnie użyć tych sił do aktualizacji pozycji i prędkości każdego atomu. Otrzymana trajektoria jest w istocie trójwymiarowym filmem opisującym konfigurację układu na poziomie atomowym w każdym punkcie symulowanego przedziału czasu [72].



Rysunek 19. Symulacja białka solwowanego przez ok. 15 000 cząsteczek wody (atomy tlenu są czerwone, a atomy wodoru białe) oraz jonów potasu i chlorku (odpowiednio fioletowe i pomarańczowe kule) [69].

Siły w symulacji MD są obliczane przy użyciu modelu znanego jako pole siłowe mechaniki molekularnej, które jest zazwyczaj dopasowane do wyników obliczeń kwantowo-mechanicznych oraz pewnych pomiarów eksperymentalnych. Typowe pole siłowe zawiera człony uwzględniające oddziaływania elektrostatyczne (kulombowskie) pomiędzy atomami, człony odpowiadające za modelowanie preferowanej długości każdego wiązania kowalencyjnego oraz dodatkowe wyrazy obejmujące kilka rodzajów oddziaływań międzyatomowych. Wśród pól siłowych najczęściej stosowanych w symulacjach MD białek należy wymienić CHARMM [73], NAMD [74], GROMACS [75], AMBER [76] oraz OPLS [77-82]. Pola sił są z natury przybliżone, więc wprowadzona przez nie niepewność powinna być brana pod uwagę przy analizie wyników symulacji. Kroki czasowe w symulacji MD muszą być odpowiednio krótkie, zwykle rzędu kilku femtosekund (10^{-15} fs). Większość zdarzeń interesujących z punktu widzenia biochemii - na przykład ważne funkcjonalnie zmiany strukturalne w białkach - zachodzi w skali czasowej nanosekund, mikrosekund lub dłuższej, a zatem przeciętna symulacja obejmuje miliony lub miliardy kroków czasowych. Symulacje MD mogą dostarczyć szeregu informacji dotyczących:

- oceny mobilności lub elastyczności różnych regionów biomolekuły [83,84];
- sprawdzenia poprawności modelowanej struktury lub nawet do jej korekty i udoładnienia [85];
- budowy lub udoskonalania modeli strukturalnych w oparciu o dane eksperymentalne;

- określenia reakcji układu biomolekularnego na pewne perturbacje, np. związane z dodaniem, usunięciem lub zastąpieniem liganda [86];
- obserwacji przebiegu procesów biomolekularnych, szczególnie wiązania ligandów, zmian konformacyjnych wywołanych przyłączeniem ligandów lub przyłożeniem napięcia, mechanizmów zwijania białek czy transportu błonowego.

Zaletami symulacji MD są, po pierwsze, możliwość określenia położenia i trajektorii każdego atomu w każdym punkcie czasu, co jest bardzo trudne w przypadku niemal każdej techniki eksperymentalnej z wyjątkiem użycia źródeł XFEL (X-ray Free Electron Laser) [87]. Po drugie, warunki symulacji są dokładnie znane i mogą być ściśle kontrolowane: wstępna konformacja białka, dobór związanych z nim ligandów, mutacje lub modyfikacje potranslacyjne, obecność innych cząsteczek w otoczeniu, stan protonacji, temperatura, napięcie membranowe i wiele innych [88,89].

3. Cel rozprawy

Głównym celem niniejszej rozprawy doktorskiej było udokładnienie modelu struktury krystalicznej kompleksu białkowego Hyp-1/ANS z dziewięciokrotną modulacją wzdłuż kierunku c w podejściu superkomórki oraz przygotowanie oprogramowania pozwalającego przeprowadzić podobne procedurę w bardziej skomplikowanym opisie w przestrzeni wielowymiarowej (3+1)D. Podjęłam również próbę zbadania uporządkowania atomowego w modulowanych strukturach białkowych oraz opracowania oryginalnego oprogramowania do szczegółowej analizy struktur o zaburzonej periodyczności. Jak zaznaczono, modulacja struktury krystalicznej jest spotykana w krystalografii małowymiarowej, ale stanowi ewenement w krystalografii białek. Zastosowanie superkomórki stanowi zweryfikowaną metodę postępowania podczas udokładniania nielicznych przypadków modulowanych struktur białkowych, jednakże wiąże się z ogromnym wzrostem liczby parametrów wymaganych podczas udokładniania oraz żmudnym i pracochłonnym procesem poprawiania przygotowanego modelu struktury. Podczas formułowania celu niniejszej rozprawy postawiono następujące pytania:

- Jakie są główne wyzwania związane z rozwiązaniem struktury modulowanych kompleksów białkowych, takich jak Hyp-1/ANS?
- Jakie są obecnie dostępne narzędzia do analizy strukturalnej modulowanych struktur białkowych i jakie są ich ograniczenia?
- Jakie są nowe podejścia do rozwiązania problemów związanych z analizą strukturalną modulowanych struktur białkowych?
- W jaki sposób przygotować oprogramowanie do analizy strukturalnej modulowanych struktur białkowych, które będzie przekształcać dane do przestrzeni wielowymiarowej, poddawać je udokładnieniu i zachowywać stereochemiczną poprawność modelu?
- Jaki wpływ mają warunki krystalizacji i wybór liganda na modulację kompleksu Hyp-1/ANS?

Wśród najważniejszych celów badawczych można wymienić:

- Udokładnienie struktury kompleksu Hyp-1/ANS z dziewięciokrotną modulacją w ujęciu superkomórki oraz jego pogłębioną analizę strukturalną w celu zbadania mechanizmów modulacji oraz sposobu jej opisu.
- Rozwój modelu struktury modulowanej kompleksów Hyp-1/ANS w oparciu o analizę wielowymiarową bez założeń dotyczących komórek nadstruktury lub współmierności wektora modulacji.
- Implementacja i doskonalenie nowego oprogramowania do udokładniania struktury w oparciu o eksperymentalne dane dyfrakcji promieniowania X. Opracowanie narzędzia do analizy i udokładniania struktury z wykorzystaniem korekty zaburzeń strukturalnych, efektów zbliżniaczenia czy nieporządku fononowego wynikającego z drgań termicznych sieci krystalicznej.

- Wykorzystanie symulacji metodami dynamiki molekularnej w celu określenia różnic w drganiach termicznych atomów oraz konformacji łańcuchów bocznych reszt aminokwasowych eksperymentalnej i zdeponowanej w bazie PDB struktury białka Hyp-1. Celem było porównanie podejścia opartego bezpośrednio na dynamice z eksperymentem, a także próba walidacji niepewnych konformacji łańcuchów bocznych.

Długofalowym celem rozprawy doktorskiej ma być pogłębienie zrozumienia fundamentalnych mechanizmów w modulowanych systemach makromolekularnych oraz opracowania oprogramowania, które umożliwi dalsze badania tych ciekawych struktur.

4. Materiały i metody

4.1. Kompleks białkowy Hyp-1/ANS

Badany kompleks został otrzymany przez dr Joannę Śliwiak i obejmował białko Hyp-1 z dziurawca zwyczajnego (*Hypericum perforatum*) krystalizowane po preinkubacji z sulfonianem 8-anilino-1-naftalenowym(ANS) wykazującym właściwości fluorescencyjne [40-42]. Hyp-1 należy do roślinnych białek związanych z patogenezą klasy 10 (Pathogenesis Related, PR-10), znanych również jako indukowalne białka związane z obroną, która może być wywołana może być indukowana działaniem fitohormonów takich jak kwas salicylowy, kwas jasmonowy lub etylen bądź poprzez bodziec mechaniczny, np. zranienie [90]. Ich cechą charakterystyczną jest, że wykazują one aktywność przeciwdrobnoustrojową, związaną m.in. z hydrolizą ściany komórkowej bakterii, toksycznością kontaktową lub prawdopodobnie zaangażowaniem w sygnalizację obronną. Białka związane z patogenezą zostały sklasyfikowane w 17 podklasach według ich aktywności biologicznej, właściwości fizykochemicznych oraz homologii sekwencji. Spośród nich, szczególnie białka grupy 10, PR-10, są zaangażowane przede wszystkim w mechanizmy obronne, ponieważ ich wytwarzanie jest indukowane przez inwazję patogenów oraz czynniki abiotyczne, takie jak susza, zimno, zasolenie, obecność miedzi, stres oksydacyjny, promieniowanie UV lub zranienie [90]. Po wielu dyskusjach w literaturze na temat biologicznej funkcji tych białek, najbardziej prawdopodobna wydaje się ich rola w wiązaniu i transporcie mediatorów metabolicznych, takich jak fitohormony, lipidy czy flawonoidy [91-94].

Tej hipotezie „pojemnika” służącego transportowi aktywnych biologicznie molekuł sprzyja kanoniczna struktura przestrzenna zwoju PR-10, który zbudowany jest wokół wewnętrznej hydrofobowej komory utworzonej przez złożony z siedmiu antyrównoległych łańcuchów arkusz β , owinięty wokół długiej i elastycznej C-terminalnej helisy $\alpha 3$ wykazującej nie tylko zmienność konformacyjną, ale i sekwencyjną [95-97]. Wewnętrzna komora hydrofobowa ma dwa wyraźne główne wejścia: wejście E1 utworzone przez pętle (oznaczone jako L3, L5, L7) i helisę $\alpha 3$, oraz wejście E2 utworzone pomiędzy tą samą helisą α a pierwszą nicią ($\beta 1$) arkusza β [110-112]. W obrębie białka Hyp-1 wyodrębniono charakterystyczny zestaw trzech miejsc zdolnych do wiązania ligandów, składający się z dwu wewnętrznych, ale dobrze oddzielonych komór (miejsca 1 i 2) oraz umiejscowionej na powierzchni głębokiej kieszeni (miejsce 3). Krystalizacja białka Hyp-1 w obecności cząsteczek ANS prowadzi do wytworzenia tetartoedrycznie zbliżniaczonych kryształów z siedmiokrotną modulacją wzdłuż wydłużonej osi c komórki elementarnej (dalej opisano ten kompleks jako 7Hyp/ANS). Powtórzenie krystalizacji Hyp-1 w obecności mieszaniny ANS i melatoniny (MEL) pozwoliło uzyskać następną modulowaną fazę kompleksu Hyp-1/ANS, w której podobny podstawowy motyw strukturalny powtarza się wzdłuż osi c dziewięć razy, dla którego przyjęto nazwę 9Hyp/ANS. Kryształy kompleksu Hyp-1/ANS otrzymano w laboratoriach Instytutu Chemii Bioorganicznej PAN w Poznaniu [40].

Szczegóły dotyczące przygotowania próbki, zastosowanych odczynników, warunków krystalizacji i uzyskanych kryształów zostały opisane w podrozdziale 5.1.

4.2. Dane eksperymentalne

Eksperymentalne dane dyfrakcji rentgenowskiej o rozdzielczości 2,4 Å dla kryształu 7Hyp/ANS zostały zmierzone przez dr Joannę Śliwiak na linii SER-CAT 19ID w APS (Advanced Photon Source) w Argonne National Laboratory (ANL) w USA, a następnie przetworzone przez prof. Mariusza Jaskólskiego i dra Mirosława Gilskiego przy użyciu programu HKL-2000 [101]. Początkowe procesowanie danych okazało się zadowalające w grupie przestrzennej P422, z R_{merge} wynoszącym 0,075 (Tabela 1). Parametr R_{merge} wskazuje jak bardzo pomiary tego samego refleksu różnią się intensywnością od średniej intensywności tego refleksu, a więc stanowi miarę zgodności wielu pomiarów tej samej wielkości fizycznej.

Dane dyfrakcyjne kryształów 9Hyp/ANS zostały zmierzone przez dr Joannę Śliwiak na linii 14.2 synchrotronu BESSY II. Dane zostały przetworzone do rozdzielczości 2,3 Å za pomocą programu HKL-3000 [102], a następnie przeskalowane w symetrii trójskośnej z R_{merge} równym 0,14. Na podstawie wstępnych wyników skalowania stwierdzono, że parametry a i b komórki elementarnej są równe i w przybliżeniu takie same jak dla wcześniej scharakteryzowanego kryształu 7Hyp/ANS, przy czym parametr c wynoszący 386 Å był 9/7 razy dłuższy niż dla 7Hyp/ANS. Już na tak wczesnym etapie analizy danych dyfrakcyjnych stanowiło to przesłankę, że otrzymany kryształ wykazuje modulowaną nadstrukturę z dziewięcioma, a nie siedmioma motywami strukturalnymi powtarzających się w kierunku osi c .

4.3. Rozwiązanie struktury

4.3.1. Phaser

Strukturę kryształu 7Hyp/ANS rozwiązano przy pomocy programu Phaser [103] dostępnego w ramach pakietu CCP4. Jest to program do rozwiązywania problemu fazowego dla makrocząsteczkowych struktur krystalicznych wykorzystujący metody podstawienia molekularnego oraz największej wiarygodności (metody *maximum likelihood*, ML). Oparta na prawdopodobieństwie warunkowym metoda ML poszukuje najlepszego modelu w oparciu o dane eksperymentalne oraz wiedzę uprzednią (*priors*). Algorytmy zastosowane w programie Phaser sprawdzają się w odróżnianiu poprawnych rozwiązań od szumu, natomiast w przypadku eksperymentalnego fazowania algorytmy Phasera dają lepsze fazy) niż inne algorytmy wykorzystujące średnie F i różnice anomalne ΔF . Podczas pracy nad rozwiązaniem struktury białka Hyp-1/ANS zastosowanie standardowych metod podstawienia molekularnego w połączeniu z metodą największej wiarygodności (ML) było niemożliwe, gdyż ich podstawowym założeniem jest jednolity rozkład czynnika struktury. Z tego powodu strukturę rozwiązano po dostosowaniu algorytmów ML Phasera do statystycznych efektów wynikających z niekrystalograficznej symetrii translacyjnej (tNCS).

4.4. Udokładnienie modelu

4.4.1. REFMAC5

Udokładnienie modelu struktury krystalicznej kompleksu białka Hyp-1/ANS przeprowadzono przy pomocy programu REFMAC5 [104] z pakietu CCP4. REFMAC5 wykorzystuje różne funkcje prawdopodobieństwa w zależności od rodzaju zastosowanych danych dyfrakcyjnych (amplitudy lub intensywności), obecności zbliźniaczenia oraz dostępności anomalnych danych dyfrakcyjnych SAD/SIRAS. W toku udokładnienia następuje dopasowanie parametrów modelu metodami ML do gęstości elektronowej, po czym liczone są mapy gęstości elektronowej, zwykle 2Fo-Fc (lub 2mFo-DFc) oraz mapy różnicowej Fo-Fc (lub mFo-DFc).

Mapy gęstości elektronowej umożliwiają badaczowi znalezienie obszarów, w których model nie odpowiada gęstości elektronowej. REFMAC wykorzystuje dwa podstawowe formaty danych wejściowych: plik *.pdb zawierający współrzędne atomów cząsteczki modelu w odpowiedniej orientacji względem wybranej komórki elementarnej oraz plik *.mtz określający czynniki struktury badanego kryształu.

Ponadto REFMAC5 zapewnia chemiczną i strukturalną integralność modelu poprzez użyciu więzów stereochemicznych i możliwości wyboru parametryzacji modelu. Przy rozdzielczości co najmniej 4 Å program umożliwia budowę wiarygodnego modelu struktury krystalicznej dzięki narzędziom udokładniania niskorozdzielczego, do których należą ograniczenia struktury drugorzędowej, ograniczenia do znanych struktur homologicznych, automatyczne globalne i lokalne więzy NCS (Non-Crystallographic Symmetry), więzy „jelly-body” (nałożenie więzów na część odległości międzyatomowych w modelu i ograniczenie niezależnego ruchu atomów w obrębie pewnego promienia) oraz zastosowanie nowych więzów dalekiego zasięgu na parametrach przemieszczenia atomowego (ADP) opartych na dywergencji Kullbacka-Leiblera [104]. REFMAC5. REFMAC5 oferuje dodatkowo parametryzację TLS (Translation-Libration-Screw motion, modelowanie anizotropowych ADP, dla fragmentów makromolekuły traktowanych jak bryła sztywna) oraz, jeśli dostępne są dane o wysokiej rozdzielczości, możliwość szybkiego udokładnienia anizotropowych ADP atomów. REFMAC5. REFMAC5 umożliwia w pełni automatyzowane udokładnianie struktury zbliźnionego wraz z detekcją macierzy zbliźniaczenia oraz udziału poszczególnych domen. Oceny jakości dopasowania modelu do danych dyfrakcyjnych dokonuje się przy pomocy standardowych parametrów zgodności R , R_{free} i RMSD.

4.5. Wizualizacja danych

4.5.1. Coot

Coot to program molekularno-graficzny umożliwiający budowę modeli w mapach gęstości elektronowej oraz walidację struktur makromolekuł. W przypadku struktur o rozdzielczości poniżej 2,5 Å Coot [15,105] pomaga uzupełnić wstępne modele wygenerowane metodami podstawienia molekularnego lub zbudować automatyczne modele na podstawie danych dostępnych bezpośrednio z pomiarów eksperymentalnych

(w przypadku wyższej rozdzielczości umożliwiają to programy ARP/wARP [14], RESOLVE [106] lub Buccaneer [15]). Po wczytaniu pliku *.pdb* zawierającego współrzędne atomów oraz związanego z nim pliku *.mtz* w oknie użytkownika wyświetlają się nałożone mapy gęstości elektronowej, mapy różnicowe i modele atomowe. Obszary o ładunku dodatnim (zielone) i ujemnym (czerwone) wskazują położenia, w których wpasowany model powinien mieć więcej elektronów (obszary zielone) lub mniej elektronów (obszary czerwone). Coot umożliwia ręczne manipulowanie modelami, takie jak: idealizacja, udokładnianie w przestrzeni rzeczywistej, ręczne obracanie/zamiana reszt oraz zmiana orientacji łańcucha polipeptydowego, dopasowywanie brył sztywnych, wyszukiwanie ligandów i rozpuszczalnika, modelowanie rotamerów (określone konformacje łańcuchów bocznych aminokwasów) oraz generowanie wykresów Ramachandrana. Posiada również wbudowane narzędzia do walidacji modelu. Udokładnianie następuje w sposób rekurencyjny – po każdej serii korekt następuje ponowne udokładnianie oraz obliczanie nowych map gęstości elektronowej, które znów poddaje się inspekcji, aby wskazać obszary, które wymagają dodatkowej uwagi i udokładnienia. Pod koniec udokładnienia następuje wbudowanie do modelu cząsteczek wody.

4.5.2. PyMOL

Do ilustracji modeli białka w formacie *.pdf* oraz przygotowania grafiki na potrzeby niniejszej rozprawy doktorskiej wykorzystano program PyMOL (The PyMOL Molecular Graphics System, Schrödinger, LLC) będący narzędziem szeroko stosowanym do trójwymiarowej wizualizacji białek, kwasów nukleinowych, małych cząsteczek, gęstości elektronowej, powierzchni oraz trajektorii. Oprogramowanie to umożliwia również edycję cząsteczek, generowanie obrazów techniką *ray tracing* oraz tworzenie filmów. PyMOL został oparty na kodzie przygotowanym w Pythonie, wraz z dostępem do licznych wtyczek i bibliotek Pythona. Wszechstronność PyMOLa umożliwia jego zastosowanie w modelowaniu molekularnym, wizualizacji i analizie strukturalnej modeli, modelowaniu białek i ligandów oraz badaniach przesiewowych leków.

4.5.3. UCSF Chimera

Drugim programem stosowanym do przygotowania ilustracji do niniejszej rozprawy jest UCSF Chimera [107] - wysoce rozszerzalny program do interaktywnej wizualizacji i analizy struktur molekularnych oraz powiązanych z nimi danych, szczególnie map gęstości elektronowej, dopasowania sekwencji, monitorowania wyników dokowania ligandów, śledzenia trajektorii i zespołów konformacyjnych. Środowisko to umożliwia również generowanie wysokiej jakości obrazów i animacji.

4.6. Testy zbliżniczenia

4.6.1. TRUNCATE

W programie TRUNCATE [53] przeprowadzono testy statystyczne rozkładu intensywności refleksów pod kątem badania zbliżniczenia struktury kompleksu białkowego Hyp-1/ANS. Standardowe zastosowanie programu obejmuje wczytanie pliku

z uśrednionymi natężeniami (plik wynikowy generowany w programach SCALA, SCALEPACK2MTZ lub DTREK2MTZ), a następnie wypisanie pliku zawierającego średnie amplitudy i oryginalne natężenia. Amplitudy są umieszczane w przybliżonej skali bezwzględnej przy użyciu współczynnika skali wziętego z wykresu Wilsona [52]. Program TRUNCATE wykorzystuje dwa zasadnicze sposoby obliczania amplitud z natężeń. Pierwszy z nich to po prostu obliczenie pierwiastka kwadratowego z natężeń, przy czym wszystkim ujemnym intensywnościom przypisuje się wartości zerowe. Druga procedura „truncate” oblicza najlepsze oszacowanie czynników struktury F z I , $\sigma(I)$ i rozkładu natężeń w powłokach rozdzielczości. Na skutek tego wymusza się, aby wszystkie negatywne obserwacje były dodatnie, a także zawyża się najniższe intensywności (mniejsze niż około 3σ). TRUNCATE może być używany do tworzenia użytecznych statystyk rozkładów intensywności, testów $N(z)$ oraz $S(H)$. Mogą one wskazywać prawdopodobieństwo, że dane są zbliżnione. Przy pomocy programu TRUNCATE wygenerowano wykresy skumulowanego rozkładu natężeń $N(z)$ oraz $S(H)$ oraz zaobserwowano oznaki bliźniaczości.

4.7. Symulacje białka Hyp-1 metodami dynamiki molekularnej

4.7.1. GROMACS

Symulacje modelu białka Hyp-1 metodami dynamiki molekularnej przeprowadzono przy pomocy ogólnodostępnego pakietu oprogramowania chemicznego GROMACS [76]. Program może być stosowany zarówno dla molekuł biologicznych (np. białka, lipidy) oraz systemów nie-biologicznych (np. polimerów). W niniejszej pracy pakiet GROMACS w wersji 5.0 wykorzystano do przeprowadzenia symulacji dynamiki molekularnej modeli badawczych, analizy uzyskanych trajektorii oraz analizy rozkładów kątów torsyjnych w modelu struktury białka Hyp1-/ANS.

4.8. Program do udokładniania modelu białka

4.8.1. Optymalizacja problemu przy użyciu algorytmu metaheurystycznego MEIGO

Pakiet oprogramowania MEIGO [108] (MEtaheuristics for systems biology and bioinformatics Global Optimization) stanowił metodę matematycznej optymalizacji w celu określenia minimum i punktu zbieżności udokładnianego modelu z doświadczalnymi amplitudami czynników struktury. Klasycznie w krystalografii białek stosuje się metody LSQ (least squares, metoda najmniejszych kwadratów), natomiast ostatnio wprowadzono metody optymalizacji Bayesa, aby określić, w jakim stopniu udokładniany model odpowiada danym eksperymentalnym. Z grubsza metody optymalizacji globalnej można podzielić na podejścia dokładne i stochastyczne. Metody dokładne mogą zagwarantować zbieżność do globalnego optimum, ale zwykle wiąże się z tym znaczny nakład obliczeniowy, zbyt wielki dla realistycznych zastosowań. Z kolei metody stochastyczne są często w stanie zlokalizować sąsiedztwo globalnego rozwiązania w rozsądnym czasie obliczeń, choć bez gwarancji zbieżności. Metaheurystyki (tj. heurystyki kierowane) są szczególną klasą metod stochastycznych, które okazały się bardzo dobre w szerokim zakresie zastosowań [11]. Pakiet MEIGO dostarcza najnowocześniejsze metaheurystyki (eSS, Enhanced Scatter Search

i VNS, Variable Neighbourhood Search) w wersjach open-source R (możliwe zastosowanie metody wnioskowania bayesowskiego BayesFit) oraz Matlab. MEIGO radzi sobie z najważniejszymi klasami problemów, szczególnie z (i) problemami ze zmiennymi decyzyjnymi rzeczywistymi (cNLP) i mieszanymi (MINLP) oraz (ii) problemami ze zmiennymi decyzyjnymi całkowitymi i binarnymi (IP). Co istotne, w przypadku złożonego problemu obliczeniowego MEIGO pozwala użytkownikowi na zastosowanie obliczeń równoległych z wykorzystaniem strategii kooperacyjnych [108]. MEIGO może optymalizować dowolne funkcje celu, dzięki temu ma zastosowanie do optymalizacji złożonych systemów, które mogą wymagać rozwiązywania wewnętrznych problemów (parametrów symulacji lub nawet innych problemów optymalizacyjnych) w celu uzyskania jawnych wartości funkcji celu i/lub możliwych ograniczeń. Pakiet MEIGO wykorzystuje dwie podstawowe metody metaheurystyczne: Enhanced Scatter Search (eSS) oraz Variable Neighbourhood Search (VNS).

Enhanced Scatter Search (eSS), czyli wyszukiwanie rozproszone [108], jest metaheurystyką opartą na populacji, której członkowie podlegają ewolucji w czasie pod wpływem zadanych parametrów optymalizacji. Założeniem metody jest, że rozmiar populacji N , w przeszukiwaniu rozproszonym jest niewielki, a kombinacje pomiędzy jej członkami są wykonywane systematycznie, a nie w sposób losowy. Bieżąca populacja jest nazywana „zbiorem referencyjnym” (RefSet). Metoda ulepszania opiera się na lokalnym wyszukiwaniu w celu zwiększenia zbieżności do rozwiązań optymalnych i może być stosowana z zadaną częstotliwością do członków tego RefSetu. Wśród uprawnień tej metody zastosowanych w metaheurystyce eSS można wymienić metodę zastępowania. W przeciwieństwie do oryginalnego schematu wyszukiwania rozproszonego, który wykorzystuje wymianę $\mu+\lambda$ (tzn. nowa populacja lub RefSet będzie składać się z najlepszych N rozwiązań wybranych z poprzednich członków RefSet i nowych rozwiązań potomnych), rozszerzone wyszukiwanie rozproszone wykorzystuje wymianę 1+1, podobną do strategii stosowanej w bardzo wydajnej metodzie ewolucyjnej, ewolucji różnicowej [108]. W ten sposób rozwiązanie potomne jest generowane z połączenia wybranego członka RefSetu i innego członka RefSetu oraz może zastąpić tylko tego członka RefSetu, który je wygenerował, a nie żadnego innego. Strategia ta zwiększa różnorodność i zapobiega przedwczesnej stagnacji wyszukiwania, nie pozwalając na obecność zbyt podobnych rozwiązań w RefSetu w tym samym czasie. Zaimplementowana w eSS strategia „go-beyond” pozwala wykorzystać tylko te kombinacje, które wskazują obiecujące kierunki optymalizacji, analizując czy rozwiązanie potomne jest lepsze od członka RefSetu, który je wygenerował. Proces ten jest powtarzany i za jego zakończenie uznaje się moment optymalizacji, gdy rozwiązania potomne przestają być lepsze od poprzednich. Na potrzeby optymalizacji podczas udokładniania modelu 7Hyp/ANS w nadprzestrzeni (3+1)D wykorzystano metaheurystykę eSS ze względu na rozsądny czas obliczeniowy oraz zalety metaheurystyki w zakresie wykorzystania najbardziej efektywnych punktów początkowych do przeprowadzenia lokalnych poszukiwań oraz uniknięcia przedwczesnej zbieżności.

Variable Neighbourhood Search (VNS) to metaheurystyka oparta na trajektorii do celów optymalizacji globalnej. VNS wykonuje lokalne wyszukiwanie dokonując oceny funkcji celu wokół wybranego, stabilnego rozwiązania, po czym powtarza procedurę odwiedzając różne sąsiedztwa i lokalizując różne lokalne optima, wśród których oczekuje się znalezienia optimum globalnego. VNS zazwyczaj poszukuje nowego sąsiedztwa poprzez perturbację zbioru zmiennych decyzyjnych przy użyciu kryterium odległości. Typowy schemat polega na zaburzaniu niewielkiego zbioru rozwiązań w pobliżu aktualnego, aż do momentu braku dalszej poprawy rozwiązania. Po tym kroku VNS przechodzi do eksploracji bardziej odległych sąsiedztw. Zaimplementowane zaawansowane strategie pozwalają na uniknięcie powtarzania perturbowanych zmiennych decyzyjnych w kolejnych wyszukiwaniach w obrębie sąsiadów. W razie potrzeby algorytm umożliwia lokalizację wysokiej jakości rozwiązania (nawet jeśli nie jest ono globalnym optimum) w krótkim czasie obliczeniowym. Ponadto w VNS zaadaptowano wspomnianą wcześniej strategię „go-beyond” do wykorzystania obiecujących kierunków podczas lokalnego wyszukiwania.

4.8.2. Skrypty MATLAB do przygotowania pliku wejściowego

Przygotowany zestaw skryptów został napisany w środowisku MATLAB (MathWorks) i zawierał komplet narzędzi służących do analizy modelu struktury krystalicznej białek. W jego obrębie można wyróżnić trzy warstwy dotyczące różnych etapów pracy nad plikami związanymi z modelem. Pierwsza warstwa obejmuje skrypty i funkcje, które umożliwiają ujednoczenie pierwotnego formatu plików wejściowych do formy, jaką przyjęto dla danej kategorii programów. Jako pliki wejściowe (*input*) można zastosować pliki w formatach *.cif* (współrzędne atomów) pobrane bezpośrednio z bazy PDB lub rutynowo generowane przez większość programów związanych z analizą i wizualizacją struktur krystalicznych. Jako eksperymentalne czynniki struktury, wraz z błędem doświadczalnym (zwykle dostępne w binarnym formacie *.mtz*), skrypt wykorzystuje pliki w formacie tekstowym *.txt* lub ASCII.

Przygotowany pakiet MATLAB stosuje następujące skrypty pomocnicze do odczytania pliku wejściowego *.cif* oraz wygenerowania tablicy atomów stosowanej na dalszych etapach:

- *is.amino.m* – funkcja sprawdza na podstawie zawartej w *.cif* etykiety „ATOM”, czy dany wiersz należy do listy atomów białka, czy do listy ligandów lub cząsteczek wody (etykieta „HETATM”).
- *letter_to_number.m* – funkcja zamienia oznaczenia kolejnych łańcuchów identyfikowanych w pliku *.cif* jako litery na liczby, co ułatwia zapis danych w formacie tablicy danych oraz znacznie skraca czas obliczeń. W przypadku struktury 7Hyp/ANS, 28 łańcuchów białek oznaczonych w pliku *.cif* jako A, B, ..., a, b zostaje zamienionych kolejno na 1, 2, ..., 27, 28. Skrypt rozróżnia małe i duże litery traktując każdy z łańcuchów indywidualnie.

- `save_structural_feature.m` – funkcja odcina zbędne dane (nagłówki, dane bibliograficzne, informacje o sekwencji) z pliku `.cif` i zapisuje do osobnego pliku `.txt` dane dotyczące struktury (parametry komórki elementarnej, symetria, grupa przestrzenna).
- `find_bonds.m`, `find_angles.m`, `find_torsion.m` – funkcja wyszukuje wśród list zawierających wszystkie atomy modelu atomy należące do danej molekuły (ip.. do danego aminokwasu oznaczonego jako 1), pobiera informację odpowiednio o długościach wiązań, kątów walencyjnych oraz kątów dwuściennych między tymi atomami i tworzy nowe pliki tekstowe zawierające listę długości wiązań, kątów oraz kątów dwuściennych w aktualnym modelu struktury. Zostaną one wykorzystane do tworzenia więzów w procesie udokładnienia modelu.
- `find_hydrogen.m` – pobiera listę atomów wodoru dla każdego aminokwasu z właściwego pliku `.cif` zawierającego dane stereochemiczne. Dla każdej z reszt aminokwasowych użyto modelowych danych zawartych w bibliotece więzów programu REFMAC. Funkcja wczytuje po kolei pliki `.cif` z katalogu zawierającego parametry wszystkich aminokwasów i ligandów obecnych w strukturze, pobiera dane o tym, który atom aminokwasu jest związany z wodorem, jaka jest długość i kąt wiązania z atomem wodoru oraz tworzy plik tekstowy zawierający informację o atomach w strukturze.
- `read_cif.m` – właściwy skrypt MATLAB wykorzystujący wymienione funkcje do generowania pełnej listy atomów struktury używanej w udokładnianiu. Po wczytaniu wejściowego, niezmiennego pliku `.cif` skrypt pobiera dane podstawowe o strukturze i przekształca superkomórkę 7Hyp/ANS do przestrzeni (3+1)D rzutując poszczególne komórki powiązane operacją tNCS (2-7) na łańcuchy A, B, a, b tworzące podstawowy motyw strukturalny w komórce 1. Następnie skrypt generuje listy długości wiązań, kątów i kątów torsyjnych między atomami w celu określenia więzów podczas udokładniania oraz tworzy osobną listę zawierającą atomy wodoru w modelu.

Rezultatem zastosowania skryptu `read_cif.m` na pliku wejściowym jest wygenerowanie pliku zawierającego dane o parametrach komórki elementarnej, utworzenie listy zawierającej dane o każdym atomie w modelu i osobnej listy atomów wodoru w modelu dających wkład do całkowitego czynnika struktury.

4.8.3. Przekształcanie danych do nadprzestrzeni (3+1)D

Skrypt `read_cif.m` generuje listę molekuł położonych wzdłuż osi **c**, które będą odpowiadać sobie po zrzutowaniu do nadprzestrzeni (3+1)D. Skrypt zawiera informację o czterech podstawowych łańcuchach białka w jednostce asymetrycznej komórki elementarnej 7Hyp/ANS – są to łańcuchy oznaczone A, B, a i b. Funkcja `reduce_data.m` generuje listę atomów po zrzutowaniu poszczególnych łańcuchów wzdłuż kierunku osi **c** na płaszczyznę utworzoną przez osie krystalograficzne **a** i **b** i odpowiadający im łańcuch z pierwszej komórki podstawowej. Odpowiednie kolumny w pliku przechowują informację

o łańcuchu, na który nastąpiło rzutowanie (A, B, a lub b) oraz średnim położeniu molekuly w wyniku nakładania kolejnych łańcuchów białek. Plik zawiera także listę wszystkich rzutowanych molekuł, przy czym warunkiem rzutowania było zachowanie odległości $< 6 \text{ \AA}$. Liczba wystąpień danej molekuly w okresie poszerzonym o 7 mówi o prawdopodobieństwie wystąpienia tej molekuly. Dotyczy ona ligandów oznaczonych w pliku *.cif* etykietą HETATOM i zachowuje informację o ich rozmieszczeniu w poszczególnych miejscach wiążących po rzutowaniu). Na koniec następuje zamiana oznaczeń poszczególnych łańcuchów w formie liter na odpowiadające im kolejne liczby naturalne (np. A \rightarrow 1, B \rightarrow 2tp.td.).

4.8.4. Generowanie więzów stereochemicznych

Więzy stereochemiczne dla aminokwasów były generowane na podstawie danych referencyjnych z bibliotek programu REFMAC. Dane te dla aminokwasów i innych ligandów są stosowane podczas udokładniania w programie w celu zachowania odpowiedniej geometrii i konformacji struktury białka, a także zwiększenia liczby obserwacji dla zrównoważenia ogromnej liczby parametrów modelu. Do generowania tablic atomów przechowujących informację o dopuszczalnych długościach wiązań, wartościach kątów oraz kątów torsyjnych posłużyły funkcje *find_bonds.m*, *find_angles.m*, *find_torsion.m*. Funkcje te pobierają z biblioteki dane o geometrii poszczególnych reszt aminokwasowych oraz porównują je z pełną listą atomów pobranych z wejściowego pliku *.cif*. Każda molekula w strukturze 7Hyp/ANS zostaje pogrupowana, dzięki czemu możliwe jest ustalenie, jakie będą numery aminokwasu danego typu w kolejnych łańcuchach i atomy wchodzące w jego skład. Funkcja porównuje, między którymi atomami modelu istnieje wiązanie i oblicza jego dopuszczalną długość na podstawie listy atomów. Całość zostaje zapisana do nowego pliku *.txt*, w którym 1. kolumna określa numer aminokwasu, 2. i 3. informuje, które atomy w obrębie molekuly tworzą wiązanie, a 4. i 5. kolumna przechowują długość wiązania oraz niepewność standardową długości. W analogiczny sposób generowane są tablice zawierające dane o wartościach kątów i kątów torsyjnych (przy czym będą one miały odpowiednio 1 i 2 kolumny więcej, gdyż kąty walencyjne są tworzone przez 3 atomy, a dwuścienne przez 4 atomy jak na Rys. 20).

1	1	2	3	4	117.124000	1.500000
2	1	3	2	1	109.241000	1.500000
3	1	5	2	1	110.906000	1.500000
4	1	5	2	3	109.344000	1.500000
5	1	6	5	2	113.476000	1.500000
6	1	7	6	5	112.576000	2.440000
7	1	8	7	6	100.595000	1.500000
8	2	10	11	12	117.159000	1.570000
9	2	11	10	9	109.627000	1.500000

Rysunek 20. Plik zawierający tablicę więzów ograniczających wartość kątów dla modelu. Pierwsza kolumna oznacza numer molekuly (aminokwasu lub cząsteczki liganda) w strukturze, kolumny 2, 3 i 4 zawierają numery atomów danej molekuly tworzących kąt, kolumna 5 informuje o wartości kąta, a kolumna 6 o dopuszczalnym odchyleniu wartości.

4.8.5. Udokładnianie modelu

Drugą warstwę działania oprogramowania stanowi zestaw skryptów oraz funkcji do udokładniania modelu struktury. W jego skład wchodzi podstawowe komponenty:

- `calc_basevectors.m` – na podstawie parametrów sieciowych komórki elementarnej (długości a , b , c i kąty α , β , γ) pobranych z pliku `.cif` w poprzednim kroku oblicza zestaw wektorów bazowych definiujących sieć.
- `calc_wavevectors.m` – oblicza wektory falowe.
- `generate_coefficients.m` – na podstawie listy atomów w zredukowanej komórce pozwala określić, które parametry będą udokładniane. Konwencja jest prosta – jeśli w kolumnie oznaczającej dany parametr pojawi się 1, będzie on obliczany i udokładniany, jeśli pojawi się 0 program nie będzie go uwzględniać w toku udokładniania. Skrypt pozwala swobodnie decydować, które parametry będą udokładniane, a które nie, dzięki czemu można je wprowadzać stopniowo. Do wyboru użytkownika pozostaje udokładnianie: amplitudy modulacji, przesunięcia, modulowanych ADPs, przesunięcia fazowego funkcji modulacji AMF oraz prawdopodobieństwa zbliżniczenia (Rys. 21).

```

24 - for i=1:694
25 -     for j=1:num_param_mod
26 -         if j==2 || j==3 || j==4 || j==5 || j==6 || j==7 %amplitudes
27 -             coeff_mod((i-1)*num_param_mod+j, 2)=1;
28 -         end
29 -         if j==8 || j==9 || j==10 %angles
30 -             coeff_mod((i-1)*num_param_mod+j, 2)=1;
31 -         end
32 -         if j==11 || j==12 || j==13 %shift
33 -             coeff_mod((i-1)*num_param_mod+j, 2)=0;
34 -         end
35 -         if j==14 || j==15 %ADP modulation amplitudes
36 -             coeff_mod((i-1)*num_param_mod+j, 2)=0;
37 -         end
38 -         if j==16 %displacive modulation phase
39 -             coeff_mod((i-1)*num_param_mod+j, 2)=1;
40 -         end
41 -         if j==17 %ADP modulation phase
42 -             coeff_mod((i-1)*num_param_mod+j, 2)=0;
43 -         end
44 -     end
45 - end
46 - coeff_mod(end-3, 2)=1; %corr_coeff
47 - coeff_mod(end-2, 2)=0; %twin
48 - coeff_mod(end-1, 2)=0; %twin
49 - coeff_mod(end, 2)=0; %twin

```

Rysunek 21. Plik służący do generowania listy atomów i związanych z nimi parametrów podlegających udokładnieniu. Parametr *j* oznacza numer kolumny w wyjściowym pliku, po ustawieniu wartości z 0 na 1, oznaczony parametr będzie udokładniany.

- `form_factors.m` – rozróżnia typ atomu (np. C, O lub N) i przypisuje każdemu odpowiednie atomowe czynniki rozpraszania.
- `structure_factor_protein.m` – funkcja służy do obliczania czynników struktury utworzonego modelu. Na podstawie listy atomów pobiera informację o ich współrzędnych i zwraca listę modulowanych czynników struktury oraz wywołuje funkcję, która oblicza czynnik struktury dla atomów wodoru. Biorąc pod uwagę zbliżenie kryształu i operacje symetrii, funkcja oblicza czynniki struktury osobno dla każdej z domen (wymaga to jedynie zdefiniowania ich wzajemnej orientacji), a następnie sumuje intensywności czynników struktury od wszystkich domen bliźniaczych podając całkowity czynnik struktury z uwzględnieniem modulacji.
- `calc_structure_factor_hydrogen.m` – funkcja wczytuje listę atomów wodoru wygenerowaną w poprzednim etapie i oblicza osobno dla nich czynnik struktury jako wkład do całkowitego czynnika struktury.

Udokładnianie modelu, czyli optymalizacja obliczonych czynników struktury względem eksperymentalnych, przebiegło z zastosowaniem pakietu MEIGO. Jako metaheurystykę zastosowano algorytm eSS. Górną i dolną granicę akceptowanych rozwiązań zmierzających do optymalizacji problemu stanowiły wektory zawierające listę wartości parametrów: amplitudy modulacji, przesunięcia fazowego AMF, modulowanych ADPs, prawdopodobieństwa zbliżenia itp. Metaheurystyka eSS to globalna metoda optymalizacji wykonująca lokalne przeszukiwanie od wybranych punktów początkowych

w celu przyspieszenia zbieżności do rozwiązań optymalnych. Jako opcję kryteriów przeszukiwania lokalnego wybrano zaimplementowaną w pakiecie MEIGO metodę *fmincon*. Jest to zaimplementowana w pakiecie MATLAB metoda programowania, która wykonuje nieliniową optymalizację przy nałożeniu ograniczeń liniowych i nieliniowych. Aby zastosować tę metodę, wystarczy skonfigurować wstępne ustawienia, czyli kryteria zbieżności, maksymalną liczbę iteracji oraz sposób obliczania gradientów [108].

4.8.6. Skrypty do generowania graficznej reprezentacji modelu i jego weryfikacji

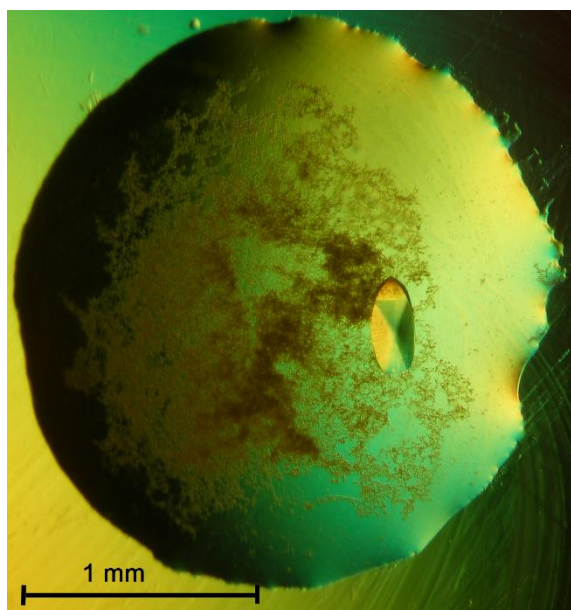
Trzecią warstwę przygotowanego oprogramowania stanowią skrypty i funkcje stosujące listę atomów po udokładnieniu i umożliwiające wizualizację modelu w celu graficznego przedstawienia efektów udokładniania, analizy ADPs w obrębie modelu, a także weryfikacji więzów stereochemicznych dla całego modelu jak i w obrębie poszczególnych reszt aminokwasowych, bądź danego typu cząsteczek liganda. W skład tej warstwy wchodzi następujące skrypty i funkcje:

- *analyse_ADp.m* – skrypt umożliwiający wygenerowanie rozkładu ADPs w modelu struktury. Jako dane wejściowe przyjmuje wygenerowany wcześniej plik z parametrami sieciowymi komórki elementarnej oraz pełną listę atomów struktury. Na przygotowany rozkład ADPs zostaje nałożony rozkład gamma z odwróconą sigmą (SIGD – Sigma-Inverted Gamma Distribution), zaproponowany jako sposób modelowania zachowania ADPs w strukturach makromolekularnych oraz parametry dopasowania tego rozkładu do istniejących ADPs.
- *plot_structure.m* – skrypt do generowania graficznej reprezentacji białka w obrębie komórki elementarnej. Umożliwia generowanie rzutu struktury wzdłuż wybranego kierunku krystalograficznego i rozróżnia typy atomów, kolorując je zgodnie z konwencją CPK (Corey’a, Paulinga, Koltuna) [109]: węgiel – barwa czarna, tlen – barwa czerwona, azot – barwa niebieska, siarka – barwa żółta, wodór – barwa biała. Zdefiniowanie odpowiedniego zakresu molekuł na podstawie ich numerów z ogólnej listy atomów generowanej w poprzednich krokach, umożliwia wizualizację wybranego fragmentu struktury, np. fragmentu łańcucha głównego, a także odtworzenie na podstawie udokładnionych czynników struktury map gęstości elektronowej dla tego fragmentu.
- *restraints_analysis.m* – skrypt służy do analizy poprawności stereochemicznej modelu. Pozwala porównać osobno długości wiązań, wartości kątów walencyjnych i kątów torsyjnych wszystkich aminokwasów danego typu lub typów ligandów względem struktury modelowej z bibliotek . WREFMAC. W tym celu wykorzystuje modelowe listy więzów generowane na początku analizy struktury i ograniczające długości wiązań oraz wartości kątów zwykłych podczas udokładniania. Zawiera również wartości kątów torsyjnych służących jako probierz walidacyjny modelu. Wykresy zawierają również rozkłady długości wiązań, kątów i kątów torsyjnych w obrębie molekuł danego typu w całej strukturze, np. rozkład długości wiązań dla całej populacji ANS w modelu.

5. Dane doświadczalne wykorzystane w rozprawie

5.1. Otrzymanie kompleksu białkowego Hyp-1/ANS

Rekombinowane białko Hyp-1 oraz jego kompleksy z ANS otrzymała dr Joanna Śliwiak. Ona też uzyskała kryształy kompleksów. Dane dyfrakcyjne dla kryształów 7Hyp/ANS zmierzył i przetworzył dr hab. Zbigniew Dauter w APS, USA. Hyp-1 wytwarzano w szczepie bakterii *Escherichia coli* DE3 przy użyciu wektora plazmidowego pET151/D zawierającego sekwencję kodującą produkcję białka Hyp-1 [41]. Po całonocnej hodowli w temperaturze 291 K, komórki bakterii odwirowano, umieszczono w roztworze buforu [500 mM NaCl, 20 mM Tris-HCl pH 8,0, 20 mM imidazolu, 3 mM merkaptoetanolu, 100 $\mu\text{g ml}^{-1}$ lizozymu z białek jaj kurzych (Sigma-Aldrich)] i poddano sonikacji w celu uwolnienia zawartości komórek. Otrzymany lizat bakterii przepuszczono przez kolumnę chromatograficzną wymywaną 500 mM roztworem imidazolu. Po przeprowadzeniu chromatografii powinowactwa, białko poddano oczyszczeniu na kolumnie wykluczającej wielkość w 3 mM buforze cytrynianowym pH 6,3 z 150 mM NaCl. Po oczyszczeniu, białko było dializowane buforem cytrynianowym, a następnie zamrożone w temperaturze 193 K. Przed krystalizacją, roztwór białka zagęszczono i poddano inkubacji jednocześnie z ANS i MEL, gdzie stosunek molowy białko : ANS w mieszaninie inkubacyjnej wynosił 1:10, natomiast stosunek ANS:MEL wynosił w zależności od kombinacji 2:1, 1:1, 1:2 lub 1:3. Próby krystalizacji dla wszystkich czterech kombinacji białko-ligand przygotowano z udziałem cytrynianu sodu jako czynnika strącającego oraz dwóch rodzajów buforu (bufor cytrynianowy o pH 6,3 lub HEPES o pH 7,5). Kryształy Hyp-1/ANS pojawiły się tylko w kombinacji zawierającej bufor HEPES, przy stosunku ANS:MEL równym 2:1 i przy stężeniu cytrynianu sodu od 1,6 M (po dwóch dniach), do 1,3 M (po dwóch miesiącach). Kształt kropli roztworu krystalizacyjnego oraz stosunek objętościowy białka do zbiornika determinował kształt powstających kryształów oraz ich zdolności dyfrakcyjne. Tym samym, gdy stosunek objętościowy białko : zbiornik wynosił 1:1, pojawiły się kryształy w kształcie soczewki, które ulegały dyfrakcji rentgenowskiej do rozdzielczości 2,3 Å (Rys. 22). Całkowitą procedurę, dobór warunków krystalizacji oraz rezultaty omówiono w [40-42].



Rysunek 22. Zdjęcie pojedynczego kryształu białka Hyp-1 współkryształowanego z ANS i MEL przy użyciu buforu HEPES i cytrynianu sodu jako czynnika strącającego roztworów w stosunku objętościowym białka i zbiornika 1:1. Zdjęcie zostało wykonane przez dr Joannę Śliwiak.

5.2. Rozwiązanie struktury Hyp-1/ANS

Dane dotyczące szczegółów dyfrakcji rentgenowskiej kryształów kompleksu 9Hyp/ANS zmierzone na linii 14.2 synchrotronu BESSY II przez dr Joannę Śliwiak zostały zamieszczone w Tabeli 1.

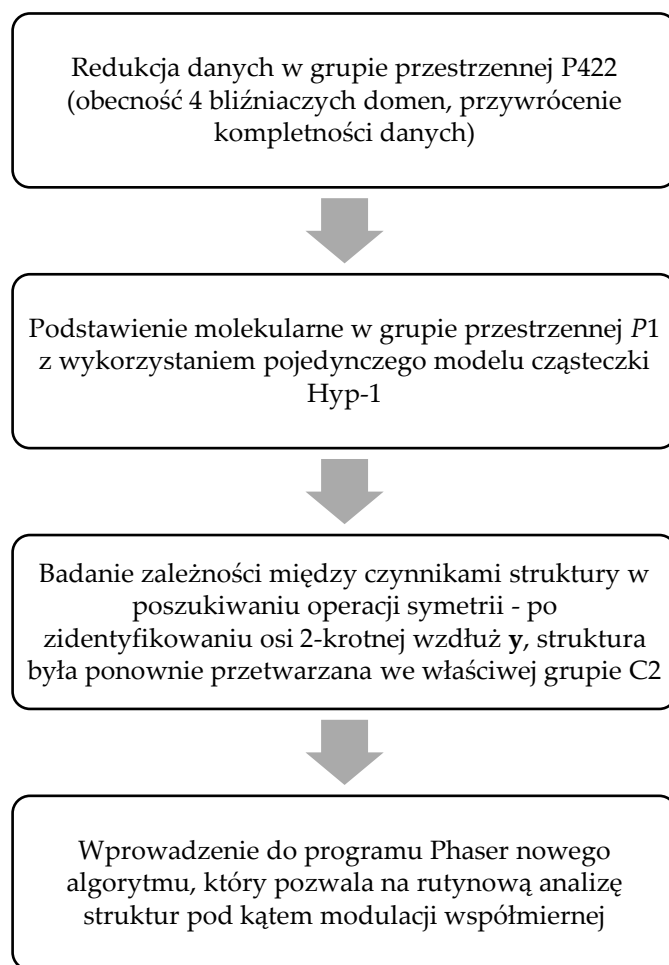
Tabela 1. Dane doświadczalne [110].

Linia pomiarowa synchrotronowa	BESSY II 14.2
Długość fali (Å)	0,9184
Temperatura pomiaru (K)	100
Krioprotektant	1,4 M cytrynian sodu
Wielkość kryształu (mm)	0,4x0,2x0,2
Grupa przestrzenna	C2
Parametry komórki elementarnej a, b, c, β (Å, °)	148,85, 148,85, 385,40, 90,00
Rozdzielczość (Å)	31,58-2,30 (2,39-2,30) ^a
Liczba refleksów zmierzonych/unikatowych	355 613/127 324
Kompletność (%)	99,3 (94,2) ^a
Wielokrotność	2,8 (2,6) ^a
$\langle I/\sigma \rangle$ stosunek sygnału do szumu	6,22 (1,28) ^a
$CC_{1/2}$	0,978 (0,620) ^a
R_{merge}^b	0,066 (0,691) ^a

^a Wartości w nawiasach odpowiadają ostatniej powłoce o najniższej rozdzielczości.

^b $R_{merge} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$, gdzie $\langle I(hkl) \rangle$ oznacza średnią intensywność danego refleksu hkl.

Podczas rozwiązania struktury 9Hyp/ANS posłużono się analogią do poprzedniego modelu 7Hyp/ANS [41]. Struktura ta przysporzyła pewnych trudności, ponieważ podczas pomiaru danych dyfrakcyjnych pierwotnie założono symetrię tetragonalną, jednak finalnie zaobserwowano, że stopień kompletności otrzymanych danych wynosi jedynie 73%. Jednakże obecność doskonałych bliźniaków tetartoedrycznych sugerowała możliwość rozszerzenia danych z symetrii tetragonalnej do symetrii jednoskośnej bez wprowadzania istotnych błędów, gdyż w przypadku doskonałego zbliźniaczenia tetartoedrycznego występowałyby zgodność otrzymanych danych z symetrią grupy punktowej 422. Następnie przeprowadzono próby podstawienia molekularnego we wszystkich prymitywnie centrowanych grupach przestrzennych zgodnych z pozorną grupą punktową 422, otrzymując potencjalne rozwiązanie w grupie przestrzennej $P4_122$, ale symetria ta została wykluczona przez silne wygaszenia $00l \neq 4n$. W połączeniu z obecnością zbliźniaczenia sugerowało to, że prawdziwa symetria jest niższa. Ze względu na trudności w wytypowaniu jednej z wielu podgrup $P422$ strukturę próbowano rozwiązać w najprostszej trójskośnej grupie przestrzennej $P1$, szukając 56 kopii Hyp-1 (7 cząsteczek białka wygenerowanych wektorem translacji $(-0,004, -0,004, 0,285)$ oraz 7 dodatkowych kopii tego rozwiązania: $7 + 7 \times 7 = 56$). Prowadzone w ten sposób poszukiwania zakończyły się sukcesem, a prawidłowa symetria $C2$ została wydedukowana na podstawie analizy symetrii obliczonych czynników struktury. Finalne rozwiązanie MR w symetrii $C2$ uzyskano poprzez wyszukiwanie czterech kopii pierwszego zestawu siedmiu cząsteczek w rozwiązaniu w grupie $P1$. Cały proces ilustruje Rys. 23.



Rysunek 23. Schemat rozwiązania struktury 7Hyp/ANS. Opracowanie własne na podstawie [40].

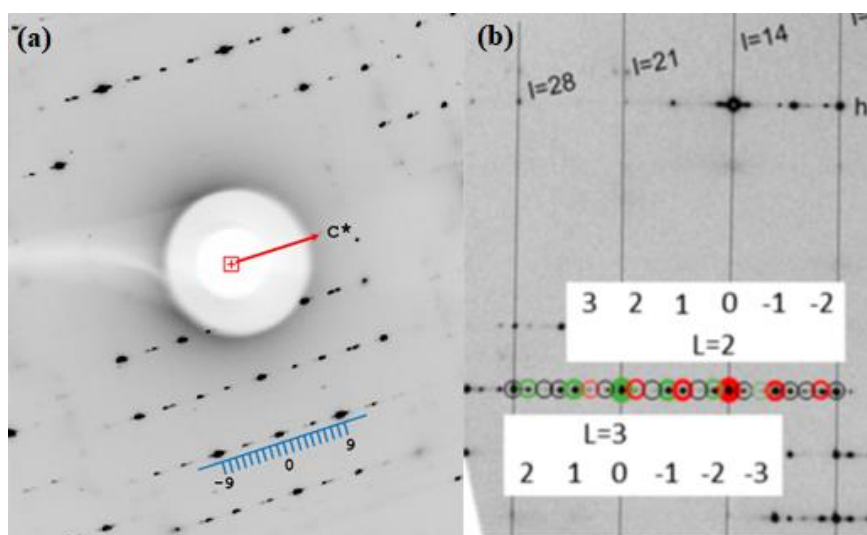
Udane rozwiązanie struktury 7Hyp/ANS pozwoliło zastosować podobną procedurę podczas procesowania i udokładnienia danych 9Hyp/ANS. Oba kryształy łączy obecność tetartoedrycznego bliźniaczenia, które utrudnia określenie prawidłowej grupy przestrzennej kryształu, dlatego rozwiązanie struktury 9Hyp/ANS analogicznie przeprowadzono metodą podstawienia molekularnego w trójskośnej grupie przestrzennej P1. Zastosowany w MR model molekularny składał się z 36 cząsteczek Hyp-1 i został przygotowany na podstawie zawartości jednostki asymetrycznej 7Hyp/ANS (siedem powtórzeń tetrad molekularnych: AB-ab, CD-YZ, EF-WX, . . . , MN-OP). Na końcu MN-OP został on przedłużony zgodnie z translacją wzdłuż c o dodatkowe kopie AB-ab i CD-YZ. Program Phaser znalazł dwie kopie tej sondy (tj. $2 \times 36 = 72$ molekuly Hyp-1) w trójskośnej komórce elementarnej. Przeprowadzone szybkie udokładnienie tego modelu w programie REFMAC zmniejszyło wskaźnik R do 0,25 i automatycznie wykryło osiem bliźniaczych domen (opcja *twin_refinement*). W przestrzeni rzeczywistej takie rozwiązanie odpowiadało sieci jednoskośnej o centrowaniu typu C tak jak w przypadku 7Hyp/ANS. Do dalszego udokładniania wybrano model składający się z jednej połowy, tj. 36 cząsteczek białka. Ze względu na obecność osi dwukrotnej w grupie przestrzennej C2, pozorne ośmiokrotne

zbliżenie zostało zredukowane do czterokrotnego, czyli tetartoedrycznego, wskazując na istnienie czterech domen bliźniaczych.

6. Wyniki

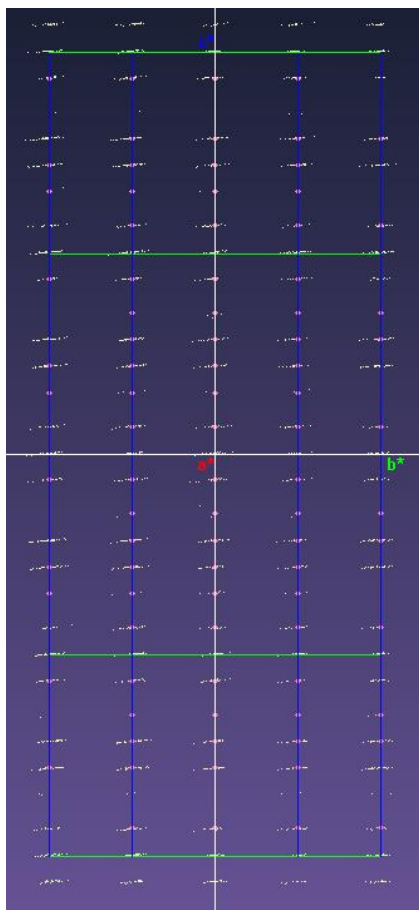
6.1. Kompleksy białkowe Hyp-1/ANS jako struktury modulowane

W przestrzeni odwrotnej modulacja struktury krystalicznej 7Hyp/ANS oraz 9Hyp/ANS przejawia się poprzez unikatowe obrazy dyfrakcyjne składający się z silniejszych głównych pików braggowskich, które są rozdzielone słabszymi refleksami satelitarnymi w kierunku c^* (Rys. 24). Udane wskaźnikowanie obrazu dyfrakcyjnego (zarówno głównych jak i satelitarnych refleksów) przy pomocy prostej trójwymiarowej siatki, w której satelity dzielą odległości pomiędzy głównymi refleksami w racjonalny sposób, wskazuje, że modulacja jest współmierna (Rys. 25).



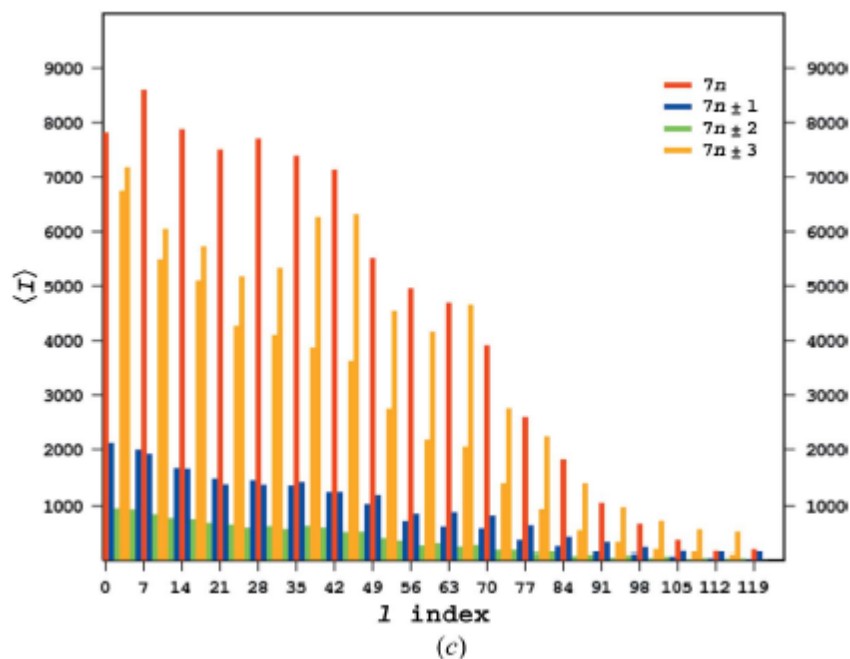
Rysunek 24. Obrazy powiększone obrazy dyfrakcyjne kryształu 9Hyp/ANS z zaznaczonymi wzdłuż kierunku c^* refleksami głównymi (-9; 9) i satelitarnymi (a) oraz analogiczny powiększony obraz z refleksami głównymi i satelitami dla 7Hyp/ANS (b). Obrazy zostały zarejestrowane przez dr Joannę Śliwiak i opisane przez dra Zbigniewa Dautera.

Tym samym można uprościć opis struktury, wprowadzając rozszerzoną (siedmiokrotnie w przypadku 7Hyp/ANS lub dziewięciokrotnie w przypadku 9Hyp/ANS) w kierunku c superkomórkę. Na tym etapie nie jest konieczne uciekanie się do bardziej rygorystycznego, ale i znacznie bardziej skomplikowanego opisu w przestrzeni wielowymiarowej, który byłby konieczny w przypadku niewspółmiernej modulacji.



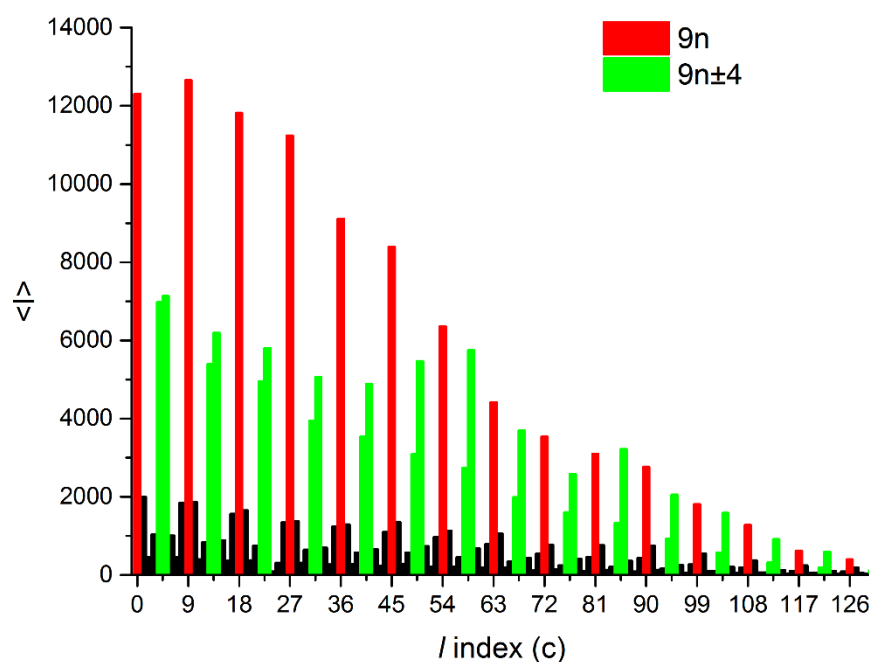
Rysunek 25. Fragment obrazu dyfrakcyjnego kryształu 7Hyp/ANS w kierunku c^* w programie CrysAlis Pro z zaznaczonymi refleksami głównymi (wzdłuż zielonych linii) oraz satelitami (różowe kropki).

Obecność elementów pseudosymetrycznych związanych z upakowaniem cząsteczek Hyp-1 wewnątrz struktury silnie wpływa na intensywność dyfrakcji. W przypadku kompleksu 7Hyp/ANS zbadano uśrednione intensywności refleksów w kolejnych warstwach o stałym wskaźniku l . Zaobserwowana powtarzająca się modulacja natężeń refleksów, z pikami oddzielonymi o $7/2$ wzdłuż kierunku c^* (Rys. 26), wskazuje na niekrytalograficzne przesunięcie wzoru molekularnego wzdłuż najdłuższego wymiaru komórki kryształu, c .



Rysunek 26. Histogram rozkładu intensywności w warstwach o tym samym wskaźniku l dla 7Hyp/ANS [40].

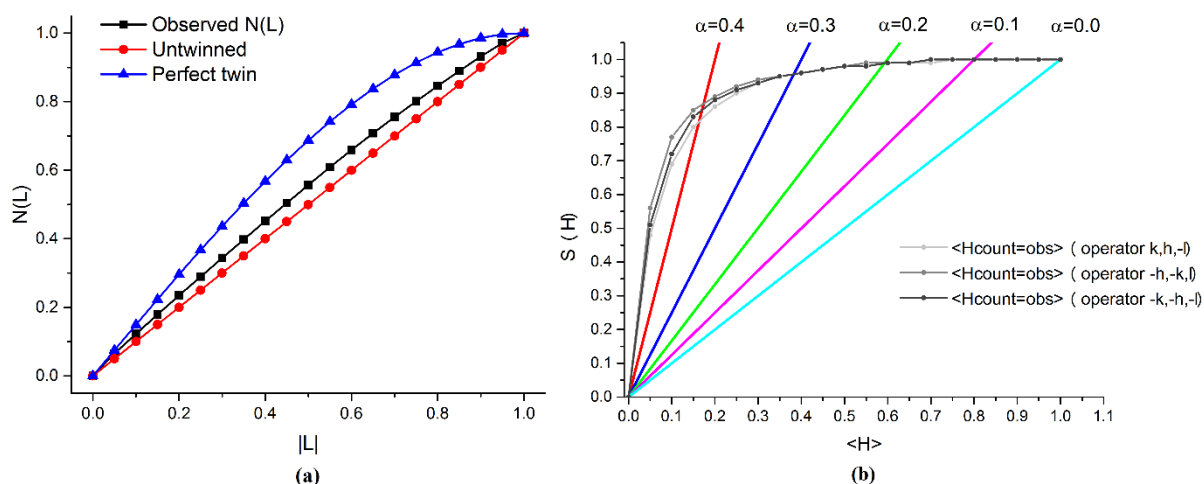
Podobna analiza intensywności pików w warstwach o stałym wskaźniku l dla struktury 9Hyp/ANS ujawnia wzór oparty na tej samej zasadzie co dla 7Hyp/ANS, ale z najsilniejszymi intensywnościami rozmieszczonymi regularnie co $9/2$ w przestrzeni odwrotnej przy kolejnych refleksach o stałym wskaźniku l (Rys. 27). Rozkład ten wynika z wskaźnikowania obrazu dyfrakcyjnego przy użyciu modelu superkomórki rozszerzonej w kierunku osi c , a tym samym obejmującej całą falę modulacji w kierunku c . W takim podejściu refleksy są wskaźnikowane (l) kolejno wzdłuż c^* , bez żadnego podziału. W tym opisie modulacja w przestrzeni rzeczywistej jest traktowana jako bardzo rozległa (siedmio- lub dziewięciokrotna) tNCS. Obecność tego tNCS jest dobrze widoczna w postaci dodatkowych silnych pików na natywnych mapach Pattersona, z maksimum mniej więcej co $1/7$ długości komórki elementarnej w kierunku c [341]. Analogicznie, obecność dziewięciokrotnej tNCS sprawia, że dla struktury 9Hyp/ANS dodatkowe silne piki pojawiają się na mapie Pattersona w odstępach $1/9$ długości komórki elementarnej wzdłuż osi c .



Rysunek 27. Histogram rozkładu intensywności w warstwach o tym samym wskaźniku l dla 9Hyp/ANS [110].

6.2. Testy zbliźniczenia

Operatory zbliźniczenia i udział poszczególnych domen dla struktury 9Hyp/ANS zostały oszacowane na podstawie skumulowanego rozkładu $S(H)$ dla par bliźniaczo powiązanych intensywności. Obliczone frakcje bliźniacze (blisko 0,45 w $S(H)$ teście dla każdego z wykrytych operatorów zbliźniczenia) wydają się nieco zawyżone w porównaniu z wartością 0,25 oczekiwaną dla idealnej tetartoedrycznie zbliźnioną strukturą. Na wykresie skumulowanego rozkładu $S(H)$ kolorowe linie ciągłe odpowiadają teoretycznym rozkładom dla frakcji bliźniaczych o udziale od 0,0 do 0,4 (łącznie wszystkie frakcje bliźniacze w kryształach sumują się do 1). Szare przerywane linie pokazują $S(H)$ dla podzbiorów refleksów o wskaźnikach $(k, h, -l)$, $(-h, -k, l)$ i $(-k, -h, -l)$, co odpowiada operatorom doskonałego tetartoedrycznego zbliźniczenia, sparowanych z podzbiorem podstawowym (h, k, l) . Kształt eksperymentalnych rozkładów sugeruje ułamek zbliźniczenia 0,45 dla każdego operatora bliźniaczego.



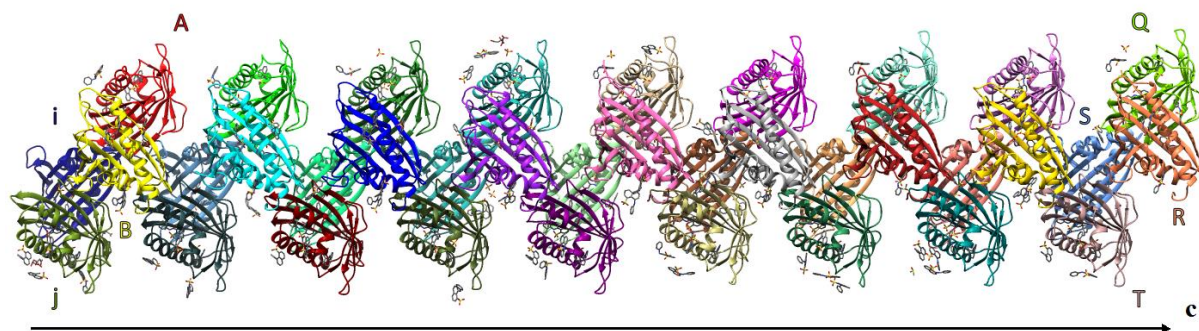
Rysunek 28. Skumulowany rozkład $N(L)$ (L-test) obliczony w programie TRUNCATE [53] dla danych 9Hyp/ANS (a). Skumulowany rozkład stosunku H (test $S(H)$) dla refleksów w zestawie danych dyfrakcyjnych 9Hyp/ANS (b).

Podobną weryfikację zbliżniczenia kryształu przeprowadzono przy użyciu L-testu [111], w którym wybrano refleksy lokalnie powiązane w przestrzeni odwrotnej do obliczenia skumulowanego rozkładu współczynnika $|L|$. Rozkład skumulowany $N(L)$ (L-test) został obliczony przy użyciu programu TRUNCATE dla przypadku danych idealnie zbliżnionych oraz całkowicie niezbliznionych, a także dla eksperymentalnych danych dyfrakcyjnymi dla 9Hyp/ANS. Na Rys. 28 niebieskie i czerwone krzywe reprezentują oczekiwane skumulowane rozkłady $|L|$ odpowiednio dla idealnie zbliżnionych i niezbliznionych danych, podczas gdy czarna linia pokazuje obliczony skumulowany rozkład dla 9Hyp/ANS, wskazując na przynajmniej częściowe zbliżniczenie kryształu. Również numeryczne wyniki L-testu dla danych procesowanych w prawidłowej jednoskośnej grupie przestrzennej ($\langle |L| \rangle = 0,459$ i $\langle L \rangle^2 = 0,290$) wskazują, że przynajmniej część kryształu nosi wyraźne cechy obecności bliźniaczych elementów symetrii.

6.3. Upakowanie molekuł w komórce elementarnej

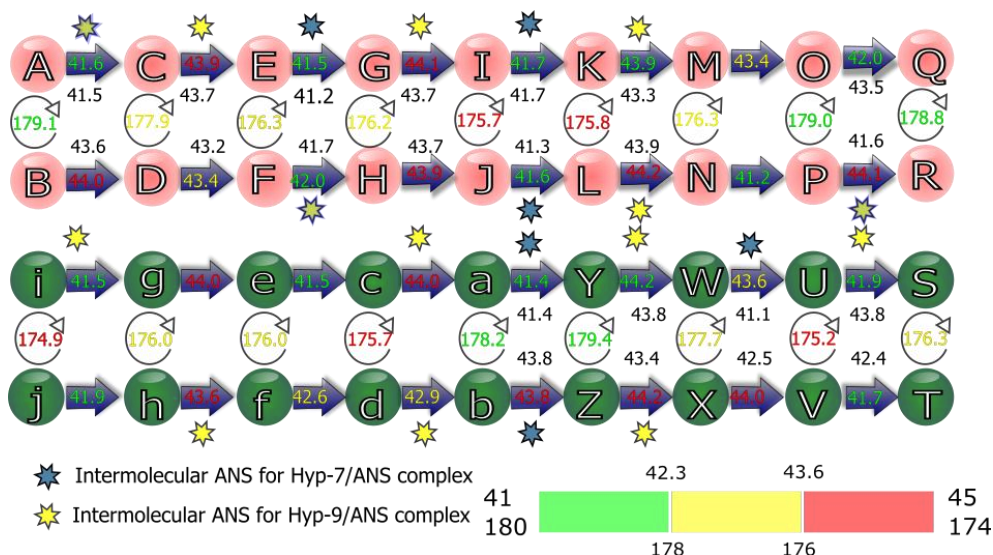
Podobnie jak większość białek grupy PR-10, Hyp-1 występuje w roztworach w postaci monomeru, co zostało potwierdzone metodami chromatograficznymi oraz elektroforezy PAGE [90]. Co więcej, należy się spodziewać, że jego funkcja biologiczna dotyczy formy monomeru. Pomimo tego zaobserwowano, że istnieje zjawisko funkcjonalnej dimeryzacji białek grupy PR-10 (np. otrzymanego z rośliny *Medicago truncatula* MtN13 [94]) oraz że w strukturze Hyp-1 z cząsteczkami PEG, znajdującymi się we wnętrzu białka, cząsteczki białka występowały w formie dimerów powiązanych mostkami S-S pomiędzy resztami cysteinowymi Cys126 [90]. W badanych strukturach 7Hyp/ANS oraz 9Hyp/ANS wszystkie kopie wielokrotne Hyp-1 są również sparowane w dimery poprzez międzycząsteczkowe oddziaływania pomiędzy atomami arkuszy $\beta 1$ sąsiednich molekuł. Łańcuchy te są antyrównoległe względem siebie oraz płynnie przechodzą w wewnątrzcząsteczkowy arkusz β od jednej cząsteczki do drugiej. Dla porządku przyjęto następujący schemat znakowania: kolejne dimery białka Hyp-1 to AB, CD, ... , ab (dla

7Hyp/ANS) oraz AB, CD, ..., ij dla 9Hyp/ANS. W przypadku struktury 9Hyp/ANS dziewięć z tych dimerów (AB, . . . , QR) ma tę samą orientację i podobne, powtarzające się odstępy równe $1/9$ wzdłuż osi c , tworząc wyraźny rząd oznaczony jako I. W wyniku działania na rząd I niekrystalograficznej osi śrubowej 2_1 równoległej do c , generowany jest rząd II składający się z dimerów ST, . . . , ij. Kolejne dimery na pseudokrystalograficznej osi $2_{1/9}$ są ze sobą powiązane operacjami obrotu o 180° i translacji o $1/18$ wzdłuż c . W ten sposób powstaje schemat upakowania molekularnego przypominający śrubę, którą można opisać jako $2_{1/9}$ (Rys. 29).



Rysunek 29. Schemat upakowania molekuł 9Hyp/ANS w obrębie komórki elementarnej wzdłuż kierunku c . Każdą z molekuł białkowych zaznaczono w innym kolorze.

Charakterystyczny wzór upakowania krystalicznego 36 cząsteczek białka Hyp-1 w dwóch kolumnach wzdłuż najdłuższego wymiaru komórki c , wraz z dużą liczbą cząsteczek w jednostce asymetrycznej stanowi cechy charakterystyczne pseudosymetrii struktury krystalicznej 9Hyp/ANS. Charakterystyczny zygzakowaty wzór dwóch kolumn w komórce elementarnej stanowi wynik działania niekrystalograficznej osi śrubowej $2_{1/9}$, która łączy obrót o 180° z translacją o $1/18$ wzdłuż osi śrubowej w kierunku $[001]$. Główna niekrystalograficzna oś śrubowa zostaje przecięta przez dwie osie dwukrotne rozmieszczone w przybliżeniu wzdłuż kierunków $[110]$ i $[\bar{1}10]$, które parami działają lokalnie odpowiednio wewnątrz i pomiędzy dimerami Hyp-1. Prawdziwa krystalograficzna oś dwukrotna (jak również prawdziwa oś śrubowa 2_1) przebiega w kierunku $[010]$ prostopadle do pseudośruby $2_{1/9}$ i ukośnie pomiędzy osiami pseudodwukrotnymi. Składowa translacyjna pseudośruby $2_{1/9}$ ma wyraźne niedoskonałości widoczne jako zmiany odległości między środkami kolejnych molekuł Hyp-1 wzdłuż rzędów I i II (Rys. 30).



Rysunek 30. Rozkład odległości i kątów skrzywienia między środkami sąsiednich łańcuchów białkowych w strukturze 9Hyp/ANS. Kolorami zielonym, żółtym i czerwonym zaznaczono miarę odstępstwa zmierzonych wartości. Przy pomocy żółtych i błękitnych gwiazdek zaznaczono lokalizację międzywęzłowych cząsteczek ANS dla struktur 7Hyp/ANS oraz 9Hyp/ANS.

Co więcej, dla modelu 9Hyp/ANS zmierzono wielkość zniekształceń wzdłuż kierunków krystalograficznych *a* i *b* wyznaczając wektory translacji między sąsiadującymi łańcuchami białka połączonymi tNCS (Rys. 31).

Kolejne łańcuchy białek w danym rzędzie	Składowa <i>x</i> wektora translacji [Å]	Składowa <i>y</i> wektora translacji [Å]	Składowa <i>z</i> wektora translacji [Å]	RMSD nałożenia węgli α łańcuchów [Å]
Rząd I				
A...C	0,98	0,34	40,87	1,01
B...D	0,12	1,11	44,14	0,80
C...E	-0,67	-0,36	44,58	0,72
D...F	-0,01	-1,30	41,60	1,04
E...G	-0,24	1,86	40,25	0,63
F...H	-0,72	1,72	43,33	0,91
G...I	0,91	2,30	44,17	0,57
H...J	2,74	-1,35	44,93	0,78
I...K	-0,67	2,31	40,83	0,61
J...L	-2,43	0,95	40,92	0,78
K...M	1,49	-2,56	44,71	0,60
L...N	2,49	-1,06	44,80	0,72
M...O	-1,35	2,35	41,35	0,73
N...P	-1,96	0,46	40,78	0,80
O...Q	1,05	-2,09	43,99	0,72
P...R	0,88	-0,28	44,38	0,64
Rząd II				

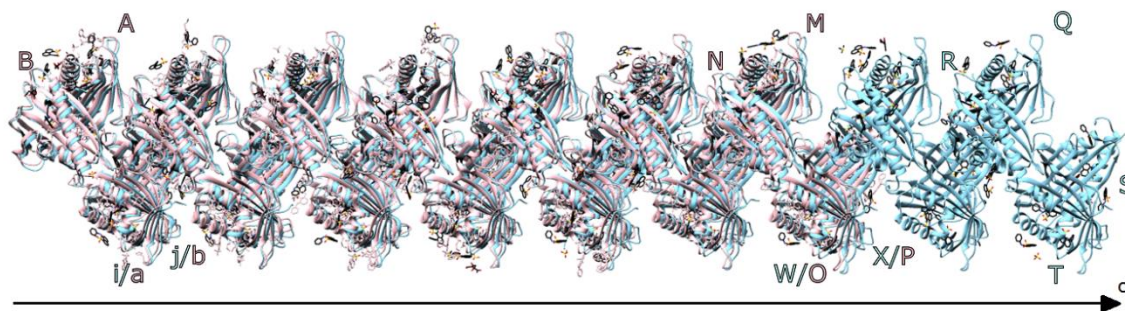
S...U	1,85	-1,75	42,56	0,72
T...V	0,58	-1,79	40,55	0,56
U...W	-1,34	1,57	43,47	0,74
V...X	-0,29	1,17	44,77	0,64
W...Y	0,48	-1,10	43,97	0,82
X...Z	-0,09	-1,24	40,52	0,73
Y...a	1,29	-0,21	41,15	0,83
Z...b	1,58	0,15	44,53	0,69
a...c	-1,28	0,67	44,56	0,59
b...d	-1,04	2,22	44,13	0,72
c...e	2,05	-0,79	40,88	0,48
d...f	1,09	-2,48	41,32	0,62
e...g	-2,39	1,31	44,58	0,50
f...h	-1,27	2,42	44,66	0,52
g...i	2,23	-1,18	41,27	0,57
h...j	0,69	-2,21	40,73	0,51

Rysunek 31. Zaburzenia porządku translacyjnego wzdłuż osi krystalograficznych a, b i c kolejnych cząsteczek Hyp-1 w rzędzie I i II, wyrażone w Å jako składowe odpowiednich wektorów translacyjnych. Translacje obliczono traktując jako punkt odniesienia atom C α Lys83 o stabilnej pozycji w białku. Naruszenia porządku translacyjnego wzdłuż a i b (oczekiwana wartość 0 Å) wyróżniono kolorem w zależności od stopnia ich nasilenia jako zielone (perturbacje rzędu 0-1,0 Å), pomarańczowe (perturbacje rzędu 1,01-2,0 Å) i czerwone (perturbacje powyżej 2,0 Å) [110].

Rys. 30 oraz Rys. 31 wskazują na niewielkie zaburzenia uporządkowania translacyjnego nie tylko wzdłuż kierunku osi c, ale także a i b, a także nieprawidłowości w rozmieszczeniu molekuł liganda ANS zarówno w obrębie, jak i pomiędzy łańcuchami białkowymi. Pomimo tego schemat odległości między środkami molekuł białka oraz populacji ANS, będące strukturalnymi manifestacjami funkcji AMF, jest ściśle odtwarzany po dziewięciu powtórzeniach głównego motywu strukturalnego, co pozwala na opis tej współmiernej strukturalnej modulacji w rozszerzonej superkomórce.

6.4. Podobieństwo struktury 9Hyp/ANS do modulowanej struktury 7Hyp/ANS

W oczywisty sposób schemat upakowania 9Hyp/ANS wykazuje wyraźne podobieństwo do układu znanego ze struktury krystalicznej kompleksu 7Hyp/ANS [41]. Oba wzory zaczynają się w tym samym punkcie komórki elementarnej wzdłuż c, dzięki czemu podział na rzędy I i II jest zachowany, przy czym główna różnica polega na tym, że motyw 9Hyp/ANS zawiera o 8 molekuł Hyp-1 więcej (28 molekuł w 7Hyp/ANS vs. 36 molekuł dla 9Hyp/ANS), a tym samym jest wydłużony o dwie tetrazy molekularne wzdłuż c (Rys. 32).



Rysunek 32. Efekt nałożenia na siebie modeli struktur 7Hyp/ANS oraz 9Hyp/ANS oraz przyjęta konwencja nazewnictwa łańcuchów. Widoczne wydłużenie w kierunku c o 2 jednostki strukturalne (8 łańcuchów białka).

Pierwsze siedem dimerów modelu 9Hyp/ANS z rzędu I, tj. AB, CD, . . . , MN, jest bezpośrednio nałożonych na tak samo oznakowane cząsteczki Hyp-1 w rzędzie I 7Hyp/ANS, a początek i osie komórek elementarnych obu struktur są uzgodnione. Dodatkowe kolejne dimery 9Hyp/ANS, tj. OP, QR i ST, UV, tworzą kolejno koniec rzędu I i początek rzędu II. Odległości pomiędzy środkami masy sąsiednich cząsteczek białka w 9Hyp/ANS wahają się od 41 do 45 Å, co stanowi zakres odległości między dimerami porównywalny z 7Hyp/ANS. Chociaż dimery białkowe w obu strukturach mają w przybliżeniu dwukrotną oś rotacji, to kąt różni się znacząco od idealnego 180° w przypadku 13 z 18 dimerów 9Hyp/ANS. Szczegółowa analiza ujawniła perturbacje pozycji białek i cząsteczek w rzędach wzdłuż osi *a* i *b* w zakresie od 0,01 do 2,74 Å, przy czym najwyższe odchylenie w kierunku *a* zaobserwowano pomiędzy łańcuchami H i J w rzędzie I. Analiza śladów atomów węgla C α łańcuchów polipeptydowych w obu strukturach nie ujawniła większych odchyień między cząsteczkami białek tych modeli ani żadnych powtarzających się wzorów zwiększonego/zmniejszonego podobieństwa. Odchylenia średniokwadratowe położenia C α nakładanych na siebie łańcuchów wynosiły od 0,47 Å (łańcuchy H i c z 7Hyp/ANS i 9Hyp/ANS) do 1,10 Å (łańcuchy D i A).

Bardziej charakterystyczną cechą różnicującą te dwie struktury jest znacznie większa liczba cząsteczek liganda w kompleksie 9Hyp/ANS. Jednakże, obsadzenie cząsteczek białka przez ligandy ANS nie wpłynęło w znaczącym stopniu na konformację łańcucha głównego. Na przykład w strukturze 9Hyp/ANS wartości RMSD C α dla łańcuchów białkowych o drastycznie różnej liczbie ligandów (np. łańcuch B, w którym siedem z ośmiu miejsc wiążących było w pełni zajętych i łańcuch Y z tylko jednym częściowo obsadzonym ligandem ANS) wynosiło zaledwie 0,68 Å, i wartość ta była równa wartości obliczonej dla dwóch podobnie obsadzonych łańcuchów B i F. Ponadto superpozycja molekuly V (cząsteczki DMS i MEL związane w miejscach nr 6 i 7) na łańcuchy f (puste miejsca wiązania nr 6 i 7) i d (gdzie miejsca nr 6 i 7 są w pełni zajęte przez ANS) ujawniła większe podobieństwo pomiędzy śladami atomów węgla C α łańcuchów V i f (0,56 Å) niż między

łańcuchami V i d (0,70 Å). Obserwacja ta sugeruje, że wiązanie składników buforu krystalizacyjnego innych niż ANS (na przykład MEL, DMS lub FLC) wywierała jedynie niewielki wpływ na charakterystyczną konformację Hyp-1. Jako kolejne podkreślenie różnic w populacji białko-ligand pomiędzy dwoma strukturami, zaobserwowano, że w strukturze 7Hyp/ANS były obecne dwie cząsteczki białka (T i V) bez żadnych wewnętrznych ligandów ANS, podczas gdy w strukturze 9Hyp/ANS znajdował się co najmniej jeden wewnętrzny ligand w każdej cząsteczce Hyp-1 (a w łańcuchu W dokładnie jeden). Porównanie śladów atomów węgla C α łańcuchów W (9Hyp/ANS) i T (7Hyp/ANS) zawierających ANS tylko w dwóch międzywęzłowych miejscach wiązania dało wartość RMSD równą 0,54 Å.

6.5. Udokładnienie modelu 9Hyp/ANS

Model 9Hyp/ANS zawierał 36 niezależnych molekuł białka w jednostce asymetrycznej, oznaczonych jako A, B, ..., Z, a, b, ..., i, j. Udokładnienie struktury 9Hyp/ANS przeprowadzono na zestawie danych dyfrakcyjnych o rozdzielczości sięgającej 2,3 Å z izotropowymi parametrami przesunięcia atomowego (ADP) i zakończono przy R_{work} i R_{free} wynoszących odpowiednio 0,226 i 0,257, co zebrano w Tabeli 2.

Tabela 2. Dane dotyczące zakończonego udokładniania 9Hyp/ANS.

Program udokładniania	REFMAC5
Liczba refleksów używanych/testowych	123350/3974
R_{work}/R_{free}^a	0,226 / 0,257
Molekuły białka/ANS w ASU	36 / 156
Liczba atomów białka/ligandów/wody/innych	44814 / 3276 / 152 / 176
Średnie ADP białka/ligandów/wody/innych (Å ²)	63,2 / 58,3 / 58,3 / 56,3
RMSD bond lengths (Å) / bond angles (°)	0,013 / 1,72
Procent par kątów ψ/φ w preferowanych/dozwolonych/zabronionych niewłaściwych obszarach wykresu Ramachandrana	92,62 / 6,87 / 0,51

Kod w bazie PDB	6SJJ
-----------------	------

^a R_{free} obliczono analogicznie jak R_{work} ze Wzoru (1) na losowo wybranym zbiorze refleksów wyłączonych z udokładniania.

Do obliczenia parametru R_{free} wykorzystano zbiór 3974 refleksów wybranych przy użyciu programu SHELXPRO [13] z cienkich powłok rozdzielczości, tak, aby uwzględnić refleksy związane symetrią niekrystalograficzną oraz refleksyrefleksów pochodzących od domen bliźniaczych. Dla cząsteczek białek zastosowano standardowe więzy stereochemiczne [112], podczas gdy więzy geometryczne dla cząsteczek ANS zostały stworzone na bazie, wykorzystując struktury soli amonowej ANS znalezionej w Cambridge Structural Database (CSD) [113] pod identyfikatorem AMANNS [114] oraz pakietu oprogramowania Mercury [115]. Wagi ograniczające dopuszczalne odstępstwa od długości wiązań, kątów i planarności płaszczyzn atomowych zostały zdefiniowane jako odchylenia standardowe wynoszące odpowiednio 0,01 Å, 3° i 0,01 Å.

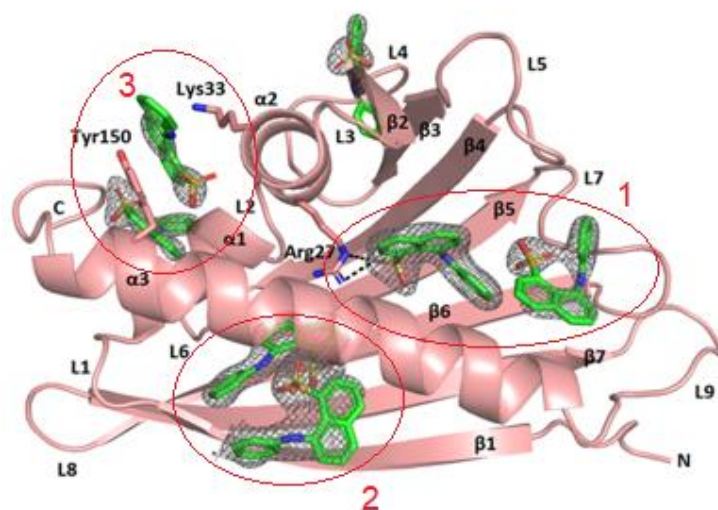
Otrzymane na początku mapy gęstości elektronowej były słabej jakości, wykazywały przerwy w obrębie łańcuchów głównych białek i rozmytą gęstość elektronową dla wielu łańcuchów bocznych. Po kilku rundach ręcznej przebudowy w programie Coot przeplatanej z automatycznym udokładnianiem REFMAC odtworzono mapy gęstości elektronowej w stanie, który umożliwił bez przerw prześledzenie wszystkich 36 łańcuchów głównych białek. Udało się również uzyskać wyraźne mapy gęstości elektronowej dla wielu łańcuchów bocznych, z wyjątkiem kilku łańcuchów bocznych fenyloalaniny i metioniny na N-końcach łańcuchów białkowych, które zostały wyłączone z modelu. Dalsza ręczna przebudowa modelu struktury w programie Coot, próby modelowania dodatkowych cząsteczek liganda/wody/innych molekuł w rozmytych obszarach gęstości elektronowej oraz próby modelowania anizotropowego ruchu/nieuporządkowania parametrami TLS nie poprawiły statystyki wskaźnika R; dlatego zdecydowano o zakończeniu udokładniania na tym etapie. Dobrą jakość modelu końcowego potwierdził dobrze uporządkowany zarys C-końcowej helisy α_3 , która w strukturach PR-10 często jest nieuporządkowana lub słabo widoczna na mapach gęstości elektronowej. Udało się również uzyskać wyraźną gęstość elektronową dla większości łańcuchów bocznych, co umożliwiło jednoznaczną identyfikację rotamerów.

Analizę poprawności modelu pod kątem stereochemicznym przeprowadzono w programie MolProbity [116]. Zgodnie z jego oceną 92,6% kątów torsyjnych φ/ψ łańcucha głównego znajduje się w preferowanych obszarach wykresu Ramachandrana, 6,9% w obszarach dozwolonych,, a jedynie 0,5% stanowią wartości błędy grube. Większość naruszeń konformacji szkieletu łańcucha polipeptydowego występuje w czterech zewnętrznych pętlach (L4, L7, L8 i L9), szczególnie przy Val60 i Glu131, których łańcuchy boczne w kilku cząsteczkach białka są zaangażowane w liczne kontakty międzycząsteczkowe.

Szczegółowa analiza różnicowych map gęstości elektronowej ujawniła lokalizację 156 cząsteczek ANS, jak również 22 innych cząsteczek pochodzących z roztworu krystalizacyjnego. W pobliżu wejścia E1 łańcucha białkowego V zaobserwowano duży obszar różnicowej gęstości elektronowej mF_o-DF_c na poziomie 11.3σ , która została wstępnie zinterpretowana jako cząsteczka liganda melatoniny (MEL) o wysokim średnim współczynniku termicznym B ($91,5 \text{ \AA}^2$). Jednakże hipoteza ta została wykluczona w wyniku dalszej analizy z wykorzystaniem różnicowych map gęstości elektronowej typu OMIT zaimplementowanych w programie Phenix [117]. Współczynnik korelacji między mapą a modelem wynosił zaledwie 0,49. Ostatecznie obszar ten został oznaczony pięcioma cząsteczkami wody jako niezidentyfikowany ligand (UNL). Jego lokalizacja różni się od położenia cząsteczek MEL w kompleksie Hyp-1-MEL, które zostały zidentyfikowane tylko w wewnętrznej komorze białka i na jego powierzchni [118]. Ostateczny model zawiera dziesięć anionów siarczanowych, które można uznać za widoczne resztki grup sulfonowych ANS lub za niezależne jony siarczanowe/fosforanowe. Stwierdzono, że pojedyncza w pełni zajęta cząsteczka HEPES (EPE) oddziaływała z resztami w pobliżu wejścia E1 łańcucha P w miejscu wiązania normalnie zarezerwowanym dla ANS. Model zawierał również siedem anionów cytrynianu sodu (FLC), który został zastosowany w buforze krystalizacyjnym, a także trzy cząsteczki dimetylosulfotlenku (DMS) użytego jako rozpuszczalnik ANS.

6.6. Miejsca wiązania ANS

Dokładna analiza struktury krystalicznej kompleksu 7Hyp/ANS, zawierającego 28 cząsteczek białka i 89 cząsteczek ANS wchodzących w interakcje białko-ligand, pozwoliła wyodrębnić osiem osobnych miejsc wiązania ANS (1-8) przedstawionych na Rys. 33 oraz ich dalszy podział na podgrupy [41,42,110,119].



Rysunek 33. Dostępne miejsca wiązania cząsteczek ANS wraz z towarzyszącą im gęstością elektronową na przykładzie łańcucha L struktury 9Hyp/ANS. Główne miejsca wiążące są

ulożone w dwóch wewnętrznych komorach (1, 2) oraz w głębokiej kieszeni powierzchniowej (3). Pozostałe miejsca 4-8 to pozycje międzywęzłowe.

Do pierwszej grupy należą trzy miejsca wiązania wewnętrznego 1-3, z których miejsce 3 znajduje się we wgłębieniu powierzchni białka, a miejsca 1 i 2 znajdują się w obszarze wewnętrznej komory hydrofobowej. Ta wewnętrzna przestrzeń jest podzielona na dwie mniejsze komory, 1 i 2, oddzielone przez Arg27 i posiadające oddzielne wejścia, odpowiednio E1 i E2 [42]. Stabilizacja cząsteczek ligandów w komorze 1 zachodzi głównie poprzez tworzenie wiązań wodorowych pomiędzy Arg27 a grupą sulfonianową ANS. Reszty hydrofobowe, takie jak Ala140 i Phe143 z helisy α_3 zapewniają środowisko hydrofobowe środowisko dla aromatycznego rdzenia ANS, podczas gdy układ dodatkowych polarnych reszt tyrozynowych, Tyr84, Tyr101 i Tyr120, oddziela cząsteczki liganda uwięzione w dwóch wewnętrznych komorach. Miejsce 3 jest tworzone w wyniku oddziaływań typu *stacking*, gdzie cząsteczka liganda jest utrzymywana w miejscu dzięki oddziaływaniom molekularnym pomiędzy łańcuchami bocznymi Tyr150 z helisy α_3 i Lys33 z helisy α_2 .

Pozostałe miejsca wiązania nr 4-8 znajdują się na powierzchni między sąsiadującymi molekułami białka i są zwykle wykorzystywane do sklejanie powierzchni trzech (lub dwóch w przypadku miejsc 7 i 8) przylegających cząsteczek Hyp-1 [119]. ANS w powierzchniowych miejscach 4, 5 i 6 łączy przylegające łańcuchy białkowe, tworząc motyw przypominający helisę. Cząsteczki ANS w miejscach 4 i 5 łączy trzy łańcuchy białkowe (dwa kolejne łańcuchy z tego samego rzędu i trzeci z przeciwległego rzędu) i są stabilizowane przez wiązania wodorowe pomiędzy atomami O ANS a grupą peptydową Gly47 z pętli L4 lub Gly110 z pętli L8. W miejscu 5 stabilizacja cząsteczki liganda następuje poprzez oddziaływania hydrofobowe z Lys21 (z helisy 1) i Ala77 (z pętli L6). W miejscu 6 cząsteczka ANS może wiązać dwa lub trzy łańcuchy białkowe w zależności od jej położenia. Cząsteczki ANS w miejscach 7 i 8, które są odpowiedzialne za łączenie dwóch sąsiednich łańcuchów białkowych, są zakotwiczone odpowiednio w pobliżu wejścia E1 i E2 i są stabilizowane przez oddziaływania z resztami z pętli L3 i L5.

6.7. Symulacje białka Hyp-1 metodami dynamiki molekularnej

6.7.1. Model Hyp-1 użyty do symulacji

Odrębne symulacje białka Hyp-1 metodami dynamiki molekularnej przeprowadzono jako metodę weryfikacji rozkładu czynników przemieszczenia atomowego ADPs oraz konformacji łańcuchów bocznych w naturalnym, bliższym biologicznemu środowisku oraz z pominięciem więzów narzuconych podczas udokładniania struktur krystalicznych. Jako współrzędne wyjściowe wykorzystano model struktury białka pozyskanego z dziurawca zwyczajnego (*Hypericum perforatum*) Hyp-1 zdeponowany w bazie PDB pod symbolem 3IE5 [90]. Dane dyfrakcyjne zostały zmierzone do rozdzielczości 1,69 Å, a model udokładniono do parametrów R_{work}/R_{free} równych odpowiednio 0,170 i 0,206.17,0 i 20,6 . Asymetryczna część komórki elementarnej zawierała dwie niezależne cząsteczki białka (A i B) tworzące dimer

poprzez oddziaływania wiązań wodorowych pomiędzy ich arkuszami β . Łańcuch A składał się ze 160 reszt aminokwasowych (159 reszt oznaczonych jako Met1-Ala159 z łańcucha głównego oraz sztucznie wprowadzonej molekuly Thr-1 pozostającej jako artefakt po etapie klonowania i oczyszczania). Natomiast cząsteczka B zawierała 164 reszty aminokwasowe, w tym 5 dodatkowych cząsteczek artefaktów (Ile-5...Thr-1). Oprócz białek, model wyjściowy zawierał różne cząsteczki glikolu polietylenowego PEG400 pochodzące z roztworu krioprotektanta oraz związane w wewnętrznej komorze i na powierzchni białka. Symulacji metodami MD poddano cały model struktury, jednak podczas analizy danych pominięto nieinteresujące cząsteczki PEG400 ze względu na ich dużą elastyczność i stosunkowo słabą widoczność w mapach gęstości elektronowej. Ostateczna asymetryczna komórka elementarna użyta do symulacji MD zawierała dwie cząsteczki Hyp-1, 258 cząsteczek wody, pojedynczy anion Cl⁻ zlokalizowany przy reszcie Lys56A oraz 10 cząsteczek glikolu polietylenowego o różnej długości łańcucha.

6.7.2. Algorytm symulacji MD białka Hyp-1

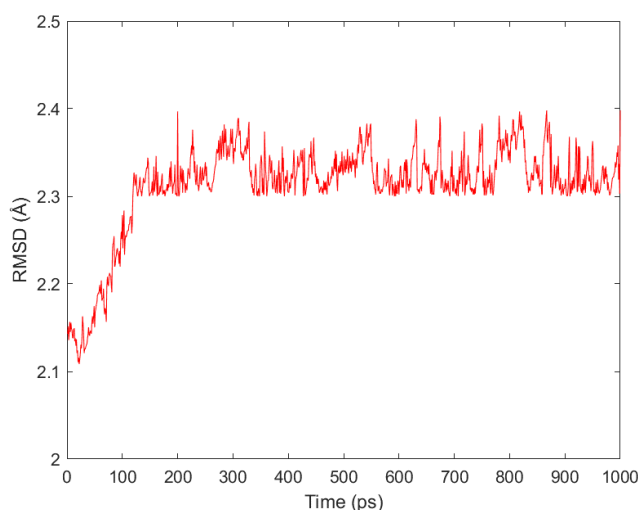
Wszystkie symulacje dynamiki molekularnej zostały przeprowadzone w otwartym pakiecie symulacyjnym GROMACS [76]. We wszystkich symulacjach wykorzystano pole siłowe ogólnego przeznaczenia OPLS-AA zawierające wszystkie niezbędne potencjały potrzebne do przeprowadzenia symulacji. Jako model wody zastosowano TIP4P-Ew [120]. Wszystkie reszty aminokwasowe Glu i Asp przyjęto jako ujemnie naładowane, podczas gdy wszystkie reszty Lys i Arg były dodatnio naładowane (dla zrównoważenia ładunku liczba ujemnie i dodatnio naładowanych reszt była zbliżona, pozostałe ładunki pochodziły m.in. od anionu Cl⁻), natomiast pozostałe reszty przyjęto jako obojętne. Wewnętrzna baza programu GROMACS zawierająca dane stereochemiczne stosowanych modeli cząsteczek została uzupełniona o topologie różnych glikoli obecnych w symulowanym układzie. Ładunki cząstkowe atomów w obrębie cząsteczek glikolu uzyskano poprzez dopasowanie ładunków punktowych do potencjału elektrostatycznego obliczonego na podstawie gęstości ładunku uzyskanej w wyniku obliczeń DFT [121,122] z wykorzystaniem orbitalnego zestawu bazowego 6-311G** typu gaussowskiego (EMSL Basis Set Exchange) z potencjałem wymiennie-korelacyjnym B3LYP [123,124] w pakiecie NWChem [125].

Model wyjściowy obejmował dimer Hyp-1 powtórzony czterokrotnie w komórce elementarnej, a następnie powielony w każdym kierunku (x, y, z), tworząc superkomórkę struktury 2x2x2. Ostateczny układ zawierał zatem 64 indywidualne kopie białka ($2 \times 4 \times 2 \times 2 \times 2 = 64$).

Strukturę wyjściową poddano procedurze minimalizacji energii osiągając szybko minimum bez znaczących zmian strukturalnych. Następnie przeprowadzono pierwszą symulację przy zachowaniu stałej liczby cząsteczek (N), objętości (V) oraz temperatury układu (T). Składała się ona ze 100 000 kroków trwających po 1 fs każdy, w których temperatura została ustalona na 292 K, oddziaływania krótkozasięgowe (van der Waalsa i Coulomba) uległy odcięciu na odległość 3,2 Å, natomiast oddziaływania dalekozasięgowe

zostały obliczone przy pomocy algorytmu PME (Particle Mesh Ewald) [126,127]. Do utrzymania odległości między sąsiadami danej cząsteczki i uniknięcia kolizji atomów zastosowano strukturę Verleta [128], aktualizując listę sąsiadów co 20 kroków symulacji. We wszystkich trzech wymiarach zastosowano okresowe warunki brzegowe, a prędkości początkowe atomów wygenerowano z rozkładu Maxwella przy temperaturze 292 K.

Po przeprowadzeniu pierwszej symulacji równoważącej NVT (stała liczba cząsteczek, objętość i temperatura) wygenerowaną strukturę poddano drugiej symulacji przy stałej liczbie cząsteczek (N), ciśnieniu (P) i temperaturze (T). Symulacja ta obejmowała 1 000 000 kroków po 1 fs każdy przy takich samych ustawieniach oddziaływań i potencjałów jak we wcześniejszej symulacji. Zastosowany izotropowy barostat utrzymywał ciśnienie na poziomie 1 bara. Stabilizacja średniego RMSD podczas symulacji pozwoliła określić, że pomimo jej krótkiego trwania udało się osiągnąć stabilizację systemu i zgromadzić dane (Rys. 34).



Rysunek 34. Średnie RMSD dla wszystkich atomów obliczane w czasie dla struktury białka Hyp-1 podczas trwającej 1 ns (1 000 000 kroków \times 1 fs) drugiej symulacji NPT. Przedstawione wartości zostały uśrednione dla 64 monomerów w kryształach.

Dane dotyczące zmian pozycji atomów zmierzono podczas trzech dodatkowych symulacji w skalach czasowych 100 ps, 10 ps i 1 ps, składających się z odpowiedniej liczby kroków czasowych 1 fs. Pozycje atomów były zapisywane do pliku trajektorii odpowiednio co 200 fs, 20 fs i 2 fs. Następnie dane te wykorzystano do obliczenia kątów odpowiadających poszczególnym konformacjom łańcuchów bocznych aminokwasów, czyli rotamerom (Tab. 3).

Tabela 3. Definicje atomów tworzących kąty torsyjne danej reszty aminokwasowej zaimplementowane podczas symulacji.

	Pierwszy kąt torsyjny	Drugi kąt torsyjny	Końcowy kąt torsyjny
Arg	N-CA-CB-CG	CA-CB-CG-CD CB-CG-CD-NE	CG-CD-NE-CZ
Asn	N-CA-CB-CG		CA-CB-CG-OD1
Asp	N-CA-CB-CG		CA-CB-CG-OD1
Cys	N-CA-CB-SG		
Gln	N-CA-CB-CG	CA-CB-CG-CD	CB-CG-CD-OE1
Glu	N-CA-CB-CG	CA-CB-CG-CD	CB-CG-CD-OE1
His	N-CA-CB-CG		CA-CB-CG-ND1
Leu	N-CA-CB-CG		CA-CB-CG-CD1
Lys	N-CA-CB-CG	CA-CB-CG-CD CB-CG-CD-CE	CG-CD-CE-NZ
Met	N-CA-CB-CG	CA-CB-CG-SD	CB-CG-SD-CE
Phe	N-CA-CB-CG	CA-CB-CG-CD1	
Pro	N-CA-CB-CG		
Ser	N-CA-CB-OG		
Thr	N-CA-CB-OG1		
Tyr	N-CA-CB-CG	CA-CB-CG-CD1	
Val	N-CA-CB-CG1		

Po przeprowadzeniu równoważących symulacji NVT i NPT struktura została schłodzona do 100 K (temperatura, w jakiej prowadzone są pomiary krystalograficzne białek) w kolejnej symulacji składającej się z 1 200 000 kroków po 1 fs każdy. Temperatura pozostawała na poziomie 292 K przez pierwsze 100 ps symulowanego czasu, następnie była liniowo obniżana do 100 K w ciągu 1 ns symulowanego czasu i pozostawała na poziomie 100 K przez końcowe 100 ps symulowanego czasu.

Ostatnim etapem było zmierzenie średniego kwadratu odchyień pozycji atomów od położenia równowagi do wyliczenia symulowanych ADPs. W tym celu schłodzona do 100 K struktura poddawana była symulacji składającej się z 1 000 000 kroków po 1 fs, a pozycje atomów były zapisywane do pliku trajektorii co 2 ps.

6.8. Proces przekształcania superkomórki 7Hyp/ANS w jej reprezentację w przestrzeni wielowymiarowej (3+1)D

Szczegóły dotyczące przekształcania z superkomórki do reprezentacji w nadprzestrzeni (3+1)D obejmowały tylko przypadek 7Hyp/ANS. Konwencja zapisu w przestrzeni wielowymiarowej (3+1)D jest taka, że komórka elementarna jest określana przez a_1, a_2, \dots, a_n zamiast a, b i c . Wektory w przestrzeni odwrotnej to $a_1^*, a_2^*, \dots, a_n^*$ zamiast a^*, b^* i c^* . Odległości są mierzone jako wartości x_1, x_2, \dots, x_n zamiast x, y i z , zaś refleksy są wskaźnikowane jako h_1, h_2, \dots, h_n zamiast h, k, l, m . Struktura 7Hyp/ANS składała się z siedmiokrotnego rozszerzenia (do superkomórki) komórki zawierającej cztery łańcuchy białkowe i kilka małych cząsteczek w jednostce asymetrycznej (ASU, Asymmetric Unit) komórki elementarnej. Wektor modulacji przebiegał w jednym kierunku, wzdłuż

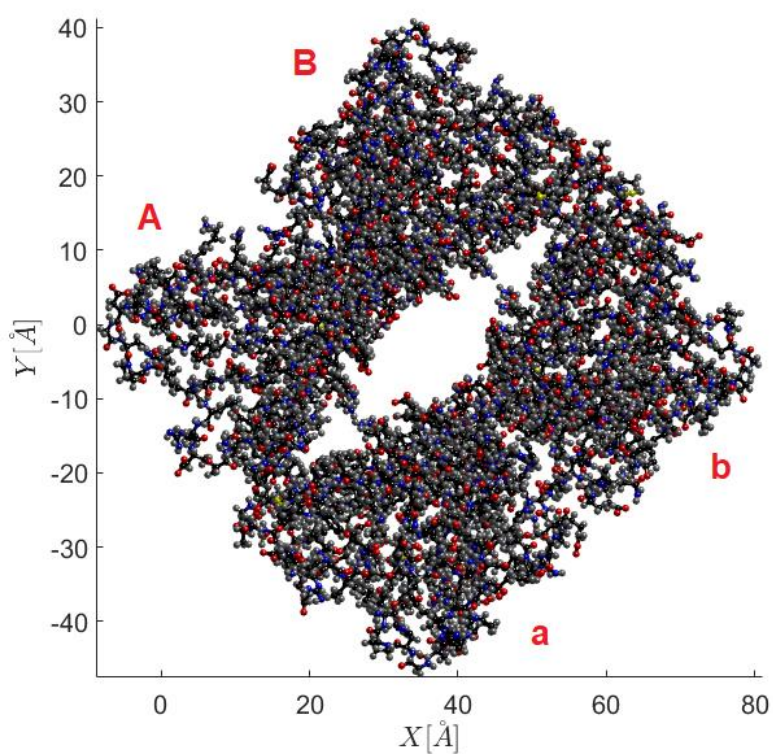
wymiaru x_3 (Rys. 29). Docelowa superkomórka zawierała 28 łańcuchów w ASU, ale łańcuchy białkowe składające się na pierwszą komórkę to A, B, a i b. Wszystkie pozostałe łańcuchy były powiązane poprzez modulację w kolejnych pojedynczych komórkach. Na przykład pierwsza lista molekuł reprezentowała łańcuchy w komórkach 1-7 (A, C, E, G, I, K, M), które były wzajemnie powiązane przez tNCS z łańcuchem A z pierwszej komórki podstawowej (Tabela 4).

Tabela 4. Łańcuchy wzajemnie powiązane operacją translacyjnej NCS w kolejnych komórkach podstawowych, zaznaczono również numery i rozmieszczenie cząsteczek ANS w poszczególnych miejscach wiązania, pozwalające obliczyć prawdopodobieństwo ich wystąpienia w zredukowanej komórce. Kolory oznaczają łańcuchy połączone wzajemnie operacją tNCS i nałożone na jeden z czterech łańcuchów komórki podstawowej: A, B, a lub b. Numery cząsteczek ANS są zgodne z ich oznaczeniami w wejściowym pliku *.cif* struktury.

Molekuła	1	2	3	4	5	6	7	8	Molekuła	1	2	3	4	5	6	7	8
A	202	203	201				204		B	20 1	20 2	20 5	20 4	20 3			
C	201	202	203				205	204	D	20 1	20 2						20 3
E	201	202	203				204 /20 5		F		20 1						
G	201	202	203				204 /20 5		H	20 2	20 3	20 1					
I	202	201	203				204 /20 5		J	20 1	20 2						
K	201	202	203						L	20 1	20 2	20 3	20 4	20 5			
M	201	202	203						N	20 1	20 2						20 3/ 20 4
O	202	201							P	20 1		20 2					
Q		201		203	202				R	20 2	20 1	20 3					20 4
S	201								T								
U	202	201		203	204		205		V	20 3						20 2	20 1
W	203	202					201 /20 4		X	20 1	20 2						20 3/ 20 4
Y	201	202		203	204				Z	20	20	20				20	

										1	2	3				4	
a		201							b	20 2	20 1						20 3/ 20 4

Do wyświetlenia danych w nadprzestrzeni (3+1)D zastosowano skrypt Matlab (MathWorks) `read_cif.m`, który został użyty do manipulacji plikami z bazy PDB, pozwalającymi wyświetlić dane w przestrzeni wielowymiarowej. Proces rozdzielania superkomórki i przekształcenia jej do przestrzeni czterowymiarowej (3+1)D rozpoczął się od podzielenia jej na 7 komórek podstawowych przy pomocy funkcji `reduce_data.m`. Liczba komórek podstawowych w dowolnym wymiarze jest równa mianownikowi wektora \mathbf{q} w tym wymiarze. Dla struktury 7Hyp/ANS liczba wszystkich komórek podstawowych wynosiła siedem i były one rozlokowane wzdłuż x_3 . Funkcja `reduce_data.m` zredukowała współrzędne atomów z poszczególnych komórek 2-7 powiązanych tNCS do współrzędnych jednej komórki podstawowej (Rys. 35, 36). Średnią pozycję atomową obliczono jako średnią z powiązanych atomów komórki podstawowej. Tak wyznaczona średnia pozycja została zastosowana do obliczenia pozycji x_4 danego atomu we wszystkich pozostałych komórkach podstawowych. Funkcja `reduce_data` określała w której komórce znajduje się odpowiednia molekula odpowiadająca tym z pierwszej komórki podstawowej i o ile należy ją przesunąć wzdłuż x_3 , aby nałożyć na analogiczną molekułę z pierwszej komórki podstawowej. Następnie zamieniono do celów obliczeniowych oznaczenia łańcuchów i pozostałych typów cząsteczek w postaci liter na liczby w kolejności rosnącej. Oddzielna tablica danych używana w innych skryptach przechowywała informację, które etykiety atomów zmieniono i na jakie.



Rysunek 35. Pełen obraz zredukowanej komórki elementarnej po nałożeniu na nią kolejnych łańcuchów z dalszych komórek podstawowych przy użyciu skryptu MATLAB. Atomy węgla są zaznaczone na czarno, azotu na niebiesko, tlenu na czerwono, siarki na żółto, a wodoru na biało.

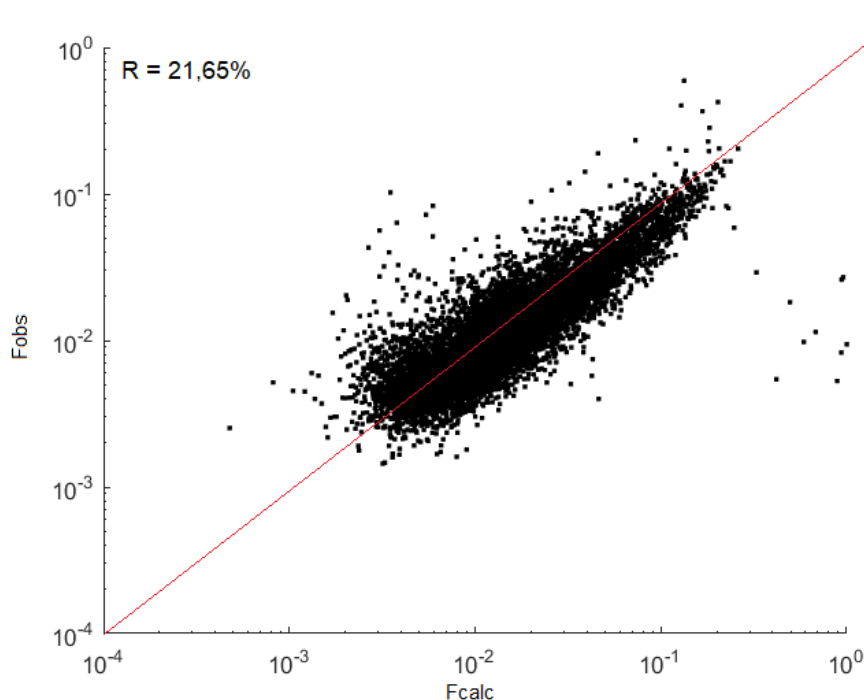


Rysunek 36. Nałożenie kolejnych łańcuchów na pierwszą komórkę 7Hyp/ANS z widokiem wzdłuż osi c. Nałożenie nastąpiło w wyniku przesunięcia łańcuchów z kolejnych komórek (2-7) o odpowiednią wielokrotność wektora translacji.

Struktura 9Hyp/ANS składała się z dziewięciokrotnego rozszerzenia komórki podstawowej w superkomórkę i podobnie jak struktura 7Hyp/ANS posiadała cztery łańcuchy białkowe w ASU komórki podstawowej. Tym samym proces przekształcania z superkomórki do przestrzeni (3+1)D byłby taki sam dla struktury 9Hyp/ANS, przy czym główna różnica polega na tym, że superkomórka 9Hyp/ANS posiada dziewięciokrotne rozwinięcie wzdłuż c , a wektor q określony dzięki wskaźnikowaniu obrazów dyfrakcyjnych w programie Crysalis Pro wynosi $4/9a_3^*$ w przeciwieństwie do $3/7a_3^*$ dla 7Hyp/ANS.

6.9. Udokładnienie modelu 7Hyp-ANS w komórce (3+1)D

Do udokładnienia modelu wykorzystano plik *.cif* dla 7Hyp/ANS zdeponowany w bazie PDB pod numerem 4N3E. Stosując opisany wyżej skrypt *read_cif.m* i wywołane w nich funkcje, otrzymano zredukowaną formę komórki elementarnej zawierającej średnie położenia atomów nałożonych na cztery łańcuchy A, B, a, b pierwszej komórki podstawowej. Jako bazę refleksów wraz z odpowiadającymi im czynnikami struktury wykorzystano informację zawartą w bazie PDB w plikach *.mtz* zapisanych z pliku binarnego do pliku tekstowego. Użyty zbiór refleksów liczył 235 350 pików uporządkowanych w pliku zgodnie z rosnącym wskaźnikiem h . Jako parametry udokładniania wybrano amplitudę modulacji, przesunięcie fazowe funkcji modulacji AMF oraz modulowane AMF. Wyniki udokładnienia przedstawiono na wykresie F_{calc}/F_{obs} , gdzie F_{calc} to czynniki struktury obliczone dla modelu 7Hyp/ANS w zredukowanej komórce elementarnej. Końcowa wartość R będąca miarą jakości dopasowania dla wygenerowanego modelu została obliczona tak jak we wzorze (1). Zależność F_{calc}/F_{obs} dla udokładnionego zbioru danych 7Hyp/ANS przedstawia Rys. 37.

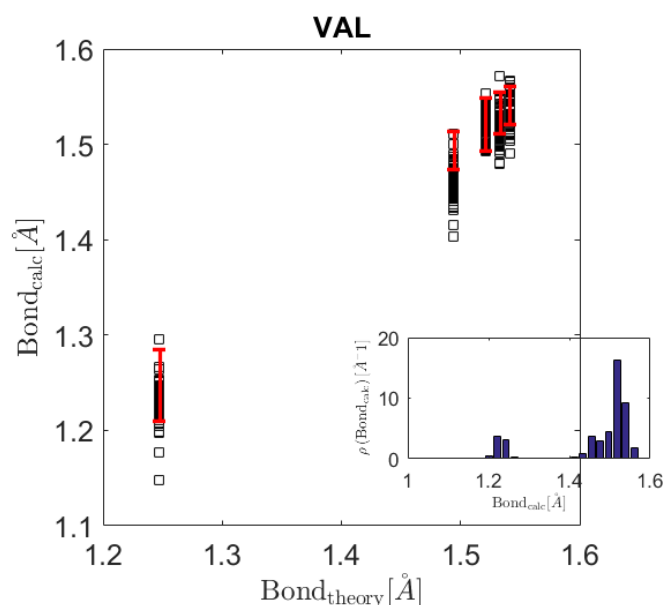


Rysunek 37. F_{calc}/F_{obs} dla zredukowanej komórki 7Hyp/ANS udokładnionej z włączoną amplitudą modulacji i przesunięciem fazowym.

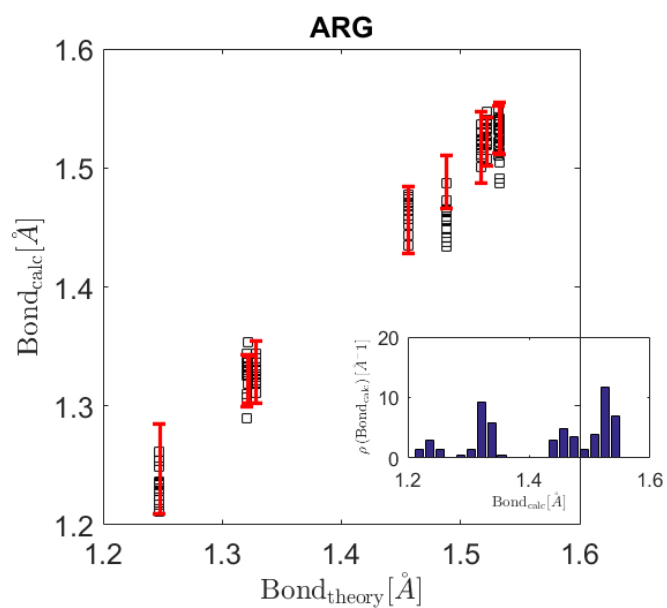
Oba zbiory danych zostały znormalizowane do najsilniejszego refleksu i tym samym przedstawione w tej samej skali. W idealnym przypadku wszystkie punkty znajdowałyby się na prostej $F_{calc} = F_{obs}$, jednak każdy pomiar amplitudy czynnika struktury obarczony był błędem. Należy również zauważyć, że każdy model pozostaje zawsze pewną teoretyczną idealizacją realnego problemu fizycznego.

6.10. Weryfikacja poprawności stereochemicznej struktury

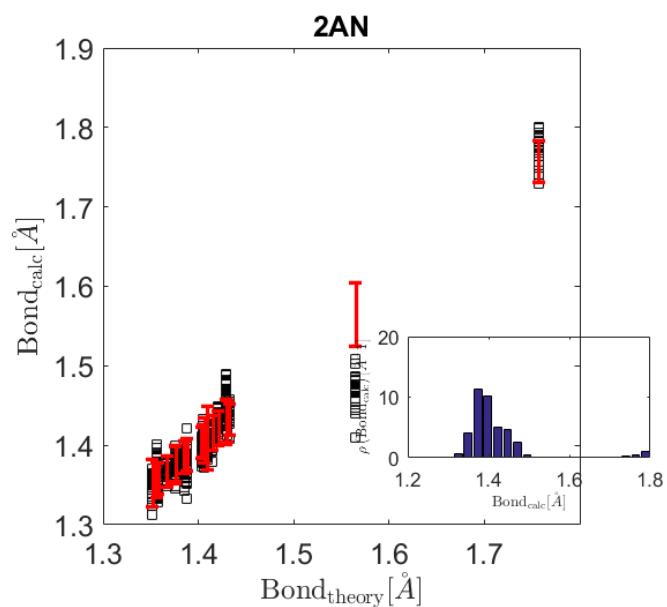
Po przeprowadzeniu udokładnienia modelu 7Hyp/ANS zastosowano skrypt `restraints_analysis.m` oraz `analyse_ADPs.m` w celu porównania obliczonych długości wiązań, wartości kątów walencyjnych oraz torsyjnych względem idealnych wartości z biblioteki danych. Porównano również rozkłady długości wiązań, kątów i kątów torsyjnych w obrębie całej populacji danego aminokwasu lub cząsteczek liganda. Dla ilustracji działania skryptu oraz określenia jakości struktury w niniejszej pracy przedstawiono takie rozkłady dla reszt aminokwasowych waliny (Val) (Rys. 38, 41, 44), argininy (Arg) (Rys. 39, 42, 45) oraz cząsteczek liganda ANS (oznaczonych jako 2AN na Rys. 40, 43, 46, gdyż pod takim symbolem cząsteczka ANS występowała w bibliotece więzów programu `.REFMAC`). Czerwone kreski na rysunkach oznaczają rozkład długości wiązań, kątów i kątów torsyjnych z biblioteki danych.



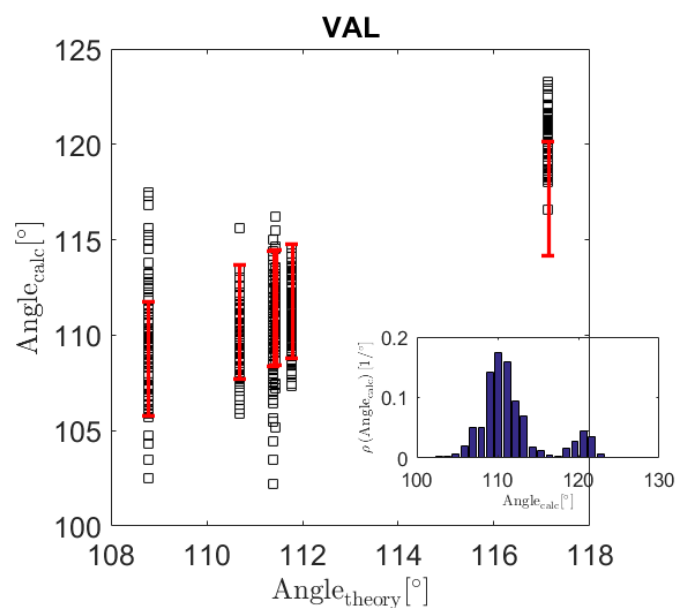
Rysunek 38. Analiza długości wiązań w udokładnionym modelu 7Hyp/ANS dla aminokwasu waliny (Val). W ramce przedstawiono rozkład obliczonych wiązań dla wszystkich molekuł Val w udokładnionym modelu.



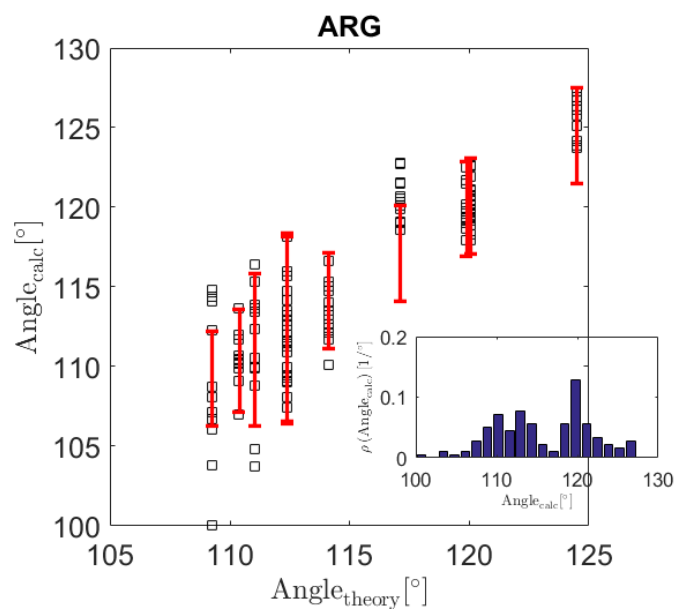
Rysunek 39. Analiza długości wiązań w udokładnionym modelu 7Hyp/ANS dla aminokwasu argininy (Arg). W ramce przedstawiono rozkład obliczonych wiązań dla wszystkich molekuł Arg w udokładnionym modelu.



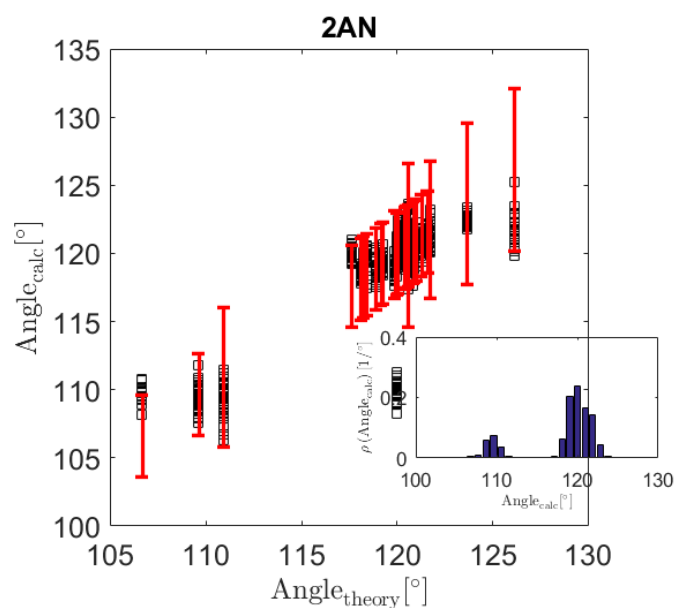
Rysunek 40. Analiza długości wiązań w udokładnionym modelu 7Hyp/ANS dla cząsteczek liganda ANS (kod w bibliotece danych 2AN). W ramce przedstawiono rozkład obliczonych wiązań dla wszystkich molekuł ANS w udokładnionym modelu.



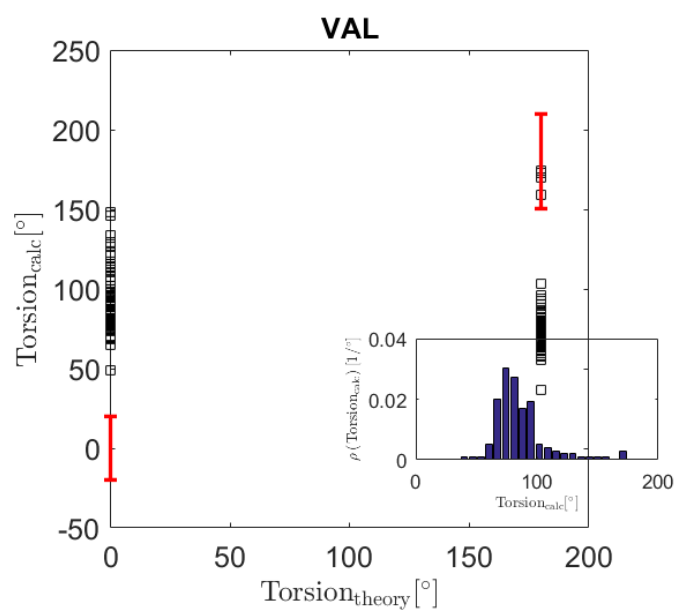
Rysunek 41. Analiza wartości kątów w udokładnionym modelu 7Hyp/ANS dla aminokwasu waliny (Val). W ramce przedstawiono rozkład obliczonych kątów dla wszystkich molekuł Val w udokładnionym modelu.



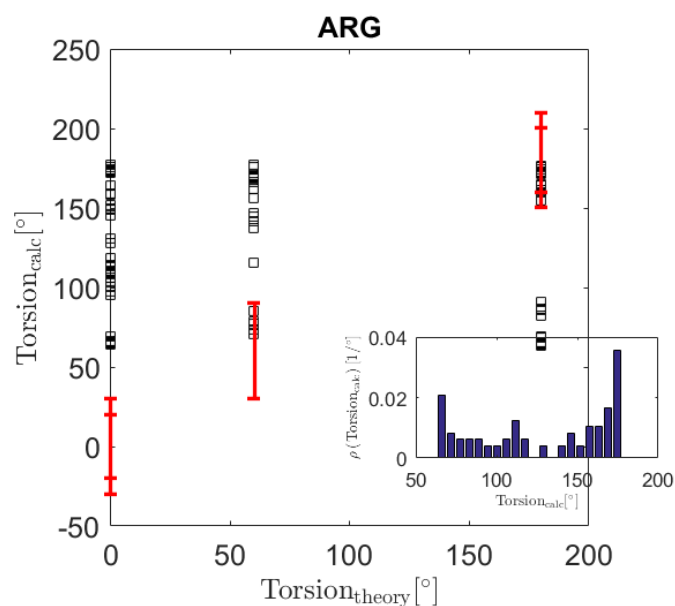
Rysunek 42. Analiza wartości kątów w udokładnionym modelu 7Hyp/ANS dla aminokwasu arginy (Arg). W ramce przedstawiono rozkład obliczonych kątów dla wszystkich molekuł Arg w udokładnionym modelu.



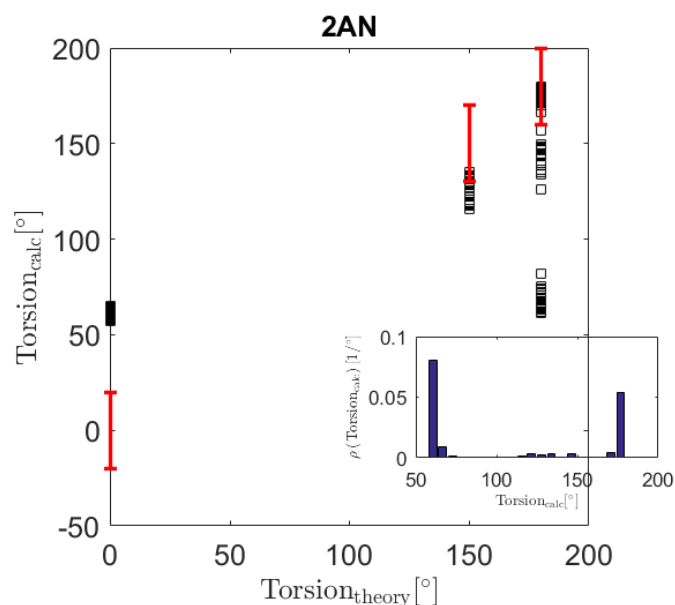
Rysunek 43. Analiza wartości kątów w udokładnionym modelu 7Hyp/ANS dla cząsteczek liganda ANS (kod w bazie 2AN). W ramce przedstawiono rozkład obliczonych kątów dla wszystkich molekuł ANS w udokładnionym modelu.



Rysunek 44. Analiza wartości kątów torsyjnych w udokładnionym modelu 7Hyp/ANS dla aminokwasu waliny (Val). W ramce przedstawiono rozkład obliczonych kątów dla wszystkich molekuł Val w udokładnionym modelu.



Rysunek 45. Analiza wartości kątów torsyjnych w udokładnionym modelu 7Hyp/ANS dla aminokwasu arginy (Arg). W ramce przedstawiono rozkład obliczonych kątów dla wszystkich molekuł Arg w udokładnionym modelu.



Rysunek 46. Analiza wartości kątów torsyjnych w udokładnionym modelu 7Hyp/ANS dla części liganda ANS (kod w bazie 2AN). W ramce przedstawiono rozkład obliczonych kątów dla wszystkich molekuł ANS w udokładnionym modelu.

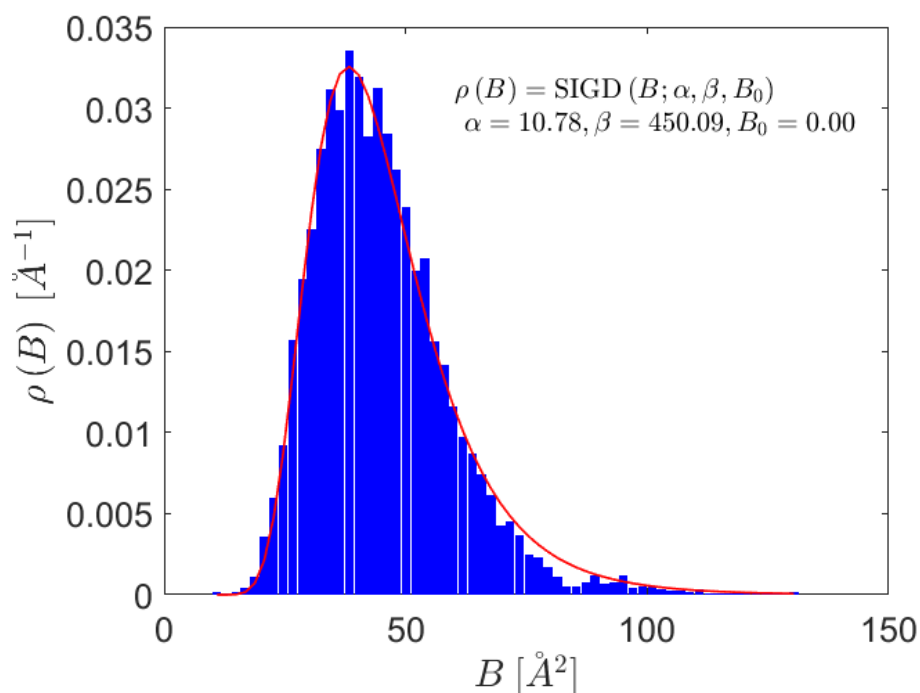
Rys. 38-43 pokazują dobrą zgodność obliczonych i modelowych długości wiązań oraz wartości kątów zwykłych dla udokładnionego modelu struktury, natomiast w przypadku kątów torsyjnych (Rys. 44-46) zaobserwowano większą dowolność przyjmowanych wartości, zarówno w przypadku aminokwasów należących do łańcucha głównego jak i swobodniej

związanych cząsteczek liganda ANS. Może to stanowić odzwierciedlenie zmian konformacyjnych zachodzących w strukturze podczas udokładniania modelu.

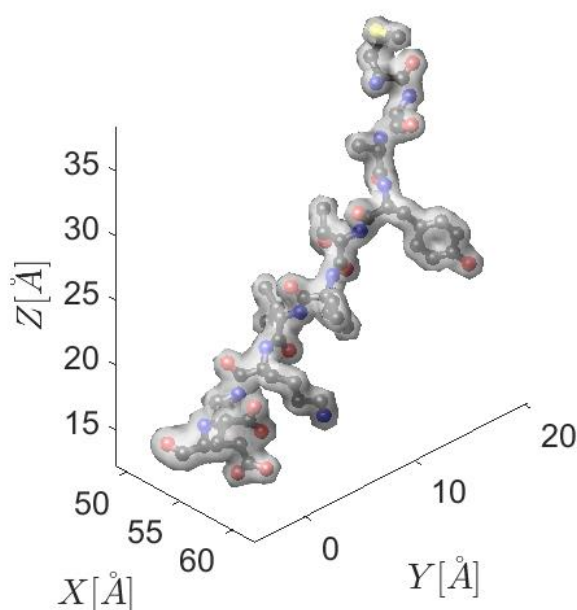
Skrypt `analyse_ADPs.m` pozwala wyznaczyć rozkład ADPs dla udokładnionej struktury wraz z dopasowanym do niego rozkładem SIGD [46]. Rozkład SIGD przyjmuje formę jak we wzorze (12):

$$\text{SIGD}(B; B_0; \alpha; \beta) = \frac{\beta^{\alpha+1}}{\Gamma(\alpha)} \frac{1}{(B-B_0)^\alpha} \exp\left(-\frac{\beta}{B-B_0}\right) \quad (12)$$

gdzie $\Gamma(\alpha)$ jest funkcją gamma, B reprezentuje wektor zawierającym zmierzone wartości ADPs, a B_0 , β i α są odpowiednio parametrami przesunięcia, skali i kształtu (Rys. 47). Na Rys. 48 przedstawiono wybrany fragment łańcucha A z wygenerowaną po udokładnieniu dzięki skryptowi `plot_structure.m` gęstością elektronową. Widoczne jest dobre dopasowanie elementów modelu do gęstości elektronowej.



Rysunek 47. Rozkład czynników temperaturowych ADPs w udokładnionym modelu 7Hyp/ANS oraz dopasowanie do niego rozkładu SIGD. W prawym górnym rogu skrypt MATLAB zwraca parametry dopasowanego rozkładu.

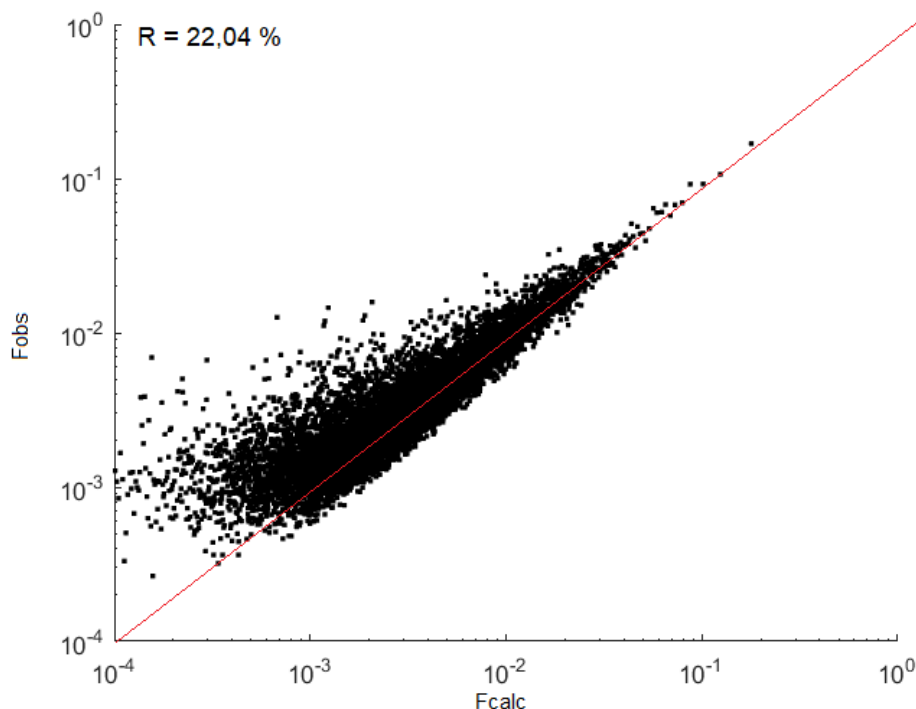


Rysunek 48. Wizualizacja wybranego fragmentu łańcucha A z komórki zredukowanej 7Hyp/ANS wraz z naniesioną na niego gęstością elektronową. Jaka mapa? Jaki kontur???

6.11. Wprowadzenie poprawek związanych z obecnością fononów

Funkcja `phonon.m` wprowadza poprawkę na obecność nieporządku wynikającego z drgań termicznych sieci krystalicznej, czyli fononów, w strukturze modulowanej. Obecność fononów i wynikający z ich obecności rozrzut położenia atomów modyfikuje klasyczny czynnik Debye'a-Wallera, a tym samym ADPs atomów w kryształach. Pomimo złamania periodyczności w strukturach modulowanych można wyróżnić obecność zarówno modów akustycznych, jak i optycznych związanych z fononami. Sprawia to, że ze względu na fonony należy wprowadzić poprawkę do czynnika struktury w formie czynnika Debye'a-Wallera. Dla takich drgań można wprowadzić opis zależności wychylenia atomu z położenia równowagi czynnikami gaussowskimi.

Opis nieporządku fononowego daje możliwość zastosowania różnych modeli. Jeśli wychylenie atomów z położenia równowagi pod wpływem drgań termicznych przybliży się funkcją harmoniczną, to w przestrzeni wektora falowego stanie się ona funkcją Bessela pierwszego rodzaju. W takim przypadku czynniki Debye'a-Wallera zastępuje się funkcjami Bessela, w którym parametrem drgań jest amplituda wychylenia atomu z położenia równowagi. Do przybliżenia nieporządku fononowego w zredukowanej strukturze 7Hyp/ANS wykorzystano podstawowy model stosujący czynniki gaussowskie, jednak funkcja `phonon.m` posiada zaimplementowane elementy umożliwiające również zastosowanie modelu z funkcjami Bessela. Na potrzeby niniejszej rozprawy doktorskiej dokonano udokładnienia modelu struktury, w którym nieporządek fononowy opisuje model z czynnikiem gaussowskim (Rys. 49). Wskaźnik R obliczono standardowo ze wzoru (1).



Rysunek 49. F_{calc}/F_{obs} dla zredukowanej komórki 7Hyp/ANS udokładnionej z włączoną amplitudą modulacji, przesunięciem fazowym oraz nieporządkiem fononowym (mod gaussowski).

7. Dyskusja

7.1. Wpływ warunków krystalizacji na modulację struktury

Podczas krystalizacji 9Hyp/ANS nieznacznie zmodyfikowano warunki krystalizacji w porównaniu do wcześniejszej modulowanej struktury kompleksu białkowego 7Hyp/ANS. Zmiana ta obejmowała dodatek melatoniny podczas inkubacji roztworu krystalizacyjnego białka z ANS, co wywołało nowy typ modulacji struktury. Stąd całkiem możliwe wydaje się otrzymanie innych wariantów modulacji nadstruktury kompleksów Hyp-1 poprzez krystalizację z użyciem innego stężenia melatoniny lub zastosowania innego dodatku do roztworu krystalizacyjnego.

Porównując parametry sieciowe 7Hyp/ANS oraz 9Hyp/ANS należałoby się spodziewać, że przy przejściu z siedmiokrotnej do dziewięciokrotnej modulacji struktury wzdłuż kierunku c , parametr sieciowy c powinien wzrosnąć $9/7$ razy, tj. z 298,56 do 383,86 Å. Tymczasem rzeczywisty parametr c kryształu 9Hyp/ANS wynosi 385,40 Å, co oznacza, że 9Hyp/ANS nie jest prostą kopią struktury 7Hyp/ANS wydłużoną o 2 jednostki strukturalne wzdłuż c , ale zupełnie odrębną strukturą modulowaną. Analogicznie, parametry a i b (spełniające zależność, że $a = b$ w przypadku obu struktur) powinny mieć takie same wartości w 7Hyp/ANS i 9Hyp/ANS. Tymczasem, wartość parametrów a i b w 9Hyp/ANS wynosi po 148,85 Å i jest większa o 1,75% niż w 7Hyp/ANS (146,29 Å). Stanowi to kolejny argument przemawiający za istnieniem rzeczywistej różnicy i odrębności obu struktur.

7.2. Artefakty pozostałe po krystalizacji

W mapach gęstości elektronowej struktury krystalicznej 9Hyp/ANS zidentyfikowano 156 cząsteczek ANS, słabo potwierdzoną melatoninę (MEL) oraz 152 cząsteczki wody. Ponadto struktura 9Hyp/ANS zawierała artefakty pozostałe po procesie krystalizacji, niekompletne cząsteczki ANS oraz inne ligandy oddziałujące z białkiem w wyróżnionych miejscach wiązania 1-8. W strukturze udało się zidentyfikować i wmodelować dziesięć anionów siarczanowych SO_4^{2-} , z których dwa zajmowały miejsce wiążące nr 2, gdzie były stabilizowane, podobnie jak cząsteczka ANS, poprzez wiązania wodorowe z atomem azotu N^{ζ} łańcucha bocznego lizyny Lys8. To silnie sugeruje szczątkowy charakter grupy sulfonianowej ANS, która była dobrze widoczna na mapach dzięki dużej liczbie elektronów i specyficznemu zakotwiczeniu pomimo nieuporządkowania pozostałej części liganda. Hipotezę tę wspierała obecność dodatnich pików gęstości elektronowej w pobliżu tych anionów siarczanowych, nawet jeśli nie można było wiarygodnie zinterpretować ich jako kompletnych cząsteczek ANS.

Pozostałe osiem anionów siarczanowych znajdowało się na powierzchni białka w obszarach normalnie zarezerwowanych dla miejsc wiązania ANS 8 (pięć jonów) i 7 (trzy jony). Zazwyczaj ich pozycja jest stabilizowana przez bezpośrednie wiązania wodorowe z dodatnio naładowanymi atomami N^{ζ} łańcuchów bocznych Lys138 i Lys139. Co ciekawe, w pięciu przypadkach znaleziono w pełni zajętą cząsteczkę ANS w odległości mniejszej niż 5 Å od anionów siarczanowych, bez żadnych bezpośrednich interakcji. Szczegółowa inspekcja obszarów różnicowej gęstości elektronowej pod koniec procesu udokładniania ujawniła pozycje siedmiu anionów cytrynianu sodu (FLC) stanowiących czynnik strącający. Cząsteczki FLC znajdowały się tylko w miejscach wiążących nr 7 i 8, tworząc bezpośrednie kontakty wodorowe z atomem $N^{\eta 1}$ łańcucha bocznego argininy Arg93 (w miejscu wiążącym nr 7) lub atomem N^{ζ} Lys134 (w miejscu nr 8). Dodatkowo wmodelowano cząsteczki sulfotlenku dimetylu (DMS) z roztworu ANS w obszarze miejsc wiążących nr 5 (łańcuch N) i 8 (łańcuchy G i V). Pojedyncze cząsteczki DMS znajdowały się w sąsiedztwie w pełni obsadzonych, dobrze widocznych cząsteczek ANS, ale udało się również zidentyfikować dwie wolne, międzywęzłowe cząsteczki DMS usytuowane pomiędzy łańcuchami N i V. Cząsteczka HEPES (EPE) pochodząca z roztworu buforowego, również zawierająca grupę sulfonianową, wyparła ANS w miejscu wiązania 7 łańcucha P. Jej dokowanie w tym miejscu było stabilizowane poprzez oddziaływanie z łańcuchem bocznym Lys138.

7.3. Analiza zbliźniczenia

Obecność elementów zbliźniczenia w kryształach 9Hyp/ANS potwierdzono wykonując dwa testy skumulowanych rozkładów intensywności zwanych L-testem oraz S(H) testem. Istotną przewagą L-testu nad testem S(H) jest jego niewrażliwość na rotację i (przy odpowiednim doborze różnicy wskaźnika l który w tym wypadku wynosił 9) na translacyjne elementy związane z NCS. Inne patologie krystalograficzne obecne w 9Hyp/ANS, takie jak translacyjny NCS, mogły zamaskować efekt zbliźniczenia poprzez

zmniejszenie intensywności całych podzbiorów refleksów. Finalnie doprowadziło to do modulacji intensywności w całych podgrupach wskaźników i dodatkowych aberracji struktury, których efekty na oczekiwany rozkład intensywności są zwykle przeciwne do tych, wynikających ze zbliźniczenia [129]. Pseudosymetria i zbliźniczenie powszechnie występują razem [130], a zatem obecność elementów pseudosymetrii wyjaśniałaby, dlaczego rozkłady intensywności są zaburzone mniej, niż można by się spodziewać w przypadku idealnego zbliźniczenia.

Testem S(H) zidentyfikowano dla każdej z grup refleksów udział frakcji bliźniaczej równy 0,45 (zamiast oczekiwanego 0,25 w przypadku idealnego zbliźniczenia tetartoedrycznego). Może to wynikać z obecności rotacyjnych elementów NCS w przestrzeni rzeczywistej, które niemal pokrywają się z potencjalnymi operatorami zbliźniczenia [131].

7.4. Obsadzenie modelu struktury cząsteczkami ANS

Chociaż miejsca wiązania ligandów są takie same zarówno w kompleksie białkowym 7Hyp/ANS jak i 9Hyp/ANS, szczegóły dotyczące lokalizacji ligandów w obrębie jednostek asymetrycznych wyraźnie różniły się w przypadku obu struktur. W strukturze 9Hyp/ANS większość cząsteczek liganda ANS (95 ze 156 zidentyfikowanych w mapach gęstości elektronowej cząsteczek) było zlokalizowanych w obrębie trzech wewnętrznych miejsc wiążących nr 1-3, przy czym 75 z nich (78,9%) występowało w pełnym obsadzeniu równym 1,0. W Tabeli 5 36 łańcuchów białkowych pogrupowano w dwa Rzędy, oznaczone jako różowe (Rząd I; AB,...QR) i zielone (Rząd II; ST,...ij). Wewnętrzne miejsca wiązania 1-3 zaznaczono na żółto. Pełne obsadzenie ligandów zaznaczono jako x lub 2x, co wskazywało liczbę cząsteczek liganda (1 lub 2) w obszarze danego miejsca wiążącego. Częściowe obsadzenie zostało oznaczone przez jego wartość ułamkową. W przypadku miejsc międzywęzłowych nr 4-8 zaznaczono również połączenie z sąsiednią cząsteczką białka, a także możliwe częściowe obsadzenie. Symbol (') wskazywał jednoznacznie obecność cząsteczek powiązanych symetrycznie z innymi miejscami wiązania. Cząsteczki ligandów inne niż ANS (DMS, FLC, EPE, SO₄) zostały oznaczone za pomocą ich skrótów i obsadzenia.

Tabela 5. Schemat wysycenia struktury 9Hyp/ANS cząsteczkami liganda [110].

Chain	Cavity-docked ANS			Superficial sites of ANS					Chain	Cavity-docked ANS			Superficial sites of ANS				
	1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8
A	x	x	0.7	-	-	-	-	0.4	B	x	x	x	g5, G6 0.7	g4, i6	-	f8	f7
C	x	x	-	-	-	-	-	0.5	D	x	x	x	-	-	0.5	-	d7
E	x	x	0.7	-	V4, T6 0.5	-	SO4 0.7	0.6	F	x	x	x	S5, Q6 0.5	S4, U6	-	b'8	b'7
G	x	x	0.7	-	-	-	0.6	0.5	H	x	x	x	-	U4, W6 0.7	-	-	Z7
I	x	x	0.7	Z5, J6 0.6	Z4, X6 0.7	-	L'8 0.7	L'7	J	x	x	x	-	W4, Y6 0.6	x	-	-
K	x	x	0.7	-	-	L4, Z5 0.7	-	-	L	x	x	x	Y5, K6	Y4, a6	-	-	-
M	x	x	-	-	-	-	-	P'7	N	x	x	x	-	DMS 0.5	x	T'8	T'7
O	x	x	0.7	-	-	-	-	-	P	x	x	x	c5, G6	c4, e6	-	j'8 EPE	j'7
Q	x	x	-	-	-	-	-	x	R	x	x	x	-	-	0.7	h'8	h'7
S	0.8	0.5	0.5	-	-	-	-	x	T	0.8	x	x	-	-	x	i'8	i'7
U	0.5	-	-	-	-	-	-	x	V	x	x	x	x	x	DMS 0.5	S'8 UNL	S'7 2x
W	0.7	-	-	-	-	-	-	-	X	x	x	x	-	-	x	U'8 2x	U'7
Y	x	-	0.6	0.7	0.7	-	0.5	x	Z	x	x	x	x	x	0.6	-	2x
a	x	x	-	-	-	0.7	x	x	b	x	x	x	K5, Y6 0.7	-	0.7	-	Y'7 2x
c	x	SO4 0.4	0.7	0.5	-	-	2x	x	d	0.5	x	x	0.7	-	x	a'8	-
e	x	-	-	-	-	0.5	FLC	0.7	f	x	x	x	x	0.5	-	-	c'7
g	x	SO4 0.5	0.7	-	-	-	x	x	h	0.7	x	x	-	-	x	f'8	f'7
i	0.7	-	-	-	-	-	FLC 0.9	FLC 0.6	j	x	x	x	x	0.7	-	x	g'7 2x

Pozostałe 61 cząsteczek liganda ANS posiadających bardzo dobrą gęstość elektronową zidentyfikowano w pięciu dodatkowych powierzchniowych miejscach wiążących nr 4-8. Żadna z molekuł białka nie była całkowicie zajęta przez cząsteczki liganda we wszystkich ośmiu miejscach wiążących, ale nie wystąpiła również pojedyncza molekula, która byłaby całkowicie nieobsadzona przez cząsteczki ANS. W dimerach AB, ..., QR tworzących rząd I, zawsze jedna z cząsteczek (A, C, E, ..., Q) miała miejsca nr 1 i 2 w pełni zajęte, natomiast miejsce nr 3 było puste lub częściowo wypełnione, a obsadzenie cząsteczką ANS wynosiło 0,7. W komplementarnych cząsteczkach tych dimerów Hyp-1 (B, D, F, ..., R) wszystkie trzy wewnętrzne miejsca wiązania ANS 1-3 były zawsze w pełni zajęte cząsteczkami ligandów posiadającymi doskonałą gęstość elektronową.

Wzór obsadzenia w obrębie rzędu II wyraźnie różni się od sytuacji napotkanej w 7Hyp/ANS, gdzie w pierwszych cząsteczkach każdego dimeru (tj. O, Q, S, U, W, Y, a) zaobserwowano wiele nieobsadzonych miejsc, przy czym miejsce 3 tych molekuł było zawsze puste. Wśród molekuł białka należących do łańcuchów T i V w ogóle nie zaobserwowano obecności wewnętrznych ligandów w miejscach nr 1-3. W przypadku 9Hyp/ANS, drugie cząsteczki z każdego dimeru (T, V, X, ..., j) tworzyły wzór wysycenia ligandami podobny do rzędu I i zawsze posiadały pełny zestaw cząsteczek ANS w trzech wewnętrznych miejscach wiążących, z wyjątkiem częściowego obsadzenia w miejscach T(1), d(1) i h(1). W pierwszych molekułach białkowych II rzędu (tj. S, U, W, ..., i) wystąpiło wiele perturbacji we wzorze nasycenia cząsteczkami ANS, a wśród nich trzy łańcuchy (U, W, e) zawierały tylko jedną, częściowo obsadzoną cząsteczkę liganda w miejscu wiążącym nr 1. Chociaż miejsce nr 1 jest zawsze zajęte przez ANS, to miejsce nr 2

w łańcuchach c i g było wypełnione anionem siarczanowym, który mógł pochodzić z buforu oczyszczającego bądź krystalizującego lub, co bardziej prawdopodobne, był jedynym widocznym sygnałem sulfonianowym pochodzącym od słabo uporządkowanej cząsteczki ANS.

W 7Hyp/ANS cząsteczki ANS generalnie obserwowane były w poszczególnych miejscach wiązania tylko w pełnym obsadzeniu, z wyjątkiem jednego częściowo zajętego miejsca nr 3 w łańcuchu R. Tymczasem w strukturze 9Hyp/ANS występuje aż 50 cząsteczek ligandów (32,0% wszystkich miejsc obsadzonych przez cząsteczki ANS) z częściowym obsadzeniem. Biorąc pod uwagę wszystkie 3 wewnętrzne miejsca wiążące w 36 łańcuchach białkowych ($36 \times 3 = 108$), stwierdzono, że miejsce nr 1 było zawsze zajęte przez cząsteczkę ANS, miejsce nr 2 miało pięć wakatów oraz dwa częściowo obsadzone aniony siarczanowe zamiast ANS, natomiast miejsce nr 3 było puste w ośmiu na 36 przypadków. W modelu 7Hyp/ANS natomiast wśród $28 \times 3 = 84$ możliwości miejsce nr 1 było puste w pięciu przypadkach, miejsce nr 2 posiadało trzy wolne miejsca, a miejsce nr 3 było puste w 15 przypadkach. Fakt, że w przeciwieństwie do 7Hyp/ANS, gdzie prawie wszystkie cząsteczki ANS występują w pełnym obsadzeniu, wiele cząsteczek ANS w modelu struktury 9Hyp/ANS wykazuje tylko częściowe obsadzenie, może być przypisany efektowi konkurencji między cząsteczkami ANS i melatoniny podczas procesu wzrostu kryształu. Schemat obsadzenia miejsc wiążących nr 4-8 na powierzchni białka jest znacznie bardziej skomplikowany – w strukturze 9Hyp/ANS zidentyfikowano 61 powierzchniowo związanych cząsteczek ANS zgrupowanych w pięciu dostępnych miejscach nr 4-8. Zaobserwowano również, że oprócz głównego liganda ANS, również inne cząsteczki pochodzące z roztworu buforu lub krystalizacyjnego takie jak EPE, DMS, FLC i prawdopodobnie MEL, są również często spotykane w miejscach powierzchniowych. Ogólnie rzecz biorąc, w 7 łańcuchach białkowych zidentyfikowano dziesięć cząsteczek ANS, przy czym występowały one z najwyższą częstotliwością w miejscu nr 7 (pięć cząsteczek ligandów). W łańcuchach V z II rzędu oraz i z I rzędu, dwa ligandy inne niż ANS były równocześnie przyłączone w miejscach nr 6 i 7 oraz w miejscach nr 7 i 8. Liczba powierzchniowych cząsteczek ligandów w miejscach wiążących 4-8 zmieniała się w zależności od łańcucha wzdłuż rzędów. Z reguły każde z powierzchniowych miejsc wiązania posiadało wiele wakatów w każdym z białkowych rzędów. Podobnie jak w 7Hyp/ANS, nie stwierdzono obecności łańcucha białkowego posiadającego cząsteczki liganda związane w każdym z jego powierzchniowych miejsc wiążących. Częstość obsadzenia miejsc międzywęzłowych rośnie począwszy od miejsca nr 4 (zajętego w 13 łańcuchach) do miejsca nr 8 (29 łańcuchów). Nałożenie wszystkich cząsteczek ANS na jeden wspólny ślad $C\alpha$ pochodzący od kolejnych łańcuchów Hyp-1 wskazało na stabilność położenia ligandów w miejscach wiążących nr 1, 2, 3, 4 i 5. W przypadku miejsc nr 6, 7 i 8 występowała znaczna zmienność pozycyjna i dowolność konformacyjna cząsteczek ANS, co było widoczne jako niedoskonałe nakładanie się poszczególnych cząsteczek liganda.

Niemniej, ich położenie względem odpowiednich elementów struktury drugorzędowej białka zostało zachowane.

7.5. Konformacja cząsteczek ANS i ich wpływ na formowanie nadstruktury

Geometrię cząsteczek ANS analizowano poprzez badanie trzech kątów torsyjnych, τ_1 (C2-C1-S-O), τ_2 (C7-C8-N-C11) i τ_3 (C8-N-C11-C), które opisują orientację grupy sulfonianowej i podstawnika anilinowego względem płaszczyzny naftalenu, a także całkowitą orientację pierścienia fenyłowego cząsteczki aniliny. Konformacja cząsteczek ANS zależała od kombinacji kątów torsyjnych 1, 2 i 3 oraz była zależna od miejsca wiązania, które wymuszało lokalną deformację ligandów. Zazwyczaj cząsteczki ANS cechowała znaczna zmienność rotacyjna, wyrażona faktem że kąt torsyjny τ_1 osiągał wartości bliskie 0° , co wynikało z lokalnej geometrii anionu grupy RSO_3^- . Zaobserwowano, że wartości kątów torsyjnych nr 3 były bliskie 0° , co jest sprzeczne z geometrią referencyjnej struktury AMANNS (27°), chociaż wartości pozostałych kątów 1 i 2 odpowiadają dość dobrze geometrii referencyjnej struktury liganda użytej podczas udokładnienia (Tabela 6).

Tabela 6. Zaobserwowane konformacje ANS opisane przez średnie wartości kątów torsyjnych τ_1 , τ_2 i τ_3 ($^\circ$) oraz ich odchylenia standardowe dla ligandów zgrupowanych w odpowiednich miejscach wiązania. Dane porównano z modelową cząsteczką ANS (oznaczoną w bazie CSD, Cambridge Structural Database jako AMANNS).

Kąt torsyjny \ Miejsce	1	2	3	4/5/6	7/8	AMANNS
C2-C1-S-O ^a (τ_1)	-2(2) 3(2)	0(2) 1(2)	-1(2) 2(2)	-2(3) -5(2)	2(10) -1(2)	1
C7-C8-N-C11 (τ_2)	48(7) 38(2)	32(4) 12(1)	20(8) -2(6)	28(10) -21(5)	18(6) 11(6)	42
C8-N-C11-C ^b (τ_3)	1(1) 11(3)	1(1) 24(5)	-1(1) -3(11)	-1(2) -31(4)	2(1) 1(12)	27

^a Atom O sulfonianu, który minimalizuje $|\tau_1|$. ^b Atom C aniliny, który minimalizuje $|\tau_3|$.

Konformacje ANS opisane średnimi wartościami kątów torsyjnych τ_1 , τ_2 i τ_3 wraz z ich odchyleniami standardowymi dla ligandów zgrupowanych w poszczególnych miejscach wiązania zebrano w Tabeli 6. Jako odniesienie wykorzystano modelową strukturę AMANNS z bazy danych CSD. Powierzchniowe miejsca wiązania odpowiadające za łączenie dwóch lub trzech sąsiadujących łańcuchów struktury 9Hyp/ANS zostały podzielone na dwie grupy. Ligandy 4, 5 i 6 odpowiadają za łączenie trzech sąsiadujących łańcuchów Hyp-1. Powoduje to, że numeracja jest płynna i dana cząsteczka ANS może być oznaczona jako 4, 5 lub 6, w zależności od najbliższego miejsca wiązania w obrębie łańcucha Hyp-1, z którym oddziałuje. Podobnie połączono miejsca nr 7 i 8, ponieważ łączą dwa łańcuchy białkowe. Otrzymane wyniki porównano z danymi uzyskanymi dla 7Hyp/ANS (zaznaczone w Tabeli kursywą). Znaczne odchylenie standardowe kątów torsyjnych w powierzchniowych cząsteczkach ANS w miejscach nr 4/5/6 (trzy łańcuchy) oraz 7/8 (dwa łańcuchy) wskazują na dużą swobodę konformacyjną cząsteczek ANS w tych miejscach oraz

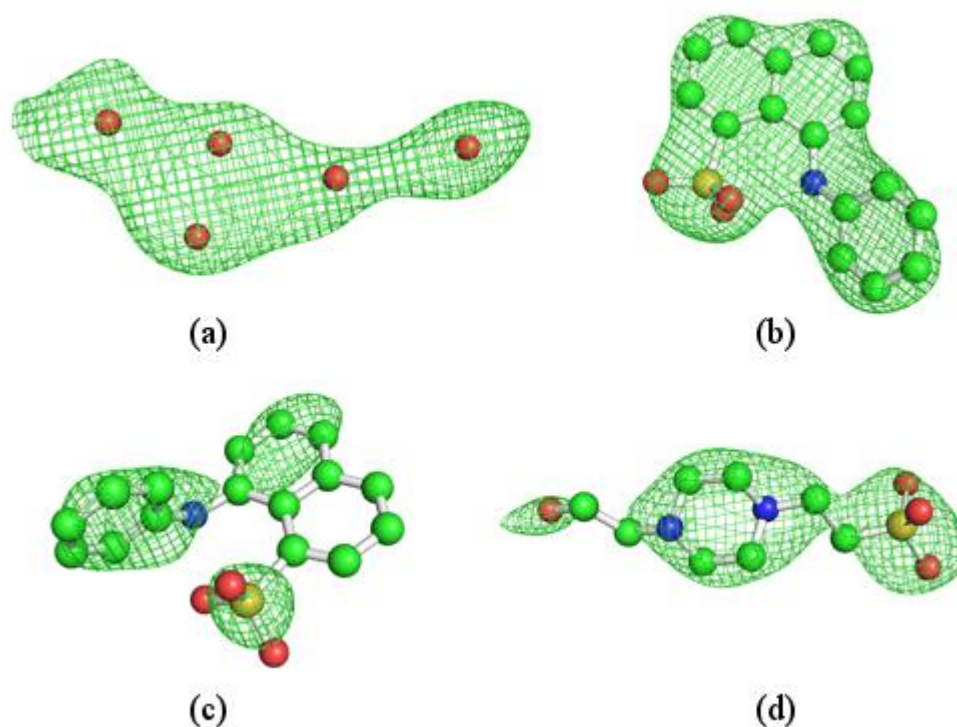
znaczące odkształcenie od modelu referencyjnego. Oddziaływania z łańcuchami bocznymi aminokwasów w miejscach wiązania pozwala ustabilizować cząsteczkę ANS i dopasować jej geometrię do istniejących ograniczeń sterycznych. Dane zebrane w Tabeli 6 wspierają hipotezę dotyczącą roli ANS w łączeniu („sklejaniu”) sąsiadujących łańcuchów białka 9Hyp/ANS oraz ich roli w formowaniu unikalnej struktury przestrzennej tego kompleksu białkowego w przestrzeni rzeczywistej superkomórki.

7.6. Identyfikacja obecności liganda melatoniny w mapach gęstości elektronowej

Białka należące do grupy PR-10, w tym Hyp-1, cechuje obecność hydrofobowej wnęki, która może służyć jako miejsce dokowania i wiązania specyficznych małowcząsteczkowych ligandów o znaczeniu biologicznym. Ostatnie badania strukturalne prowadzone przez Śliwiak i in. [118] wykazały, że białka PR-10 mogą specyficznie wiązać fitohormon o właściwościach antyoksydacyjnych, melatoninę (MEL) [131,132]. Badania kompleksu białka Hyp-1 z melatoniną prowadzone dzięki metodom krystalograficznym potwierdziły istnienie wyróżnionych miejsc wiązania liganda 1, 2 i 3, które wcześniej były znane jedynie z kompleksu 7Hyp/ANS. Analiza struktury kompleksu 7Hyp/ANS wykazała jednak, że w przeciwieństwie do liganda MEL, cząsteczka ANS jest zdolna do wiązania z białkiem również w miejscach powierzchniowych. Testy wypierania ADA potwierdziły, że nawet w wysokich stężeniach MEL nie jest w stanie zastąpić cząsteczki ANS w białku, co objawia się silnym sygnałem fluorescencji pochodzącym od ANS we wszystkich warunkach testowych [40]. Aby ocenić ewentualną konkurencję pomiędzy ANS i MEL o wewnętrzne miejsca wiązania Hyp-1, w przypadku 9Hyp/ANS dr Joanna Śliwiak podjęła próbę krystalizacji białka w obecności mieszaniny obu ligandów – MEL i ANS. W rezultacie otrzymano kolejną modulowaną strukturę krystaliczną kompleksu Hyp-1/ANS, co nasuwa wniosek, że cząsteczka ANS może być lepszym spoiwem dla molekuł Hyp-1 niż naturalny hormon melatonina i wspomagać proces łączenia łańcuchów w złożony wzór upakowania. Ponieważ w drodze eksperymentu nie udało się uzyskać kryształów Hyp-1-ANS metodą współkrystalizacji przy stosunkach MEL:ANS wyższych niż 1:2 (np. 1:1, 2:1 lub 3:1) wywnioskowano, że zachodzi silna konkurencja o dostęp do miejsc wiążących między MEL a ANS oraz ANS nie mógł być wypierany przez MEL, kiedy jego stężenie było co najmniej dwukrotnie wyższe niż MEL w warunkach współkrystalizacji.

Niemniej jednak, wśród licznych (178) cząsteczek ligandów struktury 9Hyp/ANS, wskazano obecność co najmniej jednego miejsca, V(7), w którym mogło nastąpić wiązanie melatoniny. Wewnętrzne miejsce wiążące były już zablokowane przez ANS tak jak zaobserwowano w strukturze 7Hyp/ANS. Porównanie struktury z 9Hyp/ANS wykazało, że MEL nie posiada zdolności do zastąpienia ANS w charakterze czynnika łączącego sąsiednie łańcuchy białka. Stwierdzono, że kształt gęstości elektronowej w miejscu V(7) był zgodny z kształtem liganda MEL, szczególnie biorąc pod uwagę płaski pierścień indolowy MEL. Jednakże próby modelowania cząsteczki ANS bądź superpozycji MEL i ANS w tym miejscu nie przyniosły rezultatu, a nawet pogorszyły parametry udokładnienia. Pomimo tych

trudności pojedyncza w pełni obsadzona cząsteczka melatoniny została wstępnie wymodelowana w tym obszarze, zaś tę interpretację weryfikowano przy użyciu różnicowych map gęstości elektronowej typu OMIT wygenerowanych przy użyciu programu .polder z pakietu Phenix. Istotną zaletą tej metody nad innymi powszechnie stosowanymi metodami generowania map gęstości elektronowej stanowi wyłączenie rozpuszczalnika z wybranego obszaru OMIT, co skutkuje poprawą interpretacji powstałej mapy. Analiza map polder OMIT MEL wskazała, że gęstość elektronowa pierwotnie przypisana MEL prawdopodobnie oznacza cząsteczkę rozpuszczalnika lub szum zamiast liganda. Tym samym uznano, że obecność pojedynczej cząsteczki MEL w tym miejscu nie znajduje potwierdzenia w danych eksperymentalnych. Jak widać na Rys. 50a, z powodu niepewnej interpretacji region ten został zamaskowany kilkoma cząsteczkami wody oznaczonymi jako UNL. W równoległych badaniach analiza map polder potwierdziła obecność ANS w miejscach o dobrze zdefiniowanej gęstości elektronowej (Rys. 50b), jak również w kilku bardziej problematycznych i trudnych do interpretacji pozycjach (Rys. 50c). Mapy gęstości elektronowej typu OMIT potwierdziły również obecność prawidłowo modelowanej pojedynczej cząsteczki HEPES w miejscu wiązania P(7) (Rys. 50d) oraz będącej artefaktem krystalizacji cząsteczki cytrynianu (FLC) z roztworu krystalizacyjnego.



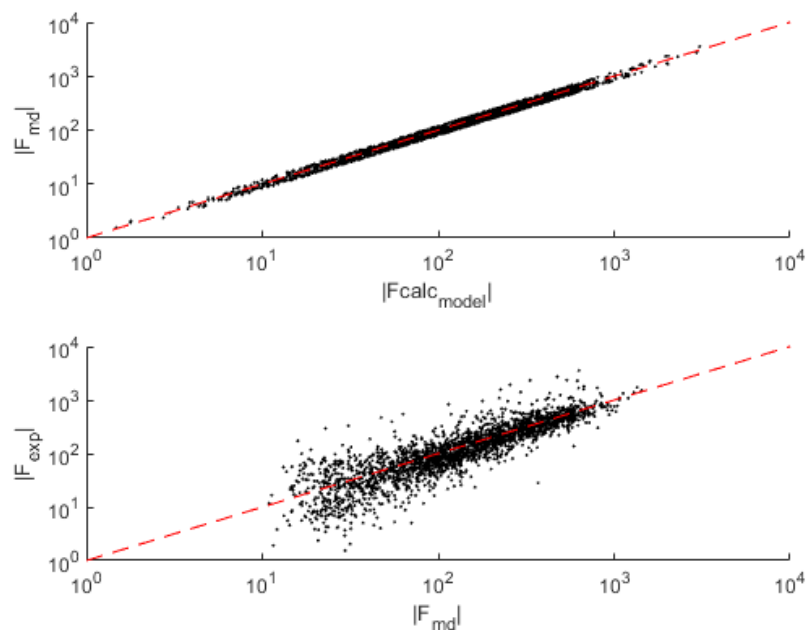
Rysunek 50. Mapy OMIT konturowane przy 3.0σ dla wybranych ligandów 9Hyp/ANS: (a) gęstość elektronowa w miejscu V(7) wstępnie interpretowana jako MEL, ale ostatecznie oznaczona pięcioma cząsteczkami wody (czerwone kule); (b) ANS w miejscu P(1) z dobrą definicją w gęstości elektronowej; (c) słabo zdefiniowana cząsteczka ANS w miejscu S(2); oraz (d) pojedyncza cząsteczka HEPES w miejscu P(7).

Chociaż podejrzenie obecności pojedynczej cząsteczki MEL w mapach gęstości elektronowej nie znalazło potwierdzenia, stwierdzono, że zmiana upakowania kryształu oraz wysycenia ANS w strukturze 9Hyp/ANS jest pośrednią wskazówką wpływu melatoniny na proces współkryształizacji oraz tworzenia struktur modulowanych w kompleksach Hyp-1 z ligandami. W strukturze 7Hyp/ANS cząsteczki ANS zawsze występowały w mapach gęstości elektronowej w pełnym obsadzeniu z jednym wyjątkiem na poziomie 0,5 i nie było nawet śladu resztkowych gęstości elektronowych ANS w pustych miejscach. W strukturze 9Hyp/ANS zaobserwowano, że aż 50 ze 156 cząsteczek ANS (32%) występuje w częściowym obsadzeniu. Nie można tego uzasadnić porównywalną rozdzielczością obu struktur (2,4 vs 2,3 Å). Obecność MEL w mieszaninie krystalizacyjnej i efekt konkurencji mogły osłabić wiązanie ANS z Hyp-1, prowadząc w niektórych przypadkach do częściowego obsadzenia. Dodatkowo, nawet nieznacząca konkurencja między MEL i ANS przy wewnętrznych miejscach wiążących mogła prowadzić do wyparcia części ANS i zwiększenia liczby molekuł dostępnych dla wiązań między łańcuchami białka. Tym samym mogło to doprowadzić do zwiększenia potencjału ANS do łączenia sąsiednich molekuł Hyp-1 oraz formowania bardziej skomplikowanego upakowania w komórce elementarnej.

7.7. Porównanie eksperymentalnego i symulowanego metodami MD modelu struktury Hyp-1

Porównanie wyjściowej, eksperymentalnej struktury białka Hyp-1 oraz modelu uzyskanego metodami dynamiki molekularnej miało na celu stwierdzenie, czy przeprowadzona symulacja nie wprowadziła znaczących, niepożądanych zmian konformacyjnych w wyjściowym modelu struktury. W tym celu porównano czynniki struktury dla refleksów zmierzonych doświadczalnie oraz wygenerowanych po symulacji MD. Dla eksperymentalnego modelu Hyp-1 zmierzono 39 745 niezależnych refleksów, z czego 1590 wykorzystano jako zbiór testowy do obliczenia parametru R_{free} . Z symulacji uśredniono po wszystkich 64 molekułach Hyp-1 modelu uzyskano zbiór 39 737 refleksów przy pomocy programu phenix.fmodel. Z porównania danych wykluczono refleksy, dla których błąd względny przekraczał 20%. Tym samym porównaniu poddano 24080 refleksów o akceptowalnym stopniu zgodności, których zestawienie pokazano w skali logarytmicznej na odpowiednich wykresach F_{calc}/F_{md} i F_{md}/F_{exp} (Rys. 51), gdzie:

- F_{calc} oznacza czynniki struktury obliczone dla udokładnianego modelu Hyp-1 przed symulacją,
- F_{md} – czynniki struktury obliczone dla modelu Hyp-1 po symulacji MD,
- F_{exp} – pierwotne czynniki struktury uzyskane w drodze eksperymentu krystalograficznego.



Rysunek 51. Wykresy F_{calc}/F_{md} (czynniki struktury wyjściowego modelu dla symulacji vs. modelu po symulacji) i F_{md}/F_{exp} (czynniki struktury modelu po symulacji vs. eksperymentalne) dla struktury białka Hyp-1.

Między wszystkimi danymi F_{md} i F_{calc} obliczono współczynnik rozbieżności R równy 10,6%, podczas gdy analogiczny parametr R obliczony między danymi F_{exp} i F_{md} osiągnął wartość 23,6%. Pomimo, że czynniki struktury dla analogicznych refleksów modelu eksperymentalnego i symulowanego zostały dopasowane z niewielkim rozrzutem, na wykresie F_{md}/F_{exp} zaobserwowano rozmycie. Prawdopodobnie jest to efekt niedoskonałego dopracowania modelu struktury lub zjawiska wielokrotnego rozpraszania. W obrębie oryginalnego modelowego dimeru Hyp-1, obie cząsteczki A i B nakładają się z wysokim RMSD ich atomów $C\alpha$ równym 1,21 Å, chociaż typowe rozbieżności strukturalne pomiędzy białkami należącymi do grupy PR-10 zawierają się w zakresie od 1,64 do nawet 2,76 Å [90]. Obliczony w programie *gesamt* RMSD dla atomów $C\alpha$ wynosi 2,38 Å pomiędzy modelem eksperymentalnym a otrzymaną średnią niskoenergetyczną konformacją Hyp-1 po MD. Ta rozbieżność wynika z ogólnej elastyczności zwoju PR-10 oraz średnich rozmiarów białka Hyp-1. Główne elementy struktury drugorzędowej zostały dobrze zachowane, natomiast, zgodnie z oczekiwaniami, większość zmian konformacyjnych zaszła w mniej stabilnych regionach pętli L3 i L5 [90]. Zmienność strukturalna Hyp-1 podczas symulacji była ograniczona więzami narzuconymi na długość kontaktów międzycząsteczkowych oraz krótkozasięgowych oddziaływań van der Waalsa.

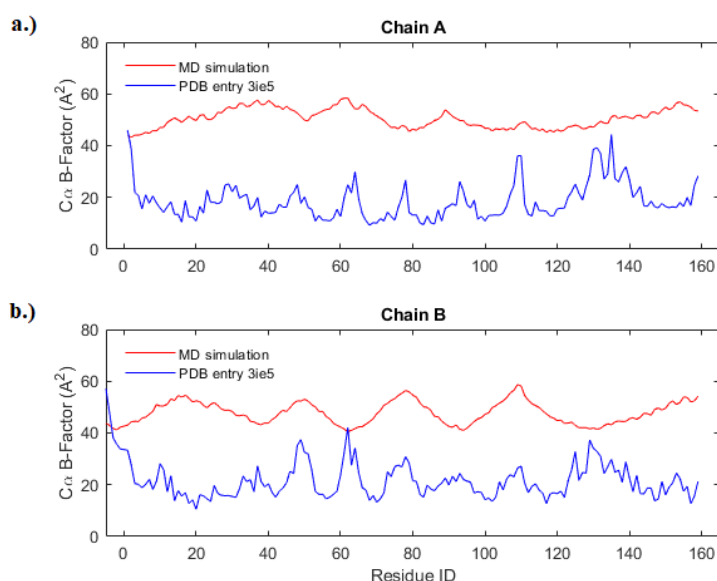
7.8. Porównanie otrzymanych i doświadczalnych ADPs

Symulacja wyjściowego modelu konformacyjnego Hyp-1 w celu obliczenia ADPs, obejmowała kilka kroków: i) trwający 100 ps etap stabilizacji temperatury układu na 292 K, ii) 1 ns symulacji równowagowej oraz iii) 1,2 ns schładzania struktury do 100 K, po

którym następowało próbkowanie fluktuacji pozycji atomów w strukturze białka. Zazwyczaj wartość RMSD pozwala na rozróżnienie pomiędzy ograniczonymi i ruchomymi częściami cząsteczki (wyższe wartości dotyczą części cząsteczki o większej swobodzie ruchu), co prowadziło do obserwacji redukcji ruchów elastycznych atomów węgla C γ (określających położenie grupy aminowej względem karboksylowej w reszcie aminokwasowej) pod wpływem pola siłowego OPLS-AA [79]. Tym samym potwierdzono, że przy danych ustawieniach symulacyjnych, udało się wygenerować zestaw zróżnicowanych strukturalnie zespołów o dużej heterogeniczności konformacyjnej, niezależnej od ruchów ciał sztywnych czy defektów sieci. Ponieważ jako model wyjściowy wykorzystano określoną krystalograficznie konformację Hyp-1, założono, że uzyskane w symulacji 64 pojedyncze cząsteczki białka (32 dimery) są wystarczającą liczbą do przedstawienia różnorodności konformacyjnej kryształów Hyp-1. ADPs zostały wyliczone standardowo na podstawie wzoru (1). Błąd standardowy pomiaru (SE) wyznaczono jako odchylenie standardowe średniej:

$$SE = \sqrt{\sum(B_i - \bar{B})^2 / N(N - 1)} \quad (13)$$

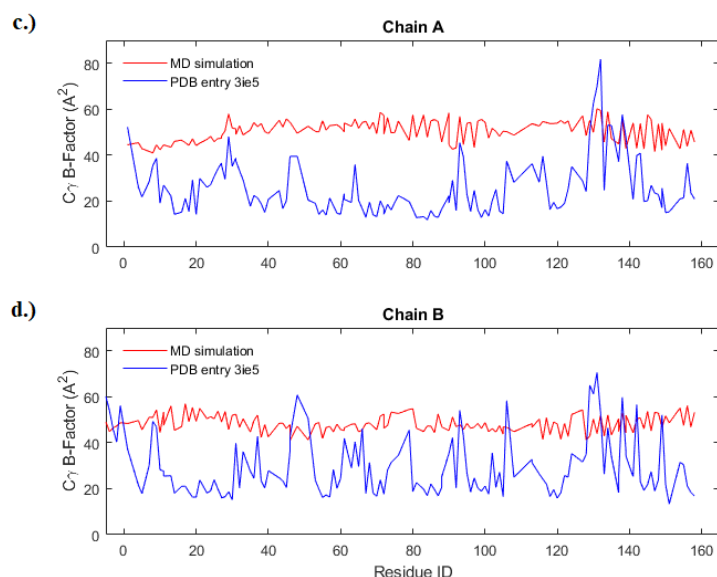
Gdzie B_i oznacza ADP i-tej konformacji (od 1 do 64), (\bar{B}) są średnie ADPs struktury po 64 symulowanych konformacjach Hyp-1, a N jest liczbą konformacji w polu symulacji ($N = 64$).



Rysunek 52. Wykresy ADPs atomów C α modelu Hyp-1 z podziałem na łańcuch A (a) oraz B (b) wygenerowane bezpośrednio z atomowo-pozycyjnego RMSD (linia czerwona) oraz eksperymentalne z PDB (linia niebieska).

Obliczone średnie izotropowe ADPs dla eksperymentalnego modelu Hyp-1 wynosiły 27,5 Å², podczas gdy średnie ADPs dla poszczególnych łańcuchów A i B wynosiły odpowiednio 23,7 Å² i 26,5 Å² (Rys. 52). Wysoka labilność i ruchliwość długich łańcuchów ligandów PEG została odzwierciedlona przez ich podwyższone średnie współczynniki B w zakresie od 49 do 71 Å² dla wewnętrznych cząsteczek PEG [45]. Tak wysokie wartości

były prawdopodobnie efektem braku silnych, kierunkowych oddziaływań białko - ligand, świadczących o dokładnym składzie roztworu PEG. W niektórych przypadkach reprezentacja liganda na mapach gęstości elektronowej była tak niepewna, że niemożliwe było stwierdzenie, czy obserwowana gęstość elektronowa odpowiada pełnowymiarowej cząsteczce, czy tylko fragmentowi wydłużonego, nieuporządkowanego łańcucha.



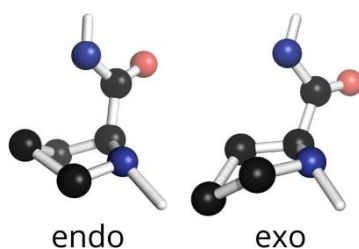
Rysunek 53. Wykresy ADPs atomów C γ łańcuchów bocznych modelu Hyp-1 z podziałem na łańcuch A (a) oraz B (b) wygenerowane bezpośrednio z atomowo-pozycyjnego RMSD (linia czerwona) oraz eksperymentalne z PDB (linia niebieska).

Rozkład eksperymentalnych ADPs odzwierciedlał różnicę w swobodzie konformacyjnej pomiędzy ograniczonymi segmentami łańcucha głównego (Rys. 512) a różnorodnością konformacyjną łańcucha bocznego (można to zaobserwować śledząc zwiększone wartości ADPs atomów C γ z łańcuchów bocznych na Rys. 53). Najwyższe wartości ruchliwości termicznej atomów C γ zaobserwowano dla łańcuchów bocznych Glu132, Arg93, Glu106 i Asp48 zaangażowanych w kontakty wodorowe między sobą lub z cząsteczkami wody na powierzchni białka (Rys. 53). Obliczony na podstawie całej symulacji atomowej RMSD w każdym stadium zachowywał niskie wartości, przy czym jego średnia wynosiła $\sim 0,2$ Å w całym polu symulacji z maksimum przy $\sim 0,25$ Å. Dlatego też obliczony rozkład ADPs jest bardziej jednorodny z ogólnie zwiększonymi wartościami współczynników termicznych w porównaniu z eksperymentalnym. Przy zastosowaniu pola siłowego OPLS-AA i w nanosekundowej skali czasu, obliczone średnie ADP były równe $50,5$ Å² (dla łańcucha A) oraz $47,8$ Å² (dla łańcucha B). Ich RMSD porównujące eksperymentalne i obliczone w wyniku symulacji ADPs atomów C α osiągały wartości (zapisane jako średnia \pm SE) $31, \pm 2,63$ Å² w łańcuchu A i $26,2 \pm 2,2$ Å² w łańcuchu B, podczas gdy całkowite RMSD pomiędzy eksperymentalnymi i obliczonymi ADP atomów C α było równe $30,5 \pm 1,7$ Å².

Całkowity RMSD obliczony tylko dla atomów C γ łańcuchów bocznych w modelu osiągał wartość 25,9 \pm 1,6 Å². Zaobserwowano, że średnie i SE dla ADPs wzrastają z czasem symulacji [45] - obliczone odchylenia pomiędzy eksperymentalnymi i obliczonymi ADP białka Hyp-1 można wyjaśnić jako wynik dłuższej trajektorii 1,2 ns.

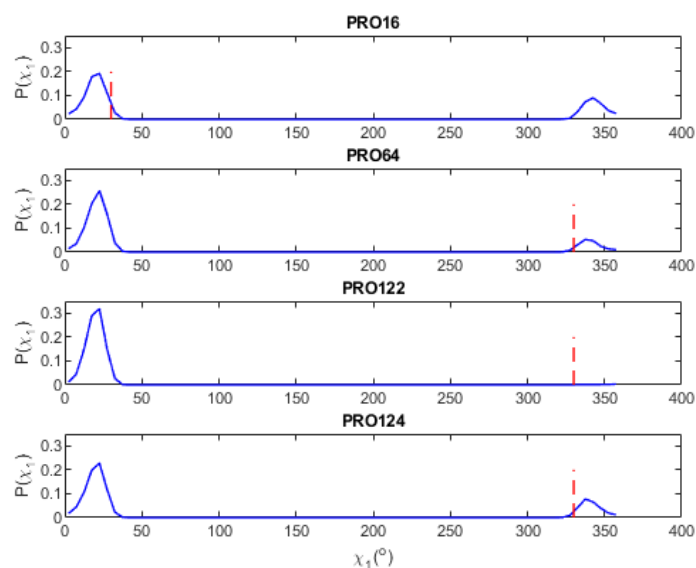
7.9. Symulowane rozkłady prawdopodobieństwa kątów łańcucha bocznego P(χ_1) dla Pro, Ser, Cys, Val i Thr

Sekwencja aminokwasowa białka Hyp-1 zawiera 7 reszt proliny (Pro), 3 reszty seryny (Ser), 2 cysteiny (Cys), 18 waliny (Val) oraz 9 reszt treoniny (Thr). Symulowane rozkłady prawdopodobieństwa kątów torsyjnych w obrębie krótkich łańcuchów bocznych tych aminokwasów opisano dla pojedynczego kąta χ_1 . Cząsteczki proliny mają tendencję do przyjmowania konformacji endo lub exo w zależności od tego, czy atom C γ znajduje się powyżej lub poniżej płaszczyzny (Rys. 54) pierścienia tworzonej przez inne atomy: C α , C β , C δ i N [133]. Ze względu na ograniczenia obliczeniowe pakietu GROMACS (brak możliwości obliczania ujemnych wartości kątów) obliczone kąty torsyjne zawierały się w zakresie 0°-360°, podczas gdy kąty torsyjne reszt aminokwasowych mają zakres -180° – 180°. Tym samym, w przypadku kątów torsyjnych o wartościach powyżej 180° należy odjąć od ich wartości 360°, aby otrzymać wartości zgodne z Penultimate Rotamer Library [136]. W tekście zawarto zarówno wartości obliczone (zaznaczone czerwonymi kreskami na rysunkach), jak i wartości zgodne z Penultimate Rotamer Library (w nawiasach okrągłych). Prawdopodobieństwa obliczone w poszczególnych przedziałach sumowano, tak że $\sum_{x_i} P(x_i) = 1$.



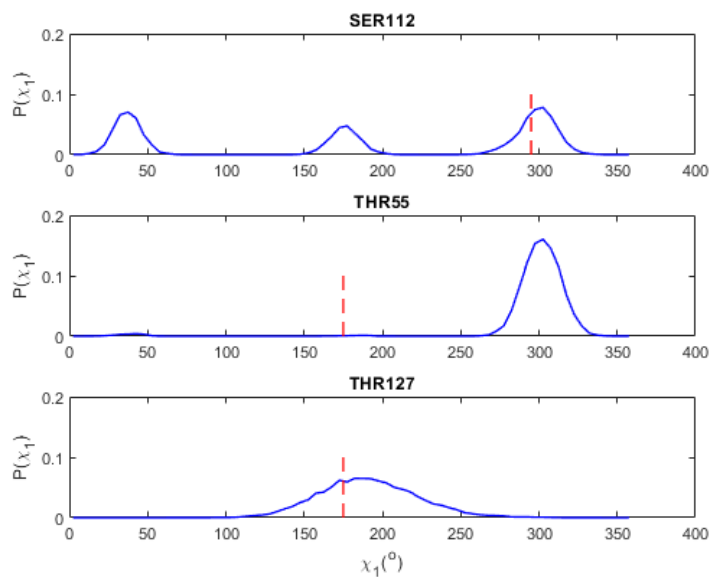
Rysunek 54. Ilustracja konformacji endo i exo reszt proliny.

W ramach modelu eksperymentalnego dimeru Hyp-1 zaobserwowano, że prolina przyjmuje 7 razy konformację C γ -endo i 7 razy konformację C γ -exo odpowiednio przy $\chi_1=30^\circ$ i 330° (-30°). Jednakże, obserwowane po symulacji MD rozkłady ujawniły przejścia pomiędzy stanami endo i exo dla Pro16, Pro64, Pro122 i Pro124 (Rys. 55) zaznaczone przez przewidywane piki P(χ_1) przy wartościach χ_1 przeciwnych do tych wynikających z konformacji struktury eksperymentalnej.



Rysunek 55. Rozkłady gęstości prawdopodobieństwa skręcenia łańcucha bocznego $P(\chi_1)$ (niebieskie linie ciągłe) dla Pro16, Pro64, Pro122 i Pro124. Wartości kątów χ_1 obserwowane w eksperymentalnym modelu Hyp-1 oznaczono jako przerywane czerwone linie [137].

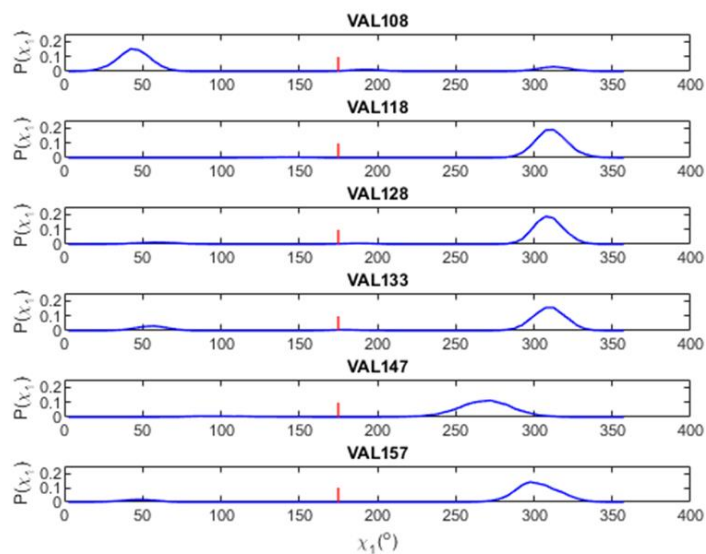
Chociaż atom C γ w większości reszt Pro występował tylko w jednej z konformacji endo-exo zarówno dla cząsteczek białka A jak i B, to ich rozkłady $P(\chi_1)$ wskazały, że konwersja do preferowanej energetycznie $\chi_1 = 30^\circ$ (forma endo) nie była całkowita. Jedynie dla Pro122, pik przy $\chi_1 = 330^\circ$ (-30°) był prawie nieobecny, co wskazywało na niskie powinowactwo do tej konformacji. Ze względu na położenie większości reszt proliny w obszarach pętli struktury Hyp-1 (z wyjątkiem Pro16 z krótkiej helisy α_1), posiadają one dużą swobodę konformacyjną ułatwiającą transformację do bardziej stabilnej konformacji endo.



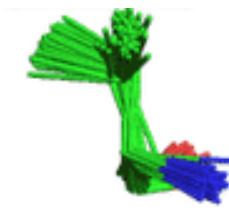
Rysunek 56. Rozkłady gęstości prawdopodobieństwa skręcenia łańcucha bocznego $P(\chi_1)$ (niebieskie linie ciągłe) dla Ser112, Thr55 i Thr127. Wartości kątów χ_1 obserwowane w eksperymentalnym modelu Hyp-1 oznaczono jako przerywane czerwone linie [137].

Dla reszt seryny i cysteiny można oczekiwać trzech wysoce prawdopodobnych konformacji łańcucha bocznego przy $\chi_1 = 62^\circ$ (forma p), $\chi_1 = 183^\circ$ (-177°) (forma t) i $\chi_1 = 295^\circ$ (-65°), konformacja m. W przypadku Cys nie natrafiono na żadne rozbieżności pomiędzy eksperymentalną i przewidywaną formą rotameryczną aminokwasu. Potwierdza to, że dobrane na podstawie analizy map gęstości elektronowej rotamery zgadzają się z istniejącymi w symulowanym modelu wolnym od narzuconych więzów stereochemicznych podczas udokładnienia. Jak pokazano na Rys. 56, dysproporcja pomiędzy eksperymentalną i symulowaną formą rotameru została znaleziona w Ser112 z pętli L8, gdzie wysoce prawdopodobna konformacja p jest oczekiwana przy $\chi_1 = 295^\circ$. Obliczony rozkład $P(\chi_1)$ ujawnił obecność trzech pików przy 295° , 183° i 30° , co sugerowało tworzenie różnych minimów energetycznych dla każdej konformacji. Odchylenie wartości pików przy $\chi_1 = 30^\circ$ od preferowanej wartości 62° wskazało na istnienie geometrycznych zniekształceń w stosunku do eksperymentalnie zdefiniowanego modelu lub energetyczne powinowactwo do jednego ze stanów niebędących rotamerem.

Trzy dominujące preferowane konformacje łańcuchów bocznych Val i Thr występują przy $\chi_1 = 60^\circ$, 175° i 300° (-60°) (Rys. 57). Natywne sekwencja białka Hyp-1 zawierała 18 reszt waliny, z czego 16 z nich, niezależnie od ich lokalizacji w strukturze drugorzędowej, przyjmowała dominujący rotamer t. W drodze symulacji MD zaobserwowano różnice pomiędzy eksperymentalnymi i symulowanymi konformacjami łańcuchów bocznych dla 17 z 18 reszt waliny białka Hyp-1, co oznaczało, że przewidywany rozkład $P(\chi_1)$ prawie zawsze posiadał pik przy ok. 300° (-60°), co odpowiadało drugiemu najbardziej prawdopodobnemu rotamerowi m.



Rysunek 57. Rozkłady gęstości prawdopodobieństwa skręcenia łańcucha bocznego $P(\chi_1)$ (niebieskie linie ciągłe) dla Val108, Val118, Val128, Val133, Val147 oraz Val157. Wartości kątów χ_1 obserwowane w eksperymentalnym modelu Hyp-1 oznaczono jako przerywane czerwone linie [137].

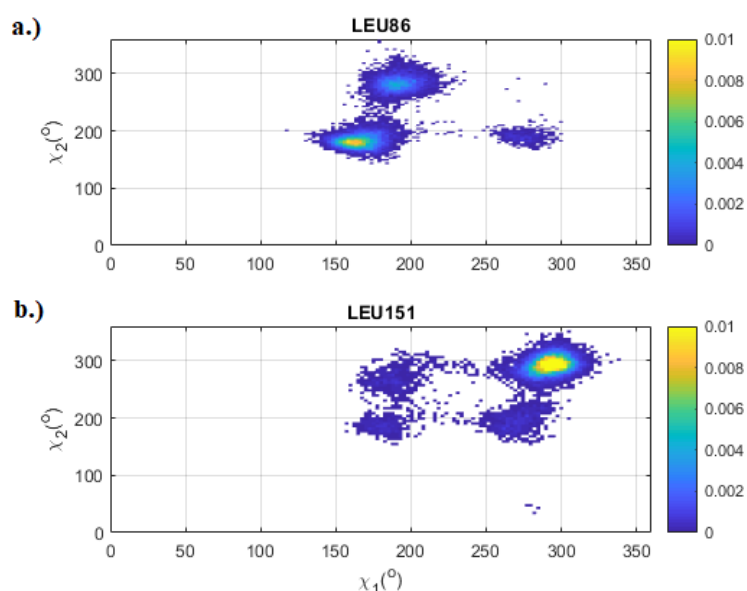


Rysunek 58. Wynik nakładania analogicznych łańcuchów bocznych Val108 w 64 symulowanych konformacjach Hyp-1.

Występowanie mniejszej liczby stanów rotamerycznych wraz z obecnością rzadszych konformacji wskazywało na dynamiczne zmiany konformacyjne w obrębie struktury Hyp-1 podczas symulacji MD (Rys. 58). Ta transformacja konformacyjna w kierunku stanów o niższej energii została wyrażona przez przesunięcie prawdopodobieństwa z rotameru t do m i wynikającą z tego zmianę preferencji χ_1 . Sekwencja Hyp-1 obejmowała 9 reszt aminokwasowych Thr, dla których 8 reszt wykazało inny symulowany rotamer z prawdopodobieństwami przesuniętymi znacząco z preferowanych p (49% prawdopodobieństwa) i m (43% prawdopodobieństwa) do trzeciej w kolejności obsadzenia konformacji Thr formy t. Dla rozgałęzionej przy atomie węgla $C\beta$ reszty aminokwasowej Thr55 zauważono odwrócenie pomiędzy dwoma dominującymi stanami rotacyjnymi, podobnie jak w przypadku wielu reszt Val. Dla Thr127 w mniej uporządkowanym regionie pętli, zaobserwowano spadek prawdopodobieństwa t, a następnie przesunięcie w kierunku preferowanego rotameru m.

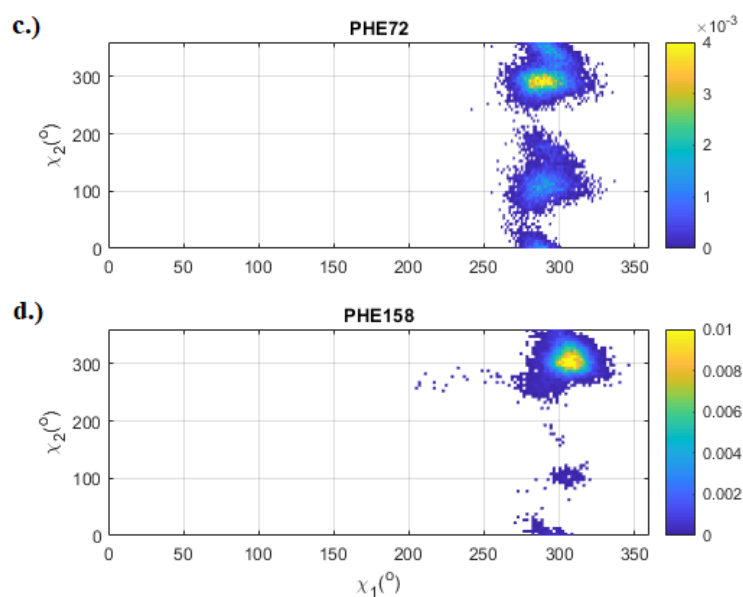
7.10. Symulowane bimodalne rozkłady prawdopodobieństwa kątów łańcucha bocznego $P(\chi_1, \chi_2)$ dla Leu, Phe, Tyr, His, Asp i Asn

Preferowane konformacje łańcucha bocznego dla leucyny (Leu), kwasu asparaginowego (Asp), asparaginy (Asn) oraz reszt aromatycznych fenyloalaniny (Phe), tyrozyny (Tyr) i histydyny (His) modelowano przy użyciu bimodalnych rozkładów kątów dwuściennych $P(\chi_1, \chi_2)$. Sekwencja Hyp-1 zawiera 10 hydrofobowych reszt Leu, jednakże tylko 2 z nich wykazały różną formę rotameru w eksperymentalnych i symulowanych modelach struktury. Dla rozgałęzionej przy atomie węgla $C\beta$ reszty aminokwasowej Leu86, χ_1 i χ_2 były oczekiwane odpowiednio w pobliżu 175° i 65° , podczas gdy obliczony rozkład wskazywał na przejście χ_2 do $\sim 180^\circ$, co skutkowało rzadkimi konformacjami łańcucha bocznego (Rys. 59). W oryginalnej strukturze Leu86 została wprowadzona do Hyp-1 jako zamiennik dla Ile86 i odgrywał rolę katalitycznego miejsca aktywnego w strukturze, co zgadza się z faktem występowania słabo obsadzonego rotameru, gdyż oznacza to wyższą energię tego obszaru. Rysunek 58b pokazuje, że kolejny rotamer inny niż w modelu eksperymentalnym zaobserwowano dla Leu151 w rejonie helisy α_3 , gdzie obliczony rozkład χ_1 jest skupiony wokół rotamerycznej wartości 295° (-65°) z nierotamerycznym pikiem χ_2 w pobliżu 300° (-60°). Taka anomalia konformacyjna łańcucha bocznego sugerowała wysokoenergetyczne perturbacje w Leu151, co potwierdziła zwiększona wartość ADP dla węgla $C\gamma$ tej reszty. Symulacje MD wykazały znaczną swobodę konformacyjną łańcucha bocznego Leu151, sugerując jego możliwą rolę w rozpoznawaniu substratów. W obrębie powiązanych reszt leucyny (Leu) i izoleucyny (Ile) nie znaleziono różnic między formami rotamerów zdefiniowanymi eksperymentalnie oraz na drodze symulacji.



Rysunek 59. Dwuwymiarowa mapa rozkładów gęstości prawdopodobieństwa konformacji łańcucha bocznego $P(\chi_1, \chi_2)$ dla Leu86, Leu151 (a,b). Wartości prawdopodobieństwa w każdym z przedziałów $5^\circ \times 5^\circ$ rosną od głębokiego niebieskiego do żółtego zgodnie z mapą kolorów [137].

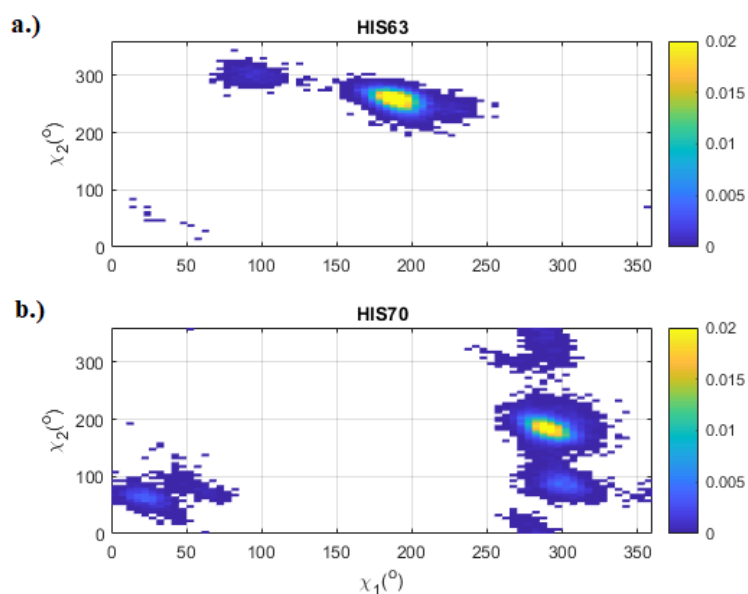
Aminokwas Phe z pierścieniem aromatycznym w strukturze Hyp-1 był reprezentowany przez 8 reszt, w większości ulokowanych w hydrofobowej wewnętrznej wnęce. Porównanie eksperymentalnych rotamerów z ich symulowanymi formami ujawniło różnice w przypadku Phe72 i Phe158 (Rys. 60). Zarówno doświadczalne jak i obliczone χ_1 osiągało pik przy ok. 295° (-65°), natomiast obserwowany rozkład χ_2 posiadał maksimum w pobliżu 270° (-90°). Drugi gęsto obsadzony region χ_2 obejmował obszar w pobliżu 90° , natomiast para χ_2 z modelu eksperymentalnego powinna posiadać piki w okolicach 330° (-30°) i 150° , jednakże wbrew oczekiwaniom obliczony rozkład $P(\chi_2)$ ujawnił małe piki prawdopodobieństwa w tych regionach. Subtelną różnicę zaobserwowano dla symulacji C-końcowej reszty Phe158 z preferencją dla rozkładu kąta χ_2 do wartości $\sim 300^\circ$ (-60°). Wcześniejsza eksploracja bogactwa konformacji Phe w strukturach białkowych ujawniła, że $\chi_1 = 300^\circ$ mogła być silnie ograniczana przez kolizje steryczne pomiędzy pierścieniem aromatycznym a grupą karbonylową sąsiednich reszt. Przy pomocy bezpośredniej symulacji ruchów białka metodami MD stwierdzono, że rotametr m-85 jest energetycznie korzystną konformacją łańcucha bocznego Phe72, co było niejasne podczas udokładniania arbitralnie wybranych konformacji łańcuchów bocznych z różnymi obsadzeniami poszczególnych form.



Rysunek 60. Dwuwymiarowa mapa rozkładów gęstości prawdopodobieństwa konformacji łańcucha bocznego $P(\chi_1, \chi_2)$ dla Phe72, Phe158 (c,d). Wartości prawdopodobieństwa w każdym z przedziałów $5^\circ \times 5^\circ$ rosną od głębokiego niebieskiego do żółtego zgodnie z mapą kolorów [137].

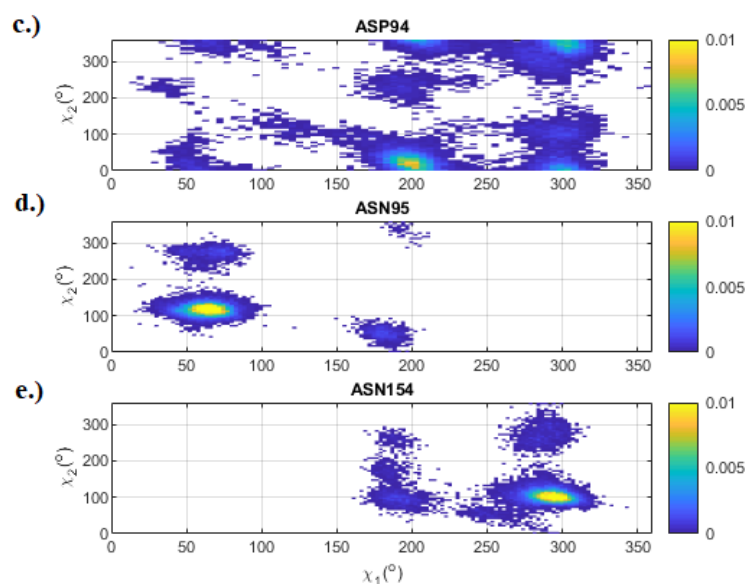
Wśród 6 reszt His z sekwencji białka Hyp-1, symulacja MD ujawniła 2 główne odchylenia od eksperymentalnej stereochemii łańcucha bocznego (Rys. 61). Dla His63, silne piki przy $\chi_1 = 180^\circ$ i $\chi_2 = 260^\circ$ odpowiadały rotamerowi t-80 z modelu struktury. Jednakże zaobserwowano również dużą plamę w rozkładzie $P(\chi_1, \chi_2)$ w pobliżu $\chi_1 = 295^\circ$ i $\chi_2 = 290^\circ$ odpowiadającą dominującemu stanowi m-70 zajmującemu statystycznie 29% populacji rotameru, chociaż konformacja ta była nieobecna w początkowym zestawie danych. Drugą zauważalną różnicę stanowiły konformacje His70 z wewnętrznej wnęki białka, dla której

eksperymentalnie dobrano drugi najbardziej preferowany rotamer m80 ($\chi_1 = 295^\circ$, $\chi_2 = 80^\circ$). Obliczony rozkład wykazywał preferencję rzadziej obserwowanej konformacji m170 (7% prawdopodobieństwa), co pokrywało się z obserwacją pików χ_1 i χ_2 w pobliżu 295° i 165° . W przypadku Tyr nie zauważono różnic pomiędzy eksperymentalną i obliczoną formą rotameryczną. Pomimo strukturalnego podobieństwa na podstawie przeprowadzonych symulacji stwierdzono, że Tyr jest ogólnie bardziej stabilną resztą w strukturze białka Hyp-1 o niskiej wrażliwości na energetyczne zmiany konformacyjne.



Rysunek 61. Dwuwymiarowa mapa rozkładów gęstości prawdopodobieństwa konformacji łańcucha bocznego $P(\chi_1, \chi_2)$ dla His63, His70 (a,b). Wartości prawdopodobieństwa w każdym z przedziałów $5^\circ \times 5^\circ$ rosną od głębokiego niebieskiego do żółtego zgodnie z mapą kolorów [137].

Chociaż natywna sekwencja Hyp-1 zawierała 6 cząsteczek Asp, znaleziono tylko jedną rozbieżność pomiędzy eksperymentalnymi i symulowanymi rotamerami Asp94. Podobnie jak w modelu Hyp-1 pochodzącym z PDB, należałoby się spodziewać dwóch silnie obsadzonych skupisk wokół $\chi_1 = 290^\circ$ (-70°) i $\chi_2 = 345^\circ$ (-15°) odpowiadających najbardziej preferowanej konformacji m-20. Tymczasem na Rys. 62c można dostrzec znaczne rozmycie dozwolonych kątów łańcucha bocznego w obliczonym rozkładzie $P(\chi_1, \chi_2)$ Asp94, gdzie silne piki występują przy $\chi_1 \sim 200^\circ$ i χ_2 w pobliżu 0° . Wysoki poziom swobody konformacyjnej Asp94 spowodował trudności z określeniem preferowanych wartości poszczególnych kątów i jednoznacznej interpretacji konformacji. Co ciekawe, stwierdzono na podstawie symulacji, że przejścia do rzadkich form rotamerów są energetycznie faworyzowane, co stanowi odmienny rezultat niż w eksperymentalnych technikach krystalograficznych.



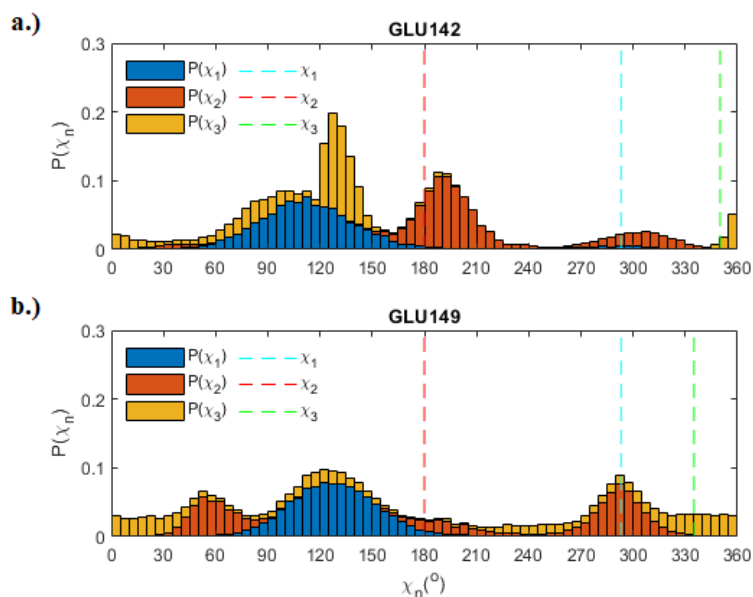
Rysunek 62. Dwuwymiarowa mapa rozkładów gęstości prawdopodobieństwa konformacji łańcucha bocznego $P(\chi_1, \chi_2)$ dla Asp94, Asn95 oraz Asn154 (c-e). Wartości prawdopodobieństwa w każdym z przedziałów $5^\circ \times 5^\circ$ rosną od głębokiego niebieskiego do żółtego zgodnie z mapą kolorów [137].

Podobne wnioski płyną z analizy rozkładów $P(\chi_1, \chi_2)$ dla Asn95 i Asn154 (Rys. 62d,e). Obie reszty występowały w nietypowych konformacjach p-10 i m-80, które powinny mieć wysokie prawdopodobieństwo odpowiednio w pobliżu $\chi_1 = 62^\circ$ i 295° . Tymczasem obliczenia przy pomocy symulacji MD wskazały, że to wartość kąta χ_2 determinuje formę rotameru i odgrywa kluczową rolę w badaniu preferowanych konformacji Asn. Tym samym Asn95 ze swoim pikiem χ_2 w okolicy 90° przyjmuje bardziej popularną konformację p30, a dla Asn154, preferowany jest rzadki rotamer m120 ze względu na pik $P(\chi_2)$ w pobliżu 100° . Znanym problemem z konformerami Asn jest to, że ich rozkład $P(\chi_2)$ jest zazwyczaj szeroki i niejednoznaczny [134]. Wyznaczony rozkład $P(\chi_1, \chi_2)$ wskazywał istnienie dobrze odizolowanych, lokalnych skupisk wartości χ_2 wyraźnie określające preferowany rotamer Asn. W przypadku Asp94 zaobserwowano brak takich wyraźnych klastrów, co wynika z dużej mobilności częściowo nieuporządkowanych łańcuchów bocznych.

7.11. Symulowane rozkłady prawdopodobieństwa kątów łańcucha bocznego $P(\chi_1, \chi_2, \chi_3)$ dla Glu, Gln i Met

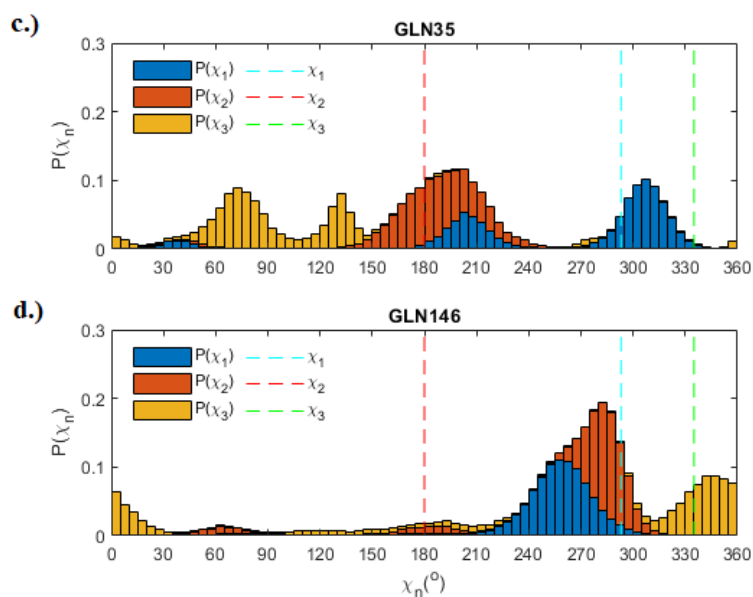
Dla każdej z 19 reszt Glu i 3 reszt Gln w modelu Hyp-1 określono trójmodowe rozkłady gęstości prawdopodobieństwa kątów torsyjnych χ_1, χ_2, χ_3 w pojedynczym łańcuchu Hyp-1 (Rys. 63). Odchylenia odnotowano dla 6 polarnych, powierzchniowych aminokwasów Glu modelowanych w najczęściej występujących rotamerach, mt-10 (prawdopodobieństwo 33%) lub tt0 (24% wszystkich Glu) wg Penultimate Rotamer Library [20]. Poszerzenie rozkładów poszczególnych kątów torsyjnych χ komplikowało wskazanie poprawnego rotameru. Zarówno dla eksperymentalnego modelu struktury Hyp-1 jak

i symulowanego modelu, dozwolone konformacje Glu były mocno uzależnione od preferowanego oddziaływania z sąsiednimi cząsteczkami.



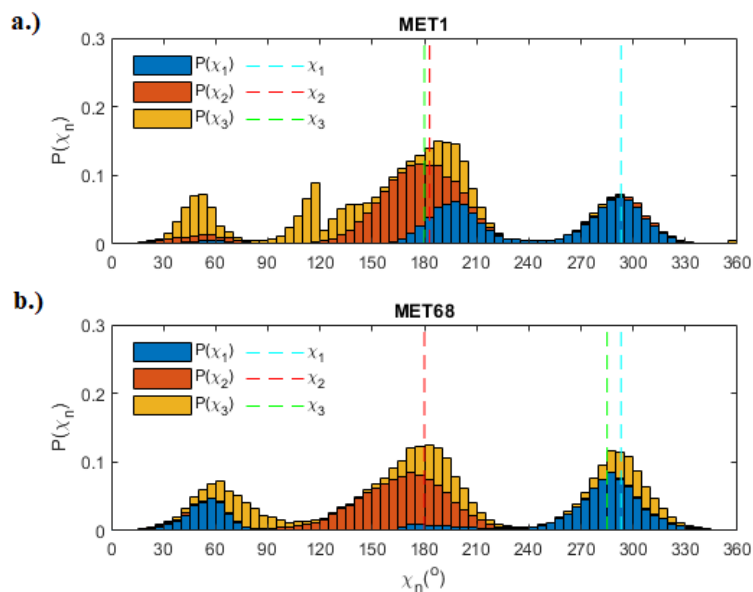
Rysunek 63. Wykres słupkowy rozkładów prawdopodobieństwa dla kątów torsyjnych χ_1 , χ_2 , χ_3 łańcucha bocznego Glu142 (a) i Glu149 (b). Wartości prawdopodobieństwa $^\circ$ dla poszczególnych kątów torsyjnych oznaczono odpowiednio jako niebieskie, pomarańczowe i żółte. Odpowiednie wartości kątów χ_1 , χ_2 , χ_3 z eksperymentalnego modelu Hyp-1 przedstawiają jasnoniebieskie, pomarańczowe i jasnozielone linie przerywane [137].

Porównanie eksperymentalnych i symulowanych modeli Hyp-1 wskazało różnice w preferowanych konformacji łańcuchów bocznych Gln35 i Gln146 (Rys. 64). Na podstawie struktury doświadczalnej Hyp-1 spodziewano się silnych pików w rozkładzie $P(\chi_1, \chi_2, \chi_3)$ dla najczęściej występującego rotameru mt-30 w pobliżu $\chi_1 = 290^\circ$, $\chi_2 = 180^\circ$ i $\chi_3 = 330^\circ$. Tymczasem w modelu po symulacji faworyzowane były mniej popularne rotametry tt0, mm-40 i mm100 stabilizowane przez oddziaływania z cząsteczkami rozpuszczalnika w boksie symulacyjnym. Preferencja do przyjmowania rzadkich rotamerów w długich łańcuchach bocznych Glu i Gln była bezpośrednim efektem ich podatności na dynamiczne zmiany konformacji, powierzchniową ekspozycję oraz możliwy brak silnych oddziaływań wewnątrzcząsteczkowych [135]. W praktyce takie łańcuchy boczne często modeluje się w mapach gęstości elektronowej jako kombinację jednego lub kilku najczęściej występujących rotamerów o częściowym obsadzeniu, podczas gdy symulacje MD wskazują na istnienie pewnej preferowanej energetycznie formy. Analiza rotamerów Glu/Gln wskazała, że w ten sposób podczas udokładniania modeli struktury białek pewne energetycznie preferowane konformacje mogą zostać pominięte.



Rysunek 64. Wykres słupkowy rozkładów prawdopodobieństwa skręcenia łańcucha bocznego χ_1 , χ_2 , χ_3 dla Gln35, Gln146 (c,d). Wartości prawdopodobieństwa dla poszczególnych kątów torsijskich oznaczono odpowiednio jako niebieskie, pomarańczowe i żółte. Odpowiednie wartości kątów χ_1 , χ_2 , χ_3 z eksperymentalnego modelu Hyp-1 zostały przedstawione kolejno jako jasnoniebieskie, pomarańczowe i jasnozielone linie przerywane [137].

Sekwencja aminokwasowa Hyp-1 zawierała tylko 2 reszty Met - N-terminalną Met1 na początku łańcucha polipeptydowego i przez to silnie ruchomą - oraz Met68. Zgodnie z danymi eksperymentalnymi oczekiwano pików rozkładu $P(\chi_1, \chi_2, \chi_3)$ przy $\chi_1 = 293^\circ$, $\chi_2 = 180^\circ$ i $\chi_3 = 180^\circ$ bądź $\chi_3 = 285^\circ$. Tym samym to wartość χ_3 jednoznacznie określała dostępny rotamer, gdyż pozostałe kąty torsijskie były takie same dla obu konformacji, mtt i mtm. W modelu MD oba symulowane rotamery Met odbiegały od doświadczalnych, głównie w kierunku bardziej powszechnej konformacji mtp, na co wskazuje położenie kąta torsijskiego χ_3 w pobliżu 75° (Rys. 65).

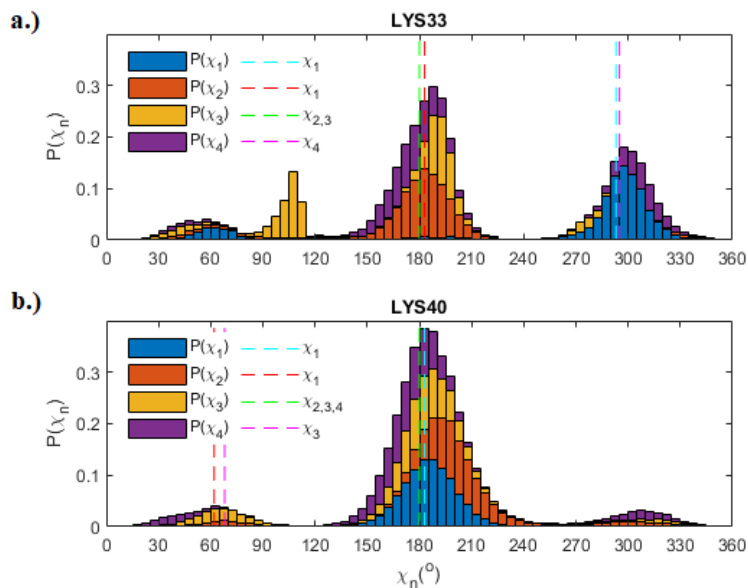


Rysunek 65. Wykres słupkowy rozkładów prawdopodobieństwa skręcenia łańcucha bocznego χ_1 , χ_2 χ_3 dla Met1 (a) oraz Met68 (b). Wartości prawdopodobieństwa dla poszczególnych kątów torsyjnych oznaczono odpowiednio jako niebieskie, pomarańczowe i żółte. Odpowiednie wartości kątów χ_1 , χ_2 χ_3 z eksperymentalnego modelu Hyp-1 zostały przedstawione kolejno jako jasnoniebieskie, pomarańczowe i jasnozielone linie przerywane [137].

Duża swoboda konformacyjna reszt Met mogła być efektem braku oddziaływań kierunkowych stabilizujących łańcuch boczny. W symulowanym modelu wybór rotameru był podyktowany zarówno wiązaniem z cząsteczkami rozpuszczalnika, jak i silnymi ograniczeniami odpychającymi nałożonymi podczas symulacji. Szczególnie istotna wydaje się konformacja Met68, która jest zaangażowana w wiązanie liganda i tworzenie kompleksu Hyp-1.

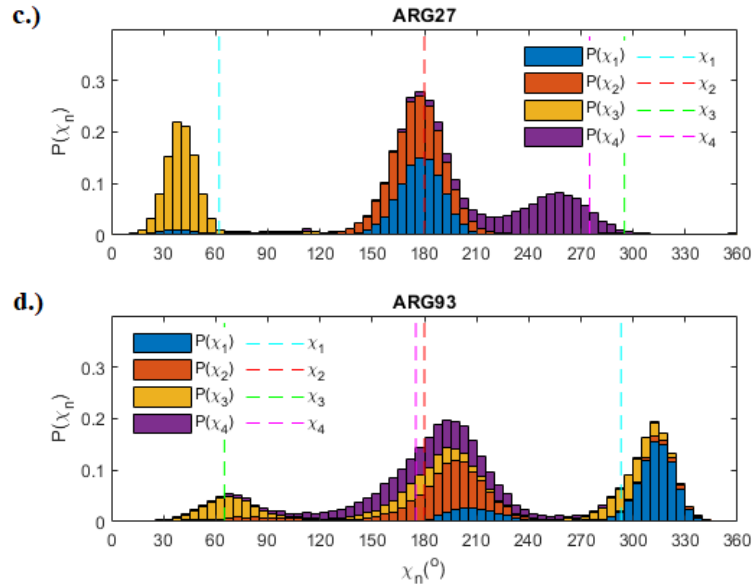
7.12. Symulowane rozkłady prawdopodobieństwa kątów łańcucha bocznego $P(\chi_1, \chi_2, \chi_3, \chi_4)$ dla Lys i Arg

Długie łańcuchy boczne Arg i Lys charakteryzujące się czterema kątami torsyjnymi mogą występować w wielu kombinacjach. Zauważono, że spośród 16 reszt Lys w pojedynczej cząsteczce Hyp-1, wszystkie przyjmowały rzadkie stany rotacyjne, z rotamerem mttt (20% populacji) jako najczęstszym. W strukturze Hyp-1 łańcuchy boczne Lys miały tendencję do tworzenia częściowo uporządkowanych supramolekularnych połączeń z ligandami PEG, co po raz pierwszy zaobserwowano dla Lys33 w cząsteczce A (Rys. 66). Specyficzna interakcja z ligandami PEG ogranicza możliwości przyjmowania rzadkich konformacji przez Lys33, jednakże symulacje MD sugerowały, że zmiany konformacyjne w łańcuchu bocznym Lys33A poprawiały oddziaływania wodorowe z sąsiednimi cząsteczkami i łańcuchami bocznymi innych aminokwasów.



Rysunek 66. Wykres słupkowy rozkładów prawdopodobieństwa skręcenia łańcucha bocznego χ_1 , χ_2 , χ_3 , χ_4 dla Lys33 (a) i Lys40 (b). Każdy rozkład prawdopodobieństwa χ_i został znormalizowany tak, że $\sum \chi_i P(\chi_i) = 1$. Wartości prawdopodobieństwa w obrębie przedziałów $5^\circ \times 5^\circ$ dla oddzielnych kątów skręcenia oznaczono odpowiednio jako niebieskie, pomarańczowe, żółte i fioletowe [137].

Dodatkowo naładowane reszty Lys mogą być łatwo zastąpione przez Arg. Obliczony rozkład prawdopodobieństwa kątów torsyjnych uwzględnił wartości czterech kątów i ujawnił istotne zmiany konformacyjne w Arg27 i Arg93. Rozkład $P(\chi_1, \chi_2, \chi_3, \chi_4)$ kątów torsyjnych łańcucha bocznego Arg27 wyraźnie wskazywał najbardziej obsadzone skupiska związane z każdym kątem torsyjnym (Rys. 67). Chociaż Arg27 w eksperymentalnej strukturze Hyp-1 była modelowana w konformacji podwójnej ze względu na brak oddziaływań stabilizujących, podczas symulacji okazało się, że najbardziej preferowanym rotamerem jest ttp-105. Zatem wiązania wodorowe z sąsiadującymi Tyr85 oraz Leu24 mogą odpowiadać za stabilizację symulowanego rotameru i zakotwiczenie go wewnątrz białka.



Rysunek 67. Wykres słupkowy rozkładów prawdopodobieństwa skrećenia łańcucha bocznego χ_1 , χ_2 , χ_3 , χ_4 dla Arg27 (c) oraz Arg93 (d). Każdy rozkład prawdopodobieństwa χ_i został znormalizowany tak, że $\sum_{\chi_i} P(\chi_i) = 1$. Wartości prawdopodobieństwa w obrębie przedziałów $5^\circ \times 5^\circ$ dla oddzielnych kątów skrećenia oznaczono odpowiednio jako niebieskie, pomarańczowe, żółte i fioletowe [137].

W obrębie rozkładu prawdopodobieństwa skrećenia łańcucha bocznego Arg93 znaleziono rozległe obszary występowania kątów torsyjnych χ_3 i χ_4 (Rys. 67). Zatem różnorodność konformacyjna Arg93 wyraża się jako mieszanina rotamerów o różnych gęstościach. Symulacja MD może zidentyfikować pewne korzystne regiony w krajobrazie energetycznym struktury, a tym samym pomóc w identyfikacji najbardziej prawdopodobnej konformacji łańcucha bocznego i uzasadnić ją na gruncie oddziaływań z innymi cząsteczkami. Tym samym jest to metoda pomocna w identyfikacji cząsteczek lub ich konformacji w słabo zdefiniowanych obszarach gęstości elektronowej względem eksperymentalnego opisu struktury rozwiązanej metodami krystalografii rentgenowskiej.

7.13. Wyniki udokładnienia w nadprzestrzeni (3+1)D oraz korekt fononowych

Opisany w rozprawie proces przekształcenia superkomórki 7Hyp/ANS do nadprzestrzeni ukazuje zaletę tej metody w zakresie analizy modulacji struktury oraz zaburzeń jej periodyczności. Wprowadzenie zredukowanej komórki ułatwia udokładnianie struktury, które w podejściu superkomórki było przeprowadzane poprzez długie, ręczne sprawdzenie łańcuchów białek, poprawianie zdeformowanych fragmentów głównego łańcucha polipeptydowego oraz łańcuchów bocznych, inspekcję fragmentów map gęstości elektronowej o niepewnym zapełnieniu oraz dopasowywaniu do nich cząsteczek i fragmentów łańcucha. Przedstawiona reprezentacja modelu struktury 9Hyp/ANS w nadprzestrzeni (3+1)D wskazała, że funkcje AMF są okresowe wzdłuż kierunku \mathbf{a}_4 . Sugeruje to, że ruchy zachodzące w rozszerzonej superkomórce między sąsiadującymi komórkami nie są przypadkowe i determinowane wyłącznie obecnością „sklejającego”

molekuły białka liganda ANS, ale tworzą bardziej skomplikowany układ związany z funkcją AMF. Podczas nakładania kolejnych komórek zaobserwowano odchylenia średnich pozycji atomowych sąsiadujących komórek powiązanych operacją tNCS, co było zgodne z pokazanymi na Rys. 31 odchyleniami wektorów translacji, nie tylko wzdłuż spodziewanego kierunku *c*, ale także osi *a* i *b*. Analizując te różnice można wskazać, że modulacja struktury 9Hyp/ANS nie ma charakteru płynnego, ponieważ przy jednym przejściu z komórki do komórki występują większe odchylenia pozycji atomowych, a przy innych mniejsze. Jest to zgodne z obserwacją obrazów dyfrakcyjnych, gdzie niektóre satelity wyższych rzędów miały wyższą intensywność i były lepiej widoczne na obrazach niż inne, słabiej widoczne. Niemniej, redukcja do nadprzestrzeni (3+1)D potwierdziła, że modulacji przebiega względem średniej struktury, a zatem wektor modulacji ma tylko jedną znaczącą składową względem *a*^{*}. Podobnej analizie można poddać strukturę 9Hyp/ANS, wprowadzając podział superkomórki na 9, zamiast 7 mniejszych komórek. Zastosowanie zautomatyzowanych skryptów MATLAB ułatwia taką analizę w przyszłości przy niewielkiej modyfikacji parametrów związanych z doбором wektora modulacji oraz użycia nowego pliku źródłowego *.cif*.

W wyniku udokładnienia modelu struktury 7Hyp/ANS w nadprzestrzeni (3+1)D otrzymano model o wskaźniku rozbieżności 0,2165, tj. nieco niższy niż finalnie zdeponowany w bazie PDB model w ujęciu superkomórki ($R = 0,226$). Niewielka poprawa wyników udokładniania wskazuje, że jest to obiecująca droga do dalszej analizy tego układu. Kolejne iteracje udokładnienia, zmiany udokładnianych parametrów lub wprowadzanie dalszych poprawek strukturalnych, związanych z funkcją modulacji bądź drganiami termicznymi mogą przynieść dalszą redukcję wskaźnika rozbieżności *R*. Przeprowadzona analiza poprawności stereochemicznej modelu wskazuje, że wciąż ma on pewne rozbieżności w obrębie kątów torsyjnych oraz długości wiązań, co powinno być wyeliminowane przy kolejnych fazach udokładnienia. Możliwa jest również rozbudowa i optymalizacja pakietu skryptów celem wydajniejszej optymalizacji długości wiązań i kątów zwykłych modelu względem danych eksperymentalnych.

Wprowadzone poprawki fononowe stanowią innowacyjne podejście do skorelowanych drgań sieci krystalicznej w obrębie modeli struktury białek. Udokładnienie z włączonym modelowaniem fononów jako członów gaussowskich (Rys. 49) wywołało niewielki wzrost wskaźnika rozbieżności *R*, jednak można spodziewać się, że dalsze udokładnienie z włączoną poprawką fononową powinno przynieść poprawę zbieżności danych struktury. Innym rozwiązaniem może być zastosowanie w przyszłości bardziej złożonej poprawki opartej na funkcjach Bessela, która jest zaimplementowana w przygotowanym pakiecie skryptów MATLAB.

8. Podsumowanie

W połowie lat 90-tych krystalografia makrocząsteczek otrzymała ogromny impuls do rozwoju. Pomimo niewątpliwych postępów metod krystalizacji białek, technik

eksperymentalnych czy rozwiązywania i udokładniania modeli struktur, krystalografia makromolekuł nadal boryka się z wieloma patologiami kryształów. Jedną z takich anomalii kryształów białek jest modulacja struktury, dość często spotykana w krystalografii małych cząsteczek, ale praktycznie niespotykana wśród makrocząsteczek. W przypadku modulacji struktury, analiza i przetwarzanie danych są szczególnie trudne ze względu na naruszenie podstawowego założenia o izotropowym rozkładzie Wilsona mierzonych intensywności refleksów. Prawdopodobnie przypadki modulacji struktury białek zdarzają się częściej, niż mówią o tym doniesienia naukowe, ponieważ w momencie ich napotkania dane takie są odrzucane jako pochodzące z rzekomo zdeformowanego kryształu lub niemożliwe do prawidłowego wskaźnikowania. W związku z tym, nowo odkryte modulowane struktury białek wymagają głębszego zbadania i stworzenia specjalistycznego oprogramowania do rutynowej analizy takiego zjawiska.

Wśród najważniejszych osiągnięć badawczych zawartych w przedstawionej rozprawie doktorskiej chciałabym wymienić:

1. Udokładnienie i dogłębną analizę strukturalną modelu białka 9Hyp/ANS z dziewięciokrotną modulacją struktury wzdłuż osi c oraz 36 niezależnymi łańcuchami białka w rozszerzonej komórce elementarnej. Podczas badania potwierdzono modulację struktury obserwując na obrazach dyfrakcyjnych refleksy główne i satelity, a także śledząc cykliczne zmiany wartości uśrednionej intensywności refleksów w warstwach o stałym wskaźniku l . Model struktury 9Hyp/ANS rozwiązano przy pomocy dostępnego oprogramowania w ogromnej superkomórce oraz zdeponowano w bazie PDB z identyfikatorem 6SJJ.
2. Opracowanie kompleksowego pakietu skryptów w środowisku Matlab służącego rozwiązywaniu oraz udokładnianiu modelu białka Hyp-1/ANS w oparciu o metodę analizy wielowymiarowej 3D+1. Pakiet umożliwia wprowadzenie danych w postaci powszechnie stosowanego w krystalografii formatu CIF zawierającego położenia atomów struktury krystalicznej. Zawiera również specjalistyczne skrypty do analizy więzów stereochemicznych modelu oraz własne oprogramowanie graficzne, co umożliwia generowanie struktury bez konieczności korzystania z innych programów.
3. Wprowadzenie analizy wielowymiarowej w przestrzeni 3D+1 do opisu struktury krystalicznej 7Hyp/ANS oraz udokładnienia danych. Umożliwiło to opis struktury z zastosowaniem niewspółmiernego wektora modulacji oraz całkowite odrzucenie założeń periodyczności struktury. W ujęciu wielowymiarowym nie jest wymagane rozszerzenie do superkomórki, a całą strukturę rozwiązano w jednej komórce wprowadzając niewspółmierną modulację z wektorem $3/7c$. Podczas udokładniania otrzymano wskaźnik rozbieżności R na poziomie 21,65%, czyli lepszy niż dla modelu udokładnianego w ujęciu superkomórki.
4. Przetestowanie pakietu na zestawie danych dyfrakcyjnych modulowanego kompleksu białka 7Hyp/ANS otrzymanych we współpracy z Instytutem Chemii

Bioorganicznej PAN w Poznaniu. Testy zostały przeprowadzone zarówno na komputerach Wydziału Fizyki i Informatyki Stosowanej AGH, jak również we współpracy z ACK Cyfronet AGH (superkomputery Prometeusz i Ares).

5. Przeprowadzenie symulacji metodami dynamiki molekularnej modelu białka Hyp-1 bez dodatku liganda ANS oraz porównanie wyników otrzymanych na drodze symulacji (ruchów termicznych, zmian konformacji łańcucha głównego i bocznych) z danymi wynikami uzyskanymi metodami krystalografii rentgenowskiej.

Przedłożona rozprawa stanowi istotny wkład w rozwój metod analizy struktur białek, szczególnie modulowanych lub z różnego rodzaju zaburzeniami ścisłej periodyczności (np. zbliżniaczenie kryształów). Całkowite odrzucenie założenia o periodyczności i opisanie struktury makrocząsteczkowej w przestrzeni wielowymiarowej stanowi innowacyjny i pionierski krok dla lepszego zrozumienia modulowanych układów białkowych. Wyniki otrzymane w ramach rozprawy są całkowicie unikatowe. Podobną próbę opisu w przestrzeni wielowymiarowej tak złożonej struktury przedstawiono już co prawda raz w literaturze [43], to jednak nikt dotąd nie przedstawił udokładnienia tak skonstruowanego modelu. Wybrany układ, tj. kompleks białka Hyp-1 z ANS, ma też znaczenie aplikacyjne w medycynie, ze względu na biologiczną zdolność białka Hyp-1 do wiązania ligandów, np. cząsteczek fitohormonów. Znajomość szczegółowej struktury atomowej pozwala lepiej określić właściwości fizyko-chemiczne oraz biochemiczne białka, które rzutują na jego potencjalne zastosowanie.

9. Bibliografia

- [1] Bragg, W. L. (1913). Proc. R. Soc. Lond. A89, 248–277.
- [2] Ewald, P.P. (1913). Physik. Z. 14, 465-472.
- [3] Debye, P. (1913).- 348 (1): 49–92.
- [4] Debye P, Scherrer, P. (1916). Phys Zeit. 17:277.
- [5] Hull, A. W. (1917). Phys Rev, 10:661.
- [6] Niggli, P. (1919). *Geometrische Kristallographie des Diskontinuums*. Leipzig: Gebrüder Bornträger.
- [7] Patterson, A. L. (1935). Z. Kristallogr. 90, 517–542.
- [8] Palatinus, L. & Chapuis, G. (2007). J. Appl. Cryst. 40, 786–790.
- [9] Clifton, N.J. (2017). *Protein Crystallography. Methods and Protocols. Methods in molecular biology*, edited by A. Wlodawer, Z. Dauter & M. Jaskolski, New York: Humana Press
- [10] Rupp, B. (2009). *Biomolecular crystallography: principles, practice and application to structural biology*, Garland Science, New York
- [11] Jaskolski, M., Dauter, Z. Postępy Biochemii 62(3) (2016), 401-407.
- [12] Liebschner, D., Afonine, P. V., Baker, M. L., Bunko' czi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J. & Adams, P. D. (2019). Acta Cryst. D75, 861–877.
- [13] Sheldrick, G. M. (2015). Acta Cryst. C71, 3–8.
- [14] Langer G, Cohen SX, Lamzin VS, Perrakis A. (2008). Nat. Protoc. 3, 1171-1179.
- [15] Cowtan, K. (2006) Acta Cryst. D62, 1002-1011.
- [16] Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). Acta Cryst. D66, 486–501.
- [17] Ramachandran, G.N., Ramakrishnan, C., Sasisekharan, V. (1963). J. Mol Biol 7, 95 99.
- [18] Hooft, R.W.W., Vriend, G., Sander, C. & Abola, E.E. (1996). Nature 381, 272.
- [19] Lovell, S.C., Word, J.M., Richardson, J.S. & Richardson, D.C. (2000). Proteins. 40, 389-408.
- [20] Ghent, F.A. (1861). Calaverite. American Journal of Science, 45, 314-317.
- [21] Donnay, J. D. H. (1935). Ann. Soc. Geol. Belg. B55, 222–230.
- [22] Preston, G. D. (1938). Proc. R. Soc. London, 167, 526–538.
- [23] Chapuis, G. (2003). Crystallographic excursion in superspace. Crystal Engineering, 6, 187-195.
- [24] Smaalen, S. van (2004). Z. Kristallogr. 219, 681–691.
- [25] Smaalen, S. van (2007). Incommensurate Crystallography. Oxford University Press.
- [26] Janner, A. & Janssen, T. (1977). Phys. Rev. B, 15, 643–658.

- [27] Janner, A. & Janssen, T. (1980a). *Acta Cryst.* A36, 399–408.
- [28] Janner, A. & Janssen, T. (1980b). *Acta Cryst.* A36, 408–415.
- [29] Wolff, P. M. de (1974). *Acta Cryst.* A30, 777–785.
- [30] Petříček, V., Dušek, M. & Palatinus, L. (2014). *Z. Kristallogr.* **229**, 345–352.
- [31] Petříček, V., Fuksa, J. & Dušek, M. (2010). *Acta Cryst.* A**66**, 649–655.
- [32] Zheng-qing, F. & Hai-fu, F. (1994). *J. Appl. Cryst.* **27**, 124–127.
- [33] Petříček, V., Coppens, P. & Becker, P. (1985). *Acta Cryst.* A**41**, 478–483.
- [34] Lovelace, J. J., Narayan, K., Chik, J. K., Bellamy, H. D., Snell, E. H., Lindberg, U., Schutt, C. E. & Borgstahl, G. E. O. (2004). *J. Appl. Cryst.* **37**, 327–330.
- [35] Porta, J., Lovelace, J. J., Schreurs, A. M. M., Kroon-Batenburg, L. M. J. & Borgstahl, G. E. O. (2011). *Acta Cryst.* D**67**, 628–638.
- [36] Porta, J., Lovelace, J. & Borgstahl, G. E. O. (2017). *J. Appl. Cryst.* **50**, 1200–1207.
- [37] Lovelace, J. J. & Borgstahl, G. E. O. (2020). *Crystallogr. Rev.* **26**, 3–50.
- [38] Lovelace, J., Petříček, V., Murshudov, G. & Borgstahl, G. E. O. (2019). *Acta Cryst.* D**75**, 852–860.
- [39] Lovelace, J. J., Simone, P. D., Petříček, V. & Borgstahl, G. E. O. (2013). *Acta Cryst.* D**69**, 1062–1072.
- [40] Sliwiak, J., Dauter, Z. & Jaskolski, M. (2016). *Front. Plant Sci.* **7**, 668.
- [41] Sliwiak, J., Dauter, Z., Kowiel, M., McCoy, A. J., Read, R. J. & Jaskolski, M. (2015). *Acta Cryst.* D**71**, 829–843.
- [42] Sliwiak, J., Jaskolski, M., Dauter, Z., McCoy, A. J. & Read, R. J. (2014). *Acta Cryst.* D**70**, 471–480.
- [43] Lovelace, J.J. & Borgstahl, G.E.O. (2021). *Acta Cryst.* D**77**, 790–798.
- [44] Feig, M. & Mirjalili V. (2016). *Proteins.* **84**, 282–292.
- [45] Pang, Y.-P. (2015). *Biochem. Biophys. Res. Commun.* **458**, 352–355.
- [46] Pang, Y.-P. (2016). *Heliyon* **2**(9), <https://doi.org/10.1016/j.heliyon.2016.e00161>.
- [47] Parthasarathy S. & Murthy M.R.N. (1997). *Protein Sci.* **6**, 2561–2567.
- [48] Masmaliyeva, R. C. & Murshudov, G. N. (2019). *Acta Cryst* D**75**, 505–518.
- [49] Caldararu, O., Kumar, R., Oksanen, E., Logan, D.T. & Ryde, U. (2019). *Phys. Chem. Chem.Phys.* **21**, 18149.
- [50] Parsons, S. (2003). *Acta Cryst.* D**59**, 1995–2003.
- [51] Duisenberg, A. J. M., Kroon-Batenburg, L. M. J. & Schreurs, A. M. M. (2003). *J. Appl. Cryst.* **36**, 220–229.
- [52] Wilson, A. J. C. (1949). *Acta Crystallographica* **2** 318–321.
- [53] French, S. & Wilson, K. (1978). *Acta Cryst.* A**34**, 517–525.
- [54] Zwart, P. H., Grosse-Kunstleve, R. W. & Adams, P. D. (2005). *CCP4 Newsl. Protein Crystallogr.* **43**, contribution 7.
- [55] McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- [56] Shechtman D., Blech I., Gratias D., Cahn J. W., *Phys. Rev. Lett.* **53** (1984) 1951.

- [57] Strzalka, R., Buganski, I., Wolny, J. (2016). *Crystals*. 6(9):104. <https://doi.org/10.3390/cryst6090104>
- [58] Buganski, I., Strzalka, R. & Wolny, J. (2019). *Acta Cryst.* A75, 352-361.
- [59] Wolny, J., Buganski, I., Kuczera, P. & Strzalka, R. (2016). *J. Appl. Crystallogr.* 49, 2106-2115.
- [60] Strzalka, R., Buganski, I., Smietanska, J. & Wolny, J. (2020). *Arch. Metall. Mater.* 65(1), 291-294.
- [61] Steurer, W. (2018). *Acta Cryst.* A74, 1-11.
- [62] Penrose, R. (1974). *Bulletin of the Institute of Mathematics and Its Applications*, 10, 266-271.
- [63] Takakura, H., Yamamoto, A., Sato, T. J., Tsai, A. P., Ozawa, Y., Yasuda, N. & Toriumi, K. (2006). *Philosophical Magazine*, 86:3-5, 621-627.
- [64] Chodyń, M.- Rozprawa doktorska (2018), WFiIS AGH Kraków.
- [65] Lee, M.R., Tsai, J., Baker, D. & Kollman, P.A. (2001). *J. Mol. Biol.* 313, 417-430.
- [66] Karplus M. & McCammon J.A. (2002). *Nat. Struct. Biol.* 9, 646-652.
- [67] Dror, R.O., Dirks, R.M., Grossman, J.P., Xu H. & Shaw, D.E. (2012). *Annu. Rev. Biophys.* 41, 429-452.
- [68] McCammon J.C., Gelin B.R. & Karplus M. (1977). *Nature*. 267, 585-590.
- [69] Patodia, S., Bagaria, A. & Chopra, D. (2014). *J. Chem. Phys. Biophys.* 4, <https://doi.org/10.4172/2161-0398.1000166>.
- [70] Karplus M. & Kuriyan J. (2005). *PNAS* 102, 6679-6685.
- [71] Allen F., Almasi G., Andreoni W., Beece D., Berne B.J., Bright A. et al. (2001). *IBM Syst. J.*, 40, 310-327.
- [72] Shaw D.E., Maragakis P., Lindorff-Larsen K., Piana S., Dror R.O., Eastwood M.P., et al. (2010). *Science*. 330, 341-346.
- [73] Freddolino P.L., Arkhipov A.S., Larson S.B., McPherson A. & Schulten K. (2006). *Structure*. 14, 437-449.
- [74] Brooks, B.R., Brucoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. & Karplus, M. (1983). *J. Comput. Chem.* 4, 187-217.
- [75] Phillips J.C., Braun R., Wang W., Gumbart J., Tajkhorshid E., Villa E., Chipot C., Skeel R.D., Kale L. & Schulten K. (2005). *J. Comput. Chem.* 26, 1781-1802.
- [76] Abraham, M. J., Murtola, T., Schulz, R., Pall, S., Smith, J. C., Hess, B. & Lindahl, E. (2015). *SoftwareX* 1, 19-25.
- [77] 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. (1995). *J. Am. Chem. Soc.* 117, 5179-5197.
- [78] Jorgensen, W. L. & Tirado-Rives, J. (1988). *J. Am. Chem. Soc.* 110, 1657-1666.
- [79] Jorgensen, W. L., Maxwell, D. S. & Tirado-Rives, J. (1996). *J. Am. Chem. Soc.* 118, 11225-11236.
- [80] Jorgensen, W. L. & McDonald, N. A. (1998). *J. Mol. Struct.:THEOCHEM* 424, 145-155.

- [81] Jorgensen, W. L. & McDonald, N. A. (1998). *J. Chem. Phys.* B102, 8049-8059.
- [82] Kaminski, G., Friesner, R. A., Tirado-Rives, J. & Jorgensen, W. L. (2001). *J. Chem. Phys.* B105, 6474-6487.
- [83] Rizzo, R. C. & Jorgensen, W. L. (1999). *J. Am. Chem. Soc.* 121, 4827-4836.
- [84] Piana, S. (2011). *Biophys. J.* 101, 1015.
- [85] Izmailov, S.A., Podkorytov, I.S. & Skrynnikov, N.R. (2015). *Sci. Rep.* 7, 9293, doi: <https://doi.org/10.1038/s41598-017-09229-7>.
- [86] Stumpff-Kane, A.W., Maksimiak, K., Lee, M.S. & M. Feig. (2007). *Proteins Struct. Funct. Bioinforma.*, 70, 1345-1356
- [87] Wereszczynski, J. & McCammon, J. A. (2012). *Q. Rev. Biophys.* 45, 1–25.
- [88] Bernardi, R. C., Melo, M. C. R. & Schulten, K. (2015). *Biochim Biophys Acta.* 1850(5), 872-877.
- [89] Wong-ekkabut, J. & Karttunen, M. (2016). *Biochimica et Biophysica Acta.* 1858, 2529–2538.
- [90] Michalska, K., Fernandes, H., Sikorski, M. M. & Jaskolski, M. (2010). *J. Struct. Biol.* 169, 161–171.
- [91] Fernandes, H., Michalska, K., Sikorski, M. & Jaskolski, M. (2013). *FEBS J.* 280, 1169–1199.
- [92] Fernandes, H., Pasternak, O., Bujacz, G., Bujacz, A., Sikorski, M. M. & Jaskolski, M. (2008). *J. Mol. Biol.* 378, 1040–1051.
- [93] Fernandes, H., Bujacz, G., Bujacz, A., Jelen, F., Jasinski, M., Kachlicki, P., Otlewski, J., Sikorski, M.M. & Jaskolski, M. (2009). *FEBS J.* 276, 1596–1609.
- [94] Ruszkowski, M., Sliwiak, J., Ciesielska, A., Barciszewski, J., Sikorski, M. & Jaskolski, M. (2014). *Acta Cryst. D70*, 2032–2041.
- [95] Ruszkowski, M., Szpotkowski, K., Sikorski, M. & Jaskolski, M. (2013). *Acta Cryst. D69*, 2365–2380.
- [96] Gajhede, M., Osmark, P., Poulsen, F. M., Ipsen, H., Larsen, J. N., Joost van Neerven, R. J., Schou, C., Løwenstein, H. & Spangfort, M. D. (1996). *Nat. Struct. Biol.* 3, 1040–1045.
- [97] Biesiadka, J., Bujacz, G., Sikorski, M. M. & Jaskolski, M. (2002). *J. Mol. Biol.* 319, 1223–1234.
- [98] Pasternak, O., Bujacz, G. D., Fujimoto, Y., Hashimoto, Y., Jelen, F., Otlewski, J., Sikorski, M. M. & Jaskolski, M. (2006). *Plant Cell*, 18, 2622–2634.
- [99] Chwastyk, M., Jaskolski, M. & Cieplak, M. (2014). *FEBS J.* 281, 416–429.
- [100] Chwastyk, M., Jaskolski, M. & Cieplak, M. (2016). *Proteins*, 84, 1275–1286.
- [101] Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* 276, 307–326.
- [102] Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M. (2006). *Acta Cryst. D62*, 859–866.
- [103] McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* 40, 658–674.

- [104] Murshudov, G.N., Skubak, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F. & Vagin, A.A. (2011). *Acta Cryst. D*67, 355-367.
- [105] Emsley P. & Cowtan K. (2004). *Acta Cryst. D*60, 2126-2132.
- [106] Terwilliger, T. C. (2000). *Acta Cryst. D*56, 965-972.
- [107] Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. *J Comput Chem.* 2004 Oct;25(13):1605-12.
- [108] Egea, J.A., Henriques, D., Cokelaer, T. (2014). *BMC Bioinformatics* **15**, 136.
- [109] Corey, R. B., Pauling, L. (1953). *Review of Scientific Instruments* 24, 8, 621-627.
- [110] Smietanska, J., Sliwiak, J., Gilski, M., Dauter, Z., Strzalka, R., Wolny, J. & Jaskolski, M. (2020). *Acta Cryst. D*76, 653-667.
- [111] Padilla, J. E. & Yeates, T. O. (2003). *Acta Cryst. D*59, 1124-1130.
- [112] Engh, R. A. & Huber, R. (1991). *Acta Cryst. A*47, 392-400.
- [113] Groom, C. R., Bruno, I. J., Lightfoot, M. P. & Ward, S. C. (2016). *Acta Cryst. B*72, 171-179.
- [114] Weber, L. D. & Tulinsky, A. (1980). *Acta Cryst. B*36, 611-614.
- [115] Macrae, C. F., Sovago, I., Cottrell, S. J., Galek, P. T. A., McCabe, P., Pidcock, E., Platings, M., Shields, G. P., Stevens, J. S., Towler, M. & Wood, P. A. (2020). *J. Appl. Cryst.* 53, 226-235.
- [116] Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst. D*66, 12-21.
- [117] Liebschner, D., Afonine, P. V., Moriarty, N. W., Poon, B. K., Sobolev, O. V., Terwilliger, T. C. & Adams, P. D. (2017). *Acta Cryst. D*73, 148-157.
- [118] Sliwiak, J., Sikorski, M. & Jaskolski, M. (2018). *FEBS J.* 285, 1907-1922.
- [119] Sliwiak, J., Dauter, Z., Jaskolski, M. (2013). *Biotechnologia* 94(1), 47-50.
- [120] Horn, H. W., Swope, W. C., Pitera, J. W., Madura, J. D., Dick, T. J., Hura, G. L. & Head-Gordon, T. (2004). *J. Chem. Phys.* 120, 9665-9678.
- [121] Hohenberg, P. & Kohn, W. (1964). *Phys. Rev.* B136, B864-B871.
- [122] Kohn, W. & Sham, L. J. (1965). *Phys. Rev.* A140, A1133-A1138.
- [123] Becke, A. D. (1988). *Phys. Rev.* A38, A3098-3100.
- [124] Lee, C., Yang, W., Parr, R. G. (1988). *Phys. Rev.* B37, B785-789.
- [125] Valiev, M., Bylaska, E. J., Govind, N., Kowalski, K., Straatsma, T. P., van Dam, H. J. J., Wang, D., Nieplocha, J., Apra, E., Windus, T. L. & de Jong, W. A. (2010). *Comp. Phys. Commun.* 181, 1477-1489.
- [126] Darden, T., York, D. & Pedersen, L. (1993). *J. Chem. Phys.* 98, 10089-10092.
- [127] Essmann, U., Perera, L. & Berkowitz, M. L. (1995). *J. Chem. Phys.* 103, 8577-8593.
- [128] Pall, S. & Hess, B. (2013). *Comp. Phys. Commun.* 184, 2641-2650.
- [129] Clifton, N. J., Whitfield, J. H., Sanchez-Romero, I., Herde, M. K., Henneberger, C., Janovjak, H. & Jackson, C. J. (2017). *Methods Mol. Biol.* 1596, 71-87.
- [130] Roversi, P., Blanc, E., Johnson, S. & Lea, S. M. (2012). *Acta Cryst. D*68, 418-424.
- [131] Arnao, M. & Hernandez-Ruiz, J. (2015). *J. Pineal Res.* 59, 133-150.

- [132] Arnao, M. & Hernaández-Ruiz, J. (2018). *Ann. Bot.* 121, 195–207.
- [133] Kuzmanic, A., Pannu, N. & Zagrovic, B. (2014). *Nat. Commun.* 5, <https://doi.org/10.1038/ncomms4220>.
- [134] Mantsyzov A. B., Savelyev O. Y., Ivantcova P. M., Bräse S., Kudryavtsev K. V. & Polshakov V. I. (2018). *Front. Chem.* 6, <https://doi.org/10.3389/fchem.2018.00091>.
- [135] Shapovalov, M. V. & Dunbrack, R. L. Jr. (2011). *Structure.* 19, 844–858.
- [136] Towse C-L., Rysavy S. J., Vulovic I. M. & Daggett V. (2016). *Structure.* 24, 187-199.
- [137] Smietanska, J., Kozik, T., Strzalka, R., Buganski, I., Wolny, J. (2021). *Crystals.* 11(1):43.

10. Dorobek naukowy

Publikacje:

Structural disorder in quasicrystals / R. STRZAŁKA, I. BUGAŃSKI, J. ŚMIETAŃSKA, J. WOLNY // *Archives of Metallurgy and Materials* / Polish Academy of Sciences. Committee of Metallurgy. Institute of Metallurgy and Materials Science – 2020 vol. 65(1), s. 291–294.

From a single slit to periodic, modulated, and quasiperiodic crystals – a physical space diffraction analysis of aperiodic systems / Radosław STRZAŁKA, Ireneusz BUGAŃSKI, Joanna ŚMIETAŃSKA, Janusz WOLNY // *Journal of Physics. Conference Series* ; ISSN 1742-6588. – 2020 vol. 1458 art. no. 012004, s. 1–6.

A new modulated crystal structure of the ANS complex of the St John's wort Hyp-1 protein with 36 protein molecules in the asymmetric unit of the supercell / Joanna ŚMIETAŃSKA, Joanna Śliwiak, Mirosław Gilski, Zbigniew Dauter, Radosław STRZAŁKA, Janusz WOLNY, Mariusz Jaskólski // *Acta Crystallographica. Section D* – 2020 vol. 76 pt. 7, s. 653–667.

*Molecular dynamics investigation of phenolic oxidative coupling protein Hyp-1 derived from *Hypericum perforatum** / Joanna ŚMIETAŃSKA, Tomasz KOZIK, Radosław STRZAŁKA, Ireneusz BUGAŃSKI, Janusz WOLNY // *Crystals* – 2021 vol. 11 iss. 1 art. no. 43, s. 1–22.

Konferencje:

60th Polish Crystallographic Meeting, 28-29 czerwca 2018, Wrocław, Polska, **nagroda za najlepszy poster** *Protein Crystal with ninefold structure modulation and 36 protein molecules in the asymmetric unit*

31-st European Crystallographic Meeting, ECM31, Oviedo, Hiszpania 22-27 sierpnia 2018, prezentacja *Protein crystal with ninefold structure modulation and 36 protein molecules in the asymmetric unit*

31-st European Crystallographic Meeting, ECM31 Oviedo, Spain 22-27 sierpnia 2018, poster *Statistical method of structural description of complex systems*

14th International Conference on Quasicrystals (ICQ14), 26-31 maja 2019, Kranjska Gora, Słowenia, poster *Structural aspects of modulated Hyp-1 - PR-10 protein from St. John's wort*

61 Konwersatorium Krystalograficzne, 27-28 czerwca 2019, Wrocław, Polska, poster *Structural aspects of modulated Hyp-1 - PR-10 protein from St. John's wort*

32nd European Crystallographic Meeting (ECM32), 18-23 sierpnia, Wiedeń, Austria, poster *Novel approach to structure determination of complex protein system Hyp-1/ANS*

Joint Polish-German Crystallographic meeting, 24–27 lutego, 2020, Wrocław, Poland, poster *Molecular dynamics investigation of the phenolic oxidative coupling protein Hyp-1*

Joint Polish-German Crystallographic meeting, 24–27 lutego, 2020, Wrocław, Poland, poster *Multidimensional analysis vs. statistical method of diffraction and structure description of aperiodic systems*

ECMetAC Days 2020, 7-10 grudnia, konferencja online, prezentacja *Molecular dynamics simulation for validation of the phononic ADPs in Hyp-1 protein*

25th IUCR Congress, 14-22 sierpnia 2021, Praga, Czechy, poster *A novel, statistical approach for structure determination of modulated pathogenesis-related protein (Hyp-1) complex with ANS*

ECMetAC Days 2021, 6-9 grudnia, konferencja online, **nagroda za najlepszą prezentację** *Looking at the modulated structure of pathogenesis-related protein (Hyp-1) complex with ANS within the higher dimensional superspace approach*

Aperiodic 2022, 20-24 czerwca, 2022, konferencja online, prezentacja *Towards an understanding of structure modulation in macromolecular systems of the Hyp1/ANS protein complex*

European Crystallographic Meeting (ECM33), 23-27 sierpnia 2022, Paryż, Francja, poster *Towards understanding of structure modulation in macromolecular system of Hyp-1/ANS protein complex*

Granty:

Badania strukturalne kwazikryształów dekadonalnych i ikozaedrycznych - nieporządek atomowy, budowa klastrowa i własności fizyczne, grant NCN OPUS nr 2019/33/B/ST3/02063, wykonawca

Przesuwanie granic krystalografii: ku zrozumieniu modulacji struktury w układach makromolekularnych, grant NCN PRELUDIUM nr 2020/37/N/ST3/01501, **kierownik**

Staż i wyjazdy zagraniczne:

UAM, Wydział Chemii, Poznań, Polska, pobyt w terminie 1 marca-30 kwietnia 2018

UAM, Wydział Chemii, Poznań, Polska, pobyt w terminie 4-8 lutego 2019

UAM, Wydział Chemii, Poznań, Polska, pobyt w terminie 20-24 stycznia 2020

Institute of Physics FZU, Praga, Czechy, pobyt w terminie 6 września 2021-8 października 2021

Warsztaty *Remote X-ray data collection from European synchrotrons at the Weizmann Institute of Science*, Rehovot, Izrael, 14-18 maja 2018

International School on Aperiodic Crystals (ISAC4), Portbail, Francja, 8-14 września 2019