

Łukasz Dziadek

**Badania właściwości dynamicznych białek w różnych otoczeniach
metodami gruboziarnistej oraz pełnoatomowej dynamiki molekularnej**

Promotor pracy:

dr hab. Adam Sieradzan, prof. UG

Promotor pomocniczy:

dr Krzysztof Bojarski

Gdańsk 2025

Pragnę serdecznie podziękować osobom, które przyczyniły się do powstania tej pracy.

Doktorowi habilitowanemu Adamowi Sieradzanowi, promotorowi mojej pracy doktorskiej, dziękuję za czuwanie nad tą rozprawą przez cały okres moich studiów doktoranckich aż po etap publikacji. Za pomoc w przeprowadzeniu analiz i licznych dyskusji oraz za nieustającą motywację do krytycznego spojrzenia na problematykę badawczą. Doktorowi Krzysztofowi Bojarskiemu za pomoc, a także za dobre słowo i wsparcie nie tylko naukowe.

Również pragnę podziękować wszystkim pracownikom katedry Chemii Teoretycznej, a w szczególności Profesorowi Adamowi Liwo, Profesorowi Cezaremu Czaplewskiemu oraz Doktorowi habilitowanemu Arturowi Giełdoniowi za nieocenianą pomoc, cenne wskazówki i liczne konsultacje naukowe oraz współpracę przez okres studiów.

Dziękuję również wszystkim znajomym i wszystkim, którzy okazywali wsparcie w trakcie studiów i pisanie rozprawy doktorskiej. W szczególności pragnę podziękować moim rodzicom, za cierpliwość do mojej osoby i wsparcie.

Dedykuję tę pracę świętej pamięci dziadkowi Romanowi Dziadek.

Badania te zrealizowałem w ramach Grantu PPN/BFR/2020 /1 /00041 (dla Ł. J. Dziadek i A. Giełdonia) z Narodowej Agencji Wymiany Akademickiej (NAWA) pt.: "Machine learning combined with coarse-grained modeling of structure and dynamics of proteins". Dziękuję Wydziałowi Chemii Uniwersytetu Gdańskiego za udostępnienie serwera etoh i piasek do przeprowadzenia obliczeń komputerowych. Zasoby obliczeniowe udostępniło również Centrum Informatycznego Trójmiejskiej Akademickiej Sieci Komputerowej (CI TASK) w Gdańsku.

Spis treści

Wykaz skrótów	6
Streszczenie	8
Abstract	10
1 Część teoretyczna	12
1.1 Elastyczność białek	12
1.2 Błony lipidowe	15
1.3 Glikozaminoglikany	16
1.4 Dynamika molekularna	18
1.5 Gruboziarniste pola siłowe	20
1.5.1 Model UNRES	21
1.5.2 Model CABS	23
1.5.3 Model NOLB	25
1.5.4 Model MARTINI	27
1.6 Pełnoatomowe pola siłowe	29
1.6.1 Pole siłowe AMBER	30
1.6.2 Pole siłowe CHARMM	31
1.7 Potencjał średnich sił	33
1.8 Metody półempiryczne	33
2 Cel pracy	35
3 Metody	37
3.1 UNRES oraz UNRES z funkcją więzów na strukturę drugorzędową	37

3.2	CABS-flex	38
3.3	NOLB	38
3.4	Wprowadzenie błony w formie jawnej do modelu UNRES	38
3.4.1	Obliczanie hiperpowierzchni energii potencjalnej	38
3.4.2	Potencjały średniej siły	39
3.4.3	Dopasowanie funkcji analitycznych	41
3.4.4	Optymalizacja wagi i badanie fluktuacji	44
3.5	Obliczanie fluktuacji układów białko-glikozaminoglikan	45
3.5.1	ff14SB/GLYCAM06j	45
3.5.2	CHARMM36m	46
3.6	Walidacja oraz analiza wyników symulacji	46
3.6.1	Wyznaczanie profili fluktuacji	47
3.6.2	Korelacja danych eksperymentalnych z teoretycznymi	48
3.6.3	Usuwanie reszt aminokwasowych z sekwencji białek oraz normalizacja danych	49
3.6.4	Analiza istotności statystycznej	50
3.6.5	Analiza skośności układu	50
3.6.6	Określenie podobieństwa strukturalnego białek otoczonych błoną lipidową	50
4	Wyniki	52
4.1	Wpływ końcowych reszt aminokwasowych na profile fluktuacji	52
4.2	Porównanie przewidywanych i eksperymentalnych profili RMSFN w zestawie te- stowym	55
4.2.1	Dokładność metod teoretycznych względem eksperymentalnych wyznacza- nia profili fluktuacji RMSFN oraz struktury drugorzędowej białek	55
4.2.2	Rozkład współczynników korelacji względem liczby zliczeń	59
4.2.3	Analiza skośności układów	62
4.2.4	Analiza dystrybuanty rozkładów	63

4.2.5	Zależność współczynników korelacji od wielkości białka	65
4.3	Analiza profili RMSFN dla wybranych białek	67
4.4	Analiza właściwości PMF oraz ocena dopasowania funkcji analitycznych dla ukła- dów cząsteczek UNRES-MARTINI	72
4.5	Analiza przewidywanych względem eksperymentalnych profili fluktuacji białek otoczonych błonami lipidowymi	77
4.6	Wpływ wiązania pomiędzy GAGami a białkami na profile fluktuacji reszt amino- kwasowych	83
4.7	Różnica wartości profili fluktuacji między formą związaną i niezwiązaną białek . .	88
5	Podsumowanie i wnioski	92
6	Literatura	95
7	Załączniki	113

Wykaz skrótów

AA - (ang. All-Atom) pełnoatomowe

BU model - (ang. Bottom-Up model) metoda wyznaczania potencjałów poprzez tworzenie podukładów o coraz to większej złożoności, które są następnie łączone razem do osiągnięcia wymaganej funkcjonalności

CG - (ang. Coarse-Grained) gruboziarniste

COM - (ang. block's center of mass) środek ciężkości

DSSP - (ang. Dictionary of Secondary Structure in Proteins) metoda określania struktur drugorzędowych w białkach z zastosowaniem statystyki

ECM - (ang. Extra Cellular Matrix) macierz zewnątrzkomórkowa

FF - (ang. Force Field) pole siłowe

HF - (ang. Hartree-Fock) - metoda do określania funkcji falowej i energii układu kwantowego

LG - (ang. Levitt-Gerstein metric) metryka Levitta-Gersteina

MC - (ang. Monte Carlo) metoda Monte Carlo

MD - (ang. Molecular Dynamics) dynamika molekularna

MNDO - (ang. Modified Neglect of Differential Overlap) nieortogonalizowany zmodyfikowany model pomijania różniczkowego nakładania się

NDDO - (ang. Neglect of Diatomic Differential Overlap) metoda pomijania dwuatomowego nakładania się różniczkowego

NMA - (ang. Normal Mode Analysis) analiza modów normalnych

NMR - (ang. Nuclear Magnetic Resonance) spektroskopia magnetycznego rezonansu jądrowego

NPT - (ang. isobaric-isothermal ensemble) układ izolowany o ustalonym ciśnieniu (P) , ustalonej liczbie atomów (N) i ustalonej temperaturze (T)

NOESY - (Nuclear Overhauser Effect Spectroscopy) spektroskopia jądrowa z wykorzystaniem efektu Overhausera

PES - (ang. Potential Energy Surfaces) powierzchnie energii swobodnej

PMF - (ang. Potential of Mean Force) potencjał średniej siły

RTB - (ang. rotations-translations of blocks) obroty translacji bloków

QM - (ang. Quantum Mechanics) mechanika kwantowa

r_p - (ang. Pearson correlation coefficient) współczynnik korelacji Pearsona

r_s - (ang. Spearman rank) współczynnik korelacji Spearmana

SMD (ang. Solvation Model Density) model solwatacji oparty na gęstości

SC - (ang. Side Chain) łańcuch boczny

WHAM - (ang. Weighted Histogram Analysis Method) metoda ważonych histogramów

X-ray - (ang. X-radiation) rentgenografia strukturalna

Streszczenie

Białka stanowią fundament dla prawidłowej pracy fizjologicznej organizmów, a ich funkcje są w głównej mierze kształtowane przez sekwencję aminokwasów. Sekwencja aminokwasowa nie tylko określa trójwymiarową strukturę białka, ale także elastyczność konformacyjną i dynamiczne zachowanie w określonym środowisku. Elastyczność białka jest istotnym elementem, który określa, w jaki sposób struktura białka wpływa na jego funkcje. Z tego powodu znalezienie elastycznych obszarów w strukturze białkowej jest kluczowe dla zrozumienia roli biologicznej białek. Eksperymentalne badanie elastyczności białek jest często trudne lub wręcz niemożliwe, co sprawia, że podejścia obliczeniowe są szczególnie cenne. Obliczenia przy użyciu klasycznych narzędzi modelowania o rozdzielczości pełno-atomowej pozwalają odwzorować wszystkie atomy badanego układu, co wymaga przeprowadzenia kosztownych obliczeniowo symulacji. W latach 70. XX wieku powstały gruboziarniste modele, które pozwalają zmniejszyć ten czas, grupując kilka atomów jako jedno centrum oddziaływania, pozwalając przyspieszyć obliczenia.

W prezentowanej pracy doktorskiej przeprowadziłem symulacje komputerowe dla zestawu 100 testowanych białek, wykorzystując wybrane modele gruboziarniste (UNRES-flex, UNRES-DSSP-flex, CABS-flex oraz NOLB) do określenia dokładności w przewidywaniu elastyczności fragmentów łańcuchów aminokwasowych białek na podstawie ich struktur eksperymentalnych (uzyskanych ze spektroskopii magnetycznego rezonansu jądrowego oraz rentgenografii strukturalnej). Na podstawie wyników symulacji wyznaczyłem średnie odchylenie standardowe fluktuacji i na ich podstawie obliczyłem wartości średnich współczynników korelacji Pearsona i Spearmana w stosunku do wartości fluktuacji eksperymentalnych. Następnie wykorzystując średnie wartości współczynników korelacji określiłem, która z metod gruboziarnistych najlepiej sprawdza się w przewidywaniu elastyczności białek dla danego typu struktury drugorzędowej (α , β , $\alpha + \beta$) oraz metody uzyskania struktury (zespołów NMR oraz X-ray). Do zweryfikowania wyników wykorzystałem odpowiednie testy statystyczne (test studenta oraz dwukierunkowa analiza wariancji). Również przeprowadziłem analizę współczynników korelacji względem liczby zliczeń, skośności oraz dystrybuanty rozkładów oraz określiłem zależność wartości współczynników korelacji od wielkości białek.

W kolejnej części przeprowadziłem symulację wykorzystując półempiryczną metodę PM7 oraz

pełnoatomowe pole siłowe ff14sb w pakiecie AMBER, uzyskując hiperpowierzchni energii potencjalnej, na podstawie których wyznaczyłem potencjały średniej siły, a następnie dopasowałem do nich funkcje analityczne dla badanych układów pomiędzy cząsteczkami z modelu UNRES oraz cząsteczkami z modelu MARTINI. Wyznaczyłem oddziaływania pomiędzy 2 cząsteczkami hydrofobowymi, 2 cząsteczkami polarnymi oraz 2 cząsteczkami naładowanymi o różnym ładunku, aby określić, która z zastosowanych metod (PM7 czy AA) jest lepsza do symulowania różnego rodzaju oddziaływań oraz wyznaczyłem potencjały średniej siły i dopasowałem do nich funkcje analityczne. Ponadto przeprowadziłem symulacje elastyczności białek otoczonych błoną lipidową za pomocą nowego pola UNRES-MARTINI.

W ostatniej części przeprowadziłem symulacje wykorzystując pełnoatomowe pola siłowe ff14sb/GLYCAM06j z pakietu AMBER oraz CHARMM36m do określenia, która spośród zastosowanych metod lepiej odzwierciedla wpływ glikozaminoglikanów na elastyczność białek w formie związanej (kompleksu) w porównaniu do form niezwiązanych, analizując uzyskane profile fluktuacji.

Uzyskane wyniki pozwoliły mi określić, która z metod gruboziarnistych może skutecznie odtworzyć elastyczność białek ze względu na występujący typ struktury drugorzędowej oraz metodę uzyskania struktury białek. Umożliwiły również porównanie półempirycznej metody PM7 z pełnoatomową ff14sb pod względem symulowania różnego rodzaju oddziaływań, a także określić dokładność przewidywania elastyczności białek w błonach lipidowych za pomocą nowego pola siłowego UNRES-MARTINI. Określiłem również, które z zastosowanych pełnoatomowych pól siłowych najlepiej odzwierciedla wpływ GAGów na elastyczność białek. Przeprowadzone przeze mnie symulacje pozwoliły określić, które z zastosowanych metod oraz modeli są w stanie prawidłowo przewidywać elastyczność struktur białkowych.

Abstract

Proteins are essential for proper, physiological functioning of organisms, and their functions are largely determined by the amino acid sequence, which governs not only proteins three-dimensional structure but also their conformational flexibility and dynamic behavior in specific environments. Protein flexibility is a crucial factor that influences the way protein structure is correlated with function. Therefore, identifying flexible regions within these macromolecules is essential to understand their biological functions. Experimental investigation of protein flexibility is often difficult or even impossible, which makes computational approaches particularly valuable. Computations using classical atomic-resolution modeling tools allow for mapping all the atoms of the studied system, which requires large computational cost. In the 1970s, coarse-grained models have been developed that allow for reducing this time by grouping several atoms as one interacting center, allowing for computational speed-up.

In this doctoral thesis, I carried out computer simulations for a test set of 100 proteins using selected coarse-grained models (UNRES-flex, UNRES-DSSP-flex, CABS-flex and NOLB) to evaluate their accuracy in predicting the flexibility of amino acid chain fragments based on experimental structures (obtained from nuclear magnetic resonance spectroscopy and X-ray crystallography). Based on the computational results, I calculated standard deviations of fluctuations, followed by Pearson and Spearman correlation coefficients with respect to the experimental fluctuation values. Using average correlation coefficients, I identified which coarse-grained method most accurately predicts protein flexibility depending on the type of secondary structure (α , β , $\alpha + \beta$) and the structure determination method (NMR ensembles or X-ray crystallography). I analyzed statistical significance using the Student's t-test and two-way ANOVA. In addition, the correlation coefficients I analyzed with respect to the numbers of count, the skewness, and the cumulative distribution functions, as well as the relationship between the size of the protein and the correlation values.

In the next part, I carried out simulations using the semi-empirical PM7 method and the ff14SB all-atom force field within the AMBER package obtaining a hypersurface of potential energy, based on which, I determined the potentials of mean force and then fitted analytical functions to them for the studied systems (between molecules from the UNRES model and molecules from the MARTINI

model). I fitted analytical functions to the resulting PMF, comparing interactions between two hydrophobic, two polar, and two charged molecules in the UNRES and MARTINI models to determine which method more accurately represents specific interaction types. Furthermore, I simulated protein flexibility simulations in a lipid membrane environment using the new UNRES-MARTINI force field and compared to the existing UNRES model.

In the last part, I was using all-atom force fields ff14SB/GLYCAM06j (AMBER) and CHARMM36m simulations to assess which method better captures the influence of glycosaminoglycans (GAG) on the flexibility of GAG-bound in complex proteins and compared with unbound proteins, based on fluctuation profiles.

The obtained results allowed me to identify which coarse-grained method most effectively reproduce protein flexibility depending on the type of secondary structure and the structural determination method. I also compared PM7 method with ff14SB in terms of describing various types of interaction centers in UNRES-MARTINI. I also helped to determine which all-atom force field best reflects the influence of GAGs on protein flexibility. These insights can significantly enhance the integration of protein flexibility into drug design workflows using computationally efficient theoretical methods.

Rozdział 1

Część teoretyczna

1.1 Elastyczność białek

Białka należą do grupy związków chemicznych, które często bada się pod względem ich elastyczności strukturalnej przeprowadzając symulacje komputerowe z wykorzystaniem metod obliczeniowych^{1,2}. Pierwsze informacje na temat zależności pomiędzy strukturą a funkcją białek zostały zauważone na początku lat 50-tych XX wieku³. Dziś wiemy, że zrozumienie zależności pomiędzy strukturą przestrzenną białek a ich aktywnością (z ang. structure-activity relationship), jest niezbędne do właściwego zrozumienia, w jaki sposób białka działają⁴. Struktury większości białek są wysoce dynamiczne, a ich funkcje biologiczne w dużym stopniu zależą od elastyczności. Dynamiczna natura białek (odnosząc się do zmian strukturalnych, którym ulegają w różnych warunkach) umożliwia im zmianę konformacji w odpowiedzi na inne cząsteczki lub zmiany środowiskowe. Ta zdolność jest kluczowa dla różnych procesów biologicznych i biochemicznych, w tym przekazywania sygnału, rozpoznawania antygenów, transportu białek, czy też katalizy enzymatycznej⁵. Elastyczność białek może prowadzić do drobnych zmian, takich jak ruch kilku łańcuchów bocznych aminokwasów w celu dostosowania się do małego substratu lub do poważniejszych transformacji, takich jak fałdowanie białek ułatwione przez obecność określonego ligandu⁵. Informacje o strukturze nie są wystarczające do pełnego zrozumienia funkcjonowania badanego białka⁵, dlatego określa się ruchliwość fragmentów struktury, które wpływają na prawidłowe ich funkcjonowanie. Jedną z najważniejszych zmiennych wpływających na ruchliwość fragmentów białkowych jest ich kontakt z rozpuszczalnikiem (w którym zachodzi proces dyfuzji). Zatem do czynników wpływających na lokalne zmiany konformacyjne należy powłoka hydratacyjna białka, kształt oraz polarność rozpuszczalnika⁶.

Dynamika białek jest niezbędna do procesów katalizy, ponieważ uczestniczy w rozpoznawaniu, wiązaniu i uwalnianiu ligandu⁴. Jest także zaangażowana w różne procesy, w tym regulację allosteryczną oraz stabilność białka⁷⁻¹². Ponadto zmiany konformacyjne są odpowiedzialne za nieprawidłowe fałdowanie i agregację białek, które przyczyniają się między innymi do rozwoju chorób neurodegeneracyjnych¹³. Aby wyjaśnić funkcje białka, informacje na temat elastyczności są tak samo ważne, jak informacje na temat jego struktury^{9,10}. Elastyczność białek jest niezbędna również do zrozumienia sposobów, w jaki leki wpływają na organizmy, jak są transportowane, czy metabolizowane. Istotna jest kinetyka wiązania leku i lokalizacji miejsca wiązania¹⁴. Elastyczność struktur białkowych (ich konformacji) pozwala na zwiększone powinowactwo między lekiem a jego celem molekularnym, co jest związane z "dopasowaniem indukowanym", podczas którego lek (ligand) wiąże się do receptora w konformacji najniższej energii białka, po czym następuje zmiana konformacyjna w celu dostosowania się do liganda¹⁴.

Opis dynamicznych cech białek można badać przy użyciu metod eksperymentalnych. Współczynnik B¹⁵, znany również jako współczynnik temperaturowy lub parametr przemieszczenia atomowego, jest najczęstszym deskryptorem elastyczności dostępnym dla większości rozwiązanych struktur białkowych, który można uzyskać z krystalografii rentgenowskiej. Odzwierciedla tłumienie rozpraszania promieni rentgenowskich spowodowane ruchem termicznym^{16,17}. Współczynnik B jest związany z rozdzielczością, więc może mieć wysoką wartość ze względu na dużą ruchliwość białka i niską rozdzielczość struktury. Krystalografia rentgenowska ma pewne ograniczenia, zwłaszcza na początkowym etapie procesu, który wymaga uzyskania wysokiej jakości kryształu białka¹⁸. Krystalografia rentgenowska zakłada, że potencjał w którym porusza się atom jest harmoniczny. Zgodnie z przybliżeniem sferycznym przemieszczenie jest izotropowe. Wiele badań i obliczeń dotyczących dynamiki molekularnej sprzeciwia się przybliżeniom izotropowym¹⁹. Upakowanie, ruchy ciał sztywnych cząsteczek²⁰, czy też rozdzielczość krystaliczna wpływa na czynniki krystalograficzne B, co ogranicza zastosowanie tego współczynnika do badania dynamiki²¹. Czyni to współczynnik B parametrem opisującym ruchliwe elementy białka tylko w pewnym przybliżeniu.

Następnym przykładem metody eksperymentalnej, dzięki której możliwe jest uzyskanie dynamiki białek w roztworze^{22,23} jest spektroskopia jądrowego rezonansu magnetycznego (NMR). Wüthrich i Wagner²⁴, Campbell i współpracownicy²⁵, oraz Snyder i inni²⁶ przeprowadzili pierwsze ekspe-

rymenty z NMR, które wykazały, że białka nie są statycznymi bytami; są raczej podatne na znaczne zmiany strukturalne. Wiele zmieniło się pod względem pełnienia funkcji NMR w badaniach dynamiki biomolekularnej od lat 70. Wielowymiarowe badania NMR wykorzystują przesunięcia chemiczne w dwóch lub więcej wymiarach, aby uzyskać znacznie większą rozdzielczość struktur białek^{27,28}.

Bez wątpienia spektroskopia NMR oferuje niezwykle dokładną i solidną metodą eksperymentalną do pomiaru wolnych ruchów białek w rozpuszczalniku lub roztworze. Jednak standardowe metody NMR często wymagają skomplikowanych i zaawansowanych analiz danych i mają wiele ograniczeń (takich jak współczynniki relaksacji i konieczność wykonywania dodatkowych czasochłonnych pomiarów NMR)^{29–32}. Za pomocą metod obliczeniowych, takich jak analiza modów normalnych³³ lub dynamiki molekularnej¹⁵ można uzupełnić informacji dotyczącego elastyczności białek z metody NMR.

Ograniczenia strukturalne używane do określenia struktury białek głównie pochodzą z sygnałów NOESY (z ang. Nuclear Overhauser Effect Spectroscopy). Do ograniczeń strukturalnych zalicza się odległości między protonami, kąty dwuścienne, wzajemną orientację wiązań³⁴. Problem z tym eksperymentem pojawia się podczas, gdy ruch w pośredniej skali czasu powoduje poszerzenie linii, co może utrudnić identyfikację szczytów poprzecznych NOESY (szybkie ruchy redukują NOE)³⁵. Eksperymenty NOE wymagają skomplikowanych warunków eksperymentalnych i są kosztowne zarówno pod względem czasu, jak i zasobów materiałowych.

Równolegle do ostatnich postępów w technikach eksperymentalnych, badanie elastyczności jest jednym z głównych celów metod teoretycznych. Już w latach 70, dwaj teoretycy Arieh Warshel i Michael Levitt³⁶ przewidzieli fałdowanie białek, wykorzystując konformację białek i minimalizację energii. W badaniach teoretycznych³⁷ i eksperymentalnych³⁸ wykorzystano uproszczony opis trójstanowy do określenia i zbadania struktury i dynamiki białek. Dwa z nich, α -helisy i β -harmonijki (złożone z β -nici), to powtarzalne struktury, które są stabilizowane przez wewnątrzcząsteczkowe wiązania wodorowe. Struktury te są połączone z innymi bardziej zmiennymi strukturami, takimi jak losowe kłębki (z ang. random coil)³⁷.

Dynamika molekularna (MD)³⁹ w reprezentacji pełnoatomowej to jedna z metod powszechnie stosowanych. Metoda ta pokazuje elastyczność białka w warunkach fizjologicznych. Symulacje MD w porównaniu do eksperymentów są mniej kosztowne i czasochłonne⁴⁰. Metody gruboziarniste (CG)

w połączeniu z prostymi potencjałami^{41,42} oferują tańszą alternatywę dla pełnoatomowego MD. Stosując CG, tracimy szczegóły atomowe, aby uzyskać obliczeniową prostotę w przedstawieniu elastyczności białek^{43,44}.

1.2 Błony lipidowe

Struktury lipidów występujące w naturze są tak różnorodne, jak różne role, jakie lipidy odgrywają w procesach komórkowych. Lipidy odgrywają główną rolę w funkcjonowaniu komórek, tworząc podwójną warstwę lipidową, która chroni komórki i organelle subkomórkowe przed przepuszczalnością⁴⁵. Funkcje komórkowe takie jak przekazywanie sygnału, plastyczność błony i transport błonowy są zmienne ze względu na zmiany w składzie i organizacji lipidów. Lipidy również są ważne ze względu na interakcje z białkami błonowymi⁴⁶. Kompleks białek błonowych jest kluczowy dla dostosowania metabolizmu białek i lipidów w odpowiedzi na sygnały wewnętrzne i zewnętrzne⁴⁷, dlatego badanie lipidów, a w szczególności błon lipidowych pod względem ich interakcji z białkami jest niezwykle istotne.

Od początku lat 70 właściwości fizyczne dwuwarstwy były szeroko badane. Poczyniło to pierwszy krok w zrozumieniu zawiłych zachowań błon komórkowych^{48–51}. Ich fascynująca morfologia wynika z płynności, czyli zdolności cząsteczek w nich zawartych do przemieszczania się na boki wzdłuż membran. Błony komórkowe aktywnie uczestniczą w zmianach morfologicznych w komórkach i organellach, takich jak podział komórek, endocytoza, autofagia i wzrost neuronów⁵². Określenie, w jaki sposób lipidy i białka oddziałują ze sobą, jest ważnym tematem badań naukowych, ponieważ jest niezbędne do rozwikłania struktury i funkcji białek błonowych.⁵³ Elastyczność strukturalna białek związanych z błoną wpływa na stabilność błony i odgrywa rolę regulującą interakcji lipid-białko. Niewystarczająca sztywność błony może powodować uszkodzenie lipidów, białek i DNA, a następnie śmierć komórki⁵⁴.

Obecnie możliwe jest przeprowadzenie symulacji dynamiki molekularnej dowolnego strukturalnie scharakteryzowanego białka otoczonego błoną lipidową⁵⁵ w skali czasowej do $\sim 0,1\mu s$ dzięki znacznej poprawie mocy obliczeniowej i opracowaniu bardziej skutecznych metodologii. Badania α -hemolizyny, jednego z najbardziej dokładnie zbadanych kanałów błonowych, pokazują postęp w

symulacji MD dla białek błonowych⁵⁵. Współczesne symulacje MD zapewniają wystarczające próbkowanie, aby dokładnie odtworzyć właściwości elektryczne⁵⁵. Precyzja, która została osiągnięta przez symulacje MD, jest tak niezwykła, że można je uznać za narzędzie do obrazowania w skali nano⁵⁶.

Ponadto symulacje MD wyjaśniają zasady fizyczne, które stanowią podstawę ważnych procesów membranowych. Na przykład przypisanie odpowiednich stanów jonizacji do miareczkowalnych grup aminokwasowych ujawnia selektywne kanały jonowe⁵⁷⁻⁶⁰. W początkowych badaniach nad układami lipidowymi wykorzystano uproszczone modele lipidów, w których często nie brano pod uwagę rozpuszczalnika⁶¹⁻⁶³. Jednak z czasem modele rozszerzono, aby reprezentowały wszystkie atomy cząsteczek lipidów z uwzględnieniem wody^{64,65}. Wyniki przeprowadzonych badań^{64,65} pokazują, że MD może szczegółowo opisywać ruchy białek i lipidów⁶⁶.

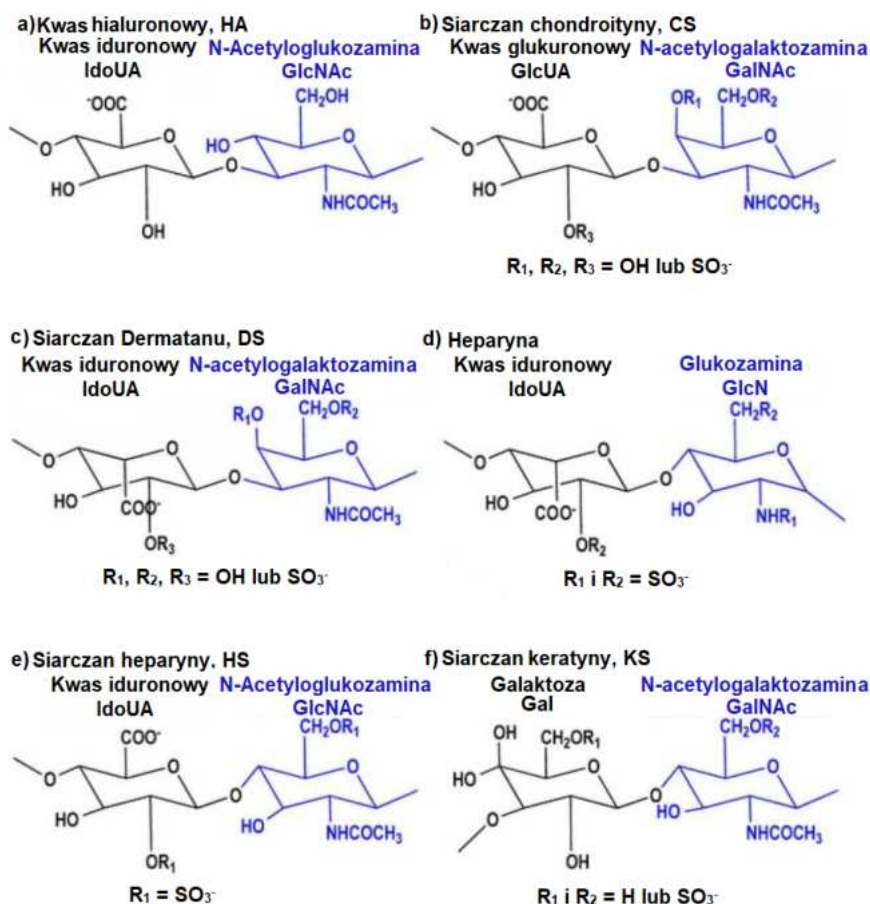
MD jest doskonałym sposobem oceny dynamiki lipidów, za pomocą którego możliwe jest zbadanie zachowania cząsteczek rozpuszczalnika zarówno w, jak i w pobliżu dwuwarstw. Możliwe jest również zbadanie różnic w zachowaniu różnych rodzajów lipidów w odniesieniu do ich struktury i dynamiki rozpuszczalnika⁶⁶.

1.3 Glikozaminoglikany

”Glikany” są złożonymi koniugatami oraz rodzajem węglowodanów. Należą do związków zbudowanych z wielu cząsteczek cukrów prostych połączonych wiązaniami glikozydowymi. Glikany są niezbędne w wielu interakcjach komórka-komórka i komórka-macierz, które odgrywają kluczową rolę w rozwoju i działaniu złożonych organizmów wielokomórkowych⁶⁷.

Natomiast glikozaminoglikany (GAGi) to długie, nierozgałęzione, ujemnie naładowane polisacharydy, które składają się z naprzemiennie powtarzalnych jednostek dwucukrowych⁶⁸. Pomimo, że GAGi wykazują dużą różnorodność strukturalną, można między nimi zaobserwować podobieństwa. Dzięki temu można wyróżnić takie klasy, jak i; siarczan chondroityny (CS), siarczan dermatanu (DS), siarczan keratanu (KS), heparyna (HP), siarczan heparanu (HS). Kwas hialuronowy (HA) jest jedynym niesiarczanowanym i wolnym GAGiem, który jest niezwiązany z proteoglikanami (czyli makrocząsteczkami, których białko rdzeniowe jest połączone wiązaniem kowalencyjnym chociaż z jednym łańcuchem GAG)⁶⁹. Siarczanowanie to złożony proces biotransformacji (polega na dodaniu

polarnej grupy siarczanowej). Może występować w różnych pozycjach w szkielecie GAGu i reguluje sygnały zewnątrzkomórkowe, takie jak interakcje komórka-komórka i komórka-molekuła. W rezultacie zostały zidentyfikowane różne wzorce siarczanowe narządów i komórek w trakcie ich rozwoju⁷⁰. Wzorce siarczanowe GAGów zostały przedstawione na rysunku 1.1⁷⁰.



Rysunek 1.1: Struktury jednostek powtarzalnych w różnych GAG-ach. (A) HA jest niesiarczanowanym polimerem z [4]- β -GlcUA-(1-3)- β -GlcNAc-(1-) jako jednostką powtarzalną. (B) CS jest polimerem siarczanowanego [4]- β -GlcUA-(1-3)- β -GalNAc-(1-). W zależności od wzoru siarczanowania, CS można klasyfikować do różnych podtypów (CSA, CSC, CSD i CSE). (C) DS składa się z powtarzającej się jednostki dimerycznej [4]- α -IdoUA-(1-3)- β -GalNAc-(1-3]. L-IdoUA jest epimerem D-GlcUA, a reszta GalNAc jest siarczanowana w różnych pozycjach podobnie do CS. (D) Główną jednostką disacharydową heparyny jest [4]- α -IdoUA-(1-4)- α -GlcN-(1-]. Kwas iduronowy jest siarczanowany przy O2, a jednostka GlcN jest N- i O6-siarczanowana. (E) Główną jednostką disacharydową HS jest [4]- α -GlcUA-(1-4)- β -GlcNAc-(1-]. Reszta GlcNAc jest na ogół siarczanowana przy N2, z lub bez siarczanowania O6. (F) W typowym KS podjednostki disacharydowe składają się z Gal i GlcNAc-6-siarczanu. Reszty Gal mogą mieć lub nie siarczanowania O6. "Adapted with permission (CC BY 4.0) from Rawat PS, Seyed Hameed AS, Meng X, Liu W. Utilization of glycosaminoglycans by the human gut microbiota: participating bacteria and their enzymatic machineries, Gut Microbes, 2022, 14 (1),2068367. Copyright 2022 National Center of Biotechnology Information ."

GAGi mogą wiązać się z białkami i zmieniać ich elastyczność konformacyjną. Badanie ich elastyczności pozwala określić, w jaki sposób struktura białka zmienia się pod wpływem GAGów, a także pozwala zbadać elastyczności struktur białkowych w środowisku silnie naładowanych cząsteczek.

GAGi odgrywają ważną rolę w wielu procesach biologicznych. Na przykład; GAGi oddziałują z chemokinami⁷¹ oraz receptorami czynników wzrostu^{72,73}. Zaburzenie tych oddziaływań może prowadzić do rozwoju poważnych chorób, takich jak nowotwory⁷⁴, choroby Alzheimer'a i Parkinsona⁷⁵ oraz zaburzenia regeneracji tkanek⁷⁶.

Jednym z problemów, który występuje podczas badania GAGów jest ich znaczna długość. Dlatego do ich symulowania stosuje się coraz częściej modelowanie gruboziarniste, na podstawie którego skatalogowano łańcuchy heparyny zawierające od 6 do 68 merów⁷⁷. Dodatkowo opracowano algorytmy dedykowane do generowania biblioteki niesiarczanowanej chondroityny zawierającej od 10 do 200 merów⁷⁸.

Możliwości przewidywania miejsc wiązania heparyny w białkach, identyfikowania biologicznie aktywnych konformacji GAG, analizowania sił międzycząsteczkowych regulujących powinowactwo wiązania GAG z białkiem oraz zrozumienia specyficzności i selektywności interakcji GAGów z białkami znacznie wzrosły dzięki wykorzystaniu dynamiki molekularnej^{79,80}.

1.4 Dynamika molekularna

Jedną z metod, która jest stosowana do badania elastyczności białek jest MD. Co istotne, MD nie zakłada występowania tylko harmonicznym drgań (odkształceń). W dynamice molekularnej rozwiązywanie numerycznych klasycznych równań ruchu powoduje zmianę czasową układu składającego się z wielu atomów⁸¹. Druga zasada dynamiki Newtona (równanie 1.1) daje w takim układzie siłę wypadkową $\vec{F}_i$, która jest sumą wszystkich sił działających na i-ty atom:

$$\vec{F}_i = m_i \vec{a}_i \quad (1.1)$$

Gdzie masa i przyspieszenie atomu to następująco m_i i a_i . Wektor położenia i-tego atomu $\vec{r}_i$ w danym momencie czasu t wyznacza przyspieszenie za pomocą równania 1.2:

$$\vec{a}_i = \frac{d^2 \vec{r}_i}{dt^2} \quad (1.2)$$

Znajomość sił, jakimi działają na siebie poszczególne atomy, jest niezbędna do wykonania obliczeń numerycznych dotyczących ewolucji czasowej układu. Odpowiedni potencjał (fizyczny) jest kluczowy dla poprawnego obliczenia sił działających pomiędzy atomami, które opisują najczęściej oddziaływania dla par lub w trójce atomów. Siłę oddziaływania w parze atomów opisuje równanie 1.3:

$$\vec{F}_{ij} = -\nabla V(\vec{r}_{ij}) \quad (1.3)$$

Gdzie $\vec{F}_{ij}$ jest siłą, z jaką oddziałują na siebie atomy w parze i-j, $V(\vec{r}_{ij})$ jest potencjałem opisującym oddziaływanie między tymi atomami, a $\vec{r}_{ij}$ jest wektorem określającym odległość między tymi atomami. Siła wypadkowa (równanie 1.4), która działa na i-ty atom w N-atomowym układzie, jest sumą wszystkich sił, z jakimi oddziałuje on z innymi atomami w układzie.

$$\vec{F}_i = \sum_{\substack{j=1 \\ j \neq i}}^N \vec{F}_{ij} \quad (1.4)$$

Znając potencjały opisujące oddziaływania między atomami, warunki początkowe i brzegowe, nie jesteśmy w stanie określić położenia atomów i ich prędkości w dowolnym momencie czasu, a równania Newtona rozwiązuje się za pomocą metody różnic skończonych. Polega ona na rozwiązywaniu równań po kolei dla określonych, niewielkich kroków czasowych Δt ^{82,83}, dlatego ustanowienie odpowiedniego kroku czasowego jest kluczowe. Jeśli wyniki zostaną otrzymane przy użyciu zbyt długiego kroku czasowego, będą one niefizyczne, a układ może się nawet rozpaść. Przyjmuje się, że okres wibracji o najwyższej częstotliwości w układzie powinien być około dziesięciokrotnie dłuższy od kroku czasowego⁸².

Dynamika molekularna to technika obliczeniowa wykorzystująca ruchy jonów, wody, małych cząsteczek i makrocząsteczek lub bardziej złożonych układów, takich jak całe wirusy, aby odtworzyć zachowanie środowiska biologicznego, w tym cząsteczek wody i błon lipidowych. Ruchy strukturalne, takie jak te zależne od temperatury i substancji rozpuszczonej lub rozpuszczalnika, są bardzo ważne dla badania kompleksów ligand-białko lub białko-białko. W tym kontekście symulacje

MD są bardzo przydatne, ponieważ ta metoda może modelować istotnie biologiczne ruchy. Wykorzystanie symulacji MD do projektowania leków pokazują również wnęki strukturalne, które są niezbędne do projektowania nowych struktur z większym powinowactwem do celu molekularnego⁸⁴. Dodatkowo symulacje MD są wykorzystywane do pełnego badania struktury i dynamiki cząsteczek biologicznych, ich kompleksów oraz zmian konformacyjnych w białkach i kwasach nukleinowych⁸⁵. Symulacje są przeprowadzane przez badanie położenia centrów, które mają zdolność interakcji ze sobą w czasie, co pozwala im określić dynamiczną ewolucję systemu. Symulacje MD dają obraz zmian struktury, naśladując zmiany w strukturze cząsteczek biologicznych w danym czasie. Dane te ułatwiają zrozumienie funkcji biologicznych⁸⁵. Symulacje dynamiki molekularnej nie ograniczają się tylko do pełnoatomowych pól siłowych (uwzględniających w obliczeniach wszystkie atomy) ale również można używać gruboziarnistych pól siłowych, które upraszczają grupę atomów w postać "pseudoatomów".

1.5 Gruboziarniste pola siłowe

Ostatnie postępy w technologii i metodologii pozwoliły na symulację układów makromolekularnych przy użyciu pól pełnoatomowych w biologicznie istotnych skalach czasowych⁸⁶⁻⁸⁸. Możliwe jest symulowanie zmian konformacyjnych w małych i średnich białkach na krótkie okresy czasu, takich jak milisekundy jak i również w dłuższych zakresach czasowych (w sekundach)⁸⁹⁻⁹³. Przewidywanie struktury i przeszukiwanie przestrzeni konformacyjnej dużych kompleksów makromolekularnych w długich skalach czasowych przy rozdzielczości pełnoatomowej jest obecnie poza zasięgiem. Z tego powodu opracowanie uproszczonych modeli jest szczególnie ważne w kontekście interpretacji danych eksperymentalnych⁹⁴⁻⁹⁶, które stają się coraz bardziej dostępne w wysokiej rozdzielczości. Dane eksperymentalne dają częściowy wgląd w niektóre elementy układów makromolekularnych, ale nie dają bezpośrednio pełnej reprezentacji dynamicznej, a symulacje mogą pomóc w uzyskaniu bardziej kompleksowego zrozumienia⁹⁷⁻⁹⁹. Uproszczone modele mogą zapewnić wgląd w ogólne zasady fizykochemiczne, regulujące systemy biofizyczne, zachodzących zmian na poziomie molekularnym¹⁰⁰.

W naszym życiu istnieje wiele ważnych procesów biologicznych, takich jak fałdowanie białek^{101,102}, wiązania/oddziaływania międzycząsteczkowego^{103,104}, zjawiska błonowe z udziałem białek

lek,^{105,106} które mają bezpośredni wpływ na zdrowie^{107,108}, a są trudne do zbadania przeprowadzając symulacje pełnoatomowe¹⁰⁹. W związku z tym, wykorzystanie CG okazuje się przydatne, przede wszystkim w badaniach o długich skalach czasowych i wielkich układów, które wykraczają poza standardowe możliwości pól pełnoatomowych¹¹⁰, poprzez redukcję liczby stopni swobody reprezentujących badany układ. Zastosowanie symulacji CG pozwala również "wygładzić" krajozobraz energii potencjalnej systemu, umożliwiając przeprowadzenie obliczeniowo niedrogich symulacji^{111–113}. CG zostało wykorzystane po raz pierwszy przez Smita¹¹⁴, zwłaszcza w systemach lipidowych, i obecnie jest szeroko stosowane do badania białek¹¹⁵, kwasów nukleinowych¹¹⁶, błon lipidowych¹¹⁷, węglowodanów¹¹⁸, czy też wody¹¹⁹.

Zaproponowano kilka modeli, aby stworzyć efektywne funkcje energii CG dla dużych układów molekularnych, które odtwarzają cechy strukturalne modeli atomistycznych (BU)^{120–125} lub odtwarzają właściwości makroskopowe jednego lub wielu układów^{97–99,101,126–128}. W modelach CG popularne podejścia BU odtwarzają kanoniczny rozkład konfiguracji, który został określony przez model pełnoatomowy.

1.5.1 Model UNRES

Jednym z powszechnie stosowanych modeli CG jest UNRES (ang. UNited RESidue)¹²⁹ (rysunek 1.2¹³⁰), w którym główny łańcuch polipeptydowy jest reprezentowany przez sekwencję atomów węgla $C\alpha$, gdzie centrami interakcji są scalone łańcuchy boczne i scalone grupy peptydowe. Efektywna energia oddziaływania została opracowana poprzez rozwinięcie potencjału średniej siły w funkcji kumulantów klastra Kubo¹³¹. Funkcja energii modelu UNRES jest wyrażona równaniem 1.5:

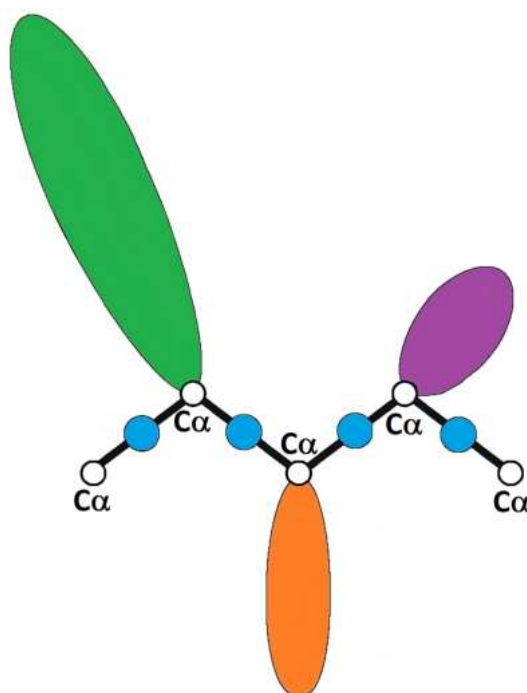
$$\begin{aligned}
U = & w_{SC} \sum_{i < j} U_{SC_i SC_j} + w_{SCp} \sum_{i \neq j} U_{SC_i p_j} + w_{pp}^{VDW} \sum_{i < j-1} U_{p_i p_j}^{VDW} + w_{pp}^{el} f_2(T) \sum_{i < j-1} U_{p_i p_j}^{el} \\
& + w_{tor} f_2(T) \sum_i U_{tor}(\gamma_i, \theta_i, \theta_{i+1}) + w_b \sum_i U_b(\theta_i) + w_{rot} \sum_i U_{rot}(\theta_i, \alpha_{CS_i}, \beta_{SC_i}) \\
& + w_{bond} \sum_i U_{bond}(d_i) + w_{corr}^{(3)} f_3(T) U_{corr}^{(3)} + w_{turn}^{(3)} f_3(T) U_{turn}^{(3)}
\end{aligned} \tag{1.5}$$

Gdzie $U_{SC_i SC_j}$ przedstawia średnią swobodną energię oddziaływań hydrofobowych (hydrofilowych) między łańcuchami bocznymi, $U_{SC_i p_j}$ oznacza potencjał objętości wykluczonej oddziaływań łańcuch boczny-grupa peptydowa, $U_{p_i p_j}$ opisuje potencjał oddziaływań grupa peptydowa-grupa peptydowa,

VDW oraz el to odpowiednio wkłady van der waalsa oraz elektrostatyczne. $U_{bond}(d_i)$ to proste lub multimodalne potencjały harmoniczne rozciągania wiązania wirtualnego, gdzie d_i jest długością i -tego wiązania wirtualnego, U_{tor} , U_b i U_{rot} to odpowiednio potencjały torsyjne pochodzące od obrotów wokół wirtualnych wiązań $C^\alpha \cdots C^\alpha$, wkłady pochodzące od odkształcania wirtualnych kątów walencyjnych, wkłady rotameryczne odpowiadające zmianom położenia centrów łańcuchów bocznych względem szkieletu peptydowego, zaś $U_{corr}^{(3)}$ i $U_{turn}^{(3)}$ to odpowiednio wkłady uwzględniające sprzężenie między lokalnymi oddziaływaniami łańcucha głównego i oddziaływaniami elektrostatycznymi. W modelu UNRES rozpuszczalnik jest w formie niejawnej. Wkłady do energii efektywnej (równanie 1.5) przedstawiają interakcje między białkiem a rozpuszczalnikiem w formie niejawnej, głównie wkład $U_{SC_iSC_j}$. Czynniki f_n opisuje zależność efektywnej funkcji energii od temperatury, co wskazuje, że jest to potencjał średniej siły, czyli ograniczona energia swobodna, a nie energia potencjalna. Równanie 1.6 przedstawia czynnik $f_n(T)$ ¹³²:

$$f_n(T) = \frac{\ln[\exp(1 + \exp(-1))]}{\ln \{ \exp[T/T_0]^{n-1} + \exp[-(T/T_0)^{n-1}] \}} \quad (1.6)$$

Gdzie $T_0 = 300K$. W UNRES jest wykorzystywana głównie dynamika molekularna Langevina, która została zastosowana we wcześniejszych pracach^{133,134}. UNRES z powodzeniem został zaaplikowany do badania dynamiki i termodynamiki fałdowania białek¹³⁵, przewidywaniu struktury białek¹³⁶ i rozwiązywaniu problemów biologicznych¹³⁰.



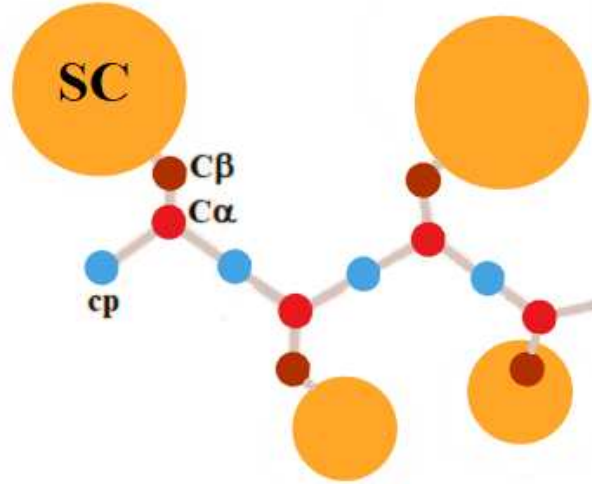
Rysunek 1.2: Gruboziarnisty model UNRES dla przykładowego łańcucha polipeptydowych, którym jasnoniebieskimi kulami oznaczono scalone grupy peptydowe pomiędzy kolejnymi atomami $C\alpha$ oznaczonymi białymi kulami, zaś zjednoczone łańcuchy boczne połączone z atomami $C\alpha$ oznaczono jako sferoidy o różnych kolorach (zielony, pomarańczowy oraz fioletowy), a także wielkościach.

1.5.2 Model CABS

Kolejnym przykładem modelu CG średniej rozdzielczości¹³⁷ jest CABS (rysunek 1.3¹³⁸), w którym atomy $C\alpha$ reprezentują kolejno połączone atomy. Pomiędzy atomami $C\alpha$ znajdują się grupy peptydowe, a każdy łańcuch boczny jest reprezentowany przez dwa centra oddziaływań; przez atomy $C\beta$ oraz przez zjednoczoną grupę obejmującą odpowiednie atomy łańcucha bocznego następujące po $C\beta$. Modelowane łańcuchy polipeptydowe są nałożone na sześcienną siatkę o dużej gęstości. Model ten spełnia wymagania odwracalności mikroskopowej¹³⁹, wykorzystując dynamikę Monte Carlo z asymetrycznym schematem Metropolis. Reprezentacja sieciowa umożliwia bardzo szybkie próbkowanie przestrzeni konformacyjnej, ponieważ umożliwia wstępne obliczenie większości składowych energii. DSSP automatycznie określa dane o strukturze drugorzędowej dla modelu CABS¹⁴⁰. Dane struktury drugorzędowej są uproszczone do reprezentacji helisy, harmonijki lub losowego kłębaka¹⁴⁰. Energię wyraża równanie 1.7:¹³⁹

$$V(q) = w_{SDD}E_{SSD} + w_{SSI}E_{SSI} + w_{HB}E_{HB} + w_RE_R + w_{LR}E_{LR} \quad (1.7)$$

Gdzie E_{SSD} (o wadze $w_{SSD}=1,0$) to energia oddziaływań niezależnych od sekwencji krótkiego zasięgu, E_{SSI} (o wadze $w_{SSI}=0,375$) to energia oddziaływań zależnych od sekwencji krótkiego zasięgu, E_{HB} (o wadze $w_{HB}=1,0$) to energia wiązania wodorowego, E_R (o wadze $w_R=1,0$) to energia oddziaływań odpychających, a E_{LR} (o wadze $w_{LR}=2,0$) to energia długozasięgowych oddziaływań parami, obliczona po zsumowaniu wszystkich oddziaływań par.



Rysunek 1.3: Gruboziarnisty model CABS przedstawiony za pomocą czterech atomów (lub pseudoatomów): $C\alpha$ (czerwone kule), $C\beta$ (brązowe kule), środek masy grupy łańcucha bocznego SC (pomarańczowe kule) i środek wiązania peptydowego cp (jasnoniebieskie kule).

Model CABS wykorzystuje się w symulacjach dynamiki białek uzyskanych z zespołów NMR¹⁴¹, do badania interakcji między białkami, a także do przewidywania elastyczności białek^{138–140}. Jest istotnym elementem systemu CABS-DOCK, który obejmuje również dokowanie peptydów¹⁴⁰.

W regionach uznawanych za sztywne, takich jak odpowiadające dobrze zdefiniowanym elementom struktury drugorzędnej¹⁴², CABS-flex może wykrywać dynamiczne, niejasne zmiany, które mogą być istotne biologicznie, w przeciwieństwie do predyktorów fluktuacji opartych na sekwencji. Można znaleźć ruchy, które są ważne dla funkcjonowania białek, wykorzystując uzyskane profile fluktu-

acji. Najbardziej ruchliwe fragmenty strukturalne mogą być potencjalnymi miejscami dokowania molekularnego¹⁴³.

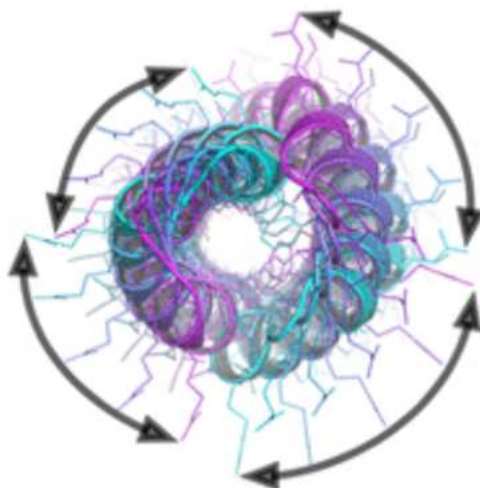
1.5.3 Model NOLB

Model NOLB³³ przedstawiony schematycznie na rysunku 1.4 jest kolejnym przykładem modelu CG. Opiera się on na technice analizy modów normalnych (NMA)³³. Jest to model sieci elastycznej, w którym energia układu jest wyrażona równaniem 1.8:

$$E_{TOT} = \sum_{\substack{i < j \\ d_{ij} < d_{cut}}} \frac{\gamma}{2} (d_{ij}(\mathbf{q}) - d_{ij}^0)^2 \quad (1.8)$$

Gdzie $\mathbf{q}$ jest wektorem współrzędnych uogólnionych, $d_{ij}(\mathbf{q})$ jest odległością pomiędzy i-tym oraz j-tym atomem, d_{ij}^0 jest odległością w strukturze odniesienia, zaś γ jest stałą siłową.

Korzystając z tego wzoru wyznacza się macierz drugich pochodnych energii względem współrzędnych $\mathbf{q}$, do uzyskania modów normalnych.



Rysunek 1.4: Ilustracja ruchów bloków cząsteczki białka uzyskanych poprzez użycie metody NOLB na przykładzie domeny cytoplazmatycznej chemoreceptora bakteryjnego z *Thermotoga maritima*, nieliniowej ekstrapolacji ruchu białka (o kodzie PDB: 2CH7¹⁴⁴) w formie zwiniętego kłębka. Przedstawia ruchy skręcania konformacji białka przy różnych amplitudach odkształceń nałożonych na siebie. Kolorami oznaczono wartości całkowitego odkształcenia (średnich odchylenia kwadratowego). Strzałki określają trajektorie poszczególnych atomów. "Reprinted with permission from Alexandre Hoffmann, Sergei Grudinin, NOLB: Nonlinear Rigid Block Normal-Mode Analysis Method, J. Chem. Theory Comput. 2017, 13, 5, 2123–2134. Copyright 2017 American Chemical Society."

Z uwagi na to, że najbardziej interesujące są ruchy części białka, które nie zmieniają swojej wewnętrznej geometrii (bloków), w tym modelu zbiór współrzędnych został zawężony do położen centrów bloków oraz współrzędnych opisujących obroty tych bloków względem ich centrów (środków mas).

Metoda NOLB opiera się na nieliniowej ekstrapolacji chwilowych kierunków ruchu opisanych przez mody normalne obliczone w podprzestrzeni RTB (z ang. rotations-translations of blocks, obroty translacji bloków)³³. W obliczeniu tych modów normalnych wykorzystuje się diagonalizację macierzy sztywności (ważonej masą) rzutowanej przez RTB.

$$P^T K_W P = \tilde{L} \tilde{\Lambda} \tilde{L} \quad (1.9)$$

Gdzie $\tilde{L}$ jest macierzą składającą się z normalnych trybów RTB i macierzą wartości własnych diagonalnych $\tilde{\Lambda}$, które są odpowiadające im, a P to macierzy projekcji. To porównanie można zapisać jako (równanie 1.10):

$$K_W = (P \tilde{L}) \tilde{\Lambda} (P \tilde{L})^T \quad (1.10)$$

Dzięki temu wszystkie atomy L^w (we współrzędnych ważonych masą) można uzyskać jako projekcję normalnych modów RTB $\tilde{L}$, zgodnie z równaniem 1.11:

$$L^W = P \tilde{L} \quad (1.11)$$

To równanie zostało wykorzystane w pierwotnej metodzie RTB do obliczenia liniowej deformacji początkowej struktury wzdłuż kierunków L^w . Niemniej jednak do obliczenia bardziej naturalnej deformacji nieliniowej należy użyć macierzy projekcji P .

Kontynuując, należy pamiętać, że rozmiar wektora trybu normalnego RTB L_i^w wynosi $6n$, gdzie n jest liczbą sztywnych bloków. W tym wektorze każda z sześciu dodatkowych współrzędnych odpowiada konkretnemu sztywnemu blokowi. Pierwsze trzy współrzędne zapewniają chwilowe przemieszczenie sztywnego bloku COM (z ang. block's center of mass, środek masy bloku), a kolejne

trzy współrzędne zapewniają chwilową oś obrotu sztywnego bloku, która przechodzi przez jego COM. Chwilowe prędkości liniowe i kątowne poszczególnych sztywnych bloków można zrozumieć, wprowadzając zależność czasową zmiennych i biorąc pochodną czasową wektora trybu normalnego RTB. Zaczynając od chwilowych prędkości liniowych i kątowych sztywnych bloków, ostatni etap tej metody obejmuje ekstrapolację ruchu bloków o dużej amplitudzie³³.

Metoda NMA jest stosowana do badań wielu układów biomolekularnych, a przede wszystkich białek³³, RNA i DNA³³, a także do badań wiązania pomiędzy białek a ligandem¹⁴⁵, interakcji pomiędzy białkami¹⁴⁶ i związania białek¹⁴⁷.

1.5.4 Model MARTINI

Ostatnim przykładem modelu CG w tej pracy jest MARTINI¹⁴⁸ (Rysunek 1.5), który początkowo został opracowany w celu zbadania właściwości błon lipidowych^{149,150}.

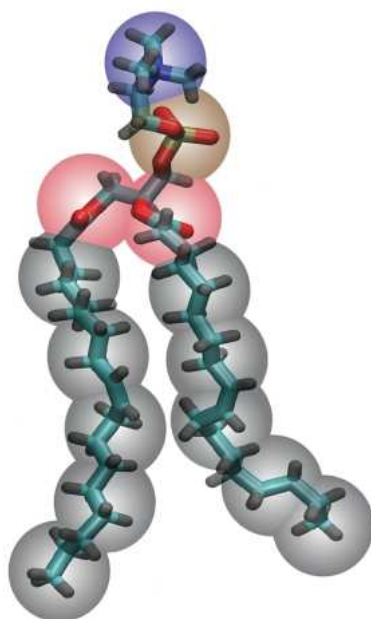
W modelu MARTINI pojedyncze centrum interakcji reprezentuje cztery ciężkie atomy wraz z związanymi wodorami¹⁵¹. Mapowanie cztery do jednego zostało wybrane jako najlepszy sposób na połączenie wydajności obliczeniowej z zachowaniem właściwości chemicznych charakteru centrum. Ponieważ cztery cząsteczki wody są mapowane na kulkę wody w CG, ten wybór jest zgodny z mapowaniem. Jedna kulka CG reprezentująca zarówno jon, jak i pierwszą powłokę hydratacyjną. Ogólne podejście mapowania cztery do jednego jest zbyt ogólne, aby pokazać geometrię małych fragmentów lub cząsteczek przypominających pierścien, takich jak cholesterol, benzen i kilka aminokwasów¹⁵¹. W rezultacie cząsteczki przypominające pierścien są mapowane na jedną cząsteczkę MARTINI przy użyciu dwóch atomów, które nie są atomami wodoru.

W tym modelu oddziaływania daleko-zasięgowe zostały sparametryzowane tak, aby odtwarzały eksperymentalne wartości energii swobodnej przejścia pomiędzy fazą polarną i niepolarną dużej liczby związków chemicznych. Ponadto dane z symulacji pełnoatomowych były wykorzystywane do uzyskania wkładów lokalnych do efektywnej energii oddziaływania. Model MARTINI można zatem zakwalifikować do obu podejść top-down i bottom-up. Zatem może tworzyć użyteczny pomost między skalami makroskopowymi i atomistycznymi¹⁵².

Efektywna funkcja energii została wyrażona równaniem 1.12:

$$E = E_{bond} + E_{bend} + E_{tors} + E_{LJ} + E_{ele} \quad (1.12)$$

Gdzie E_{bond} to potencjał harmoniczny rozciągania wirtualnego wiązania, E_{bend} to potencjał harmoniczny zginania wirtualnego kąta walencyjnego, E_{tors} to multimodalny potencjał torsyjny, E_{LJ} to potencjał Lennarda-Jonesa 6-12 dla niezwiązanych interakcji, a E_{ele} to kulombowski wkład do energii potencjalnej.



Rysunek 1.5: Gruboziarnisty model MARTINI przedstawiający strukturę lipidu. Reprezentacja wszystkich atomów jest pokazana za pomocą małych patyczków, podczas gdy reprezentacja gruboziarnista jest przedstawiona w postaci dużych kul. "Reprinted with permission (CC-BY) from Ryan Bradley and Ravi Radhakrishnan, Coarse-Grained Models for Protein-Cell Membrane Interactions, Polymers 2013, 5, 890-936. Copyright 2013 Polymers."

Wyróżnia się cztery kategorie centrów oddziaływań w modelu MARTINI: polarne (P), niepolarne (N), apolarne (C) i naładowane (Q)¹⁴⁸. Każda z klas Q i N ma cztery podtypy, które są powiązane z ich zdolnością do udziału w tworzeniu wiązania wodorowego: donor i akceptor, donor, akceptor oraz ich brak. Główną różnicą między tymi podtypami jest ich siła interakcji ze sobą¹⁴⁸. Każda z klas P i C obejmuje pięć podtypów, które odzwierciedlają stopniowe przejście od słabych do silnych cech polarnych bądź niepolarnych¹⁴⁸. MARTINI wykorzystuje związane i niezwiązane formy funkcyjne potencjału, które są powszechnie stosowane w pełnoatomowych polach sił. To sprawia, że

model jest łatwy do wprowadzenia w nowoczesnych programach dynamiki molekularnej, takich jak GROMACS¹²⁷, GROMOS¹⁵³ i NAMD^{154,155}.

Model ten pozwala określić zachowanie dużych agregatów lipidowych w przestrzeni i skalach czasowych nieosiągalnych dla pełnoatomowych symulacji MD, przy jednoczesnym zachowaniu wystarczającej rozdzielczości i specyficzności chemicznej, aby uzyskać mikroskopijny i dynamiczny obraz wciąż niedostępny w eksperymentach. Wykazano, że pole siłowe MARTINI można zastosować do wielu procesów lipidowych, takich jak tworzenie fazy żelowej i ciekłej¹⁵⁶. Zastosowanie pola siłowego z biegiem lat rozszerzyło się do najbardziej powszechnych biomolekuł, takich jak białka^{157,158}, cukry^{118,159}, nukleotydy^{160,161} i kilka ważnych kofaktorów¹⁶², a także wiele cząsteczek niebiologicznych, w tym polimery syntetyczne^{163–165} i nanocząstki¹⁶⁶. Wykorzystano model MARTINI do przewidywania różnych procesów biomolekularnych, w tym wiązania białko-lipid¹⁶⁷. W niektórych przypadkach ta metoda została również wykorzystana do wiązania lipidów do miejsc głęboko osadzonych w białkach^{168,169}. Niemniej jednak przykłady wiązania białko-ligand z użyciem pola MARTINI są nadal rzadkie^{170,171}.

1.6 Pełnoatomowe pola siłowe

W przeciwieństwie do gruboziarnistych pól siłowych, pełnoatomowe są złożonymi modelami, w których każdy atom jest reprezentowany jako osobne centrum oddziaływania. Początkowe prędkości przypisane są do każdego atomu losowo na początku trajektorii symulacji. Aby uniknąć błędów w całkowaniu tych równań, krok czasu integracji jest zwykle ustalany na 1 do 2 fs dla systemów biologicznych z ograniczeniem na wodory, gdzie 2 fs to przybliżona skala czasowa najszybszego ruchu cząsteczki (drgania wiązania)⁸². Celem zastosowania tego rodzaju pól jest badanie struktury molekularnej, dynamiki, a także właściwości układów na poziomie atomowym.

Pełnoatomowe pola siłowe wykorzystują parametry z danych eksperymentalnych i obliczeń mechaniki kwantowej do dokładnego modelowania interakcji między atomami. Wśród pól siłowych obejmujących wszystkie atomy można wyróżnić m. in. CHARMM¹⁷², który jest często używany do symulacji lipidów¹⁷³, kwasów nukleinowych¹⁷⁴ i innych makrocząsteczek oraz¹⁷⁵ oraz AMBER¹⁷⁶, który jest często używany do symulacji małych cząsteczek¹⁷⁷, kwasów nukleinowych i białek¹⁷⁸.

Powstało wiele prac porównujących oba te pola^{179–181}. W pracy prof. Zhang i inni¹⁸², moc pre-

dykcyjna tych dwóch pól sił została porównana, w której zbadano wpływ zmian konformacyjnych reszty alaniny w monomerycznym GAGu, ale do tej pory nie przeprowadzono podobnych badań dotyczących struktur białko-GAG.

1.6.1 Pole siłowe AMBER

AMBER¹⁷⁶ (z ang. Assisted Model Building and Energy Refinement) to zestaw molekularnych pól siłowych, które są wykorzystywane do modelowania biocząsteczek. Podstawowe pole siłowe zostało zaimplementowane za pomocą równania 1.13¹⁷⁶ do obliczenia funkcji energii potencjalnej, które zachowuje najprostrze funkcjonalne formy cząsteczek w skondensowanych fazach.

$$\begin{aligned}
V(r) = & \sum_{\text{wiązania}} K_b(b - b_0)^2 + \sum_{\text{kąty walencyjne}} K_\theta(\theta - \theta_0)^2 \\
& + \sum_{\text{kąty dwuściennne}} \left(\frac{V_n}{2}\right)(1 + \cos[n\phi - \delta]) \\
& + \sum_{\text{oddziaływania niewiążące}} (A_{ij}/r_{ij}^{12}) - (B_{ij}/r_{ij}^6) + (q_i q_j / r_{ij})
\end{aligned} \tag{1.13}$$

Gdzie b_0 oraz θ_0 to odpowiednio referencyjna długość wiązania i kąt wiązania; K_b , K_θ oraz V_n są stałymi siłowymi, n to krotność, ϕ to kąt torsyjny, a δ jest kątem fazowym, b , θ oznaczają odpowiednio długości wiązań, kąty walencyjne, A_{ij} oraz B_{ij} to odpowiednie parametry do potencjału Lennarda-Jonesa, q_i i q_j to odpowiednie cząstkowe ładunki na atomach i oraz j oraz r_{ij} jest wektorem określającym odległość między tymi atomami. Sumowanie odbywa się po wkładach długości wiązań reprezentuje energię między atomami o wiązaniach kowalencyjnych, po wkładach kątów wiązań przedstawia energię wynikającą z geometrii orbitali elektronowych biorących udział w wiązaniu kowalencyjnym, po wkładach kątów dwuściennych (z ang. torsions) przedstawia energię kątów dwuściennych wiązania ze względu na kolejność wiązania (np. wiązania podwójne) i sąsiednie wiązania lub wolne pary elektronów, zaś ostatni człon reprezentuje energię niewiążącą pomiędzy wszystkimi parami atomów¹⁷⁶.

Do badań kwasów nukleinowych oraz białek zalecane było stosowania pola siłowego ff14SB¹⁸³, który powstał na podstawie pola ff99SB¹⁸⁴ wykazujący słabości w preferencjach drugorzędnej struktury rotameru łańcucha bocznego i szkieletu¹⁸⁵. W porównaniu z ff99SB, model pola ff14SB popra-

wia odtwarzanie eksperymentalnych geometrii. AMBER ff14SB oferuje niezwykle stabilną dynamikę szkieletu zwiniętych białek i poprawia stabilność helikalną, co czyni go lepszym niż poprzednie pola siłowe AMBER¹⁸⁵. Pole siłowe ff14SB wymaga zastosowania modelu wody TIP3P, ponieważ ten 3 punktowy model wody był najczęściej używany do polepszania wcześniejszych pól siłowych pakietu AMBER¹⁸⁶. Najnowsze pole siłowe ff19SB¹⁸⁷ zostało zoptymalizowane do badań białek^{188,189}. Właściwości pola ff19SB odzwierciedlają różnice w mapie Ramachandrama dla aminokwasów i najlepiej łączy się z modelem wody OPC¹⁸⁷. GLYCAM06j¹⁹⁰ to pole siłowe z parametrami do modelowania glikokoniugatów i węglowodanów. Opiera się na podstawie zastosowań GLYCAM93^{191,192}. Wiązania, stałe siłowe odkształcenia kątów walencyjnych i właściwości elektrostatyczne w najnowszej wersji GLYCAM06j zostały uzyskane za pomocą obliczeń QM, ponieważ parametry te są trudne do uzyskania eksperymentalnie¹⁹⁰. Pole siłowe GLYCAM06j sprawdza się w przypadku odtworzenia właściwości fazy gazowej małych cząsteczek testowych¹⁹⁰, a ponadto odtwarza populacje rotamerów małych cząsteczek i reprezentacyjnych bloków budulcowych biopolimerów w wodzie w formie niejawnej, oraz właściwości sieci krystalicznej¹⁹⁰. Pole siłowe AMBER (ff14SB) w połączeniu z GLYCAM jest stosowane w badaniach GAGów, czego przykładem jest badanie interakcji GAGów z prokaepsyną S podczas dojrzewania, analizy dynamiki, wiązania energii swobodnej i potencjalnej regulacji allosterycznej¹⁹³.

1.6.2 Pole siłowe CHARMM

CHARMM (z ang. Chemistry at HARvard Macromolecular Mechanics) jest to powszechnie używany zestaw pełnoatomowych pól siłowych dynamiki molekularnej, którego funkcja energii potencjalnej została wyrażona równaniem 1.14¹⁷².

$$\begin{aligned}
V = & \sum_{\text{wiązania}} K_b(b - b_0)^2 + \sum_{\text{kąty walencyjne}} K_\theta(\theta - \theta_0)^2 + \sum_{\text{niewłaściwe kąty torsyjne}} K_\varphi(\varphi - \varphi_0)^2 \\
& + \sum_{\text{kąty dwuścienne}} \sum_{n=1}^6 K_{\phi n}(1 + \cos(\phi_n - \delta_n)) + \sum_{\text{Urey-Bradley}} K_u(u - u_0)^2 \\
& + \sum_{\text{oddziaływania elektrostatyczne}} \frac{q_i q_j}{4\pi D r_{ij}} \\
& + \sum_{\text{oddziaływania van der Waalsa}} \varepsilon_{ij} \left[\left(\frac{R_{min,ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{min,ij}}{r_{ij}} \right)^6 \right] \tag{1.14}
\end{aligned}$$

Gdzie b_0 , θ_0 , φ_0 , to referencyjna długości wiązań, kąty walencyjne, niewłaściwe kąty torsyjne; b , θ , φ , ϕ , u oraz u_0 oznaczają odpowiednio długości wiązań, kąty walencyjne, kąty torsyjne, kąty dwuścienne wyznaczone z geometrii molekularnej oraz wkład równowagi Ureya-Bradleya (potencjał harmoniczny na odległości między końcowymi atomami zaangażowanymi w kąt walencyjny (1,3)), a K_b , K_θ , K_φ oraz K_u są odpowiednimi stałymi siły, $K_{\phi,n}$ to amplituda, a ϕ_n to faza n -tej krotności, D jest stałą dielektryczną oraz q_i i q_j to odpowiednie częściowe ładunki atomowe na atomach i oraz j . ε_{ij} to głębokość studni potencjału, a $R_{min,ij}$ to promień w potencjale Lennarda-Jonesa 6-12 stosowany do opisu oddziaływań van der Waalsa, r_{ij} to odległość między atomami i oraz j .

Pełnoatomowe pole siłowe CHARMM jest powszechnie wykorzystywane w modelowaniu molekularnym i symulacjach molekularnych, aby tworzyć złożone, jak i nieuporządkowane zespoły konformacyjne szkieletu polipeptydowego dla wewnętrznie nieuporządkowanych peptydów i białek (IDP)¹⁹⁴. CHARMM36m dostosowuje cząstkowe ładunki atomowe i dopasowuje je do interakcji między substancją rozpuszczoną (wodą) a fragmentami węglowodanów, które zostały obliczone za pomocą metod mechaniki kwantowej¹⁹⁵. W jednym z badań nad GAGami porównano dynamikę molekularną (MD) Amber ff14SB i CHARMM36m w celu zbadania populacji dimeru HP36, stabilności i kinetyki tworzenia dimeru¹⁹⁶. Wykazano, że CHARMM36m dobrze reprezentuje eksperymentalne populacje monomerów HP36 przy niższych stężeniach. W tym porównaniu z AMBER ff14SB, to CHARMM36m udowadnia, że daje niższe współczynniki asocjacji (wyższe populacje monomerów HP36).

1.7 Potencjał średnich sił

Profil energii swobodnej wzdłuż wybranej współrzędnej nazywany jest potencjałem średniej siły (PMF)¹⁹⁷, który wyznacza się za pomocą równania 1.15.

$$G(\xi) = -k_B T \ln [p(\xi)] \quad (1.15)$$

Gdzie k_B to stała Boltzmanna, T oznacza temperaturę, $p(\xi)$ jest rozkładem prawdopodobieństwa względem wybranej współrzędnej. Współrzędną może być kąt torsyjny, czy odległość między dwoma atomami.

Badania obliczeniowe często wykorzystują potencjały średniej siły do opisanie różnych procesów, takich jak oddziaływania hydrofobowe¹⁹⁸, transfer protonów¹⁹⁹, czy też przenikanie jonów przez kanały błonowe²⁰⁰. Jedną z właściwości potencjału średniej siły, jest to, że pozwala określić parametry fizykochemiczne badanych układów. Właściwości fizykochemiczne, takie jak stałe asocjacji, można mierzyć za pomocą PMF. Jest to metoda, która może również być użyta do określenia bezwzględnej energii swobodnej wiązania dla układów białko-ligand²⁰¹. Następnie, w oparciu o dane eksperymentalne, można przeprowadzić walidację zastosowanych pól siłowych. Określanie PMF jest przydatne, kiedy bezpośredni eksperyment jest trudny do wykonania (np. badania asocjacji hydrofobowej) lub gdy chcemy zmniejszyć koszty badań²⁰². Symulacje dynamiki molekularnej lub mechaniki kwantowej są używane do uzyskania PMF, aby określić, jak zmienia się energia układu w funkcji dobrze zdefiniowanej współrzędnej reakcji, reprezentowaną przez jedną lub więcej zmiennych^{203,204}.

1.8 Metody półempiryczne

Metody półempiryczne zostały opracowane w celu uproszczenia metody Hartre–Focka (HF). Takie metody mogą znacznie skrócić czas obliczeń, zastępując pewne całki składające się na procedurę HF parametrami empirycznymi i eksperymentalnymi. Naukowcy początkowo opracowali rodzinę półempirycznych metod wywodzących się z nakładania się różniczkowego pomijania dwuatomowego (NDDO)^{205,206}, ale później Dewar i Thiel zmodyfikowali metodę pośredniego zaniedbania różniczkowego nakładania się (MNDO)^{207,208}. W późniejszym czasie wprowadzono wiele modyfikacji

tego podstawowego schematu NDDO, takich jak AM1²⁰⁹, PM3^{210,211}, PM6²¹² itd.

W związku z tym, że metody stawały się coraz bardziej zaawansowane, każda z nowych metod wykazywała poprawę opisu energii i geometrii układu w porównaniu z poprzednimi. Na przykład w PM6, średni błąd bezwzględny energii interakcji dla prostych związków organicznych spadł o około 30% w porównaniu z PM3. Te zalety sprawiły, że każda nowa metoda do modelowania układów chemicznych była bardziej atrakcyjna niż jej poprzednie wersje²¹³.

PM7²¹³, najnowsza wersja tej rodziny metod półempirycznych, ma większą dokładność i większy zakres zastosowań w porównaniu do swoich poprzedników²¹³. Mając na celu poprawę przestrzeni parametrów, PM7 została sparametryzowana przy użyciu eksperymentalnych i wysokopoziomowych danych referencyjnych *ab initio*, a następnie rozszerzona o nowy typ danych referencyjnych. Modele struktur krystalicznych i ciepła tworzenia ciał stałych zostały wykorzystane do testowania nowo opracowanej techniki. W porównaniu z PM6 do zestawu przybliżeń wprowadzono dwie zmiany; poprawa opisu oddziaływań niekowalencyjnych oraz poprawa dwóch drobnych błędów w formalizmie NDDO. Przy przewidywaniu geometrii i entalpii tworzenia układów organicznych w fazie gazowej metoda PM7 daje średnie błędy o 5% i 10% mniejsze niż w przypadku PM6. To znaczące ulepszenie czyni metodę PM7 wiarygodną opcją do obliczania tych właściwości fizykochemicznych w chemii organicznej²¹⁴.

Rozdział 2

Cel pracy

Elastyczność, która umożliwia określonym molekułom adaptację, reakcję lub interakcję z innymi substancjami na wiele sposobów, w różnych warunkach, jest niezbędna do zrozumienia funkcjonowania biomolekuł. Celem mojej pracy było określenie elastyczności białek, białek otoczonych błoną lipidową oraz zbadanie wpływ GAGów na elastyczność białek. Do tego celu wykorzystałem teoretyczne podejścia wykorzystujące gruboziarniste oraz pełnoatomowe pola siłowe do zbadania elastyczności za pomocą średnich standardowych odchyłeń fluktuacji.

W pierwszej części tej pracy zbadalem wiarygodności przewidywań elastyczności białek wykonanych za pomocą czterech różnych metod gruboziarnistych. Następnie określiłem, która metoda najlepiej sprawdza się w przewidywaniu elastyczności struktur białek pochodzących z zespołów NMR oraz X-ray, oraz występującego typu struktury drugorzędowej, na podstawie współczynników korelacji pomiędzy teoretycznymi a eksperymentalnymi profilami fluktuacji. Następnie sprawdziłem, od czego są zależne uzyskane współczynniki korelacji przeprowadzając dwukierunkową analizę wariancji oraz testy Studenta.

W następnym kroku przeprowadziłem analizę rozkładu współczynników korelacji względem liczby zliczeń, skośność układów oraz dystrybuanty rozkładów, aby określić asymetrię rozkładów, a także określić różnice między rozkładami, wyznaczając tym samym najlepszą metodę do badania elastyczności białek w zależności od eksperymentalnej metody określania struktury oraz występującej struktury drugorzędowej. Następnie sprawdziłem, czy uzyskane wartości współczynników korelacji mogą zależeć od długości łańcucha białkowego. Zestawiłem ze sobą uzyskane profile fluktuacji dla wybranych białek, aby zobrazować różnice w przewidywaniu elastyczności białek i wytypować najlepszą, uniwersalną metodę.

W dalszej części rozszerzyłem pole siłowe UNRES wprowadzając błony lipidowej w formie jawnej (z ang. explicit) używając modelu MARTINI. W tym celu obliczyłem PES między centrami UNRES-MARTINI, wyznaczyłem PMF, dopasowałem funkcje analityczne, które wprowadziłem do modelu UNRES. W przypadku białek otoczonych błoną lipidową określiłem, w jaki sposób metoda UNRES-flex potrafi przewidywać ich elastyczność przy użyciu nowego pola siłowego z błoną lipidową.

W ostatnim części tej pracy określiłem, które z zastosowanych pełnoatomowych pól siłowych, ff14SB/GLYCAM06j z pakietu AMBER oraz CHARMM36m, najlepiej odwzorowuje elastyczność białek w formie związanej (kompleksu) oraz niezwiązanej z GAGami. Określiłem to na podstawie porównania profili fluktuacji uzyskanych teoretycznie z danymi eksperymentalnymi.

Rozdział 3

Metody

3.1 UNRES oraz UNRES z funkcją więzów na strukturę drugorzędową

W moich badaniach zastosowałem model UNRES, przeprowadziłem kanoniczne symulacje Langevina MD, do przewidywania fluktuacji białek. Wykonałem krótkie symulacje MD, z krokiem czasowym 4,89 fs, z łączną liczbą kroków 200 000, co daje długość trajektorii około 1 ns . Jednak każda symulacja faktycznie odpowiadała μs czasu laboratoryjnego, ponieważ jednostka czasu UNRES wynosi około 1000 jednostek czasu laboratoryjnego¹²⁹. W symulacjach dynamiki Langevina termostat ustawiłem na temperaturę 300 K, a tarcie wody skalowałem o współczynnik 0,01, podobnie jak w pracy M. Khalili i inni¹³³.

Przeprowadziłem trzy oddzielne symulacje MD dla każdego z zestawu stu wybranych białek. Określiłem tę metodę jako "UNRES-flex", która została wykorzystana do symulacji przeprowadzonych w dalszej części pracy. Przeprowadziłem również symulacje MD modelu UNRES z DSSP przy tych samych ustawieniach, co standardowe symulacje UNRES MD. Jako "UNRES-DSSP-flex" określiłem metodę do przewidywania elastyczności białka z funkcją definiowania struktur drugorzędowych.

Ograniczenia struktury drugorzędowej zostały nałożone na wybrane kąty, ustalone przez DSSP²¹⁵ opierając się na wiązaniach wodorowych łańcucha głównego, zgodnie z definicją zawartą w modelu elektrostatycznym²¹⁶. Nałożone ograniczenia były między wirtualnym wiązaniem szkieletu $C\alpha \cdots C\alpha \cdots C\alpha \cdots C\alpha$. Zastosowałem ograniczniki kwartalne o płaskim dnie ze stałą siłą równą 50 kcal/mol/rad⁴ z zakresem płaskiego dna czwartego rzędu wynoszącym $50^\circ \pm 20^\circ$ dla regionów α -helikalnych i $180^\circ \pm 40^\circ$ dla regionów β -harmonijki.

3.2 CABS-flex

Model CABS-flex¹³⁷ wykorzystałem do przeprowadzenia symulacji w domyślnych ustawieniach programu. Więzy zostały wygenerowane dla par reszt („tryb ss2”), dla których przypisuje regularną strukturę drugorzędową (helikalną lub harmonijkową) w odległości 3,8Å a 8,0Å pomiędzy węglami C α ... C α (opcja „gap3”). Autorzy zalecają ustawić temperaturę na 1,4, ponieważ wartość 1,0 jest zwykle porównywalna z temperaturą kryształu, a wartość 2,0 zazwyczaj powoduje całkowite rozwinięcie małych łańcuchów białkowych pozbawionych więzów¹³⁸. Przeprowadziłem symulacje dla zestawu 100 wybranych białek w trzech niezależnych symulacjach Monte Carlo.

3.3 NOLB

Zastosowanie modelu NOLB różni się w porównaniu do reszty modeli, ponieważ jest on deterministyczny, dlatego przeprowadziłem pojedyncze symulacje. Działanie algorytmu zatrzymuje się, gdy liczba iteracji przekracza 100 (wartość domyślna) lub względne odkształcenie jest mniejsze niż 1×10^{-6} . W swoich obliczeniach NOLB wykorzystuje schemat iteracyjny. Każda iteracja tworzy nowy model sieci sprężystej, który odzwierciedla ostatnią przewidywaną konformację, a następnie generuje nowe mody normalne, tworząc dziesięć modów o najniższej częstotliwości względem struktury wyjściowej.

3.4 Wprowadzenie błony w formie jawnej do modelu UNRES

3.4.1 Obliczanie hiperpowierzchni energii potencjalnej

Do określenia siły oddziaływania układu zbudowanego z cząsteczek modelu UNRES i cząsteczek modelu MARTINI wyznaczyłem potencjały średniej siły, gdzie powierzchnie energii potencjalnej obliczyłem na siatce. Gęstość siatki ustaliłem dla r' od 1Å do 13Å z krokiem 0,2Å, β od 15° do 165° przy $\Delta\beta = 30^\circ$ oraz γ od 0° do 300° przy $\Delta\gamma = 60^\circ$ (rysunek 3.1). Na podstawie powierzchni energii potencjalnej (PES), obliczonych za pomocą półempirycznej metody mechaniki kwantowej PM7²¹³ oraz modelu wody SMD²¹⁷, wyznaczyłem potencjały średniej siły między centrami interakcji UNRES oraz MARTINI.

Symulacje wykorzystujące półempiryczną metodę PM7 i symulacje dynamiki molekularnej AM-

BER zostały przeze mnie wykorzystane do uzyskania PES dla oddziaływań UNRES-MARTINI. Metoda PM7 jest stosowana z modelem wody SMD, w którym rozpuszczalnik w formie niejawnej jest przedstawiany jako medium dielektryczne z napięciem powierzchniowym na granicy substancji rozpuszczonej-rozpuszczalnika²¹⁷.

W związku z tym, że wcześniejsze potencjały modelu UNRES otrzymywane były poprzez symulacje wykorzystujące wszystkie atomy²¹⁸, postanowiłem otrzymać profile PMF za pomocy metody pełnoatomowej. W kolejnym kroku przeprowadziłem symulacji pomiędzy centrami oddziaływań cząsteczek uzyskanych modelem UNRES oraz cząsteczek uzyskanych modelem UNRES dla różnych odległości, za pomocą pełnoatomowego pola siłowego ff14SB z pakietu AMBER16¹⁸⁷ do obliczeń PMF, na podstawie których wyprowadziłem wkłady energetyczne dla oddziaływań pomiędzy centrum UNRES oraz MARTINI.

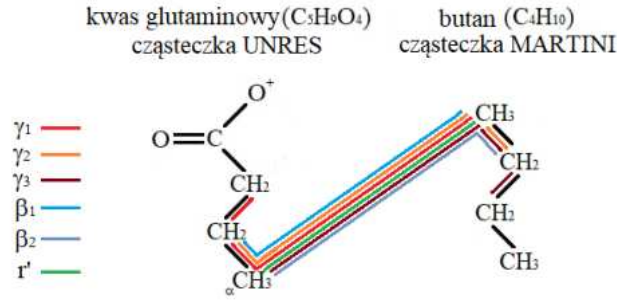
Każdy z badanych układów wprowadziłem do pudełka periodycznego, w którym cząsteczki wody zostały przedstawione za pomocą modelu wody TIP3P²¹⁹, a także umieściłem przeciwjon Na^+ lub Cl^- , dzięki któremu możliwe jest utrzymanie elektroobojętności. Dla każdego badanego układu systemu przeprowadziłem symulacje próbkowania parasolowego z wykorzystaniem potencjału harmonicznego do regulowania odległości między atomami najbliższymi centrów mas danej cząstki, gdzie próbkowanie nastąpiło w odległościach od 3Å do 15Å, w odstępach co 1Å. Przeprowadziłem minimalizację struktur startowych do 15 Å pomiędzy centrami interakcji, kolejnie ogrzewałem układ do temperatury 300K w warunkach NVT (25ps), następnie wykonałem równowagowanie układu dla wszystkich odległości w warunkach NPT (800ps), zaś symulacje dynamiki molekularnej dla każdej odległości przeprowadziłem w warunkach NPT (1500ps).

3.4.2 Potencjały średniej siły

PMF obliczyłem na podstawie powierzchni energii potencjalnej PM7 stosując równanie 3.1:

$$F(\alpha_i, r_j) = -RT \ln \sum_{r'} \sum_{\beta_1} \sum_{\beta_2} \sum_{\gamma_1} \sum_{\gamma_2} \sum_{\gamma_3} e^{-\frac{1}{RT} \varepsilon(\alpha, r)} \quad (3.1)$$

Gdzie r' to odległość pomiędzy pierwszym ciężkim atomem cząsteczki MARTINI i atomem Ca modelu UNRES, γ_1 jest kątem dwuściennym pomiędzy pierwszym atomem centrum MARTINI oraz



Rysunek 3.1: Zobrazowanie zmiennych użytych do obliczenia potencjału średnich sił na przykładzie łańcucha bocznego kwasu glutaminowego (cząsteczka UNRES) oraz butanu (cząsteczka MARTINI).

atomami $C\alpha$, $C\beta$ i $C\gamma$ centrum UNRES, γ_2 to kąt dwuścienny między pierwszym i drugim ciężkim atomem centrum MARTINI oraz $C\alpha$ i $C\beta$ centrum UNRES, γ_3 to kąt dwuścienny pomiędzy pierwszym, drugim i trzecim atomem centrum MARTINI oraz atomem $C\alpha$ centrum UNRES, β_1 oraz β_2 oznaczają kąty płaskie pomiędzy $C\alpha$ i $C\beta$ oraz pierwszym atomem centrum MARTINI, a także pierwszym i drugim ciężkim atomem centrum MARTINI w stosunku do odpowiedniego atomu $C\alpha$ centrum UNRES (określone zmienne przestawiłem na rysunku 3.1). Wartości α oraz r wyznacza się za pomocą równania 3.2 oraz 3.3:

$$\alpha_i - \frac{\Delta\alpha}{2} \leq \alpha < \alpha_i + \frac{\Delta\alpha}{2} \quad (3.2)$$

$$r_j - \frac{\Delta r}{2} \leq r < r_j + \frac{\Delta r}{2} \quad (3.3)$$

Gdzie α to kąt utworzony przez atom $C\alpha$ między środkiem masy UNRES a środkiem masy MARTINI, zaś r to odległość między środkami mas dwóch cząsteczek. Zdefiniowałem szerokości przedziałów dla zmiennych współczynników r i α na $0,2 \text{ \AA}$ i 15° .

Przekształciłem PMF w taki sposób, żeby najmniejsza wartość była równa zeru, co wyraziłem za pomocą równania 3.4:

$$F(\alpha_i, r_j)_{min} = 0 \quad (3.4)$$

W przypadku symulacji pełnoatomowych zastosowałem metodę analizy ważonych histogramów (WHAM)²²⁰ do wyznaczenia potencjałów średniej siły badanych układów, która pozwoliła usunąć

wkłady energetyczne pochodzące z wiezów nałożonych podczas próbkowania przestrzeni konformacyjnej badanych układów.

3.4.3 Dopasowanie funkcji analitycznych

Po uzyskaniu PMF w następnej kolejności dopasowałem odpowiednie funkcje analityczne opisane w tej sekcji. Jest to konieczne, ponieważ w polu siłowym UNRES jest przeprowadzana dynamika molekularna, która wymaga, aby funkcja energii była opisywana za pomocą różniczkowalnych funkcji. Dodatkowym efektem jest wygładzenie hiperpowierzchni energii potencjalne, co ułatwia ich interpretację.

W serii publikacji prof. Makowskiego^{221–223} zostały określone wzory matematyczne stosowane w przypadku PMF dla par hydrofobowych substancji rozpuszczonych w środowisku wodnym, które zastosowałem w swoich badaniach. PMF dla rozpuszczonej substancji niepolarnej (n), a także innych centrów (X) znajdujących się w wodzie²²⁴, który został opisany za pomocą równania 3.5:

$$W_{nX} = E_{GBerne} + \Delta F_{cav} \quad (3.5)$$

Gdzie E_{GBerne} jest składnikiem energii van der Waalsa. Aby obliczyć wkład E_{GBerne} , wykorzystuje się potencjał Gay-Berna (równanie 3.6)²²⁵.

$$E_{VDW} = E_{GBerne} = 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}^0}{r_{ij} - \sigma_{ij} + \sigma_{ij}^0} \right)^{12} - \left(\frac{\sigma_{ij}^0}{r_{ij} - \sigma_{ij} + \sigma_{ij}^0} \right)^6 \right] \quad (3.6)$$

Gdzie r_{ij} jest odległością między środkami cząstek, σ_{ij} jest odległością, na której energia E_{VDW} przyjmuje wartość zero, niezależnie od orientacji cząstek, σ_{ij}^0 to odległość bez wkładu energii (przyjmuje wartość 0) dla orientacji cząstek bok do boku (z ang. side to side), zaś ε_{ij} odpowiada głębokości studni, która zależy od względnej orientacji cząsteczek van der Waalsa. Równania 3.7–3.10 wyrażają orientację zależności dla ε_{ij} oraz σ_{ij} :

$$\varepsilon_{ij} = \varepsilon(\omega_{ij}^{(1)}) = \varepsilon_{ij}^0 \varepsilon_{ij}^{(1)} \quad (3.7)$$

wraz z

$$\varepsilon_{ij}^{(1)} = [1 - \chi_{ij}^{(1)}]^2 \quad (3.8)$$

$$\sigma_{ij} = \sigma_{ij}^0 [1 - \chi_{ij}^{(1)}]^{-1/2} \quad (3.9)$$

wraz z:

$$\omega_{ij}^{(1)} = \hat{u}_{ij}^{(1)} \cdot \hat{r}_{ij} = \cos \alpha_{ij} \quad (3.10)$$

Gdzie $\hat{u}_{ij}$ wskazują wektory jednostkowe, które są ułożone wzdłuż głównych osi oddziałujących miejsc, zwłaszcza dla osi C α - łańcuch boczny (SC), wektor $\hat{r}_{ij}$ łączy środki oddziałujących miejsc, zaś r_{ij} to odległość między centrami cząsteczek. Parametr $\chi_{ij}^{(1)}$ oznacza ilościową anizotropię w odległości van der Waalsa dla centrum UNRES (które w przypadku centrum MARTINI jest kuliste). Podobnie $\chi_{ij}^{(1)}$ opisuje anizotropię głębokości studni van der Waalsa.

Udział energii tworzenia wnęki (ΔF_{cav}) dla cząstek sferoidalnych został określony za pomocą modelu Gaussa, nakładania się asocjacji hydrofobowej. Jest to wkład, który opisuje wydatek energetyczny związany złączeniem cząsteczek wody wokół hydrofobowego dimeru:

$$\Delta F_{cav} = \frac{\alpha_{ij}^{(1)} \left[(x \cdot \lambda)^{1/2} + \alpha_{ij}^{(2)} x \lambda - \alpha_{ij}^{(3)} \right]}{1 + \alpha_{ij}^{(4)} (x \cdot \lambda)^{12}} \quad (3.11)$$

wraz z:

$$x = \frac{r_{ij}}{\sqrt{\sigma_i^2 + \sigma_j^2}} \quad (3.12)$$

$$\lambda = [1 - \chi_{ij}^{\prime\prime(1)} \omega_{ij}^{(1)2}]^{1/2} \quad (3.13)$$

Gdzie $\chi^{\prime\prime(1)}$ reprezentuje anizotropię ΔF_{cav} . Parametry σ_i^1 i σ_j^2 to odpowiednio odległości kontaktowe między środkami cząstek i oraz j . Ponadto dopasowanie wyrażenia analitycznego na energię swobodną dwóch łańcuchów bocznych oddziałujących w wodzie do profili PMF określa się za pomocą metody najmniejszych kwadratów do parametrów $\alpha^{(1)}$, $\alpha^{(2)}$, $\alpha^{(3)}$ i $\alpha^{(4)}$ wraz z σ_i^1 , σ_j^2 i $\chi^{\prime\prime(1)}$. Parametry $\alpha^{(1)}$ - $\alpha^{(4)}$ nie mają łatwej interpretacji fizycznej.

Wzór 3.14 opisuje PMF par polarnych substancji rozpuszczonych (W_{pp}) w wodzie:

$$W_{pp} = E_{GBerne} + \Delta F_{cav} + E_{pp} \quad (3.14)$$

Ponadto energię potencjalną wynikającą z interakcji między dwoma biegunami głów cząstek (z ang. headgroups) dipol-dipol, określa E_{pp} (równanie 3.15):

$$E_{pp} = \frac{\omega_{p1}}{R_{ij}^3}(-3\cos\alpha_{ij}) - \frac{\omega_{p2}}{R_{ij}^6}(4 + 6\cos^2\alpha_{ij}) \quad (3.15)$$

Gdzie R_{ij} jest odległością między środkami grup polarnych, a ω_{p1} i ω_{p2} oznaczają parametry określone za pomocą procedury dopasowywania najmniejszych kwadratów zastosowanej do wyrażeń analitycznych pasujących do PMF, α_{ij} jest kątem między osiami $C\alpha \cdots SC$ łańcuchów bocznych.

Wzór 3.16 określa PMF dla naładowanej-polarnej substancji rozpuszczonych (W_{cp}) w wodzie, który przedstawia się następująco:

$$W_{cp} = E_{GBerne} + E_{pol} + \Delta F_{cav} + E_{cp} \quad (3.16)$$

Wzór 3.17 przedstawia wkład polaryzacyjny rozpuszczalnika (E_{pol}) do interakcji między cząstkami naładowanymi a polarnymi:

$$E_{pol} = \alpha_{ij}^{pol} \left(\frac{1}{f_{GB}(r_{ij}'')} \right)^4 + \alpha_{ji}^{pol} \left(\frac{1}{f_{GB}(r_{ji}'')} \right)^4 \quad (3.17)$$

Gdzie α_{ij}^{pol} i α_{ji}^{pol} są powiązane z polaryzacją niepolarnych składowych w łańcuchach bocznych i oraz j , r_{ij}'' to odległość między środkami ładunku łańcucha bocznego oraz i miejscem niepolarnym łańcucha bocznego j , zaś r_{ji}'' to odległość między środkami ładunku łańcucha bocznego j a miejscem niepolarnym łańcucha bocznego i , gdzie f_{GB} wyraża się za pomocą równania 3.18:

$$f_{GB}(r_{ij}') = \sqrt{r_{ij}'^2 + a_i a_j \exp\left(-\frac{r_{ij}'^2}{4R_i R_j}\right)} \quad (3.18)$$

Gdzie a_i oraz a_j są promieniami Borna miejsc i oraz j .

Zaś wzór 3.19 przedstawia potencjał interakcji E_{cp} , który charakteryzuje interakcje pomiędzy miejscami naładowanymi i polarnymi:

$$E_{cp} = w_{dip}^{(1)} \frac{q \cdot \cos\alpha}{R_{ij}^2} - w_{dip}^{(2)} \cdot \frac{q^2}{R_{ij}^4} \quad (3.19)$$

Gdzie q oznacza ładunek naładowanej "głowy" cząsteczki, a R_{ij} oznacza odległość między środkami głów cząstek, zaś $w_{dip}^{(1)}$ oraz $w_{dip}^{(2)}$ określa się metodą najmniejszych kwadratów poprzez dopasowanie wyników analitycznych wyrażenia z PMF

Potencjał średnich sił pomiędzy parą naładowanych substancji rozpuszczonych w wodzie (W_{cc}), określa się jako suma energii oddziaływania elektrostatycznego między ich naładowanymi komponentami poprzez równanie 3.20:

$$W_{cc} = E_{GBerne} + E_{el} + E_{GBorn} + \Delta F_{cav} \quad (3.20)$$

Wkład kulombowski w interakcji pomiędzy naładowanymi cząstkami jest określany jako E_{el} (równanie 3.21):

$$E_{el} = 332 \left(\frac{q_i q_j}{r'_{ij}} \right) \quad (3.21)$$

Gdzie r'_{ij} to odległość między głowami naładowanych miejsc w łańcuchach bocznych SC_i i SC_j , a q_i i q_j oznaczają ładunki związane z miejscami i oraz j . Współczynnik 332 jest współczynnikiem przeliczeniowym energii w kcal/mol. Równanie 3.22 przedstawia potencjał uogólnionego Borna:

$$E_{GBorn} = -332 \left(\frac{q_i q_j}{f_{GB}(r'_{ji})} \right) \left(\frac{1}{\epsilon_{in}} - \frac{\exp(-\kappa f_{GB} r'_{ji})}{\epsilon_{out}} \right) \quad (3.22)$$

Gdzie κ jest czynnikiem ekranowania, a ϵ_{in} jest stałą dielektryczną wewnątrz układu (ϵ_{out} jest stałą dielektryczną wody). Metoda najmniejszych kwadratów Marquardta zastosowałem do określenia wszystkich parametrów²²⁶.

3.4.4 Optymalizacja wagi i badanie fluktuacji

W kolejnym badaniu przetestowałem wagi, czyli czynniki skalujące energię oddziaływania ($\omega_{UNRES-MARTINI}$ z równania 3.23, porównaj z innymi wagami równania "1.5") $w =$ od 0,5 do 4,0 dla struktur białek odtworzonych za pomocą modelu UNRES otoczonych błoną lipidową uzyskanych za pomocą modelu MARTINI, aby zoptymalizować wagę interakcji w badanych molekułach.

$$U = U_{UNRES} + U_{MARTINI} + \omega_{UNRES-MARTINI} U_{UNRES-MARTINI} \quad (3.23)$$

Gdzie całkowita energia U składa się z U_{UNRES} , $U_{MARTINI}$, $U_{UNRES-MARTINI}$ wagi energii pola siłowego UNRES-MARTINI i oddziaływania nowo wprowadzonych potencjałów UNRES-MARTINI $\omega_{UNRES-MARTINI}$ - wagi UNRES-MARTINI.

Zastosowałem projektowaną *de novo* transbłonowych beczek beta (kod PDB: 6X9Z) i domenę ORD ludzkiego ORP8 (kod PDB: 8P7A) z membraną POPC z dostępnych w bazie PDB białek z błoną lipidową. Projektowana *de novo* transbłonową beczką beta znajdowała się wewnątrz błony, podczas gdy domena ORD ludzkiego ORP8 znajdowała się na powierzchni błony. Zbadałem również domenę C1B kinazy białkowej C w połączeniu z laktonem diacyloglicerolu (kod PDB 7LF3). Dla każdego z badanych układów wykonałem obliczenia z krokiem czasowym 0,1MTU (4,98fs) dla 1 000 000 kroków, przy rozmiarze pudełka periodycznego 75x75x120Å. Przeprowadziłem symulacje w temperaturze 300K przy użyciu termostatu Berendsena i stałej sprężenia 1,0MTU. Membrany POPC stworzyłem za pomocą CHARMM-GUI²²⁷.

Struktury reprezentatywne uzyskałem z 10 trajektorii, po usunięciu błon lipidowych, nałożeniu wszystkich konformacji białek na pierwszą konformację, a następnie użyłem algorytm klastrowania z programu CPPTRAJ²²⁸. Jako algorytm klastrowania użyłem DBSCAN²²⁹.

3.5 Obliczanie fluktuacji układów białko-glikozaminoglikan

3.5.1 ff14SB/GLYCAM06j

Zastosowałem pola siłowe ff14SB i GLYCAM06j do przeprowadzenia symulacji. Solwato wałem kompleksy FGF-1 HP dp6 (z ang. degree of polymerization), FDF-2 HP dp6 oraz CatK C4-s dp6 w oktaedrycznym pudełku periodycznym z modelem wody TIP3P, gdzie grubość warstwy cząsteczek wody liczona od powierzchni substancji rozpuszczonej wynosiła 6Å i zoboje tniałem przeciwnonami Na^+ . Parametry oraz ładunki dla grup siarczanowych GAGów zostały pozyskane z pola siłowego GLYCAM¹⁹⁰ oraz z literatury²³⁰. Przeprowadziłem minimalizację energii w dwóch etapach. Pierwszym krokiem było przeprowadzenie 5 cykli największego spadku i 10 cykli gradientu sprężonego z ograniczeniami siły harmonicznego 100 kcal/(mol·Å²) na atomach białka. Następnie 30 cykli największego spadku i 30 sprężonych cykli gradientów bez ograniczeń. W kolejnym kroku przez 10ps ogrzewałem system do 300K z siłami harmonicznymi ograniczającymi 100kcal/(mol·Å²) na atomach substancji

rozpuszczonej. Po czym równoważyłem system w izotermicznym zespole izobarycznym (NPT) przy temperaturze 300K i ciśnieniu 10^5 Pa. Na końcowym etapie w zespole NPT wykonałem trzy symulacje MD dla każdego kompleksu o czasie trwania $1\mu s$. Dla oddziaływań niezwiązanych wykorzystałem metodę Ewalda z siatką cząstek oraz algorytm SHAKE o punktem odcięcia $8\text{\AA}$ i kroku czasowym 2fs. Struktury zostały zapisane co 10ps, co dawało w sumie 10^4 struktur na symulację.

3.5.2 CHARMM36m

Pliki wejściowe przygotowałem przy użyciu serwera Charmm-gui²²⁷ do symulacji MD w polu siłowym CHARMM36m.

Solvatowałem kompleksy FGF-1 HP dp6, FDF-2 HP dp6 oraz CatK C4-s dp6 w oktaedrycznym pudełku periodycznym z modelem wody TIP3P, gdzie grubość warstwy cząsteczek wody liczona od powierzchni substancji rozpuszczone wynosiła $6\text{\AA}$ i zobojętniałem przeciwjonami Na^+ . Dla każdej symulacji wykonałem minimalizację energii w jednym etapie. Na atomy łańcucha głównego, łańcuchy boczne i kąty dwuścienne nałożyłem więzy, stosując stałą siłową. Aby zminimalizować energię, zastosowałem algorytm największego spadku z tolerancją 1×10^4 kJ/mol-nm. Następnie przeprowadziłem etap równowagowania, który miał te same więzy pozycyjne, co etap minimalizacji energii. Przeprowadziłem symulację równowagowania w zespole NVT dla 125ps w 300K i sprężeniu temperaturowym przy użyciu rozszerzonego zespołu Nose-Hoovera oraz algorytmu LINCS. Na koniec przeprowadziłem trzy symulacje MD dla każdego kompleksu z czasem trwania $1\mu s$ w zespole NPT. Użyłem algorytmów Nose-Hoovera i Parrinello-Rahmana do regulacji temperatury i ciśnienia przy stałej czasowej sprzęgania równej 1ps. Struktury zostały zapisane co 10ps, co dawało symulacji w sumie 10^4 na symulację. Przeprowadziłem te symulacje za pomocą programu GROMACS²³¹.

3.6 Walidacja oraz analiza wyników symulacji

Walidacja jest niezbędna do utrzymania integralności i dokładności uzyskanych wniosków w analizie danych. Mając na uwadze fakt, że część uzyskanych danych z analizy może wpływać na pozostałe wyniki (przeszacować bądź niedoszacować wartości), dlatego sprawdziłem poprawność ich uzyskania. Sprawdziłem, czy otrzymane dane mieszczą się w oczekiwanych granicach (np. wartości fluktuacji poniżej zera są jasnym sygnałem błędu danych, czy też wysokie wartości fluktuacji wpływają na

resztę wyników poszczególnych regionów).

3.6.1 Wyznaczanie profili fluktuacji

Do określenia elastyczności strukturalnej wykorzystałem średnie odchylenie standardowe fluktuacji (RMSF)¹⁴¹ do oceny jakości przewidywania profilów fluktuacji białek przy użyciu każdej z czterech metod rozważanych w tej pracy: UNRES-flex (także dla białek otoczonych błoną lipidową), UNRES-DSSP-flex, CABS-flex i NOLB, oraz dla białek w kompleksach z GAGami. Dla reszty aminokwasowej o indeksie i , profil RMSF jest definiowany równaniem 3.24:¹⁴¹

$$RMSF_i = \sqrt{\frac{1}{N} \sum_j^N (x_i(j) - \langle x_i \rangle)^2} \quad (3.24)$$

Gdzie $x_i(j)$ jest pozycją i - tego atomu $C\alpha$ w danym j - tym stanie lub j - tym modelu NMR, a $\langle x_i \rangle$ jest pozycją i - tego atomu $C\alpha$ uśrednioną z odpowiedniej symulacji lub zespołu NMR. Elastyczność danego związku jest widoczna w profilu RMSF³⁹.

Następnie obliczyłem profile fluktuacji białek uzyskanych z zespołów NMR, a także metod UNRES-flex, UNRES-DSSP-flex i NOLB. Najpierw z odpowiedniej symulacji węgle $C\alpha$ wszystkich struktur nałożyłem na $C\alpha$ pierwszej struktury, tak samo w przypadku NMR. Następnie obliczyłem średnią strukturę, uśredniając współrzędne kartezjańskie $C\alpha$, a następnie ponownie obliczyłem średnią strukturę, nakładając każdą strukturę z trajektorii na średnią strukturę. Ponieważ średnie struktury z drugiej iteracji były już bardzo zbliżone do tych z pierwszej, nie było potrzeby powtarzania procesu. Aby obliczyć profile RMSF, wykorzystałem średnie struktury z pierwszej iteracji jako odniesienia (równanie 3.24). Program CABS-Flex bezpośrednio wyprowadził profile RMSF uzyskane z metody CABS-flex¹³⁹.

W przypadku struktur rentgenowskich białek wartość RMSF jest powiązana ze współczynnikiem B , co jest przedstawia równanie 3.25:²³²

$$RMSF_i = \sqrt{\frac{3B_i}{8\pi^2}} \quad (3.25)$$

Gdzie B_i jest współczynnikiem B dla każdego i , który jest obliczany na podstawie współczynnika B dla każdego, indywidualnego atomu w strukturze.

W dalszej części pracy nazwałem jako “przewidywane” wartość profili fluktuacji RMSF, które zostały uzyskane przez odpowiednią metodę teoretyczną, zaś “eksperymentalne” zostały obliczone za pomocą współczynników B lub z zespołów NMR. Jeśli uwzględni się nieporządek krystalograficzny ciała sztywnego, np. dodając dodatkowe parametry nieuporządkowania ciała sztywnego dla każdej struktury PDB i optymalizując je wzajemnie za pomocą regresji²³³, można w zasadzie uzyskać nieco lepsze dopasowanie do współczynników krystalograficznych B. Jednak pominąłem dodatkową korektę nieuporządkowania w moich obliczeniach, ponieważ moim głównym celem było porównanie czterech metod symulacyjnych w badaniach nad elastycznością białek.

3.6.2 Korelacja danych eksperymentalnych z teoretycznymi

Aby ustalić jakie jest podobieństwo danych eksperymentalnych względem teoretycznych, analizowałem profile fluktuacji RMSF, które uzyskałem za pomocą odpowiednich metod, w porównaniu z profilami obliczonymi z zespołów NMR lub czynników B. Do tego wykorzystałem współczynnik korelacji liniowej Pearsona (r_p)²³⁴ i współczynnik korelacji rang Spearmana (r_s)²³⁵, które są przedstawione w równaniach 3.26 oraz 3.27. Wykorzystałem te współczynniki, aby określić stopień podobieństw profili fluktuacji eksperymentalnych z teoretycznymi.

$$r_p = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3.26)$$

Gdzie x_i oraz y_i to przewidywane i eksperymentalne wartości fluktuacji RMSF dla reszty o indeksie i , odpowiednio, $\bar{x}$ oraz $\bar{y}$ to uśrednione wartości fluktuacji RMSF dla wszystkich reszt, a n to liczba reszt.

$$r_s = 1 - \frac{\sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (3.27)$$

Gdzie d_i jest różnicą między rangami przewidywanego i eksperymentalnego profilu fluktuacji RMSF dla reszty o indeksie i . Ranga jest pozycją reszty i uzyskaną po sortowaniu reszty według wartości fluktuacji RMSF. W przypadku, gdy dwa profile są powiązane ze sobą liniowo (homotetycznie), współczynnik korelacji Pearsona jest bliski 1 lub -1, co oznacza korelację lub antykorelację. Wartość bliska zero wskazuje na brak współbieżności lub przeciwbieżności lub brak korelacji lub antykorelacji²³⁶.

3.6.3 Usuwanie reszt aminokwasowych z sekwencji białek oraz normalizacja danych

Zazwyczaj fragmenty końcowe N i C białek jednołańcuchowych wykazują znacznie większe ruchy w porównaniu do pozostałych fragmentów białek. Ta cecha jest najbardziej wyraźna w strukturach NMR. Niespecyficzne ruchy części końcowych mogą utrudniać identyfikację elastyczności w częściach pętli, która zazwyczaj jest ważna funkcjonalnie. W rezultacie analiza elastyczności białka, która obejmuje końce N i C, może powodować błędy w wynikach korelacji. Dlatego usunąłem regiony końcowe, stosując procedury opisane poniżej dla struktur rentgenowskich i zespołów NMR, aby porównać obliczone teoretycznie i eksperymentalnie profile fluktuacji dla istotnych biologicznie fragmentów. Jednak początkowe analizy przeprowadziłem dla pełnych struktur białek, a także obliczenia teoretyczne dla metod: UNRES-flex, UNRES-DSSP-flex, CABS-flex oraz NOLB.

Dla każdej struktury rentgenowskiej profil RMSF obliczyłem z współczynników B (równanie 3.25) dla pełnej struktury białka. Następnie obliczyłem średnią (dla wszystkich reszt aminokwasowych) wartość RMSF ($\overline{RMSF}$) wraz z odchyleniem standardowym ($\sigma_{\overline{RMSF}}$). Na koniec usunąłem końcowe reszty aminokwasowe w następujący sposób:

$$RMSF_i - \overline{RMSF} > 3\sigma_{\overline{RMSF}} \quad (3.28)$$

dla $i = 1, 2, \dots, lnt$ i $i = n, n - 1, \dots, n - let + 1$, gdzie lnt i let to długości wyeliminowanych reszt N- i C-końcowych białek.

Określiłem średnią strukturę i profil RMSF dla pełnych struktur białek uzyskanych z zespołów NMR, jak opisałem we wcześniejszym podrozdziale. Następnie obliczyłem średnią wartość fluktuacji RMSF, a także odchylenie standardowe. Usunąłem końcowe N oraz C reszty aminokwasowe, których wartość $RMSF_i - \overline{RMSF} > 3\sigma_{\overline{RMSF}}$. W przypadku skróconych łańcuchów białek powtórzyłem tę procedurę, ale w większości przypadków nie było potrzebne dalsze ich usuwanie.

Ponieważ różne metody powodują różne amplitudy fluktuacji RMSF, uwzględniłem również znormalizowane profile RMSF (RMSFN), które są zdefiniowane równaniem 3.29:

$$RMSFN_i = \frac{RMSF_i}{\sqrt{\sum_{i=1}^n RMSF_i^2}} \quad (3.29)$$

Gdzie $RMSF_i$ to przewidywany profil fluktuacji dla reszt o indeksie i , a n to liczba reszt.

3.6.4 Analiza istotności statystycznej

Aby ocenić zależności współczynników korelacji r_p oraz r_s od zastosowanej metody (UNRES-flex, UNRES-DSSP-flex, CABS-flex i NOLB), rodzaju struktur drugorzędowych (α , β i $\alpha + \beta$), oraz od metody eksperymentalnej (rentgenografii strukturalnej i spektroskopii magnetycznego rezonansu jądowego) wykorzystałem dwukierunkową analizę wariancji (ANOVA)²³⁷, którą przeprowadziłem przy poziomie istotności 5% wykorzystując serwer online Statistics Kingdom (dostępny pod adresem <https://www.statskingdom.com/two-way-anova-calculator/>).

3.6.5 Analiza skośności układu

Porównując ze sobą rozkłady współczynników korelacji pomiędzy profilami fluktuacji z symulacji a profilami fluktuacji z eksperymentu uzyskanych z różnych metod może dojść do sytuacji, w której badany rozkład może nie wykazywać widocznych cech modalności. Dlatego wyznacza się skośność badanego układu za pomocą równania 3.30:

$$\gamma = \frac{1}{n\sigma_{RMSF}^3} \sum_{i=1}^n (RMSF_i - \overline{RMSF})^3 \quad (3.30)$$

Gdzie n jest liczbą analizowanych struktur. Skośność pozwala określić stopień asymetrii rozkładu danych, czyli czy są one symetrycznie rozłożone wokół średniej, czy też jeden koniec jest bardziej “rozciągnięty” niż drugi.

3.6.6 Określenie podobieństwa strukturalnego białek otoczonych błoną lipidową

Najprostszym sposobem porównania struktur białkowych otoczonych błoną lipidową jest przeprowadzenie transformacji, która nakłada odpowiednie atomy jednej struktury z atomami drugiej struktury. W ten sposób możliwe jest minimalizowanie odchylenia średniokwadratowego (RMSD)²³⁸ pomiędzy współrzędnymi nałożonych na siebie struktur (równanie 3.31).

$$RMSD = \sqrt{\frac{1}{n} \sum_{i=1}^L (x_i - x_0)^2} \quad (3.31)$$

Gdzie n jest to całkowita liczba wartości, x_i jest wartością oczekiwaną, zaś x_o jest wartością obserwowaną (rzeczywistą).

W przeciwieństwie do tego TM-score²³⁹ wykorzystuje zmienioną wersję metryki Levitta-Gersteina (LG)²³⁹. Niezależnie od długości, ta metryka zmniejsza wpływ rozbieżnych par atomów w wyrównanych strukturach (równanie 3.32). Nie ma prostej zależności pomiędzy macierzą kowariancji a optymalną transformacją, która maksymalizuje metrykę LG, w przeciwieństwie do RMSD. Alternatywnie dopasowuje się różne podzbiory atomów za pomocą algorytmu Kabscha²⁴⁰, a następnie wylicza się wynik LG dla całej struktury białka²³⁹.

$$TM - score = \frac{1}{L} \sum_{i=1}^L \frac{1}{1 + (\frac{d_i}{d_0})^2} \quad (3.32)$$

Gdzie L jest długością białka, d_i jest odległością pomiędzy i -tym dopasowanym atomem $C\alpha$, a d_0 jest współczynnikiem skalującym, normalizującym dopasowania.

Rozdział 4

Wyniki

4.1 Wpływ końcowych reszt aminokwasowych na profile fluktuacji

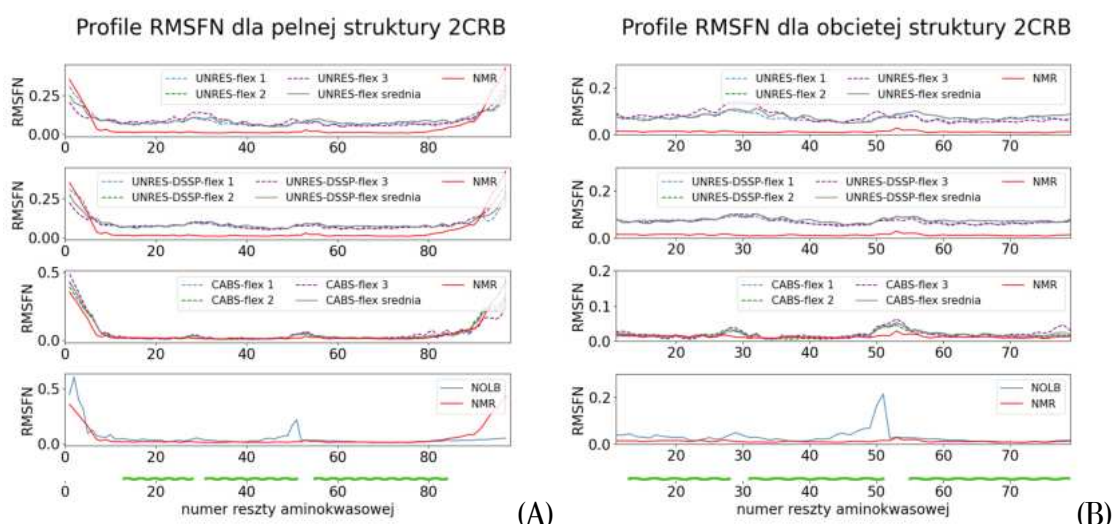
W związku z tym, że poszczególne metody gruboziarniste wykazują różne dokładności przewidywania elastyczności struktur białkowych, dlatego w swojej pracy skupiłem się na porównaniu wyników symulacji uzyskanych z metod teoretycznych z danymi eksperymentalnymi, aby określić, która spośród 4 badanych przeze mnie metod najdokładniej odwzorowuje elastyczność białek. W tym celu szczegółowo przeanalizowałem przykładowe białko ze względu na jego krótką strukturę: myślinowy czynnik wiążący receptor jądrowy 2 (o kodzie PDB: 2CRB)²⁴¹, czyli białko, które zawiera w sobie wyłącznie struktury α -helikalne. Struktura tego białka została uzyskana z zespołów NMR, którą przedstawiłem na rysunku 4.1. Metody UNRES-flex, UNRES-DSSP-flex i CABS-flex wykazały wysokie wartości współczynników r_p dla pełnej struktury analizowanego białka (od 0,89 do 0,96) w porównaniu do metody NOLB (0,56). Średnie wartości współczynników korelacji dla pełnej struktury białka (o kodzie PDB: 2CRB) znajdują się w tabeli 8.1 w załącznikach. Każdy z współczynników korelacji obliczyłem na podstawie trzech przeprowadzonych symulacji za pomocą odpowiednich metod uzyskując profile RMSFN, których wyniki uśredniłem. Wysokie wartości fluktuacji reszt aminokwasowych, które znajdują się na N- i C- końcach białek przyczyniają się do dobrej zgodności profilu eksperymentalnego RMSFN z profilami przewidywanymi RMSFN. Zgodność profili RMSFN eksperymentalnych i przewidywanych jest nadal dobra po usunięciu sekcji końcowych w przypadku metody CABS-flex, gdzie wartość współczynnika r_p waha się od 0,53 do 0,74 (pozostawiając po obcięciu sekwencję reszt aminokwasów między 8–85). Dla pozostałych metod wartość współczynnika r_p jest o wiele niższa i waha się od 0,27 do 0,52 (rysunek 4.2 B). Średnie wartości współczynników kore-

lacji dla skróconej struktury białka (kod PDB: 2CRB) znajdują się w tabeli 8.2 w załącznikach. W dalszych analizach wykorzystałem profile fluktuacji bez fragmentów końcowych, aby wyeliminować dominujący wpływ fragmentów końcowych na profile RMSE. Profile RMSFN uzyskane z każdej z trzech indywidualnych trajektorii symulowanych (UNRES-flex, UNRES-DSSP-flex i CABS-flex) porównałem, aby zobaczyć, jak poszczególna symulacja wpływa na profile fluktuacji. W związku z tym, że metoda NOLB jest metodą deterministyczną, brałem pod uwagę tylko jedną trajektorię. Różnice między profilami RMSFN obliczonymi z indywidualnych trajektorii są podobne dla pełnej struktury białka, co można zauważyć na rysunku 4.2 A. Zauważyłem, że różnice w zakresie fluktuacji są mniejsze na końcowych obszarach sekwencji białka, gdzie wartości profili fluktuacji są największe.

Tym samym różnice wartości r_p pomiędzy przeprowadzonymi trajektoriami (0,09 dla UNRES-flex; 0,06 dla UNRES-DSSP-flex i 0,08 dla CABS-flex) i r_s (0,06 dla UNRES-flex, 0,12 dla UNRES-DSSP-flex i 0,02 dla CABS-flex) są nieznaczne. Z analizy wszystkich profili RMSFN wynika, że zastosowanie jednej trajektorii może być wystarczające. Niemniej jednak należy pamiętać, że fluktuacje odcinków końcowych w znaczny sposób wpływają na wartości r_p i r_s . Różnice wartości r_p pomiędzy przeprowadzonymi trajektoriami (0,38 dla UNRES-flex, 0,22 dla UNRES-DSSP-flex i 0,19 dla CABS-flex) i r_s (0,05 dla UNRES-flex, 0,22 dla UNRES-DSSP-flex i 0,03 dla CABS-flex) stają się bardziej widoczne, gdy analizuje się bez N- i C- końcowych fragmentów. Na podstawie uzyskanych wyników stwierdzam, że przeprowadzenie wielu trajektorii jest niezbędne, aby uzyskać wiarygodne profile RMSFN. Z tego powodu w dalszych analizach wykorzystałem profile RMSFN i wielkości uzyskane na podstawie uśrednionych trajektorii. Należy również zwrócić uwagę na to, że wartości r_p są niższe w przypadku zastosowania metody NOLB. To prawdopodobnie wynika z faktu, że analiza trybu normalnego jest ważna głównie dla dobrze określonych struktur odniesienia, co nie występuje w przypadku zespołów NMR.



Rysunek 4.1: Struktura analizowanego białka z ukazaniem jego struktur drugorzędowych, na których przedstawiono: myszy czynnik wiążący receptor jądrowy 2 (o kodzie PDB: 2CRB, posiadający wszystkie- α , struktura uzyskana z zespołu NMR) Reprinted with permission from Ł. J. Dziadek, A. K. Sieradzan, C. Czaplewski, M. Zalewski, F. Banaś, M. Toczek, W. Nisterenko, S. Grudin, A. Liwo, A. Giełdoń, Assessment of Four Theoretical Approaches to Predict Protein Flexibility in the Crystal Phase and Solution, J. Chem. Theory Comput. 2024, 20, 17, 7667–7681. Licensed under CC-BY 4.0

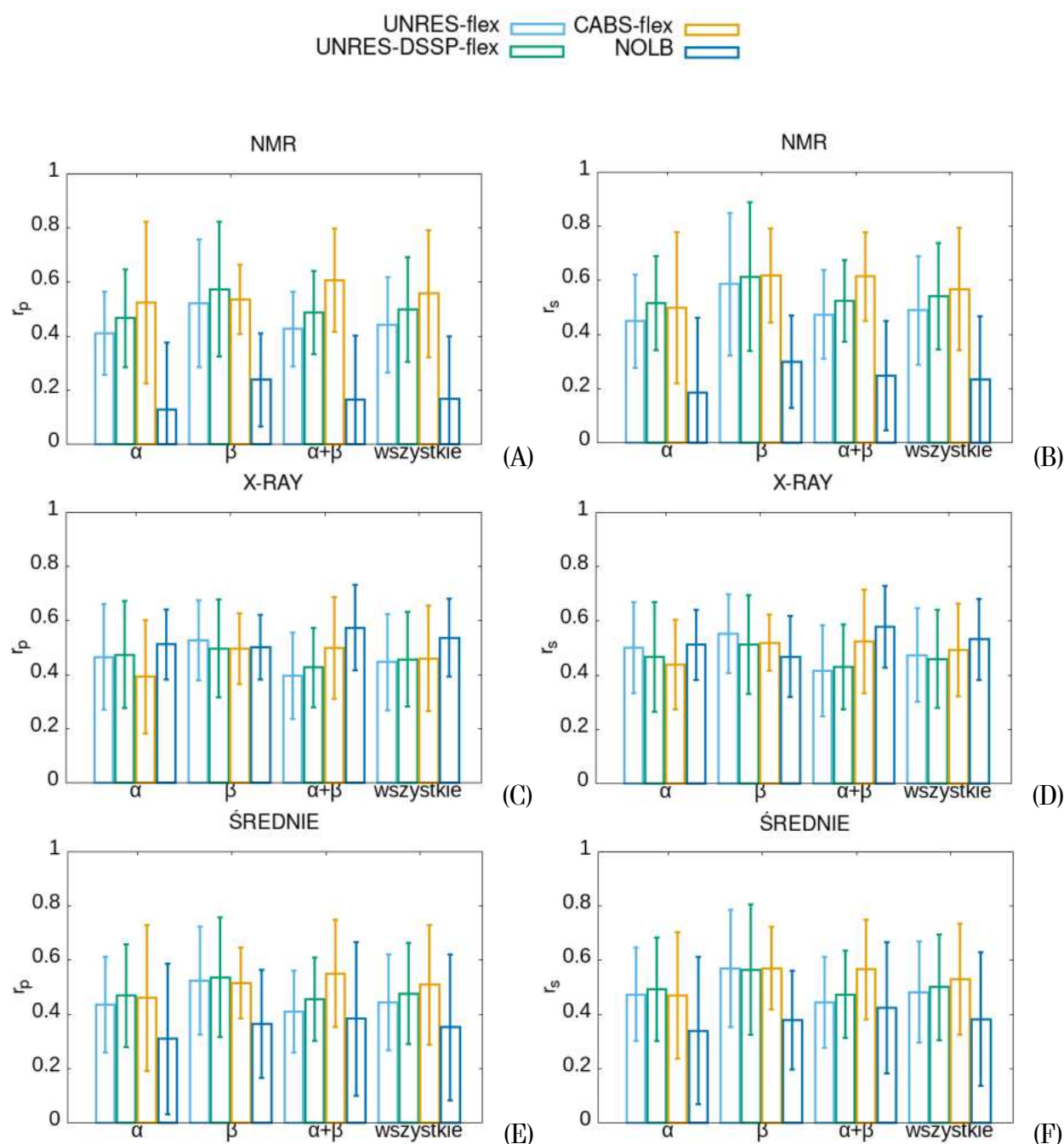


Rysunek 4.2: Profile RMSFN dla białka o kodzie PDB: 2CRB, obliczone za pomocą metod: UNRES-flex, UNRES-DSSP-flex, CABS-flex (trzy symulacje i średnia) oraz NOLB w porównaniu z profilem RMSFN obliczonym z zespołów NMR dla (A) pełnej i (B) obciętej struktury białka. Profile są rozróżniane za pomocą stylów linii i kolorów. Falowane zielone linie na dole wykresów oznaczają α -helikalne struktury. Adapted with permission from Ł. J. Dziadek, A. K. Sieradzan, C. Czaplewski, M. Zalewski, F. Banaś, M. Toczek, W. Nisterenko, S. Grudin, A. Liwo, A. Giełdoń, Assessment of Four Theoretical Approaches to Predict Protein Flexibility in the Crystal Phase and Solution, J. Chem. Theory Comput. 2024, 20, 17, 7667–7681. Licensed under CC-BY 4.0

4.2 Porównanie przewidywanych i eksperymentalnych profili RMSFN w zestawie testowym

4.2.1 Dokładność metod teoretycznych względem eksperymentalnych wyznaczania profili fluktuacji RMSFN oraz struktury drugorzędowej białek

Następną analizę przeprowadziłem aby określić, która z zastosowanych metod (UNRES-flex, UNRES-DSSP-flex, CABS-flex lub NOLB) najlepiej odtwarza profile fluktuacji względem danych eksperymentalnych na podstawie przeprowadzonych obliczeń średnich współczynników korelacji Pearsona (r_p) i Spearmana (r_s) dla pełnych oraz skróconych struktur badanych białek. Na rysunkach 7.1 (dla skróconych struktur białek) i 7.2 (dla pełnych struktur białek) w załącznikach przedstawiłem wykresy profili fluktuacji RMSFN dla wszystkich stu testowanych białek, zaś w tabelach 7.4 i 7.5 w załącznikach przedstawiłem r_p oraz r_s pomiędzy teoretycznymi a eksperymentalnymi profilami fluktuacji dla każdego badanego białka. W każdym przypadku obliczyłem również średnie wartości r_p oraz r_s (sekcja 4.3) dla podzbiorów względem eksperymentalnych metod (NMR lub X-ray), a także dla typu struktury drugorzędowej (α , β , $\alpha + \beta$) dla 100 testowanych białek (tabela 7.3 w załącznikach). W tabelach 7.6 i 7.7 w załącznikach podsumowałem średnie współczynniki korelacji. Na rysunkach 4.3 A-D przedstawiłem średnie r_p oraz r_s w zależności od metody eksperymentalnej (NMR i Xray) oraz typu struktury drugorzędowej, zaś na rysunkach 4.3 E-F przedstawiłem średnie r_p oraz r_s dla całego zbioru. Uzyskane wyniki prezentują, że wartości r_p i r_s (odpowiednio: 0,441 i 0,489 UNRES-flex; 0,498 i 0,541 UNRES-DSSP-flex; 0,558 i 0,568 CABS-flex, tabela 7.6 A w załącznikach) są wyższe dla metod UNRES-flex, UNRES-DSSP-flex oraz CABS-flex w przypadku struktur białek uzyskanych z zespołów NMR oraz dla całego zbioru białek w porównaniu do metody NOLB. W przypadku struktur uzyskanych metodą rentgenografii Xray, metoda NOLB w większości przypadków (oprócz struktur drugorzędowych β) wykazuje wyższe wartości r_p i r_s (odpowiednio: 0.512 i 0.512 α ; 0.573 0.578 $\alpha + \beta$ 0.537 i 0.532 dla wszystkich struktur, tabela 7.6 B) w porównaniu do innych metod.

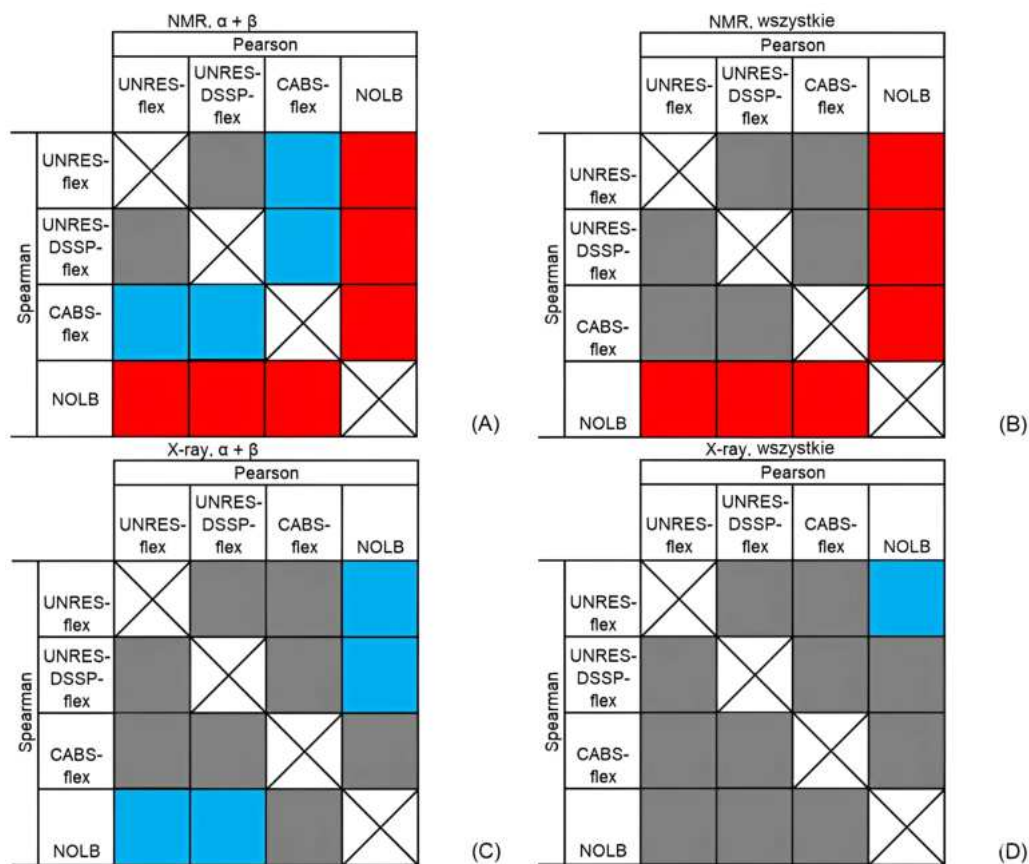


Rysunek 4.3: Wykresy słupkowe (z zaznaczonymi odchyleniami standardowymi) średnich współczynników korelacji Pearsona (A, C, E) i Spearmana (B, D, F) między profilami fluktuacji reszt aminokwasowych uzyskanych z symulacji UNRES-flex (stalowo-niebieski), UNRES-DSSP-flex (zielony), CABS-flex (pomarańczowy) i NOLB (niebieski), dla struktur NMR (A, B) i rentgenowskich (C, D) oraz niezależnie od metody określania struktury (E, F) dla skróconych struktur białek. Analizy przeprowadzono dla białek α , β i $\alpha + \beta$ oraz niezależnie od typu struktury drugorzędowej. Adapted with permission from Ł. J. Dziadek, A. K. Sieradzan, C. Czaplewski, M. Zalewski, F. Banaś, M. Toczek, W. Nisterenko, S. Grudin, A. Liwo, A. Gieldoń, Assessment of Four Theoretical Approaches to Predict Protein Flexibility in the Crystal Phase and Solution, J. Chem. Theory Comput. 2024, 20, 17, 7667–7681. Licensed under CC-BY 4.0

Następnie zastosowałem dwukierunkową analizę wariancji z następującymi dwiema kategoriami zmiennych: (I) metodą przewidywania fluktuacji i (II) typem struktur drugorzędnych, aby zweryfikować wcześniejsze stwierdzenie. Przeprowadziłem również analizy w zależności od metody eksperymentalnej (NMR i Xray). Jest to związane z tym, że eksperymentalne profile fluktuacji w strukturach uzyskanych z rentgenografii Xray są obliczane ze współczynników B i odpowiadają drganiom harmonicznym. W przypadku eksperymentalnych profili fluktuacji obliczonych struktur białek uzyskanych z zespołów NMR sytuacja jest inna. Dzieje się tak ze względu na różnice wokół średniej struktury. W tabeli 7.8 w załącznikach podsumowałem odpowiednie poziomy istotności. Na podstawie uzyskanych wyników stwierdziłem, że zależność współczynników korelacji od metody przewidywania fluktuacji jest istotna na poziomie co najmniej 0,05 (z wyjątkiem r_s i struktur rentgenowskich). Z drugiej strony zależność współczynników korelacji od typu struktury drugorzędowej jest nieistotna (z wyjątkiem r_s i struktur NMR). W tabeli 7.8 w załącznikach pokazałem, że wpływ obu kategorii zmiennych, (przedstawionych jako interakcje), na współczynniki korelacji jest minimalny lub go nie ma. W rezultacie zależność r_s i r_p od metody przewidywania fluktuacji jest potwierdzona przez analizę ANOVY, co przedstawiłem na rysunku 4.3.

Przeanalizowałem zestawy współczynników korelacji (r_p lub r_s) przy użyciu testu Studenta²⁴² osobno dla struktur uzyskanych krystalografią Xray i z zespołów NMR, aby określić różnice między cechami metod przewidywania fluktuacji stosowanych do białek o określonym typie struktury drugorzędowej. W tabelach 7.9 w załącznikach umieściłem szczegółowe informacje dotyczące wyników. Dla struktur uzyskanych z zespołów NMR, wyniki r_p lub r_s uzyskane metodą NOLB są niższe w porównaniu z innymi metodami: CABS-flex, UNRES-flex i UNRES-DSSP-flex, zaś w przypadku białek uzyskanych z rentgenografii Xray, o strukturach drugorzędowych $\alpha + \beta$, metoda NOLB wykazuje wyższe wartości r_p lub r_s w porównaniu do wyników uzyskanych z metod UNRES-flex i UNRES-DSSP-flex (tabela 7.9 C, D w załącznikach, rysunek 4.4). Warto zauważyć, że odnosiłem się wyłącznie do różnic, które zostały uznane za statystycznie istotne. W przypadku analizy białek bez rozróżnienia struktury drugorzędowej określiłem, że wyniki uzyskane z metod UNRES-flex, UNRES-DSSP-flex oraz CABS-flex wykazują wyższe wartości r_p lub r_s w porównaniu do wyników uzyskanych z metody NOLB. Na podstawie uzyskanych wyników mogę stwierdzić, że do przewidywania elastyczności struktur białek uzyskanych z zespołów NMR o strukturze drugorzędowej

$\alpha + \beta$ metoda CABS-flex wykazuje lepsze wyniki statystyczne niż metoda UNRES-flex, przy 5% lub wyższej istotności statystycznej (tabela 7.9 A w załącznikach, rysunek 4.4). Zaobserwowałem również, że metoda CABS-flex sprawdza się lepiej w porównaniu do metody UNRES-flex w przypadku białek uzyskanych z rentgenografii Xray o strukturach $\alpha + \beta$, jednak istotność statystyczna różnic między współczynnikami korelacji jest mniejsza niż 5% (tabela 7.9 C w załącznikach). W przypadku białek uzyskanych z zespołów NMR o strukturze $\alpha + \beta$, metoda CABS-flex wykazuje większe wartości r_p w porównaniu do metody UNRES-DSSP-flex przy poziomie istotności około 5% (tabela 7.9 A w załącznikach, rysunek 4.4). Z drugiej strony nie ma statystycznie istotnych różnic między uzyskanymi wartościami r_p i r_s dla metod UNRES-flex, CABS-flex i UNRES-DSSP-flex dla testowanych białek, niezależnie od typu struktury drugorzędowej. Z tego powodu mogę stwierdzić, że metody CABS-flex i UNRES-DSSP-flex przewidują profile fluktuacji z podobną dokładnością.

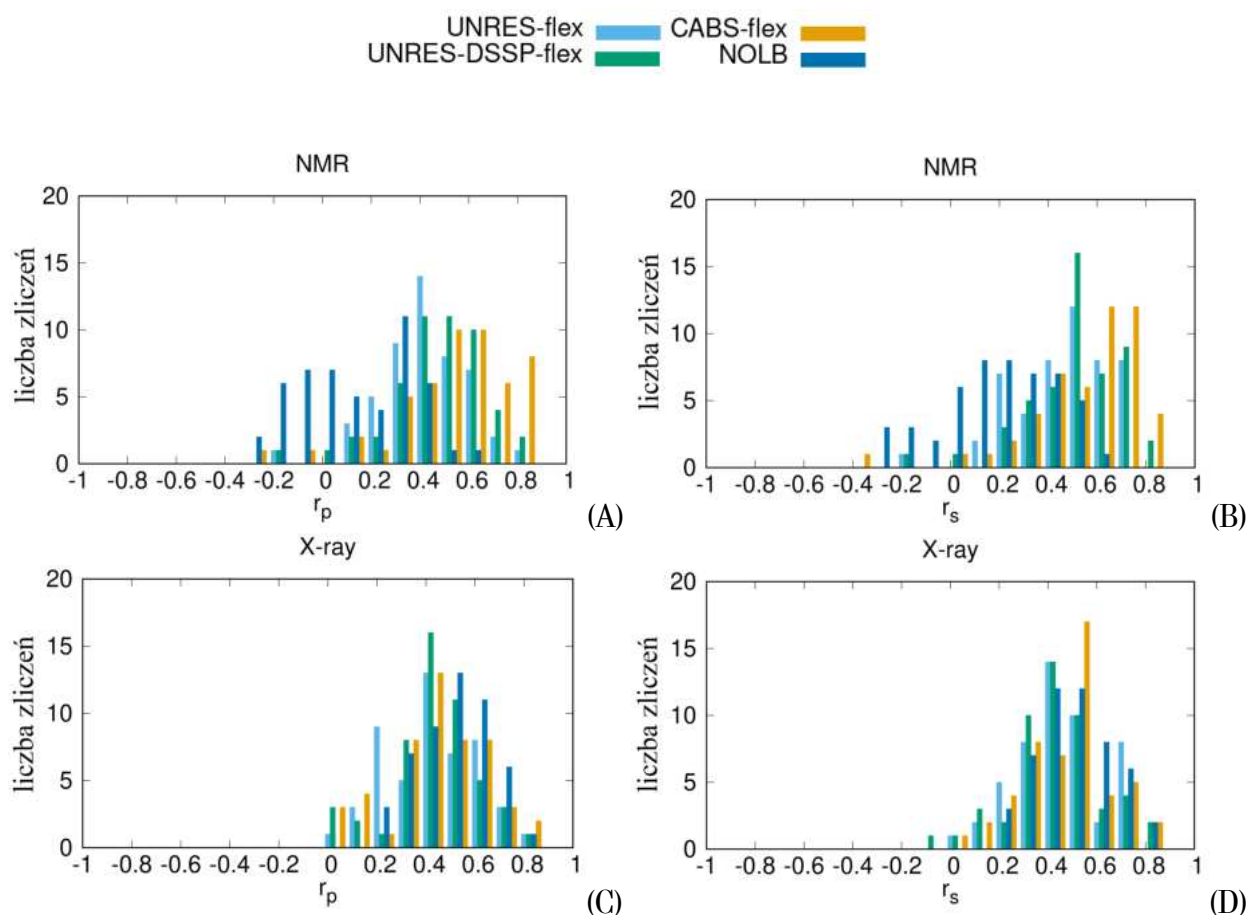


Rysunek 4.4: Zestawienie istotności statystycznej oraz znaku współczynników korelacji r_p (powyżej przekątnej) i r_s (poniżej przekątnej) odpowiadających czterem metodom przewidywania profili fluktuacji ocenianych w tej pracy dla struktur NMR (A, B) i rentgenowskich (C, D) oraz białek $\alpha + \beta$ (A, C) i wszystkich typów struktur białek (B, D). W przypadku współczynnika r_p metoda odpowiadająca konkretnemu nagłówkowi kolumny jest porównywana z metodą odpowiadającą wierszowi powyżej przekątnej, natomiast w przypadku r_s metoda odpowiadająca odpowiedniemu nagłówkowi wiersza jest porównywana z metodą odpowiadającą kolumnie poniżej przekątnej. Czerwony kolor oznacza, że różnica jest ujemna przy poziomie istotności $<0,05$, niebieski: różnica jest dodatnia przy poziomie istotności $<0,05$, szary: różnica jest statystycznie nieistotna. Adapted with permission from Ł. J. Dziadek, A. K. Sieradzan, C. Czaplewski, M. Zalewski, F. Banaś, M. Toczek, W. Nisterenko, S. Grudin, A. Liwo, A. Giełdoń, Assessment of Four Theoretical Approaches to Predict Protein Flexibility in the Crystal Phase and Solution, J. Chem. Theory Comput. 2024, 20, 17, 7667–7681. Licensed under CC-BY 4.0

4.2.2 Rozkład współczynników korelacji względem liczby zliczeń

Poprzednia analiza pozwoliła mi ocenić średnią dokładność przewidywania profili fluktuacji w zależności od rodzaju struktury drugorzędnej i metody predykcji fluktuacji dla wybranych czterech metod. Jednak sposób rozkładu współczynników korelacji, a przede wszystkim modalność i asymetria, może wskazać, czy przewidywania wartości fluktuacji zostały poprawnie przewidziane

dla większości białek. Przeprowadziłem analizę rozkładów r_p i r_s dla ich liczby zliczeń (w przedziałach o rozmiarze 0,1) obciętych struktur białek uzyskanych z rentgenografii X-ray oraz z zespołów NMR, aby określić asymetrię rozkładów. Zestawiłem liczbę zliczeń z r_p i r_s dla różnych metody na rysunku 4.5 A–D.



Rysunek 4.5: Rozkłady wartości współczynników korelacji Pearsona (r_p) lub Spearmana (r_s) pomiędzy profilami fluktuacji uzyskanymi z zespołów NMR lub współczynników B promieniowania rentgenowskiego a profilami przewidzianymi przy użyciu metod: UNRES-flex (kolumna stalowo-niebieska), UNRES-DSSP-flex (kolumna zielona), CABS-flex (kolumna pomarańczowa) i NOLB (kolumna niebieska) dla skróconych struktur białek. Adapted with permission from Ł. J. Dziadek, A. K. Sieradzan, C. Czaplewski, M. Zalewski, F. Banaś, M. Toczek, W. Nisterenko, S. Grudin, A. Liwo, A. Giełdoń, Assessment of Four Theoretical Approaches to Predict Protein Flexibility in the Crystal Phase and Solution, J. Chem. Theory Comput. 2024, 20, 17, 7667–7681. Licensed under CC-BY 4.0

Dla struktur uzyskanych z zespołów NMR rozkład r_p obliczony za pomocą metody NOLB wykazuje bimodalność, gdzie pierwsze maksimum wynosi około 0, zaś drugie w okolicy 0,3. W zasadzie rozkład r_s również jest bimodalny dla metody NOLB, gdzie druga część rozkładu jest dość szeroka. Biorąc pod uwagę wyłącznie drugą część rozkładu r_p , dokładność metody NOLB na podsta-

wie r_p wykazuje podobną dokładność w porównaniu do metody UNRES-flex, podczas gdy pierwsza część rozkładu odpowiada niewielkiej liczbie zliczeń. W związku z tym przeprowadziłem badanie elastyczności struktur białek, które odpowiadają pierwszej części rozkładu, biorąc pod uwagę środkową część układu, dla $r_p = 0$, których listę przedstawiłem w tabeli 7.10 w załącznikach. Na jej podstawie stwierdziłem, że niektóre struktury nie mają wspólnych cech, takich jak szczególny typ struktury drugorzędnej czy też długie pętle. Może to wynikać z faktu, że hamiltonian metody NOLB, który opiera się wyłącznie na odległościach międzyatomowych, powoduje gorszą wydajność ze strukturami uzyskanymi z zespołów NMR w porównaniu ze strukturami uzyskanymi z rentgenografii Xray. W związku z tym elastyczność struktur białek uzyskanych z rentgenografii Xray zależy głównie od odległości ich atomów od atomów innych struktur, gdzie siła oddziaływań specyficznych jest mniej ważna. Jest to spowodowane tym, że fragmenty struktur białek uzyskanych z rentgenografii Xray, takie jak α -helisy, są zwykle blisko innych elementów strukturalnych danego białka, a jeśli znajdują się na zewnątrz białka, są dość upakowane względem pozostałych cząsteczek białka. W przeciwieństwie do α -helis, pętle zazwyczaj nie są tak ściśle upakowane w porównaniu do innych fragmentów białka.

Wcześniejsze wnioski mogę potwierdzić na podstawie rysunków 4.5 C i D, na których przedstawiłem rozkład r_p i r_s odpowiednio ze struktur białek uzyskanych z rentgenografii Xray. Widoczne jest, że rozkłady odpowiadające wynikom metody NOLB są unimodalne dla struktur białek określonych za pomocą rentgenografii Xray i nieznacznie przesunięte w kierunku wyższych wartości w porównaniu z rozkładami otrzymanymi z innych metod. W przypadku metody CABS-flex można zaobserwować na rysunkach 4.5 A i B, że rozkłady dla obu współczynników korelacji są bimodalne i silnie przesunięte w kierunku dużych wartości sugerując, że w większości badanych białek struktur uzyskanych z zespołu NMR, metoda CABS-flex najlepiej przewiduje profile fluktuacji. Metody UNRES-flex oraz UNRES-DSSP-flex wykują unimodalność i również są przesunięte w prawym kierunku. Porównując ze sobą obie metody na podstawie liczby zliczeń dla przedziałów o wyższych wartości współczynników korelacji (powyżej 0,5) widoczne jest, że metoda UNRES-DSSP-flex jest lepsza w porównaniu do metody UNRES-flex. Na rysunkach 4.5 C i D obserwuje się, że metody UNRES-flex, UNRES-DSSP-flex oraz CABS-flex w przypadku struktur białek uzyskanych z rentgenografii Xray, liczba zliczeń dla najwyższych wartości współczynników korelacji są zliżone do siebie (dla przedziału 0,6 oraz

wyższych), a rozkłady są unimodalne. Dla r_p zauważalne jest, że w przypadku metody UNRES-flex liczba zliczeń jest większa dla mniejszych wartości współczynnika w porównaniu do reszty metod co może sugerować, że w mniejszym stopniu sprawdza się w przypadku białek X-ray w porównaniu z innymi zastosowanymi metodami.

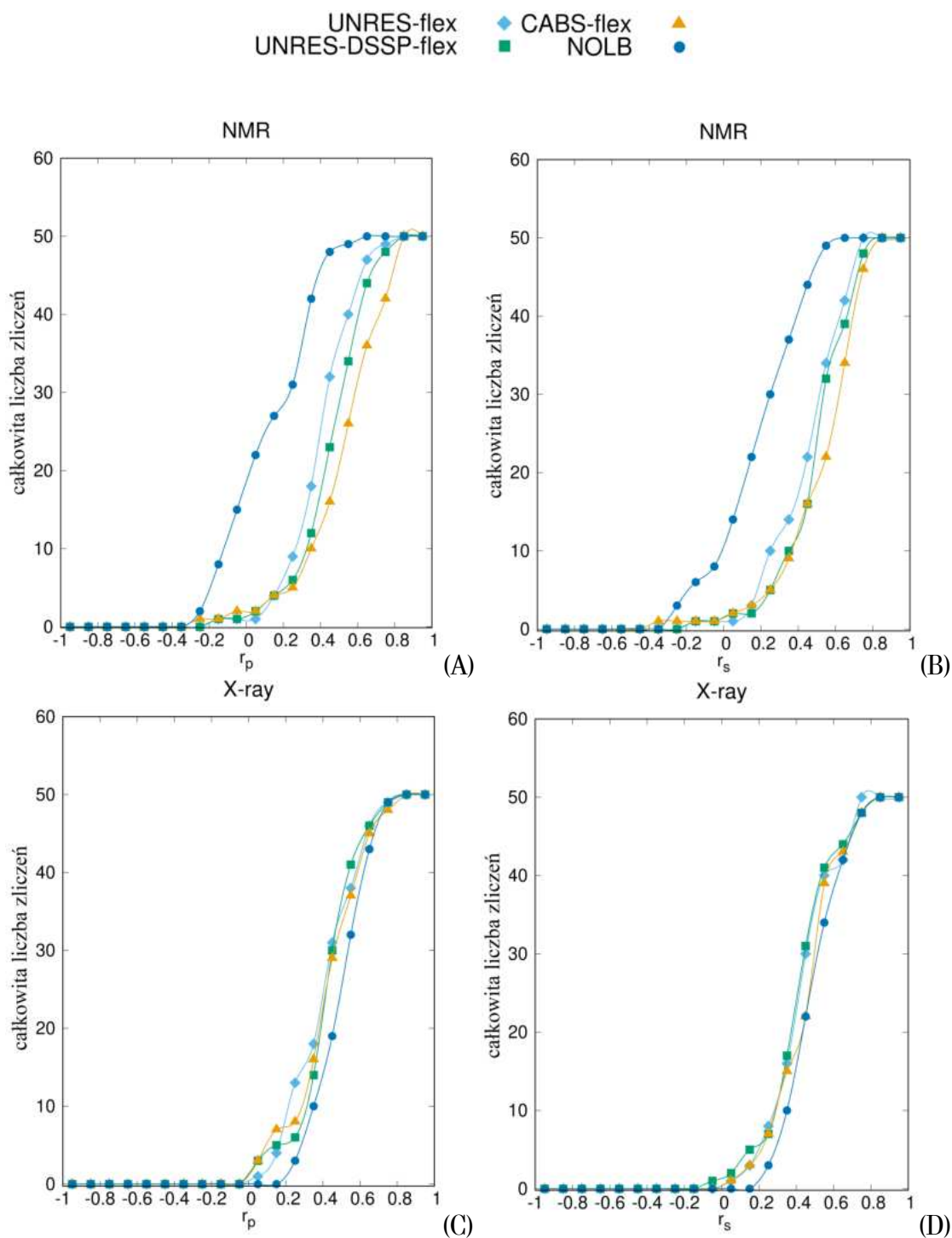
4.2.3 Analiza skośności układów

Poprzednia analiza pozwoliła mi określić, że rozkłady współczynników korelacji z metod UNRES-flex, UNRES-DSSP-flex oraz CABS-flex są lewostronnie skośne, jak również, że te metody nie wykazują bimodalności. Aby to zweryfikować obliczyłem skośności układów (sekcja 4.5). Dla współczynnika korelacji r_p wartości γ prezentują się w następujący sposób: UNRES-flex = $-0,50$, UNRES-DSSP-flex = $-0,86$, CABS-flex = $-1,16$ i NOLB = $-0,01$. Dla współczynnika korelacji r_s : UNRES-flex = $-0,79$, UNRES-DSSP-flex = $-1,09$, CABS-flex = $-1,60$ i NOLB = $-0,34$. Metoda CABS-flex wykazuje najbardziej ujemną skośność wyników korelacji, co jest zgodne z dużą gęstością znacznie dodatniego współczynnika korelacji, a także z faktem, że istnieją również wartości o małych lub nawet ujemnych wartościach. Metoda UNRES-DSSP-flex jest przykładem drugiej ujemnej skośności, która wykazuje bardziej symetryczny rozkład w porównaniu do metody CABS-flex, ze względu na przesunięty w lewo środek rozkładu (wykres 4.5). W przypadku metody NOLB skośność jest niewielka dla r_p oraz r_s , dlatego nie jest brana pod uwagę. Jest to spowodowane podobną wagą obu części rozkładu r_p (powyżej i poniżej wartości 0.2) i prawie symetrycznym, szerokim rozkładem r_s . Na podstawie przeprowadzonej analizy doszedłem do wniosku, że chociaż metody CABS-flex i UNRES-DSSP-flex zwykle dają dobre przewidywania fluktuacji dla struktur białek uzyskanych z zespołów NMR, to w niektórych przypadkach mogą mieć niskie przewidywania. Rozkłady współczynników korelacji r_p i r_s okazują się nie być znacząco skośne w przypadku struktur białek uzyskanych z rentgenografii X-ray (rysunek 4.5 C, D). W ich przypadku, wartości skośności wyglądają następująco: γ dla współczynnika korelacji r_p wyglądają następująco: UNRES-flex = $-0,02$, UNRES-DSSP-flex = $-0,46$, CABS-flex = $-0,31$ i NOLB = $-0,14$. W przypadku współczynnika korelacji r_s : UNRES-flex = $-0,06$, UNRES-DSSP-flex = $-0,21$, CABS-flex = $-0,21$ i NOLB = $0,13$. Zauważyłem, że w przeciwieństwie do rozkładów z metody NOLB, maksymalne rozkłady metody CABS-flex, UNRES-DSSP-flex i UNRES-flex są przesunięte w lewo w porównaniu z odpowiadającym strukturom białek uzyskanych z zespołów NMR. Ta cecha

prawdopodobnie wynika z przewidywania zmian w izolowanych cząsteczkach białka podczas pakowania ich w struktury krystaliczne.

4.2.4 Analiza dystrybuanty rozkładów

Poprzednia analiza pozwoliła mi określić, jaka jest skośność każdego z rozkładu współczynników korelacji oraz potwierdzić jego modalność. Oznacza to, że opierając się tylko i wyłącznie na średnich współczynnika korelacji pomiędzy teoretycznymi a eksperymentalnymi profilami fluktuacji można wyciągnąć błędne wnioski, a także niedostrzec istotnych informacji. W związku z tym przeprowadziłem analizę dystrybuanty rozkładów, których wyniki przedstawiłem na rysunku 4.6 A-D. Wartość dystrybuanty rozkładu x określiłem jako liczbę struktur, dla których x nie przekracza odpowiedniego współczynnika korelacji. W przypadku współczynników korelacji dla struktur białek uzyskanych z zespolów NMR, krzywe odpowiadające metodzie NOLB są znacznie przesunięte w kierunku niższych wartości od krzywych odpowiadających trzem pozostałym metodom przewidywania fluktuacji, co można zaobserwować na rysunku 4.6 A i B. Oznacza to, że metoda NOLB nie powinna być metodą stosowaną do przewidywania fluktuacji struktur uzyskanych z zespolów NMR, co potwierdza wniosek uzyskany z wcześniejszych analiz. Na podstawie przeprowadzonych dystrybuant w przypadku struktur uzyskanych z zespolów NMR można określić, że metoda CABS-flex klasyfikuje się jako metoda dająca najbardziej satysfakcjonujące wyniki, za nią klasyfikuje się metoda UNRES-DSSP-flex, a tuż za nią metoda UNRES-flex. Na podstawie tych informacji wnioskuję, że do przewidywania profili fluktuacji struktur białek uzyskanych z zespolów NMR, najlepszymi metodami okazują się CABS-flex oraz UNRES-DSSP-flex.

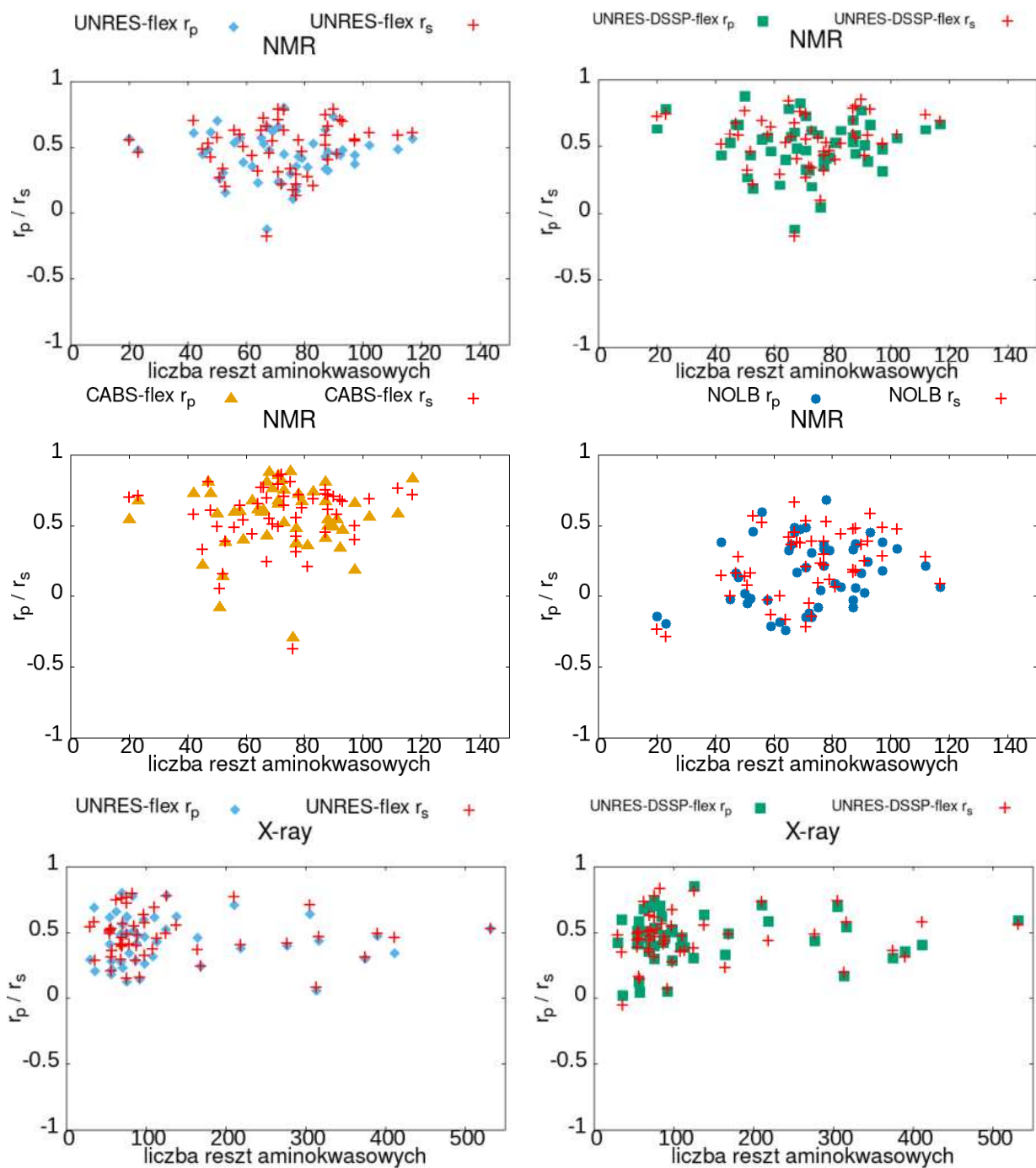


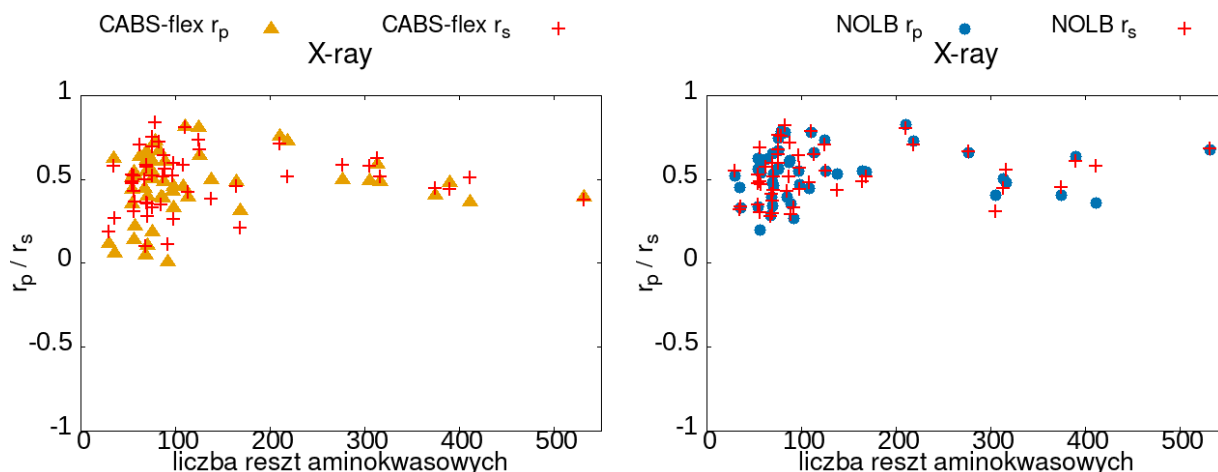
Rysunek 4.6: Dystrybuanty rozkładu średnich wartości współczynników korelacji Pearsona (r_p) i Spearmana (r_s) pomiędzy profilami fluktuacji uzyskanymi z zespołów NMR lub współczynników B promieniowania rentgenowskiego a profilami przewidzianymi przez UNRES-flex (stalowoniebieska linia i romby), UNRES-DSSP-flex (zielona linia i kwadraty), CABS-flex (pomarańczowa linia i trójkąty) oraz NOLB (niebieska linia i kropki) po usunięciu końcowych reszt aminokwasowych analizowanych białek. Adapted with permission from Ł. J. Dziadek, A. K. Sieradzan, C. Czaplewski, M. Zalewski, F. Banaś, M. Toczek, W. Nisterenko, S. Grudinin, A. Liwo, A. Giełdoń, Assessment of Four Theoretical Approaches to Predict Protein Flexibility in the Crystal Phase and Solution, *J. Chem. Theory Comput.* 2024, 20, 17, 7667–7681. Licensed under CC-BY 4.0

W przypadku struktur białek uzyskanych z rentgenografii Xray widoczne jest, że wyniki współczynników korelacji uzyskane metodą NOLB są niewiele przesunięte w kierunku wyższych wartości w porównaniu do wyników uzyskanych z pozostałych metod, co wskazuje, że metoda NOLB sprawdza się najlepiej w przypadku tych struktur. Ponieważ metoda NOLB okazała się być statystycznie lepsza tylko dla struktur uzyskanych z rentgenografii Xray białek o strukturze $\alpha + \beta$ (rysunek 4.4 C), moje obserwacje potwierdzają wnioskami z poprzedniej analizy. Jednak w porównaniu do trzech metod, najniższe współczynniki korelacji dla NOLB zaczynają się od około 0,2 dla struktur uzyskanych z rentgenografii Xray, podczas gdy dla pozostałych trzech metod zaczynają się od 0. Pomimo tego, że metoda NOLB była najlepsza spośród analizowanych metod, to metoda UNRES-DSSP-flex oraz CABS-flex przewidyuje profile fluktuacji z porównywalną dokładnością, a tuż za nimi metoda UNRES-flex.

4.2.5 Zależność współczynników korelacji od wielkości białka

W kolejnym etapie analizy określiłem, czy dokładność przewidywania elastyczności białka zależy od długości jego łańcucha. Analizę przeprowadziłem na podstawie średnich wartości r_p i r_s dla struktur skróconych białek dla każdej z czterech metod. Analizowałem oddzielnie dla struktur uzyskanych z zespołów NMR i rentgenografii Xray. Struktury uzyskane z zespołów NMR mają 20–117 reszt aminokwasowych, a struktury uzyskane z rentgenografii Xray 30–532 reszt (po obcięciu końcowych reszt aminokwasowych). Na podstawie uzyskanych wyników (rysunek 4.7) mogę stwierdzić, że w przeprowadzonej analizie nie ma zależności między długością łańcucha białek a współczynnikami korelacji r_p lub r_s . Dla wszystkich metod oprócz metody UNRES-flex, współczynniki korelacji są bardziej skoncentrowane wokół wartości $\sim 0,5$ dla łańcuchów zawierających więcej niż 200 reszt aminokwasowych, co jest najbardziej widoczne w przypadku metody CABS-flex dla struktur uzyskanych z rentgenografii Xray.

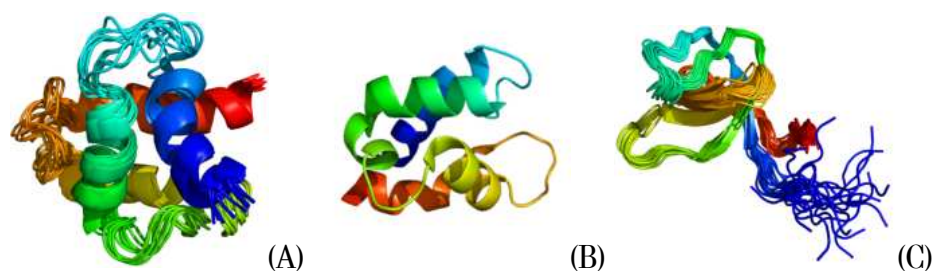




Rysunek 4.7: Zależność pomiędzy średnimi wartościami współczynników korelacji Pearsona i Spearmana obliczonych na podstawie profili fluktuacji reszt aminokwasowych białek po usunięciu końcowych reszt aminokwasowych uzyskanych z symulacji: UNRES-flex, UNRES-DSSP-flex, CABS-flex, oraz NOLB a liczbą reszt aminokwasowych w zestawie badanych białek. Adapted with permission from Ł. J. Dziadek, A. K. Sieradzan, C. Czaplewski, M. Zalewski, F. Banaś, M. Toczek, W. Nisterenko, S. Grudinin, A. Liwo, A. Gieldoń, Assessment of Four Theoretical Approaches to Predict Protein Flexibility in the Crystal Phase and Solution, J. Chem. Theory Comput. 2024, 20, 17, 7667–7681. Licensed under CC-BY 4.0

4.3 Analiza profili RMSFN dla wybranych białek

Następnie przeprowadziłem analizę dla trzech przykładowych białek, ze względu na ich typ struktury oraz długość sekwencji aminokwasowej. W tym celu zastosowałem cztery różne metody do określenia ich zdolności do przewidywania profili fluktuacji badanych białek. Do przeprowadzonej analizy wybrałem następujące białka: poliproteiny domeny M retrowirusa z wirusa mięsaków Rousa (o kodzie PDB: 1A6S, posiadające wszystkie struktury drugorzędowe α -helikalne, struktura uzyskana z zespołów NMR)²⁴³, białko wiążące wapń z jelita bydlęcego (o kodzie PDB: 3ICB, posiadające wszystkie struktury drugorzędowe α -helikalne, struktura uzyskana z rentgenografii X-ray)²⁴⁴ oraz trzecia domena SH3 białka adaptorowego Cin85 (o kodzie PDB: 2K9G, posiadające wszystkie struktury drugorzędowe β -harmonijek, struktura uzyskana z zespołów NMR)²⁴⁵. Struktury analizowanych białek przedstawiłem na rysunku 4.8 od A do C, a ich profile fluktuacji na rysunku 4.9 od A do C. W tabeli 7.11 w załącznikach przedstawiłem średnie wartości r_p oraz r_s uzyskanych pomiędzy profilami fluktuacji z eksperymentów, a profilami fluktuacji uzyskanych z zastosowanych metod dla obciętych struktur białek.



Rysunek 4.8: Struktury analizowanych białek (lub zbiorów konformacyjnych białek) w reprezentacji wstążkowej, na których przedstawiono: (A) poliproteinę domeny M retrowirusa z wirusa mięsaków Rousa (PDB: 1A6S, wszystkie- α , struktura NMR)²⁴³, (B) białko wiążące wapń zależne od witaminy D z jelita bydlęcego (PDB: 3ICB, wszystkie- α , struktura rentgenowska)²⁴⁴ i (C) trzecią domenę SH3 białka adapterowego Cin85 (PDB: 2K9G, wszystkie- β , struktura NMR)²⁴⁵. Reprinted with permission from Ł. J. Dziadek, A. K. Sieradzan, C. Czaplewski, M. Zalewski, F. Banaś, M. Toczek, W. Nisterenko, S. Grudinin, A. Liwo, A. Gieldoń, Assessment of Four Theoretical Approaches to Predict Protein Flexibility in the Crystal Phase and Solution, J. Chem. Theory Comput. 2024, 20, 17, 7667–7681. Licensed under CC-BY 4.0

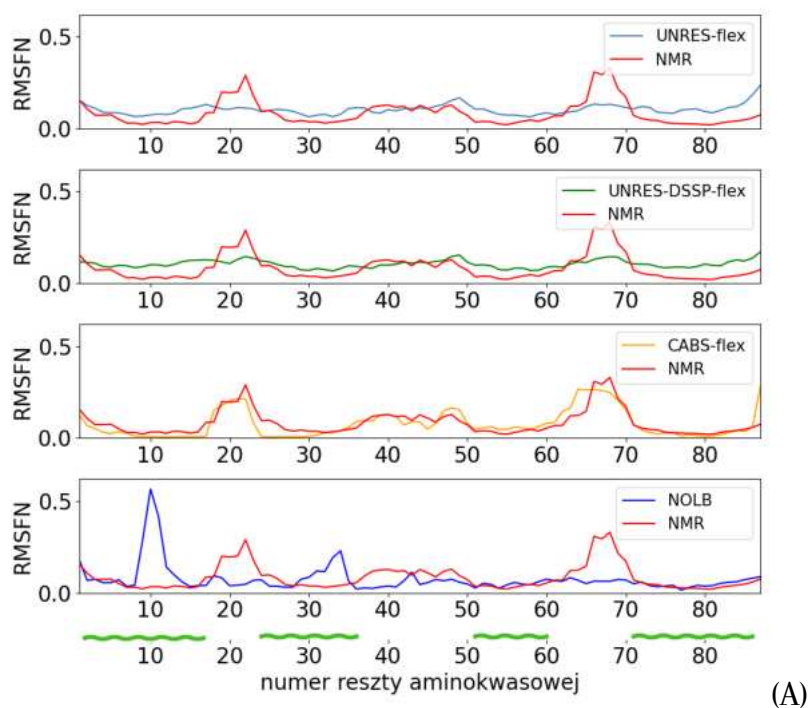
Struktura NMR białka α -helikalnego przedstawiłem na rysunku 4.9 A (kod PDB: 1A6S). Profil RMSFN, który został obliczony za pomocą NOLB, znacznie odbiega od profilu określonego na podstawie zespołów NMR. Obecność wysokiej wartości fluktuacji w N-końcowej części profilu NOLB, którego nie ma w zespole NMR, jest przyczyną niskich współczynników korelacji r_p i r_s (odpowiednio 0,13 i 0,15, tabela 7.11 w załącznikach). Odwrotną sytuację można zaobserwować w obszarze C końca profilu RMSFN z metody NOLB, gdzie nie występuje przeszacowanie wartości fluktuacji w porównaniu do wyników z pozostałych metod. Profile fluktuacji uzyskane z metod UNRES-flex, UNRES-DSSP-flex i CABS-flex są zbieżne z profilami zespołów NMR. W obszarze pierwszej i ostatniej pętli pomiędzy strukturami drugorzędowymi α widoczne jest niedoszacowane wartości fluktuacji dla wszystkich metod, z wyjątkiem metody CABS-flex, co wskazuje, że ta metoda wykazuje najlepsze dopasowanie profilu fluktuacji względem eksperymentalnego.

Drugim białkiem (rysunek 4.9 B), które szczegółowo analizowałem, jest struktura uzyskana z rentgenografii X-ray białka α -helikalnego (kod PDB: 3ICB). W tym przypadku profil RMSFN uzyskany z metody NOLB jest bardziej zgodny z profilem obliczonym na podstawie współczynników B, co znajduje odzwierciedlenie w wyższych wartościach współczynników korelacji ($r_p=0,56$ i $r_s=0,68$). W związku z tym, że obserwuje się przeszacowanie wartości fluktuacji w przypadku metod UNRES-flex, UNRES-DSSP-flex oraz CABS-flex w obszarze 18-tej reszty aminokwasowej, to w porównaniu do nich metoda NOLB wykazuje najlepsze odwzorowanie danych eksperymentalnych. Można za-

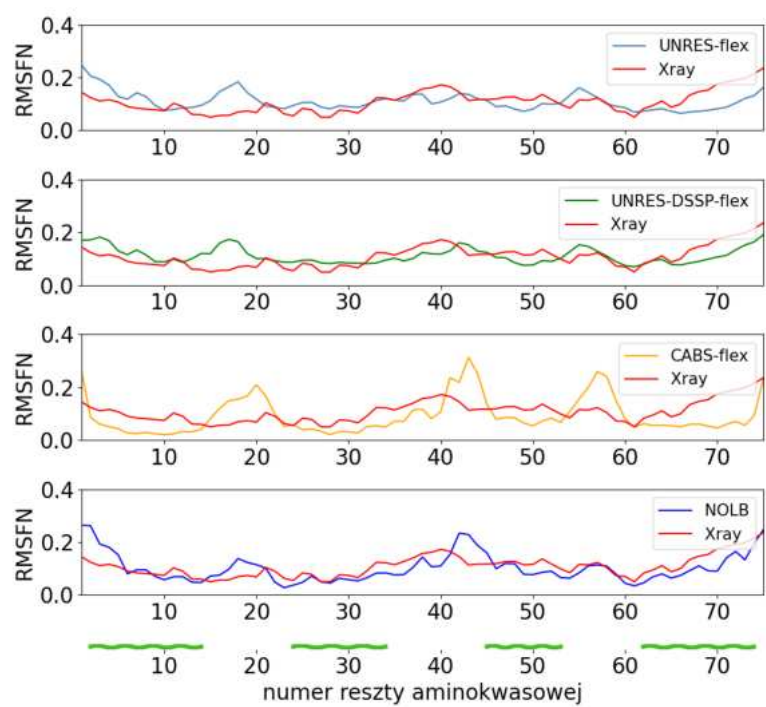
uważyć, że metoda CABS-flex przeszacowuje zakres wartości fluktuacji w okolicach 20, 43 oraz 55 reszt aminokwasowych, to przeszacowanie jest widoczne przede wszystkim w pętach pomiędzy drugorzędowymi strukturami α .

W większości przypadków metoda CABS-flex ma tendencję do dobrego przewidywania fluktuacji struktur drugorzędowych białek, a także w obszarze pętli pomiędzy nimi. Jest to pozytywna cecha, jaką może zaoferować w przeprowadzonych badaniach metodą CABS-flex, ale zdarzają się sytuacje, w których ta metoda może prowadzić do błędnych przewidywań wartości fluktuacji. Przykładem tego jest białko o strukturze NMR, białka β -harmoniki, którego profile fluktuacji przedstawiłem na rysunku 4.9 (kod PDB: 2K9G). W przeciwieństwie do innych metod, metoda CABS-flex przeszacował wartości fluktuacji dla pętli pomiędzy drugorzędowymi strukturami β w okolicy 40-50 reszty aminokwasowej (w środkowej części łańcucha). To spowodowało, że uzyskane wartości współczynników r_p i r_s (odpowiednio 0,52 i 0,71) są niższe. Metody UNRES-flex i UNRES-DSSP-flex uzyskały znaczne wyższe wartości r_p (tabela 7.11 w załącznikach).

Profile RMSFN dla obciętej struktury 1A6S

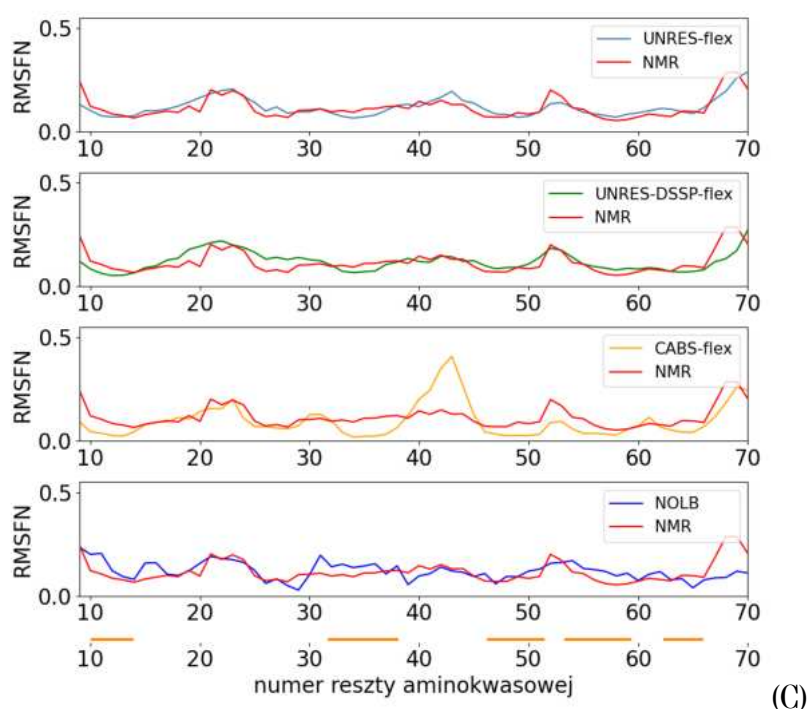


Profil RMSFN dla obciętej struktury 3ICB



(B)

Profile RMSFN dla obciążonej struktury 2K9G



(C)

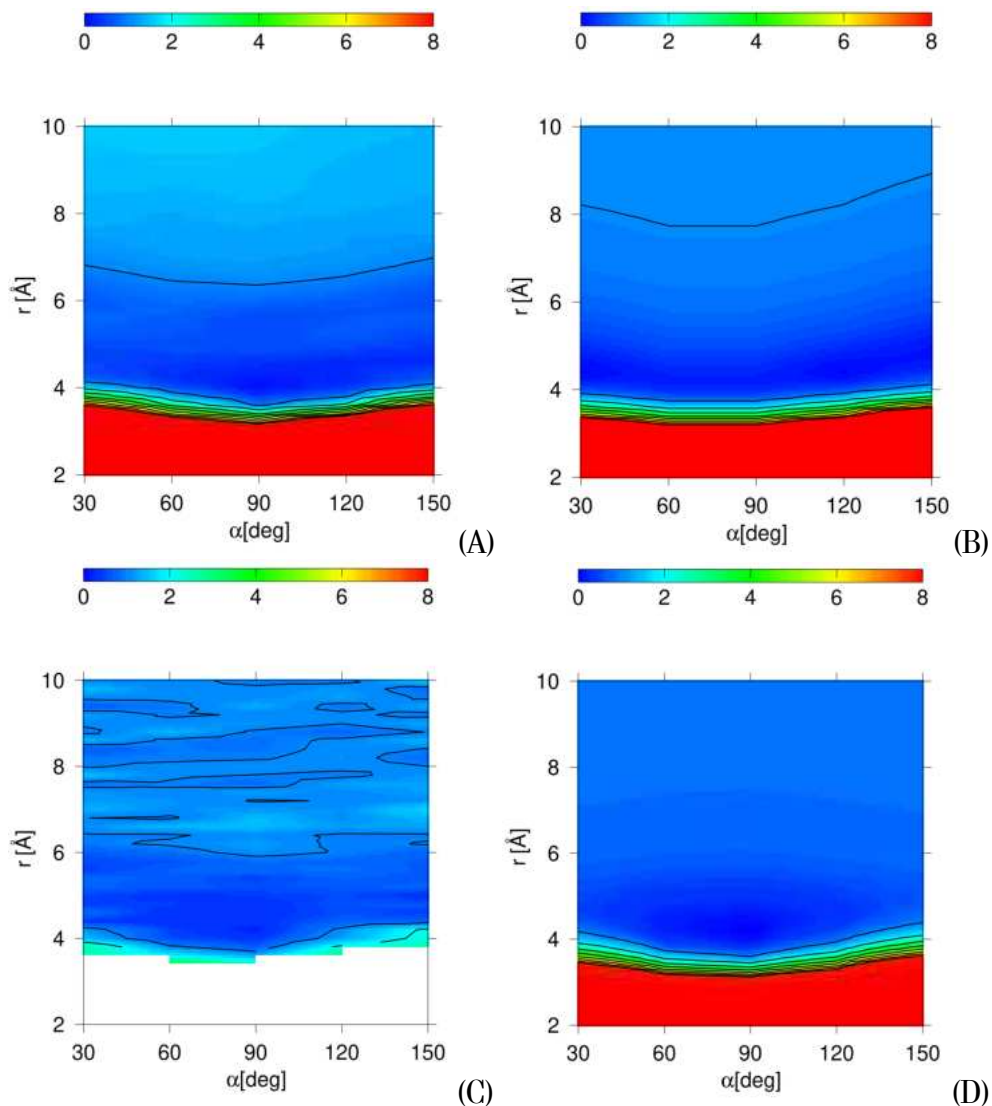
Rysunek 4.9: Profile RMSFN dla skróconych struktur białek o kodach PDB: 1A6S, 3ICB i 2K9G. Profile fluktuacji uzyskane z trzech symulacji metodą UNRES-flex zaznaczono jasnoniebieską linią, UNRES-DSSP-flex zieloną linią, CABS-flex żółtą linią, NOLB (pojedyncza symulacja) niebieską linią, zaś eksperymentalne (Xray lub NMR) za pomocą czerwonej linii. Struktury drugorzędowe zaznaczono poniżej wykresów za pomocą zielonej falowanej linii (dla α -helis) lub prostej pomarańczowej linii (dla β -harmonijek). Adapted with permission from Ł. J. Dziadek, A. K. Sieradzan, C. Czaplewski, M. Zalewski, F. Banaś, M. Toczek, W. Nisterenko, S. Grudinin, A. Liwo, A. Giełdoń, Assessment of Four Theoretical Approaches to Predict Protein Flexibility in the Crystal Phase and Solution, J. Chem. Theory Comput. 2024, 20, 17, 7667–7681. Licensed under CC-BY 4.0

4.4 Analiza właściwości PMF oraz ocena dopasowania funkcji analitycznych dla układów cząsteczek UNRES-MARTINI

W kolejnym etapie tej pracy przeprowadziłem symulacje za pomocą półempirycznej metody mechaniki kwantowej PM7 i dynamiki molekularnej za pomocą pola siłowego ff14SB w oprogramowaniu AMBER. Badanie oddziaływań między cząsteczkami jest ważne, ponieważ pomaga zrozumieć, jak zachowują się cząsteczki i jak oddziałują na siebie, aby uzyskać PES, na podstawie którego obliczyłem PMF. Do przeprowadzonych obliczeń zastosowałem 20 centrów UNRES oraz 18 centrów MARTINI, co daje razem 360 kombinacji. Gdybym chciał opisać każde oddziaływanie, to praca byłaby niezwykle obszerna. Z tego względu, zdecydowałem opisać trzy różne rodzaje interakcji ze względu na różny charakter chemiczny tych oddziaływań; pomiędzy dwiema cząsteczkami hydrofobowymi (rysunek 4.10), dwiema cząsteczkami polarnymi (rysunek 4.11) i dwiema naładowanymi cząsteczkami o różnym ładunku (rysunek 4.12). Następnie przeprowadziłem dopasowanie funkcji analitycznych do uzyskanych potencjałów średnich sił, czego celem było odwzorowanie najistotniejszych interakcji, jednocześnie umożliwiając policzenie pochodnej funkcji (warunek konieczny do implementacji do dynamiki molekularnej). Dodatkowo dopasowane funkcje analityczne ułatwiają interpretację powierzchni PMF. W ramach tego projektu przedstawiłem uzyskane parametry funkcji analitycznych opisujące interakcje pomiędzy cząsteczkami UNRES i MARTINI, aby wprowadzić błonę lipidową w formie jawnej, co było motywacją do przeprowadzenia analiz.

Dla profilu energetycznego PMF w przypadku modelu łańcucha bocznego alaniny (cząsteczki w reprezentacji modelu UNRES) i 1-propanitiolu (cząsteczki w reprezentacji modelu MARTINI) minimum globalne znajduje się w okolicy $r = 3,5 \text{ \AA}$ oraz kąta $\alpha = 90^\circ$ dla wyników uzyskanych metodą PM7, a także dla profili PMF uzyskanych z symulacji pełnoatomowej (rysunek 4.10). W porównaniu z nieco bardziej "chropowatą" powierzchnią energii potencjalnej (zawierającą wiele płytkich minimumów lokalnych) uzyskaną metodą pełnoatomową, profil energii wzdłuż współrzędnej r jest gładniejszy dla metody PM7. Może to być spowodowane zastosowaniem niejawnego modelu rozpuszczalnika w przeprowadzonych obliczeniach metodą PM7. Dopasowana funkcja analityczna oparta na potencjale Gay-Berne dokładnie oddaje ogólną topologię powierzchni PMF oraz położenie i głębokości minimum

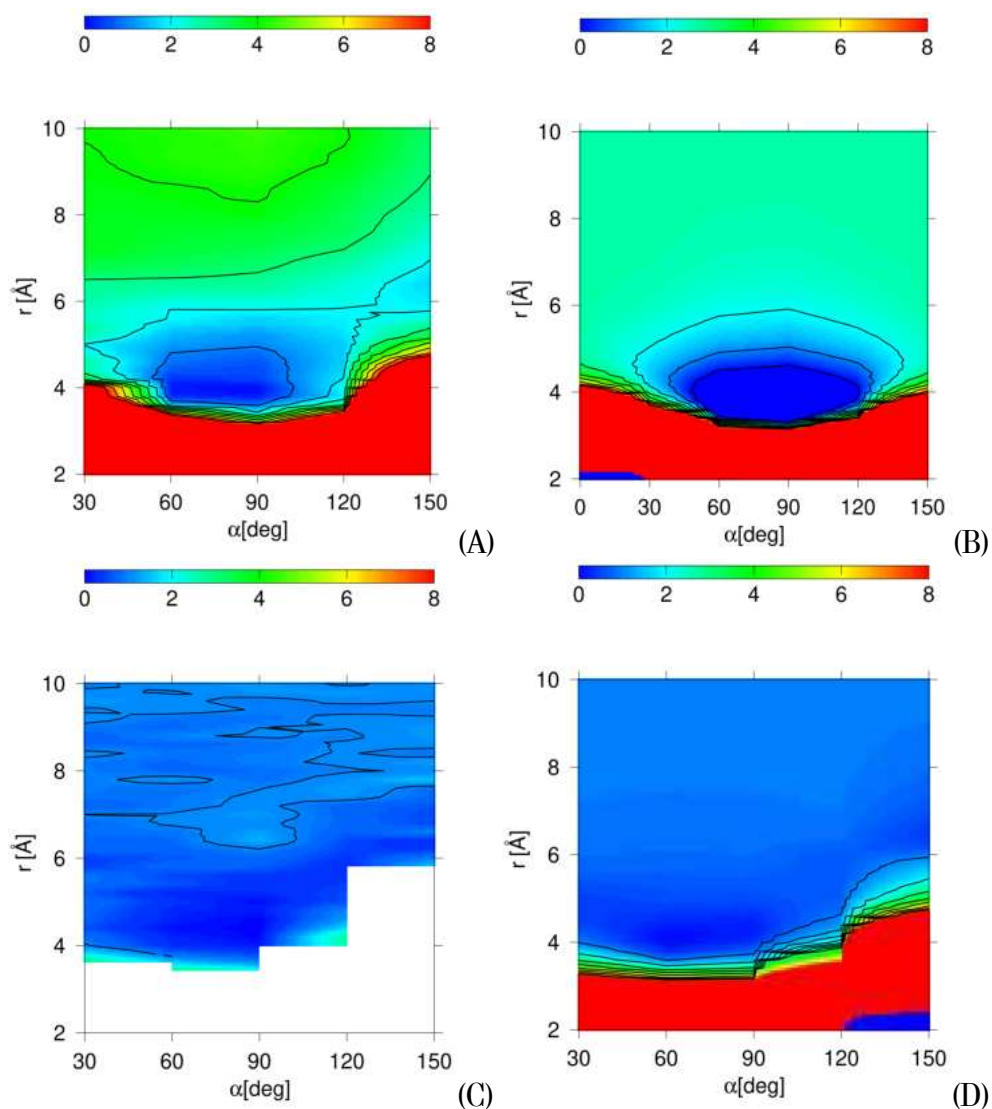
globalnego dla obu metod.



Rysunek 4.10: Potencjały średniej siły dla oddziaływania pomiędzy modelem łańcucha bocznego alaniny i 1-propanotolu otrzymane za pomocą metody PM7 (A) i symulacji pełnoatomowej (C) oraz dopasowane funkcje analityczne do odpowiednich PMF otrzymanych za pomocą metody PM7 (B) i symulacji pełnoatomowej (D). Nad wykresem została przedstawiona skala energii (kcal/mol).

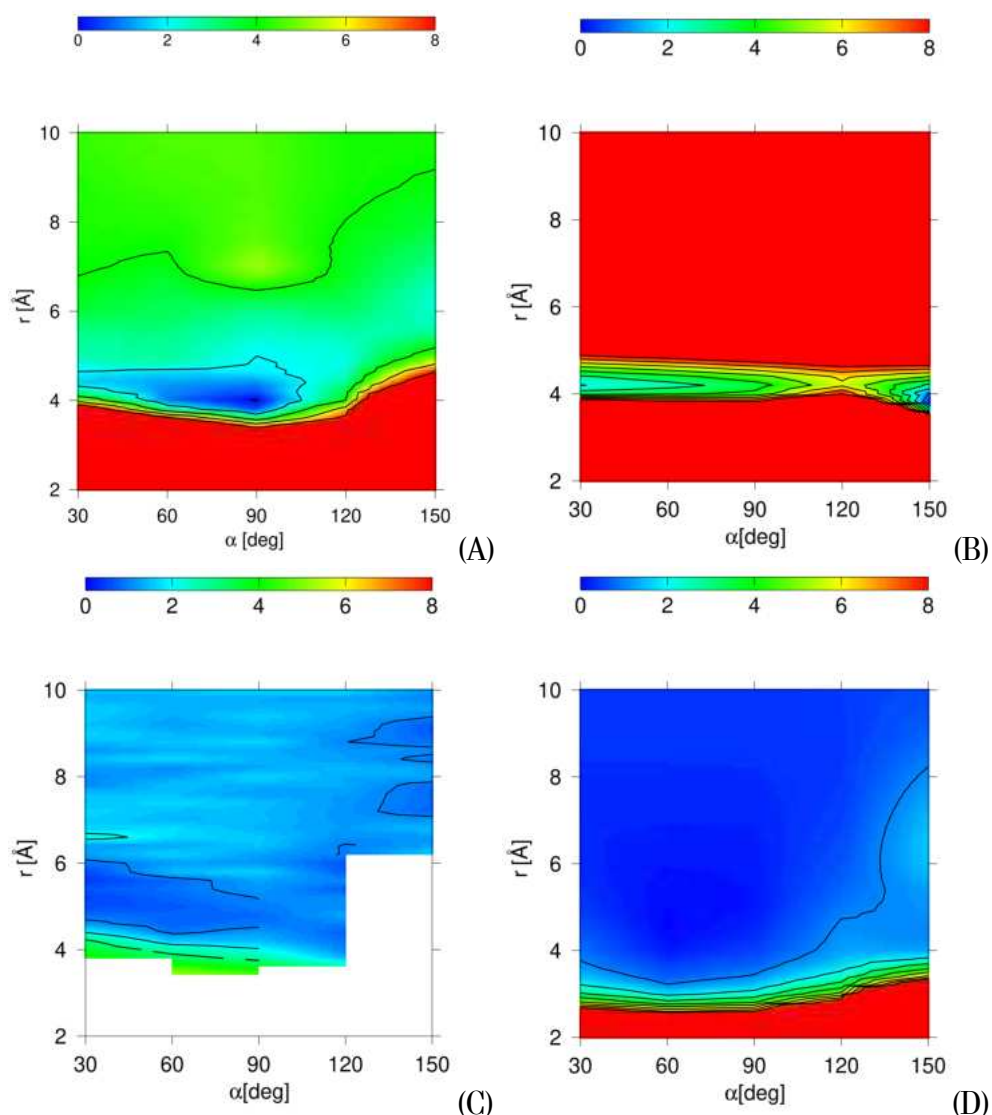
Dla wyników uzyskanych za pomocą metody PM7, profil energetyczny PMF modelu łańcucha bocznego histydyny (cząsteczki w reprezentacji modelu UNRES) i 1-propyloaminy (cząsteczki w reprezentacji modelu MARTINI) wykazuje minimum globalne w okolicy $r = 4,0 \text{ Å}$ oraz kąta $\alpha = 90^\circ$ oraz drugie minimum lokalne w okolicy $r = 6,5 \text{ Å}$ oraz kąta $\alpha = 150^\circ$ (rysunek 4.11 A). W przypadku PMF, który został uzyskany w symulacji pełnoatomowej, minimum globalne wynosi $r = 4,5 \text{ Å}$ oraz kąt $\alpha = 90^\circ$, zaś drugie minimum lokalne wynosi $r = 6,0 \text{ Å}$ oraz kąt $\alpha = 150^\circ$ (rysunek 4.11 C). Oddziały-

wania cząsteczek polarnych mają dwa różne minima, niezależnie od metody, w przeciwieństwie do oddziaływań hydrofobowych. Głębokość minimum dla metody PM7 z niejawnym rozpuszczalnikiem jest znacznie większe. W obu metodach (PM7 i pełnoatomowej) jest widoczna asymetria w hiperpowierzchni energii potencjalnej, szczególnie objawia się w obszarze zakazanym sterycznie dla r w okolicy $4,5 \text{ \AA}$ oraz $\alpha = 150^\circ$. Dla metody pełnoatomowej, dopasowane funkcje analityczne do PMF poprawnie odwzorowały i zachowały głębokość minimum. Niestety w przypadku metody PM7, drugie co do głębokości minimum (lokalne) nie jest dobrze odzwierciedlone i nie jest uwidocznione w dopasowanej funkcji analitycznej ze względu na to, że pierwsze minimum (globalne) jest zbyt głębokie w stosunku do głębokości drugiego. Bardzo głębokie minimum globalne może sprawić, że dopasowanie funkcji analitycznych nie uwzględni płytszych minimów lokalne. W efekcie drugie minimum (lokalne) będzie mniej dokładnie odwzorowane.



Rysunek 4.11: Potencjały średniej siły dla oddziaływania pomiędzy modelem łańcucha bocznego histydyny i 1-propylaminy otrzymane za pomocą metody PM7 (A) i symulacji pełnoatomowej (C) oraz dopasowane funkcje analityczne do odpowiednich PMF otrzymanych za pomocą metody PM7 (B) i symulacji pełnoatomowej (D). Nad wykresem została przedstawiona skala energii (kcal/mol).

Dla profilu energetycznego PMF, w przypadku modelu łańcucha bocznego kwasu glutaminowego (cząsteczki w reprezentacji modelu UNRES) i choliny (cząsteczki w reprezentacji modelu MARTINI), minimum globalne znajduje się w okolicy $r = 4,0 \text{ Å}$ oraz kąta $\alpha = 90^\circ$ dla wyników uzyskanych metodą PM7, a także dla profili PMF uzyskanych z symulacji pełnoatomowej (rysunek 4.12). W przypadku dopasowania funkcji analitycznych do PMF metodą PM7 widoczne jest znaczne przeszacowanie wkładu kolumbowskiego uzyskując mniej dokładny opis głębokości minimum globalnego, aczkolwiek jego położenie w okolicy 4 Å jest poprawne.



Rysunek 4.12: Potencjały średniej siły dla oddziaływania pomiędzy modelem łańcucha bocznego kwasu glutaminowego i choliny otrzymane za pomocą metody PM7 (A) i symulacji pełnoatomowej (C) oraz dopasowane funkcje analityczne odpowiednich PMF otrzymanych za pomocą metody PM7 (B) i symulacji pełnoatomowej (D). Nad wykresem została przedstawiona skala energii (kcal/mol).

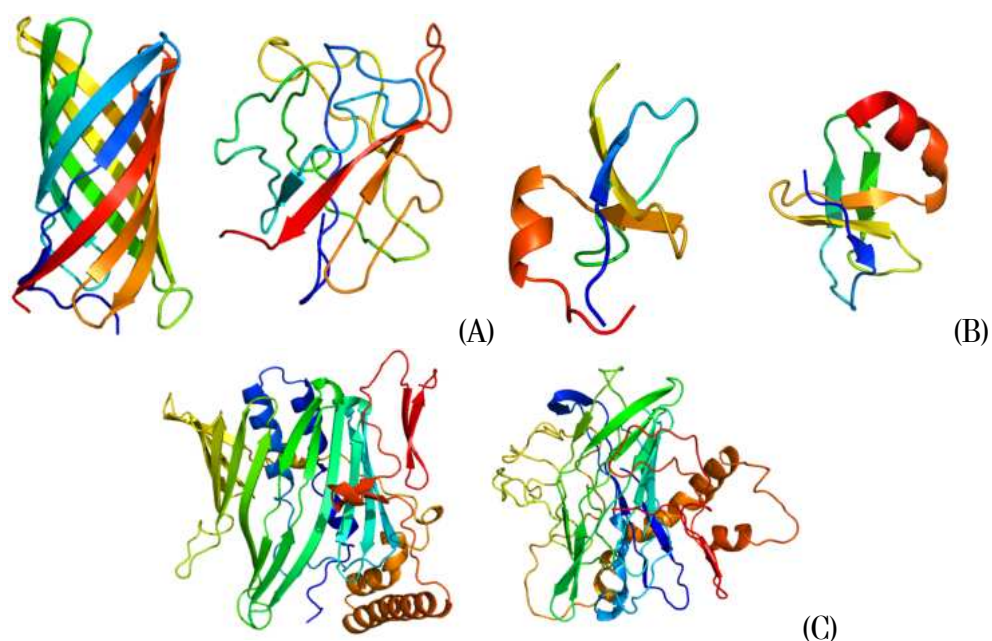
Dopasowanie funkcji analitycznych do profili energetycznych PMF pary cząsteczek polarnych substancji rozpuszczonych wymaga uwzględnienia wielu składników energetycznych, nie tylko potencjału Gay-Berne, który odzwierciedla anizotropowe właściwości cząsteczek i właściwości steryczne, ale także wkłady energetyczne opisujące udział wiązań wodorowych lub oddziaływania dipol-dipol. Funkcje analityczne dopasowane do profili PMF dla symulacji pełnoatomowych i metody PM7 wygładzają hiperpowierzchnie, jednocześnie zachowując najważniejsze cechy. PMF dla pary cząsteczek naładowanych zawiera wiele składowych do energii. Najważniejsze oddziaływania

mogą być modelowane przez potencjał Gay-Berne’a, a także elektrostatykę, która obejmuje zarówno bezpośrednie siły kulombowskie, jak i efekt solwatacji. Wkłady do energii opisują najważniejsze cechy hiperpowierzchni, biorąc pod uwagę właściwości dielektryczne wody. Należy zauważyć, że znaczne przesunięcie naładowanego centrum w modelowej cząsteczce UNRES również skutkuje pogorszeniem opisu oddziaływań cząsteczek naładowanych, podczas gdy dopasowanie funkcji analitycznych do PMF uzyskanych metodą pełnoatomową poprawnie odwzorowało minimum globalne i wygładziło funkcje PMF.

4.5 Analiza przewidywanych względem eksperymentalnych profili fluktuacji białek otoczonych błonami lipidowymi

Dla czterech różnych wag (sekcja 3.4.1 w metodach) 0,5 do 4,0 wraz z zestawem 4 różnych parametrów przetestowałem, która waga sprawdziła się najlepiej w badaniu analizowanych białek w polu siłowym UNRES-MARTINI. Obliczyłem TM-score i profile RMSD dla 3 testowanych białek, których wyniki dla różnych testowanych wag przedstawiłem w tabelach 4.1-4.3.

Przeprowadziłem analizę dla 3 wybranych białek ze względu na pełnione funkcje biologiczne : de novo zaprojektowana transbłonowa β -beczka (kod PDB: 6X9Z, wszystkie- β , struktura X-RAY)²⁴⁶, który jest białkiem błonowym, (B) Domena C1B kinazy białkowej C w kompleksie z diacyloglicerol-laktonem AJH-836 (PDB: 7LF3, α - β , struktura X-RAY)²⁴⁷, czyli białko w kompleksie z pojedynczym lipidem oraz (C) Domeny ORD ludzkiego ORP8 (PDB: 8P7A , α - β , struktura X-RAY)²⁴⁸ będący białkiem peryferyjnym. Przeprowadziłem symulacje w błonach lipidowych reprezentowanych potencjałem MARTINI. Struktury eksperymentalne oraz uzyskane z symulacji badanych białek zostały przedstawione na rysunku 4.13. W kolejnej analizie przeprowadziłem symulacje, które miały na celu wprowadzenie pola siłowego UNRES-MARTINI, aby określić, w jakim stopniu nowe pole siłowe potrafi odtworzyć eksperymentalne profile fluktuacji białek otoczonych błoną lipidową. Obliczyłem współczynniki korelacji pomiędzy teoretycznymi a eksperymentalnymi profilami fluktuacji dla skróconych struktur białek dla zastosowanej wagi $\omega_{UNRES-MARTINI} = 0,5$ (sekcja 3.4.4 w metodach).



Rysunek 4.13: Wybrane struktury białek z uwzględnieniem typu struktur drugorzędowych (po lewej stronie struktura eksperymentalna, zaś po prawej najbardziej prawdopodobna struktura uzyskana z symulacji UNRES-flex): (A) de novo zaprojektowana transbłonowa β -beczka (o kodzie PDB: 6X9Z, posiadająca wszystkie struktury drugorzędowe- β , struktura uzyskana z rentgenografii Xray) , (B) Domena C1B kinazy białkowej C w kompleksie z diacyloglicerol-laktonem AJH-836 (o kodzie PDB: 7LF3, α - β , struktura uzyskana rentgenografii Xray) oraz (C) Domeny ORD ludzkiego ORP8 (PDB: 8P7A , α - β , struktura uzyskana z rentgenografii Xray) .

Tabela 4.1: Zbiór wartości TM-score oraz RMSD (w nawiasach) dla różnych przetestowanych wag w zależności od zastosowanych metod (pełnoatomowego pola siłowe oraz półempirycznej metody PM7 z lub bez uogólnionego modelu Borna) dla analizowanego białka (kod PDB: 6X9Z).

Weight	All-atom No EGB	All-atom EGB	PM7 No-EGB	PM7 EGB
0.5	0.347 (7.69)	0.264 (12.35)	0.398 (6.91)	0.376 (8.69)
1.0	0.496 (5.11)	0.218 (13.49)	0.396 (6.91)	0.336 (8.89)
2.0	0.524 (5.01)	0.186 (16.52)	0.509 (5.74)	0.294 (9.18)
4.0	0.467 (5.84)	0.192 (16.38)	0.401 (7.43)	0.180 (16.85)

Tabela 4.2: Zbiór wartości TM-score oraz RMSD (w nawiasach) dla różnych przetestowanych wag w zależności od zastosowanych metod (pełnoatomowego pola siłowe oraz półempirycznej metody PM7 z lub bez uogólnionego modelu Borna) dla analizowanego białka (kod PDB: 7FL3).

Weight	All-atom No EGB	All-atom EGB	PM7 No-EGB	PM7 EGB
0.5	0.426 (3.97)	0.350 (5.25)	0.316 (6.46)	0.284 (7.62)
1.0	0.430 (4.50)	0.333 (7.66)	0.279 (8.00)	0.258 (9.45)
2.0	0.278 (5.75)	0.334 (6.64)	0.239 (9.79)	0.269 (7.72)
4.0	0.340 (5.75)	0.334 (9.04)	0.194 (7.47)	0.215 (8.77)

Tabela 4.3: Zbiór wartości TM-score oraz RMSD (w nawiasach) dla różnych przetestowanych wag w zależności od zastosowanych metod (pełnoatomowego pola siłowe oraz półempirycznej metody PM7 z lub bez uogólnionego modelu Borna) dla analizowanego białka (kod PDB: 8P7A).

Weight	All-atom No EGB	All-atom EGB	PM7 No-EGB	PM7 EGB
0.5	0.363 (16.46)	0.161 (25.44)	0.370 (13.64)	0.296 (17.76)
1.0	0.502 (11.89)	0.203 (21.52)	0.339 (14.81)	0.321 (16.04)
2.0	0.437 (11.42)	0.195 (21.70)	0.319 (16.58)	0.254 (17.12)
4.0	0.275 (12.08)	0.275 (15.47)	0.301 (16.30)	0.228 (19.74)

Zaobserwowałem, że dla wagi wynoszącej 2,0 i zestawu parametrów, gdy w funkcji dopasowania nie zastosowałem uogólnionego modelu Borna i parametrów pochodzących od PMF obliczonych przy użyciu pełnoatomowego pola siłowego uzyskałem najlepsze wyniki. W związku z tym, że w przypadku białka (domeny ORD ORP8) nie zaobserwowałem dysocjacji, nawet po bardzo długich symulacjach, musiałem zastosować inną wagę oraz zestaw parametrów podczas przeprowadzania analiz. Gdy nie zastosowałem uogólnionego modelu Borna w funkcji dopasowania i pełnoatomowego pola siłowego, to waga 0,5 oraz zestaw parametrów dla tej wagi okazała się najlepsza. Wynika to z faktu, że białko peryferyjne (kod PDB: 8P7A) wykazało zdolność do oddysocjowania białka od błony, a białko błonowe (kod PDB: 6X9Z) pozostaje stabilne w błonie, zaś białko 7LF3 jest w stanie oddysocjować jak i przyłączyć pojedynczy lipid.

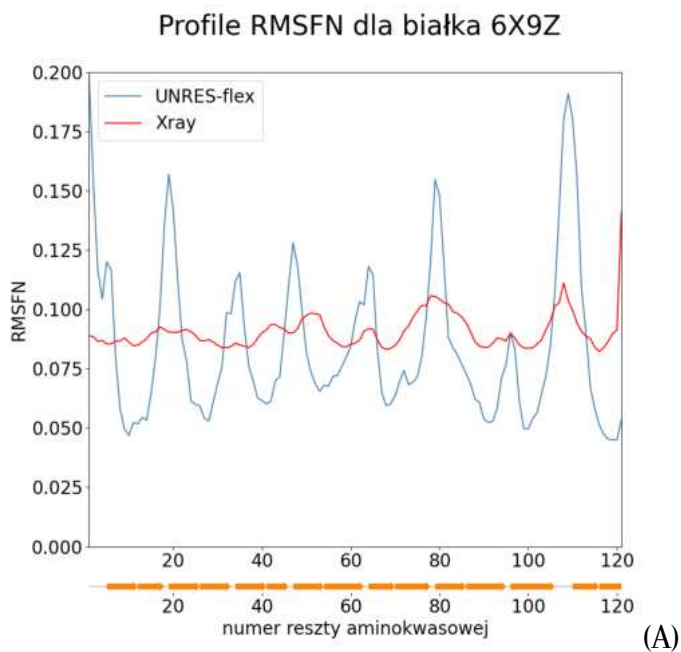
Na rysunku 4.14 A, profil fluktuacji dla drugorzędowej β -harmonijki białka uzyskanego z rentgenografii strukturalnej (kod pdb: 6X9Z). Profil RMSF obliczony metodą UNRES-flex z błoną w formie jawnej wykazuje wyraźne różnice w porównaniu z profilem fluktuacji RMSF pochodzącym z danych eksperymentalnych (Xray) w obszarach każdej pętli pomiędzy drugorzędą β -harmonijką. Odzwierciedla to rozbieżności w średnich wartościach współczynników korelacji Pearsona oraz Spearmana ($r_p = 0,34$ oraz $r_s = 0,44$). W przypadku metody eksperymentalnej jest widoczny pik w regionie N-końcowym, który nie występuje w przypadku UNRES-flex, jest to główny powód rozbieżności w wynikach profili fluktuacji w przypadku białka o kodzie PDB: 6X9Z. Podobnie region C końcowy metody UNRES-flex wykazuje niewielkie przesunięcie w porównaniu z danymi eksperymentalnymi.

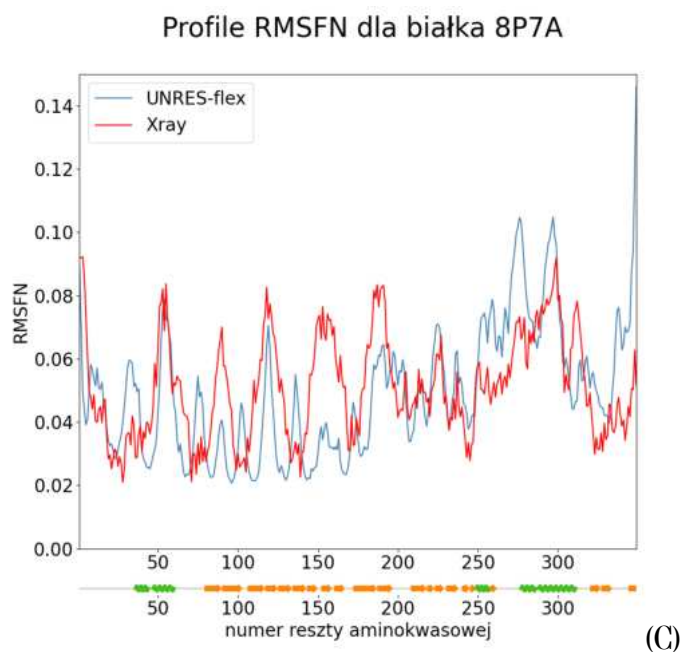
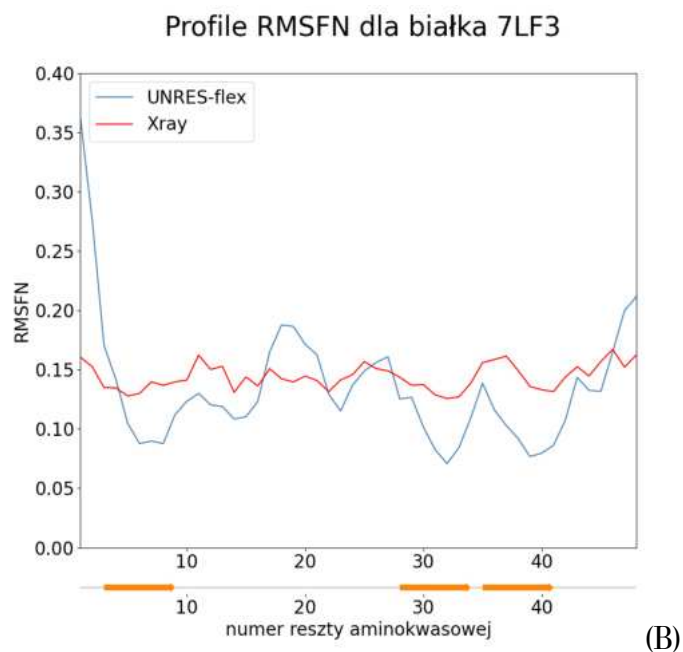
W kolejnym etapie analizowałem fluktuacje białka α -helikalnego (kod PDB: 7LF3), a uzyskane wyniki przedstawiłem na rysunku 4.14 B. W tym przypadku profil RMSF uzyskany za pomocą metody UNRES-flex jest w większym stopniu zgodny z profilem fluktuacji uzyskanym na podstawie współczynników B, co odzwierciedlają wyższe współczynniki korelacji ($r_p=0,51$ i $r_s=0,57$). Oznacza to, że pole siłowe UNRES-MARTINI lepiej przewiduje profile fluktuacji dla białek z pojedynczym lipidem w porównaniu do białek w kompleksie (kod PDB: 6X9Z).

Ostatnim przykładem jest profil RMSF białka o strukturze drugorzędowej $\alpha + \beta$ (gdzie struktura drugorzędowa α została obcięta, ze względu na bardzo wysoką fluktuację w tym obszarze, dlatego nie jest widoczna na wykresie) uzyskanego metodą rentgenografii strukturalnej (kod PDB: 8P7A) przed-

stawilem na rysunku 4.14 C. Metoda UNRES-Flex w tym przypadku zaniża wartości fluktuacji aż do 200-tnej reszty aminokwasowej. Powyżej tej reszty widoczna jest tendencja do przeszacowywania wartości fluktuacji. W porównaniu z danymi eksperymentalnymi największe rozbieżności zaobserwowałem w C-końcowym obszarze profilu RMSE. W wyniku tych niespójności uzyskano średnie wartości współczynników korelacji ($r_p=0,45$ i $r_s=0,42$) dla profili fluktuacji metody UNRES-flex i danych eksperymentalnych. Dodatkowo białko ORP było symulowane z błoną, zaś krystalizowane bez błony, co częściowo może wyjaśniać różnice

Zaimplementowanie UNRES-MARTINI umożliwiło symulowanie białek z lipidami w formie jawnej (ang. explicit). Dokładność przewidywania fluktuacji dla białek błonowych jest na podobnym poziomie co wyjściowa metoda UNRES-flex (dla białek bez lipidów). Pomimo tego, że w tym badaniu analizowałem 3 różne rodzaje białek, to dla każdego z nich zostały dobrze odwzorowane profile fluktuacji, co wnioskuje na podstawie uzyskanych wartości współczynników korelacji (6X9Z: $r_p=0,34$ i $r_s=0,44$; 7LF3: $r_p=0,51$ i $r_s=0,57$; oraz dla 8P7A: $r_p=0,45$ i $r_s=0,42$). Spośród analizowany białek mogę stwierdzić, że w przypadku białka z pojedynczym lipidem (kod pdb: 7LF3) pole siłowe UNRES-MARTINI najlepiej poradziło sobie w odtworzeniu eksperymentalnego profilu fluktuacji.



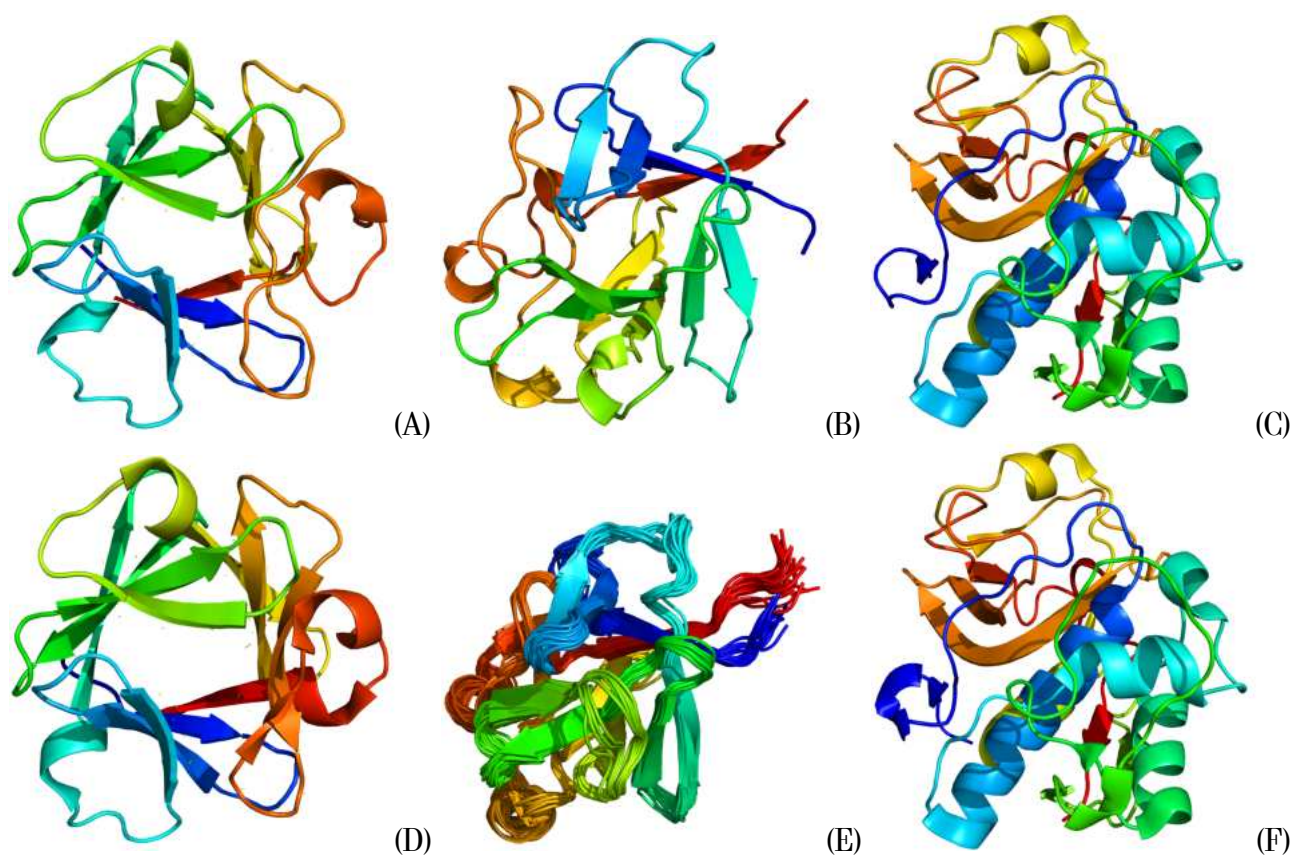


Rysunek 4.14: Profile RMSFN dla skróconych struktur białek (o kodach PDB: 6X9Z (A), 7LF3 (B) i 8P7A (C)). Profile fluktuacji zostały uśrednione z dziesięciu symulacji metodą UNRES-flex i zaznaczone jasnoniebieską linią, zaś eksperymentalne (Xray) za pomocą czerwonej linii. Struktury drugorzędowe zaznaczono poniżej wykresów za pomocą zielonej falowanej linii (dla α -helis) lub prostej pomarańczowej linii (dla β -harmonijek).

4.6 Wpływ wiązania pomiędzy GAGami a białkami na profile fluktuacji reszt aminokwasowych

W ostatnim etapie tej pracy określiłem wpływ wiązania glikozaminoglikanów (GAGów) na elastyczność struktur białkowych, korzystając z dwóch różnych pól siłowych: AMBER oraz CHARMM. Do przeprowadzenia analiz wybrałem trzy przykładowe białka ze względu na ich funkcje biologiczne: FGF-1 i FGF-2 (czynniki wzrostu fibroblastów) oraz Katepsynę K (proteazę). Poprzez wiązanie się z receptorami czynnika wzrostu fibroblastów (FGFR) i ich aktywację, FGF-y pełnią wiele funkcji biologicznych, takich jak różnicowanie, migrację, przeżycie i proliferację komórkową. FGF mogą być wykorzystywane do regeneracji uszkodzonych tkanek, takich jak naczynia krwionośne, mięśnie, tkanki tłuszczowe, ścięgna i więzadła, chrząstki, kości, zębów, nerwów i skóry, dzięki swoim funkcjom biologicznym²⁴⁹. Z kolei katepsyna K jest katepsyną cysteinową w osteoklastach, która uczestniczy w trawieniu macierzy zewnątrzkomórkowej podczas przebudowy kości. Selektywne hamowanie tej katepsyny może być kluczowe do skutecznego leczenia osteoporozy i innych chorób z nadmierną resorpcją kości²⁵⁰.

Symulacje przeprowadziłem dla kompleksów; FGF-2 (PDB: 1BFC, struktura drugorzędowa: $\alpha + \beta$, uzyskana metodą rentgenografii X-ray)²⁵¹ i FGF-1 (PDB: 2AXM, posiadające struktury drugorzędowe $\alpha + \beta$, struktura uzyskana z rentgenografii X-ray)²⁵² z heparyną dp6 oraz Katepsyna K (PDB: 4N8W, posiadające struktury drugorzędowe $\alpha + \beta$, struktura uzyskana z rentgenografii X-ray)²⁵³ z chondroityno-4-siarczanem (C4-S) dp6, oraz form niezwiązanych; FGF-2 (PDB: 1BFG, struktura drugorzędowa: $\alpha + \beta$, uzyskana metodą rentgenografii X-ray)²⁵⁴ i FGF-1 (PDB: 2K43, posiadające struktury drugorzędowe $\alpha + \beta$, struktura uzyskana z zespołów NMR) oraz Katepsyna K (PDB: 5TUN, posiadające struktury drugorzędowe $\alpha + \beta$, struktura uzyskana z rentgenografii X-ray)²⁵⁵. Struktury eksperymentalne badanych białek przedstawiłem na rysunku 4.15



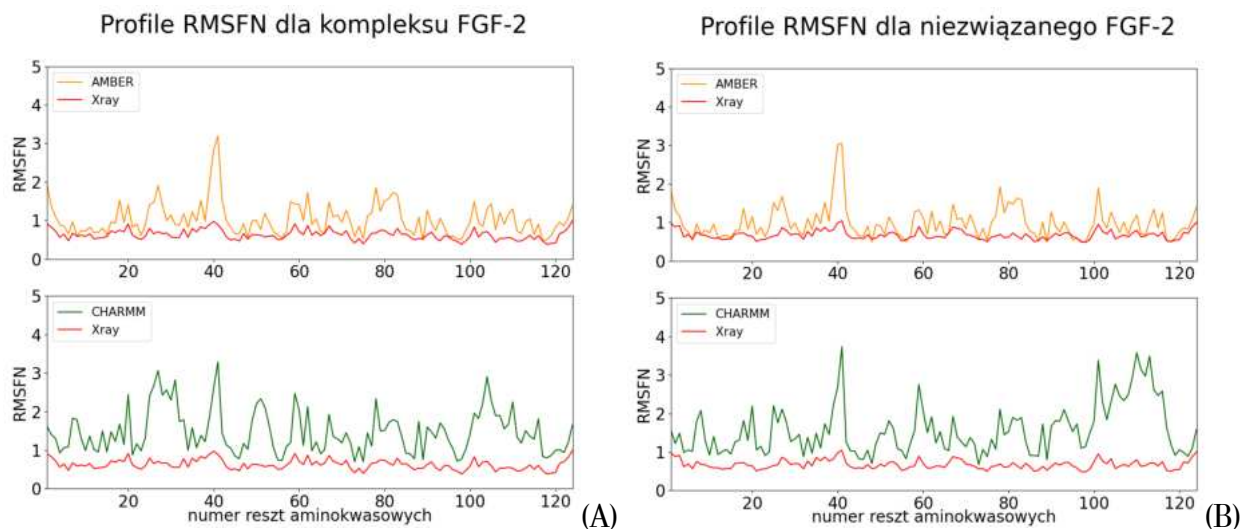
Rysunek 4.15: Struktury analizowanych białek (lub zbiorów konformacyjnych białek) w reprezentacji wstążkowej, na których przedstawiono: (A) FGF-2 (PDB: 1BFC, struktura drugorzędowa: $\alpha + \beta$, uzyskana metodą rentgenografii Xray) z heparyną dp6, (B) FGF-2 (PDB: 2AXM, posiadające struktury drugorzędowe $\alpha + \beta$, struktura uzyskana z rentgenografii Xray) z heparyną dp6, (C) Katepsyny K (PDB: 4N8W, posiadające struktury drugorzędowe $\alpha + \beta$, struktura uzyskana z rentgenografii Xray) oraz w formie niezwiązanej; (D) FGF-2 (PDB: 1BFG, struktura drugorzędowa: $\alpha + \beta$, uzyskana metodą rentgenografii Xray), (E) FGF-1 (PDB: 2K43, posiadające struktury drugorzędowe $\alpha + \beta$, struktura uzyskana z zespołów NMR) oraz (F) Katepsyny K (PDB: 5TUN, posiadające struktury drugorzędowe $\alpha + \beta$, struktura uzyskana z rentgenografii Xray).

Dla każdego pola siłowego przeprowadziłem trzy niezależne symulacje, z których obliczyłem profile fluktuacji. W większości przypadków wyniki były do siebie zbliżone, dlatego zostały uśrednione. Profile fluktuacji uzyskane na podstawie danych eksperymentalnych (z zespołów NMR i rentgenografii Xray) dodatkowo poddałem normalizacji (sekcja 4.3), a uzyskane wyniki r_p i r_s pomiędzy profilami fluktuacji uzyskanymi metodami teoretycznymi a eksperymentalnymi przedstawiłem w tabeli 4.4.

Tabela 4.4: Zestawienie współczynników korelacji Pearsona (r_p) i Spearmana (r_s) pomiędzy przewidywanymi oraz eksperymentalnymi profilami RMSF.

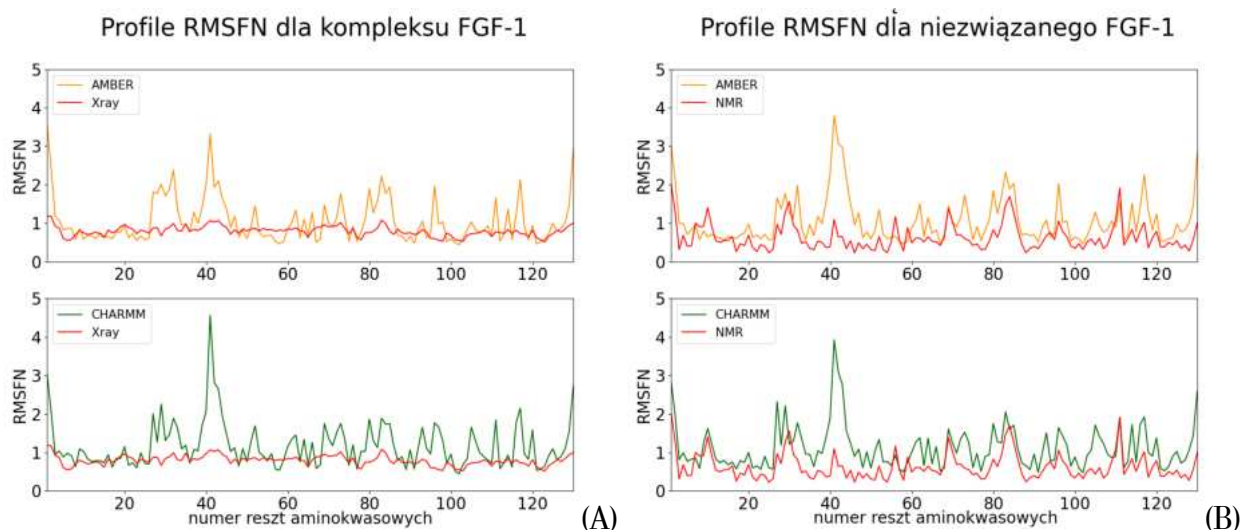
Białko	AMBER kompleks		CHARMM kompleks		AMBER niezwiązany		CHARMM niezwiązany	
	r_p	r_s	r_p	r_s	r_p	r_s	r_p	r_s
FGF-2	0.7308	0.7457	0.4840	0.5105	0.7247	0.7385	0.2559	0.3231
FGF-1	0.6109	0.5325	0.5716	0.5172	0.5346	0.5269	0.5945	0.6219
CatK	0.6243	0.6901	0.5748	0.6296	0.5547	0.6461	0.6655	0.7017

W przypadku kompleksu FGF-2 (PDB: 1BFC, struktura drugorzędowa: $\alpha + \beta$, uzyskana metodą rentgenografii Xray) z heparyną dp6, wysokie wartości r_p i r_s ($r_p = 0,73$ i $r_s = 0,75$) dla pola siłowego AMBER, wnioskuje, że to pole siłowe lepiej odwzorowuje elastyczność badanego białka niż CHARMM ($r_p = 0,48$ oraz $r_s = 0,51$). Różnice współczynników korelacji wynoszą: $\Delta r_p = 0,24$ oraz $\Delta r_s = 0,24$. Wyniki sugerują, że pole siłowe CHARMM przewiduje przeszacowanie profilu fluktuacji w większości fragmentów struktury, a w szczególności w okolicach 30 (pętla pomiędzy 2 a 3 α -helisą), 50 (pierwsza β -harmonijka), 90 (pętla pomiędzy 7 α -helisą a 2 β -harmonijką) oraz 100-110 (pętla pomiędzy 9 α -helisą a 4 β -harmonijką) reszty aminokwasowej, co przedstawiłem na rysunku 4.16. Podobną sytuację zaobserwowałem w przypadku niezwiązanej formy FGF-2 (PDB: 1BFG, struktura drugorzędowa: $\alpha + \beta$, uzyskana metodą rentgenografii Xray). Dużą korelację pomiędzy danymi teoretycznymi a eksperymentalnymi zaobserwowałem dla wyników uzyskanych polem siłowym AMBER ($r_p = 0,72$ oraz $r_s = 0,74$) w przeciwieństwie do CHARMM ($r_p = 0,26$ oraz $r_s = 0,32$), dlatego różnice korelacji są jeszcze większe: $\Delta r_p = 0,47$ oraz $\Delta r_s = 0,42$. Oznacza to, że w przypadku niezwiązanej formy FGF-2 pole siłowe CHARMM jeszcze silniej przeszacowuje wartości fluktuacji elastyczności białka. W FGF-2 miejscem wiązania heparyny jest reszta N27²⁵⁶, która została poprawnie przewidziana przez oba pola siłowe.



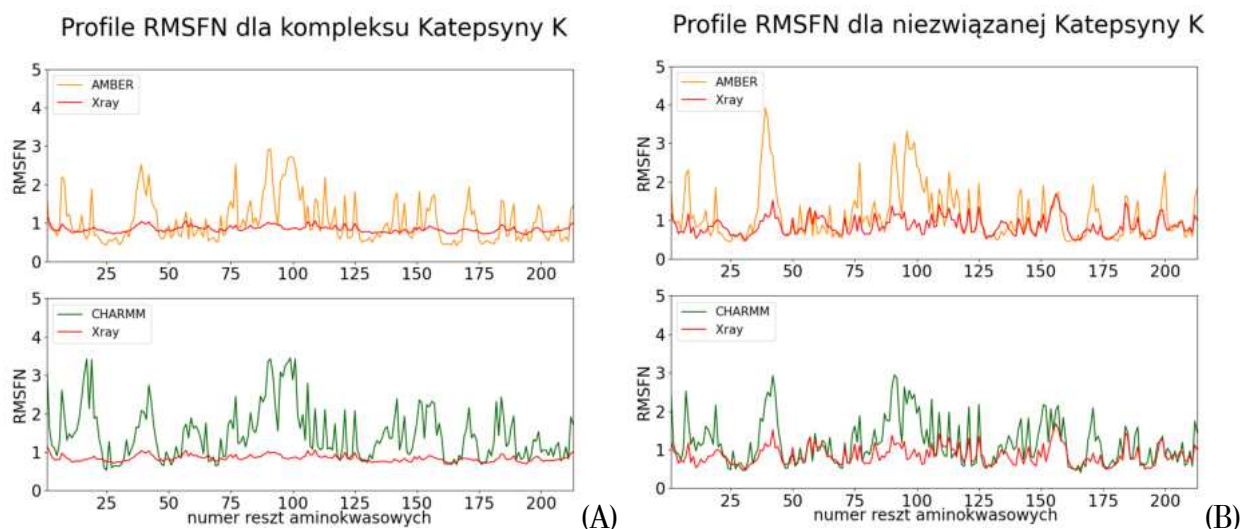
Rysunek 4.16: Profile fluktuacje reszt aminokwasowych FGF-2 w kompleksie z heparyną dp6 (A) i stanie niezwiązanym (B) w HP dp6 uzyskane z symulacji AMBER (pomarańczowa linia) i CHARMM (zielona linia) oraz profile fluktuacje reszt aminokwasowych uzyskanych z eksperymentów (czerwona linia).

Kolejnym analizowanym białkiem był FGF-1 w formie kompleksu (PDB: 2AXM, posiadające struktury drugorzędowe $\alpha + \beta$, struktura uzyskana z rentgenografii Xray) z heparyną dp6. Zarówno pole siłowe AMBER, jak i CHARMM wykazują dobrą korelację (odpowiednio $r_p = 0,61$ i $r_s = 0,53$ oraz $r_p = 0,57$ i $r_s = 0,52$) między danymi teoretycznymi a eksperymentalnymi (rysunek 4.17, tabela 5.1), z niewielkimi różnicami: $\Delta r_p = 0,04$, $\Delta r_s = 0,02$. Znaczne przeszacowanie w CHARMM widoczne jest tylko w obszarze 43–45 (pętla pomiędzy 4 a 5 α -helisą) reszty aminokwasowej. Również pole siłowe AMBER przeszacowuje fluktuacje w tym regionie. W formie niezwiązanej FGF-1 (PDB: 2K43, posiadające struktury drugorzędowe $\alpha + \beta$, struktura uzyskana z zespołów NMR), pole siłowe CHARMM lepiej odwzorowuje elastyczność białka niż AMBER na podstawie wartości korelacji (odpowiednio $r_p = 0,59$ i $r_s = 0,62$ oraz $r_p = 0,53$ i $r_s = 0,53$), gdzie różnica między uzyskanymi korelacjami jest nieco większa w porównaniu do formy kompleksu: $\Delta r_p = 0,06$, $\Delta r_s = 0,10$. Heparyna silniej wpływa na elastyczność niezwiązanej formy białka w przypadku zastosowania pola siłowego AMBER, z przeszacowaniami w obszarach 43–45 (pętla pomiędzy 4 a 5 α -helisą), 75 (pętla pomiędzy 7 a 8 α -helisą) i 118 (pętla pomiędzy 2 a 3 β -harmonijką) reszty aminokwasowej. Miejscem wiązania heparyny w FGF-1 jest reszta N26²⁵⁶, która została poprawnie zidentyfikowana przez oba pola siłowe.



Rysunek 4.17: Profile fluktuacje reszt aminokwasowych FGF-1 w kompleksie z heparyną dp6 (A) i stanie niezwiązanym (B) w HP dp6 uzyskane z symulacji AMBER (pomarańczowa linia) i CHARMM (zielona linia) oraz profile fluktuacje reszt aminokwasowych uzyskanych z eksperymentów (czerwona linia).

Trzecim białkiem była Katepsyna K w formie kompleksu (PDB: 4N8W, posiadające struktury drugorzędowe $\alpha + \beta$, struktura uzyskana z rentgenografii Xray) z chondroityno-4-siarczanem (C4-S) dp6. Pole siłowe AMBER lepiej przewiduje elastyczność białka w porównaniu do pola siłowego CHARMM na podstawie uzyskanych wartości korelacji (odpowiednio $r_p = 0,62$ i $r_s = 0,69$ oraz $r_p = 0,57$ i $r_s = 0,63$), gdzie różnice w wartościach współczynników korelacji wyglądają następująco; $\Delta r_p = 0,05$, $\Delta r_s = 0,06$. CHARMM pokazuje większy wpływ GAG-u na elastyczność, z przeszacowaniami w obszarach 15-20 (pętla pomiędzy 1 α -helisą a 1 β -harmonijką), 150-160 (pętla pomiędzy 6 a 7 α -helisą) oraz 190 (pętla pomiędzy 8 a 9 α -helisą) reszty aminokwasowej (rysunek 4.18). W formie niezwiązanej (PDB: 5TUN, posiadające struktury drugorzędowe $\alpha + \beta$, struktura uzyskana z rentgenografii Xray), CHARMM przewiduje elastyczność lepiej niż AMBER na podstawie uzyskanych wartości korelacji (odpowiednio $r_p = 0,67$ i $r_s = 0,70$ oraz $r_p = 0,55$ i $r_s = 0,65$), gdzie różnica korelacji pomiędzyimi zastosowanymi polami siłowymi wynosi odpowiednio: $\Delta r_p = 0,11$, $\Delta r_s = 0,06$. AMBER przeszacowuje fluktuacje w obszarach 30-40 (2 β -harmonijka), 75 (2 β -harmonijka), 80-100 (pętla pomiędzy 2 α -helisą a 6 β -harmonijką) i 200 (pętla pomiędzy 9 α -helisą a 9 β -harmonijką) reszty aminokwasowej. Miejscami wiązania C4-S w Katepsynie są reszty Arg8, Lys9 i Lys10²⁵⁷, które poprawnie zostały przewidziane przez oba pola siłowe.



Rysunek 4.18: Profile fluktuacje reszt aminokwasowych katespiny K w kompleksie z chondroityno-4-siarczanem (C4-S) dp6 (A) i stanie niezwiązanym (B) w HP dp6 uzyskane z symulacji AMBER (pomarańczowa linia) i CHARMM (zielona linia) oraz profile fluktuacje reszt aminokwasowych uzyskanych z eksperymentów (czerwona linia).

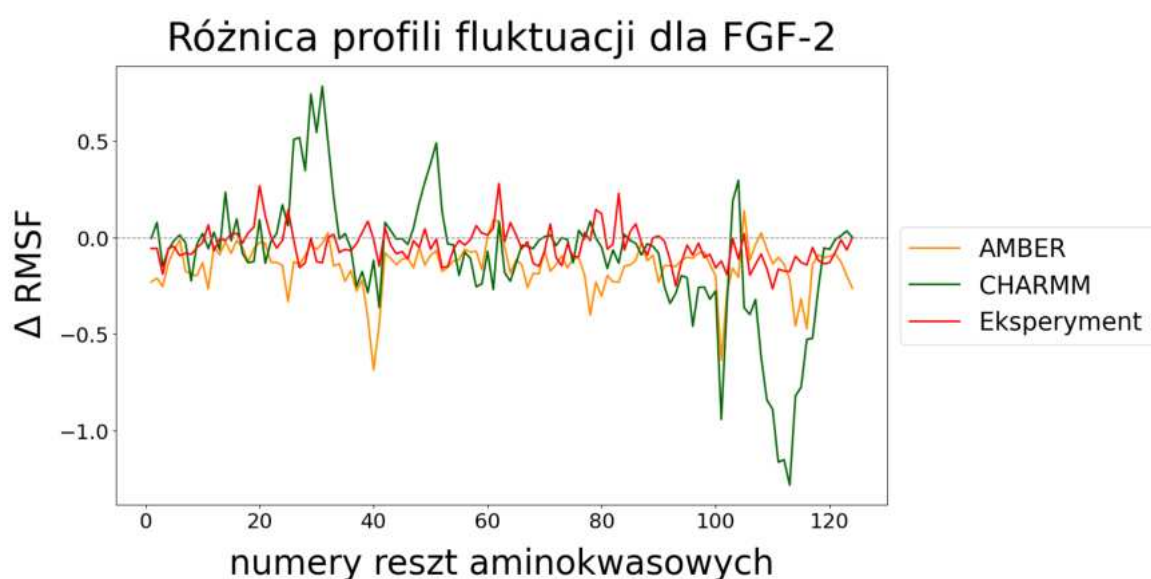
Na podstawie uzyskanych wyników stwierdzam, że pole siłowe AMBER lepiej sprawdza się do badania wpływu wiązania GAGów na elastyczność struktur białek w formie związanej i niezwiązanej. Główne niedokładności przedwydowania profili fluktuacji w obu tych polach między wynikają z przeszacowania ruchliwości pętli.

4.7 Różnica wartości profili fluktuacji między formą związaną i niezwiązaną białek

W celu potwierdzenia poprzednich wniosków przeanalizowałem również różnice w profilach fluktuacji pomiędzy formą niezwiązaną i związaną (kompleksu) każdego z badanych białek, których wartości r_p oraz r_s zestawilem w tabeli 4.5.

W pierwszej części przeanalizowałem różnice w profilach fluktuacji dla białka FGF-2 w formie niezwiązanej oraz związanej (kompleksu) z heparyną dp6, co przedstawiłem na rysunku 4.19. Widoczne jest, że w przypadku pola siłowego CHARMM występuje gwałtowny wzrost wartości różnicy fluktuacji w obszarze 25–35 (znajdujący się blisko Heparyny) oraz w rejonie 50 (oddalony od Heparyny) reszty aminokwasowej, co nie znajduje odzwierciedlenia w danych eksperymentalnych. Wzrost różnicy fluktuacji obserwuje się w okolicach 30 (blisko Heparyny), 60 (oddalony od Heparyny) i 100

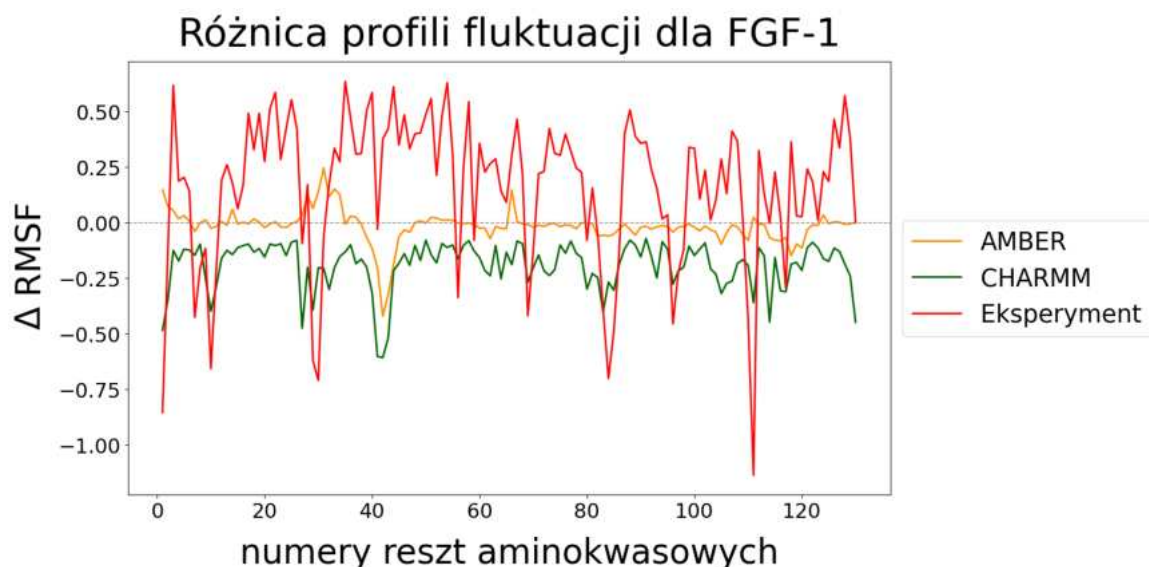
(oddalony od Heparyny) reszty w przypadku pola siłowego AMBER. Natomiast bardzo duży spadek jest zauważalny w obszarach 80–100 (oddalony od Heparyny) oraz 105–120 (oddalony od Heparyny) reszty aminokwasowej dla pola siłowego CHARMM. Tylko obszar 25–35 reszty aminokwasowej znajduje się w pobliżu Heparyny, co potwierdza jego miejsce wiązania z poprzedniej analizy. Wskazuje to, że pole siłowe AMBER lepiej odwzorowuje zmiany elastyczności struktury białka FGF-2 w wyniku wiązania z GAG-iem. Z kolei CHARMM silnie uwidacznia wpływ heparyny na elastyczność struktury białka FGF-2, co może wynikać z przeszacowania efektu interakcji.



Rysunek 4.19: Profile różnic fluktuacji struktury białka FGF-2 pomiędzy formą niezwiązaną i związaną z Heparyną dp6 uzyskanych z symulacji pól siłowych AMBER (pomarańczowa linia) i CHARMM (zielona linia) oraz dla eksperymentów (czerwona linia).

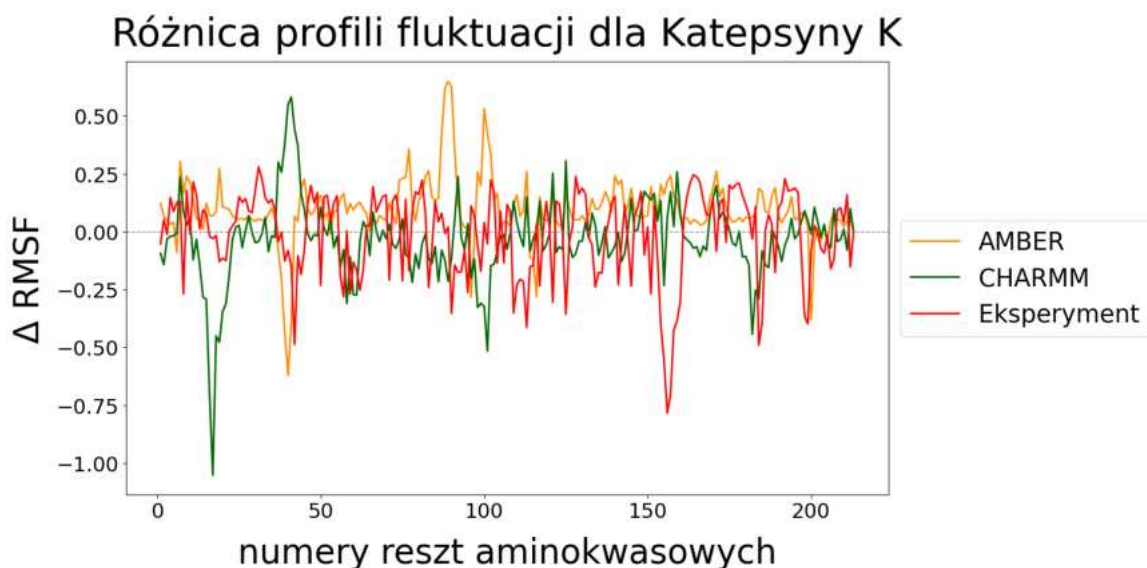
W kolejnej analizie zestawilem różnice profili fluktuacji dla białka FGF-1, co przedstawiłem na rysunku 4.20. Analiza ta jest trudniejsza do jednoznacznego opisanie ze względu na znaczne różnice w profilach fluktuacji uzyskanych z danych eksperymentalnych. W przypadku pola siłowego CHARMM zaobserwowałem spadek wartości różnicy fluktuacji w okolicach 30 (blisko Heparyny), 40 (oddalona od Heparyny), 95 (oddalona od Heparyny) i w okolicy 120 (oddalona od Heparyny) reszty aminokwasowej. Z kolei pole siłowe AMBER wykazuje wyraźny wzrost wartości różnicy fluktuacji w obszarze 30 (blisko Heparyny) oraz niewielki wzrost w okolicy 70 (oddalona od Heparyny) reszty aminokwasowej, zaś duży spadek w okolicy 40 (oddalona od Heparyny) oraz niewielki spadek w okolicy 120 (oddalona od Heparyny) reszty aminokwasowej. Tylko obszar 30 reszty aminokwasowej

znajduje się w pobliżu Heparyny, co potwierdza jego miejsce wiązania z poprzedniej analizy. Na podstawie tych obserwacji stwierdzam, że pole siłowe AMBER dla obu form białka FGF-1 stosunkowo oddaje zmiany fluktuacji podczas wiązania GAGu, z wyjątkiem wyżej wymienionych obszarów.



Rysunek 4.20: Profile różnic fluktuacji struktury białka FGF-1 pomiędzy formą niezwiązaną i związaną z Heparyną dp6 uzyskanych z symulacji pól siłowych AMBER (pomarańczowa linia) i CHARMM (zielona linia) oraz dla eksperymentów (czerwona linia).

W ostatniej analizie porównałem różnice profili fluktuacji dla katepsyny K, co przedstawiłem na rysunku 4.21. Z analizy wynika, że dla pola siłowego CHARMM obserwuje się znaczący spadek wartości różnicy fluktuacji w rejonach 25 (oddalona od C4-S), a także mniejsze spadki w obrębie 60 (oddalona od C4-S), 100 (oddalony od C4-S) oraz 180 (oddalona od C4-S) reszty aminokwasowej. Z kolei pole siłowe AMBER wykazuje spadek wartości różnicy fluktuacji w obszarach 40-50 (oddalony od C4-S), 90-100 (oddalony od C4-S), 120 (oddalona od C4-S) oraz 200 (oddalona od C4-S) reszty aminokwasowej, a także wzrost w okolicy 70-100 (oddalony od C4-S) reszty aminokwasowej. W tym przypadku przeszacowania i niedoszacowania różnic wartości fluktuacji błędnie przewidują miejsce wiązania białka z GAGiem. Porównując ze sobą dane uzyskane z pola siłowego AMBER oraz CHARMM oraz uzyskanych wartości r_p oraz r_s zebranych w tabeli 4.5 mogę stwierdzić, że w oba pola w zbliżonych stopniu przewidują wpływ GAGu na elastyczność katepsyny K.



Rysunek 4.21: Profile różnic fluktuacji struktury białka Katepsyny K formą niezwiązaną i związaną z chondroityno-4-siarczanem (C4-S) dp6 symulacji pól siłowych AMBER (pomarańczowa linia) i CHARMM (zielona linia) oraz dla eksperymentów (czerwona linia).

Tabela 4.5: Zestawienie współczynników korelacji Pearsona (r_p) i Spearmana (r_s) pomiędzy przewidywanymi oraz eksperymentalnymi profilami RMSF dla różnic pomiędzy formą niezwiązaną i związaną (kompleksu) każdego z badanych białek.

Białko	AMBER		CHARMM	
	r_p	r_s	r_p	r_s
FGF-2	0.1068	0.0867	0.3002	0.2555
FGF-1	-0.1219	0.0791	0.4617	0.4804
CatK	-0.1682	-0.2955	-0.1762	-0.1874

Wizualna analiza różnic fluktuacji może być myląca, dlatego konieczne jest, aby ją zweryfikować za pomocą metod statystycznych. Biorąc pod uwagę uzyskane wartości współczynników korelacji pomiędzy przewidywanymi a eksperymentalnymi profilami fluktuacji dla różnic między formą niezwiązaną a formą związaną (kompleksową) każdego z badanych białek mogę stwierdzić, że pole siłowe CHARMM jest dokładniejszy w przewidywaniu wpływu GAGów na zmianę elastyczności białek odnosząc się do pola siłowego AMBER.

Rozdział 5

Podsumowanie i wnioski

Przeprowadzone symulacje wykorzystujące gruboziarniste oraz pełnoatomowe pola siłowe pozwoliły mi zweryfikować i określić elastyczności występujące w zestawie badanych białek, białek otoczonych błoną lipidową a także określić wpływ GAGów na elastyczność białek. Uzyskane przeze mnie wyniki pozwalają mi potwierdzić spełnienie założonych celów, a także wyciągnąć wymienione wnioski :

1. Biorąc pod uwagę fakt, że końcowe reszty aminokwasowe wpływają na wartości r_p oraz r_s i są dominujące, stwierdzam, że do badania elastyczności białek należy usunąć ich końcowe fragmenty, aby uzyskać wiarygodne i miarodajne analizy.

2. Zobrazowanie uśrednionych współczynników korelacji na podstawie przeprowadzonych symulacji metod teoretycznych w odniesieniu do metod eksperymentalnych pozwoliło mi na stwierdzenie, że przewidywanie elastyczności białek zależy od typu struktury drugorzędowej, a także od metody uzyskania struktury białka (zespołów NMR lub krystalografii X-ray)

3. Dwukierunkowa analiza wariancji pozwoliła mi potwierdzić, że uzyskane wartości współczynników korelacji zależą od metody przewidywania fluktuacji struktur białkowych.

4. Przeprowadzone testy studenta pozwoliły określić, że metody CABS-flex, UNRES-flex i UNRES-DSSP-flex sprawdzają się do przewidywania elastyczności struktur białek pochodzących z zespołów NMR, zaś metoda NOLB jest stanowczo lepszą metodą w przypadku przewidywania elastyczności struktur białek pochodzących z rentgenografii X-ray. Uzyskane istotności statystyczne wskazują, że metoda CABS-flex oraz UNRES-DSSP-flex lepiej przewiduje elastyczność struktur białek uzyskanych z zespołów NMR o strukturze drugorzędowej $\alpha + \beta$ w porównaniu do reszty zastosowanych metod, zaś metoda NOLB lepiej sprawdza się w przypadku struktur białek uzyskanych metodą rentgenografii X-ray o strukturze drugorzędowej $\alpha + \beta$.

5. Na podstawie uzyskanych wyników z rozkładu współczynników korelacji udowodniłem, że metoda NOLB nie powinna być używana do badania elastyczności struktur białek pochodzących z zespołów NMR.

6. Przeprowadzona analiza skośności układów potwierdziła wcześniejsze stwierdzenie, że metoda CABS-flex oraz UNRES-DSSP-flex sprawdzają się lepiej w przypadku badania elastyczności struktur białek uzyskanych z zespołów NMR, zaś metoda NOLB ma przewagę nad innymi metodami w przewidywaniu elastyczności struktur białek uzyskanych metodą rentgenografii Xray.

7. Wyniki analizy dystrybuanty rozkładów ponownie potwierdziły, że metoda NOLB jest najlepszą metodą do przewidywania elastyczności struktur białek uzyskanych metodą rentgenografii Xray, zaś metoda CABS-flex okazała się najlepsza do przewidywania elastyczności struktur białek pochodzących z zespołów NMR.

8. Wykazałem, że nie istnieje jakakolwiek zależność pomiędzy uzyskanymi wartościami współczynników korelacji na podstawie przeprowadzonych symulacji metodami teoretycznymi a ilością reszt aminokwasowych występujących w łańcuchu białkowym.

9. W oparciu o przeprowadzoną analizę profili fluktuacji dla wybranych białek potwierdzam wcześniejsze ustalenia, że do określenia elastyczności struktur białek uzyskanych metodą rentgenografii Xray najlepiej sprawdza się metoda NOLB i nie powinna być brana pod uwagę w przypadku białek uzyskanych z zespołów NMR, dla których najlepiej sprawdza się metoda CABS-flex, a tuż za nią metoda UNRES-DSSP-flex.

10. Obliczyłem PES, PMF i na ich podstawie dopasowałem funkcje analityczne dla oddziaływań par cząsteczek UNRES-MARTINI. Mogę stwierdzić, że dla cząsteczek hydrofobowych, potencjał Gay-Berne'a potrafi dobrze odwzorować mapę energetyczną potencjału średnich sił, w przypadku oddziaływań cząsteczek polarnych następuje "wygładzenie" w okolicach minimum globalnego, zaś dla oddziaływań cząsteczek naładowanych w przypadku metody PM7 dopasowanie funkcji analitycznych jest problematyczne, co jest związane z tym, że są silnie odpychane cząsteczki o jednoimiennych ładunkach lub bardzo silnym przyciąganiem różnoimiennych.

11. Udowodniłem, że nowe pole siłowe UNRES-MARTINI ma porównywalną dokładność w przewidywaniu elastyczności struktur białek związanych z lipidami jak wyjściowa metoda UNRES-flex dla białek w wodzie, niezależnie od rodzaju analizowanego białka.

12. Ostatnia analiza jednoznacznie wskazuje, że pole siłowe z pakietu AMBER zapewnia bardziej spójne wyniki wartości fluktuacji względem eksperymentów dla banych białek FGF-1, FGF-2 i katepsyny K w stanie niezwiązanych i związanym, zaś na podstawie uzyskanych różnic wartości współczynników korelacji pomiędzy przewidywanymi a eksperymentalnymi profilami fluktuacji między formą niezwiązaną a formą związaną (kompleksową), pole siłowe CHARMM lepiej przewiduje wpływ GAGów na elastyczność badanych białek.

Rozdział 6

Literatura

- [1] Kuriata, A.; Gierut, A. M.; Oleniecki, T.; Ciemny, M. P.; Kolinski, A.; Kurcinski, M.; Kmiecik, S. *Nucleic acids research* **2018**, *46*, W338–W343.
- [2] Dziadek, Ł.; Sieradzan, A.; Czaplewski, C.; Zalewski, M.; Banas, E.; Toczek, M.; Nisterenko, W.; Grudinin, S.; Liwo, A.; Gieldon, A. *Journal of Chemical Theory and Computation* **2024**, *20*, 7667–7681.
- [3] Anfinsen, C. B.; Redfield, R. R. *Advances in protein chemistry* **1956**, *11*, 1–100.
- [4] Bornot, A.; Etchebest, C.; De Brevern, A. G. *Proteins: Structure, Function, and Bioinformatics* **2011**, *79*, 839–852.
- [5] Teilum, K.; Olsen, J. G.; Kragelund, B. B. *Cellular and Molecular Life Sciences* **2009**, *66*, 2231–2247.
- [6] Frauenfelder, H.; Chen, G.; Berendzen, J.; Fenimore, P. W.; Jansson, H.; McMahon, B. H.; Stroer, I. R.; Swenson, J.; Young, R. D. *Proceedings of the National Academy of Sciences* **2009**, *106*, 5129–5134.
- [7] Eisenmesser, E. Z.; Millet, O.; Labeikovsky, W.; Korzhnev, D. M.; Wolf-Watz, M.; Bosco, D. A.; Skalicky, J. J.; Kay, L. E.; Kern, D. *Nature* **2005**, *438*, 117–121.
- [8] Boehr, D. D.; McElheny, D.; Dyson, H. J.; Wright, P. E. *Science* **2006**, *313*, 1638–1642.
- [9] Boehr, D. D.; Dyson, H. J.; Wright, P. E. *Chemical Reviews* **2006**, *106*, 3055–3079.

- [10] Peng, T.; Zintsmaster, J. S.; Namanja, A. T.; Peng, J. W. *Nature Structural & Molecular Biology* **2007**, *14*, 325–331.
- [11] Boehr, D. D.; Wright, P. E. *Science* **2008**, *320*, 1429–1430.
- [12] Dunker, A. K.; Oldfield, C. J.; Meng, J.; Romero, P.; Yang, J. Y.; Chen, J.; Vacic, V.; Obradovic, Z.; Uversky, V. N. *BMC Genomics* **2008**, *9*, S1.
- [13] Dobson, C. M. *Nature* **2003**, *426*, 884–890.
- [14] Teague, S. J. *Nature Reviews Drug Discovery* **2003**, *2*, 527–541.
- [15] Vander Meersche, Y.; Cretin, G.; De Brevern, A. G.; Gelly, J.-C.; Galochkina, T. *Journal of Molecular Biology* **2021**, *433*, 166882.
- [16] Sun, Z.; Liu, Q.; Qu, G.; Feng, Y.; Reetz, M. T. *Chemical Reviews* **2019**, *119*, 1626–1665.
- [17] Carugo, O. *BMC Bioinformatics* **2018**, *19*, 1–9.
- [18] Dropka, N.; Gradwohl, K.-P. *Encyclopedia of Condensed Matter Physics* **2024**, 231–247.
- [19] Kurplus, M.; McCammon, J. A. *Annual Review of Biochemistry* **1983**, *52*, 263–300.
- [20] Wall, M. E.; Wolff, A. M.; Fraser, J. S. *Current opinion in structural biology* **2018**, *50*, 109–116.
- [21] Eyal, E.; Gerzon, S.; Potapov, V.; Edelman, M.; Sobolev, V. *Journal of Molecular Biology* **2005**, *351*, 431–442.
- [22] Ishima, R.; Torchia, D. A. *Nature Structural Biology* **2000**, *7*, 740–743.
- [23] Kay, L. E. *Nature Structural Biology* **1998**, *5*, 513–517.
- [24] Wüthrich, K.; Wagner, G. *Trends in Biochemical Sciences* **1978**, *3*, 227–230.
- [25] Iain Donald, C.; Christopher Martin, D.; Robert Joseph Paton, W. *Proceedings of the Royal Society of London. Series B. Biological Sciences* **1975**, *189*, 503–509.
- [26] Snyder, G. H.; Rowan III, R.; Karplus, S.; Sykes, B. D. *Biochemistry* **1975**, *14*, 3765–3777.

- [27] Bax, A. *Current Opinion in Structural Biology* **1994**, 4, 738–744.
- [28] Pardi, A. *Methods in enzymology* **1995**, 261, 350–380.
- [29] Palmer, A. G.; Kroenke, C. D.; Patrick Loria, J. *Methods in Enzymology* **2001**, 339, 204–238.
- [30] Fushman, D.; Cahill, S.; Cowburn, D. *Journal of Molecular Biology* **1997**, 266, 173–194.
- [31] Korzhnev, D. M.; Orekhov, V. Y.; Arseniev, A. S. *Journal of Magnetic Resonance* **1997**, 127, 184–191.
- [32] Emwas, A. H.; Roy, R.; McKay, R. T.; Tenori, L.; Saccenti, E.; Gowda, G. A. N.; Raftery, D.; Alahmari, F.; Jaremko, L.; Jaremko, M.; Wishart, D. S. *Metabolites* **2019**, 9, 123.
- [33] Hoffmann, A.; Grudinin, S. *Journal of Chemical Theory and Computation* **2017**, 13, 2123–2134.
- [34] Jardetzky, O. *Progress in Biophysics and Molecular Biology* **1996**, 65, 171–219.
- [35] Salvatella, X. *Advances in Experimental Medicine and Biology* **2014**, 805, 67–85.
- [36] Levitt, M.; Warshel, A. *Nature* **1975**, 253, 694–698.
- [37] Graveur, P. et al. *Frontiers in Molecular Biosciences* **2015**, 2, 20.
- [38] Harder, M. E.; Deinzer, M. L.; Leid, M. E.; Schimerlik, M. I. *Protein Science: A Publication of the Protein Society* **2004**, 13, 2207–2222.
- [39] McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* **1977**, 267, 585–590.
- [40] Rueda, M.; Ferrer-Costa, C.; Meyer, T.; Pérez, A.; Camps, J.; Hospital, A.; Gelpí, J. L.; Orozco, M. *Proceedings of the National Academy of Sciences* **2007**, 104, 796–801.
- [41] Dokholyan, N. V.; Buldyrev, S. V.; Stanley, H. E.; Shakhnovich, E. I. *Folding and Design* **1998**, 3, 577–587.
- [42] Tirion, M. M. *Physical Review Letters* **1996**, 77, 1905–1908.
- [43] Emperador, A.; Carrillo, O.; Rueda, M.; Orozco, M. *Biophysical Journal* **2008**, 95, 2127–2138.
- [44] Ma, J. *Structure* **2005**, 13, 373–380.

- [45] Cockcroft, S. *Essays in Biochemistry* **2021**, *65*, 813–845.
- [46] Bogdanov, M.; Mileykovskaya, E.; Dowhan, W. *Subcellular Biochemistry* **2008**, *49*, 197–239.
- [47] Saxton, R. A.; Sabatini, D. M. *Cell* **2017**, *168*, 960–976.
- [48] Evans, E. A. *Biophysical journal* **1974**, *14*, 923–931.
- [49] Helfrich, W. *Zeitschrift für Naturforschung C* **1973**, *28*, 693–703.
- [50] Canham, P. *Journal of Theoretical Biology* **1970**, *26*, 61–81.
- [51] Singer, S. J.; Nicolson, G. L. *Science* **1972**, *175*, 720–731.
- [52] Overhoff, M.; De Bruyckere, E.; Kononenko, N. L. *Journal of Neurochemistry* **2021**, *157*, 263–296.
- [53] Coskun, Ü.; Simons, K. *Structure* **2011**, *19*, 1543–1548.
- [54] Giri, R. P.; Mukhopadhyay, M. K.; Sanyal, M. K.; Bose, D.; Chakrabarti, A.; Quan, P.; Bu, W.; Lin, B. *The Journal of Physical Chemistry Letters* **2022**, *13*, 11430–11437.
- [55] Gumbart, J.; Wang, Y.; Aksimentiev, A.; Tajkhorshid, E.; Schulten, K. *Current opinion in structural biology* **2005**, *15*, 423–431.
- [56] Li, Y.; Xu, J.; Li, D. *Microfluidics and Nanofluidics* **2010**, *9*, 1011–1031.
- [57] Vrouenraets, M.; Miedema, H. *European Biophysics Journal* **2010**, *39*, 1563–1571.
- [58] Varma, S.; Jakobsson, E. *Biophysical Journal* **2004**, *86*, 690–704.
- [59] Varma, S.; Chiu, S.-W.; Jakobsson, E. *Biophysical Journal* **2006**, *90*, 112–123.
- [60] Maffeo, C.; Bhattacharya, S.; Yoo, J.; Wells, D.; Aksimentiev, A. *Chemical Reviews* **2012**, *112*, 6250–6284.
- [61] Kox, A.; Michels, J.; Wiegel, F. *Nature* **1980**, *287*, 317–319.
- [62] Van der Ploeg, P.; Berendsen, H. *Jorunal of Chemical Physics* **1982**, *76*, 3271–3276.
- [63] Van der Ploeg, P.; Berendsen, H. J. *Jorunal of Molecular Physics* **1983**, *49*, 233–248.

- [64] Jönsson, B.; Edholm, O.; Teleman, O. *Journal of Chemical Physics* **1986**, *85*, 2259–2271.
- [65] Egberts, E.; Berendsen, H. *Journal of Chemical Physics* **1988**, *89*, 3718–3732.
- [66] Tieleman, D.; Marrink, S.; Berendsen, H. *Biochimica et Biophysica Acta (BBA) - Reviews on Biomembranes* **1997**, *1331*, 235–270.
- [67] Murrey, H. E.; Hsieh-Wilson, L. C. *Chemical Reviews* **2008**, *108*, 1708–1731.
- [68] Imberty, A.; Lortat-Jacob, H.; Pérez, S. *Carbohydrate Research* **2007**, *342*, 430–439.
- [69] Perez, S.; Makshakova, O.; Angulo, J.; Bedini, E.; Bisio, A.; de Paz, J. L.; Fadda, E.; Guerrini, M.; Hricovini, M.; Hricovini, M.; others *Jacs Au* **2023**, *3*, 628–656.
- [70] Soares Da Costa, D.; Reis, R. L.; Pashkuleva, I. *Annual Review of Biomedical Engineering* **2017**, *19*, 1–26.
- [71] Almond, A. *Current Opinion in Structural Biology* **2018**, *50*, 58–64.
- [72] Sankaranarayanan, N. V.; Nagarajan, B.; Desai, U. R. *Current Opinion in Structural Biology* **2018**, *50*, 91–100.
- [73] Joseph, P. R.; Mosier, P.; Desai, U.; Rajarathnam, K. *Biochemical Journal* **2015**, *472*, 121–133.
- [74] Rother, S.; Samsonov, S. A.; Hofmann, T.; Blaszkiewicz, J.; Köhling, S.; Moeller, S.; Schnabelrauch, M.; Rademann, J.; Kalkhof, S.; von Bergen, M.; Pisabarro, M. T.; Scharnweber, D.; Hintze, V. *Acta Biomaterialia* **2016**, *45*, 143–154.
- [75] Li, Z.; Kienetz, M.; Cherney, M. M.; James, M. N.; Brömme, D. *Journal of Molecular Biology* **2008**, *383*, 78–91.
- [76] Van der Smissen, A.; Samsonov, S.; Hintze, V.; Scharnweber, D.; Moeller, S.; Schnabelrauch, M.; Pisabarro, M. T.; Anderegg, U. *Acta Biomaterialia* **2013**, *9*, 7775–7786.
- [77] Samsonov, S. A.; Bichmann, L.; Pisabarro, M. T. *Journal of Chemical Information and Modeling* **2015**, *55*, 114–124.

- [78] Whitmore, E. K.; Vesenka, G.; Sihler, H.; Guvench, O. *Biomolecules* **2020**, *10*, 537.
- [79] Pomin, V. H.; Wang, X. *Molecules* **2018**, *23*, 2314.
- [80] Nagarajan, B.; Holmes, S. G.; Sankaranarayanan, N. V.; Desai, U. R. *Current Opinion in Structural Biology* **2022**, *74*, 102356.
- [81] Allen, M. P.; others *Computational Soft Matter: from Synthetic Polymers to Proteins* **2004**, *23*, 1–28.
- [82] Leach, A. R. *Molecular Modelling: Principles and Applications*; Pearson education, 2001.
- [83] Haile, J. M. *Molecular Dynamics Simulation: Elementary Methods*; John Wiley & Sons, Inc., 1992.
- [84] Hernández-Rodríguez, M.; C. Rosales-Hernández, M.; E. Mendieta-Wejbe, J.; Martínez-Archundia, M.; Correa Basurto, J. *Current Medicinal Chemistry* **2016**, *23*, 3909–3924.
- [85] Badar, M. S.; Shamsi, S.; Ahmed, J.; Alam, M. A. *Springer* **2022**, 131–151.
- [86] Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. *Science* **2011**, *334*, 517–520.
- [87] Buch, I.; Harvey, M. J.; Giorgino, T.; Anderson, D. P.; De Fabritiis, G. *Journal of Chemical Information and Modeling* **2010**, *50*, 397–403.
- [88] Shirts, M.; Pande, V. S. *Science* **2000**, *290*, 1903–1904.
- [89] Dror, R. O.; Pan, A. C.; Arlow, D. H.; Borhani, D. W.; Maragakis, P.; Shan, Y.; Xu, H.; Shaw, D. E. *Proceedings of the National Academy of Sciences* **2011**, *108*, 13118–13123.
- [90] Shukla, D.; Meng, Y.; Roux, B.; Pande, V. S. *Nature Communications* **2014**, *5*, 3397.
- [91] Plattner, N.; Noé, F. *Nature Communications* **2015**, *6*, 7653.
- [92] Plattner, N.; Doerr, S.; De Fabritiis, G.; Noé, F. *Nature Chemistry* **2017**, *9*, 1005–1011.
- [93] Paul, F.; Wehmeyer, C.; Abualrous, E. T.; Wu, H.; Crabtree, M. D.; Schöneberg, J.; Clarke, J.; Freund, C.; Weikl, T. R.; Noé, F. *Nature Communications* **2017**, *8*, 1095.
- [94] Clementi, C. *Current Opinion in Structural Biology* **2008**, *18*, 10–15.

- [95] Saunders, M. G.; Voth, G. A. *Annual Review of Biophysics* **2013**, *42*, 73–93.
- [96] Noid, W. G. *The Journal of Chemical Physics* **2013**, *139*, 090901.
- [97] Matysiak, S.; Clementi, C. *Journal of Molecular Biology* **2004**, *343*, 235–248.
- [98] Matysiak, S.; Clementi, C. *Journal of Molecular Biology* **2006**, *363*, 297–308.
- [99] Chen, J.; Chen, J.; Pinamonti, G.; Clementi, C. *Journal of Chemical Theory and Computation* **2018**, *14*, 3849–3858.
- [100] Wang, J.; Olsson, S.; Wehmeyer, C.; Pérez, A.; Charron, N. E.; De Fabritiis, G.; Noé, F.; Clementi, C. *ACS Central Science* **2019**, *5*, 755–767.
- [101] Clementi, C.; Nymeyer, H.; Onuchic, J. N. *Journal of Molecular Biology* **2000**, *298*, 937–953.
- [102] Liwo, A.; Khalili, M.; Scheraga, H. A. *Proceedings of the National Academy of Sciences* **2005**, *102*, 2362–2367.
- [103] Souza, P. C.; Thallmair, S.; Conflitti, P.; Ramírez-Palacios, C.; Alessandri, R.; Raniolo, S.; Limongelli, V.; Marrink, S. J. *Nature Communications* **2020**, *11*, 3714.
- [104] Roel-Touris, J.; Don, C. G.; V. Honorato, R.; Rodrigues, J. P.; Bonvin, A. M. *Journal of Chemical Theory and Computation* **2019**, *15*, 6358–6367.
- [105] Louhivuori, M.; Risselada, H. J.; van der Giessen, E.; Marrink, S. J. *Proceedings of the National Academy of Sciences* **2010**, *107*, 19856–19860.
- [106] Davies, K. M.; Anselmi, C.; Wittig, I.; Faraldo-Gómez, J. D.; Kühlbrandt, W. *Proceedings of the National Academy of Sciences* *109*, 13602–13607.
- [107] Zheng, W.; Tsai, M.-Y.; Chen, M.; Wolynes, P. G. *Proceedings of the National Academy of Sciences* **2016**, *113*, 11835–11840.
- [108] Pak, A. J.; Yu, A.; Ke, Z.; Briggs, J. A.; Voth, G. A. *Nature Communications* **2022**, *13*, 1002.
- [109] Barnoud, J.; Monticelli, L. *Molecular Modeling of Proteins* **2015**, 125–149.

- [110] Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; De Vries, A. H. *The Journal of Physical Chemistry B* **2007**, *111*, 7812–7824.
- [111] Kamerlin, S. C.; Vicatos, S.; Dryga, A.; Warshel, A. *Annual Review of Physical Chemistry* **2011**, *62*, 41–64.
- [112] Ayton, G. S.; Noid, W. G.; Voth, G. A. *Current Opinion in Structural Biology* **2007**, *17*, 192–198.
- [113] Zhou, H.-X. *Current Opinion in Structural Biology* **2014**, *25*, 67–76.
- [114] Smit, B.; Hilbers, P. A. J.; Esselink, K.; Rupert, L. A. M.; Van Os, N. M.; Schlijper, A. G. *Nature* **1990**, *348*, 624–625.
- [115] Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A. E.; Kolinski, A. *Chemical Reviews* **2016**, *116*, 7898–7936.
- [116] Maciejczyk, M.; Spasic, A.; Liwo, A.; Scheraga, H. A. *Journal of Computational Chemistry* **2010**, *31*, 1644–1655.
- [117] Stevens, M. J. *The Journal of Chemical Physics* **2004**, *121*, 11942–11948.
- [118] López, C. A.; Rzepiela, A. J.; de Vries, A. H.; Dijkhuizen, L.; Hunenberger, P. H.; Marrink, S. J. *Journal of Chemical Theory and Computation* **2009**, *5*, 3195–3210.
- [119] Hadley, K. R.; McCabe, C. *Molecular Simulation* **2012**, *38*, 671–681.
- [120] Lyubartsev, A. P.; Laaksonen, A. *Physical Review E* **1995**, *52*, 3730–3737.
- [121] Müller-Plathe, F. *ChemPhysChem* **2002**, *3*, 754–769.
- [122] Praprotnik, M.; Site, L. D.; Kremer, K. *Annual Review of Physical Chemistry* **2008**, *59*, 545–571.
- [123] Izvekov, S.; Voth, G. A. *The Journal of Physical Chemistry B* **2005**, *109*, 2469–2473.
- [124] Wang, Y.; Noid, W. G.; Liu, P.; Voth, G. A. *Physical Chemistry Chemical Physics* **2009**, *11*, 2002.
- [125] Shell, M. S. *The Journal of Chemical Physics* **2008**, *129*, 144108.

- [126] Nielsen, S. O.; Lopez, C. F.; Srinivas, G.; Klein, M. L. *The Journal of Chemical Physics* **2003**, *119*, 7043–7049.
- [127] Marrink, S. J.; De Vries, A. H.; Mark, A. E. *The Journal of Physical Chemistry B* **2004**, *108*, 750–760.
- [128] Davtyan, A.; Schafer, N. P.; Zheng, W.; Clementi, C.; Wolynes, P. G.; Papoian, G. A. *The Journal of Physical Chemistry B* **2012**, *116*, 8494–8503.
- [129] Czaplewski, C.; Karczyńska, A.; Sieradzan, A. K.; Liwo, A. *Nucleic Acids Research* **2018**, *46*, W304–W309.
- [130] Ślusarz, R.; Lubecka, E. A.; Czaplewski, C.; Liwo, A. *Frontiers in Molecular Biosciences* **2022**, *9*, 1071428.
- [131] Kubo, R. *Journal of the Physical Society of Japan* **1962**, *17*, 1100–1120.
- [132] Liwo, A.; Khalili, M.; Czaplewski, C.; Kalinowski, S.; Oldziej, S.; Wachucik, K.; Scheraga, H. A. *The Journal of Physical Chemistry B* **2007**, *111*, 260–285.
- [133] Khalili, M.; Liwo, A.; Rakowski, F.; Grochowski, P.; Scheraga, H. A. *The Journal of Physical Chemistry B* **2005**, *109*, 13785–13797.
- [134] Khalili, M.; Liwo, A.; Jagielska, A.; Scheraga, H. A. *The Journal of Physical Chemistry B* **2005**, *109*, 13798–13810.
- [135] Scheraga, H. A.; Khalili, M.; Liwo, A. *Annual Review of Physical Chemistry* **2007**, *58*, 57–83.
- [136] Ma, P.; Li, D.; Brüschweiler, R. *Proteins: Structure, Function, and Bioinformatics* **2023**, *91*, 847–855.
- [137] Kurcinski, M.; Oleniecki, T.; Ciemny, M. P.; Kuriata, A.; Kolinski, A.; Kmiecik, S. *Bioinformatics* **2019**, *35*, 694–695.
- [138] Kuriata, A.; Gierut, A. M.; Oleniecki, T.; Ciemny, M.; Kolinski, A.; Kurcinski, M.; Kmiecik, S. *Nucleic Acids Research* **2018**, *46*, W338–W343.

- [139] Kolinski, A. *Acta Biochimica Polonica* **2004**, *51*, 349–371.
- [140] Jamroz, M.; Kolinski, A.; Kmiecik, S. *Nucleic Acids Research* **2013**, *41*, W427–W431.
- [141] Jamroz, M.; Kolinski, A.; Kmiecik, S. *Bioinformatics* **2014**, *30*, 2150–2154.
- [142] Badaczewska-Dawid, A. E.; Kolinski, A.; Kmiecik, S. *Protein Structure Prediction* **2020**, *2165*, 337–353.
- [143] Dosztanyi, Z.; Meszaros, B.; Simon, I. *Briefings in Bioinformatics* **2010**, *11*, 225–243.
- [144] Park, S.-Y.; Borbat, P. P.; Gonzalez-Bonet, G.; Bhatnagar, J.; Pollard, A. M.; Freed, J. H.; Bilwes, A. M.; Crane, B. R. *Nature Structural & Molecular Biology* **2006**, *13*, 400–407.
- [145] Bueno, J. G. R.; Borelli, G.; Corrêa, T. L. R.; Fiamenghi, M. B.; José, J.; De Carvalho, M.; De Oliveira, L. C.; Pereira, G. A. G.; Dos Santos, L. V. *Biotechnology for Biofuels* **2020**, *13*, 145.
- [146] Torchala, M.; Gerguri, T.; Chaleil, R. A. G.; Gordon, P.; Russell, E.; Keshani, M.; Bates, P. A. *Proteins: Structure, Function, and Bioinformatics* **2020**, *88*, 962–972.
- [147] Hura, G. L.; Hodge, C. D.; Rosenberg, D.; Guzenko, D.; Duarte, J. M.; Monastyrskyy, B.; Grudin, S.; Kryshchuk, A.; Tainer, J. A.; Fidelis, K.; Tsutakawa, S. E. *Proteins: Structure, Function, and Bioinformatics* **2019**, *87*, 1298–1314.
- [148] Bruininks, B. M. H.; Souza, P. C. T.; Marrink, S. J. *Springer New York* **2019**, *2022*, 105–127.
- [149] Marrink, S. J.; Mark, A. E. *Journal of the American Chemical Society* **2003**, *125*, 11144–11145.
- [150] Marrink, S. J.; Mark, A. E. *Journal of the American Chemical Society* **2003**, *125*, 15233–15242.
- [151] Marrink, S. J.; Tieleman, D. P. *Chemical Society Reviews* **2013**, *42*, 6801–6822.
- [152] Alessandri, R.; Grünwald, E.; Marrink, S. J. *Advanced Materials* **2021**, *33*, 2008635.
- [153] Baron, R.; Trzesniak, D.; De Vries, A. H.; Elsener, A.; Marrink, S. J.; Van Gunsteren, W. F. *ChemPhysChem* **2007**, *8*, 452–461.

- [154] Shih, A. Y.; Arkhipov, A.; Freddolino, P. L.; Schulten, K. *The Journal of Physical Chemistry B* **2006**, *110*, 3674–3684.
- [155] Shih, A. Y.; Freddolino, P. L.; Arkhipov, A.; Schulten, K. *Journal of Structural Biology* **2007**, *157*, 579–592.
- [156] Marrink, S. J.; Risselada, J.; Mark, A. E. *Chemistry and Physics of Lipids* **2005**, *135*, 223–244.
- [157] Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S.-J. *Journal of Chemical Theory and Computation* **2008**, *4*, 819–834.
- [158] De Jong, D. H.; Singh, G.; Bennett, W. F. D.; Arnarez, C.; Wassenaar, T. A.; Schäfer, L. V.; Periole, X.; Tieleman, D. P.; Marrink, S. J. *Journal of Chemical Theory and Computation* **2013**, *9*, 687–697.
- [159] López, C. A.; Bellesia, G.; Redondo, A.; Langan, P.; Chundawat, S. P. S.; Dale, B. E.; Marrink, S. J.; Gnanakaran, S. *The Journal of Physical Chemistry B* **2015**, *119*, 465–473.
- [160] Uusitalo, J. J.; Ingólfsson, H. I.; Akhshi, P.; Tieleman, D. P.; Marrink, S. J. *Journal of Chemical Theory and Computation* **2015**, *11*, 3932–3945.
- [161] Uusitalo, J. J.; Ingólfsson, H. I.; Marrink, S. J.; Faustino, I. *Biophysical Journal* **2017**, *113*, 246–256.
- [162] De Jong, D. H.; Liguori, N.; Van Den Berg, T.; Arnarez, C.; Periole, X.; Marrink, S. J. *The Journal of Physical Chemistry B* **2015**, *119*, 7791–7803.
- [163] Rossi, G.; Fuchs, P. F. J.; Barnoud, J.; Monticelli, L. *The Journal of Physical Chemistry B* **2012**, *116*, 14353–14362.
- [164] Rossi, G.; Monticelli, L.; Puisto, S. R.; Vattulainen, I.; Ala-Nissila, T. *Soft Matter* **2011**, *7*, 698–708.
- [165] Alessandri, R.; Uusitalo, J. J.; De Vries, A. H.; Havenith, R. W. A.; Marrink, S. J. *Journal of the American Chemical Society* **2017**, *139*, 3697–3705.

- [166] Monticelli, L. *Journal of Chemical Theory and Computation* **2012**, *8*, 1370–1378.
- [167] Corradi, V.; Sejdiu, B. I.; Mesa-Gallosa, H.; Abdizadeh, H.; Noskov, S. Y.; Marrink, S. J.; Tieleman, D. P. *Chemical reviews* **2019**, *119*, 5775–5848.
- [168] Arnarez, C.; Mazat, J.-P.; Elezgaray, J.; Marrink, S.-J.; Periole, X. *Journal of the American Chemical Society* **2013**, *135*, 3112–3120.
- [169] Van Eerden, F. J.; Melo, M. N.; Frederix, P. W.; Periole, X.; Marrink, S. J. *Nature Communications* **2017**, *8*, 15214.
- [170] Negami, T.; Shimizu, K.; Terada, T. *Journal of Computational Chemistry* **2014**, *35*, 1835–1845.
- [171] Negami, T.; Shimizu, K.; Terada, T. *Chemical Physics Letters* **2020**, *742*, 137144.
- [172] Vanommeslaeghe, K.; MacKerell, A. *Biochimica et Biophysica Acta (BBA) - General Subjects* **2015**, *1850*, 861–871.
- [173] Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell Jr, A. D.; Pastor, R. W. *The Journal of Physical Chemistry B* **2010**, *114*, 7830–7843.
- [174] MacKerell Jr, A. D.; Banavali, N.; Foloppe, N. *Biopolymers: Original Research on Biomolecules* **2000**, *56*, 257–265.
- [175] Guvench, O.; Mallajosyula, S. S.; Raman, E. P.; Hatcher, E.; Vanommeslaeghe, K.; Foster, T. J.; Jamison, F. W.; MacKerell Jr, A. D. *Journal of Chemical Theory and Computation* **2011**, *7*, 3162–3180.
- [176] Case, D. A.; Aktulga, H. M.; Belfon, K.; Ben-Shalom, I. Y.; Berryman, J. T.; Brozell, S. R.; Cerutti, D. S.; Cheatham III, T. E.; Cisneros, G. A.; Cruzeiro, V. W. D.; others *Amber 2023*; 2023.
- [177] Lin, F.-Y.; MacKerell, A. D. *Biomolecular Simulations: Methods and Protocols* **2019**, 21–54.
- [178] MacKerell Jr, A. D.; Nilsson, L. *Current Opinion in Structural Biology* **2008**, *18*, 194–199.

- [179] Smith, M. D.; Rao, J. S.; Segelken, E.; Cruz, L. *Journal of Chemical Information and Modeling* **2015**, *55*, 2587–2595.
- [180] Chen, W.; Shi, C.; MacKerell Jr, A. D.; Shen, J. *The Journal of Physical Chemistry B* **2015**, *119*, 7902–7910.
- [181] Martin, M. G. *Fluid Phase Equilibria* **2006**, *248*, 50–55.
- [182] Zhang, S.; Andrews, B.; Schweitzer-Stenner, R.; Urbanc, B. *The Journal of Physical Chemistry B* **2020**, *124*, 11600–11616.
- [183] Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Structure, Function, and Bioinformatics* **2006**, *65*, 712–725.
- [184] Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. *Proteins* **2010**, *78*, 1950–1958.
- [185] Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. *Journal of chemical theory and computation* **2015**, *11*, 3696–3713.
- [186] Abriata, L. A.; Dal Peraro, M. *Computational and Structural Biotechnology Journal* **2021**, *19*, 2626–2636.
- [187] Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguetta, L.; Huang, H.; Miguez, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; Simmerling, C. *Journal of Chemical Theory and Computation* **2020**, *16*, 528–552.
- [188] Breyfogle, K. L.; Blood, D. L.; Rosnik, A. M.; Krueger, B. P. *The Journal of Physical Chemistry B* **2023**, *127*, 5772–5788.
- [189] Song, G.; Zhong, B.; Zhang, B.; Rehman, A. U.; Chen, H.-F. *Journal of Chemical Information and Modeling* **2023**, *63*, 1602–1614.
- [190] Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. *Journal of Computational Chemistry* **2008**, *29*, 622–655.

- [191] Tempel, W.; Tschampel, S.; Woods, R. J. *Journal of Biological Chemistry* **2002**, *277*, 6615–6621.
- [192] Hemmingsen, L.; Madsen, D. E.; Esbensen, A. L.; Olsen, L.; Engelsen, S. B. *Carbohydrate Research* **2004**, *339*, 937–948.
- [193] Bojarski, K. K.; Samsonov, S. A. *Journal of Molecular Graphics and Modelling* **2023**, *120*, 108406.
- [194] Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; De Groot, B. L.; Grubmüller, H.; MacKerell, A. D. *Nature Methods* **2017**, *14*, 71–73.
- [195] Grothaus, I. L.; Bussi, G.; Colombi Ciacchi, L. *Journal of Chemical Information and Modeling* **2022**, *62*, 4992–5008.
- [196] Andrews, B.; Long, K.; Urbanc, B. *The Journal of Physical Chemistry B* **2021**, *125*, 6897–6911.
- [197] Torrie, G. M.; Valleau, J. P. *Journal of Computational Physics* **1977**, *23*, 187–199.
- [198] Harris, N. C.; Song, Y.; Kiang, C.-H. *Physical Review Letters* **2007**, *99*, 068101.
- [199] Lu, H.; Isralewitz, B.; Krammer, A.; Vogel, V.; Schulten, K. *Biophysical Journal* **1998**, *75*, 662–671.
- [200] Dudko, O. K.; Hummer, G.; Szabo, A. *Proceedings of the National Academy of Sciences* **2008**, *105*, 15755–15760.
- [201] Baştuğ, T.; Chen, P.-C.; Patra, S. M.; Kuyucak, S. *The Journal of Chemical Physics* **2008**, *128*.
- [202] Makowska, J.; Makowski, M.; Liwo, A.; Chmurzyński, L. *Journal of Computational Chemistry* **2005**, *26*, 235–242.
- [203] Branduardi, D.; Gervasio, F. L.; Parrinello, M. *The Journal of Chemical Physics* **2007**, *126*.
- [204] Doudou, S.; Burton, N. A.; Henchman, R. H. *Journal of Chemical Theory and Computation* **2009**, *5*, 909–918.
- [205] Pople, J. A.; Santry, D. P.; Segal, G. A. *The Journal of Chemical Physics* **1965**, *43*, S129–S135.

- [206] Pople, J. A.; Beveridge, D. L.; Dobosh, P. A. *The Journal of Chemical Physics* **1967**, *47*, 2026–2033.
- [207] Dewar, M. J. S.; Thiel, W. *Journal of the American Chemical Society* **1977**, *99*, 4907–4917.
- [208] Dewar, M. J. S.; Thiel, W. *Journal of the American Chemical Society* **1977**, *99*, 4899–4907.
- [209] Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *Journal of the American Chemical Society* **1985**, *107*, 3902–3909.
- [210] Stewart, J. J. P. *Journal of Computational Chemistry* **1989**, *10*, 209–220.
- [211] Stewart, J. J. P. *Journal of Computational Chemistry* **1989**, *10*, 221–264.
- [212] Stewart, J. J. P. *Journal of Molecular Modeling* **2007**, *13*, 1173–1213.
- [213] Stewart, J. J. P. *Journal of Molecular Modeling* **2013**, *19*, 1–32.
- [214] Hostaš, J.; Řezáč, J.; Hobza, P. *Chemical Physics Letters* **2013**, *568–569*, 161–166.
- [215] Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577–2637.
- [216] Hagler, A. T.; Dauber, P.; Lifson, S. *Journal of the American Chemical Society* **1979**, *101*, 5131–5141.
- [217] Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *The Journal of Physical Chemistry B* **2009**, *113*, 6378–6396.
- [218] Sieradzan, A. K.; Czaplewski, C.; Krupa, P.; Mozolewska, M. A.; Karczyńska, A. S.; Lipska, A. G.; Lubecka, E. A.; Gołaś, E.; Wirecki, T.; Makowski, M.; others *Protein Folding: Methods and Protocols* **2022**, 399–416.
- [219] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *The Journal of Chemical Physics* **1983**, *79*, 926–935.
- [220] Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *Journal of Computational Chemistry* **1992**, *13*, 1011–1021.

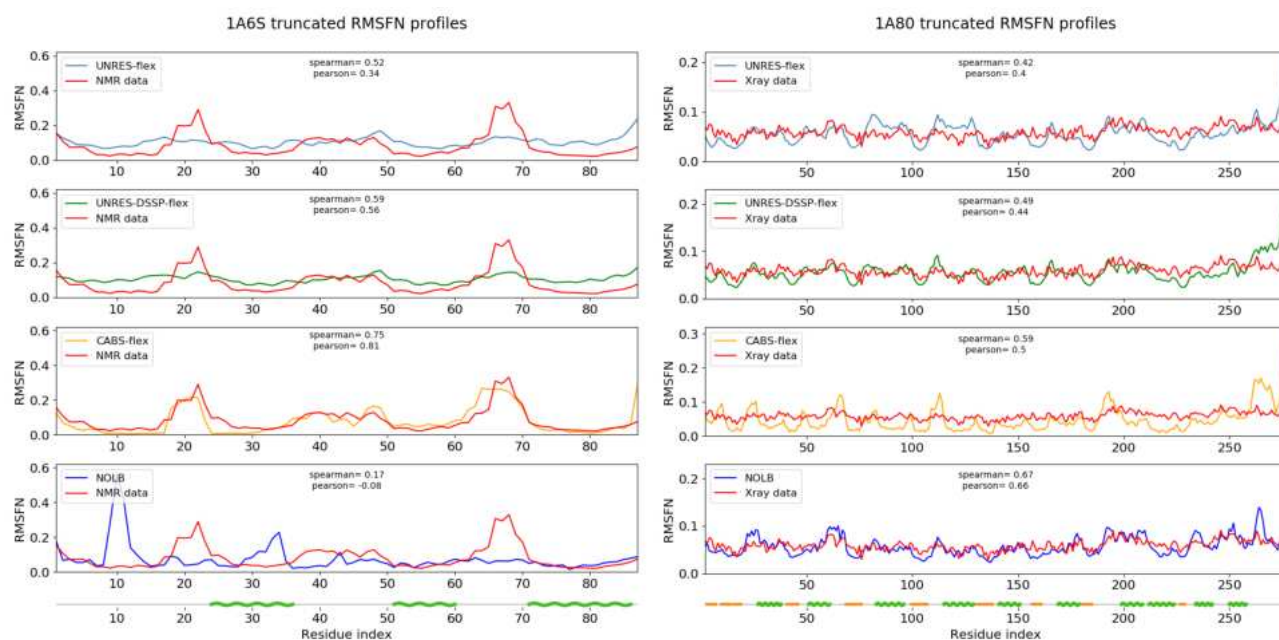
- [221] Makowski, M.; Liwo, A.; Scheraga, H. A. *The Journal of Physical Chemistry B* **2007**, *111*, 2910–2916.
- [222] Makowski, M.; Liwo, A.; Maksimiak, K.; Makowska, J.; Scheraga, H. A. *The Journal of Physical Chemistry B* **2007**, *111*, 2917–2924.
- [223] Makowski, M.; Sobolewski, E.; Czaplewski, C.; Liwo, A.; Oldziej, S.; No, J. H.; Scheraga, H. A. *The Journal of Physical Chemistry B* **2007**, *111*, 2925–2931.
- [224] Makowski, M.; Sobolewski, E.; Czaplewski, C.; Oldziej, S.; Liwo, A.; Scheraga, H. A. *The Journal of Physical Chemistry B* *112*, 11385–11395.
- [225] Gay, J. G.; Berne, B. J. *The Journal of Chemical Physics* **1981**, *74*, 3316–3319.
- [226] Modest, M. F. *Radiative Heat Transfer* **2013**, 779–802.
- [227] Jo, S.; Kim, T.; Iyer, V. G.; Im, W. *Journal of Computational Chemistry* **2008**, *29*, 1859–1865.
- [228] Roe, D. R.; Cheatham III, T. E. *Journal of Chemical Theory and Computation* **2013**, *9*, 3084–3095.
- [229] Gholizadeh, N.; Saadatfar, H.; Hanafi, N. *The Journal of supercomputing* **2021**, *77*, 6214–6235.
- [230] Huige, C. J. M.; Altona, C. *Journal of Computational Chemistry* **1995**, *16*, 56–79.
- [231] Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. *SoftwareX* **2015**, *1-2*, 19–25.
- [232] Kuzmanic, A.; Zagrovic, B. *Biophysical Journal* **2010**, *98*, 861–871.
- [233] Dehouck, Y.; Bastolla, U. *Integrative Biology* **2017**, *9*, 627–641.
- [234] Ahlgren, P.; Jarneving, B.; Rousseau, R. *Journal of the American Society for Information Science and Technology* **2003**, *54*, 550–560.
- [235] Shanmugam, R. *Journal of Statistical Computation and Simulation* **2005**, *79*, 1275–1276.
- [236] De Winter, J. C. E.; Gosling, S. D.; Potter, J. *Psychological Methods* **2016**, *21*, 273–290.

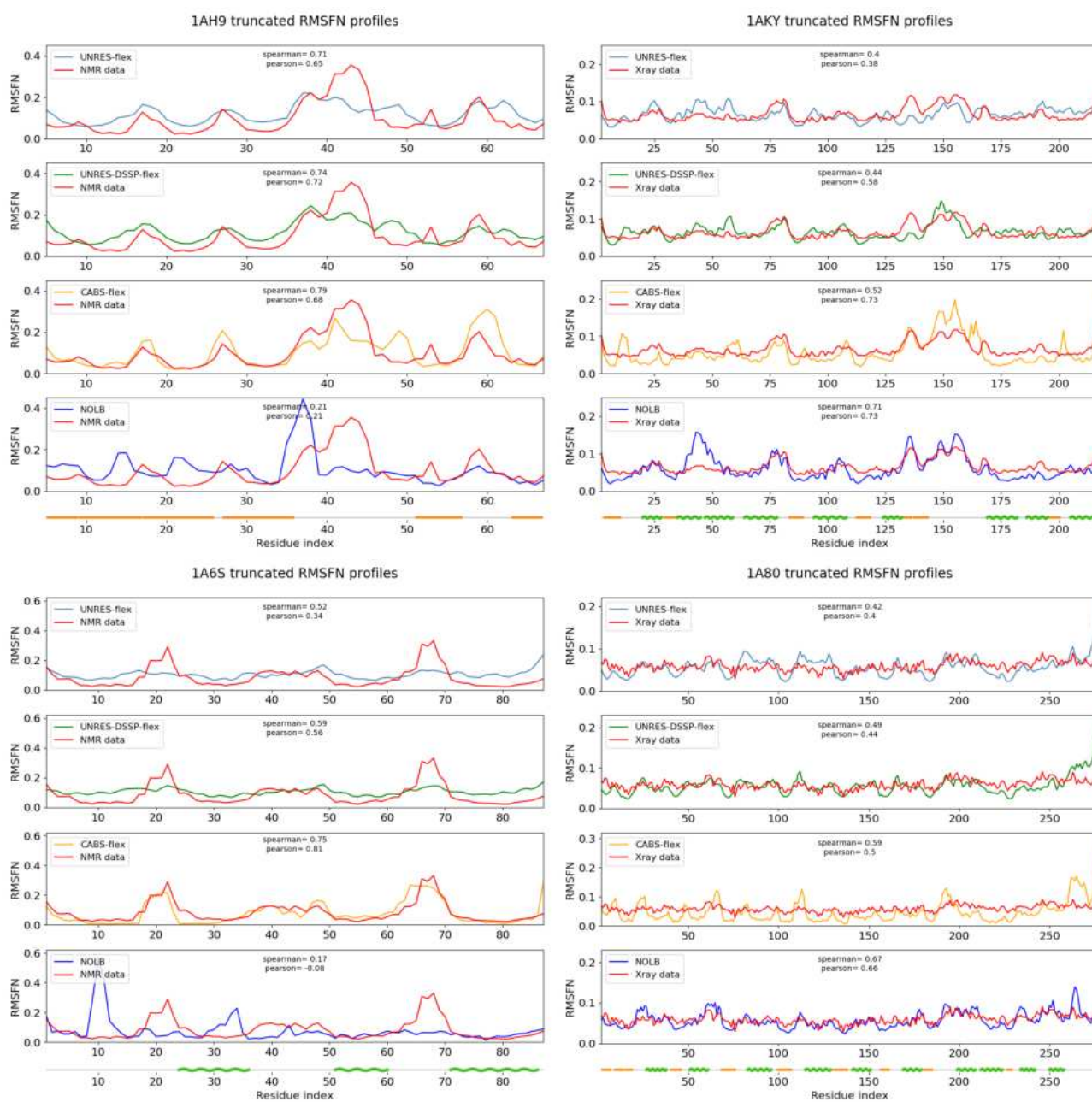
- [237] Statistics Kingdom *Statistics Kingdom [web application]* **2017**,
- [238] Maiorov, V. N.; Crippen, G. M. *Journal of Molecular Biology* **1994**, *235*, 625–634.
- [239] Hung, L.-H.; Samudrala, R. *Bioinformatics* **2012**, *28*, 2191–2192.
- [240] Lawrence, J.; Bernal, J.; Witzgall, C. *Journal of Research of the National Institute of Standards and Technology* **2019**, *124*, 124028.
- [241] Suetake, H. F. Y. S., T. <https://doi.org/10.2210/pdb2crb/pdb>.
- [242] David, H. A.; Gunnink, J. L. *The American Statistician* **1997**, *51*, 9.
- [243] McDonnell, J. M.; Fushman, D.; Cahill, S. M.; Zhou, W.; Wolven, A.; Wilson, C. B.; Nelle, T. D.; Resh, M. D.; Wills, J.; Cowburn, D. *Journal of Molecular Biology* **1998**, *279*, 921–928.
- [244] Szebenyi, D. M.; Moffat, K. *Journal of Biological Chemistry* **1986**, *261*, 8761–8777.
- [245] Philippe, D. L.; Ladbury, J. E.; Pfuhl, M. <https://doi.org/10.2210/pdb2k9g/pdb>.
- [246] Vorobieva, A. A.; White, P.; Liang, B.; Horne, J. E.; Bera, A. K.; Chow, C. M.; Gerben, S.; Marx, S.; Kang, A.; Stiving, A. Q.; Harvey *Science* **2021**, *371*, eabc8182.
- [247] Katti, S. S.; Krieger, I. V.; Ann, J.; Lee, J.; Sacchettini, J. C.; Igumenova, T. I. *Nature Communications* **2022**, *13*, 2695.
- [248] Eisenreichova, A.; Klima, M.; Anila, M.; Koukalova, A.; Humpolickova, J.; Różycki, B.; Boura, E. *Cells* **2023**, *12*, 1974.
- [249] Yun, Y.-R.; Won, J. E.; Jeon, E.; Lee, S.; Kang, W.; Jo, H.; Jang, J.-H.; Shin, U. S.; Kim, H.-W. *Journal of tissue engineering* **2010**, *1*, 218142.
- [250] Lecaille, F.; Brömme, D.; Lalmanach, G. *Biochimie* **2008**, *90*, 208–226.
- [251] Faham, S.; Hileman, R.; Fromm, J.; Linhardt, R.; Rees, D. *Science* **1996**, *271*, 1116–1120.
- [252] DiGabriele, A. D.; Lax, I.; Chen, D. I.; Svahn, C. M.; Jaye, M.; Schlessinger, J.; Hendrickson, W. A. *Nature* **1998**, *393*, 812–817.

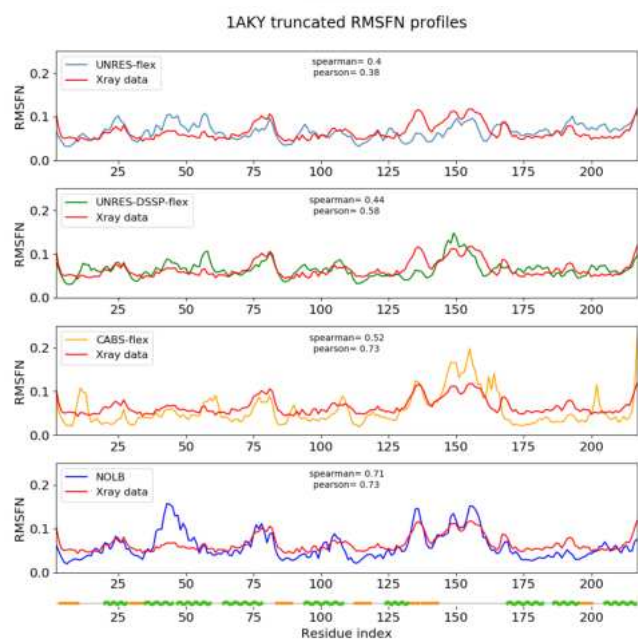
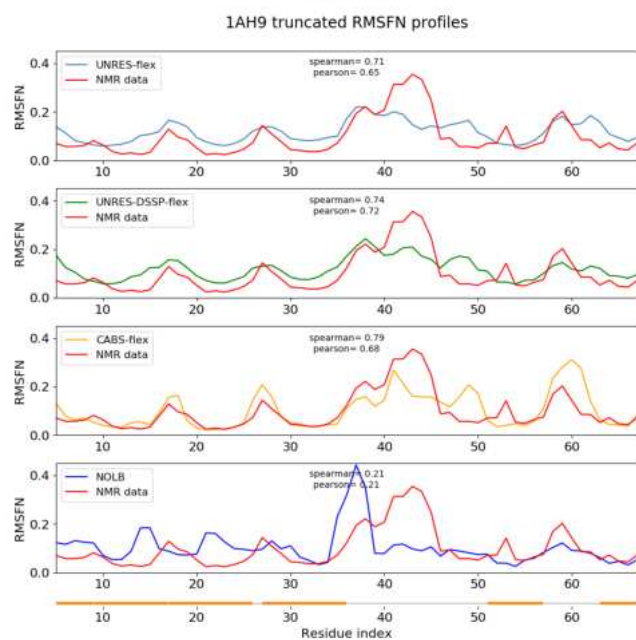
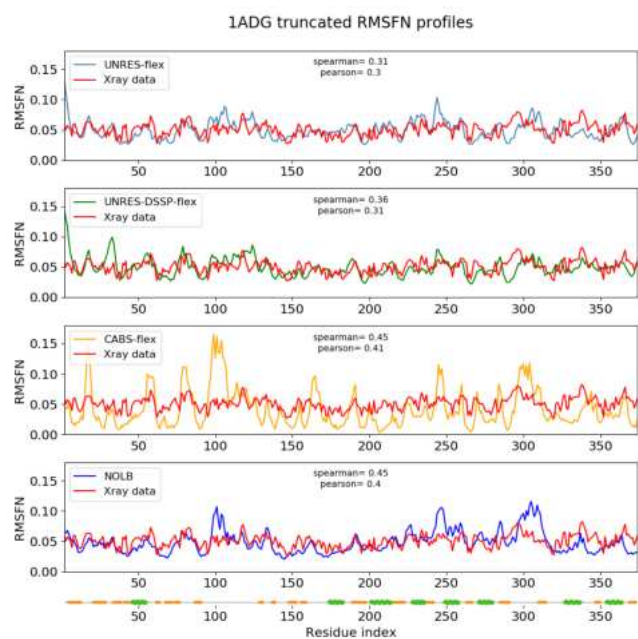
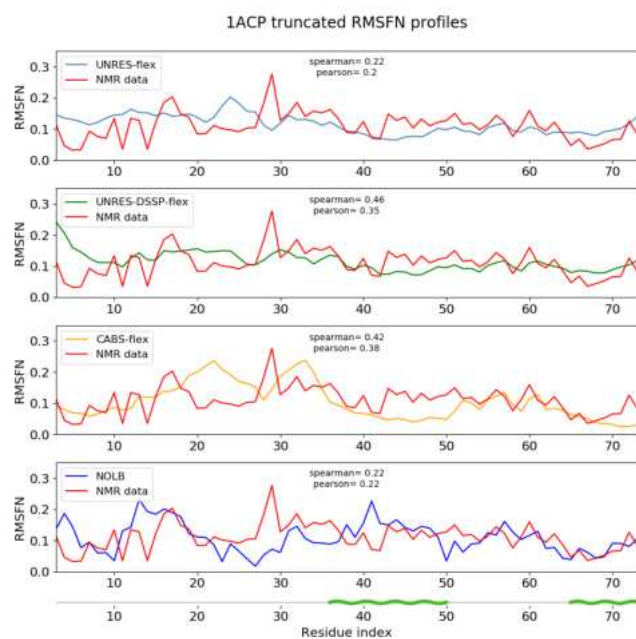
- [253] Aguda, A. H.; Panwar, P.; Du, X.; Nguyen, N. T.; Brayer, G. D.; Brömme, D. *Proceedings of the National Academy of Sciences* **2014**, *111*, 17474–17479.
- [254] Ago, H.; Kitagawa, Y.; Fujishima, A.; Matsuura, Y.; Katsube, Y. *The Journal of Biochemistry* **1991**, *110*, 360–363.
- [255] Law, S.; Andrault, P.-M.; Aguda, A. H.; Nguyen, N. T.; Kruglyak, N.; Brayer, G. D.; Brömme, D. *Biochemical Journal* **2017**, *474*, 851–864.
- [256] Raman, R.; Venkataraman, G.; Ernst, S.; Sasisekharan, V.; Sasisekharan, R. *Proceedings of the National Academy of Sciences* **2003**, *100*, 2357–2362.
- [257] Bojarski, K. K.; David, A.; Lecaille, F.; Samsonov, S. A. *Carbohydrate Research* **2024**, 109201.

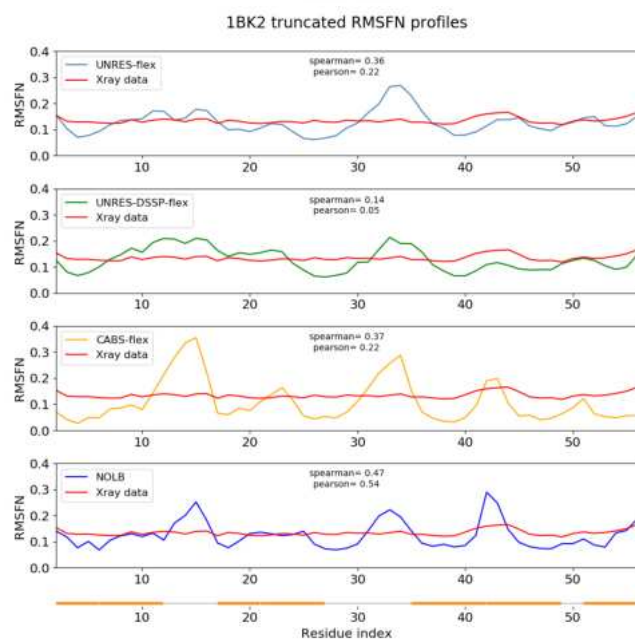
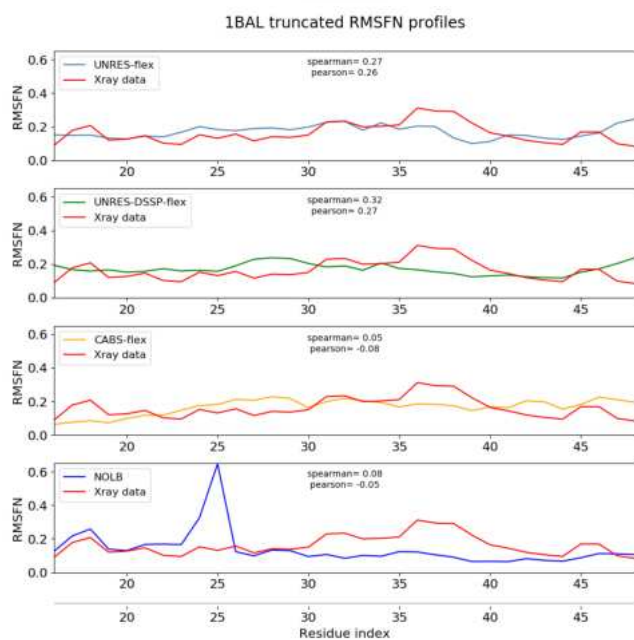
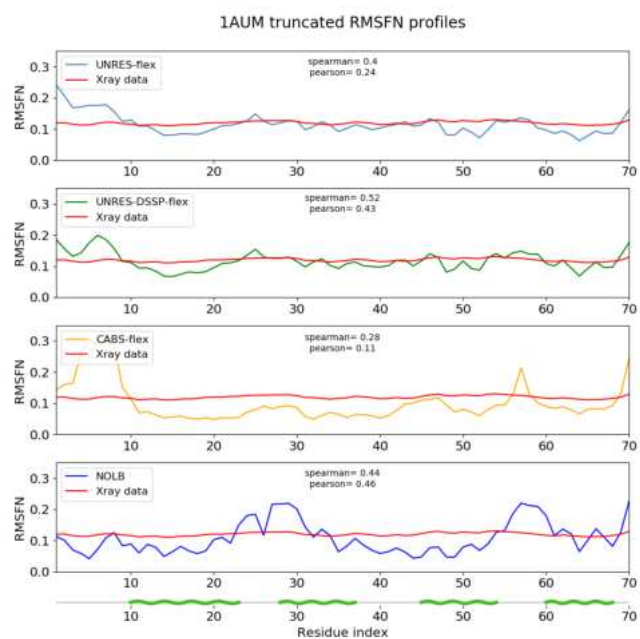
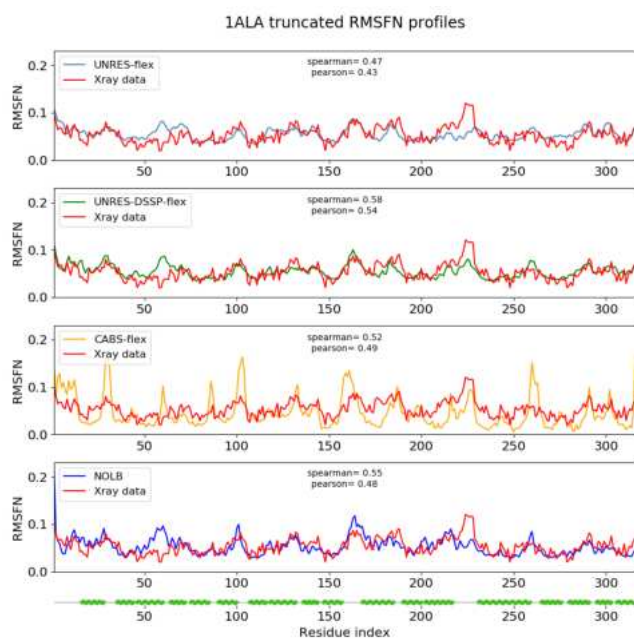
Rozdział 7

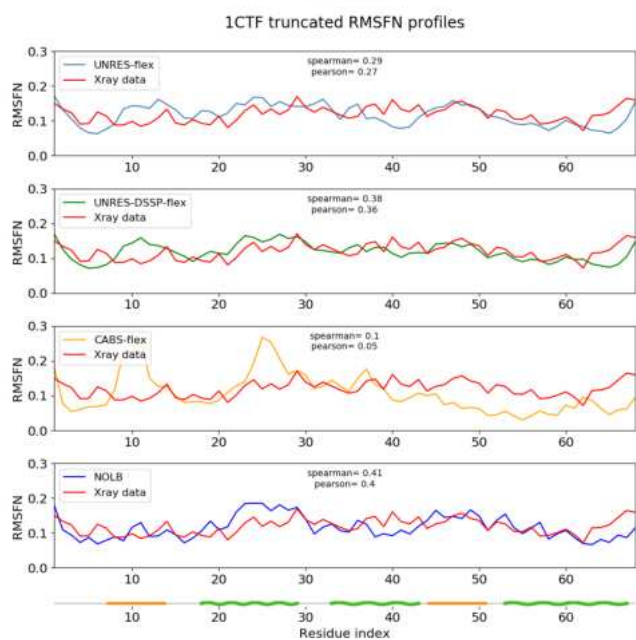
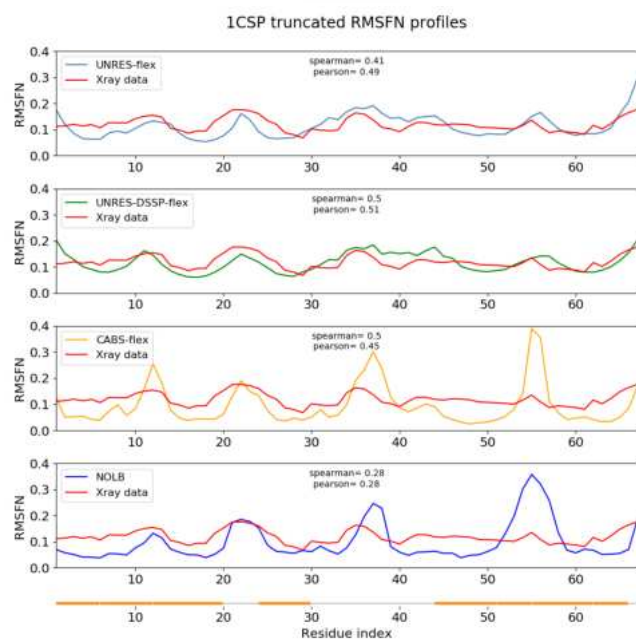
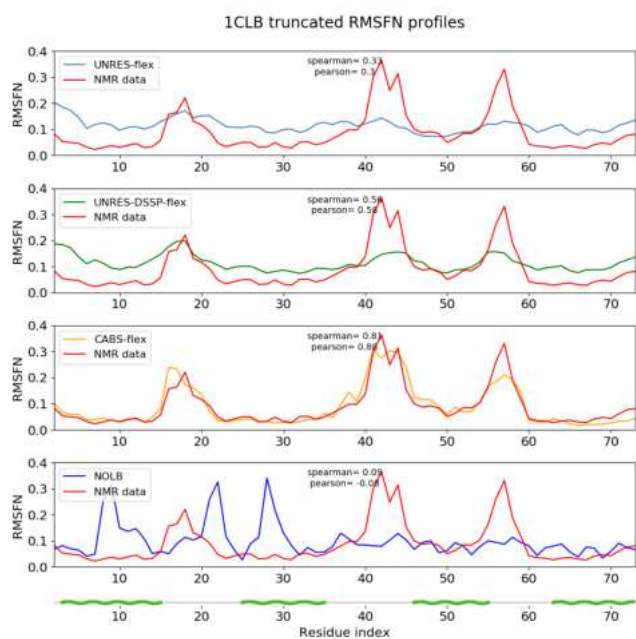
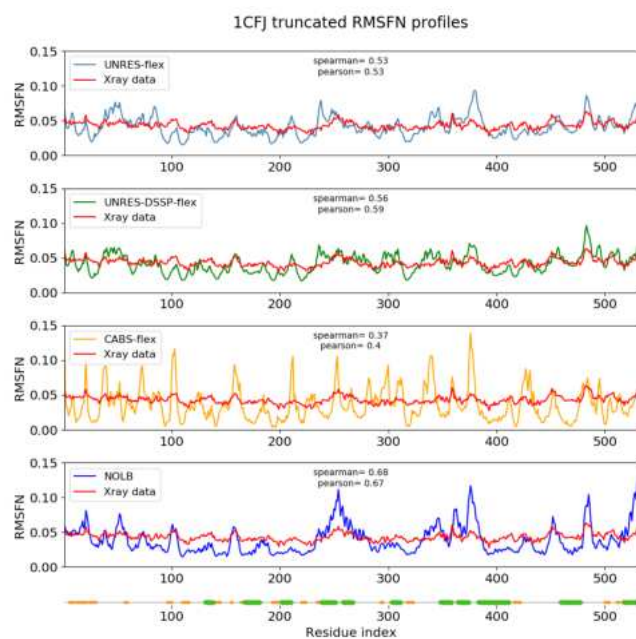
Załączniki

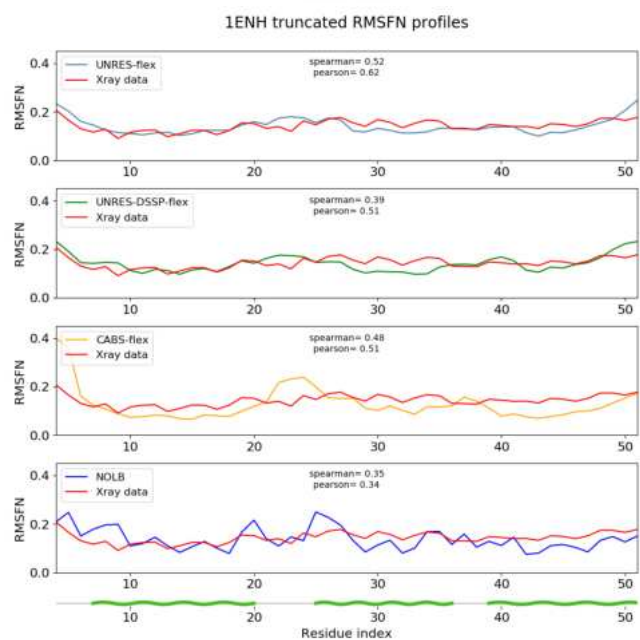
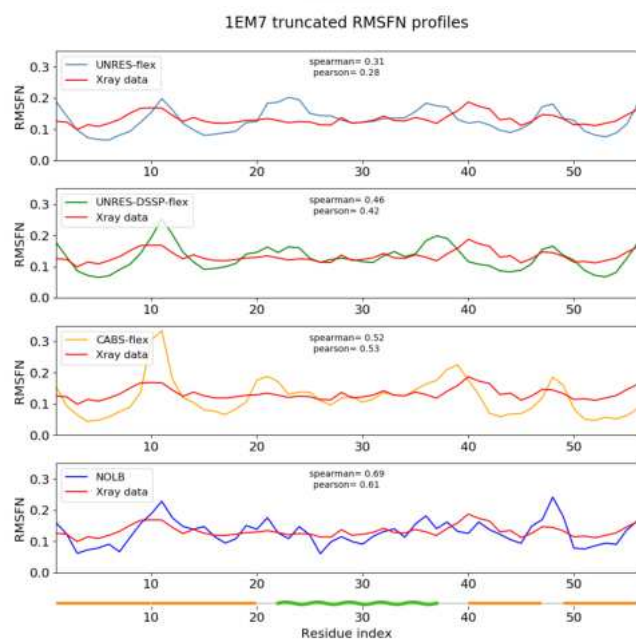
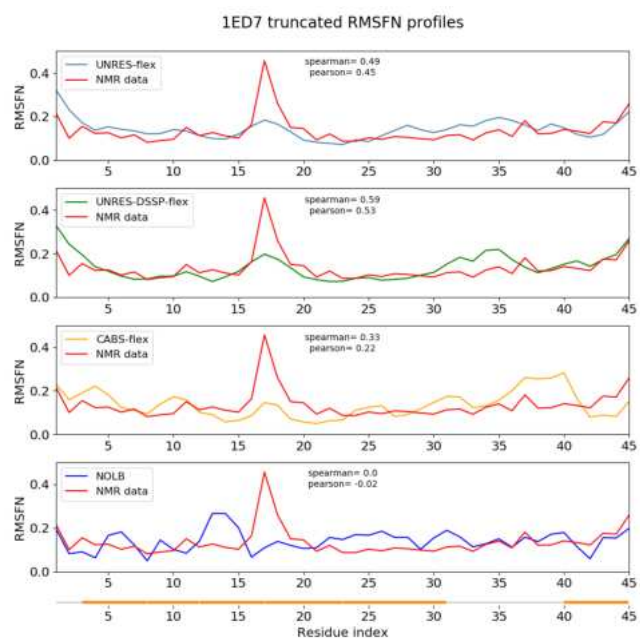
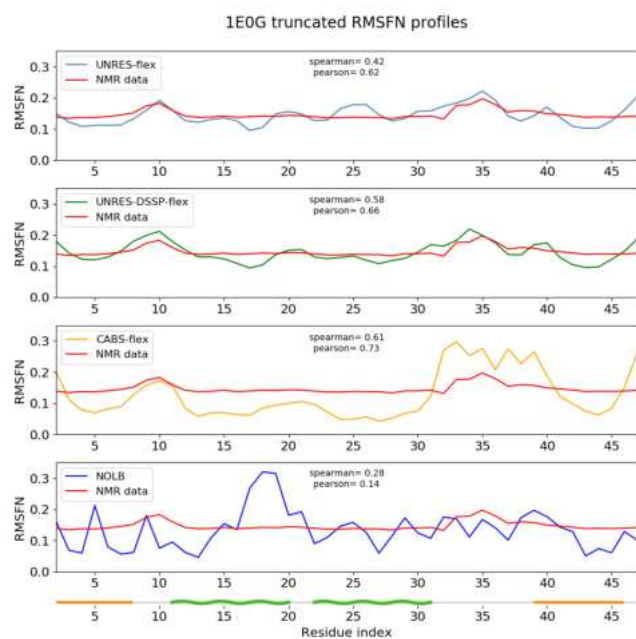


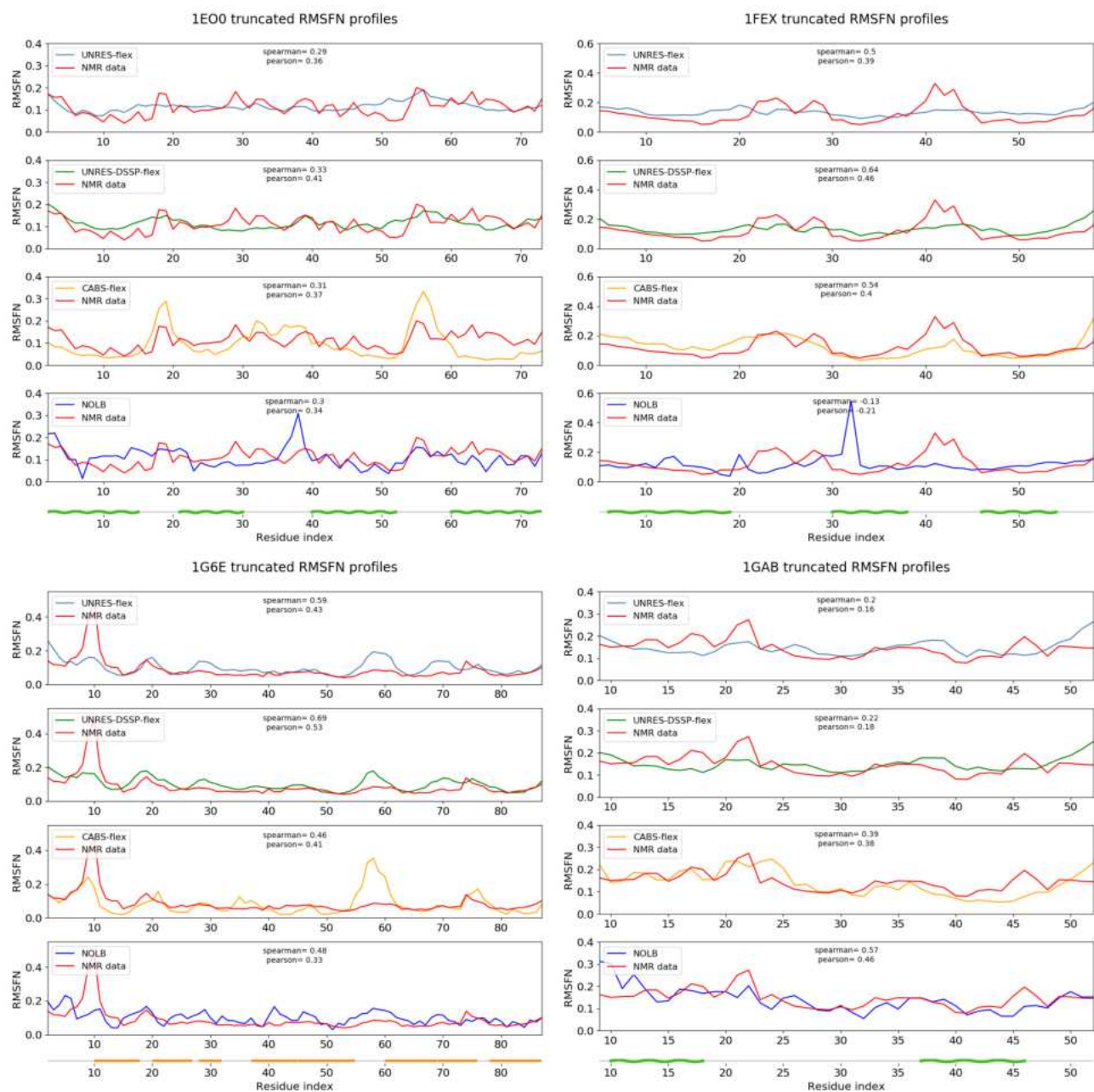


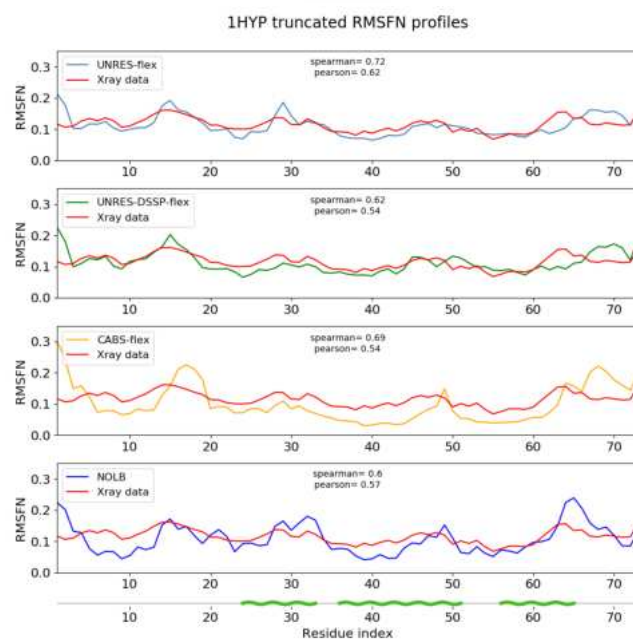
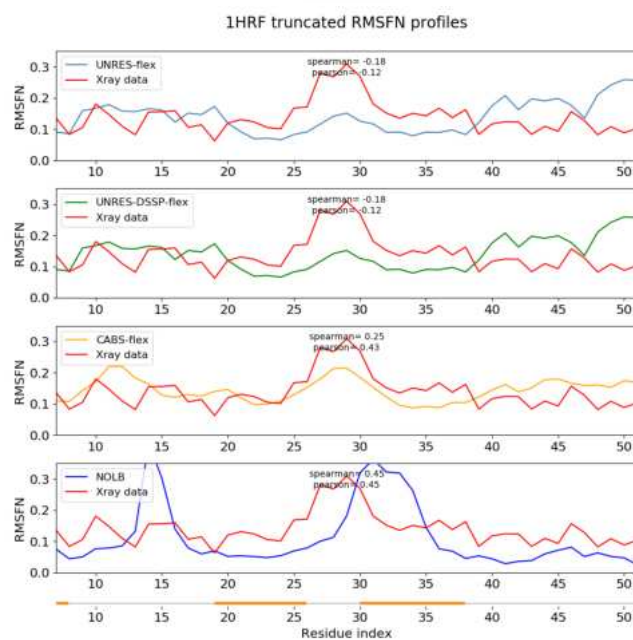
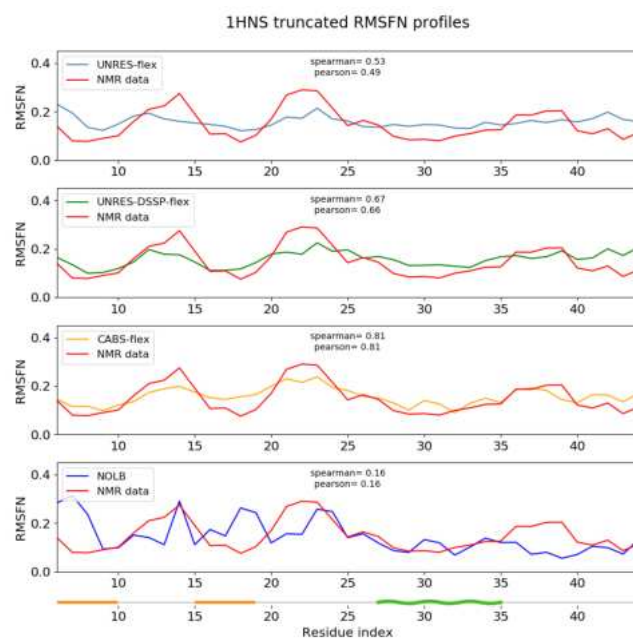
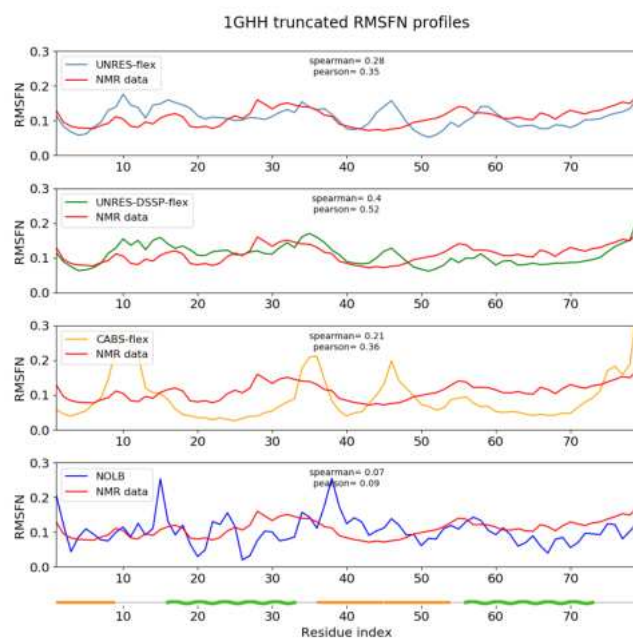


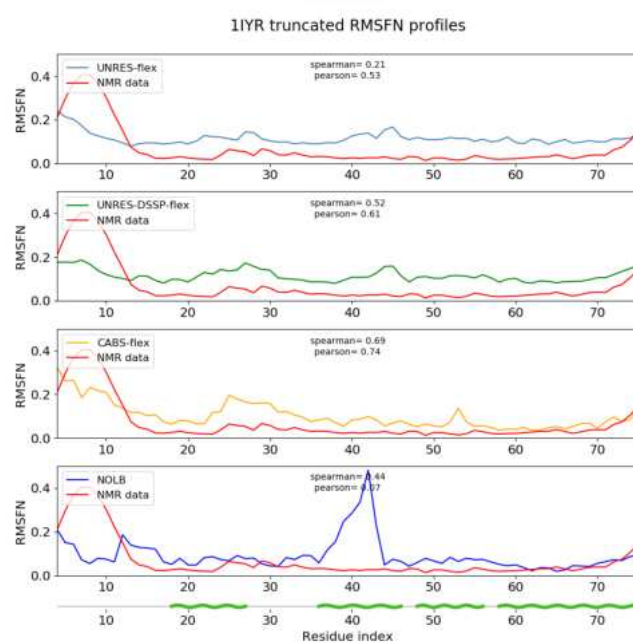
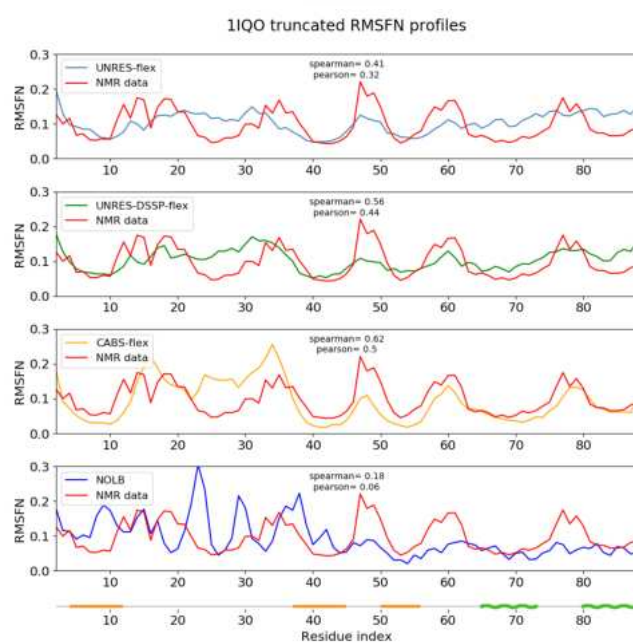
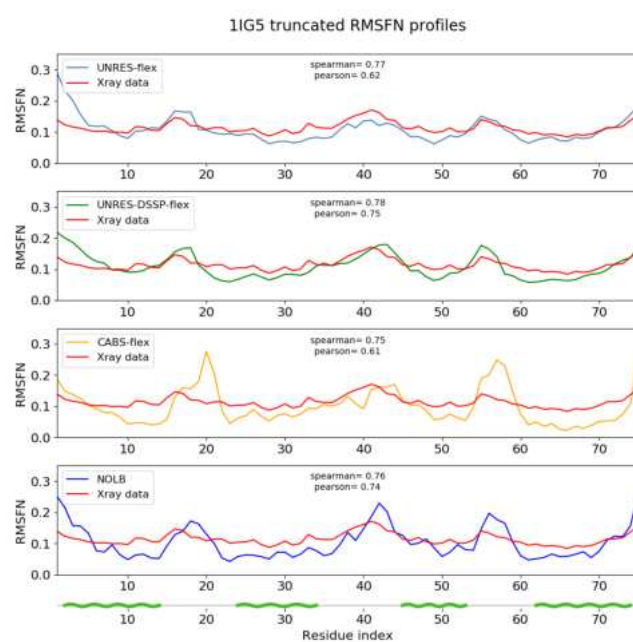
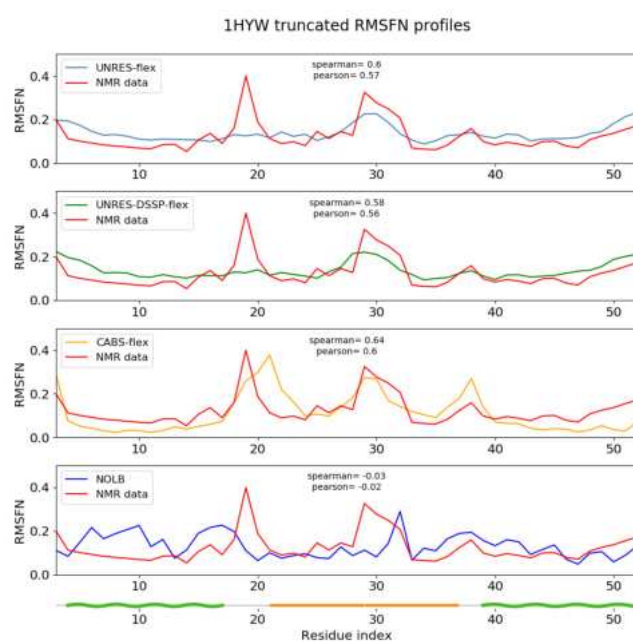


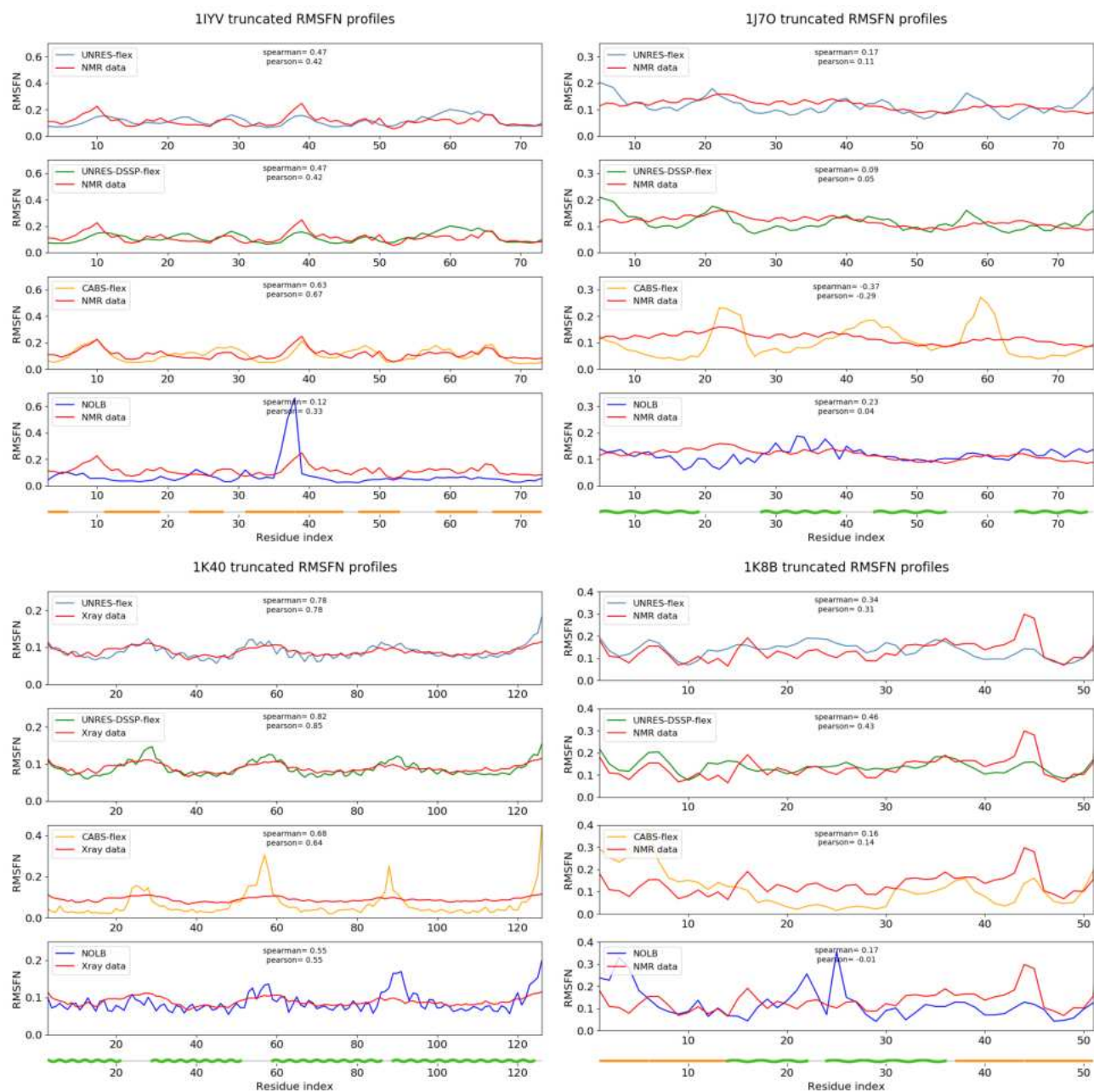


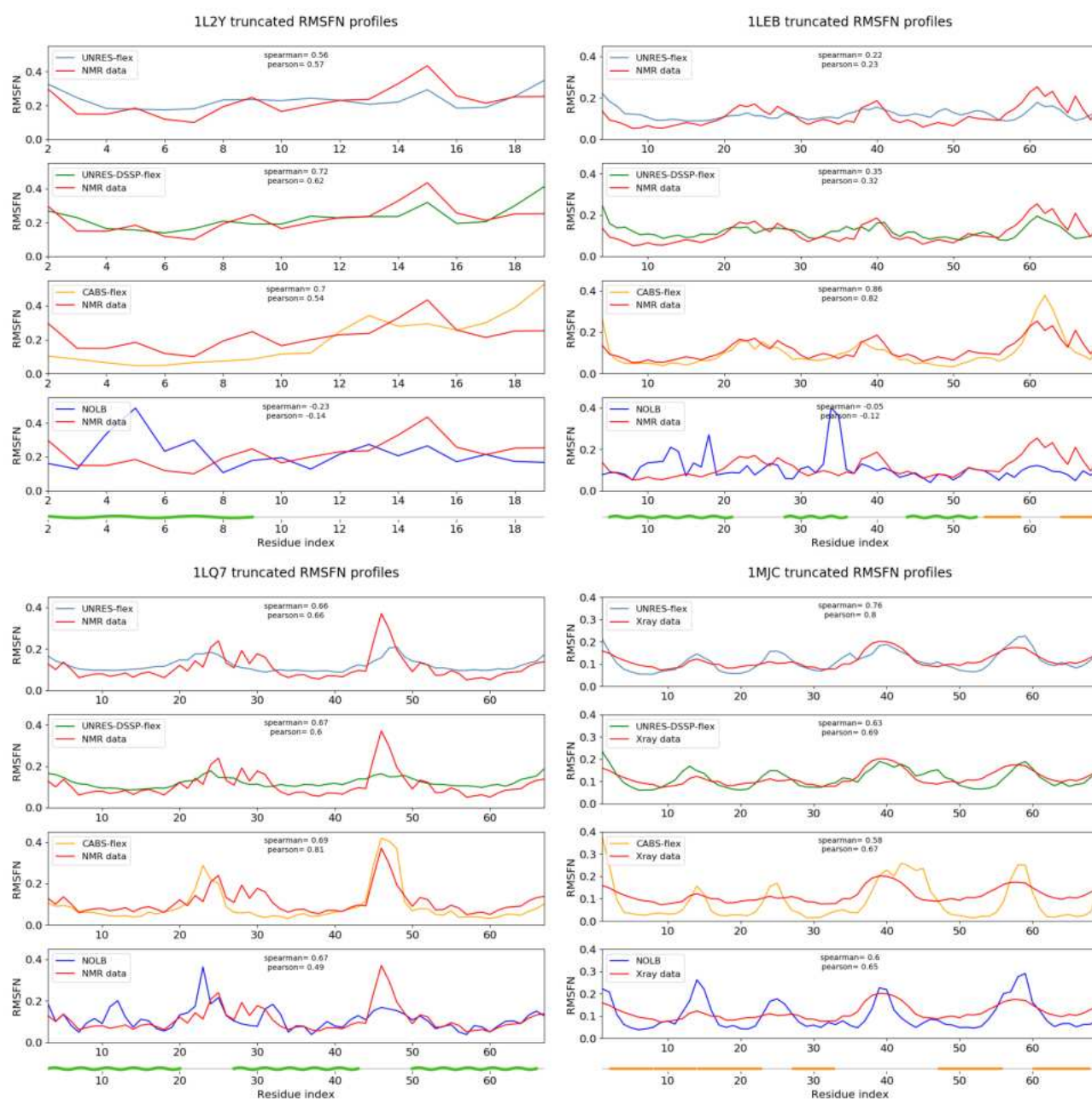


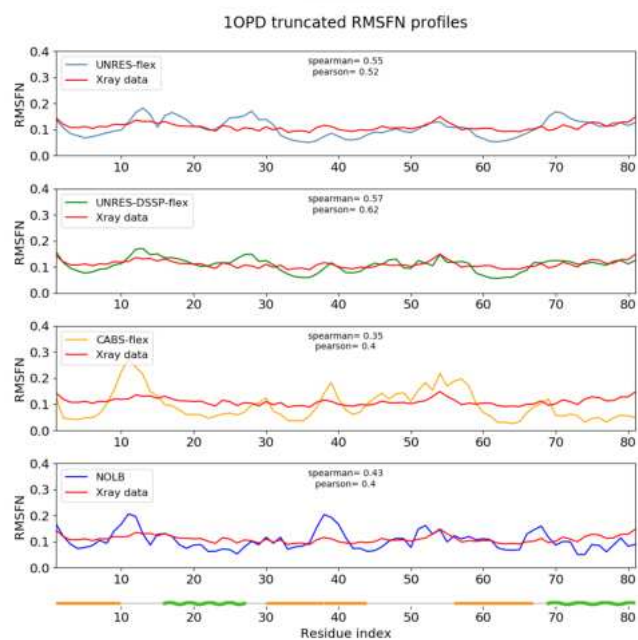
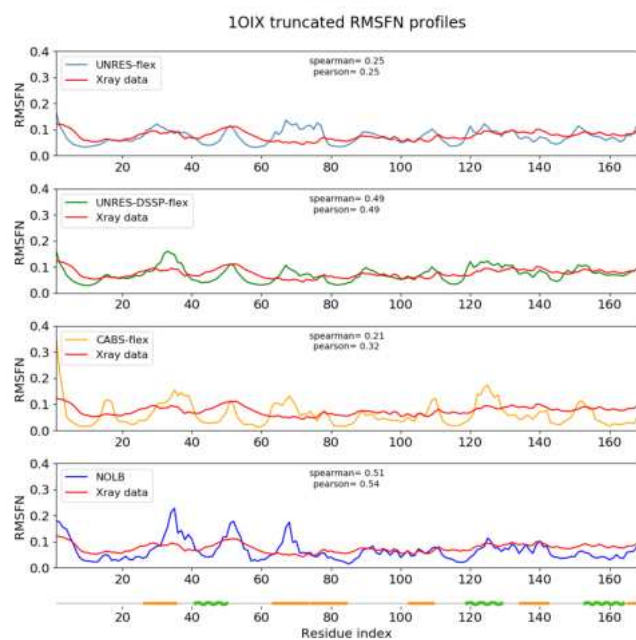
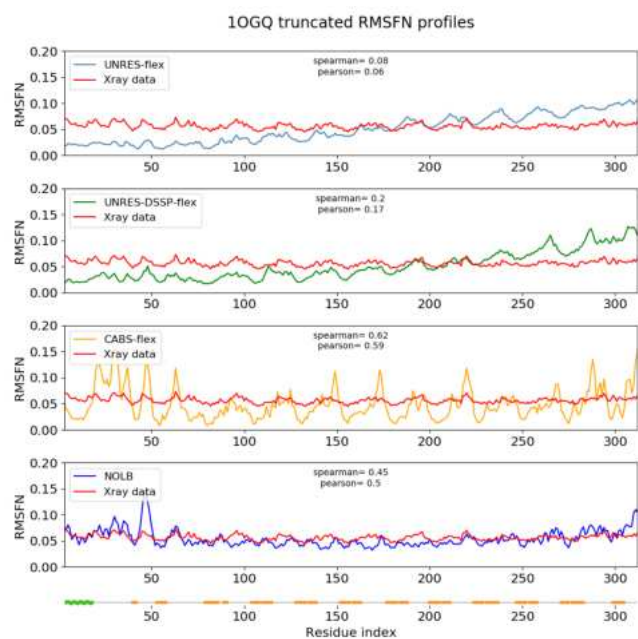
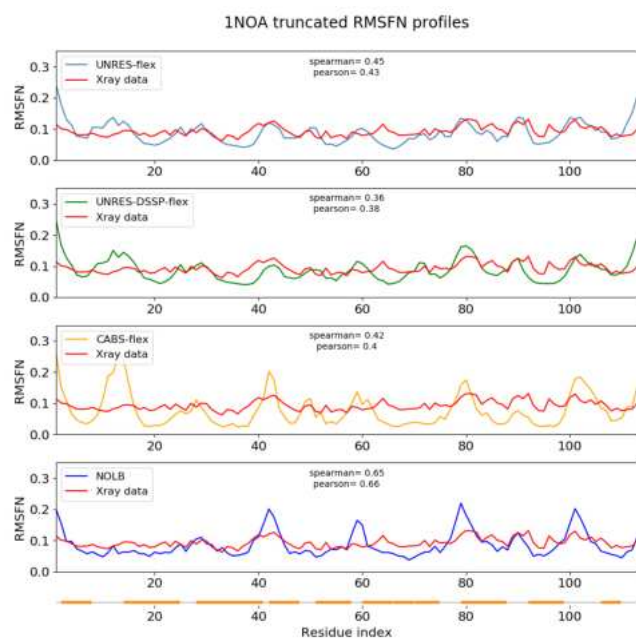


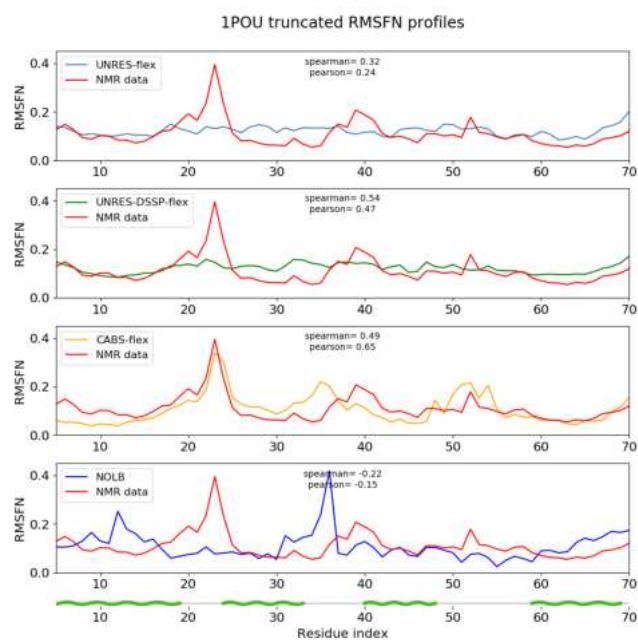
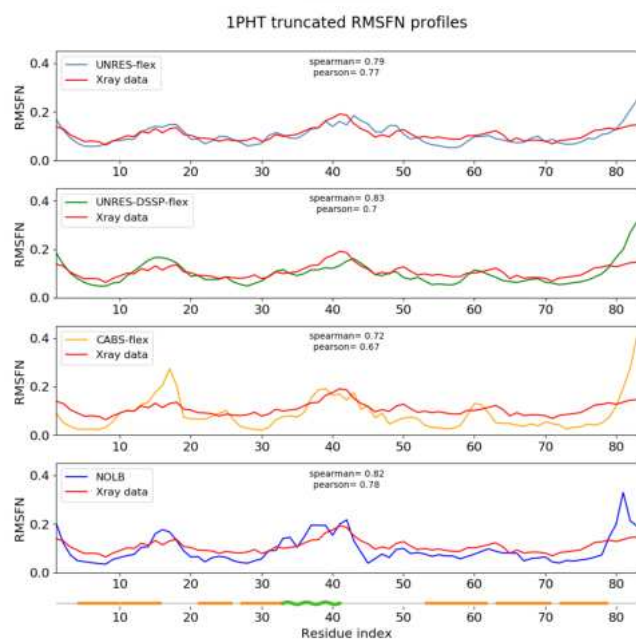
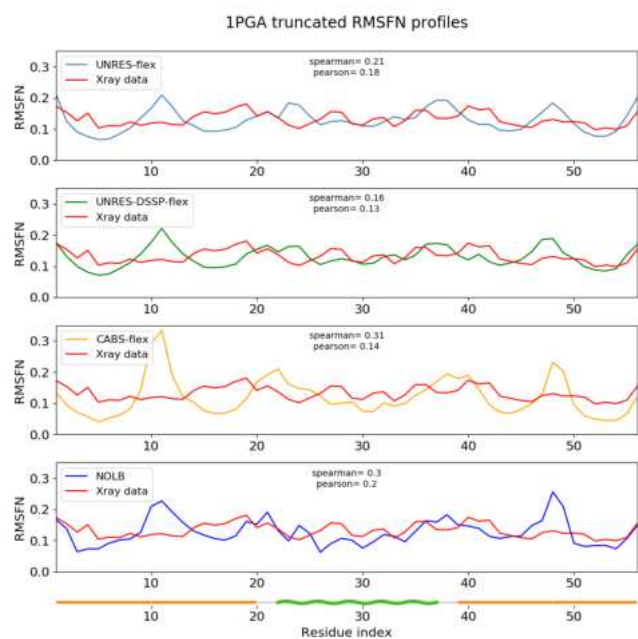
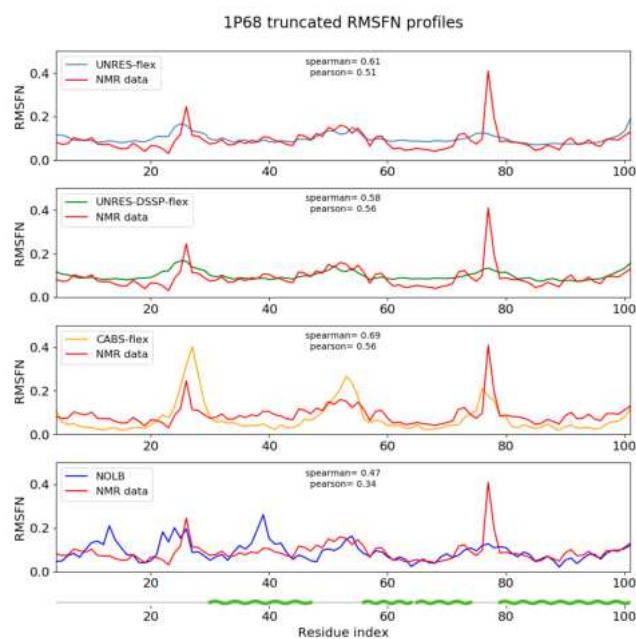


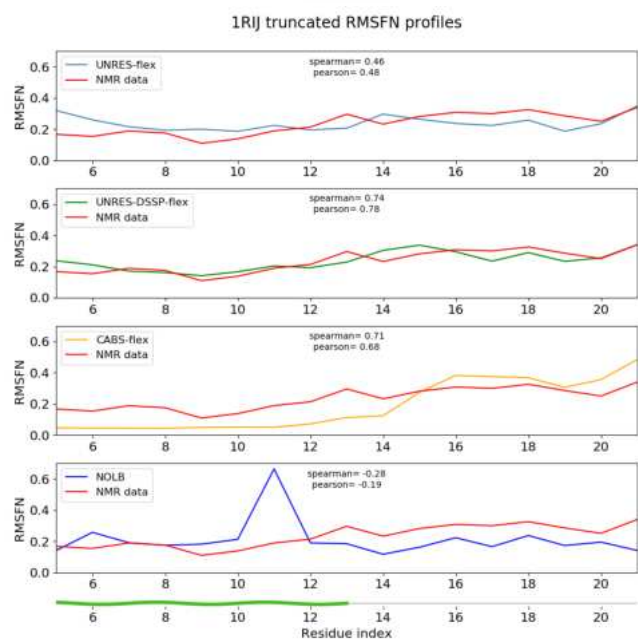
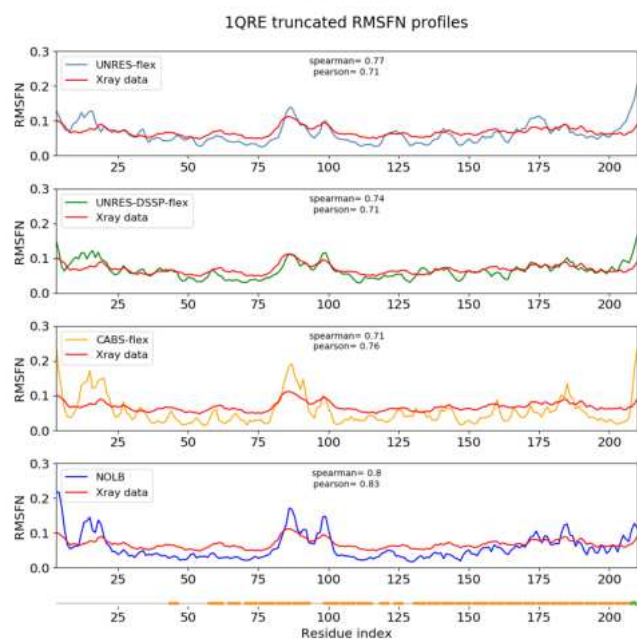
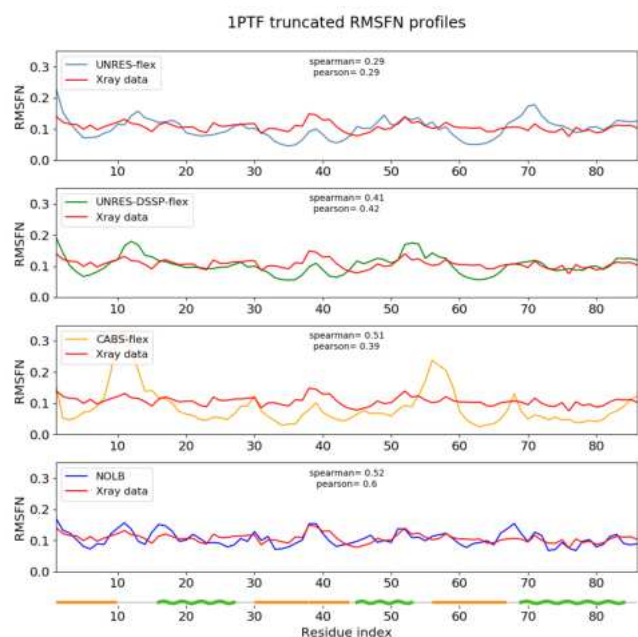
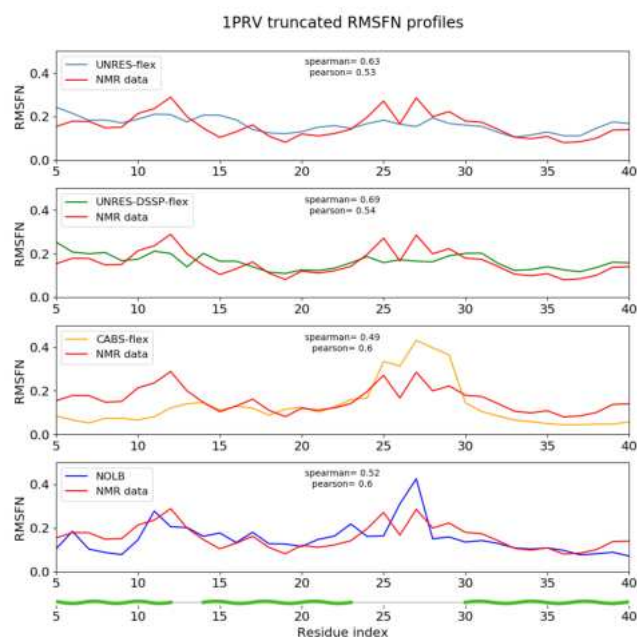


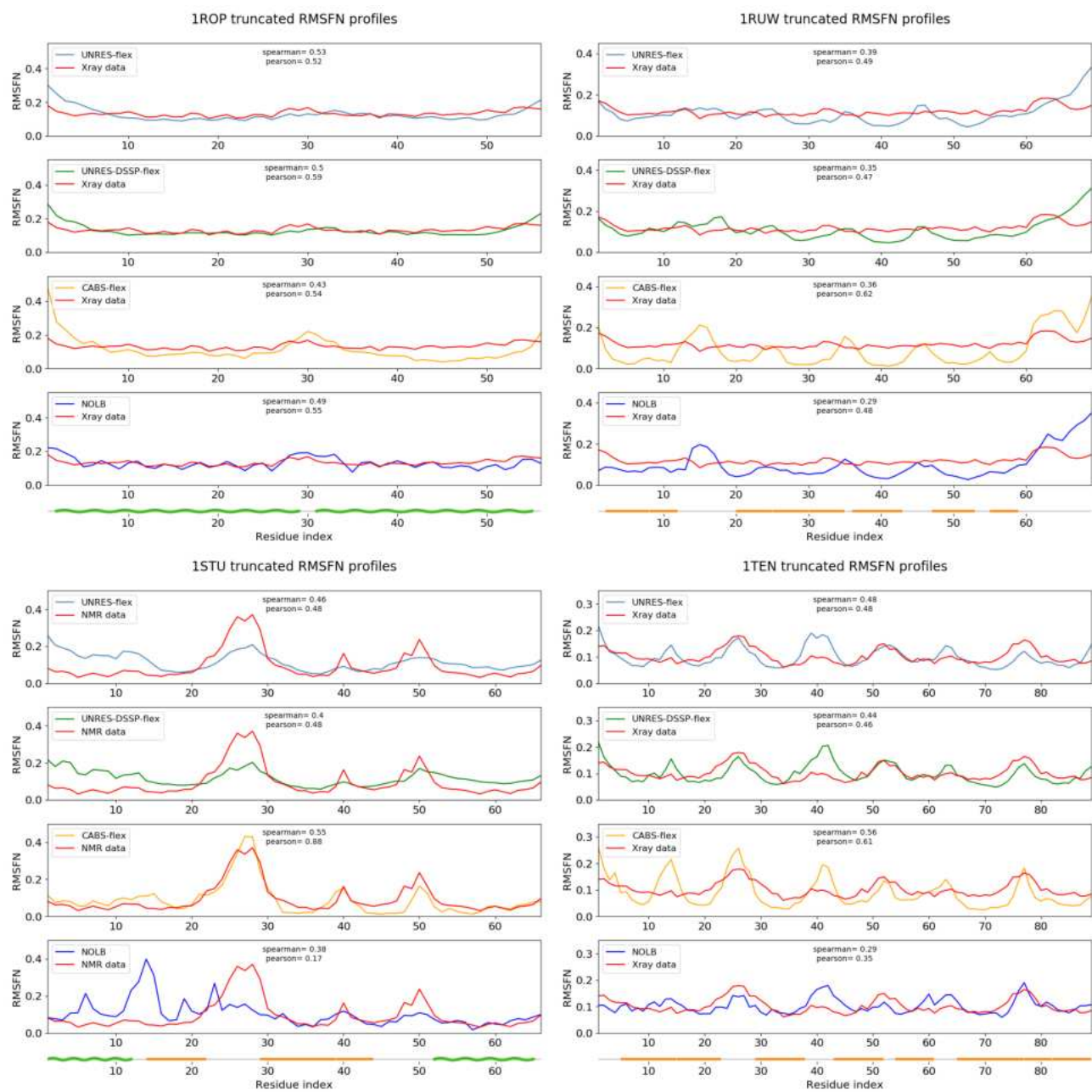


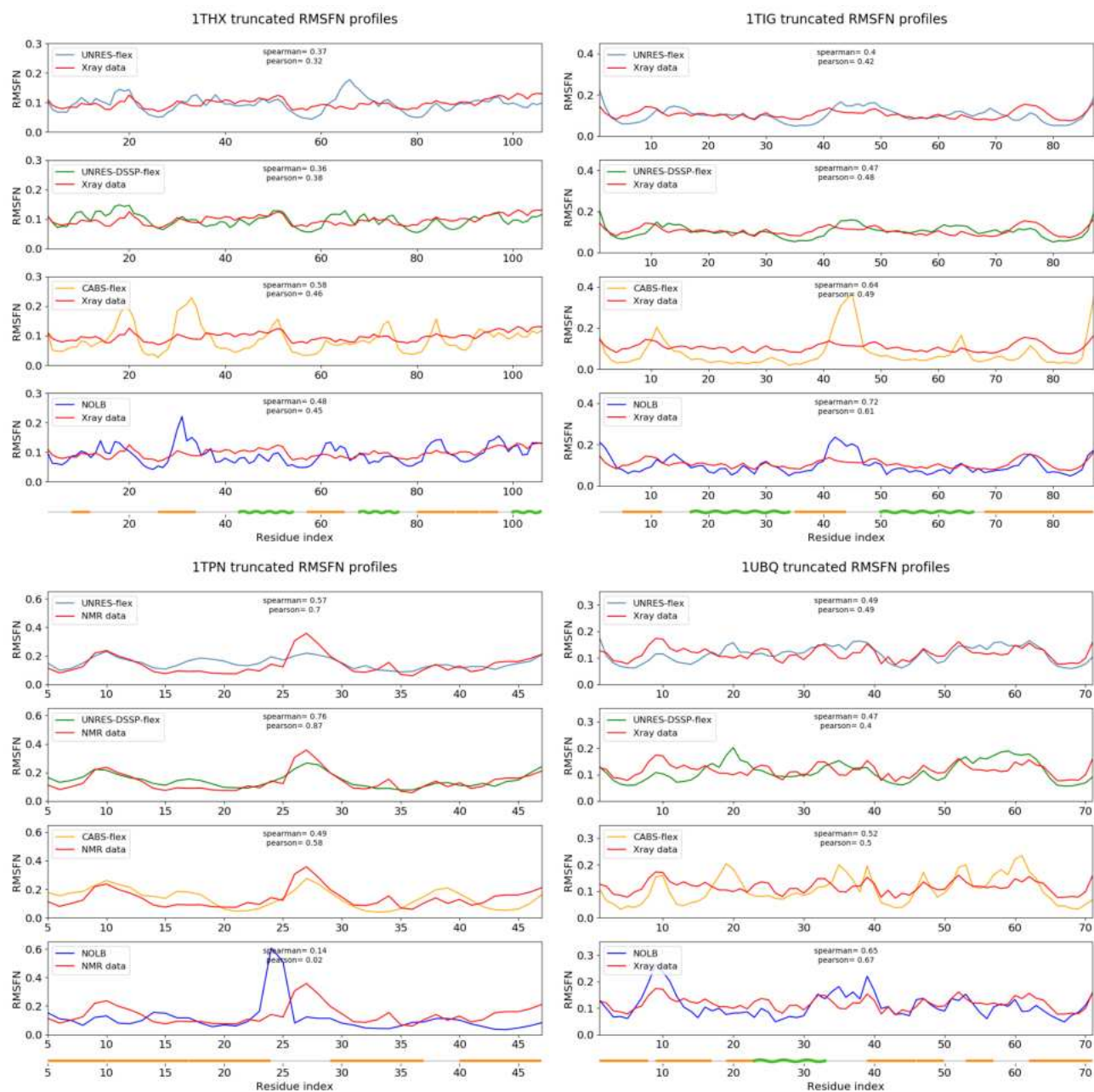


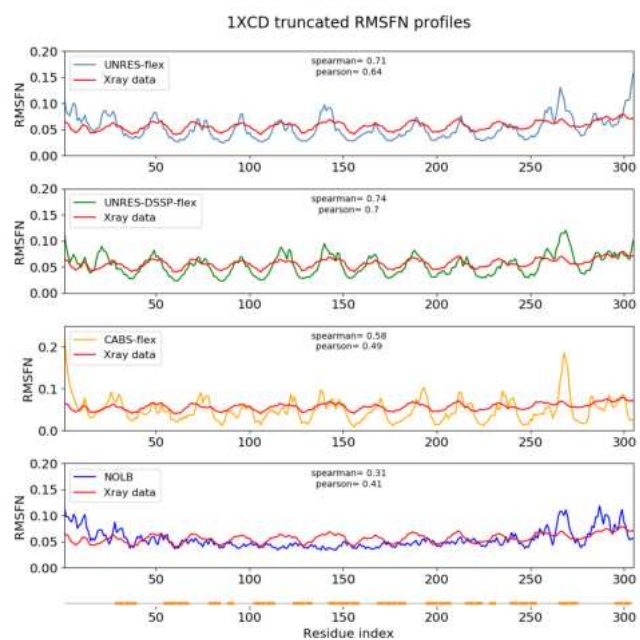
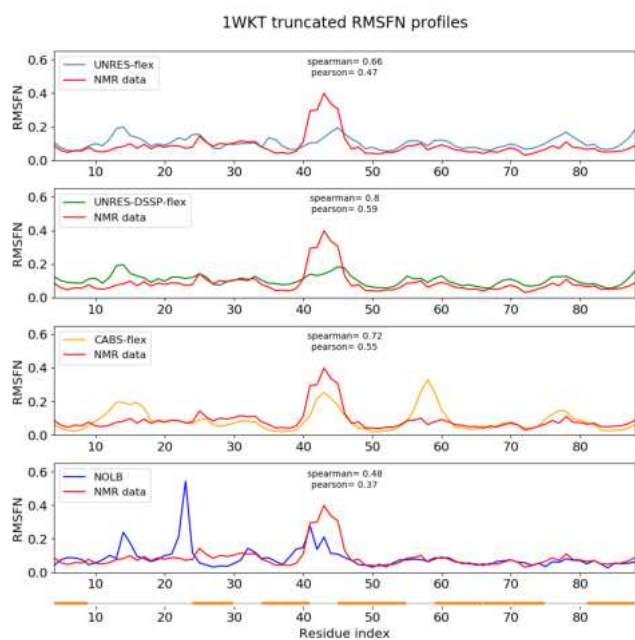
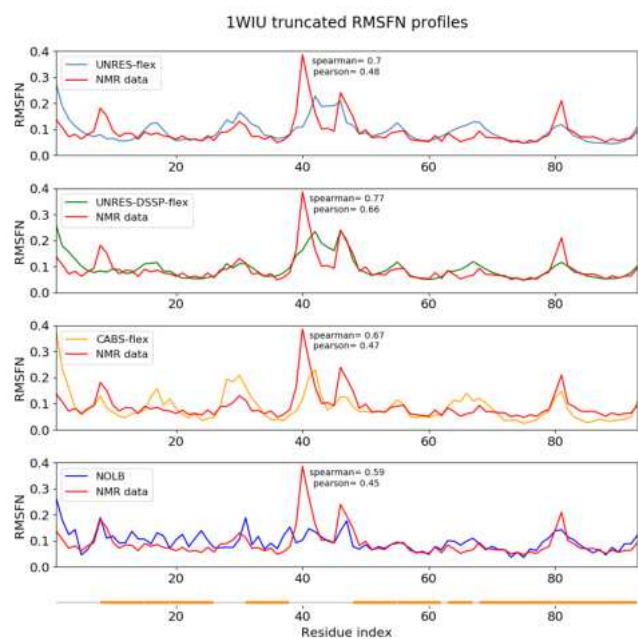
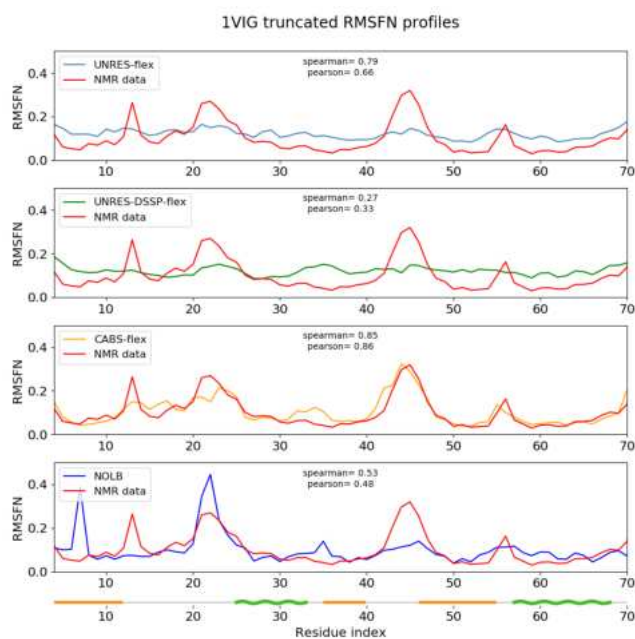


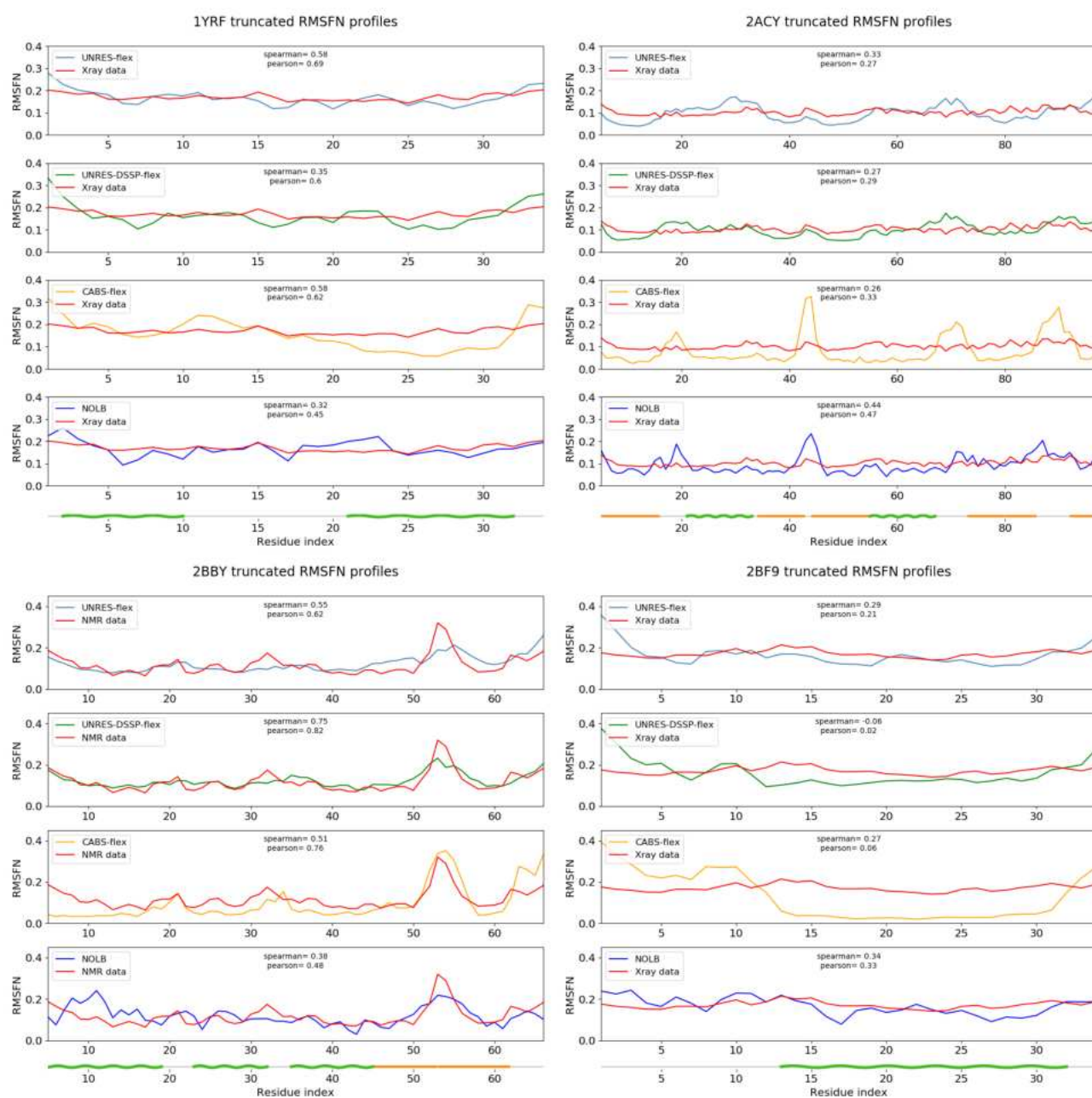


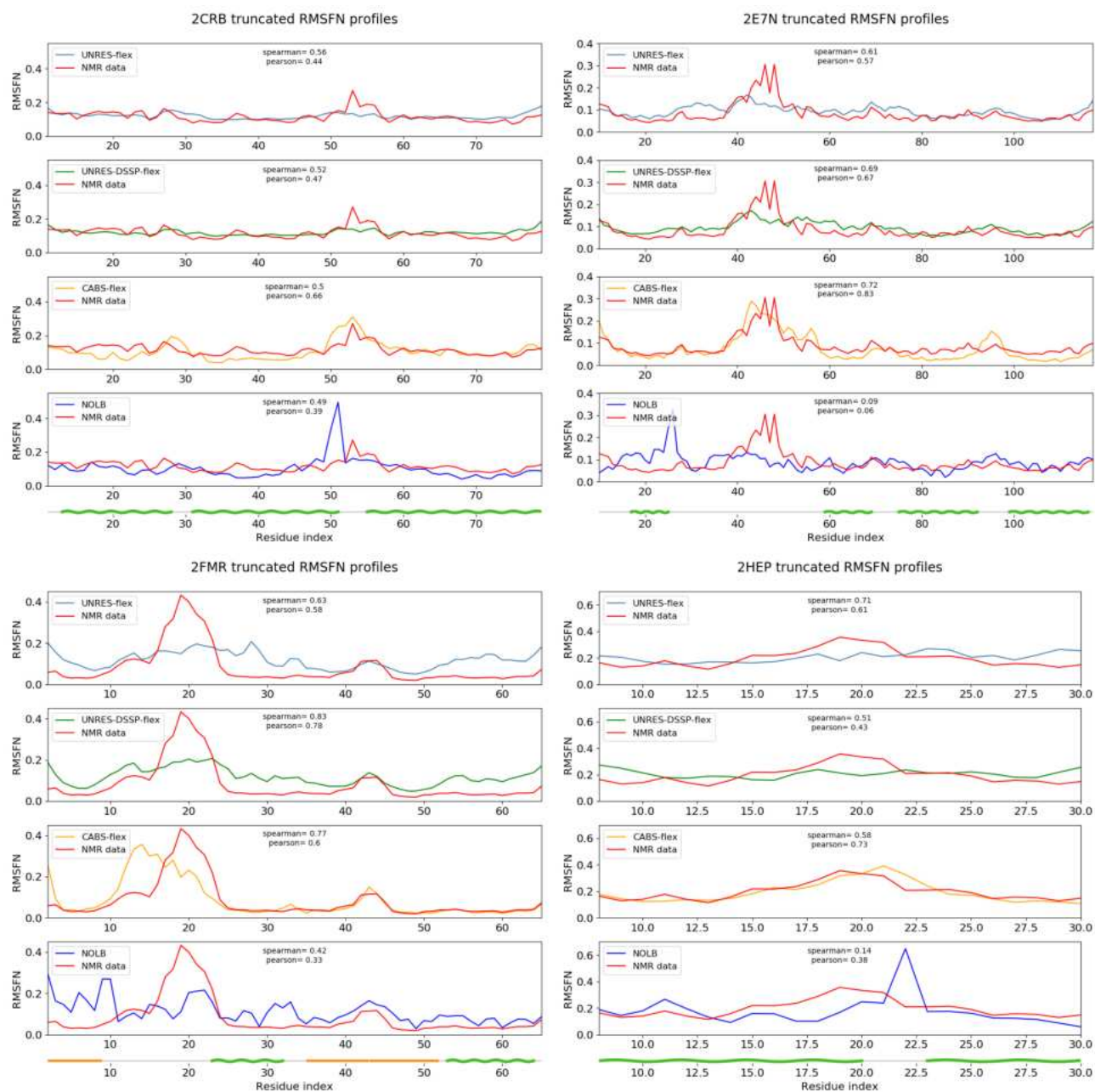


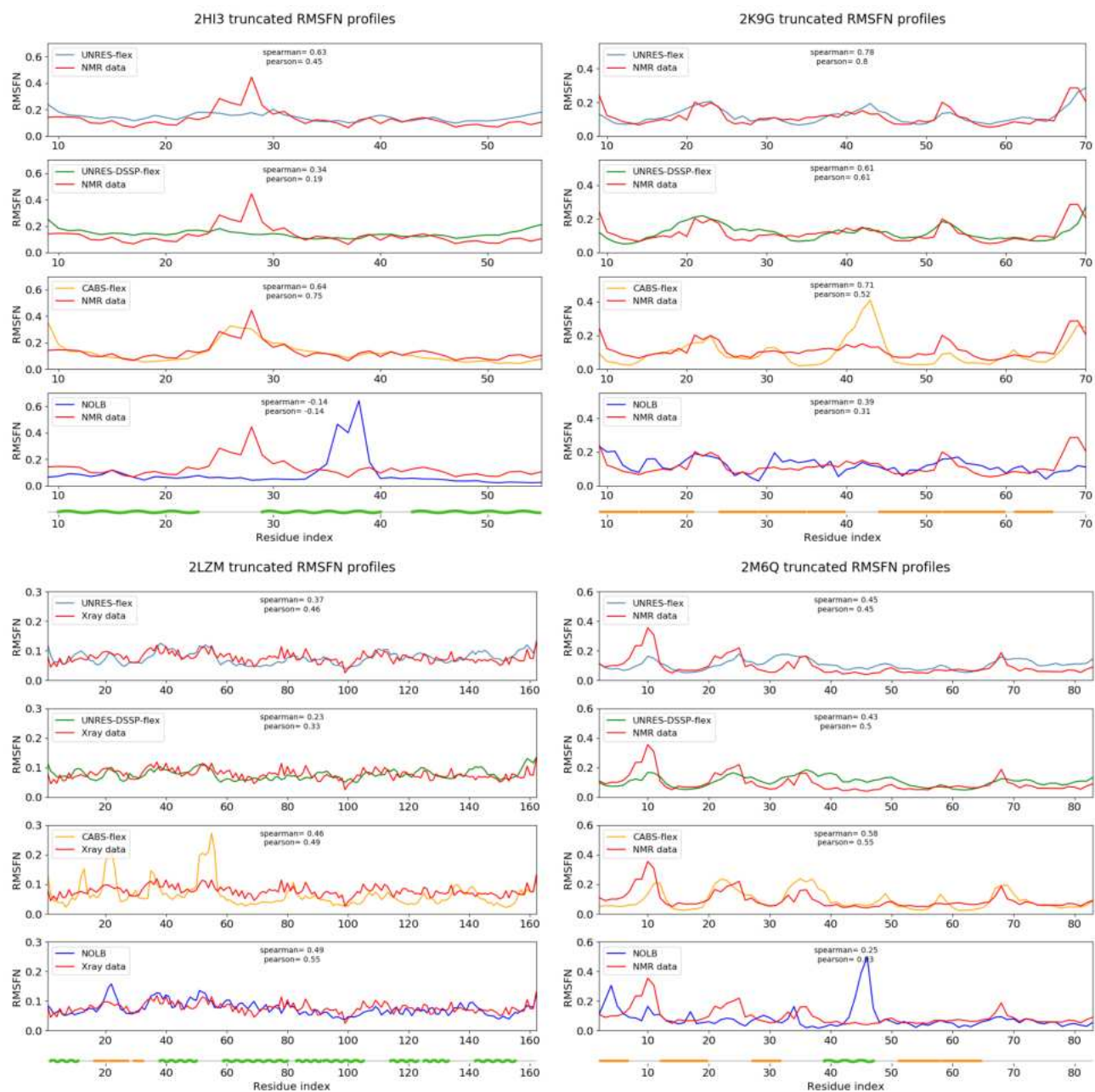


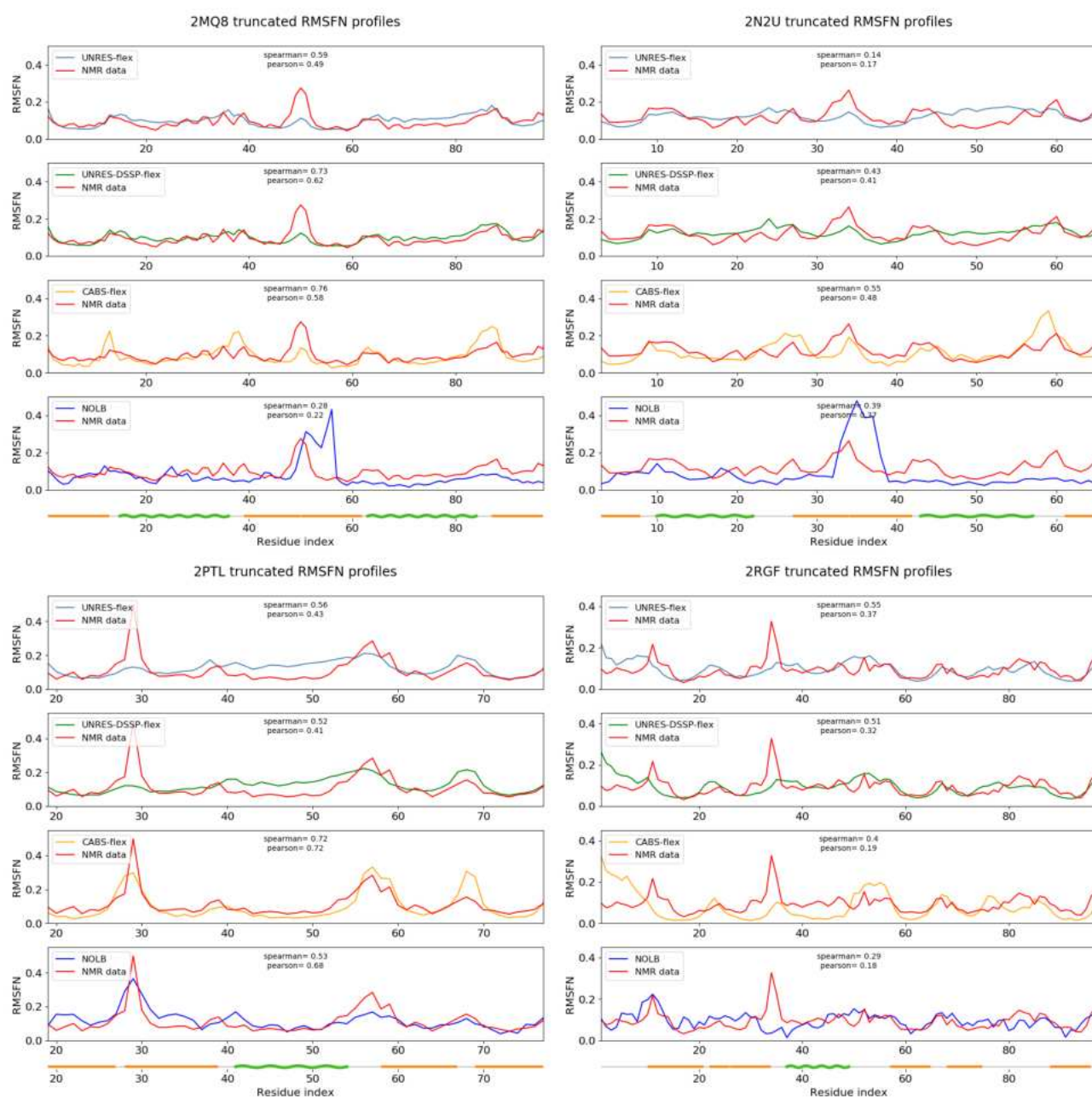


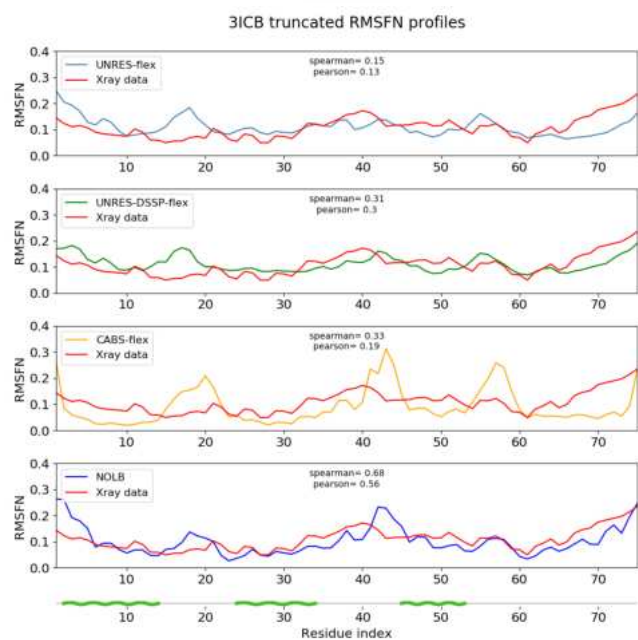
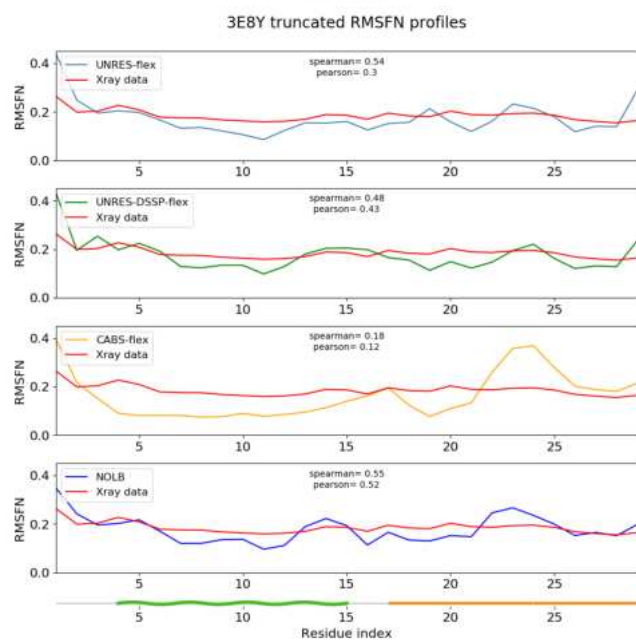
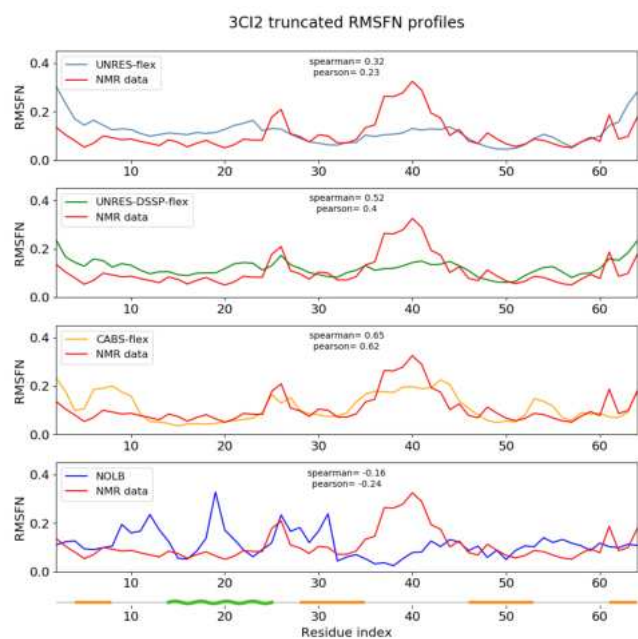
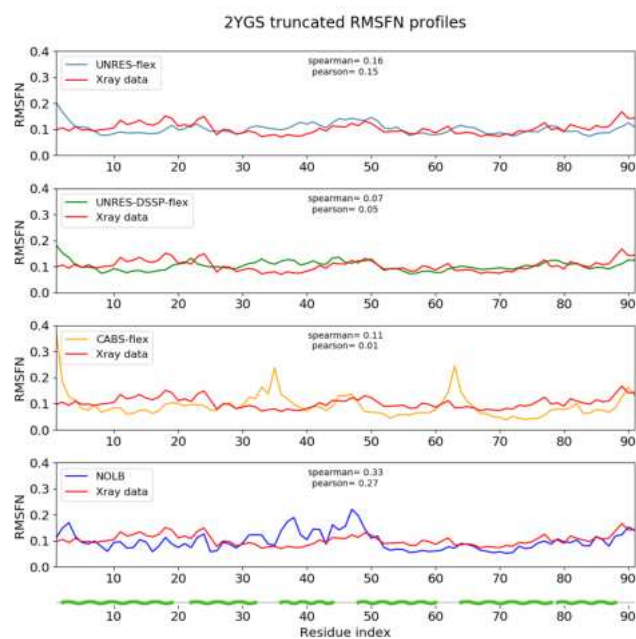


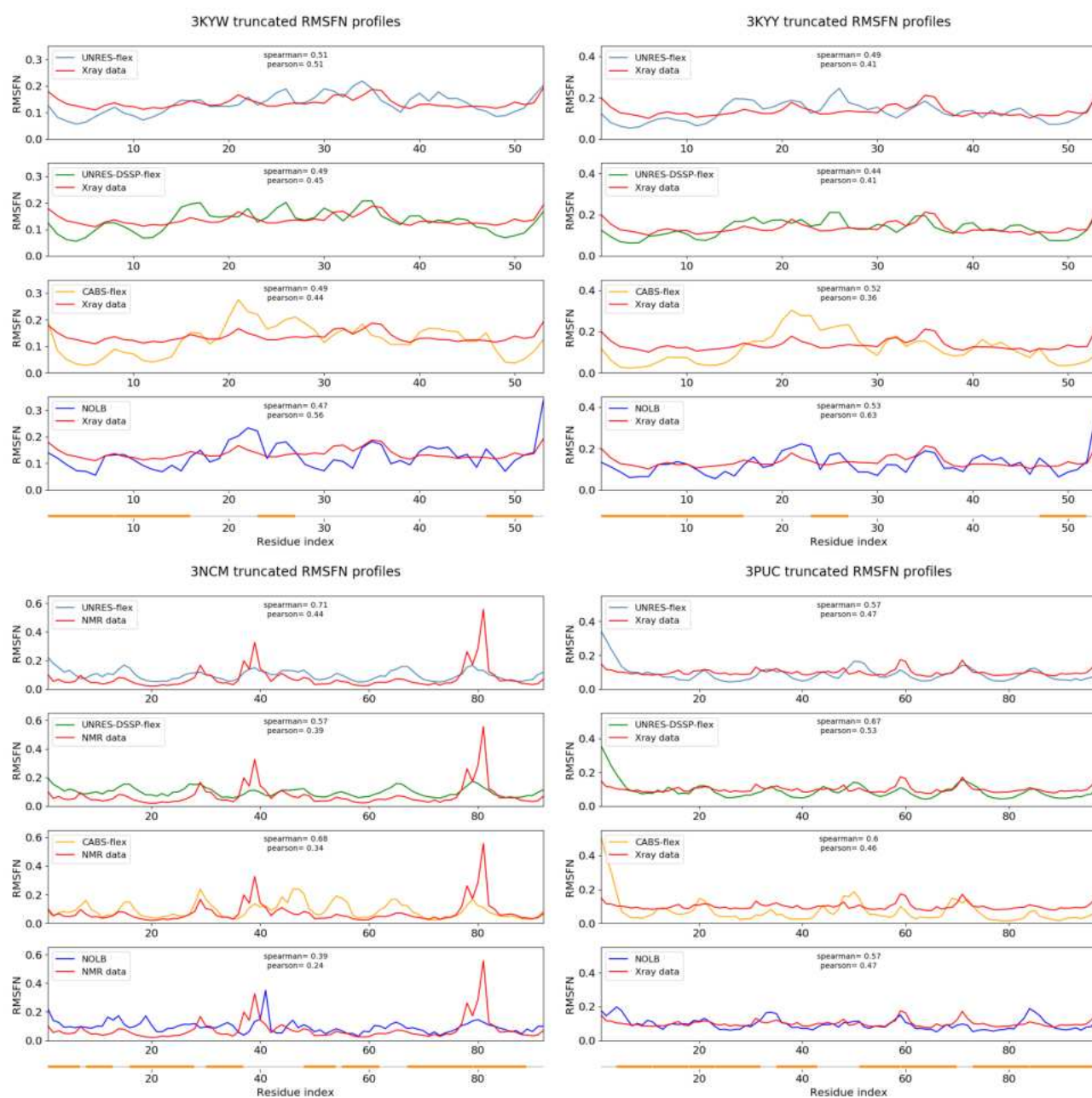


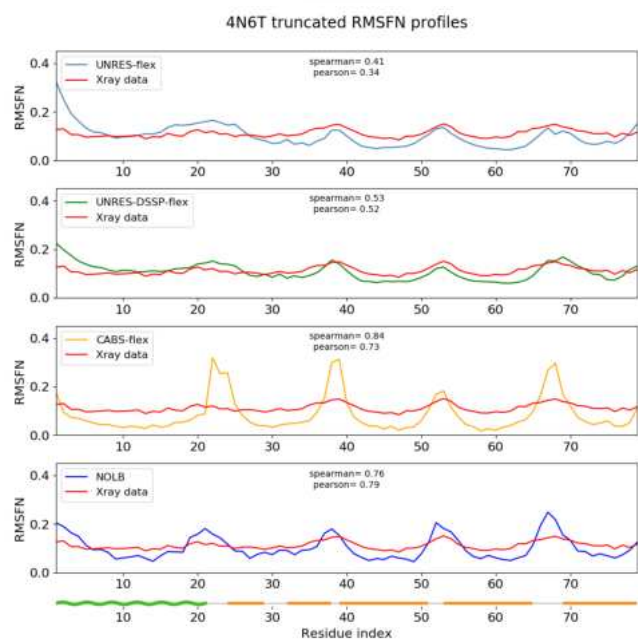
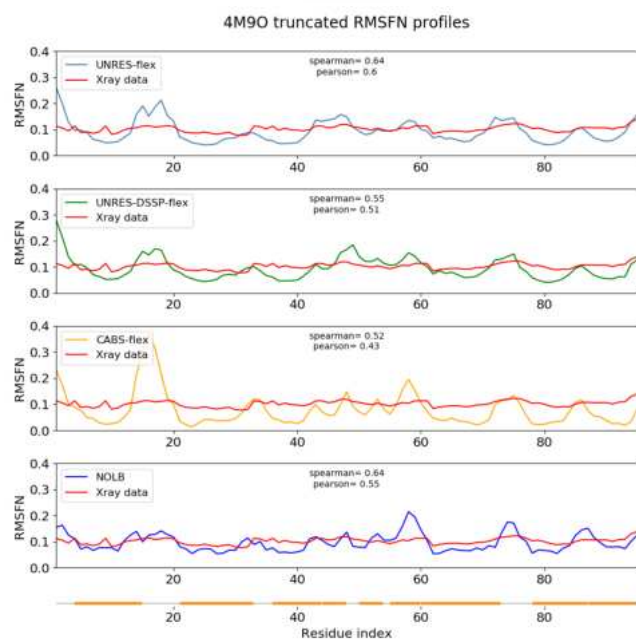
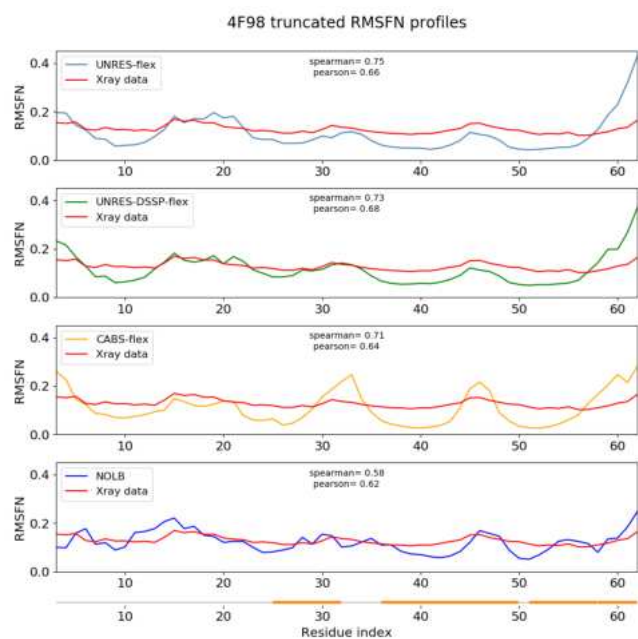
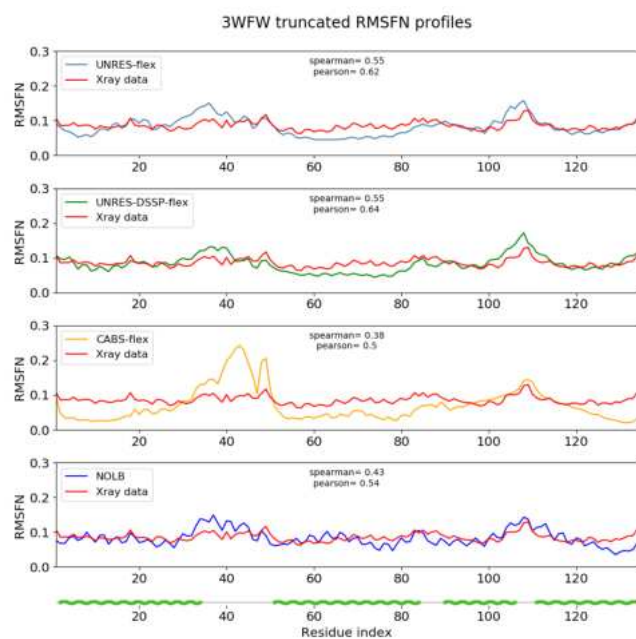


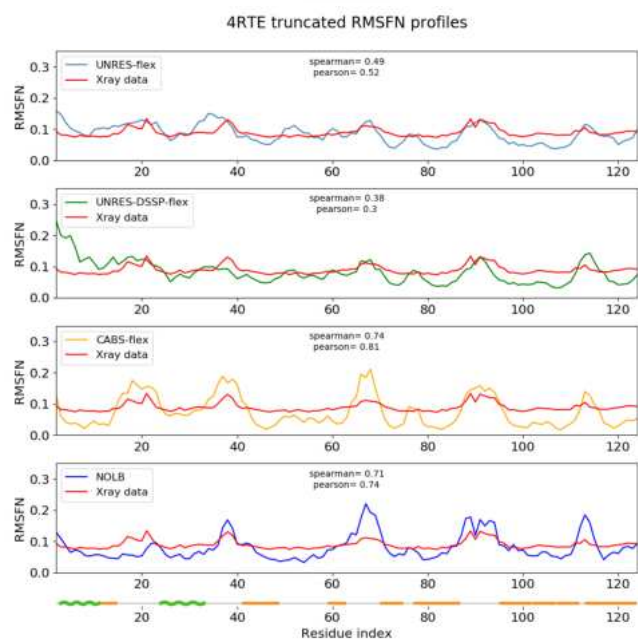
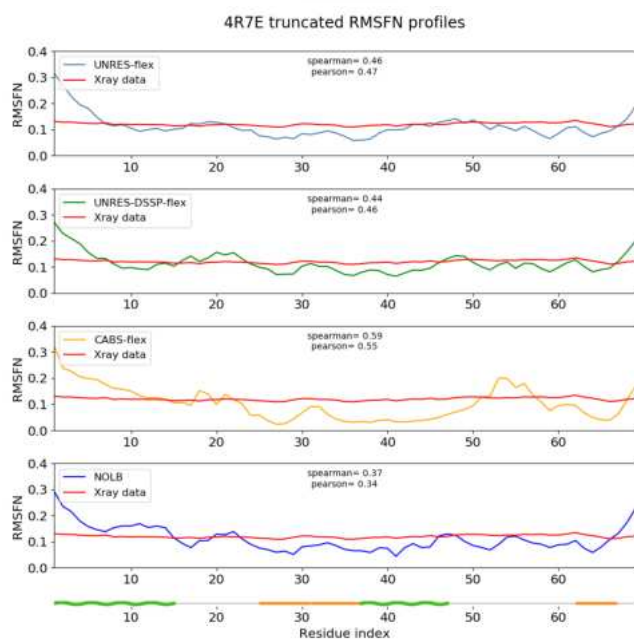
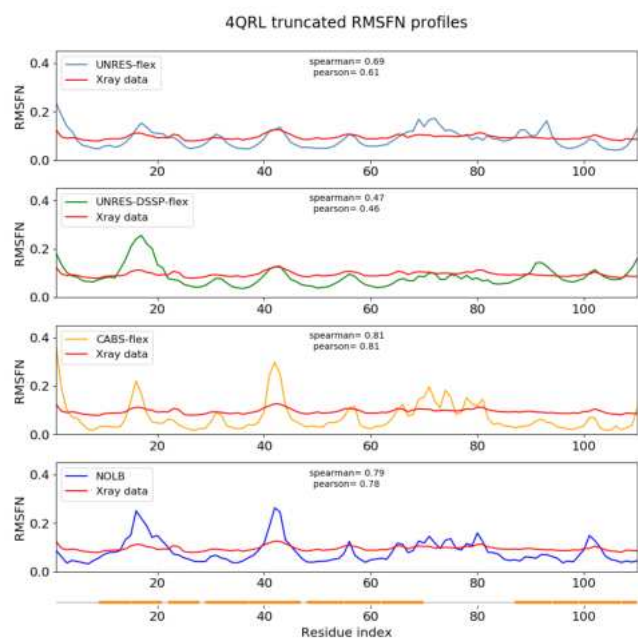
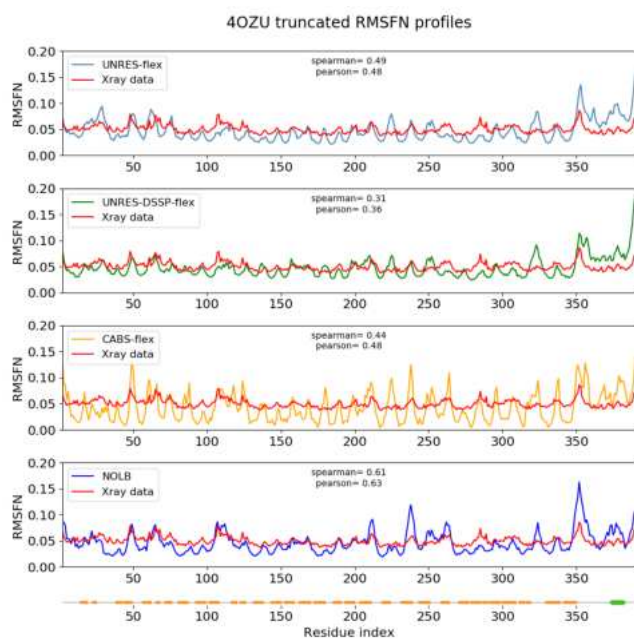


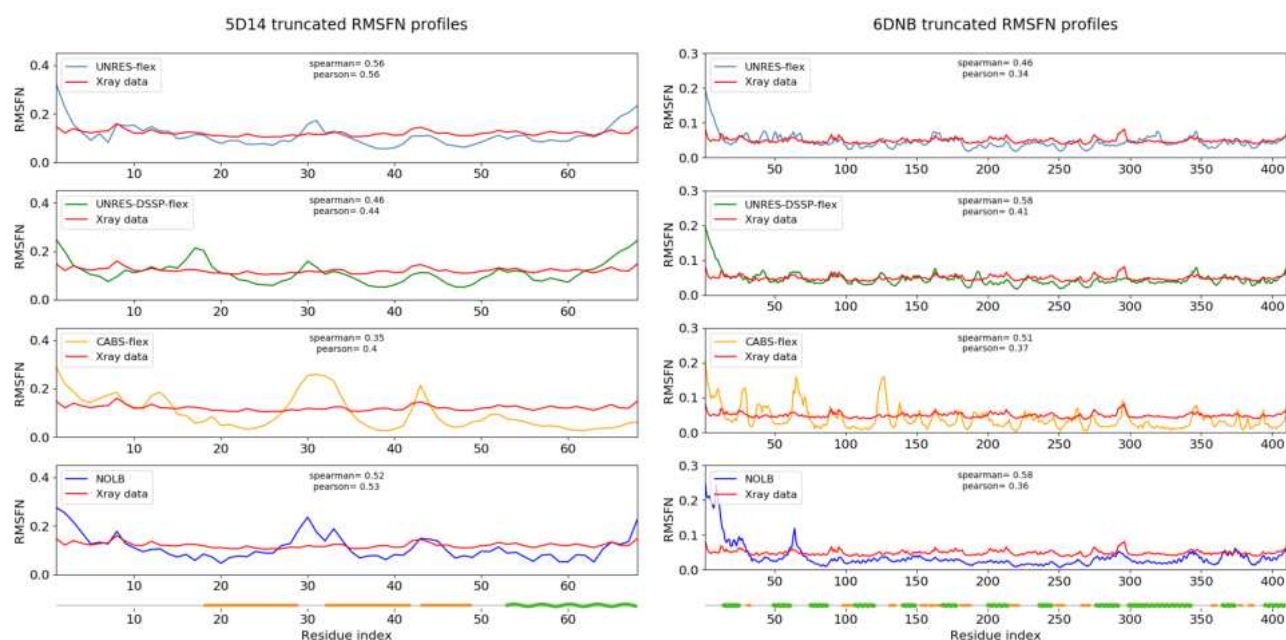




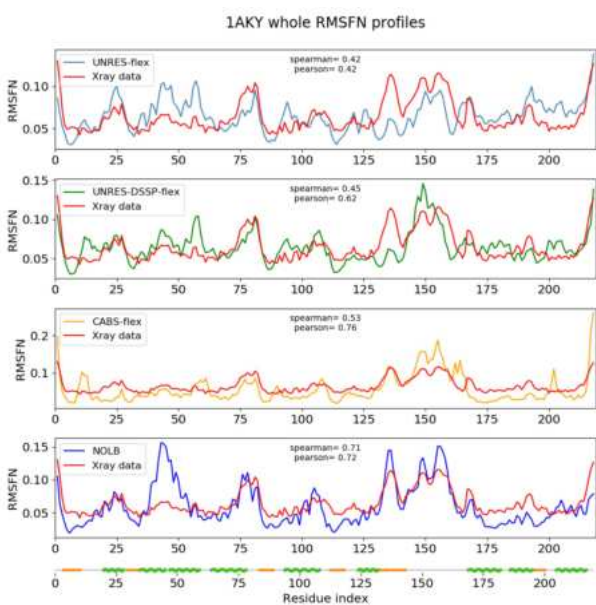
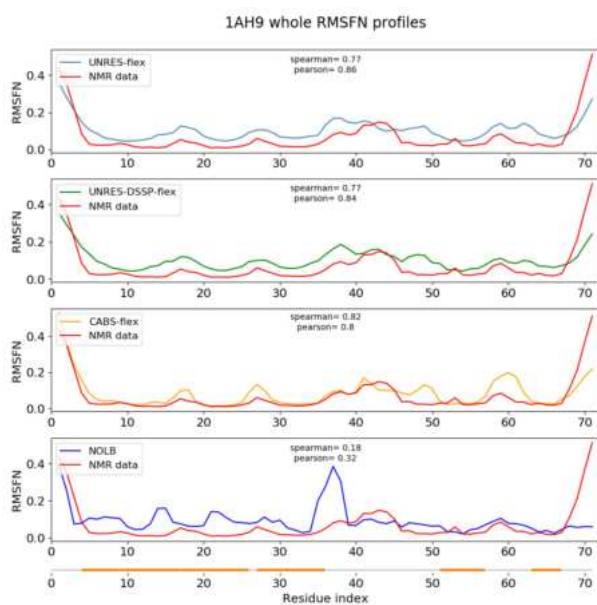
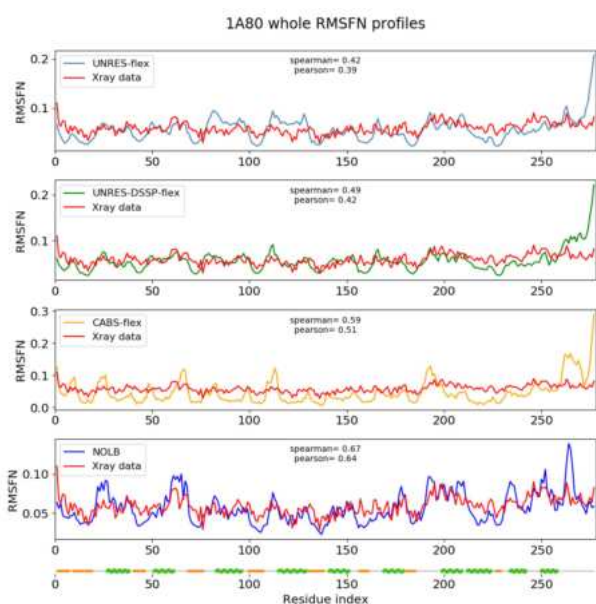
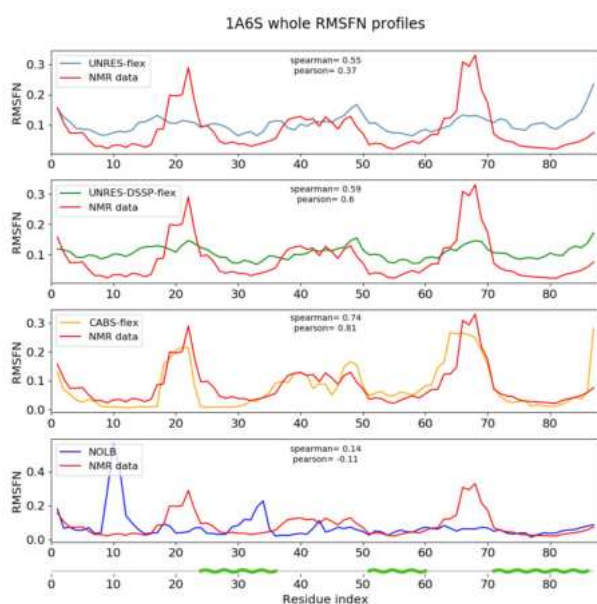


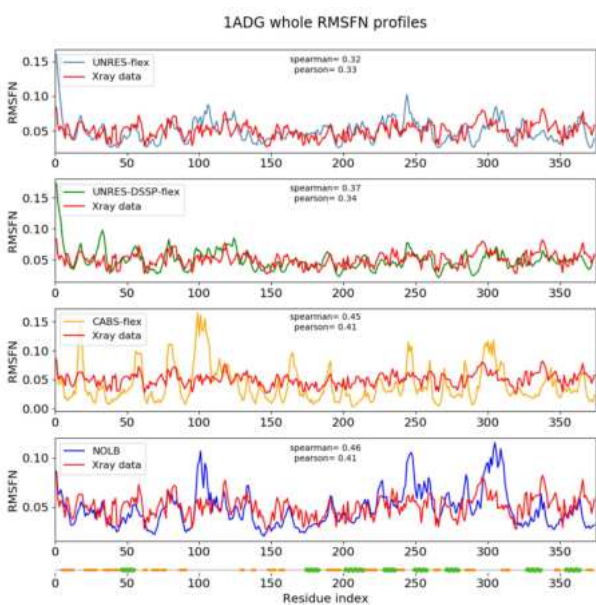
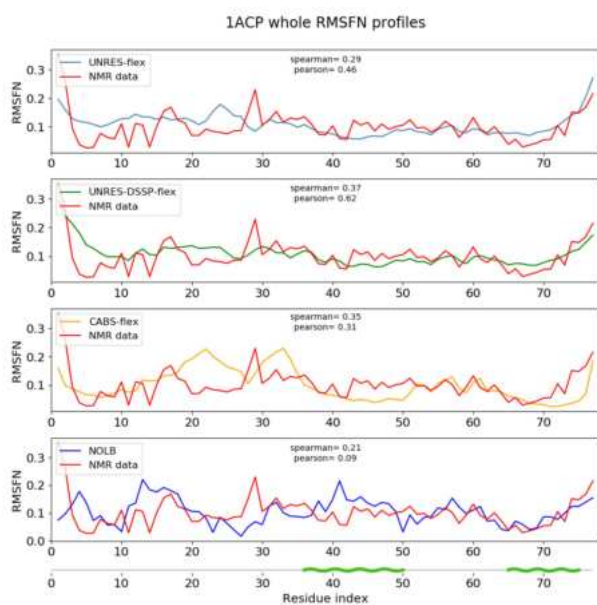
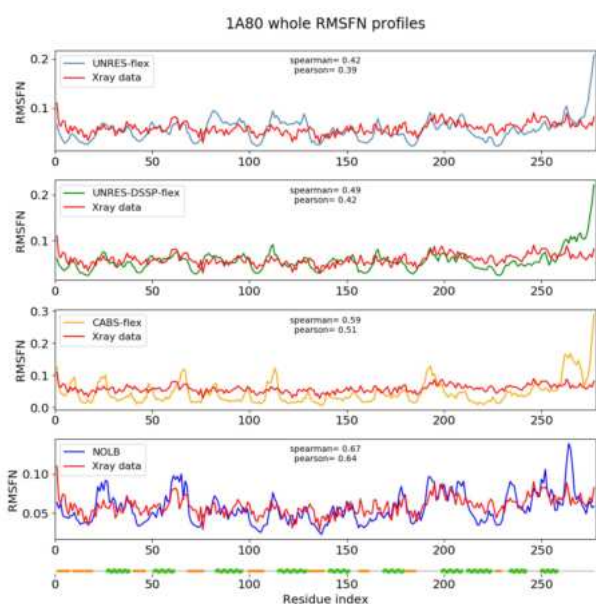
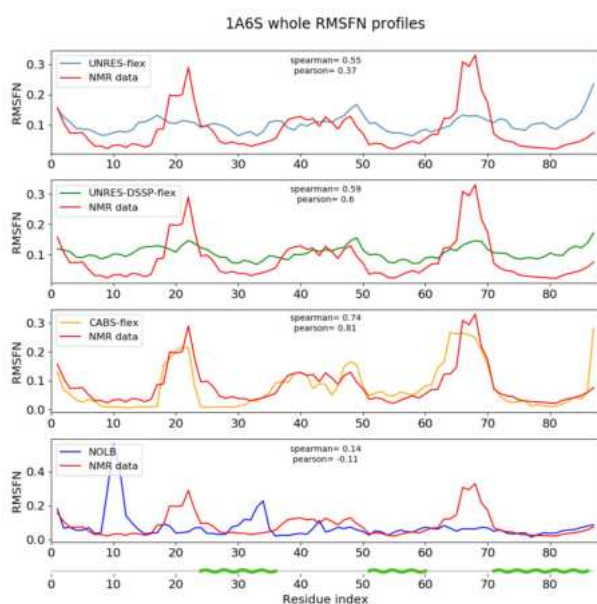


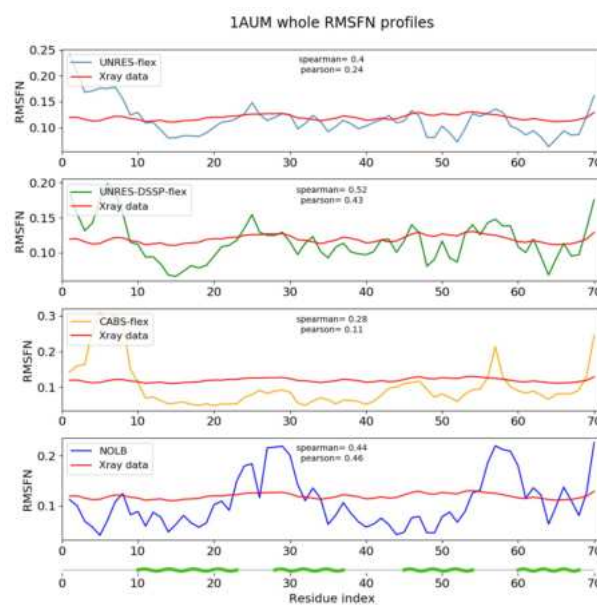
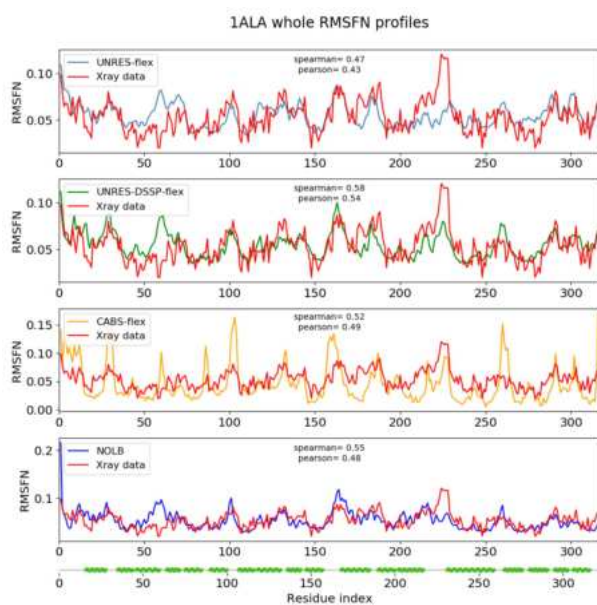
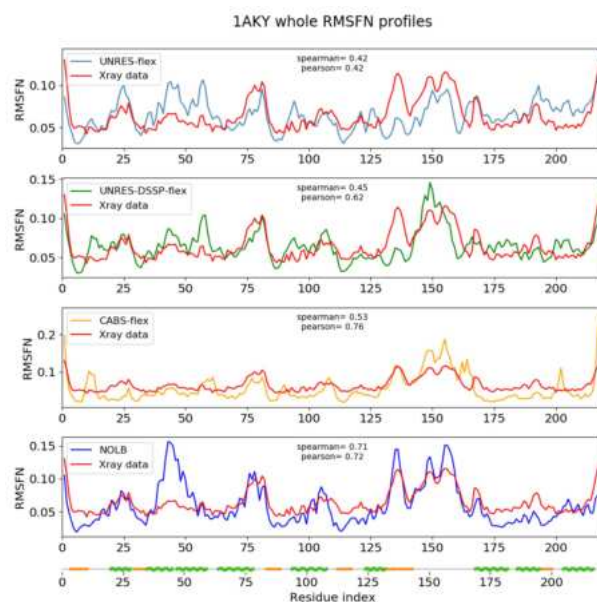
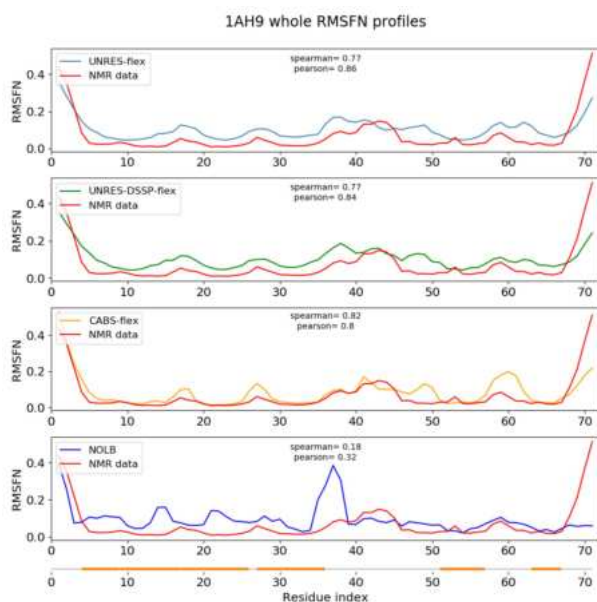


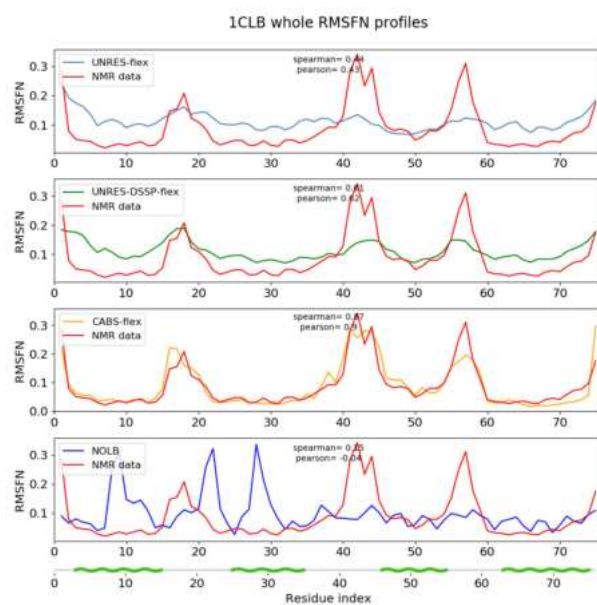
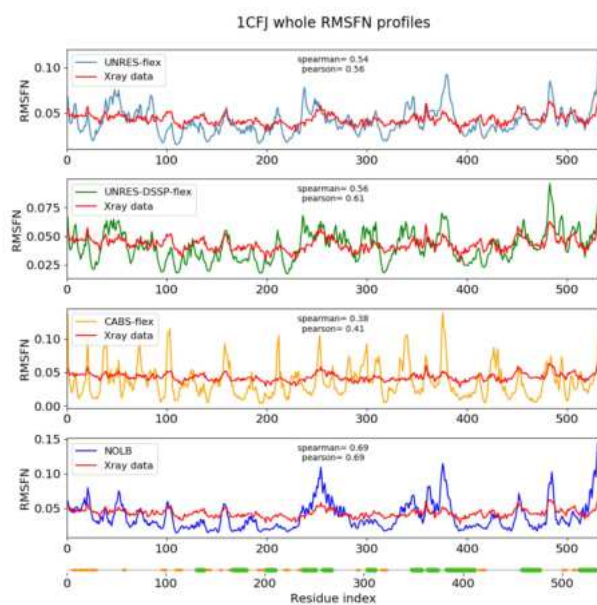
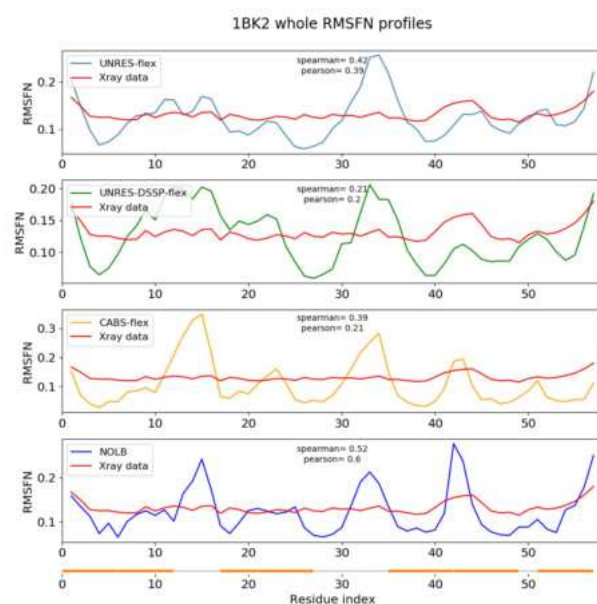
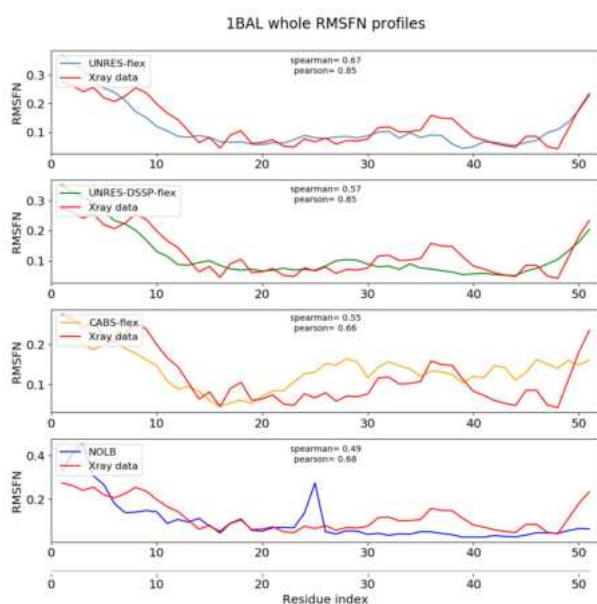


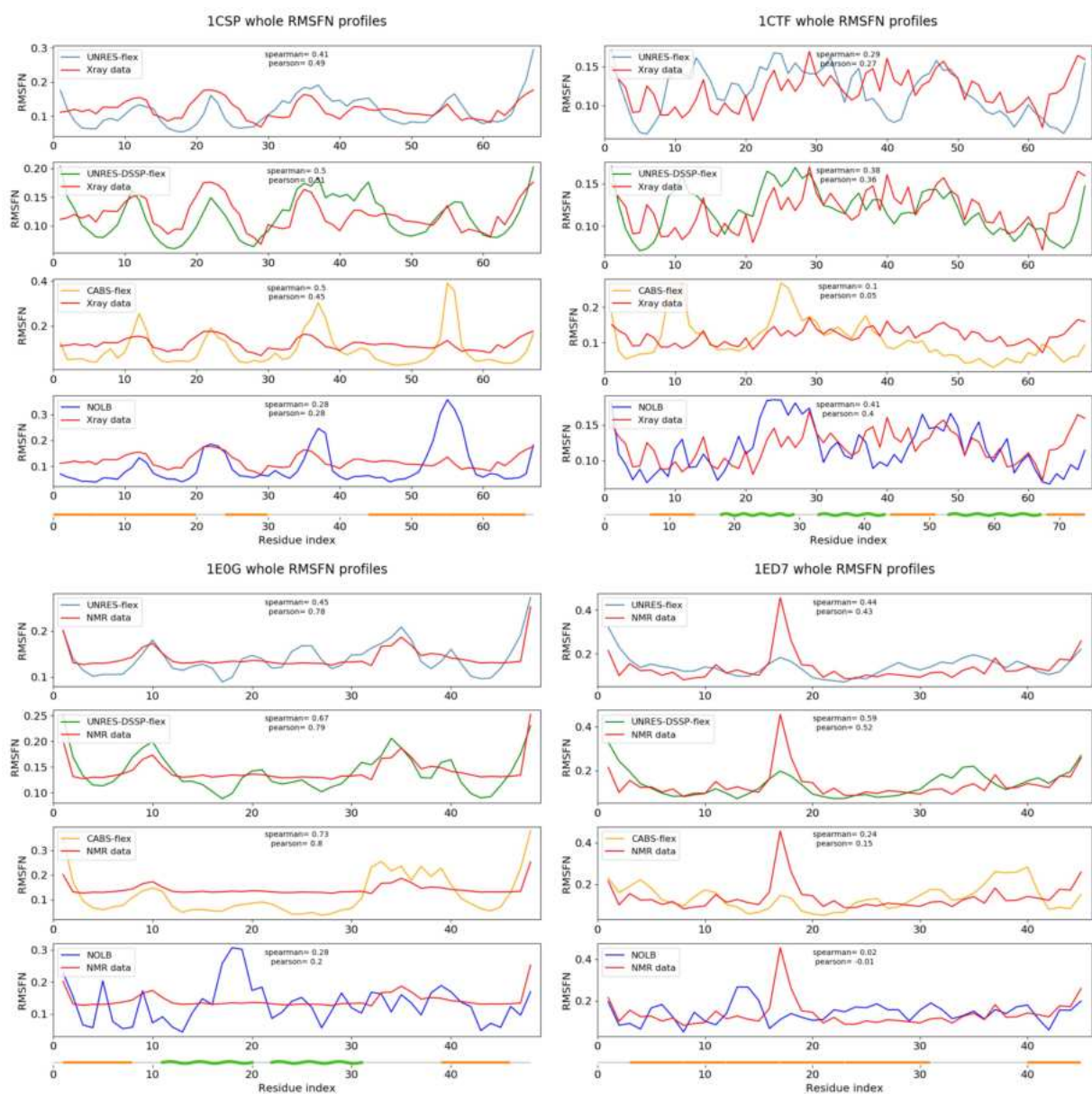
Rysunek 7.1: Profile fluktuacji po normalizacji (RMSFNN) obliczone na podstawie zespołów NMR lub współczynników B promieniowania rentgenowskiego (czerwona linia) i odpowiadające im profile fluktuacji przewidziane za pomocą metod: UNRES-flex (jasnoniebieska linia), UNRES-DSSP-flex (zielona linia), CABS-flex (linia żółta) oraz NOLB (linia niebieska). Każdy profil fluktuacji został uśredniony na podstawie trzech symulacji, z wyjątkiem NOLB. Dla każdego profilu fluktuacji został określony współczynniki korelacji Pearson (r_p) i Spearman (r_s), które określają zależności pomiędzy profilami fluktuacji przewidywanych (teoretycznymi) a eksperymentalnymi. Ciągłe pomarańczowe i falowane zielone linie w dolnej części informują o występujących strukturach drugorzędowych: α -helisy lub/i β -harmonijki. Wykresy dotyczą obciętych struktur białek. "Reprinted with permission (CC-BY 4.0.) from Ł. J. Dziadek, A. K. Sieradzan, C. Czaplewski, M. Zalewski, F. Banaś, M. Toczek, W. Nisterenko, S. Grudin, A. Liwo, A. Gieldoń, Assessment of Four Theoretical Approaches to Predict Protein Flexibility in the Crystal Phase and Solution, J. Chem. Theory Comput. 2024, 20, 17, 7667–7681 Copyright 2024 Journal of Chemical Theory and Computation."

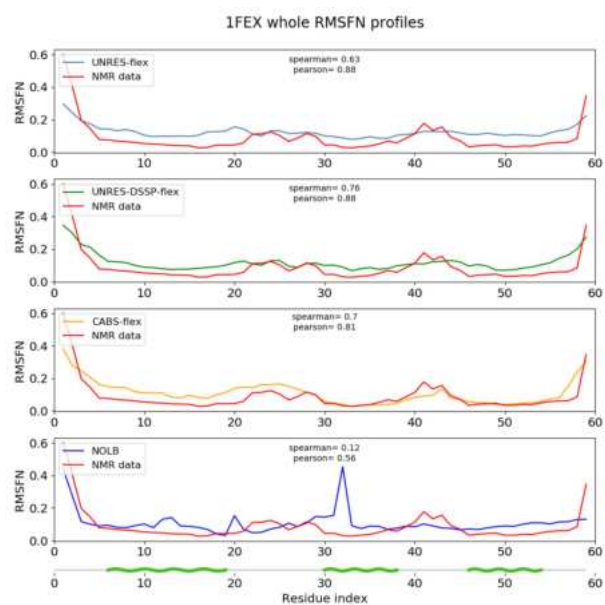
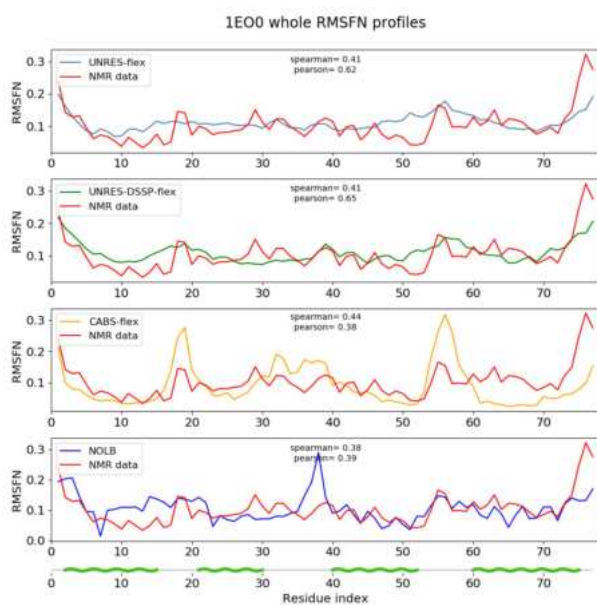
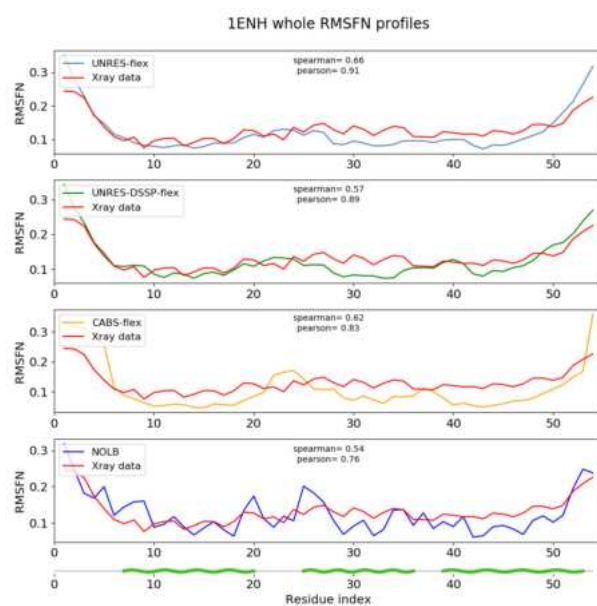
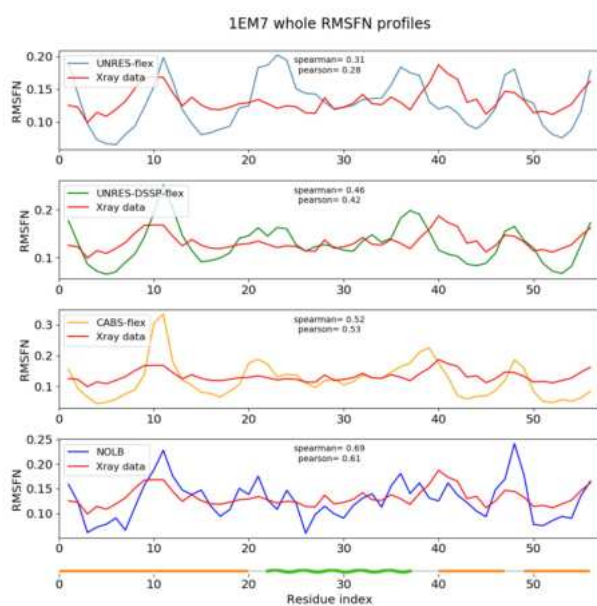


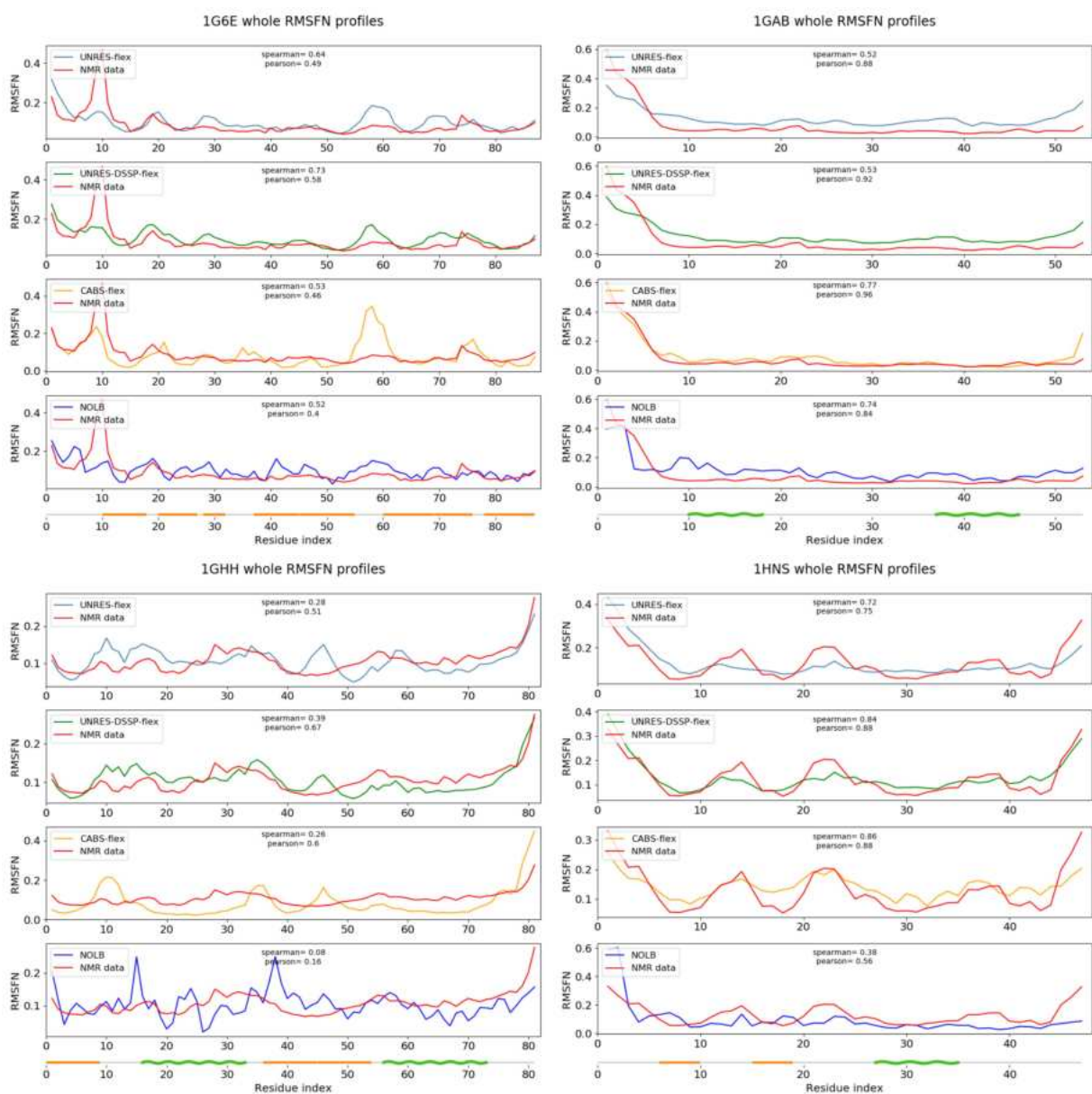


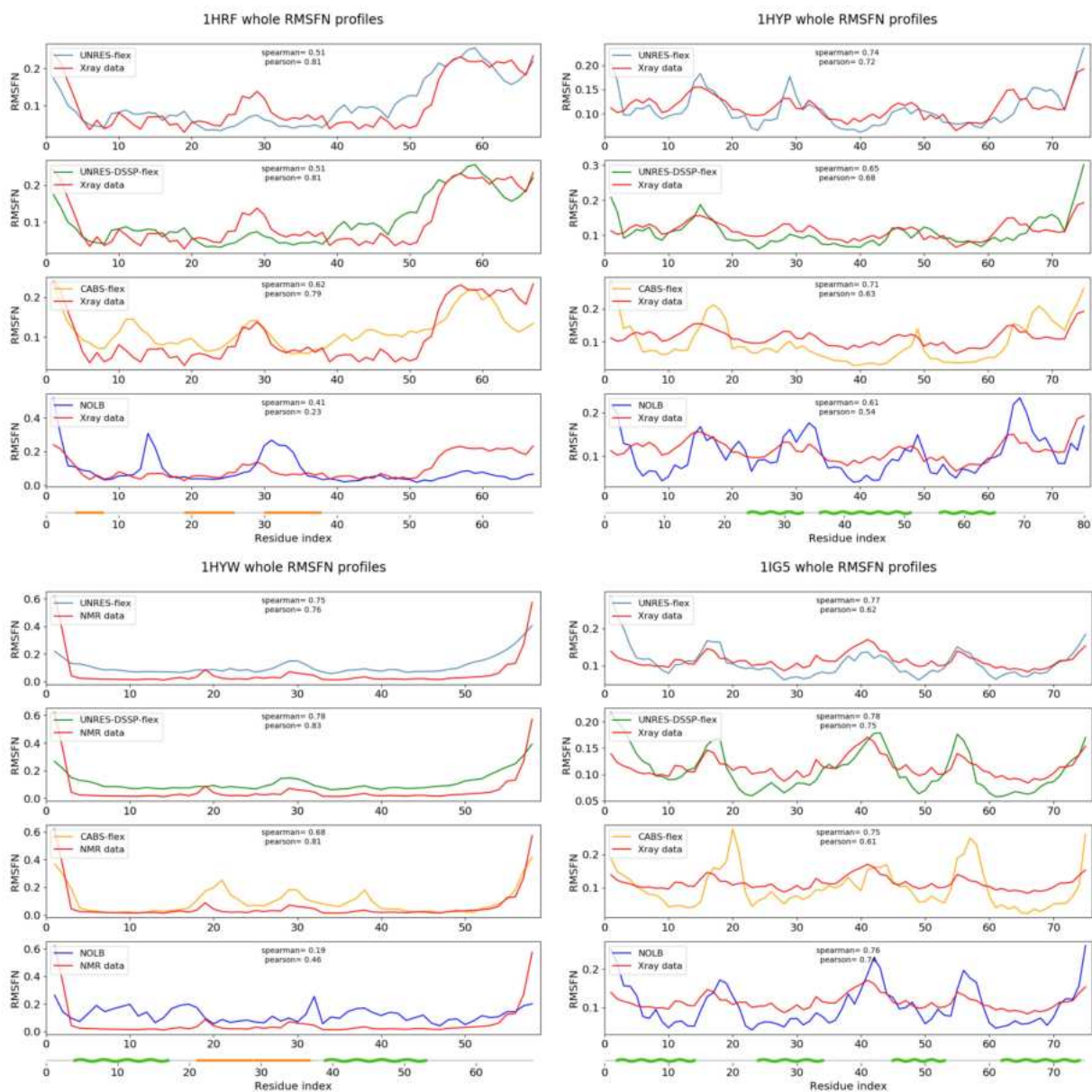


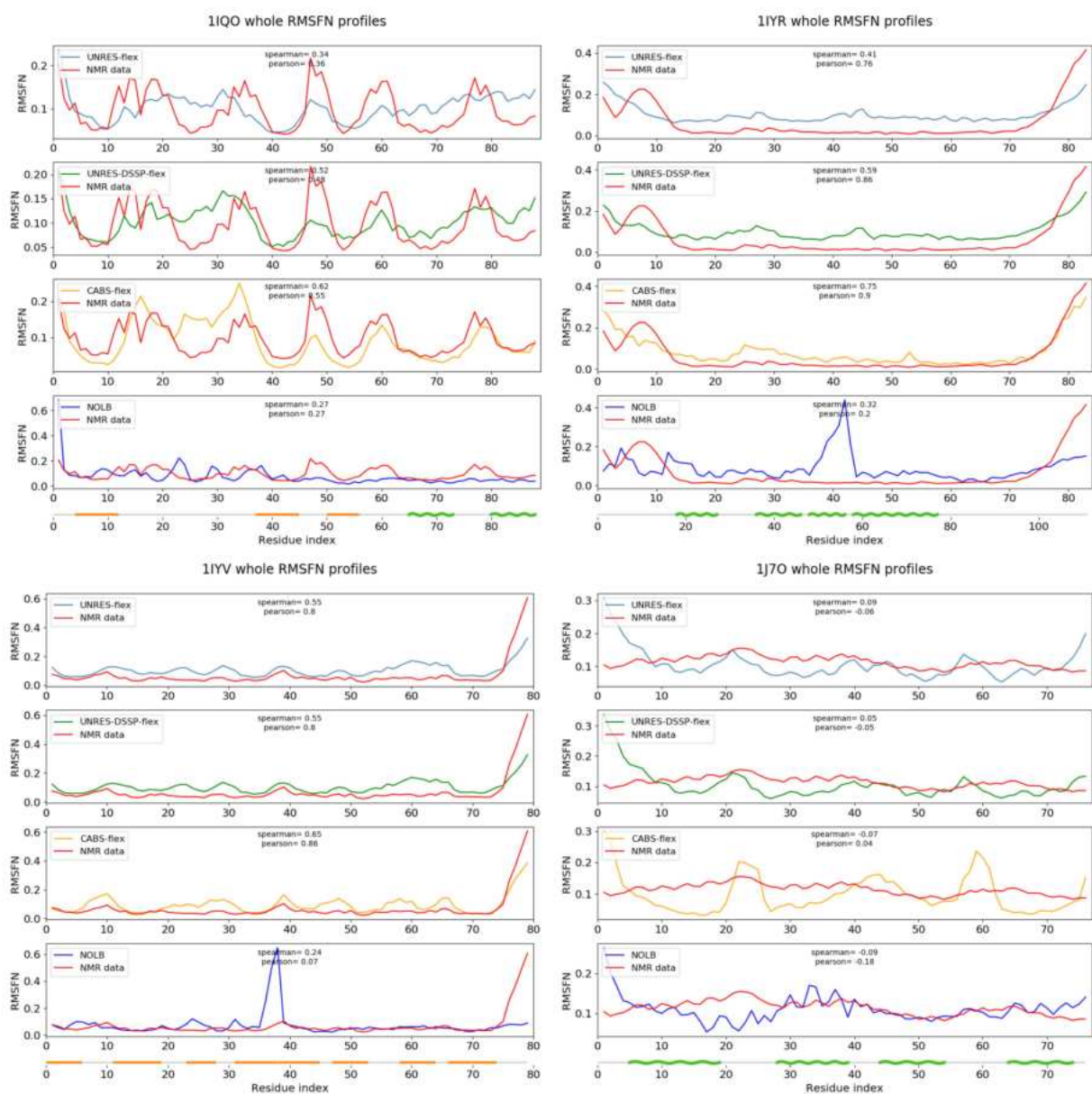


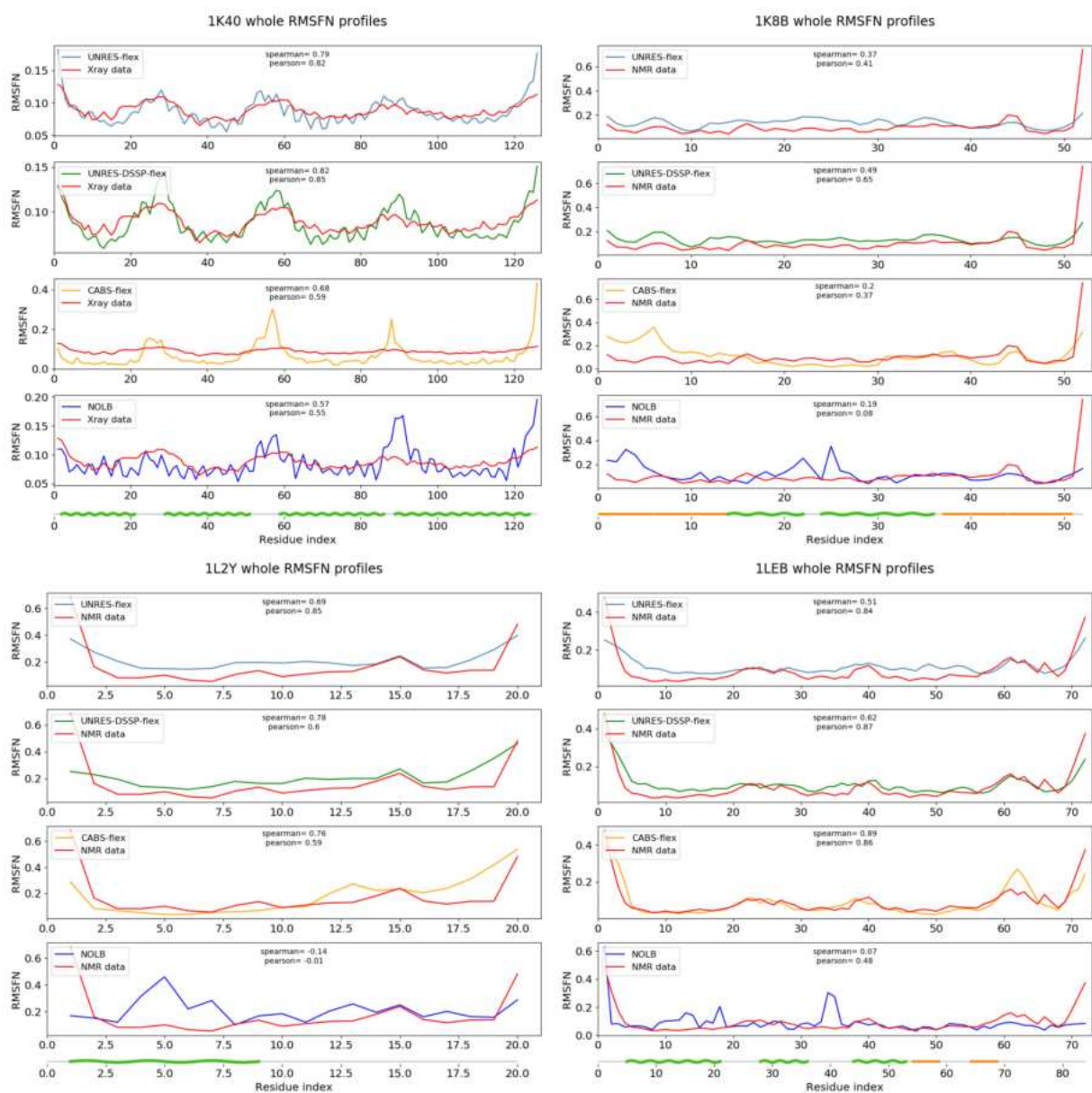


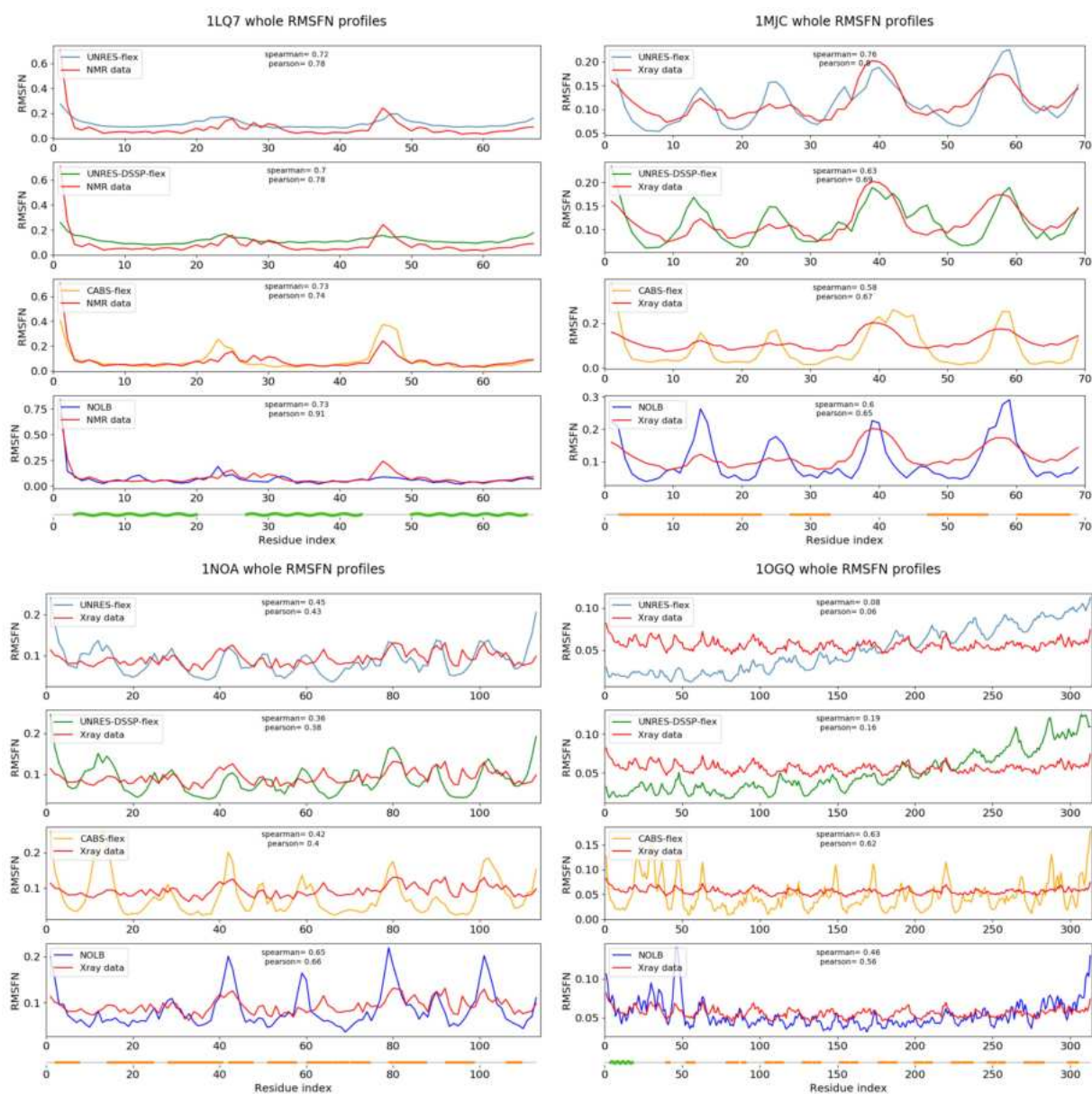


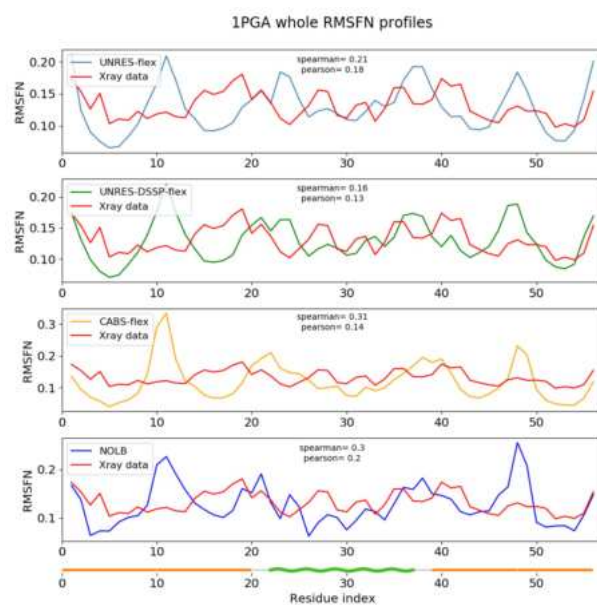
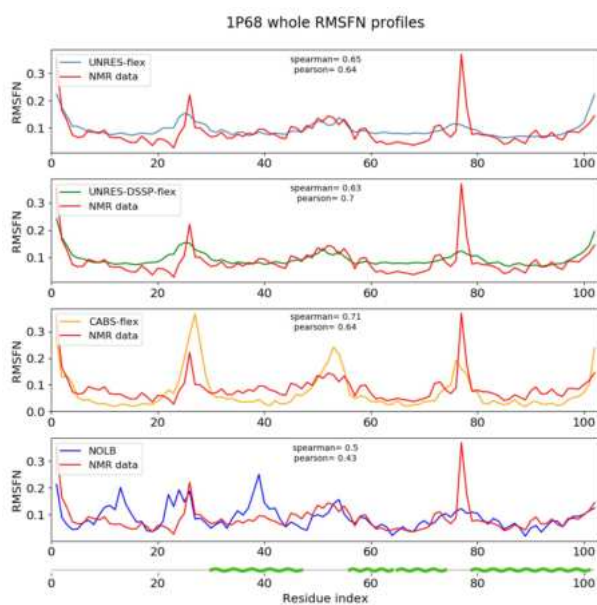
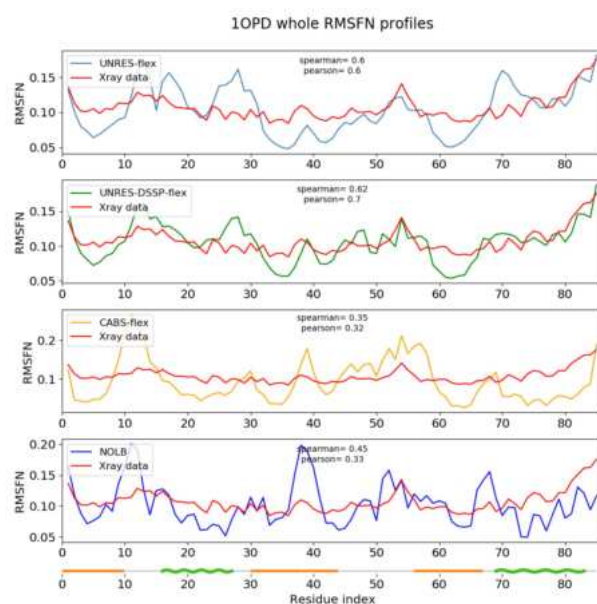
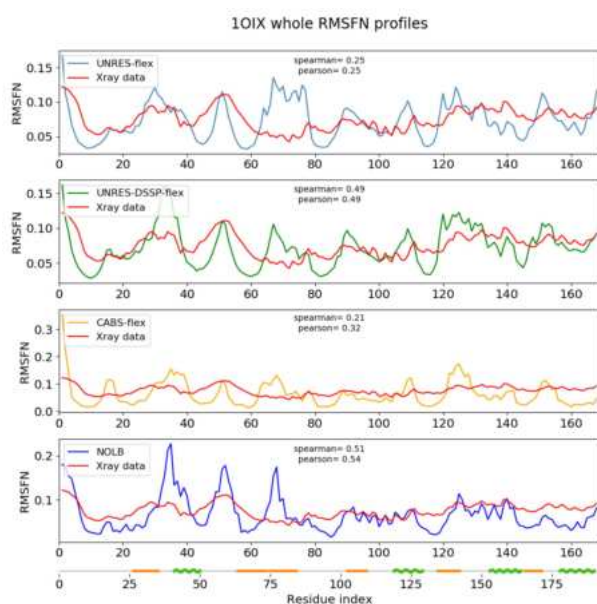


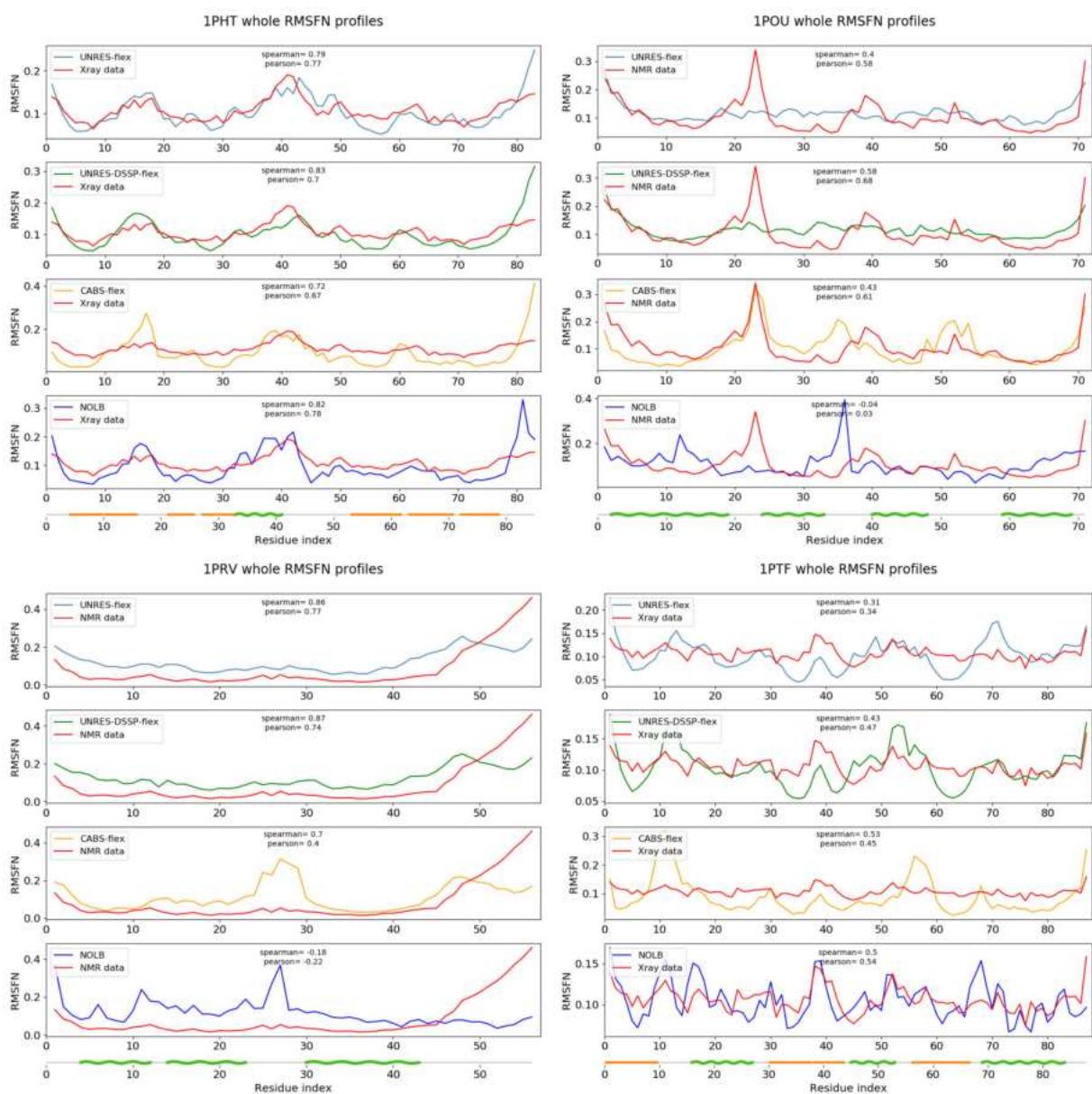


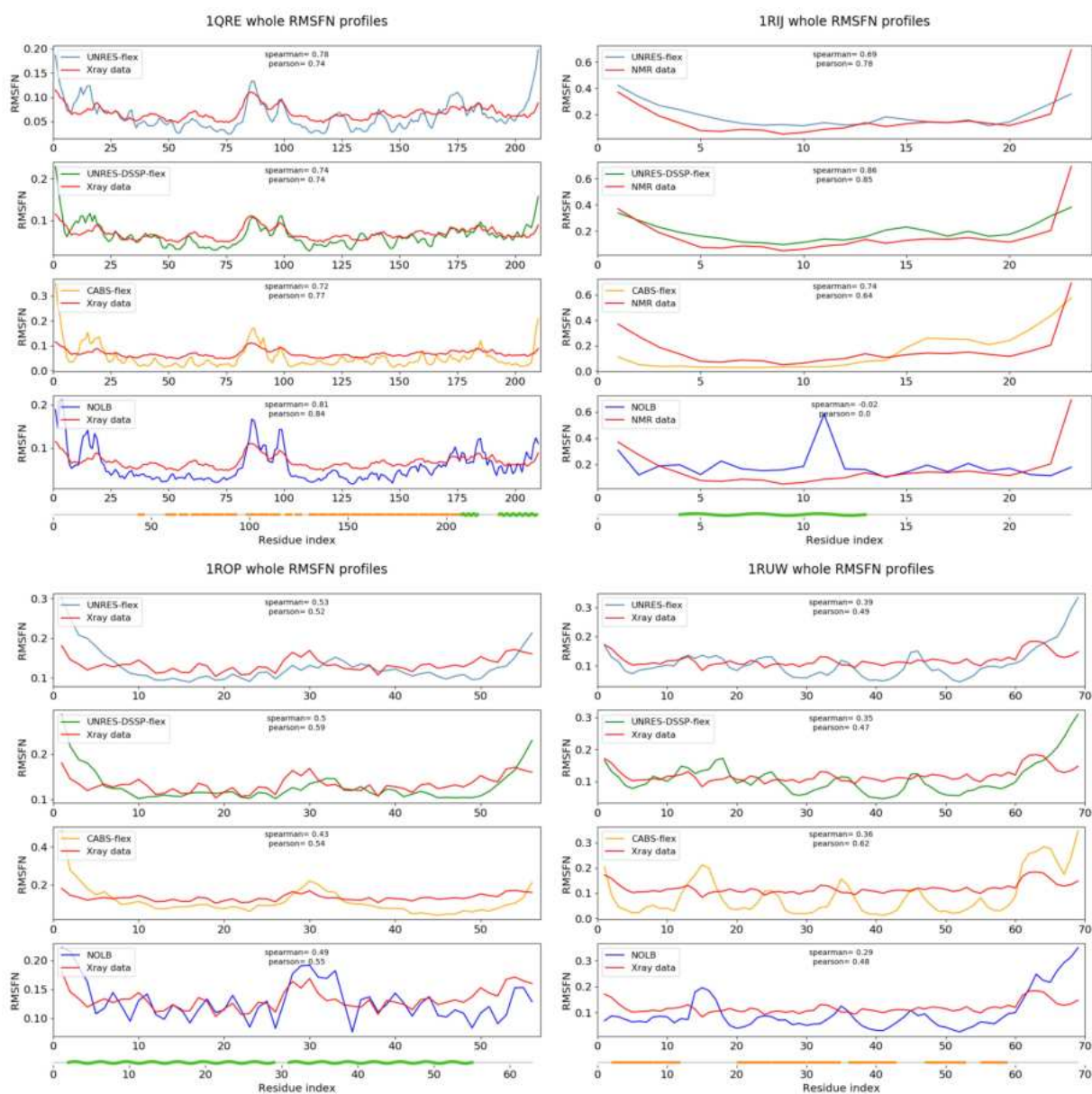


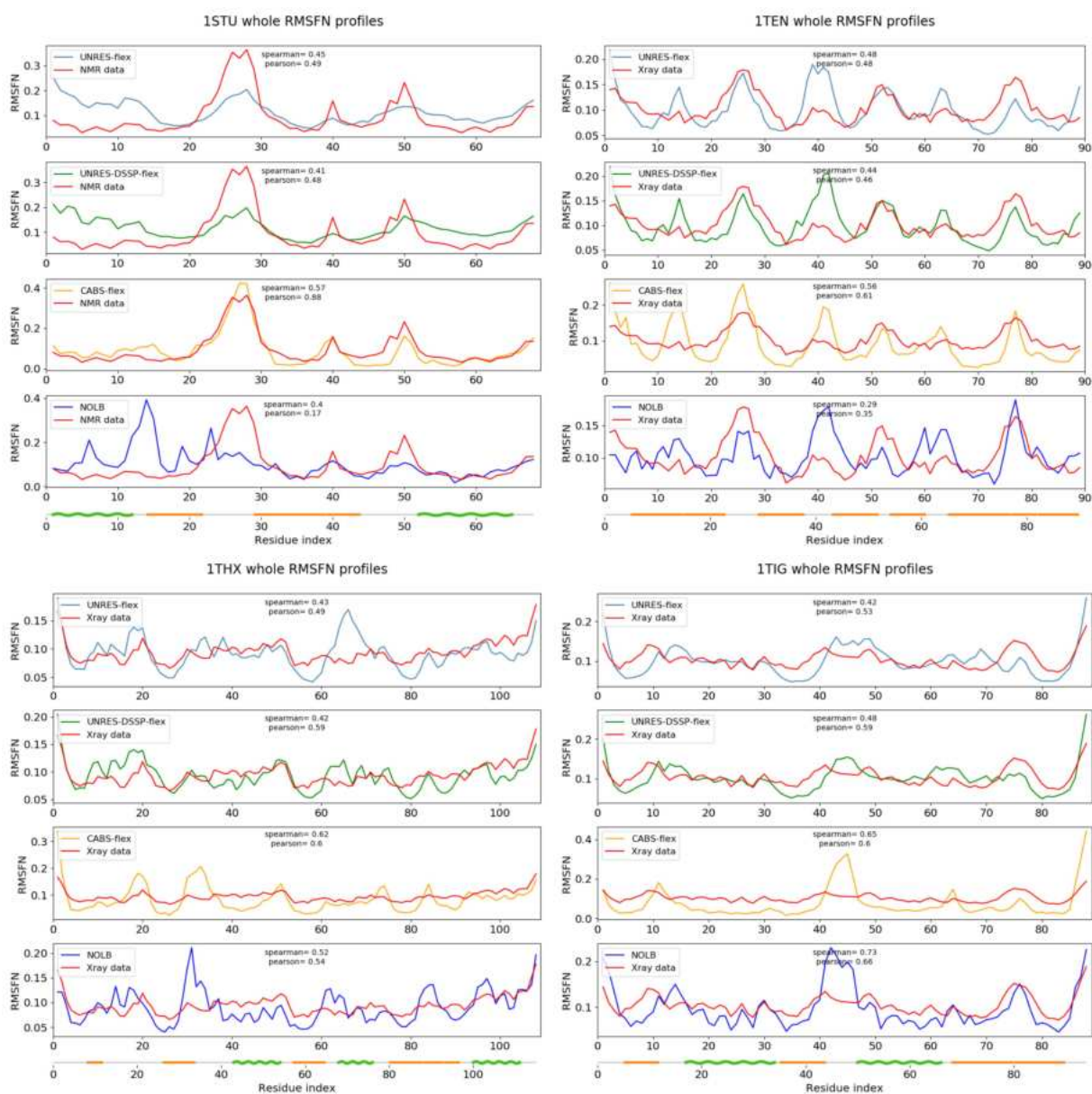


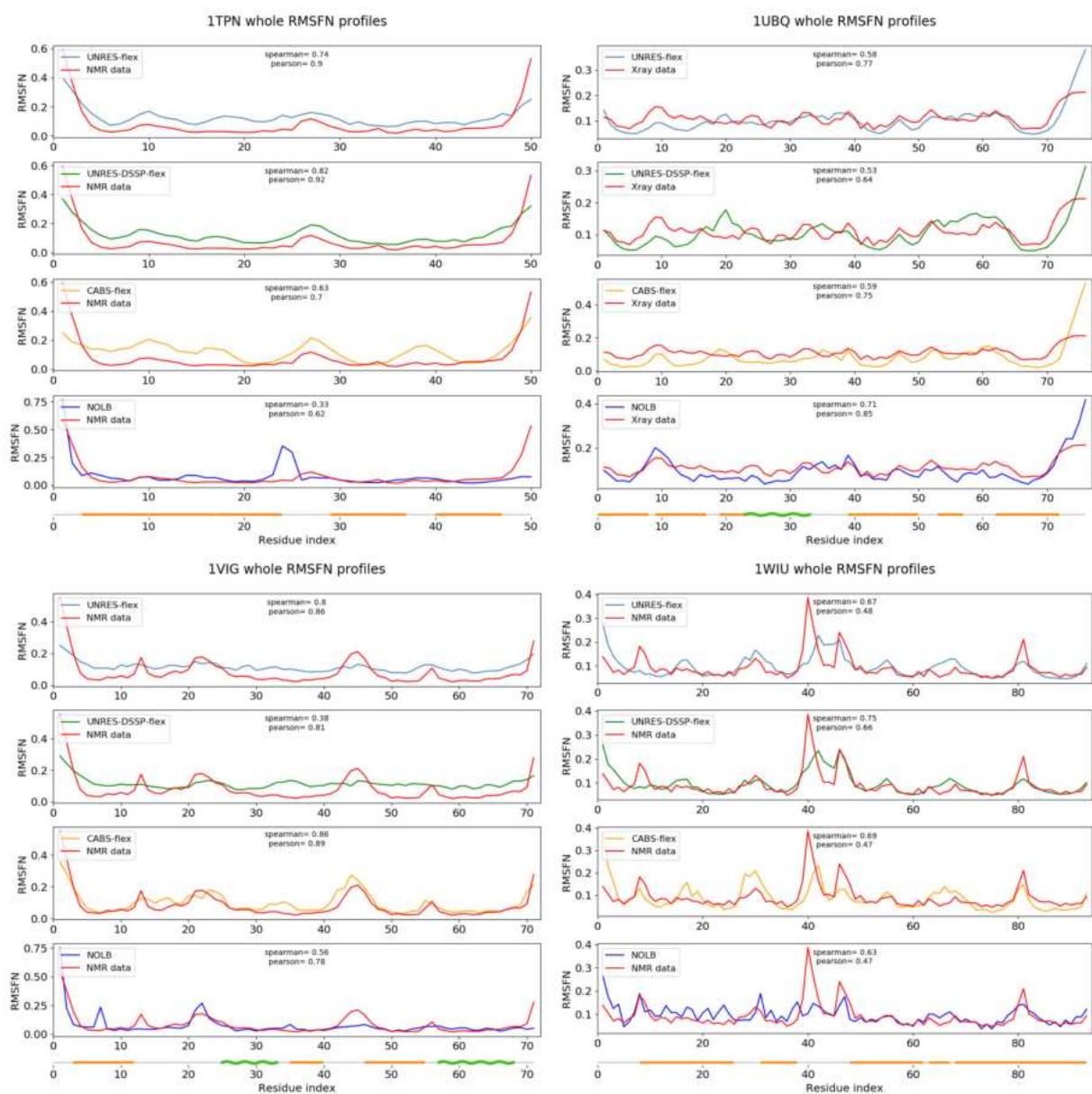


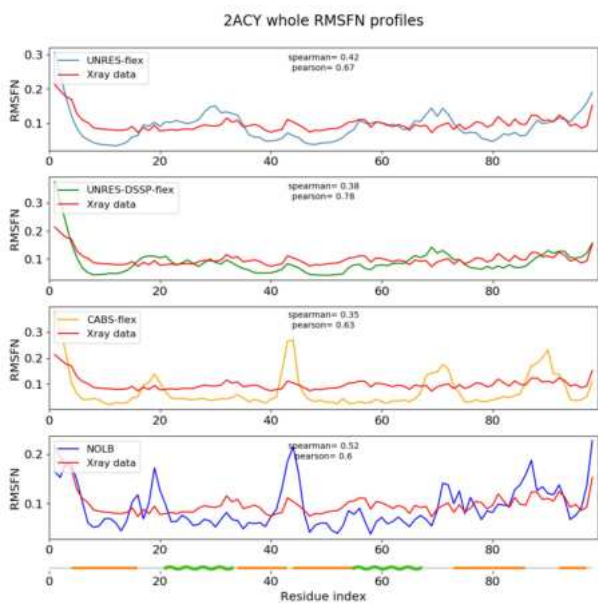
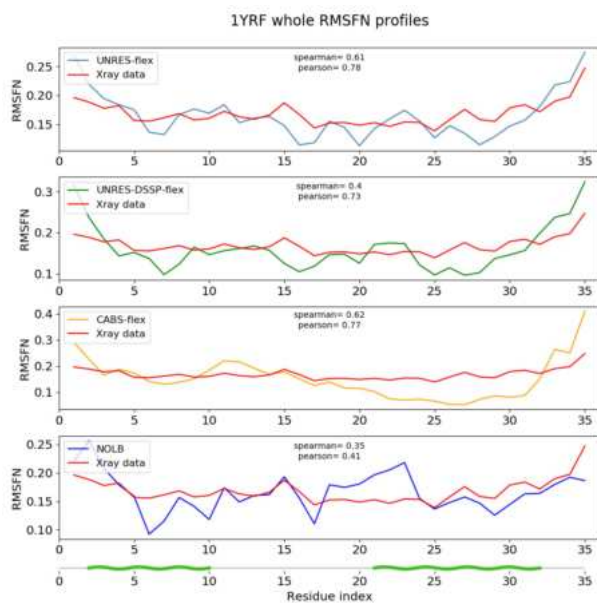
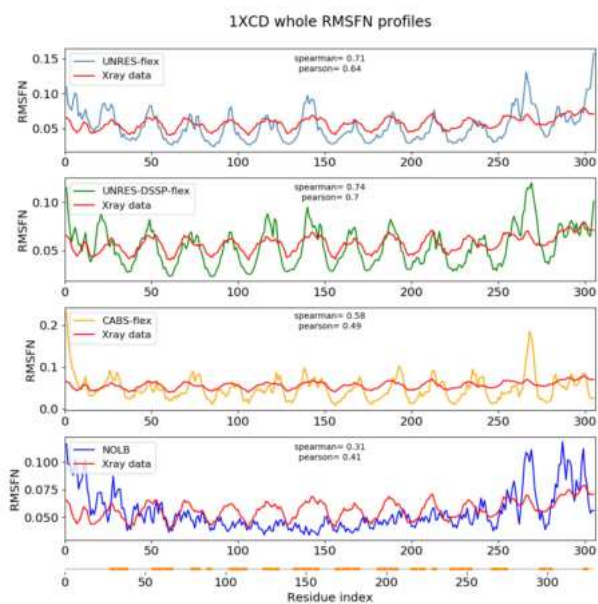
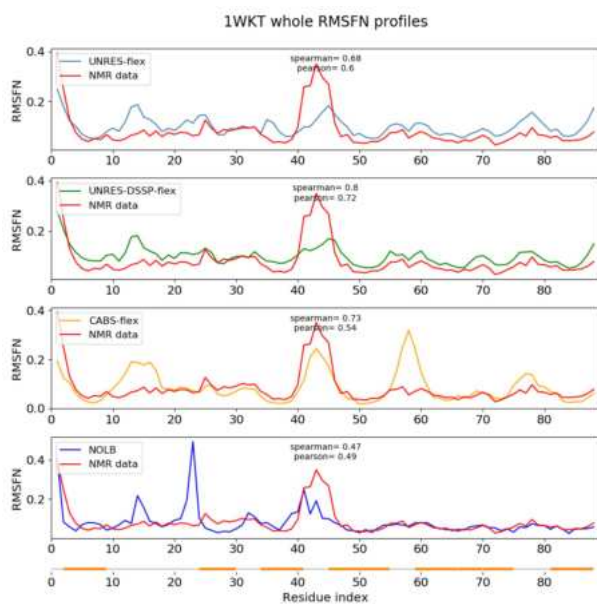


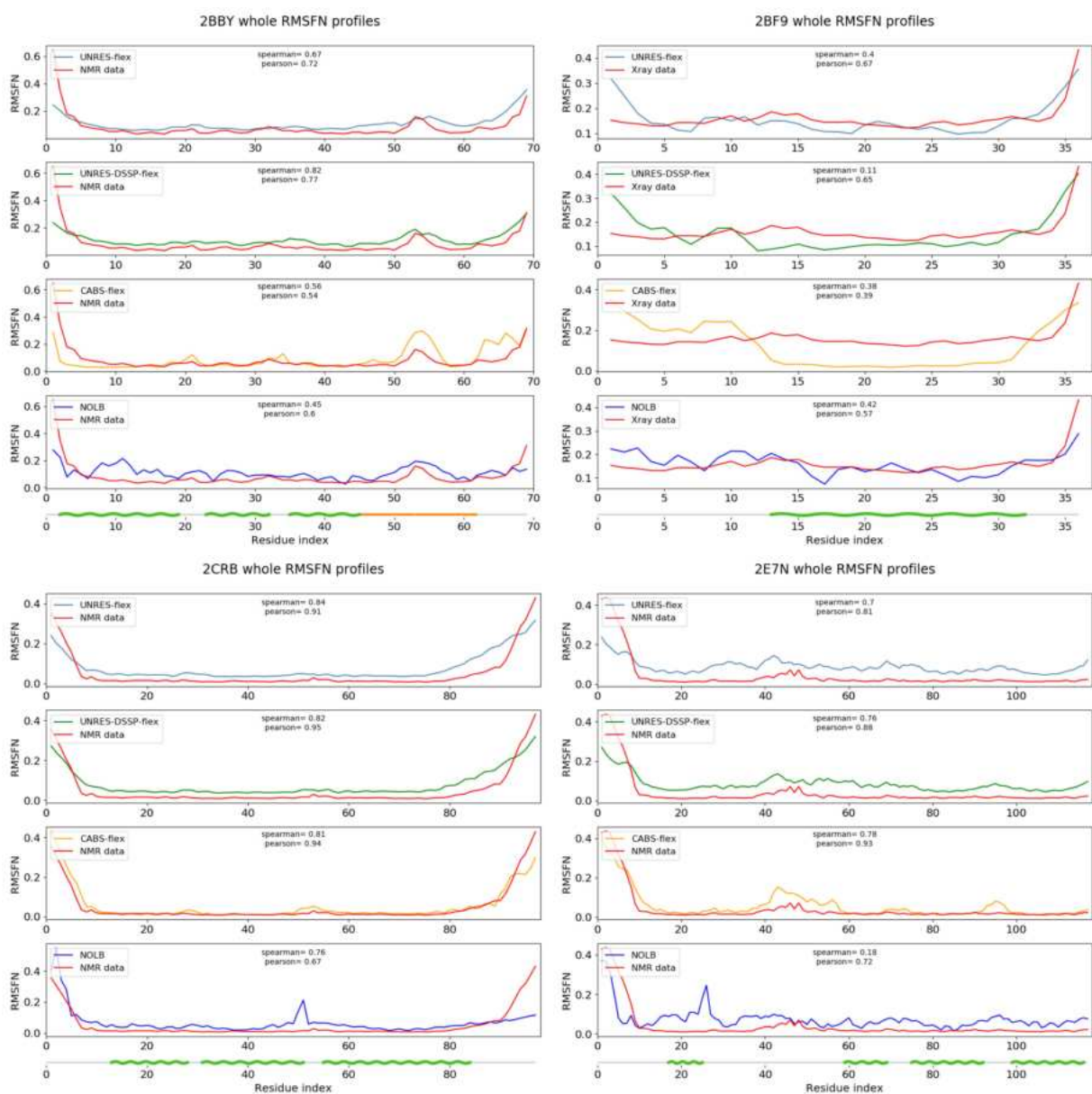


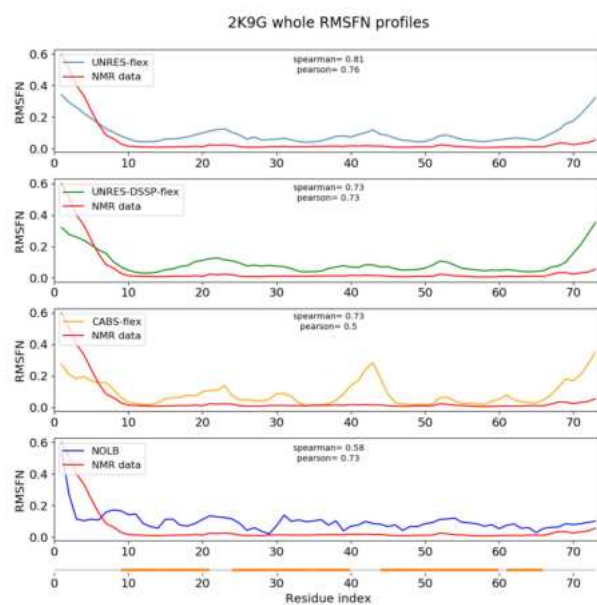
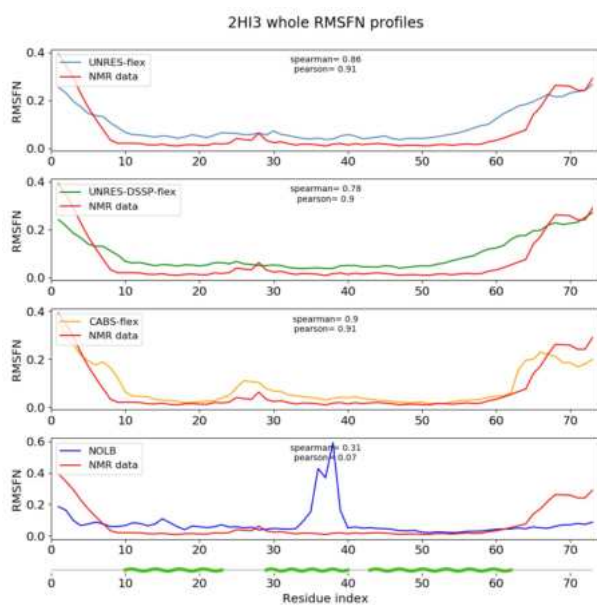
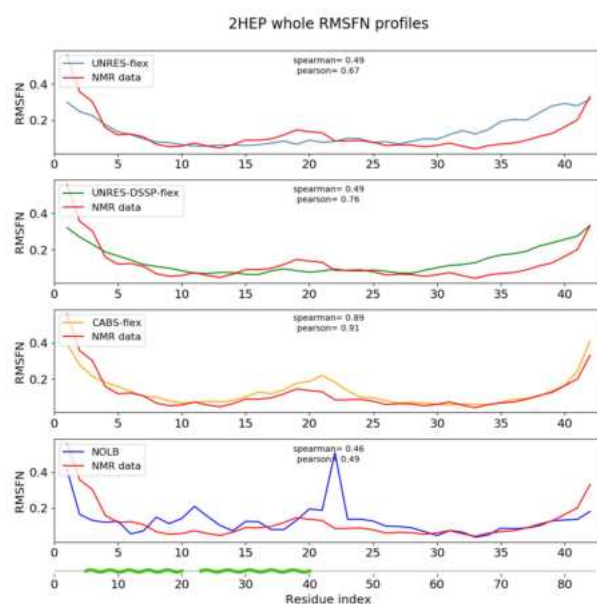
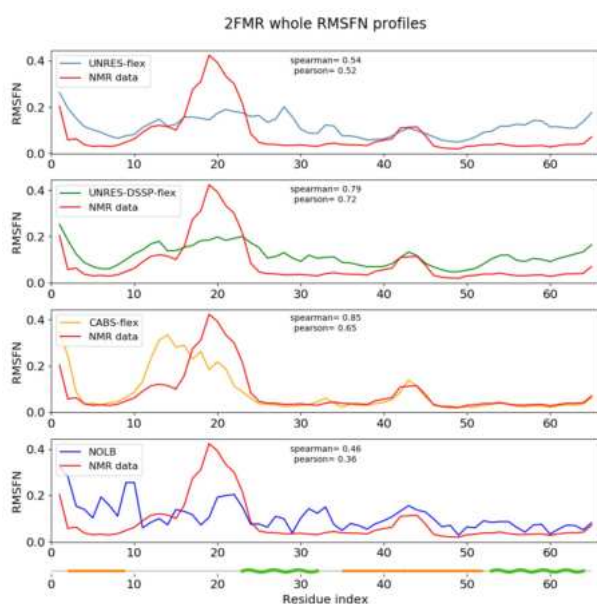


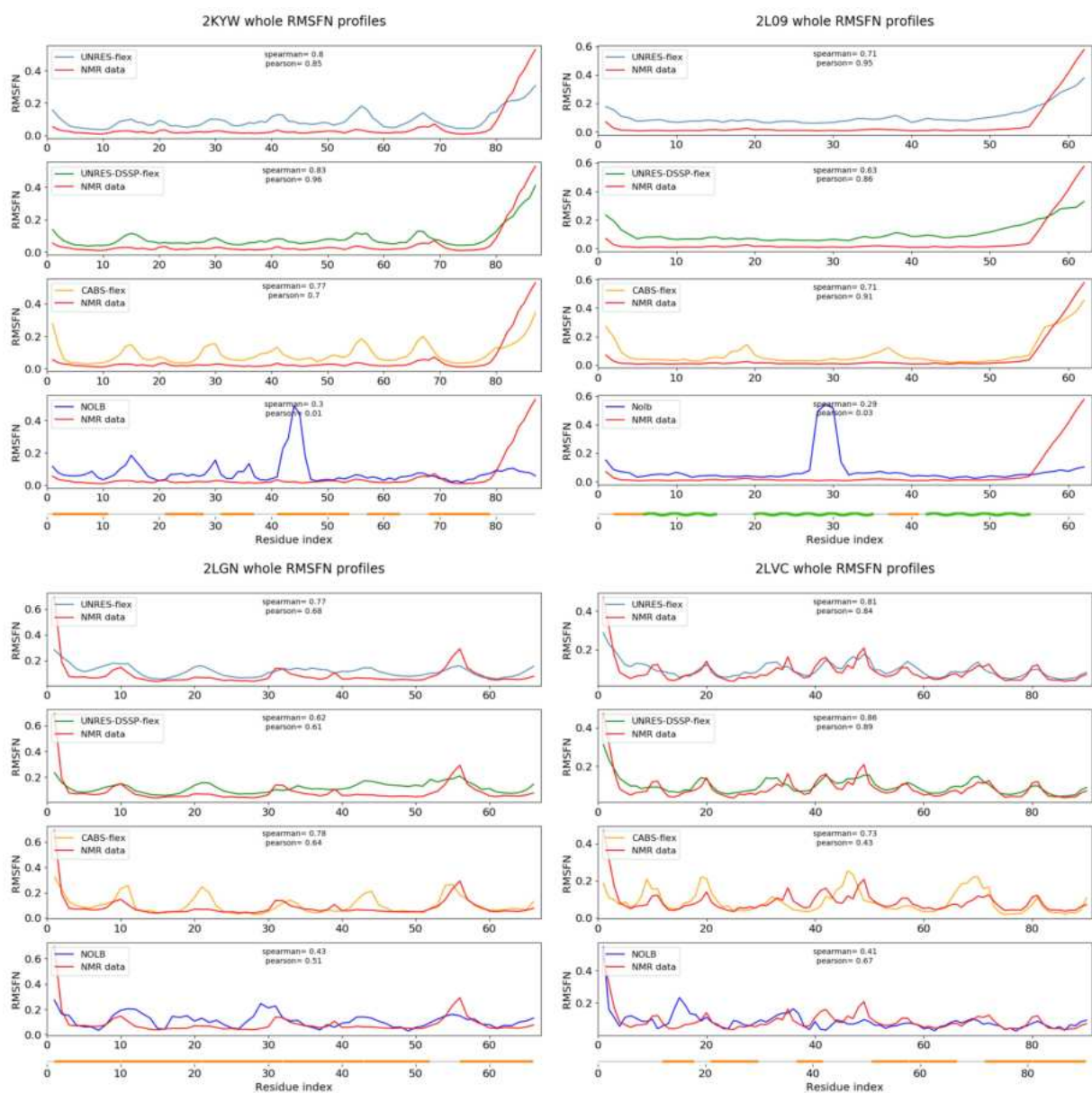


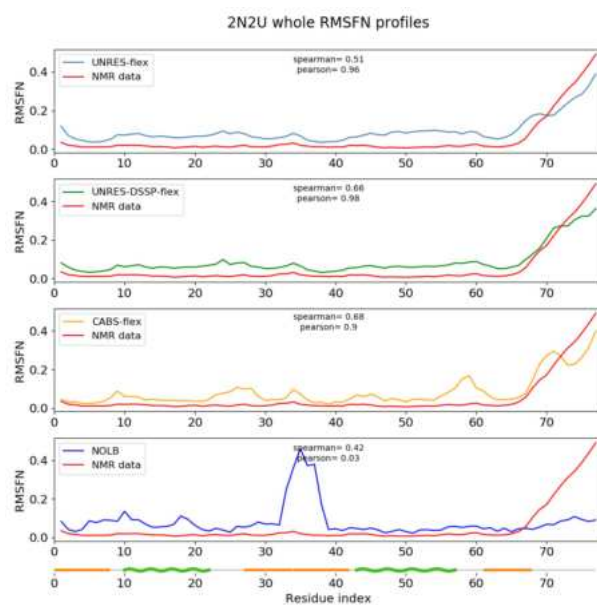
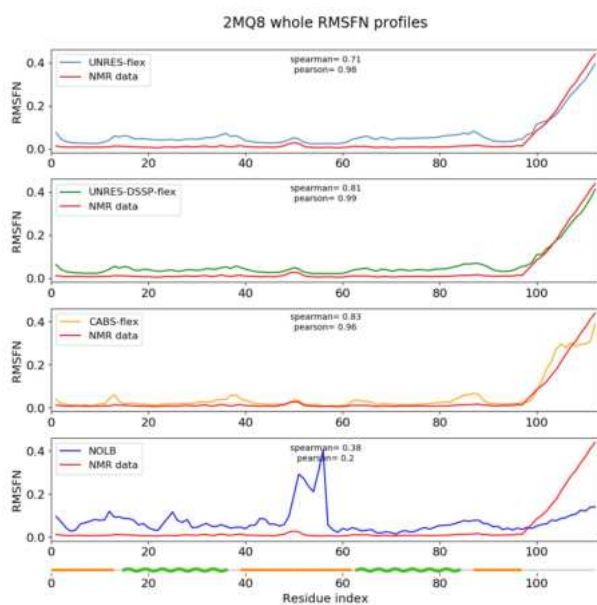
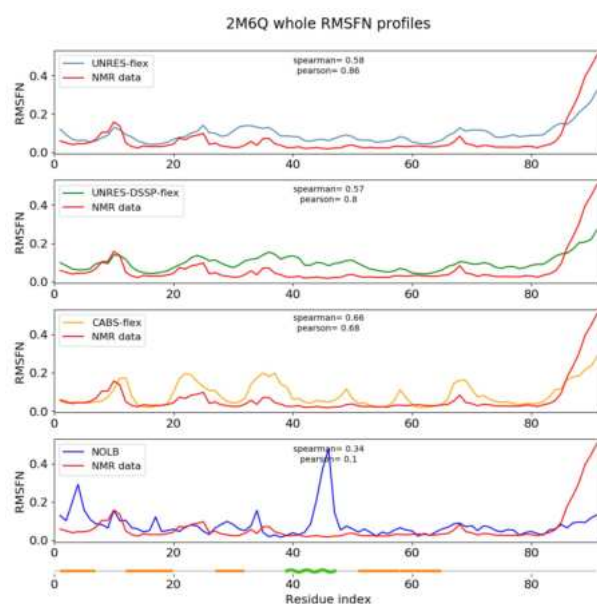
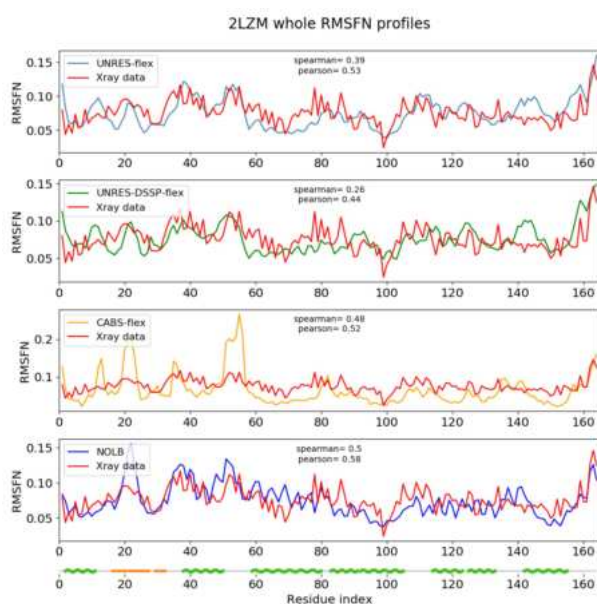


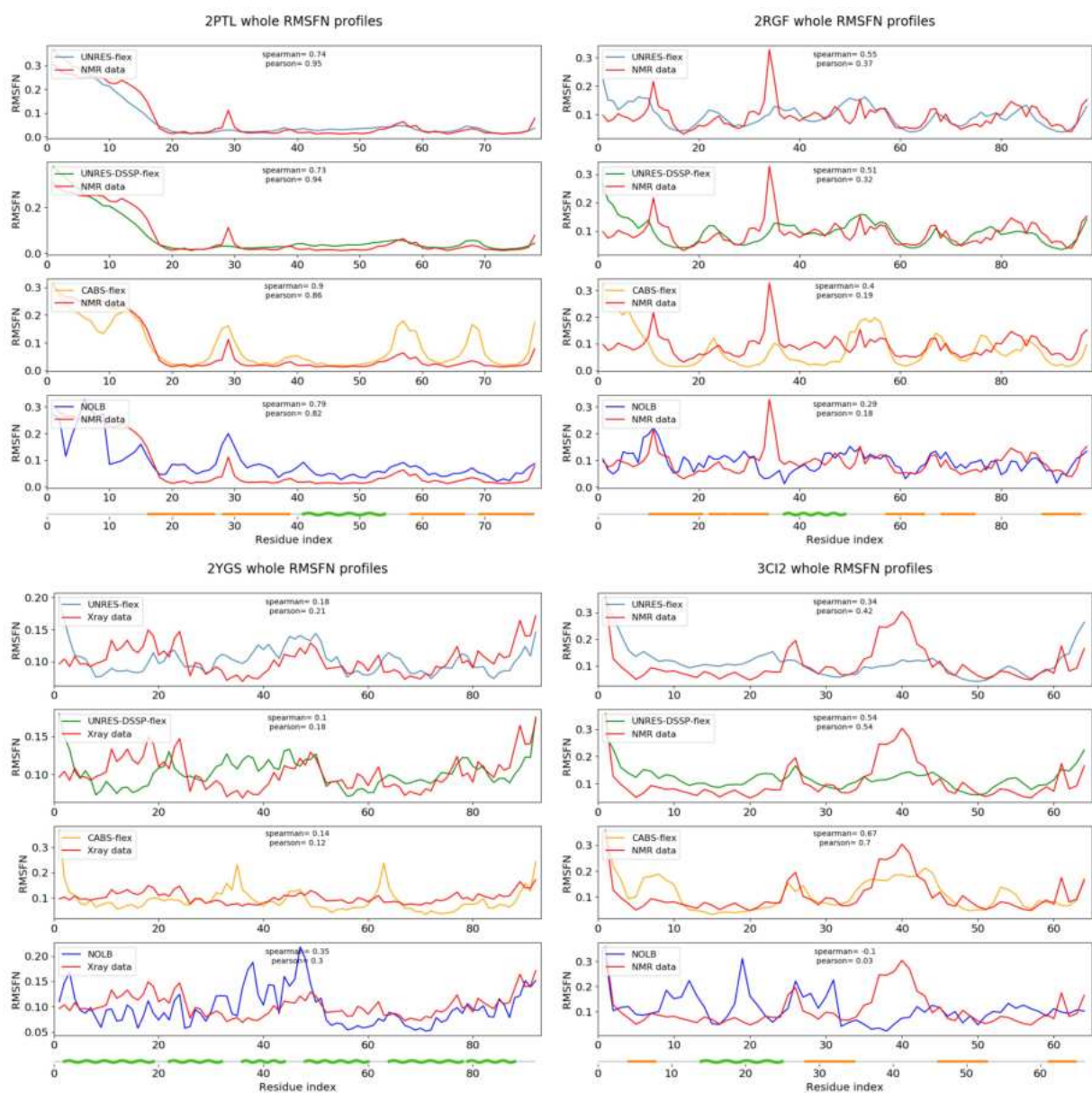


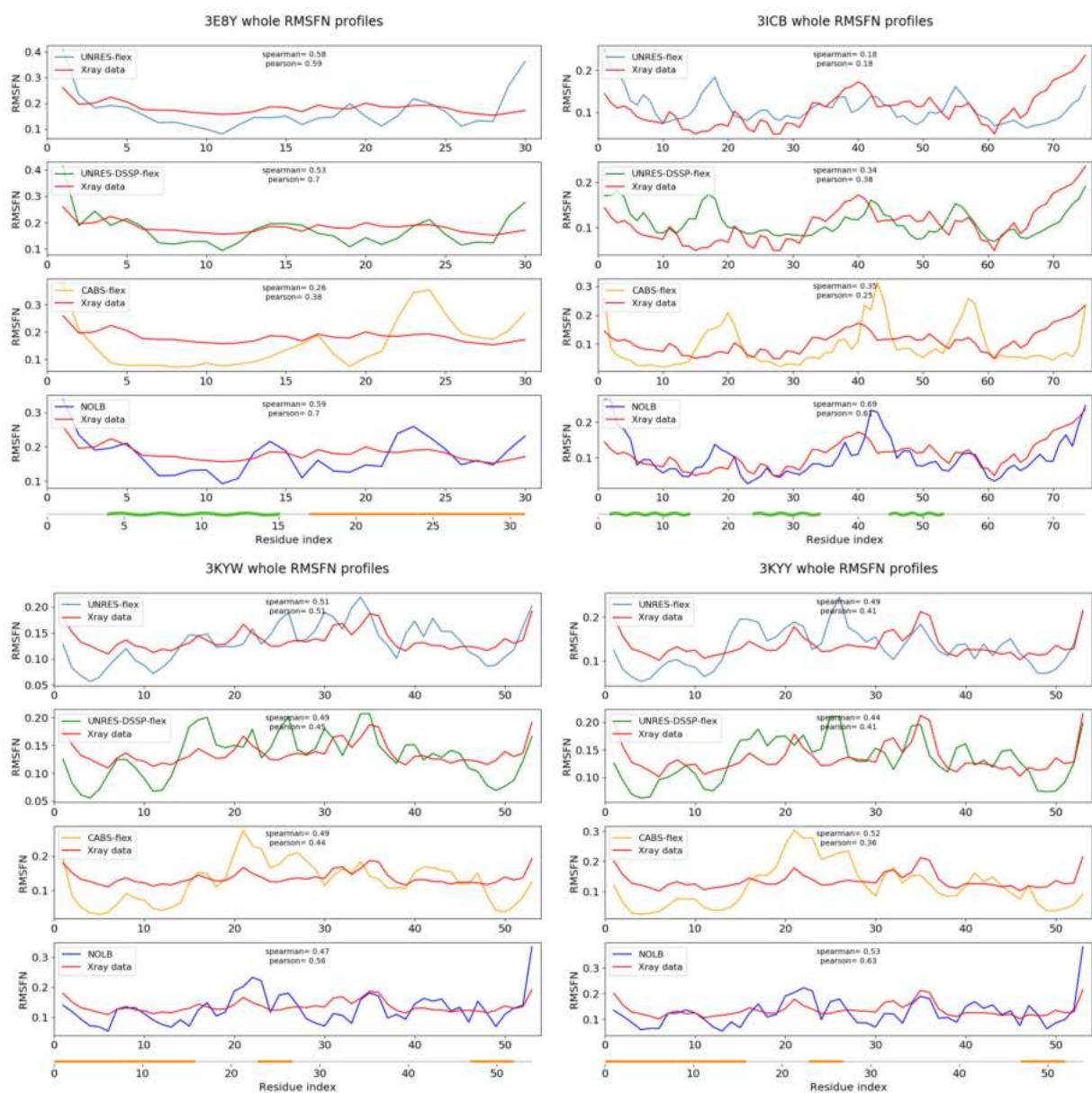


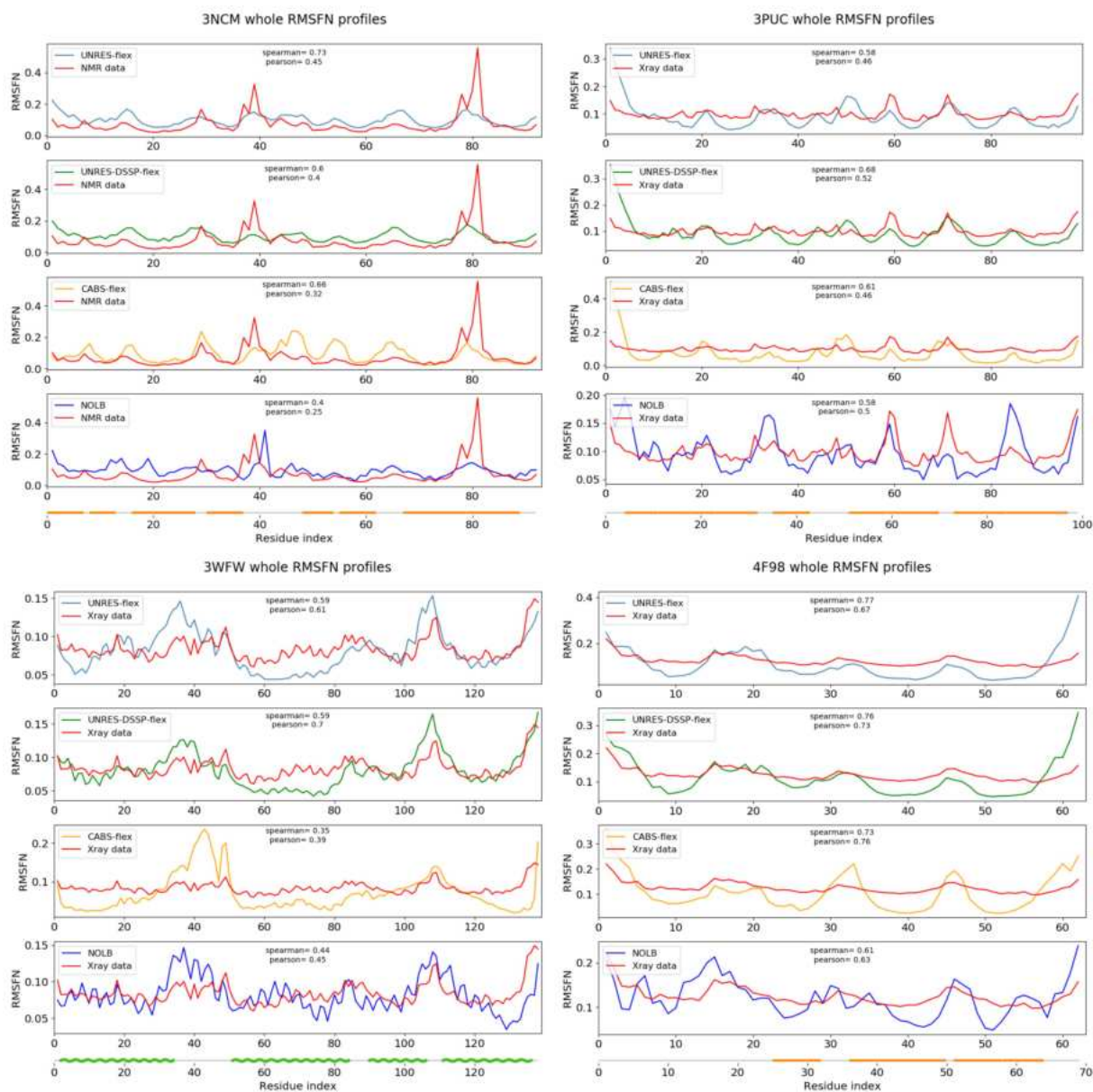


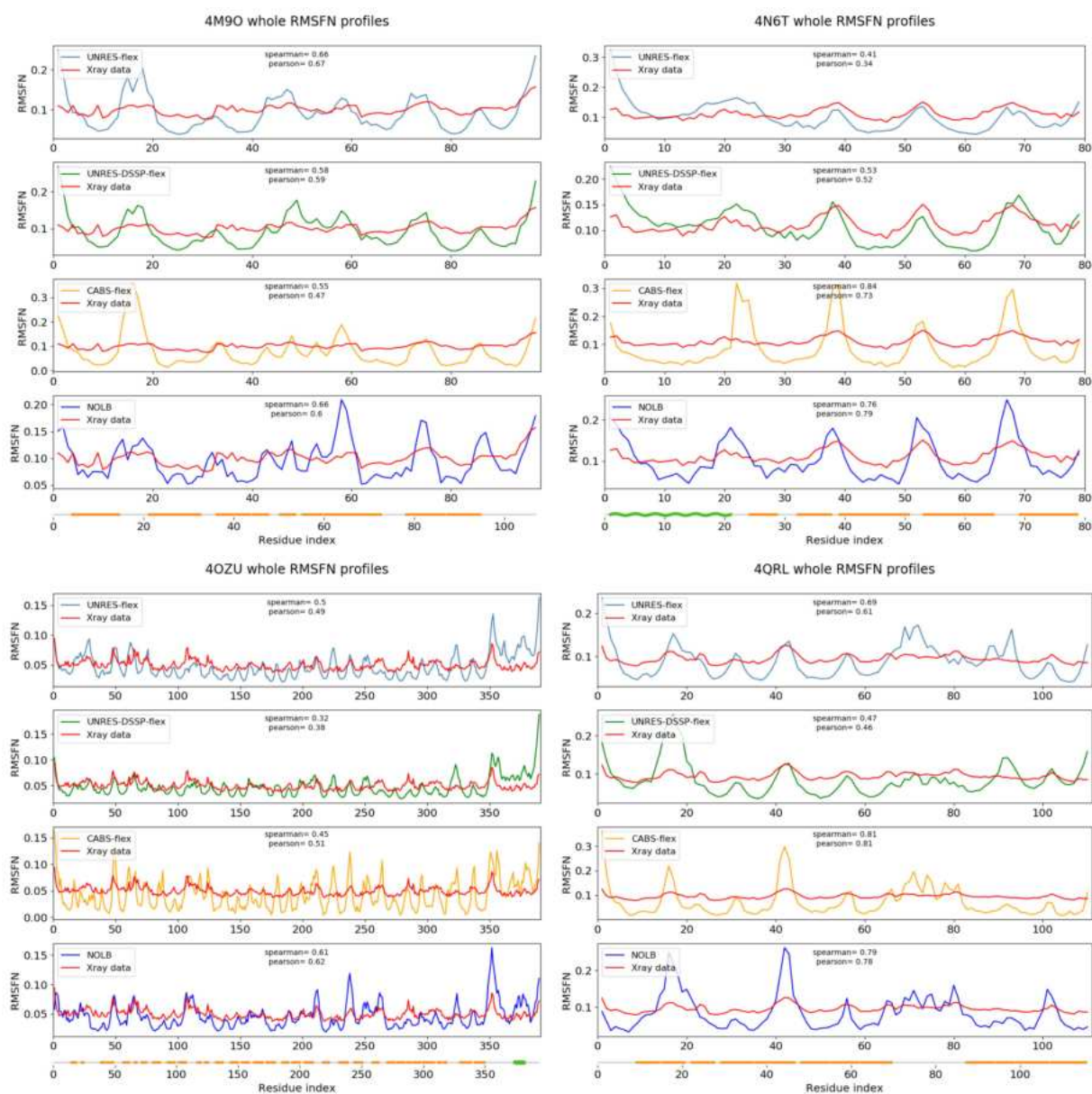


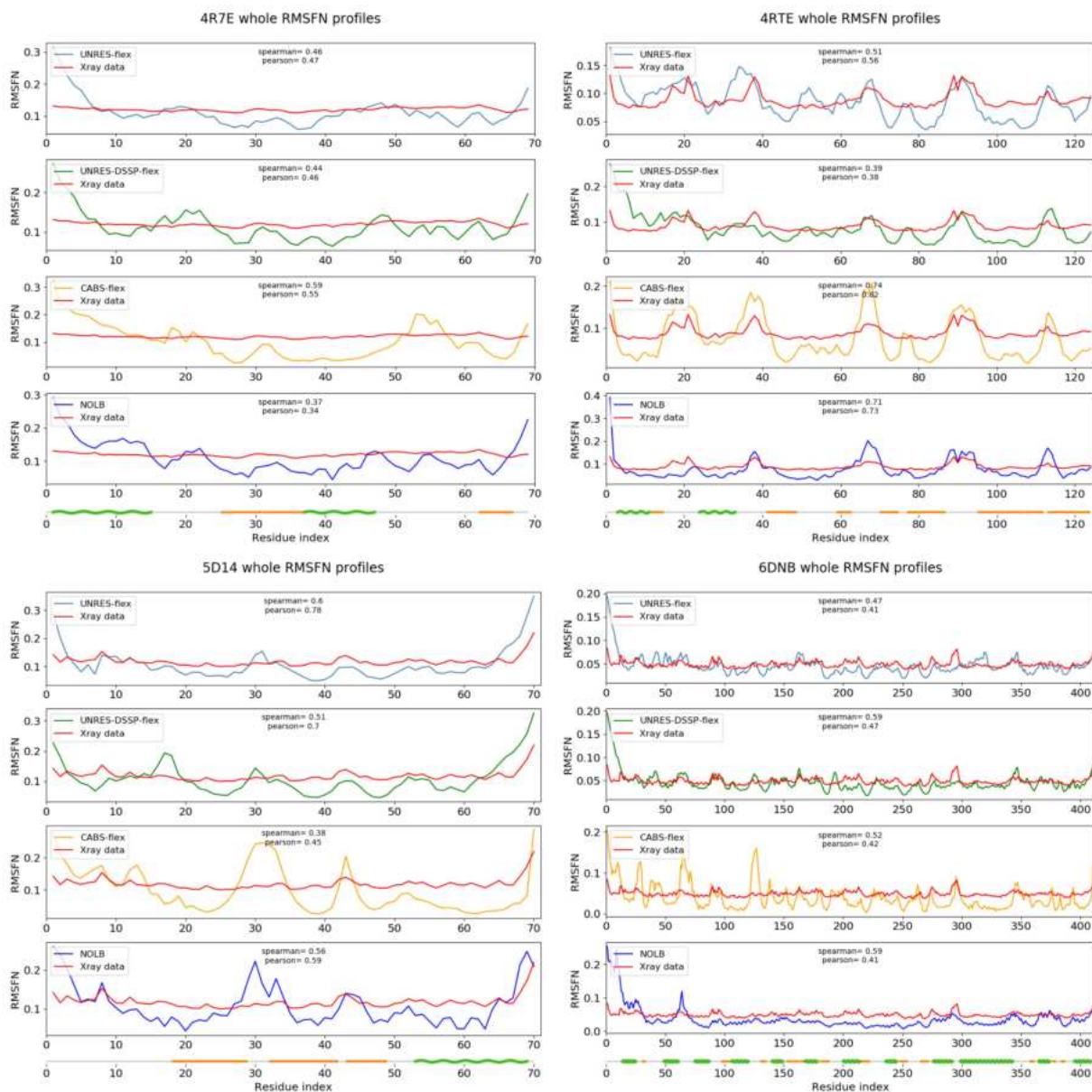












Rysunek 7.2: Profile fluktuacji po normalizacji (RMSFN) obliczone na podstawie zespołów NMR lub współczynników B promieniowania rentgenowskiego (czerwona linia) i odpowiadające im profile fluktuacji przewidziane za pomocą metod: UNRES-flex (jasnoniebieska linia), UNRES-DSSP-flex (zielona linia), CABS-flex (linia żółta) oraz NOLB (linia niebieska). Każdy profil fluktuacji został uśredniony na podstawie trzech symulacji, z wyjątkiem NOLB. Dla każdego profilu fluktuacji został określony współczynniki korelacji Pearson (r_p) i Spearman (r_s), które określają zależności pomiędzy profilami fluktuacji przewidywanych (teoretycznymi) a eksperymentalnymi. Ciągłe pomarańczowe i falowane zielone linie w dolnej części informują o występujących strukturach drugorzędowych: α -helisy lub/i β -harmonijki. Wykresy dotyczą pełnych struktur białek. "Reprinted with permission (CC-BY 4.0.) from Ł. J. Dziadek, A. K. Sieradzan, C. Czaplewski, M. Zalewski, F. Banaś, M. Toczek, W. Nisterenko, S. Grudinin, A. Liwo, A. Gieldoń, Assessment of Four Theoretical Approaches to Predict Protein Flexibility in the Crystal Phase and Solution, J. Chem. Theory Comput. 2024, 20, 17, 7667–7681 Copyright 2024 Journal of Chemical Theory and Computation."

Tabela 7.1: Zbiór średnich wartości współczynników korelacji Pearsona (r_p) i Spearmana (r_s) obliczonych pomiędzy profilem eksperymentalnym RMSFN a profilami RMSFN przewidzianymi metodami: UNRES-flex, UNRES-DSSP-flex, CABS-flex oraz NOLB dla pierwszych, drugich i trzecich symulacji oraz średnich z symulacji (dla metody NOLB tylko dla jednej symulacji) dla pełnej struktury białka o kodzie PDB: 2CRB.

	1 symulacja		2 symulacja		3 symulacja		średnia z symulacji	
	r_p	r_s	r_p	r_s	r_p	r_s	r_p	r_s
UNRES-flex	0,93	0,61	0,94	0,58	0,85	0,56	0,94	0,58
UNRES-DSSP-flex	0,89	0,71	0,95	0,62	0,95	0,68	0,95	0,62
CABS-flex	0,94	0,87	0,96	0,78	0,87	0,75	0,94	0,81
NOLB	0,56	0,53	-	-	-	-	-	-

Tabela 7.2: Zbiór średnich wartości współczynników korelacji Pearsona (r_p) i Spearmana (r_s) obliczonych pomiędzy profilem eksperymentalnym RMSFN a profilami RMSFN przewidzianymi metodami: UNRES-flex, UNRES-DSSP-flex, CABS-flex oraz NOLB dla pierwszych, drugich i trzecich symulacji oraz średnich z symulacji (dla metody NOLB tylko dla jednej symulacji) dla skróconej struktury białka o kodzie PDB: 2CRB.

	1 symulacja		2 symulacja		3 symulacja		średnia z symulacji	
	r_p	r_s	r_p	r_s	r_p	r_s	r_p	r_s
UNRES-flex	0,43	0,60	0,34	0,27	0,35	0,43	0,44	0,56
UNRES-DSSP-flex	0,47	0,53	0,27	0,21	0,52	0,51	0,47	0,52
CABS-flex	0,74	0,68	0,60	0,43	0,53	0,33	0,66	0,50
NOLB	0,39	0,49	-	-	-	-	-	-

Tabela 7.3: Lista badanych struktur białek z bazy PDB (PDB ID) z odpowiednią liczbą reszt aminokwasowych dla pełnej struktury białka (res.), liczbą reszt aminokwasowych po obcięciu końcowych reszt (res.o) i zakresami numeru reszt po obcięciu (zak.o). Białka zostały pogrupowane na 3 klasy ze względu na występującą strukturę drugorzędową. Kody PDB zapisane pogrubioną czcionką odpowiadają strukturom spektroskopii magnetycznego rezonansu jądowego, zaś kody zapisane zwykłą czcionką odpowiadają strukturom rentgenografii strukturalnej.

[h]											
α				β				$\alpha+\beta$			
PDB ID	res.	res.o	zak.o	PDB ID ^a	res.	res.o	zak.o	PDB ID ^a	res.	res.o	zak.o
1A6S	87	87	1-87	1AH9	71	63	5-67	1A80	277	276	2-277
1ACP	77	71	3-73	1BK2	57	55	2-56	1ADG	374	373	2-374
1ALA	316	316	1-316	1CSP	67	67	1-67	1AKY	218	216	2-217
1AUM	70	70	1-70	1ED7	45	45	1-45	1CFJ	532	530	1-530
1BAL	51	33	116-48	1HRF	67	45	7-51	1CLB	75	72	2-73
1ENH	54	48	4-51	1IYV	79	71	3-73	1CTF	68	68	1-68
1EOO	77	73	2-73	1MJC	69	69	1-69	1E0G	48	46	2-47
1FEX	59	54	5-58	1NOA	113	113	1-113	1EM7	56	56	1-56
1GAB	53	44	9-52	1RUW	69	69	1-69	1G6E	87	86	2-87
1HNS	47	39	6-44	1TEN	89	89	1-89	1GHH	81	80	1-79
1HYP	75	73	1-73	1TPN	50	43	5-47	1HYW	58	52	3-52
1IYR	83	72	4-75	1WIU	93	93	1-93	1IG5	75	75	1-75
1J7O	76	71	5-75	1WKT	88	85	4-88	1IQO	88	87	2-88
1K40	126	124	3-126	1XCD	305	305	1-305	1K8B	52	51	1-51
1L2Y	20	18	2-19	2K9G	73	62	9-70	1LEB	72	66	4-69
1LQ7	67	65	3-67	2KYW	87	78	2-79	1OGQ	313	310	3-312
1P68	102	98	4-101	2LGN	66	65	2-66	1OIX	168	168	1-168
1POU	71	66	5-70	2LVC	90	87	4-90	1OPD	85	81	1-81
1PRV	56	36	5-40	3PUC	98	97	1-97	1PGA	56	56	1-56
1RIJ	23	17	5-21	4F98	62	60	3-62	1PHT	83	83	1-83
1ROP	56	56	1-56	4M9O	97	95	1-95	1PTF	87	86	1-86
1YRF	35	34	1-34					1QRE	210	208	3-210
2BF9	36	34	1-34					1STU	68	66	1-66
2CRB	97	69	11-79					1THX	108	104	3-106
2E7N	117	108	10-117					1TIG	88	87	1-87
2HEP	42	23	8-30					1UBQ	76	71	1-71
2HI3	73	47	9-55					1VIG	71	67	4-70
2YGS	92	91	1-91					2ACY	98	93	5-97
3ICB	75	74	1-75					2BBY	69	62	5-66
3WFW	138	134	1-134					2FMR	65	64	2-65
								2L09	62	48	3-50
								2LZM	164	162	1-162
								2M6Q	91	82	2-83
								2MQ8	112	97	1-97
								2N2U	77	63	3-65

α				β				$\alpha+\beta$			
PDB ID	res.	res. _o	zak. _o	PDB ID ^a	res.	res. _o	zak. _o	PDB ID ^a	res.	res. _o	zak. _o
								2PTL	78	59	19-77
								2RGF	97	97	1-97
								3CI2	64	63	2-64
								3E8Y	30	29	1-29
								3KYW	54	53	1-53
								3KYY	54	53	1-53
								3NCM	92	92	1-92
								4N6T	79	79	1-79
								4OZU	390	389	2-390
								4QRL	110	110	1-110
								4R7E	69	69	1-69
								4RTE	124	123	2-124
								5D14	70	68	1-68
								6DNB	411	409	2-410

Tabela 7.4: Zbiór wartości Współczynników korelacji Pearsona (r_p) i Spearmana (r_s) obliczonych na podstawie profili RMSFN przewidzianych dla obciążonych struktur białek dla współczynników B (dla struktur uzyskanych rentgenografią strukturalną) lub zespołów NMR (dla struktur uzyskanych spektroskopią magnetycznego rezonansu jądowego) oraz profili RMSFN przewidzianych za pomocą metod: UNRES-flex, UNRES-DSSP-flex, CABS-flex oraz NOLB.

PDB ID	UNRES-flex		UNRES-DSSP-flex		CABS-flex		NOLB	
	r_s	r_p	r_s	r_p	r_s	r_p	r_s	r_p
1A6S	0.52	0.34	0.59	0.56	0.75	0.81	0.17	-0.08
1A80	0.42	0.40	0.49	0.44	0.59	0.50	0.67	0.66
1ACP	0.22	0.20	0.46	0.35	0.42	0.38	0.22	0.22
1ADG	0.31	0.30	0.36	0.31	0.45	0.41	0.45	0.40
1AH9	0.71	0.65	0.74	0.72	0.79	0.68	0.21	0.21
1AKY	0.40	0.38	0.44	0.58	0.52	0.73	0.71	0.73
1ALA	0.47	0.43	0.58	0.54	0.52	0.49	0.55	0.48
1AUM	0.40	0.24	0.52	0.43	0.28	0.11	0.44	0.46
1BAL	0.27	0.26	0.32	0.27	0.05	-0.08	0.08	-0.05
1BK2	0.36	0.22	0.14	0.05	0.37	0.22	0.47	0.54
1CFJ	0.53	0.53	0.56	0.59	0.37	0.40	0.68	0.67
1CLB	0.33	0.30	0.56	0.58	0.81	0.88	0.09	-0.08
1CSP	0.41	0.49	0.50	0.51	0.50	0.45	0.28	0.28
1CTF	0.29	0.27	0.38	0.36	0.10	0.05	0.41	0.40
1E0G	0.42	0.62	0.58	0.66	0.61	0.73	0.28	0.14
1ED7	0.49	0.45	0.59	0.53	0.33	0.22	0.00	-0.02
1EM7	0.31	0.28	0.46	0.42	0.52	0.53	0.69	0.61
1ENH	0.52	0.62	0.39	0.51	0.48	0.51	0.35	0.34
1EO0	0.29	0.36	0.33	0.41	0.31	0.37	0.30	0.34
1FEX	0.50	0.39	0.64	0.46	0.54	0.40	-0.13	-0.21
1G6E	0.59	0.43	0.69	0.53	0.46	0.41	0.48	0.33
1GAB	0.20	0.16	0.22	0.18	0.39	0.38	0.57	0.46
1GHH	0.28	0.35	0.40	0.52	0.21	0.36	0.07	0.09
1HNS	0.53	0.49	0.67	0.66	0.81	0.81	0.16	0.16
1HRF	-0.18	-0.12	-0.18	-0.12	0.25	0.43	0.45	0.45
1HYP	0.72	0.62	0.62	0.54	0.69	0.54	0.60	0.57
1HYW	0.60	0.57	0.58	0.56	0.64	0.60	-0.03	-0.02
1IG5	0.77	0.62	0.78	0.75	0.75	0.61	0.76	0.74
1IQO	0.41	0.32	0.56	0.44	0.62	0.50	0.18	0.06
1IYR	0.21	0.53	0.52	0.61	0.69	0.74	0.44	0.07
1IYV	0.47	0.42	0.47	0.42	0.63	0.67	0.12	0.33
1J7O	0.17	0.11	0.09	0.05	-0.37	-0.29	0.23	0.04
1K40	0.78	0.78	0.82	0.85	0.68	0.64	0.55	0.55
1K8B	0.34	0.31	0.46	0.43	0.16	0.14	0.17	-0.01
1L2Y	0.56	0.57	0.72	0.62	0.70	0.54	-0.23	-0.14
1LEB	0.22	0.23	0.35	0.32	0.86	0.82	-0.05	-0.12
1LQ7	0.66	0.66	0.67	0.60	0.69	0.81	0.67	0.49
1MJC	0.76	0.80	0.63	0.69	0.58	0.67	0.60	0.65

PDB ID	UNRES-flex		UNRES-DSSP-flex		CABS-flex		NOLB	
	r_s	r_p	r_s	r_p	r_s	r_p	r_s	r_p
1NOA	0.45	0.43	0.36	0.38	0.42	0.40	0.65	0.66
1OGQ	0.08	0.06	0.20	0.17	0.62	0.59	0.45	0.50
1OIX	0.25	0.25	0.49	0.49	0.21	0.32	0.51	0.54
1OPD	0.55	0.52	0.57	0.62	0.35	0.40	0.43	0.40
1P68	0.61	0.51	0.58	0.56	0.69	0.56	0.47	0.34
1PGA	0.21	0.18	0.16	0.13	0.31	0.14	0.30	0.20
1PHT	0.79	0.77	0.83	0.70	0.72	0.67	0.82	0.78
1POU	0.32	0.24	0.54	0.47	0.49	0.65	-0.22	-0.15
1PRV	0.63	0.53	0.69	0.54	0.49	0.60	0.52	0.60
1PTF	0.29	0.29	0.41	0.42	0.51	0.39	0.52	0.60
1QRE	0.77	0.71	0.74	0.71	0.71	0.76	0.80	0.83
1RIJ	0.46	0.48	0.74	0.78	0.71	0.68	-0.28	-0.19
1ROP	0.53	0.52	0.50	0.59	0.43	0.54	0.49	0.55
1RUW	0.39	0.49	0.35	0.47	0.36	0.62	0.29	0.48
1STU	0.46	0.48	0.40	0.48	0.55	0.88	0.38	0.17
1TEN	0.48	0.48	0.44	0.46	0.56	0.61	0.29	0.35
1THX	0.37	0.32	0.36	0.38	0.58	0.46	0.48	0.45
1TIG	0.40	0.42	0.47	0.48	0.64	0.49	0.72	0.61
1TPN	0.57	0.70	0.76	0.87	0.49	0.58	0.14	0.02
1UBQ	0.49	0.49	0.47	0.40	0.52	0.50	0.65	0.67
1VIG	0.79	0.66	0.27	0.33	0.85	0.86	0.53	0.48
1WIU	0.70	0.48	0.77	0.66	0.67	0.47	0.59	0.45
1WKT	0.66	0.47	0.80	0.59	0.72	0.55	0.48	0.37
1XCD	0.71	0.64	0.74	0.70	0.58	0.49	0.31	0.41
1YRF	0.58	0.69	0.35	0.60	0.58	0.62	0.32	0.45
2ACY	0.33	0.27	0.27	0.29	0.26	0.33	0.44	0.47
2BBY	0.55	0.62	0.75	0.82	0.51	0.76	0.38	0.48
2BF9	0.29	0.21	-0.06	0.02	0.27	0.06	0.34	0.33
2CRB	0.56	0.44	0.52	0.47	0.50	0.66	0.49	0.39
2E7N	0.61	0.57	0.69	0.67	0.72	0.83	0.09	0.06
2FMR	0.63	0.58	0.83	0.78	0.77	0.60	0.42	0.33
2HEP	0.71	0.61	0.51	0.43	0.58	0.73	0.14	0.38
2HI3	0.63	0.45	0.34	0.19	0.64	0.75	-0.14	-0.14
2K9G	0.78	0.80	0.61	0.61	0.71	0.52	0.39	0.31
2KYW	0.75	0.63	0.78	0.69	0.72	0.67	0.19	-0.02
2L09	0.43	0.36	0.29	0.20	0.44	0.68	0.00	-0.18
2LGN	0.72	0.53	0.56	0.57	0.77	0.61	0.37	0.36
2LVC	0.79	0.73	0.85	0.77	0.71	0.49	0.36	0.16
2LZM	0.37	0.46	0.23	0.33	0.46	0.49	0.49	0.55
2M6Q	0.45	0.45	0.43	0.50	0.58	0.55	0.25	0.03
2MQ8	0.59	0.49	0.73	0.62	0.76	0.58	0.28	0.22
2N2U	0.14	0.17	0.43	0.41	0.55	0.48	0.39	0.37
2PTL	0.56	0.43	0.52	0.41	0.72	0.72	0.53	0.68

PDB ID	UNRES-flex		UNRES-DSSP-flex		CABS-flex		NOLB	
	r_s	r_p	r_s	r_p	r_s	r_p	r_s	r_p
2RGF	0.55	0.37	0.51	0.32	0.40	0.19	0.29	0.18
2YGS	0.16	0.15	0.07	0.05	0.11	0.01	0.33	0.27
3CI2	0.32	0.23	0.52	0.40	0.65	0.62	-0.16	-0.24
3E8Y	0.54	0.30	0.48	0.43	0.18	0.12	0.55	0.52
3ICB	0.15	0.13	0.31	0.30	0.33	0.19	0.68	0.56
3KYW	0.51	0.51	0.49	0.45	0.49	0.44	0.47	0.56
3KYY	0.49	0.41	0.44	0.41	0.52	0.36	0.53	0.63
3NCM	0.71	0.44	0.57	0.39	0.68	0.34	0.39	0.24
3PUC	0.57	0.47	0.67	0.53	0.60	0.46	0.57	0.47
3WFW	0.55	0.62	0.55	0.64	0.38	0.50	0.43	0.54
4F98	0.75	0.66	0.73	0.68	0.71	0.64	0.58	0.62
4M9O	0.64	0.60	0.55	0.51	0.52	0.43	0.64	0.55
4N6T	0.41	0.34	0.53	0.52	0.84	0.73	0.76	0.79
4OZU	0.49	0.48	0.31	0.36	0.44	0.48	0.61	0.63
4QRL	0.69	0.61	0.47	0.46	0.81	0.81	0.79	0.78
4R7E	0.46	0.47	0.44	0.46	0.59	0.55	0.37	0.34
4RTE	0.49	0.52	0.38	0.30	0.74	0.81	0.71	0.74
5D14	0.56	0.56	0.46	0.44	0.35	0.40	0.52	0.53
6DNB	0.46	0.34	0.58	0.41	0.51	0.37	0.58	0.36
Średnia:	0.48	0.44	0.44	0.48	0.53	0.51	0.38	0.35

Tabela 7.5: Współczynniki korelacji Pearsona (r_p) i Spearmana (r_s) obliczone na podstawie profili RMSFN przewidzianych dla pełnych struktur białek dla współczynników B (dla struktur uzyskanych metodą X-ray) lub zespołów NMR (dla struktur uzyskanych metodą NMR) oraz profili RMSFN przewidzianych za pomocą metod: UNRES-flex, UNRES-DSSP-flex, CABS-flex oraz NOLB.

PDB ID	UNRES-flex		UNRES-DSSP-flex		CABS-flex		NOLB	
	r_s	r_p	r_s	r_p	r_s	r_p	r_s	r_p
1A6S	0.52	0.34	0.59	0.56	0.75	0.81	0.17	-0.08
1A80	0.42	0.39	0.49	0.42	0.59	0.51	0.67	0.64
1ACP	0.38	0.59	0.55	0.67	0.36	0.30	0.25	0.18
1ADG	0.32	0.33	0.37	0.34	0.45	0.41	0.46	0.41
1AH9	0.77	0.84	0.79	0.82	0.83	0.79	0.19	0.31
1AKY	0.42	0.42	0.45	0.62	0.53	0.76	0.71	0.72
1ALA	0.47	0.43	0.58	0.54	0.52	0.49	0.55	0.48
1AUM	0.40	0.24	0.52	0.43	0.28	0.11	0.44	0.46
1BAL	0.71	0.94	0.77	0.96	0.47	0.68	0.73	0.81
1BK2	0.42	0.39	0.21	0.20	0.39	0.21	0.52	0.60
1CFJ	0.54	0.56	0.56	0.61	0.38	0.41	0.69	0.69
1CLB	0.40	0.46	0.60	0.64	0.82	0.87	0.12	-0.07
1CSP	0.41	0.49	0.50	0.51	0.50	0.45	0.28	0.28
1CTF	0.29	0.27	0.38	0.36	0.10	0.05	0.41	0.40
1E0G	0.49	0.78	0.63	0.77	0.65	0.84	0.33	0.25
1ED7	0.49	0.45	0.59	0.53	0.33	0.22	0.00	-0.02
1EM7	0.31	0.28	0.46	0.42	0.52	0.53	0.69	0.61
1ENH	0.66	0.91	0.57	0.89	0.62	0.83	0.54	0.76
1EO0	0.40	0.59	0.43	0.66	0.36	0.31	0.38	0.39
1FEX	0.62	0.88	0.73	0.88	0.64	0.79	0.07	0.57
1G6E	0.60	0.49	0.70	0.57	0.47	0.44	0.49	0.40
1GAB	0.52	0.88	0.54	0.92	0.64	0.95	0.71	0.84
1GHH	0.30	0.52	0.42	0.67	0.24	0.57	0.10	0.16
1HNS	0.72	0.82	0.81	0.93	0.86	0.84	0.41	0.62
1HRF	0.55	0.87	0.55	0.87	0.69	0.79	0.39	0.05
1HYP	0.74	0.72	0.65	0.68	0.71	0.63	0.61	0.54
1HYW	0.71	0.76	0.70	0.82	0.73	0.81	0.19	0.46
1IG5	0.77	0.62	0.78	0.75	0.75	0.61	0.76	0.74
1IQO	0.43	0.41	0.57	0.50	0.63	0.53	0.21	0.29
1IYR	0.47	0.79	0.68	0.91	0.79	0.91	0.53	0.22
1IYV	0.55	0.80	0.55	0.80	0.65	0.86	0.24	0.07
1J7O	0.33	0.79	0.25	0.83	-0.15	0.52	0.37	0.67
1K40	0.79	0.82	0.82	0.85	0.68	0.59	0.57	0.55
1K8B	0.37	0.41	0.49	0.65	0.20	0.37	0.19	0.08
1L2Y	0.68	0.84	0.77	0.60	0.76	0.58	-0.12	-0.01
1LEB	0.51	0.85	0.62	0.88	0.89	0.86	0.07	0.48
1LQ7	0.69	0.77	0.70	0.78	0.72	0.72	0.69	0.92
1MJC	0.76	0.80	0.63	0.69	0.58	0.67	0.60	0.65

PDB ID	UNRES-flex		UNRES-DSSP-flex		CABS-flex		NOLB	
	r_s	r_p	r_s	r_p	r_s	r_p	r_s	r_p
1NOA	0.45	0.43	0.36	0.38	0.42	0.40	0.65	0.66
1OGQ	0.08	0.06	0.19	0.16	0.63	0.62	0.46	0.56
1OIX	0.25	0.25	0.49	0.49	0.21	0.32	0.51	0.54
1OPD	0.60	0.60	0.62	0.70	0.35	0.32	0.45	0.33
1P68	0.65	0.67	0.63	0.72	0.72	0.64	0.48	0.42
1PGA	0.21	0.18	0.16	0.13	0.31	0.14	0.30	0.20
1PHT	0.79	0.77	0.83	0.70	0.72	0.67	0.82	0.78
1POU	0.43	0.58	0.62	0.67	0.51	0.66	-0.07	0.02
1PRV	0.87	0.77	0.88	0.74	0.67	0.40	-0.24	-0.23
1PTF	0.31	0.34	0.43	0.47	0.53	0.45	0.50	0.54
1QRE	0.78	0.74	0.74	0.74	0.72	0.77	0.81	0.84
1RIJ	0.78	0.90	0.85	0.86	0.58	0.43	-0.02	0.03
1ROP	0.53	0.52	0.50	0.59	0.43	0.54	0.49	0.55
1RUW	0.39	0.49	0.35	0.47	0.36	0.62	0.29	0.48
1STU	0.48	0.49	0.44	0.49	0.58	0.88	0.40	0.18
1TEN	0.48	0.48	0.44	0.46	0.56	0.61	0.29	0.35
1THX	0.43	0.49	0.42	0.59	0.62	0.60	0.52	0.54
1TIG	0.42	0.53	0.48	0.59	0.65	0.60	0.73	0.66
1TPN	0.72	0.91	0.85	0.92	0.65	0.70	0.36	0.62
1UBQ	0.58	0.77	0.53	0.64	0.59	0.75	0.71	0.85
1VIG	0.82	0.88	0.38	0.82	0.87	0.88	0.58	0.78
1WIU	0.70	0.48	0.77	0.66	0.67	0.47	0.59	0.45
1WKT	0.69	0.63	0.82	0.75	0.73	0.54	0.50	0.52
1XCD	0.71	0.64	0.74	0.70	0.58	0.49	0.31	0.41
1YRF	0.61	0.78	0.40	0.73	0.62	0.77	0.35	0.41
2ACY	0.42	0.67	0.38	0.78	0.35	0.63	0.52	0.60
2BBY	0.67	0.72	0.82	0.77	0.56	0.54	0.45	0.60
2BF9	0.40	0.67	0.11	0.65	0.38	0.39	0.42	0.57
2CRB	0.84	0.91	0.82	0.95	0.81	0.94	0.76	0.67
2E7N	0.69	0.81	0.75	0.88	0.78	0.93	0.17	0.72
2FMR	0.64	0.61	0.84	0.80	0.78	0.63	0.44	0.40
2HEP	0.94	0.96	0.91	0.97	0.48	0.76	0.05	0.31
2HI3	0.89	0.93	0.81	0.92	0.84	0.90	0.14	0.06
2K9G	0.86	0.77	0.76	0.74	0.80	0.51	0.53	0.73
2KYW	0.82	0.85	0.84	0.96	0.79	0.70	0.30	0.01
2L09	0.73	0.95	0.67	0.86	0.66	0.91	0.25	0.02
2LGN	0.73	0.66	0.58	0.59	0.78	0.63	0.40	0.50
2LVC	0.81	0.84	0.86	0.88	0.71	0.40	0.42	0.69
2LZM	0.39	0.53	0.26	0.44	0.48	0.52	0.50	0.58
2M6Q	0.58	0.86	0.57	0.80	0.66	0.68	0.34	0.10
2MQ8	0.73	0.98	0.83	0.99	0.84	0.96	0.37	0.20
2N2U	0.51	0.96	0.65	0.98	0.67	0.90	0.41	0.03
2PTL	0.80	0.95	0.78	0.94	0.85	0.85	0.74	0.81

PDB ID	UNRES-flex		UNRES-DSSP-flex		CABS-flex		NOLB	
	r_s	r_p	r_s	r_p	r_s	r_p	r_s	r_p
2RGF	0.55	0.37	0.51	0.32	0.40	0.19	0.29	0.18
2YGS	0.18	0.21	0.10	0.18	0.14	0.12	0.35	0.30
3CI2	0.35	0.45	0.54	0.56	0.67	0.71	-0.11	0.05
3E8Y	0.58	0.59	0.53	0.70	0.26	0.38	0.59	0.70
3ICB	0.18	0.18	0.34	0.38	0.35	0.25	0.69	0.61
3KYW	0.51	0.51	0.49	0.45	0.49	0.44	0.47	0.56
3KYY	0.49	0.41	0.44	0.41	0.52	0.36	0.53	0.63
3NCM	0.71	0.44	0.57	0.39	0.68	0.34	0.39	0.24
3PUC	0.58	0.46	0.68	0.52	0.61	0.46	0.58	0.50
3WFW	0.59	0.61	0.59	0.70	0.35	0.39	0.44	0.45
4F98	0.77	0.67	0.76	0.73	0.73	0.76	0.61	0.63
4M9O	0.66	0.67	0.58	0.59	0.55	0.47	0.66	0.60
4N6T	0.41	0.34	0.53	0.52	0.84	0.73	0.76	0.79
4OZU	0.50	0.49	0.32	0.38	0.45	0.51	0.61	0.62
4QRL	0.69	0.61	0.47	0.46	0.81	0.81	0.79	0.78
4R7E	0.46	0.47	0.44	0.46	0.59	0.55	0.37	0.34
4RTE	0.51	0.56	0.39	0.38	0.74	0.82	0.71	0.73
5D14	0.60	0.78	0.51	0.70	0.38	0.45	0.56	0.59
6DNB	0.47	0.41	0.59	0.47	0.52	0.42	0.59	0.41
Średnia:	0.56	0.62	0.58	0.65	0.58	0.59	0.43	0.45

Tabela 7.6: Średnie współczynniki korelacji Pearsona (r_p) i Spearmana (r_s) pomiędzy profilami fluktuacji uzyskanymi za pomocą metod eksperymentalnych (X-ray lub NMR) oraz profilami fluktuacji za pomocą metod doświadczalnych (UNRES-flex, UNRES-DSSP-flex, CABS-flex oraz NOLB) obliczonymi dla obciętych struktur białek (wartości w nawiasach oznaczają odchylenia standardowe).

A: Struktury uzyskane metodą NMR								
Struktury drugorzędowe								
	α		β		$\alpha + \beta$		wszystkie	
	r_p	r_s	r_p	r_s	r_p	r_s	r_p	r_s
UNRES-flex	0.410 (0.153)	0.449 (0.172)	0.522 (0.236)	0.586 (0.264)	0.427 (0.138)	0.474 (0.163)	0.441 (0.176)	0.489 (0.200)
UNRES-DSSP-flex	0.466 (0.181)	0.516 (0.174)	0.574 (0.248)	0.613 (0.275)	0.488 (0.154)	0.525 (0.151)	0.498 (0.194)	0.541 (0.197)
CABS-flex	0.524 (0.300)	0.498 (0.279)	0.536 (0.129)	0.618 (0.173)	0.608 (0.191)	0.614 (0.163)	0.558 (0.235)	0.568 (0.226)
NOLB	0.128 (0.249)	0.185 (0.277)	0.238 (0.172)	0.300 (0.171)	0.166 (0.237)	0.247 (0.202)	0.167 (0.233)	0.234 (0.234)

B: Struktury uzyskane metodą X-ray								
Struktury drugorzędowe								
	α		β		$\alpha + \beta$		wszystkie	
	r_p	r_s	r_p	r_s	r_p	r_s	r_p	r_s
UNRES-flex	0.465 (0.195)	0.501 (0.169)	0.527 (0.149)	0.553 (0.145)	0.396 (0.159)	0.417 (0.168)	0.447 (0.178)	0.474 (0.173)
UNRES-DSSP-flex	0.473 (0.198)	0.467 (0.202)	0.496 (0.181)	0.512 (0.182)	0.426 (0.148)	0.431 (0.157)	0.457 (0.176)	0.460 (0.182)
CABS-flex	0.392 (0.209)	0.439 (0.165)	0.497 (0.131)	0.519 (0.104)	0.499 (0.189)	0.524 (0.190)	0.460 (0.194)	0.492 (0.171)
NOLB	0.512 (0.130)	0.512 (0.129)	0.502 (0.120)	0.468 (0.150)	0.573 (0.158)	0.578 (0.150)	0.537 (0.145)	0.532 (0.149)

C: Wszystkie struktury								
Struktury drugorzędowe								
α		β		$\alpha + \beta$		wszystkie		
r_p	r_s	r_p	r_s	r_p	r_s	r_p	r_s	
UNRES-flex	0.436 (0.176)	0.474 (0.173)	0.525 (0.200)	0.570 (0.216)	0.410 (0.150)	0.444 (0.169)	0.444 (0.177)	0.482 (0.187)
UNRES-DSSP-flex	0.469 (0.189)	0.493 (0.190)	0.537 (0.222)	0.565 (0.241)	0.455 (0.154)	0.474 (0.161)	0.477 (0.186)	0.500 (0.194)
CABS-flex	0.461 (0.269)	0.470 (0.234)	0.517 (0.131)	0.571 (0.153)	0.550 (0.198)	0.566 (0.183)	0.509 (0.221)	0.530 (0.204)
NOLB	0.310 (0.278)	0.340 (0.273)	0.364 (0.199)	0.380 (0.182)	0.384 (0.284)	0.425 (0.242)	0.352 (0.268)	0.383 (0.246)

Tabela 7.7: Średnie współczynniki korelacji Pearsona (r_p) i Spearmana (r_s) pomiędzy profilami fluktuacji uzyskanymi za pomocą metod eksperymentalnych (X-ray lub NMR) oraz profilami fluktuacji za pomocą metod doświadczalnych (UNRES-flex, UNRES-DSSP-flex, CABS-flex oraz NOLB) obliczonymi dla pełnych struktur białek (wartości w nawiasach oznaczają odchylenia standardowe).

Struktury uzyskane metodą X-ray								
Struktury drugorzędowe								
α		β		$\alpha + \beta$		wszystkie		
r_p	r_s	r_p	r_s	r_p	r_s	r_p	r_s	
UNRES-flex	0.759 (0.171)	0.625 (0.184)	0.735 (0.152)	0.698 (0.116)	0.681 (0.212)	0.580 (0.148)	0.724 (0.188)	0.624 (0.164)
UNRES-DSSP-flex	0.803 (0.130)	0.680 (0.163)	0.774 (0.131)	0.723 (0.122)	0.715 (0.192)	0.623 (0.131)	0.763 (0.161)	0.668 (0.148)
CABS-flex	0.672 (0.217)	0.590 (0.245)	0.600 (0.183)	0.695 (0.129)	0.706 (0.213)	0.666 (0.164)	0.669 (0.212)	0.642 (0.200)
NOLB	0.361 (0.340)	0.282 (0.293)	0.358 (0.272)	0.356 (0.159)	0.292 (0.242)	0.318 (0.192)	0.334 (0.293)	0.312 (0.234)

Struktury uzyskane metodą NMR								
Struktury drugorzędowe								
α		β		$\alpha + \beta$		wszystkie		
r_p	r_s	r_p	r_s	r_p	r_s	r_p	r_s	
UNRES-flex	0.560 (0.210)	0.529 (0.166)	0.552 (0.126)	0.563 (0.142)	0.453 (0.182)	0.434 (0.170)	0.512 (0.190)	0.494 (0.172)
UNRES-DSSP-flex	0.594 (0.175)	0.503 (0.178)	0.524 (0.154)	0.525 (0.173)	0.492 (0.167)	0.449 (0.154)	0.535 (0.174)	0.484 (0.170)
CABS-flex	0.464 (0.196)	0.464 (0.160)	0.513 (0.149)	0.528 (0.106)	0.541 (0.197)	0.537 (0.186)	0.508 (0.191)	0.509 (0.167)
NOLB	0.562 (0.131)	0.537 (0.116)	0.518 (0.126)	0.478 (0.156)	0.589 (0.165)	0.590 (0.147)	0.565 (0.149)	0.548 (0.145)

Wszystkie struktury								
Struktury drugorzędowe								
α		β		$\alpha + \beta$		wszystkie		
r_p	r_s	r_p	r_s	r_p	r_s	r_p	r_s	
UNRES-flex	0.665 (0.215)	0.580 (0.182)	0.648 (0.167)	0.634 (0.146)	0.559 (0.227)	0.502 (0.176)	0.618 (0.217)	0.559 (0.180)
UNRES-DSSP-flex	0.704 (0.185)	0.596 (0.192)	0.655 (0.190)	0.628 (0.178)	0.595 (0.211)	0.530 (0.168)	0.649 (0.203)	0.576 (0.184)
CABS-flex	0.574 (0.232)	0.530 (0.218)	0.559 (0.173)	0.615 (0.145)	0.617 (0.220)	0.597 (0.188)	0.588 (0.217)	0.575 (0.196)
NOLB	0.456 (0.281)	0.403 (0.260)	0.434 (0.230)	0.415 (0.169)	0.451 (0.252)	0.464 (0.217)	0.449 (0.259)	0.430 (0.228)

Tabela 7.8: Wyniki ANOVA przedstawione jako poziomy istotności (wyrażone jako wartości p), współczynników korelacji Pearsona (r_p) i Spearmana (r_s) pomiędzy profilami fluktuacji obliczonymi za pomocą metod teoretycznych: UNRES-flex, UNRES-DSSP-flex, CABS-flex i NOLB oraz odpowiadające im profile fluktuacji za pomocą metod eksperymentalnych dla obciążonych struktur białek. Wartości w kolumnach „metoda” i „struktura drugorzędowa” wskazują istotność dla zależności współczynnika korelacji od zastosowanej metody przewidywania profilu fluktuacji i typu struktury drugorzędowej (odpowiednio α , β lub $\alpha + \beta$). Wartości w kolumnie „interakcja” wskazują na istotność jednoczesnej zależności współczynnika korelacji od metody przewidywania profilu fluktuacji oraz typu struktury drugorzędowej. Wartości p niższe od 0,05 (wskazują istotność statystyczną) zaznaczono czerwoną czcionką, a wyższe wartości (odpowiadające niskiej istotności statystycznej lub jej brakiem) oznaczono zieloną czcionką.

poziom istotności	r_p/r_s	metoda	struktura drugorzędowa	interakcja
	struktury uzyskane metodą NMR			
	r_p	<0.001	0.082	0.82
	r_s	<0.001	0.016	0.94
	struktury uzyskane metodą X-ray			
	r_p	0.040	0.42	0.14
	r_s	0.17	0.60	0.069
	wszystkie struktury			
	r_p	<0.001	0.074	0.37
	r_s	<0.001	0.024	0.19

Tabela 7.9: Istotność statystyczna różnic (wyrażonych jako wartości p) dla współczynników korelacji Pearsona (r_p) i Spearmana (r_s) profili fluktuacji dla obciążonych struktur białek, obliczonych za pomocą metod: UNRES-flex, UNRES-DSSP-flex, CABS-flex i NOLB zostały określone za pomocą testu t dla dwóch próbek w zależności od zastosowanej metody oraz rodzaju struktury drugorzędowej. Wartości p ze znakiem „-” oznaczają, że odpowiednio współczynniki korelacji są większe dla metody w lewej kolumnie (Metoda 1), natomiast znak „+” oznacza, że współczynniki korelacji są większe dla metody w prawej kolumnie (Metoda 2). Wartości p niższe niż 0,05 (wskazujące istotność statystyczną) zaznaczono czerwoną czcionką, a wyższe wartości (odpowiadające słabej istotności statystycznej lub jej brakowi) oznaczono zieloną czcionką.

A: Struktury uzyskane metoda NMR, współczynnik korelacji Pearson (r_p)						
poziom istotności	Metoda 1	Metoda 2	Rodzaj struktury drugorzędowej			
			α	β	$\alpha + \beta$	all
	UNRES-flex	UNRES-DSSP-flex	+ 0.31	+ 0.64	+ 0.22	+ 0.42
		CABS-flex	+ 0.15	+ 0.88	+ 0.0024	+ 0.35
		NOLB	- <0.001	- 0.0060	- <0.001	- <0.001
	UNRES-DSSP-flex	CABS-flex	+ 0.47	- 0.67	+ 0.046	+ 0.36
		NOLB	- <0.001	- 0.0022	- <0.001	- <0.001
	CABS-flex	NOLB	- <0.001	- <0.001	- <0.001	- <0.001

B: Struktury uzyskane metoda NMR, współczynnik korelacji Spearman (r_s)						
poziom istotności	Metoda 1	Metoda 2	Rodzaj struktury drugorzędowej			
			α	β	$\alpha + \beta$	all
	UNRES-flex	UNRES-DSSP-flex	+ 0.31	+ 0.64	+ 0.22	+ 0.42
		CABS-flex	+ 0.15	+ 0.88	+ 0.0024	+ 0.35
		NOLB	- <0.001	- 0.0060	- <0.001	- <0.001
	UNRES-DSSP-flex	CABS-flex	+ 0.47	- 0.67	+ 0.046	+ 0.36
		NOLB	- <0.001	- 0.0022	- <0.001	- <0.001
	CABS-flex	NOLB	- <0.001	- <0.001	- <0.001	- <0.001

C: Struktury uzyskane metoda X-ray, współczynnik korelacji Pearson (r_p)						
poziom istotności	Metoda 1	Metoda 2	Rodzaj struktury drugorzędowej			
			α	β	$\alpha + \beta$	all
	UNRES-flex	UNRES-DSSP-flex	+ 0.91	- 0.70	+ 0.52	+ 0.65
		CABS-flex	- 0.30	- 0.66	+ 0.061	+ 0.24
		NOLB	+ 0.41	- 0.70	+ <0.001	+ 0.0073
	UNRES-DSSP-flex	CABS-flex	- 0.25	- 0.99	+ 0.17	+ 0.35
		NOLB	+ 0.50	- 0.94	+ 0.0035	+ 0.31
	CABS-flex	NOLB	+ 0.052	+ 0.94	+ 0.18	+ 0.001

D: Struktury uzyskane metoda X-ray, współczynnik korelacji Spearman (r_s)						
poziom istotności	Metoda 1	Metoda 2	Rodzaj struktury drugorzędowej			
			α	β	$\alpha + \beta$	all
	UNRES-flex	UNRES-DSSP-flex	- 0.60	- 0.61	+ 0.78	- 0.70
		CABS-flex	- 0.28	- 0.58	+ 0.061	+ 0.22
		NOLB	+ 0.84	- 0.24	+ 0.0021	+ 0.25
	UNRES-DSSP-flex	CABS-flex	- 0.66	+ 0.92	+ 0.092	+ 0.39
		NOLB	+ 0.45	- 0.58	+ 0.0033	+ 0.26
	CABS-flex	NOLB	+ 0.16	- 0.41	+ 0.30	+ 0.29

E: Wszystkie struktury, współczynnik korelacji Pearson (r_p)						
poziom istotności	Metoda 1	Metoda 2	Rodzaj struktury drugorzędowej			
			α	β	$\alpha + \beta$	all
	UNRES-flex	UNRES-DSSP-flex	+ 0.44	+ 0.86	+ 0.19	+ 0.41
		CABS-flex	+ 0.63	- 0.90	+ <0.001	+ 0.38
		NOLB	- 0.023	- 0.015	- 0.61	- 0.31
	UNRES-DSSP-flex	CABS-flex	- 0.89	- 0.74	+ 0.019	+ 0.43
		NOLB	- 0.0053	- 0.013	- 0.17	- 0.10
	CABS-flex	NOLB	- 0.020	- 0.0063	- 0.0034	- 0.0090

F: Wszystkie struktury, współczynnik korelacji Spearman (r_s)						
poziom istotności	Metoda 1	Metoda 2	Rodzaj struktury drugorzędowej			
			α	β	$\alpha + \beta$	all
	UNRES-flex	UNRES-DSSP-flex	+ 0.66	- 0.94	+ 0.40	+ 0.59
		CABS-flex	- 0.94	- 1.0	+ 0.0027	+ 0.49
		NOLB	- 0.014	- 0.0045	- 0.69	- 0.34
	UNRES-DSSP-flex	CABS-flex	- 0.65	+ 0.93	+ 0.021	+ 0.40
		NOLB	- 0.0067	- 0.0091	- 0.28	- 0.14
	CABS-flex	NOLB	- 0.031	- <0.001	- 0.0043	- 0.012

Tabela 7.10: Kody PDB struktur uzyskanych metodą NMR o niskim podobieństwie profili fluktuacji uzyskanych metodami teoretycznymi (UNRES-flex, UNRES-DSSP-flex, CABS-flex oraz NOLB) w porównaniu do metody eksperymentalnych, oszacowanych z zespołów NMR.

metoda	struktura z $r_p \leq 0.2$
UNRES-flex	1ACP, 1BAL, 1GAB, 1HRF, 1J7O, 1LEB, 1POU, 2N2U, 3CI2
UNRES-DSSP-flex	1BAL, 1GAB, 1HRF, 1J7O, 2HI3, 2L09
CABS-flex	1BAL, 1ED7, 1J7O, 1K8B, 2RGF
NOLB	1A6S, 1ACP, 1AH9, 1BAL, 1CLB, 1E0G, 1ED7, 1FEX, 1GHH, 1HNS, 1HYW, 1IQO, 1IYR, 1J7O, 1K8B, 1L2Y, 1LEB, 1POU, 1RIJ, 1STU, 1TPN, 2E7N, 2HI3, 2KYW, 2L09, 2LVC, 2M6Q, 2MQ8, 2RGF, 3CI2, 3NCM

Tabela 7.11: Zbiór średnich wartości współczynników korelacji Pearsona (r_p) i Spearmana (r_s) obliczonych pomiędzy profilem eksperymentalnym RMSFN a profilami RMSFN przewidzianymi metodami: UNRES-flex, UNRES-DSSP-flex, CABS-flex oraz NOLB skróconych strukt białek o kodach PDB: 1A6S, 3ICB oraz 2K9G.

	1A6S		3ICB		2K9G	
	r_p	r_s	r_p	r_s	r_p	r_s
UNRES-flex	0,34	0,52	0,13	0,15	0,80	0,78
UNRES-DSSP-flex	0,56	0,53	0,30	0,31	0,61	0,61
CABS-flex	0,81	0,75	0,19	0,33	0,52	0,71
NOLB	-0,08	0,17	0,56	0,68	0,31	0,39